# SYSTEMATIC PHENOMICS ANALYSIS OF AUTISM RISK GENES AND VARIANTS

by

Troy A. McDiarmid

B.Sc., Carleton University, 2014

# A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

# THE REQUIREMENTS FOR THE DEGREE OF

# DOCTOR OF PHILOSOPHY

in

# THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Neuroscience)

# THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2020

© Troy A. McDiarmid, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

## SYSTEMATIC PHENOMICS ANALYSIS OF AUTISM RISK GENES AND VARIANTS

submitted by	Troy A. McDiarmid	in partial fulfillment of the requirements for
the degree of	Doctor of Philosophy	
in	Neuroscience	

## **Examining Committee:**

Dr. Catharine Rankin, Professor, Department of Psychology, UBC Supervisor

Dr. Kurt Haas, Professor, Department of Cellular and Physiological Sciences, UBC Supervisory Committee Member

Dr. Paul Pavlidis, Professor, Department of Psychiatry & Michael Smith Laboratories, UBC Supervisory Committee Member

Dr. Philip Hieter, Professor, Department of Biochemistry & Michael Smith Laboratories, UBC University Examiner

Dr. Freda Miller, Professor, Department of Medical Genetics & Michael Smith Laboratories University Examiner

## Additional Supervisory Committee Members:

Dr. Kota Mizumoto, Assistant Professor, Department of Zoology Supervisory Committee Member

## Abstract

Two major challenges facing the genetics of Autism Spectrum Disorders (ASD) are the large and growing number of candidate risk genes and gene variants of unknown functional significance. The goals of this dissertation were to combine emerging methods in CRISPR-Cas9 genome engineering with machine vision phenomics to gain insight into the functions of ASD risk genes and the functional impact of specific variants. I developed a pipeline to discover the functions of ASD risk genes by obtaining strains of the genetic model organism Caenorhabditis elegans with inactivating mutations in each gene and observed the phenotypic consequences using machine vision. I quantified 26 phenotypes spanning morphology, locomotion, tactile sensitivity, and learning in >27,000 animals representing 135 genotypes (98 strains with mutations in different genes and 37 strains with additional alleles of a subset of these genes), allowing us to identify disruptions in habituation (a neural circuit's plastic ability to decrease responding to repeated sensory stimuli) as a common impairment. I then clustered genes by similarity in phenomic profiles and used epistasis analysis to discover parallel networks centered on CHD8•chd-7 and NLGN3•nlg-1 that underlie mechanosensory hyper-responsivity and impaired habituation learning. Next, I demonstrated how this database can facilitate experiments that determine the functional consequences of missense variants and whether phenotypic alterations are reversible. Further, I developed a broadly applicable CRSIRP-Cas9 genome editing strategy to replace C. elegans genes with human genes that allows for in vivo analysis of human genetic variation with unprecedented precision. Finally, I contributed to the development of a multi-model system pipeline for high-confidence assessment of missense variants in the ASD risk gene PTEN. This work charts the phenotypic landscape of ASD-associated genes, offers in vivo variant functional assays, and potential therapeutic targets for ASD.

# Lay Summary

Although >100 genes have been implicated in the etiology of autism spectrum disorder (ASD), we need to better understand their functions to reveal mechanisms underlying ASD and develop treatments. In this dissertation, I developed an experimental pipeline to discover functions of ASD-associated genes by inactivating each gene in the genetic model organism *Caenorhabditis elegans* and observing the phenotypic consequences using machine vision. I quantified 26 phenotypes spanning morphology, locomotion, tactile sensitivity, and learning in >27,000 animals representing 135 genotypes (98 strains with mutations in different genes and 37 strains with additional alleles of a subset of these genes), allowing us to identify disruptions in habituation (a neural circuit's plastic ability to decrease responding to repeated sensory stimuli) as a shared impairment. I then demonstrated how this database can facilitate experiments that determine the functional consequences of missense variants and whether phenotypic alterations are reversible.

## Preface

Portions of the introductory chapter are modified from a review article on which I am the primary author: **McDiarmid, T. A.,** Yu, A. J., & Rankin, C. H. (2018). Beyond the response— High throughput behavioral analyses to link genome to phenome in *Caenorhabditis elegans*. *Genes, Brain and Behavior*, 17(3), e12437.

The work in chapter 2, entitled "Systematic phenomics analysis of autism-associated genes reveals parallel networks underlying reversible impairments in habituation" has been published as:

**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. *Proceedings of the National Academy of Sciences*, 117(1), 656-667.

I was the lead investigator for this project. I designed and performed all experiments and data analysis for this manuscript with the following exceptions: Confocal imaging experiments in Figure 2.12 and 2.14 were done with co-author Joseph Liang. Confocal imaging experiments in Figure 2.15 were done by coauthors Ardalan Hendi and Dr. Gregory Mullen under the supervision of coauthors Dr. Kota Mizumoto and Dr. James Rand, respectively. Non-habituation behavioral assays in Figure 2.13 were conducted by coauthors Dr. Eleanor Mathews and Dr. Gregory Mullen under the supervision of coauthor of coauthor Dr. James Rand. The statistical analysis and data visualization pipelines for Figures 2.1-2.7 were generated with coauthors Manuel Belmadani and Dr. Paul Pavlidis. The CRISPR-Cas9 variant lines in Figure 2.12 were generated by coauthor Dr. Wan-Rong Wong. Coauthor Fabian Meili assisted with plasmid construction

under the supervision of coauthor Dr. Kurt Haas. I wrote the manuscript and prepared the

figures. Dr. Catharine Rankin was the supervisory author throughout.

The work in chapter 3, entitled "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" has been adapted from work published as:

**McDiarmid, T. A.,** Au, V., Loewen, A. D., Liang, J., Mizumoto, K., Moerman, D. G., & Rankin, C. H. 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*, 11(12), dmm036517, 2018.

## And:

McDiarmid, T.A., Au, V., Moerman, D.G., & Rankin, C. H. (2020). Peel-1 negative selection promotes screening-free CRISPR-Cas9 genome editing in *Caenorhabditis elegans*. *PLOS ONE*, 15(9), e0238950, 2020.

I was the lead investigator for these projects. I performed experiments and data analysis for these manuscripts with the following exceptions: Genome editing and efficiency analysis in Figures 3.3, 3.5, 3.7 and 3.8 were done with coauthor Vinci Au under the supervision of coauthor Dr. Don Moerman. Coauthors Aaron Loewen and Joseph Liang assisted with behavioral experiments and transgenesis, respectively. The human gene replacement strategy was designed with guidance from Dr. Kota Mizumoto and Dr. Don Moerman. I wrote the manuscript and prepared the figures. Dr. Catharine Rankin was the supervisory author throughout.

## The work in chapter 4, entitled "Multi-model functionalization of disease-associated PTEN

## missense mutations identifies multiple molecular mechanisms underlying protein

## dysfunction" has been published as:

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., Allan, D. W., Pavlidis, P.†, & Haas, K.† Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nature Communications*, 11(11):2073, 2020.

I am a co-first author on this collaborative project. I designed and performed all C.

elegans experiments and contributed to figure preparation, data analysis and interpretation.

# **Table of Contents**

Abstrac	iii		
Lay Su	nmaryiv		
Preface	v		
Table o	f Contents viii		
List of [	List of Tables xiii		
List of ]	Figures xiv		
Acknow	vledgements xvii		
Dedicat	Dedication xix		
Chapte	r 1: Introduction 1		
1.1	Clinical overview of Autism Spectrum Disorders 1		
1.2	Early evidence of genetic contributions to ASD		
1.3	The contribution of <i>de novo</i> coding variants to ASD		
1.4	The emerging genetic architecture of ASD – a spectrum of risk7		
1.5	Insights into the biology of ASD and the quest for convergence		
1.6	Human phenotyping studies		
1.7	Model system investigations of ASD risk gene function		
1.8	The growing challenge of variants of uncertain significance		
1.9	Major challenges facing the genetics of ASD		
1.10	Finding function in novel targets – <i>C. elegans</i> as a model system		
1.11	Scientific motivations		

Chapter	c 2: Systematic phenomics analysis of Autism-associated genes reveals parallel
network	xs underlying reversible impairments in habituation
2.1	Introduction
2.2	Results
2.2	.1 ASD-associated genes are highly conserved to <i>C. elegans</i>
2.2	.2 Quantitative phenotypic profiles identify a set of shared functions among ASD-
ass	ociated genes
2.2	.3 Phenotypic profiles of strains with mutations in ASD-associated genes define
sha	red and unique functions
2.2	A phenomic database of strains with mutations in ASD-associated genes
2.2	.5 Phenomic profiles can be leveraged for <i>in vivo</i> variant functional assays
2.2	.6 CRISPR-Cas9 auxin inducible degradation reveals <i>nlg-1</i> phenotypes are
rev	ersible by adult-specific re-expression
2.3	Discussion
2.3	.1 Phenomic clustering and epistasis to map genetic networks among ASD-
ass	ociated genes in vivo
2.3	.2 Phenomic profiles can be leveraged for <i>in vivo</i> assays of missense variant effect72
2.3	.3 Harnessing phenotypic profiles for tests of adult reversibility
2.3	.4 Conclusions
2.4	Materials and Methods75
2.4	.1 Ortholog identification and strain selection
2.4	.2 Animals
2.4	.3 Microbe strains
	ix

	2.4.4	Behavioral assays	77
	2.4.5	Multi-Worm Tracker behavioral analysis and statistics	79
	2.4.6	Clustering analyses	80
	2.4.7	CRISPR-Cas9 missense variant lines	81
	2.4.8	List of strains generated in this work	82
	2.4.9	Strain and plasmid generation	83
	2.4.10	Genotype confirmation	84
	2.4.11	Auxin administration	85
	2.4.12	Confocal imaging	85
	2.5 Sup	oplemental information	86
Cł	hapter 3: (	CRISPR-Cas9 human gene replacement and phenomic characterization in	
Ca	aenorhabdi	itis elegans to understand the functional conservation of human genes and	
Ca de	<i>tenorhabdi</i> cipher var	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance	89
Ca de	aenorhabdi ecipher var 3.1 Intr	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance	<b> 89</b> 90
Ca de	ecipher var 3.1 Intr 3.2 Res	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance roduction	<b> 89</b> 90 93
Ca de	<b>ccipher var</b> 3.1 Intr 3.2 Res 3.2.1	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen	<b> 89</b> 90 93 e
Ca de	<b>ceipher var</b> 3.1 Intr 3.2 Res 3.2.1 with a si	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen ingle copy of its human ortholog	<b> 89</b> 90 93 e 93
Ca de	ecipher var 3.1 Intr 3.2 Res 3.2.1 with a si 3.2.2	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen ingle copy of its human ortholog <i>PTEN</i> as a prototypic disease-associated gene for targeted human gene	<b>89</b> 90 93 e 93
Ca de	ecipher var 3.1 Intr 3.2 Res 3.2.1 with a si 3.2.2 replacen	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen ingle copy of its human ortholog <i>PTEN</i> as a prototypic disease-associated gene for targeted human gene nent	90 93 e 93
Ca de	ecipher var 3.1 Intr 3.2 Res 3.2.1 with a si 3.2.2 replacen 3.2.3	<i>itis elegans</i> to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen ingle copy of its human ortholog <i>PTEN</i> as a prototypic disease-associated gene for targeted human gene nent An automated chemotaxis paradigm reveals a conserved nervous system role	<b>89</b> 90 93 e 93 97 for
<i>Ca</i>	ecipher var 3.1 Intr 3.2 Res 3.2.1 with a si 3.2.2 replacen 3.2.3 PTEN in	itis elegans to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen ingle copy of its human ortholog <i>PTEN</i> as a prototypic disease-associated gene for targeted human gene nent An automated chemotaxis paradigm reveals a conserved nervous system role n controlling NaCl preference	<b>89</b> 90 93 e 93 93 for 100
<i>Ca</i>	ecipher var 3.1 Intr 3.2 Res 3.2.1 with a si 3.2.2 replacen 3.2.3 PTEN in 3.2.4	it is elegans to understand the functional conservation of human genes and riants of uncertain significance roduction sults A general genome editing strategy for direct replacement of a <i>C. elegans</i> gen ingle copy of its human ortholog <i>PTEN</i> as a prototypic disease-associated gene for targeted human gene nent An automated chemotaxis paradigm reveals a conserved nervous system role n controlling NaCl preference Complete deletion of the <i>daf-18</i> ORF causes strong NaCl avoidance that is no	89 90 93 e 93 for 100 ot

	3.2.5	A streamlined human gene replacement strategy functionally replaces daf-18	with
	human <i>I</i>	PTEN	105
	3.2.6	Peel-1 negative selection promotes screening-free CRISPR-Cas9 genome edit	ing
	in Caene	orhabditis elegans	107
	3.2.7	daf-18 deletion causes mechanosensory hyporesponsivity that is rescued by	
	targeted	replacement with human PTEN	112
	3.2.8	Assessment of PTEN missense variants using CRISPR-Cas9 human gene	
	replacen	nent	115
3	3.3 Dis	cussion	118
	3.3.1	Comparing CRISPR-Cas9 targeted human gene replacement with orthology-	
	based va	ariant assessment methods	120
	3.3.2	Combining human gene replacement and automated phenomic characterizatio	n to
	discover	conserved gene functions and establish variant functional assays	122
	3.3.3	Further applications of targeted human gene replacement	124
3	3.4 Ma	terials and Methods	126
	3.4.1	Strains and culture	126
	3.4.2	Strain and plasmid generation	128
	3.4.3	Peel-1 induction	130
	3.4.4	CRISPR Screening and quantification for Peel-1-DMS	130
	3.4.5	Genotype confirmation	131
	3.4.6	RNA extraction, library preparation, and cDNA amplification	132
	3.4.7	Protein structure modeling and visualization	133
	3.4.8	NaCl chemotaxis behavioral assays	133
			xi

3.4.	.9 Mechanosensory habituation behavioral assays	
3.4.	.10 Multi-worm tracker behavioral analysis and statistics	135
Chapter	r 4: Multi-model functionalization of disease-associated PTEN missense mu	itations
identifie	es multiple molecular mechanisms underlying protein dysfunction	136
4.1	Introduction	136
4.2	Results and discussion	
4.2.	.1 PTEN variants impact sensory processing and behavior in <i>C. elegans</i>	
4.3	Cross-model comparisons reveal multiple molecular mechanisms of protein	
dysfu	nction stemming from missense variants in PTEN	138
4.4	Materials and Methods	
4.4.	.1 PTEN Variant selection and annotation	
4.4.	.2 <i>C. elegans</i> chemotaxis assay	144
4.4.	.3 Data modeling and variant effects analysis	145
4.5	Supplemental information	147
Chapter	r 5: Conclusion	149
5.1	Insights into the convergent functions of ASD risk genes	150
5.2	Parallel pathways among high-confidence ASD risk genes	151
5.3	Deciphering variants of uncertain significance in PTEN	152
5.4	Future directions	
Referen	ICes	156

# List of Tables

Table 1.1 Select discoveries in C. elegans research.	25
Table 2.1 Phenotypic feature descriptions.	40
Table 3.1 CRISPR reagent sequences used to generate deletion alleles for Peel-1-DMS	
validation	. 129

# List of Figures

Figure 1.1 Core autism spectrum disorder symptoms
Figure 1.2 The inheritance patterns of syndromes with known genetic etiology and high
incidence of autism, as well as that of genes recently identified to be associated with autism 5
Figure 1.3 Points of convergence among ASD risk genes
Figure 1.4 The Multi-Worm Tracker
Figure 2.1 ASD-associated gene ortholog identification and phenomic characterization pipeline.
Figure 2.2 Phenotypic features display varied moderate correlations that correspond to feature
subclasses
Figure 2.3 Phenotypic profiles enable rapid reverse genetic screens to identify shared functions
among ASD-associated genes
Figure 2.4 Quantitative phenotypic profiles enable rapid reverse genetic screens to identify
shared functions among ASD-associated genes
Figure 2.5 Phenotypic profiles of strains with mutations in ASD-associated genes define shared
and unique functions
Figure 2.6 Phenotypic profiles of strains with mutations in ASD-associated genes define shared
and unique functions and phenotypic modularity
Figure 2.7 A phenomic database of strains with mutations in ASD-associated genes
Figure 2.8 Phenotypic correlations and t-SNE clustering of 135 strains harboring mutations in
ASD-associated gene orthologs

Figure 3.8 Peel-1-DMS attenuates array-driven overpopulation/starvation and promotes
screening-free genome editing at diverse loci
Figure 3.9 daf-18 deletion causes mechanosensory hyporesponsivity that is rescued by targeted
replacement with human PTEN 114
Figure 3.10 Human gene replacement and <i>in vivo</i> phenotypic assessment accurately identifies
functional consequences of the pathogenic PTEN-G129E variant 116
Figure 3.11 Functional assessment of ASD risk variants of uncertain significance using CRISPR-
Cas9 human gene replacement
Figure 3.12 A conceptual framework for <i>in vivo</i> functional analysis of human genetic variation
using C. elegans
Figure 4.1 Functional impact of PTEN variants on sensory behavior in C. elegans
Figure 4.2 Cross-model correlations reveal stability-dependent and stability-independent
mechanisms of PTEN variant dysfunction
Figure 4.3 Functional and stability data identifies distinct mechanisms of molecular dysfunction.
Figure 4.4 PTEN variant classification based on 9 variant functional assays

## Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Catharine Rankin, for being such a supportive mentor and cultivating a lab where students are encouraged to follow their passion. I've learned so many lessons from her, but perhaps the most important for me was to not get bent out of shape when experiments don't work out as planned, and to instead be excited at the opportunity to learn and discover. She's inspired my sense of wonder in science, and I'll always be grateful.

I would also like to thank all of the members of my supervisory committee. First, I want to thank Dr. Kurt Haas, for invaluable mentorship in scientific strategy, and for showing me the fun in planning ambitious projects and the grit required to see them through. I'm also grateful to Dr. Kota Mizumoto, for introducing me to genomic engineering and for always being excited to brainstorm new experimental methods. And to Dr. Paul Pavlidis, thank you for always encouraging me to think about how my work relates to the bigger picture, for teaching me the high-throughput mindset and how to think statistically.

I'm also thankful for the opportunity to have worked with Dr. Don Moerman. He's been an amazing role model for me who epitomizes the *C. elegans* researcher we all strive to be: sharing, open, kind, and wise. I'll do be best to live up to the ideals of the worm community that he exemplified as I move forward in my scientific career.

To the members of the Rankin lab past and present, my graduate cohort, the variant collaboration, the worm community, and my Cold Spring Harbor communities, it's been a privilege getting to learn from and do science with you all. I feel incredibly fortunate to have made lasting friendships during my time here at UBC. I would especially like to thank Dr. Evan Ardiel, for being a close friend and mentor who is always excited to nerd out about some new data no matter where we are in the world. I'm also thankful to Manuel Belmadani and Vinci Au, for everything they taught me and for an incredible experience in doing truly collaborative science with true friends. I've been lucky enough to have the opportunity to mentor several talented undergraduates during my time at UBC, many of whom have contributed directly to this thesis and become colleagues and friends, thank you all!

Before UBC, I'm grateful for those who started my academic journey in neuroscience, in particular Dr. Kim Hellemans, for being such an inspiring lecturer, and Dr. Alfonso Abizaid, for helping me secure a funded spot in his lab that I might not have been able to get a start in research without. They also both taught me the important lesson that scientists can be cool.

I would also like to thank my martial arts coach Peter Tremblay, who instilled my dedication to learning and self-improvement. Whenever someone has complimented my work ethic or perspective it's almost always when I am practicing something he taught.

Most importantly, I want to thank my family, for understanding and supporting me on this journey. My parents both sacrificed so much to give my sister and I an incredible childhood, and I won't stop striving to make the most of the opportunities they've worked so hard to give me.

Finally, I would like to thank my fiancée Christine. I wouldn't be where I am or who I am without her and I can't wait to start our next adventure together.

To Christine

# **Chapter 1: Introduction**

## 1.1 Clinical overview of Autism Spectrum Disorders

Autism Spectrum Disorders (ASD) encompass a clinically heterogeneous group of neurodevelopmental disorders defined by early onset deficits in two core symptom domains: 1) social communication and interaction and 2) restrictive, repetitive behaviors or atypical responses to sensory information (Fig. 1.1; American Psychiatric Association, 2013; Lord et al., 2020). The diagnostic criteria of ASD have evolved over time, from an initially narrow and rigid set of symptom requirements to include a progressively broader spectrum of conditions with wide-ranging degrees of impairment (American Psychiatric Association, 2013; Kanner, 1943; Lord et al., 2020; Volkmar and McPartland, 2014). For example, The 5<sup>th</sup> and current edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) unites autistic disorder, Asperger's disorder, childhood disintegrative disorder, and pervasive developmental disorder-not otherwise specified (PDD-NOS), which were previously related but separate diagnoses, under the single grouping of Autism Spectrum Disorder (American Psychiatric Association, 2013; Volkmar and McPartland, 2014). Further, sensory processing abnormalities such as hyper- or hyporesponsitivity, once considered a frequent comorbidity (>90% of cases), are now considered part of the core diagnostic criteria (integrated into the restrictive and repetitive behavior domain; American Psychiatric Association, 2013; de la Torre-Ubieta et al., 2016).



#### Figure 1.1 Core autism spectrum disorder symptoms.

The core symptoms listed in the Venn diagram represent the common features required to receive an ASD diagnosis. All three social communication and interaction symptoms and two of four symptoms in the restrictive/repetitive behavior domain must be present to receive a diagnosis in the DSM-V criteria. Select comorbidities, proposed clinical biomarkers, and genetic variant classes are listed. SNV = Single Nucleotide Variant, EEG = Electroencephalogram (Adapted with permission from Veenstra-Vanderweele and Blakely, 2012).

While initially rare, with prevalence estimates on the order of 1 in 1000 as recent as the 1990s, ASD diagnosis rates have risen sharply and continue to rise, with current prevalence estimates of approximately 1 in 54 children under 8 years of age in the USA (global estimates

are similar at approximately 1-1.5%; Gaugler et al., 2014; Iossifov et al., 2014; Lord et al., 2020; Maenner et al., 2020; De Rubeis et al., 2014). There are substantial differences in the rate of ASD diagnosis between males and females, with males receiving a diagnosis approximately 4.3 times more often than females (Maenner et al., 2020). However, disentangling the biological, environmental, and/or diagnostic factors responsible for rising ASD prevalence estimates and male bias remains an area of active investigation (de la Torre-Ubieta et al., 2016; Robinson et al., 2013; Werling and Geschwind, 2013). Comorbidities are also frequent in ASD, which include but are not limited to epilepsy, motor problems (e.g. hypotonia or motor delay), gastrointestinal issues, and sleep disturbances (de la Torre-Ubieta et al., 2016; Volkmar and McPartland, 2014). Another salient comorbidity is intellectual disability (ID), present in ~35% of cases (de la Torre-Ubieta et al., 2016; Maenner et al., 2020; Volkmar and McPartland, 2014). Many of these comorbidities can present challenges in diagnosis and disease modelling as they overlap phenotypes observed in other neuropsychiatric disorders (e.g. Schizophrenia and attention deficit hyperactivity disorder). In particular, intellectual disability presents a major challenge in the clinical setting as it can severely confound diagnostic instruments designed to assess core ASD symptom domains (de la Torre-Ubieta et al., 2016; Veenstra-Vanderweele and Blakely, 2012; Volkmar and McPartland, 2014). Despite these challenges, behavioral criteria remain essential for diagnosis as there is currently no neuropathology or biomarker that can be used to definitively diagnose an individual with ASD. Although the burden of ASD is often high for individuals and families, the development of definitive diagnostics and effective pharmacological interventions has been hindered by the fact that we understand little about the biological mechanisms that cause ASD. The most promising direction for research into the

mechanisms of ASD has stemmed from the observation that they have a strong genetic component.

## 1.2 Early evidence of genetic contributions to ASD

A key early insight that revealed genetic contributions to ASD came from the observation that several rare medical genetic syndromes with diverse modes of inheritance have high penetrance for an ASD diagnosis (**Fig. 1.2**; Betancur, 2011; de la Torre-Ubieta et al., 2016; Zafeiriou et al., 2013). For example, ~40% of people with Fragile X syndrome, a rare neurodevelopmental disorder caused by mutations in the *FMRP* gene, also have a diagnosis of ASD (Fu et al., 1991; Verkerk et al., 1991). In a more extreme example, ~60-80% of individuals with Timothy syndrome, caused by mutations in the *CACNA1C* gene (Splawski et al., 2004), meet criteria for an ASD diagnosis (de la Torre-Ubieta et al., 2016; Zafeiriou et al., 2013). However, it is important to note that each of these syndromes are rare in the population, and none have been shown to account for >1% of ASD (Betancur, 2011; de la Torre-Ubieta et al., 2016; Zafeiriou et al., 2013). Together, analysis of these syndromes indicated that individual rare genetic variants could substantially increase risk for ASD. However, they also revealed that there is no single risk gene that accounts for all cases, instead there is an extraordinary degree of genetic heterogeneity underlying ASD (**Fig 1.2**).



# Figure 1.2 The inheritance patterns of syndromes with known genetic etiology and high incidence of autism, as well as that of genes recently identified to be associated with autism.

(A) The blue stars indicate a causal allele and the red pie charts indicate a small proportion of risk. Most dominant disorders stem from *de novo* variants. Autosomal recessive, autosomal dominant and X-linked inheritance patterns best fit a major gene model, whereas a polygenic model is best represented by additive risk. (B) The types of genetic variation (left and middle) and the developmental disorders (right) associated with autism. Examples of genes that have been associated with ASD are also indicated. (Adapted with permission from de la Torre-Ubieta et al., 2016).

An additional early insight into the genetic basis of ASD was the observation that it is highly heritable, with monozygotic concordance estimates of ~70-90% (Bailey et al., 1995; Fernandez and Scherer, 2017; Folstein and Rutter, 1977; Hallmayer et al., 2011; Lichtenstein et al., 2010; Pettersson et al., 2019; Ronald and Hoekstra, 2011; Rosenberg et al., 2009; Tick et al., 2016). The heritability estimate of ASD was, and remains, among the highest of any behaviorally diagnosed disorder, and has spurred large-scale efforts in search of additional genetic risk factors (de la Torre-Ubieta et al., 2016; Pettersson et al., 2019). Many of the early genome-wide investigations used microarray-based genome-wide association (GWAS) methods. These approaches are typically designed to identify common risk variants (usually single nucleotide variants present in 1% or more of the population, called Single Nucleotide Polymorphisms or SNPs) that on their own confer little risk for ASD but are thought to work additively to contribute substantial risk in aggregate (**Fig 1.2**). While early GWAS studies provided compelling evidence for the contribution of common variants *en masse* to ASD, for a variety of reasons including limited sample sizes, they did not conclusively identify any specific reproducible common risk variants at genome-wide significance (Devlin et al., 2011; Gaugler et al., 2014; Grove et al., 2019; Ma et al., 2009; The Autism Spectrum Disorders Working Group of The Psychiatric Genomics, 2017).

### 1.3 The contribution of *de novo* coding variants to ASD

ASD risk gene discovery has advanced rapidly in recent years, stemming in large part from the seminal discovery that *de novo* variants, which are new variants present in the affected child but not in either unaffected parent, are an important source of genetic risk (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Ronemus et al., 2014; Sanders et al., 2015). Pioneering studies first established that *de novo* structural variants (Sebat et al., 2007), and subsequently protein-altering single nucleotide variants (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012), occur at higher frequency in affected individuals compared to unaffected family members or controls (**Fig 1.3**), and identified individual risk loci with strong

evidence for their association with ASD (For review see: Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Ronemus et al., 2014; Sanders et al., 2015). The analysis of de novo variants in increasingly large family cohorts using whole exome and whole genome sequencing technology represent the most successful efforts in ASD risk gene discovery to date (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Ronemus et al., 2014; Sanders et al., 2015). A recent large-scale aggregation of exome sequencing data (35,584 exomes including 11,986 with ASD) identified 102 high-confidence risk genes through this approach alone (Satterstrom et al., 2020). Indeed, there are now well over 100 risk genes and loci with established ties to the disorder, many of which are already being used in diagnosis (Fernandez and Scherer, 2017; Satterstrom et al., 2020; Schaaf et al., 2020). It is estimated that de novo coding variants contribute to approximately 30% of all ASD cases and that several hundred independent loci contribute risk through this mechanism, suggesting there are still many important risk genes to discover (Iakoucheva et al., 2019; Iossifov et al., 2014; Ronemus et al., 2014). Similar to the monogenic syndromes described earlier, none of these genes have shown complete penetrance for, or specificity to ASD (they have also implicated in other neurodevelopmental disorders); and no single gene has been shown to account for >1% of cases, reinforcing the genetic heterogeneity underlying ASD (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Satterstrom et al., 2020; Schaaf et al., 2020).

#### **1.4** The emerging genetic architecture of ASD – a spectrum of risk

There is now strong evidence that both rare and common variants contribute risk for ASD (Grove et al., 2019; Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). At one end of the

genetic risk spectrum are the monogenic disorders, where a single gene is a major contributor to risk (Fig 1.2). At the other extreme is polygenic risk, stemming from the aggregate effects of thousands of common variants (Fig 1.2). Between these two extremes lies a collection of variants that have yet to be well characterized (Iakoucheva et al., 2019). Most rare variants identified to date are *de novo*, but there has also been success in identifying rare inherited variants (Fig 1.2; Devlin and Scherer, 2012; Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Sanders et al., 2015). These variants are either recessive or partially penetrant, allowing the variant to be inherited from parents who do not meet criteria for ASD (Fig 1.2). Further, GWAS studies have recently reached sufficient sample sizes (a metanalysis of 18,831 cases and 27,969 controls) to identify the first five credible common risk variants (Fig 1.2; Grove et al., 2019). There are likely to be thousands of common risk variants, all with small effects (Grove et al., 2019; Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). The contribution of all SNPs to a trait can be summarized using a polygenic risk score (PRS; Dudbridge, 2013), and polygenic risk scores were recently shown to predict ASD liability in multiple independent cohorts (Grove et al., 2019; Iakoucheva et al., 2019).

There are now clear examples that indicate common variants and rare variants can work additively to increase risk for ASD (Iakoucheva et al., 2019; Weiner et al., 2017). Further, rare variants can also work additively with other rare variants (i.e. oligogenic inheritance, 2 or more 'hits'). In analyses of individuals with a rare structural variant or *de novo* variant that increases risk for ASD with incomplete penetrance, the presence of a secondary rare variant in the genetic background was predictive of increased clinical severity (Girirajan et al., 2010; Pizzo et al., 2019). Similarly, a common variant derived PRS for ASD predicted clinical severity (ASD

diagnosis rate) in rare variant carriers (Weiner et al., 2017). Taken together, current evidence suggests that in ASD, as in other psychiatric disorders, genetic contributions are multifactorial even in the context of a large-effect rare variant. Importantly, the overall contribution of common variants is proportional to effect of the rare variant, with common variation playing a more minor role in those who carry highly penetrant rare variant (Bergen et al., 2019; Iakoucheva et al., 2019; Ronemus et al., 2014).

During the last decade much of the genetic architecture of ASD has come to light (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Sanders et al., 2015; Schaaf et al., 2020). The identification of specific high-confidence risk genes and variants represents a major advance that is already transforming diagnosis and clinical practice (Schaaf et al., 2020). However, at first, the identification of a list of risk genes represents a 'parts list' that does not immediately translate into a mechanistic understanding of ASD. This understanding could be revealed by determining the functions of these genes in neurodevelopment and behavior.

### 1.5 Insights into the biology of ASD and the quest for convergence

As increasingly large sets of ASD risk genes have been identified they have been repeatedly queried against existing databases to gain insight into their functions (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Sanders et al., 2015; Satterstrom et al., 2020). A finding that has been replicated multiple times is that several high-confidence risk genes play important roles in an encouragingly small set of broadly defined biological processes, including most prominently gene expression regulation (e.g. chromatin modification) and neuronal/synaptic communication (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; De Rubeis et al., 2014;

Sanders et al., 2015; Satterstrom et al., 2020). Many ASD risk genes also appear to be coexpressed in the developing human nervous system, particularly in cortical excitatory and inhibitory neuronal lineages (Chang et al., 2015; Grove et al., 2019; Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Lin et al., 2015; Parikshak et al., 2013; Satterstrom et al., 2020; Willsey et al., 2013; Xu et al., 2014). There is also evidence to suggest that certain ASD risk genes converge on specific signaling pathways involved in the regulation of cell proliferation and differentiation, including Wnt, mitogen activated protein kinase (MAPK), and mammalian target of rapamycin (mTOR) signaling (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; O'Roak et al., 2012; Winden et al., 2018). Of note, post-mortem transcriptomic analyses of idiopathic ASD have reported downregulation of genes implicated in neuronal communication and upregulation genes enriched in activated microglia and astrocytes, the latter of which are not directly affected by any identified large-effect risk variants. This has led some to speculate that these immune changes are a reaction to the developmental perturbations induced by core risk factors (Gupta et al., 2014; Voineagu et al., 2011).

These findings immediately lead to the question of whether ASD risk genes converge through a mechanism whereby the regulatory genes control the expression of the neuronal communication genes. Initial studies using targeted methods focused on select high-confidence risk suggested that this may indeed be the case. For example, chromatin immunoprecipitation sequencing (ChIP-seq) analyses have revealed that the proteins encoded by the high-confidence ASD risk genes *CHD8*, a chromatin remodeling factor, and *TBR1*, a transcription factor, preferentially bind and positively regulate other ASD risk genes (Cotney et al., 2015; Notwell et al., 2016; Sugathan et al., 2014). Similarly, multiple studies have shown that targets of the RNA-

binding proteins encoded by FMRP and CELF4 are enriched in ASD risk genes, and that this enrichment was stronger among their shared targets (Darnell et al., 2011; Wagnon et al., 2012). However, a recent large-scale analysis of the binding sites of 26 regulatory proteins challenged this notion. This study reported that while the regulatory targets of ASD risk genes were enriched for other ASD risk genes involved in gene regulation, neuronal communication genes were not (Satterstrom et al., 2020). Similarly, while several studies have reported an enrichment of protein-protein interactions among all ASD risk genes, more recent studies suggest that the gene expression regulation and neuronal/synaptic communication genes are enriched for physical interactions with themselves but not each other (de la Torre-Ubieta et al., 2016; O'Roak et al., 2012; Sanders et al., 2015; Satterstrom et al., 2020). Further, gene expression regulation and neuronal communication genes appear to peak in expression at different points development, with regulatory genes exhibiting a prenatal bias and neuronal communication genes exhibiting a postnatal bias (Satterstrom et al., 2020). While evidence for global convergence among ASD risk genes has eluded the field so far, these studies have nevertheless demonstrated that many ASD risk genes display convergence at multiple scales: many are co-expressed in developing cortical excitatory and inhibitory neurons and play important roles in gene regulatory networks and neuronal communication. These findings are valuable for refining investigations into the pathophysiological mechanisms of ASD by providing clues of when and where to look. However, it remains unclear precisely how these genes regulate nervous system development and function, or how their disruption perturbs the development and function of these circuits to alter behavior.



Figure 1.3 Points of convergence among ASD risk genes.

Multiple ASD risk genes interact within the context of gene regulatory networks. These are highlighted as biological processes within a single neuron (cytoplasm in dark gray and nucleus in light gray). Convergence is evident at multiple levels of interaction including DNA binding, RNA binding, and protein-protein interactions. Biological processes that are associated with ASD risk genes include the regulation of gene expression and synaptic function. ASD risk genes are expressed preferentially in the developing brain. Rare genetic variants implicated in ASD also converge upon specific signaling pathways involved in the regulation of cell proliferation and differentiation including mTOR, MAPK, and Wnt signaling (Adapted with permission from Iakoucheva et al., 2019).

## 1.6 Human phenotyping studies

The identification of rare variants with large effects that are thought to account for ASD in the affected individual offers an unprecedented opportunity to correlate genotype with specific phenotypes in an attempt to reveal gene function and potential genetic subtypes of ASD (Arnett et al., 2019; Bernier et al., 2014; Schaaf et al., 2020). There is a growing list of examples where specific phenotypes can be tied to specific risk genes (Arnett et al., 2019; Iakoucheva et al., 2019). For example, rare protein-altering variants in certain high-confidence genes such as CHD8 and PTEN predispose individuals to ASD with co-occurring macrocephaly (large head size) whereas variants in other genes such as *DYRK1A*, predispose to ASD with microcephaly (Arnett et al., 2019; Bernier et al., 2014; Butler et al., 2005; Courcet et al., 2012; Iakoucheva et al., 2019). Beyond morphological phenotypes, there are also burgeoning studies linking individual risk genes/variants with more complex psychiatric traits. One influential example from cognitive and language studies is that carriers of the reciprocal 16p11.2 deletion or duplication structural variation display mirror effects on specific linguistic skills, with deletion carriers having reduced performance, and duplication carriers having greater performance than IQ-matched controls (Hippolyte et al., 2016; Iakoucheva et al., 2019). Other features, such as seizures, motor problems, and ASD, are shared across both genotypes (Arnett et al., 2019; Duyzend and Eichler, 2015). These findings highlight the importance of fine-grained dimensional phenotyping and further suggest that the specificity of genotype-phenotype relationships can be extended to complex human psychiatric traits. Further, specific variants underlying genetic subtypes of ASD may eventually be used to design targeted therapeutics. One promising example is the high-confidence ASD risk gene SCN2A, which encodes a voltage-gate sodium ion channel. Gain-of-function variants in SCN2A predispose carriers to infantile-onset epilepsy whereas loss-of-function variants are predominantly associated with ASD, suggesting clear targets for potential therapeutic interventions (Arnett et al., 2019; Ben-Shalom et al., 2017; Sanders et al., 2018).

Importantly, all genes associated with ASD to date have also been implicated in other neurodevelopmental disorders, such as intellectual disability (Schaaf et al., 2020). However, for some ASD risk genes impacted by deleterious variants (e.g. *CHD8* and *PTEN*) the associated prevalence of ASD is higher than that of intellectual disability (Bernier et al., 2014; Satterstrom et al., 2020). In a recent and highly controversial analysis conducted as part of the large-scale exome sequencing effort mentioned earlier (Satterstrom et al., 2020), clinical phenotyping data was used in an attempt to classify specific risk genes that were more likely to be impacted in cases ascertained for ASD (53/102 high-confidence risk genes classified as "ASD predominant") versus those more frequently identified in individuals ascertained to have more severe neurodevelopmental delay and intellectual disability (49/102 high-confidence risk genes classifications in subsequent work, whereas others have argued that there is insufficient evidence to make such "autism-specific" genetic partitions at this point (Myers et al., 2020).

Many more human reverse genetic studies are now underway following the burst in ASD genetic discoveries in recent years (Arnett et al., 2019; Iakoucheva et al., 2019). These studies are remarkably valuable as the human biology they uncover is often unambiguously relevant to ASD (Arnett et al., 2019). However, human studies are inherently limited for important ethical and technical reasons that make it difficult to establish causality, control for differences in genetic and environmental background, screen for the effects of a large number of potential treatments or investigate cellular and molecular mechanisms in detail. Further, carries of individual rare variants are, by definition, rare, meaning it remains difficult to collect sufficiently large groups of patients even with rapidly declining sequencing costs and increasingly large

cohorts of affected individuals. For this reason, diverse model systems have and continue to be used to investigate the functions of ASD risk genes and to screen for potential therapeutics.

## 1.7 Model system investigations of ASD risk gene function

The last decade has seen the identification of a large number of high-confidence ASD risk genes with poorly understood biological functions alongside rapid improvements in genome engineering technology (Iakoucheva et al., 2019; Knott and Doudna, 2018; Schaaf et al., 2020). These advances have necessitated and enabled increasingly sophisticated investigations of risk gene function in a growing collection of model systems (Coll-Tané et al., 2019; Iakoucheva et al., 2019; Ijaz and Hoffman, 2016; de la Torre-Ubieta et al., 2016). These studies generally follow the same arc: a genetic perturbation is made in a model system that corresponds to an identified human risk variant and then the model system is scored for various phenotypes that provide insight into the biological functions of the gene. The most prominent model system used to study ASD risk genes to date is the mouse, though other model systems are being used increasingly frequently (Coll-Tané et al., 2019; Ijaz and Hoffman, 2016; Kepler et al., 2020; de la Torre-Ubieta et al., 2016). Of note, advances in human induced pluripotent stem cell (iPSCs) technology are making it increasingly feasible to study neurons derived from patients with identified rare mutations (or idiopathic ASD) in vitro and are quickly becoming one of the most popular model systems (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). Further, threedimensional brain organoids derived from self-assembling IPSCs are allowing unprecedented opportunities to observe and manipulate cellular processes underlying human neurodevelopment, including progenitor proliferation, migration, differentiation, and maturation (Amin and Paşca, 2018). While not yet used as widely to study ASD risk genes, high-throughput genetic model

organisms such as zebrafish (*Danio rerio*) and fruit flies (*Drosophila melanogaster*) have been used to investigate the functions of specific risk genes and have already demonstrated value in rapid whole-organism pharmacological screening (e.g. promising potential treatments for sleep disturbances and seizures stemming from variants in ASD risk genes have recently been identified using zebrafish; Coll-Tané et al., 2019; Griffin et al., 2016; Hoffiman et al., 2016; Ijaz and Hoffman, 2016; Kepler et al., 2020; van der Voet et al., 2014). Genome engineered nonhuman primates are also being used as an *in vivo* model system with complex social behavior (de la Torre-Ubieta et al., 2016; Zhou et al., 2019). Each model system has unique advantages, with iPSCs and mammalian model systems offering a higher degree of evolutionary conservation and/or behavioral similarity and high-throughput genetic model organisms offering rapid genetic manipulation and phenotypic characterization.

These studies have led to important insights into the biological functions of ASD risk genes (for detailed reviews see: Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). Of note, the majority of mouse models of ASD risk genes display altered social behavior, repetitive behaviors, and sensory processing abnormalities (Kazdoba et al., 2015; de la Torre-Ubieta et al., 2016). However, the severity of these deficits, and the presence of other features such as motor dysfunction and seizures, varies widely (de la Torre-Ubieta et al., 2016). Another finding consistently observed in human iPSCs and animal models is that disruption of ASD risk genes leads to dysregulation of other ASD risk genes in *trans*, as well as dysregulation of pathways commonly impacted by core ASD risk genes (e.g. mTOR, MAPK, Wnt) (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). That is, even ASD risk genes that are not direct cisregulatory targets of proteins encoded by ASD risk genes (such as those in the studies described

earlier) are preferentially dysregulated by inactivating ASD risk genes (Araujo et al., 2015; Gompers et al., 2017; Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016; Notwell et al., 2016; Sugathan et al., 2014). Studies in iPSCs and animal models have also consistently noted abnormalities in cell proliferation and brain growth that generally align with what has been observed in human rare variant carriers (in the relatively limited instances where human data is available for comparison; Arnett et al., 2019; Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). For example, iPSCs and animal models with disruptive variants in CHD8 and PTEN show increased cell proliferation and macrocephaly, whereas DYRK1A disruption led to reduced proliferation and microcephaly (Bellmaine et al., 2017; Bernier et al., 2014; Fotaki et al., 2002; Gompers et al., 2017; Kwon et al., 2006; Tejedor et al., 1995). Morphological changes in dendritic arborization, synapse number, and corresponding electrophysiological disruptions have also been consistently observed across iPSC and animal model systems (Iakoucheva et al., 2019; de la Torre-Ubieta et al., 2016). For example, iPSCs derived from Rett syndrome patients with loss-of-function variants in MECP2 displayed reduced dendritic arborization, decreased excitatory synapse numbers, and reduced spontaneous postsynaptic currents, whereas cellular models of MECP2 gain of function (MECP2 duplication) displayed opposite effects (Marchetto et al., 2010; Nageshappa et al., 2016).

Model systems have been used to discover numerous genotype-phenotype relationships that reveal how disruption of select ASD risk genes impact neurodevelopment. Further, they provide an important testbed to screen for the ability of potential treatments to reverse these phenotypic disruptions. However, an important takeaway from these investigations is that different ASD risk genes often have drastically different or even directly opposite effects across
molecular, cellular, and behavioral phenotypes. Taken together, these studies clearly illustrate the need to systematize model system phenotyping efforts and increase throughput to reveal the shared and unique functions of the rapidly growing list of ASD risk genes.

#### 1.8 The growing challenge of variants of uncertain significance

Beyond understanding the functions of the large and growing number of genes implicated in ASD, a second major challenge facing ASD genetics is that it is currently extremely difficult to predict the functional consequences of specific single nucleotide variants within those genes, so called "variants of uncertain significance" (Landrum et al., 2014, 2016; Starita et al., 2017). Variants identified in ASD are often so rare that they preclude traditional genetic approaches to infer pathogenicity (e.g. observation of the variant in multiple affected individuals or segregation within a pedigree; Karczewski et al., 2020; Sanders et al., 2015; Satterstrom et al., 2020; Starita et al., 2017). It is sobering to note that the majority of the thousands of de novo variants identified in individuals with ASD are currently classified as variants of uncertain significance, meaning their role as a causative agent in the disorder, or their pathogenicity, remains unknown (Iossifov et al., 2014; Ronemus et al., 2014; De Rubeis et al., 2014; Sanders et al., 2015; Satterstrom et al., 2020). This means that even for genes with well-characterized biological functions there are often hundreds of variants of uncertain significance in need of functional characterization. A complete understanding of the genetic architecture and pathophysiological mechanisms of ASD requires deciphering the functional impact of these variants.

The challenges posed by variants of uncertain significance are currently being faced by many areas of clinical genetics, and in an attempt to address these challenges multiple computational algorithms have been developed to predict the functional impact of variants (Rentzsch et al., 2019; Richards et al., 2015). These tools use a variety of annotations including evolutionary conservation, protein structural information, or frequencies of variants in putatively healthy control populations to predict whether a given variant is likely to be pathogenic (Kircher et al., 2014; Rentzsch et al., 2019; Richards et al., 2015; Starita et al., 2017; Weile and Roth, 2018). These tools are easily scalable, allowing predictions to made for any variant. However, none of these tools used in isolation or combination can accurately report on the functional consequences of a large portion of genetic variation, and their accuracy is intrinsically limited to existing experimental data (Grimm et al., 2015; Miosge et al., 2015; Starita et al., 2017). These limitations were clearly illustrated by a recent study that showed *in vivo* transgenic experimental functional assays in yeast could discern pathogenic variants in 21 human disease-associated genes with greater sensitivity and specificity than leading computational methods (Sun et al., 2016). This creates a challenging situation that necessitates experimental evaluation of the functional significance of variants of uncertain significance.

Model systems play crucial roles in experimental functional characterization of variants of uncertain significance (Boycott et al., 2020; Hmeljak and Justice, 2019; Kepler et al., 2020; Wangler et al., 2017). Governmental and private organizations are increasingly commissioning large-scale programs to support the development of collaborations that connect clinicians identifying variants of uncertain significance with model system researchers equipped to evaluate their functional impact (Boycott et al., 2020; Chong et al., 2015; Gahl et al., 2016; Wangler et al., 2017). These studies typically involve introducing the variant into a model system and screening for phenotypic divergence from controls. Variants of uncertain significance

that cause severe phenotypic disruptions (e.g. complete loss-of-function equivalent to a null allele) and/or mimic known pathogenic variants of the gene are classified as function altering and/or likely pathogenic (Boycott et al., 2020; Wangler et al., 2017). Such studies in model systems have already demonstrated their utility by discerning bona fide causal variants underlying diverse neurodevelopmental disorders and identifying potential personalized treatments targeted to the affected individuals' specific genetic variant (Boycott et al., 2020; Kepler et al., 2020; Wangler et al., 2017). In a recent illustrative example, clinical exome sequencing identified de novo missense variants of uncertain significance in CDK19 (which encodes a cyclin dependent protein kinase) in three individuals with a severe neurodevelopmental disorder associated with intellectual disability, epilepsy, and ASD (Chung et al., 2020). Drosophila harboring loss-of-function mutations in the ortholog CDK19 exhibited severe seizures and loss of synapses, consistent with a role for *CDK19* in neurodevelopment. Expression of a human wild-type/reference control cDNA was able to rescue these impairments, whereas expression of the p.Tyr32His and p.Thr196Ala variants identified in the affected individuals failed to rescue and behaved as null alleles. Taken together, these results indicated that p.Tyr32His and p.Thr196Ala are strongly function altering variants and that deleterious CDK19 variants underlie a syndromic neurodevelopmental disorder (Chung et al., 2020). With so many variants in need of functional characterization, promising results like these indicate that in addition to systematizing efforts to identify the functions of ASD risk genes, we also need to systematize the analysis of variants of uncertain significance. In order to gain more global and accurate views of the functional impact of genetic variation, these efforts would ideally employ multiple phenotypic functional assays and precise methods to model human genetic variation.

#### 1.9 Major challenges facing the genetics of ASD

In the last decade ASD has gone from one of the most mysterious and poorly understood disorders in psychiatry to one of the success stories of the post-genomic era. However, the recent burst of genetic discoveries has presented two major challenges now facing ASD genetics:

1) The large and growing number of candidate risk genes associated with ASD with poorly characterized biological functions.

2) The inability to predict the functional consequences of the large number of variants of uncertain significance.

These challenges need to be addressed if we are to transform the growing list of implicated genes and variants into a mechanistic understanding of ASD. Due to the formidable technical challenges associated with conducting studies in animal models at scale, this lag between gene discovery and functional characterization is even more pronounced when assessing the roles of ASD risk genes in complex sensory and learning behaviors. There is a great need to rapidly determine the functions of ASD-associated genes, the functional consequences of variants of uncertain significance, and delineate complex functional genetic networks among ASD-associated genes *in vivo*. What is needed is model system where a large number of genes and variants can be perturbed - and their phenotypic consequences measured - to gain insight into the multiple functions of ASD risk genes and the functional impact of specific variants.

#### 1.10 Finding function in novel targets – C. elegans as a model system

The nematode worm *Caenohabditis elegans* has a rich history as a pioneering model system for the functional characterization of disease-associated genetic variation - particularly high-throughput functional characterization of risk genes identified through genomics (Corsi et al., 2015; Kaletta and Hengartner, 2006; Kepler et al., 2020). C. elegans was first chosen by Sydney Brenner as a genetic model to study development and the nervous system, but today C. elegans is used across a wealth of diverse areas in modern biology (Brenner, 1974; Corsi et al., 2015). Table 1.1 lists a small selection of the many contributions C. elegans research has made to both basic and biomedical research (Corsi et al., 2015). Notable C. elegans discoveries relevant to human disease include the identification of presenilin as part of the gamma secretase complex, the mechanism of action of antidepressant medications, and the role of the insulin signaling pathway in normal and pathological aging (Kaletta and Hengartner, 2006; Levitan and Greenwald, 1995; Levitan et al., 1996; Murphy et al., 2003; Ranganathan et al., 2001; Table **1.1**). C. elegans has also been used to investigate the functions of ASD risk genes. For example, systematic in vivo RNA interference (RNAi) screens have revealed novel roles for many ASD risk genes in synapse development and function (Sieburth et al., 2005), and detailed investigations have elucidated novel mechanisms through which select ASD risk genes interact to orchestrate synaptic physiology (Hu et al., 2012; Pym et al., 2017; Tong et al., 2015, 2017).

Year	Discovery	References
1974	Identification of mutations that affect animal behavior	Brenner 1974 PMID: 4366476; Dusenberry <i>et al.</i> 1975 PMID: 1132687; Hart 2006

Year	Discovery	References
1975	First description of mutations that affect thermotaxis and mechanotransduction	Hedgecock and Russell 1975 PMID: 1060088; Sulston <i>et al.</i> 1975 PMID: 240872; Chalfie and Sulston 1981 PMID: 7227647; Mori and Ohshima 1995, PMID: 7630402
1977	First cloning and sequencing of a myosin gene	Macleod et al. 1977 PMID: 909083
1977	977Genetic pathways for sex determination and dosage compensation describedHodgkin and Brenner 1977 PMID: 56 2005 PMID: 18050416; Zarkower 20 18050479	
1981	1Identification of mutations affecting touch sensitivitySulston <i>et al.</i> 1975 PMID: 240872; Chalfie and Sulston 1981 PMID: 7227647	
1981	First germline stem cell niche identified	Kimble and White 1981 PMID: 7202837; Kimble and Crittenden 2005 PMID: 18050413
1983	Notch signaling, presenilins, ternary complex, and lateral inhibition roles in development described	Greenwald <i>et al.</i> 1983 PMID: 6616618; Levitan and Greenwald 1995 PMID: 7566091; Petcherski and Kimble 2000 PMID: 10830967; Greenwald and Kovall 2012 PMID: 23355521
1983	First complete metazoan cell lineage	Sulston and Horvitz 1977 PMID: 838129; Kimble and Hirsh 1979 PMID: 478167; Sulston et al. 1983 PMID: 6684600
1983	Discovery of apoptosis (cell death) genes	Hedgecock <i>et al.</i> 1983 PMID: 6857247; Ellis and Horvitz 1986 PMID: 3955651; Yuan and Horvitz 1992 PMID: 1286611; Yuan <i>et al.</i> 1993 PMID: 8242740; Conradt and Xue 2005 PMID: 18061982
1984	Identification of heterochronic genes	Ambros and Horvitz 1984 PMID: 6494891; Slack and Ruvkun 1997 PMID: 9442909
1986	First complete wiring diagram of a nervous system	White <i>et al.</i> 1986 PMID: 22462104; Jarrell <i>et al.</i> 2012 PMID: 22837521; White 2013 PMID: 23801597
1987	Discovery of the first axon guidance genes	Hedgecock <i>et al.</i> 1987, 1990 PMID: 3308403 PMID: 2310575; Culotti 1994 PMID: 7950328

Year	Discovery	References	
1987	Identification of role of Notch signaling in embryonic blastomeres	Priess <i>et al.</i> 1987 PMID: 3677169; Priess 2005 PMID: 18050407	
1988	Discovery of <i>par</i> genes, whose products affect the asymmetric distribution of cellular components in embryos	Kemphues <i>et al.</i> 1988 PMID: 3345562; Gönczy and Rose 2005 PMID: 18050411	
1988	38Identification of the first LIM and POU homeodomain transcription factorsWay and Chalfie 1988 PMID: 2898300; Finne al. 1988 PMID: 2903797; Hobert 2013 PMID 24081909		
1990	First description of a role for RAS signaling function in metazoan developmentBeitel <i>et al.</i> 1990 PMID: 2123303; Han and Sternberg 1990 2257629; Sternberg 2005 PMID 		
1990	<i>C. elegans</i> established as a model system to investigate the mechanisms of learning and memory/multiple mechanisms of habituation identified	Rankin <i>et al.</i> 1990 PMID: 2310497; Rankin & Broster 1992 PMID: 1590951; Broster & Rankin 1994 PMID: 7893394; Kindt <i>et al.</i> 2007 PMID: 17698017; Ardiel, McDiarmid & Timbers <i>et al</i> 2018 PMID: 30429311	
1993	Demonstration of a role for insulin pathway genes in regulating lifespan	Friedman and Johnson 1988 PMID 8608934; Kenyon <i>et al.</i> 1993 PMID: 8247153; Kimura <i>et al.</i> 1997 PMID: 9252323; Collins <i>et al.</i> 2007 PMID: 18381800	
1993	Identification of genes for conserved synaptic functions	Gengyo-Ando <i>et al.</i> 1993 PMID: 8398155; Richmond <i>et al.</i> 1999 PMID: 10526333; Richmond 2007 PMID: 18050398	
1993	First microRNA (lin-4) and its mRNA target (lin-14) described	Lee <i>et al.</i> 1993 PMID: 8252621; Wightman <i>et al.</i> 1993 PMID: 8252622; Vella and Slack 2005 PMID: 18050425	
1993	Identification of nonsense- mediated decay genes	Pulak and Anderson 1993 PMID: 8104846; Hodgkin 2005 PMID: 18023120	
1994	Introduction of GFP as a biological marker	Chalfie <i>et al.</i> 1994 PMID: 8303295; Boulin <i>et al.</i> 2006 PMID: 18050449	
1994	First demonstration of specific olfactory receptor/ligand pair	Sengupta <i>et al.</i> 1994 PMID: 8001144; Bargmann 2006 PMID: 18050433	

Year	Discovery	References
1998	First metazoan genome sequenced	C. elegans Sequencing Consortium 1998 PMID: 9851916; Schwarz 2005 PMID: 18023117
1998	Discovery of RNA interference (RNAi)	Fire et al. 1998 PMID: 9486653
2000	Conservation and ubiquity of miRNAs	Pasquinelli et al. 2000 PMID: 11081512
2000	Development of genome-wide RNAi screening/first full genome- wide profiling of gene function	Fraser <i>et al.</i> 2000 PMID: 11099033; Kamath <i>et al.</i> 2001 PMID: 11178279
2000	Transgenerational inheritance and its mediation by piRNA	Grishok <i>et al.</i> 2000 PMID: 10741970; Ashe <i>et al.</i> 2012 PMID: 22738725
2002	First cytoplasmic polyA polymerase ( <i>gld-2</i> ) discovered	Wang <i>et al.</i> 2002 PMID: 12239571; Kimble and Crittenden 2005 PMID: 18050413
2005	First full-genome RNAi profiling of early embryogenesis	Sönnichsen et al. 2005 PMID: 15791247
2005	First use of channelrhodopsin optogenetics in an intact animal	Nagel et al. 2005 PMID: 16360690
2011	Discovery of first nematode viruses	Félix et al. 2011 PMID: 21283608
2017	First comprehensive single-cell transcriptional profiling of a multi-cellular animal	Cao et al. 2017 PMID: 28818938; Packer et al. 2019 PMID: 31488706

 Table 1.1 Select discoveries in C. elegans research.

(Adapted with permission from Corsi et al., 2015).

These discoveries have been fueled by a strong tradition of creating comprehensive tools and datasets that are freely available for anyone to use, an approach that was instilled the worm community by its founders. *C. elegans* was the first animal to have is cell lineage traced (Sulston and Horvitz, 1977; Sulston et al., 1983), connectome mapped (White et al., 1986), and genome

sequenced (C. elegans Sequencing Consortium, 1998). It was also the first animal where genome-wide RNA interference screens were achieved and will likely be the first animal with loss-of-function mutations available for every gene (Hutter and Moerman, 2015; Kamath et al., 2003). There are already well-annotated libraries with deletion alleles available for >16,000 of the  $\sim 20,000$  protein coding genes, and the implementation of CRISPR-based genome engineering now allows targeted deletion or editing of any gene of interest in <1 month and should streamline completion of the knockout library (Au et al., 2018; Dickinson and Goldstein, 2016; Hutter and Moerman, 2015; Norris et al., 2015; Thompson et al., 2013). Recently, C. elegans also became the first animal for which the complete transcriptome of every one of its identified cells has been measured (Cao et al., 2017). This feat was just extended to include multiple developmental stages, allowing for lineage-resolved analysis of transcriptomic trajectories at single-cell resolution (Packer et al., 2019). This means that C. elegans researchers have an unmatched understanding of when every cell in the animal is born, when it dies, where it came from, what it connects to, and what genes it expresses at multiple points in development. Moreover, there are a vast set of tools to efficiently manipulate genes in a spatially and temporally precise manner and observe the effects of these manipulations on development and complex behaviors in vivo.

*C. elegans* also possess several attributes that naturally lend themselves to highthroughput biology. *C. elegans* are small (~1mm long fully grown), easy to culture in the lab, develop rapidly (~3 days from egg to egg laying adult), and exhibit a prolific hermaphroditic mode of reproduction where each adult will typically produce several hundred progeny, allowing for routine cultivation of large populations of isogenic animals (Brenner, 1974; Corsi et al.,

2015; Kaletta and Hengartner, 2006; Mcdiarmid et al., 2018). *C. elegans* are also transparent, allowing for rapid analysis of the effects of a genetic perturbation on nervous system development and function in intact animals (e.g. via fluorescent reporters of neuron morphology or activity; Chalfie et al., 1994; Corsi et al., 2015; Suzuki et al., 2003). Importantly, *C. elegans* have orthologs of >50% of human genes, and human genes have repeatedly been shown to be so structurally and functionally conserved that they can directly replace their *C. elegans* ortholog (Calahorro and Ruiz-Rubio, 2012; Kim et al., 2018; Levitan et al., 1996; McDiarmid et al., 2018a; Solari et al., 2005). Finally, CRISPR-Cas9 genome editing has proven to remarkably efficient and precise in *C. elegans*, with multiple rigorous whole genome sequencing studies revealing no evidence of off-target effects in this organism (Au et al., 2018; Chiu et al., 2013).

In addition to all of these awesome worm advantages, another primary reason I chose to use *C. elegans* to investigate the functions of ASD risk genes is because our lab developed the Multi-Worm Tracker (Swierczek et al., 2011). The Multi-Worm Tracker is a machine vision phenotyping system consisting of a high-resolution camera suspended over a petri plate of worms as well as software for real-time feature extraction and automated delivery of sensory stimuli (e.g. an automated mechanosensory "tap" delivered to the petri plate via a push solenoid; **Fig. 1.4**). The Multi-Worm Tracker is capable of creating detailed digital representations of >100 freely behaving animals while they carry out complex sensory and learning behaviors (Mcdiarmid et al., 2018; Swierczek et al., 2011). The representations are stored and can then be analyzed offline to computationally extract virtually any simple or compound phenotypic feature of choice, including measures of morphology, baseline locomotion, sensory and learning behaviors (Mcdiarmid et al., 2018; Swierczek et al., 2011). Multiplexing by running multiple

trackers in parallel allows for automated phenotypic analysis of several hundred worms simultaneously.



#### Figure 1.4 The Multi-Worm Tracker.

(Top) The Multi Worm Tracker allows for high throughput, high resolution behavioral analysis of *C. elegans*. The Multi-worm tracker delivers mechanosensory stimuli and preforms image acquisition, object selection, and parameter selection in real time while choreography software extracts detailed phenotypic information offline. (Bottom) (A) A petri plate of *C. elegans* (B) A petri plate of *C. elegans* selected for analysis by the Multi-Worm Tracker (C) A Multi-worm tracker digital representation showing the degree of phenotypic detail. The *C. elegans* response to a mechanosensory tap to the side of the Petri plate is brief backwards locomotion (from C to D). This habituates (decreases in probability, duration and speed) with repeated taps. Scale bars are 1cm, 1cm, 0.25mm, 0.25mm from left to right.

#### 1.11 Scientific motivations

At the beginning of my graduate work I saw an opportunity to combine CRISPR-Cas9 genome engineering with machine vision phenomics to gain insight into the functions of ASD risk genes and the functional impact of specific genetic variants. I developed a pipeline to discover functions of ASD risk genes by inactivating each gene in *Caenorhabditis elegans* and observing the phenotypic consequences using machine vision. I quantified 26 phenotypes spanning morphology, locomotion, tactile sensitivity, and learning in >27,000 animals representing 135 genotypes (98 strains with mutations in different genes and 37 strains with additional alleles of a subset of these genes), and identified disruptions in habituation (a neural circuit's plastic ability to decrease responding to repeated sensory stimuli) as a shared impairment. I then demonstrated how this database can facilitate experiments that determine the functional consequences of missense variants and whether phenotypic alterations are reversible. Further, I developed a broadly applicable CRSIRP-Cas9 genome editing strategy to replace C. elegans genes with human genes that allows for in vivo analysis of human genetic variation with unprecedented precision. Finally, I contributed to the development of a multi-model system pipeline for high-confidence assessment of missense variants in ASD risk genes. These contributions are presented in subsequent chapters as independent manuscripts characteristic of the field.

# Chapter 2: Systematic phenomics analysis of Autism-associated genes reveals parallel networks underlying reversible impairments in habituation.

A major challenge facing the genetics of Autism Spectrum Disorders (ASD) is the large and growing number of candidate risk genes and gene variants of unknown functional significance. Here, we used *Caenorhabditis elegans* to systematically functionally characterize ASD-associated genes in vivo. Using our custom machine vision system we quantified 26 phenotypes spanning morphology, locomotion, tactile sensitivity, and habituation learning in 135 strains (98 strains with mutations in different genes and 37 strains with additional alleles of a subset of these genes) each carrying a mutation in an ortholog of an ASD-associated gene. We identified hundreds of genotype-phenotype relationships ranging from severe developmental delays and uncoordinated movement to subtle deficits in sensory and learning behaviors. We clustered genes by similarity in phenomic profiles and used epistasis analysis to discover parallel networks centered on CHD8•chd-7 and NLGN3•nlg-1 that underlie mechanosensory hyperresponsivity and impaired habituation learning. We then leveraged our data for in vivo functional assays to gauge missense variant effect. Expression of wild-type NLG-1 in *nlg-1* mutant C. elegans rescued their sensory and learning impairments. Testing the rescuing ability of conserved ASD-associated neuroligin variants revealed varied partial loss-of-function despite proper subcellular localization. Finally, we used CRISPR-Cas9 auxin inducible degradation to determine that phenotypic abnormalities caused by developmental loss of NLG-1 can be reversed by adult expression. This work charts the phenotypic landscape of ASD-associated genes, offers in vivo variant functional assays, and potential therapeutic targets for ASD.

#### 2.1 Introduction

Autism Spectrum Disorders (ASD) encompass a clinically and genetically heterogeneous group of neurodevelopmental disorders characterized by deficits in social communication and interaction, restrictive repetitive behaviors, and profound sensory processing abnormalities (American Psychiatric Association, 2013; Fernandez and Scherer, 2017; de la Torre-Ubieta et al., 2016; Sinclair et al., 2016). The fifth edition of the Diagnostic and Statistical Manual of Mental disorders combines autistic disorder, Asperger disorder, childhood disintegrative disorder and pervasive developmental disorder not otherwise specified into the single grouping of Autism Spectrum Disorder (American Psychiatric Association, 2013). Despite extensive study, there is currently no unanimously agreed upon structural or functional neuropathology common to all individuals with ASD and there is little understanding of the biological mechanisms that cause ASD (de la Torre-Ubieta et al., 2016). The most promising avenue for research into ASD has stemmed from the observation that they have a strong genetic component, with monozygotic concordance estimates of ~70-90% and several distinct highly penetrant genetic syndromes (Fernandez and Scherer, 2017; de la Torre-Ubieta et al., 2016).

Rapid advances in copy number variation association, whole-exome, and more recently, whole-genome sequencing technology and the establishment of large sequencing consortia have dramatically increased the pace of gene discovery in ASD (Iossifov et al., 2014; De Rubeis et al., 2014; Sanders et al., 2015; Satterstrom et al., 2019; Stessman et al., 2017). There are now >100 diverse genes with established ties to ASD, many of which are being used in diagnosis. Importantly, each gene accounts for <1% of cases and none have shown complete specificity for ASD, with many implicated in multiple neurodevelopmental disorders (Fernandez and Scherer,

2017; de la Torre-Ubieta et al., 2016; Satterstrom et al., 2019). Some of these genes have fallen into an encouragingly small set of broadly defined biological processes such as gene expression regulation (e.g. chromatin modification) and synaptic neuronal communication (de la Torre-Ubieta et al., 2016; O'Roak et al., 2012; De Rubeis et al., 2014; Sanders et al., 2015; Satterstrom et al., 2019). Seminal studies using mouse models, genetically stratified populations of individuals with ASD, human induced pluripotent stem cells (iPSCs), and more recently highthroughput genetic model organisms such as *Drosophila* and zebrafish have investigated the molecular, circuit, and behavioral phenotypic disruptions that result from mutations in diverse ASD-associated genes. These systems have offered valuable insights into the biological mechanisms underlying this heterogeneous group of disorders (Amin and Paşca, 2018; Beighley et al., 2019; Bernier et al., 2014; Coll-Tané et al., 2019; Ellegood et al., 2015; Grissom et al., 2018; Kwon et al., 2006; Paşca et al., 2011; Peñagarikano et al., 2015; Silverman et al., 2010; Stessman et al., 2014; van der Voet et al., 2014; Zaslavsky et al., 2019; Zoghbi and Bear, 2012). However, thousands of additional mutations in these and many other genes have been identified in individuals with ASD and their roles as causative agents, or their pathogenicity, remains ambiguous. Thus, there are two major challenges facing ASD genetics: 1) the large, growing number of candidate risk genes with poorly characterized biological functions and; 2) the inability to predict the functional consequences of the large number of rare missense variants. Difficulties in rare missense variant interpretation stem in part from constraints on computational variant effect prediction and a paucity of in vivo experimental variant functional assays (Karczewski et al., 2019; Starita et al., 2017). This lag between gene discovery and functional characterization is even more pronounced when assessing the role of putative ASD risk genes and variants in complex sensory and learning behaviors. As such, there is a great need to rapidly

determine the functions of ASD-associated genes, the functional consequences of variants of uncertain significance, and delineate complex functional genetic networks among ASD-associated genes *in vivo*.

The genetic model organism *Caenorhabditis elegans* is a powerful system for the functional analysis of disease-associated genetic variation, particularly for high-throughput in vivo characterization of risk genes identified through genomics (Kaletta and Hengartner, 2006). *C. elegans*' fully sequenced and thoroughly annotated genome and complete connectome have fueled numerous disease discoveries, including the identification of presenilins as part of the gamma secretase complex and the role of the insulin signaling pathway in normal and pathological aging (Kaletta and Hengartner, 2006; Kenyon et al., 1993; Levitan and Greenwald, 1995; White et al., 1986). There are clear *C. elegans* orthologs for >50% of human genes, and human genes have repeatedly been shown to be so structurally and functionally conserved that they can directly replace their C. elegans ortholog (Calahorro and Ruiz-Rubio, 2012; Kaletta and Hengartner, 2006; Kim et al., 2018; McDiarmid et al., 2018a). There are well-annotated libraries of C. elegans strains with deletion alleles available for >16,000 of the  $\sim 20,000$  protein coding genes (Au et al., 2018). C. elegans small size and rapid hermaphroditic mode of reproduction (3 days from egg to egg-laying adult) allows for the routine cultivation of large numbers of isogenic animals. Further, CRISPR-Cas9 genome engineering is reliable and efficient in C. elegans, and unlike most organisms analyzed to date, rigorous whole-genome sequencing studies have revealed no significant off-target effects due to CRISPR-Cas9 genome editing in this organism (Au et al., 2018; Chiu et al., 2013). Finally, we developed the Multi-Worm Tracker (MWT), a machine vision system that allows for comprehensive phenotypic analysis of large populations of

freely behaving animals while they perform complex sensory and learning behaviors (McDiarmid et al., 2018b; Swierczek et al., 2011). Multiplexing by running several trackers in parallel allows for analysis of multiple measures of morphology, locomotion, mechanosensory sensitivity, and several forms of learning in hundreds of animals simultaneously.

Here, we developed a scalable phenomic characterization pipeline to discover functions of ASD-associated genes by systematically inactivating each gene in a model organism and observing the phenotypic consequences using machine vision (**Fig. 2.1A**). We present data summarizing scores on 26 quantitative phenotypes spanning morphology, baseline locomotion, tactile sensitivity, and habituation learning in 135 strains of *C. elegans* (98 strains with mutations in different genes and 37 strains with additional alleles of a subset of these genes) each carrying a mutation in an ortholog of an ASD-associated gene, revealing hundreds of shared and unique genotype-phenotype relationships. We clustered strains based on phenotypic similarity to discover functional interactions among ASD-associated genes, which we validate with epistasis analysis. Further, we use the phenotypic profiles for *in vivo* functional assays to assess missense variant effect, and to determine whether phenotypic alterations are reversible using targeted protein degradation methods based on degrons.

#### 2.2 Results



Figure 2.1 ASD-associated gene ortholog identification and phenomic characterization pipeline.

(A) Animals carrying mutations in orthologs of ASD-associated genes are systematically engineered using CRISPR-Cas9 or ordered from stock centers, large populations of are grown and their phenotypic profiles are characterized using The MWT. Genotype-phenotype relationships are used to cluster strains based on phenotypic similarity, establish variant functional assays, and test phenotypic reversibility. These data can serve to facilitate precision medicine drug screens in future work. (B) Databases used during risk gene list concatenation and ortholog identification. Whole-Genome Sequencing (WGS), Whole-Exome Sequencing (WES), Targeted Sequencing (TS). (C) A greater proportion of high-confidence ASD-associated genes have C. elegans orthologs (blue, 72% 18/25 SFARI Gene, 78% 80/102 Satterstrom et al. 2019) compared to all human genes (gray, 53% 10,678/23,010). (D) The MWT delivers stimuli and performs image acquisition, object selection, and parameter selection in real time while choreography software extracts phenotypic information offline. (panels) I) Petri plate of C. elegans II) C. elegans selected for analysis III) Digital representation illustrating phenotypic detail. An example behavior scored by the MWT: the C. elegans response to a mechanosensory tap to the side of the plate is brief backwards locomotion (from III to IV). Scale bars: 1cm, 1cm, 0.25mm, 0.25mm from I-IV. (E) Phenomic characterization paradigm plotted alongside a single phenotype, absolute movement speed. Following a 5min acclimation phase a further 5min period is recorded from which multiple measures of morphology and baseline locomotion are extracted. Beginning at 10min 30 mechanosensory stimuli are delivered at a 10s ISI to which the animals initially reverse and then habituate, allowing for assessment of multiple measures of initial sensitivity and learning. This is followed by a 5min recovery before administering a 31st stimulus to assess spontaneous recovery from habituation/short-term memory retention. F-I) Multiple measures of habituation of the same reversal responses exhibit different extents and rates of response decrement. Data shown as mean±s.e.m using number of plates as n.

#### 2.2.1 ASD-associated genes are highly conserved to C. elegans

We identified *C. elegans* orthologs of the 102 genes associated with increased risk for ASD (FDR  $\leq 0.1$ ) according to Satterstrom et al. (2019), the largest exome sequencing study of ASD to date (35,584 individuals, 11,986 ASD cases) (Satterstrom et al., 2019). We also identified numerous orthologs from SFARI Gene (<u>https://gene.sfari.org/</u>) "Syndromic" and "High confidence" categories and the top 50 ASD-associated genes by variant count from our ongoing meta-analysis of rare variants, Varicarta (<u>https://varicarta.msl.ubc.ca/</u>) (Belmadani et al., 2019). Of note, 72% (18/25) of genes in the SFARI High confidence category and 78% (80/102) of genes associated with ASD by Satterstrom et al. (2019) have a clear *C. elegans* ortholog according to OrthoList2 and the Alliance of Genome Resources, compendiums of human-*C. elegans* orthology prediction algorithm results. This is substantially higher than the 53% estimated genome-wide orthology between humans and *C. elegans* (10,678/23,010) (Kim et al., 2018) suggesting an exceptionally high conservation of biological processes core to ASD pathology (**Fig. 2.1B and C**).

Rapid advances in gene discovery and orthology prediction have altered the gene lists used during the course of this project. Despite this challenge, our list covers 92% (58/63) of the most strongly-associated genes from Satterstrom et al. and 100% (15/15) of the SFARI Gene high confidence category genes (**Supplemental dataset 2.1** and **Supplemental dataset 2.2**) that had an assay-suitable ortholog deletion or severe missense allele available. This mix of currently defined high- and mid-confidence ASD-associated genes offered a unique opportunity to study putative ASD-associated genes of unknown function alongside genes with known roles in neurodevelopment and sensory processing (See **Supplemental dataset 2.1** and **Supplemental dataset 2.2** for a complete listing of characterized strains, orthology relationships, and ASDassociation confidence).

We then used our MWT (**Fig. 2.1D**) to systematically characterize the 135 *C. elegans* strains with a mutation in an ortholog of an ASD-associated gene (98 strains with mutations in different genes and 37 strains with additional alleles of a subset of these genes, see **Supplemental dataset 2.1** and **Supplemental dataset 2.2**). We developed software to measure a comprehensive range of parameters while animals were subjected to an automated short-term

habituation learning behavioral paradigm (**Fig. 2.1E-I**, and see **Materials and Methods** for details). We measured habituation because it is a sensitive *in vivo* assay of synaptic function and behavioral plasticity. Habituation is also impaired in individuals with ASD, and abnormalities in tactile sensitivity are present in >95% of cases (McDiarmid et al., 2017; Orefice et al., 2016; Sinclair et al., 2016). The degree of habituation impairment in ASD correlates with the severity of social impairment, and recent studies in monogenic mouse models of ASD suggest peripheral tactile hypersensitivity and impaired habituation precede, and may even lead, to more complex cognitive and social impairments (McDiarmid et al., 2017; Orefice et al., 2016; Sinclair et al., 2016).

We wrote custom scripts to extract 26 quantitative phenotypic features that fall into 5 categories: morphology, baseline locomotion, initial sensitivity, habituation learning, and spontaneous recovery/short-term memory retention (**Fig. 2.2A** and **Table 2.1** for complete feature descriptions). Features were designed to minimize redundancy while maintaining interpretability. For example, we quantified multiple features of habituation because we have previously shown they habituate to different extents and with different time courses (**Fig. 2.1 F-I**) and because growing evidence indicates different components of a single habituating response are mediated by different molecular mechanisms (Ardiel et al., 2018; McDiarmid et al., 2018b, 2019; Randlett et al., 2019). Quantification of correlations between all possible phenotypic feature pairs revealed expected moderate correlations (e.g. between length and width) and clustering that reflected the feature categories we predefined (**Fig. 2.2B**). Digital representations of all strains are freely available in their raw and processed forms

(https://doi.org/10.5683/SP2/FJWIL8), allowing the code to be modified to extract any simple or

compound phenotypic feature of choice.

Feature	Description
Width	The distance spanned perpendicular to the major axis, in mm.
Length	The distance spanned by objects along their major axis (defined to be the axis of a least squares fit), in mm.
Morph Width	The mean width of the central 60% of the object, measured perpendicular to the skeleton. This is only available if the object has a skeleton.
Curve	The average angle, in radians, of points along the object's skeleton taken two apart. This measure is only available if skeletons exist.
Area	The area in square millimeters of objects
Midline Length	The length of the object, in mm, measured along the skeleton; this is only available if skeletons were collected (or outlines were collected, and skeletons were generated with a plugin).
Path Length	The distance that the animal has traveled. This is both signed (forwards is positive, backwards is negative) based on the results of the "bias" output and cumulative. If movement direction cannot be computed, then the length starts again from zero.
Bias	The fraction of objects moving in the dominant movement direction as opposed to the other direction. The dominant direction is labeled 1, the opposite 1, and stationary objects are 0. This is usually inaccurate unless the segment (S) option is used.
Direction	The frequency of changes in direction. Forward motion is 1; motion that backtracks is 1 at the point of turning. This metric is somewhat unreliable unless segment (S) is used.
Angular Speed	The angular speed in radians/second of objects; this is calculated over the same interval as speed, but reports the greatest difference in angle between primary axes over that time
Aspect	The ratio of the width to length

Kink	The angle in radians between the line from the first to third point of the skeleton and the fourth through last points; or the angle between first and fourth to last and third to last to last, whichever is greater. This is similar to the MWT endwiggle measure.	
Feature	Description	
Crab	The speed perpendicular to the length of the body, in mm/sec, averaged as specified by s. Note that the speed along the length of the body is sqrt(speed*speed – crab*crab).	
Baseline Speed	The speed in mm/second of objects. This speed is averaged over the duration given in the s parameter; it is equal to the distance between the most widely spaced pair of points over that period of time (half ahead of and half behind the current timepoint) divided by that duration.	
Initial Reversal Response Probability	The proportion of animals reversing to the first mechanosensory stimulus.	
Initial Reversal Response Duration	The duration of reversal responses to the first mechanosensory stimulus	
Initial Reversal Response Speed	The speed of reversal responses to the first mechanosensory stimulus	
Initial Reversal Response Distance	The distance (mm) of reversal responses to the first mechanosensory stimulus	
Habituation of Response Probability	The difference between the initial reversal response probability and the final habituated level of response probability (average of 28-30th responses).	
Habituation of Response Duration	The difference between the initial reversal response duration and the final habituated level of response duration (average of 28-30th responses).	
Habituation of Response Speed	The difference between the initial reversal response speed and the final habituated level of response speed (average of 28-30th responses).	
Habituation of Response Distance	The difference between the initial reversal response distance and the final habituated level of response distance (average of 28-30th responses).	

Table 2.1 Phenotypic feature descriptions.



Figure 2.2 Phenotypic features display varied moderate correlations that correspond to feature subclasses. (A) List of quantitative machine vision phenotypic features computationally extracted and their predefined subclasses. (B) Hierarchically clustered correlation matrix illustrating varied moderate correlations between features. Pearson's r is shown.

#### 2.2.2 Quantitative phenotypic profiles identify a set of shared functions among ASD-

#### associated genes

One way to visualize this data is as a series of 26 quantitative reverse genetic screens, where each strain is compared to wild-type on each phenotypic feature to determine if and how that gene plays a role in the biological process of interest (**Fig. 2.3** and see **Supplemental dataset 2.3**). Phenotypes shared by a large number of ASD risk genes were of particular interest, as they would suggest common biological functions among these seemingly diverse genes. We found that mutations in the vast majority of ASD-associated gene orthologs decrease multiple measures of size, most notably length (77/135, 57%), in age-synchronized populations (**Fig.** 

**2.3A**). Thus, the majority of genes putatively associated with ASD/neurodevelopmental disorders delay development or impede growth when inactivated.





Plots illustrate sample mean distance of each genotype group from wild-type. Strains outside the 95% confidence interval of the wild-type distribution are labeled and colored blue. A maximum of five strains are labeled in each direction to prevent overplotting. (A) Length by genotype. (B) Initial response duration by genotype. (C) Habituation of response probability by genotype. Error bars represent 95% confidence interval. (F) The number of

strains with normal initial responses that either impair habituation (blue) or enhance habituation (gray) across the four response metrics quantified.

Analysis of baseline locomotion revealed that mutations in a large proportion of ASDassociated gene orthologs increase forward movement bias and path length (distance travelled) (**Fig. 2.4A**). This means many strains with mutations in ASD-associated gene orthologs spent more time moving forward before reversing. The frequency and duration of spontaneous reversals is modulated by the integration of multiple cross-modal sensory inputs, (de Bono and Maricq, 2005) suggesting a widespread imbalance in the neural circuits that control spontaneous forward movement behavior toward increased activity in strains harboring mutations in orthologs of ASD-associated genes.



Figure 2.4 Quantitative phenotypic profiles enable rapid reverse genetic screens to identify shared functions among ASD-associated genes.

All plots illustrate the sample mean distance of each genotype group from wild-type. Strains outside the 95% confidence interval of the wild-type distribution are labeled and colored blue. Only a maximum of five strains are labeled in each direction per feature to prevent overplotting. (A) Movement direction bias by genotype. (B)

Spontaneous recovery of response probability by genotype. Distinct partially overlapping sets of ASD-associated gene orthologs alter initial sensitivity, habituation learning, and spontaneous recovery/memory retention, indicating genetically dissociable underlying mechanisms. Error bars represent 95% confidence intervals.

Assessing the duration, distance, and speed of reversal responses to the first mechanosensory stimulus indicated that mutations in orthologs of ASD-associated genes are approximately equally likely to result in initial hyper- or hypo-responsivity to tactile stimuli (**Fig. 2.4B**; Bidirectional analysis of initial sensitivity was precluded only for response probability due to a ceiling effect with >90% of animals responding to the initial stimulus). These results identify positive and negative roles for multiple ASD-associated genes in modulating mechanosensory processing.

Analysis of habituation learning across 135 strains revealed that mutations in many ASDassociated gene orthologs specifically impair habituation of response probability (**Fig. 2.3C**). Even after filtering out strains with abnormal initial response probability, ASD-associated gene orthologs were approximately 3-fold more likely to impair habituation than enhance it (**Fig. 2.3D** and **Supplemental dataset 2.3**). This is a remarkably specific phenotype that causes the neural circuit to continue responding in an inflexible manner, versus merely impairing the ability to detect stimuli or respond. Importantly, we only observed this phenomenon for response probability; there was no such consistent pattern of habituation impairment for the duration, distance, or speed of the same measured responses (**Fig. 2.3D** and **Supplemental dataset 2.3**). Together, these results suggest many ASD-associated genes normally mediate plasticity of the likelihood, but not vigor, of responding to mechanosensory stimuli.

Finally, we discovered that distinct sets of genes alter initial sensitivity, habituation, and retention of the same component of the same behavioral response (**Fig. 2.3B-C, Fig. 2.4**, and **Supplemental dataset 2.3**). Adding to the complexity, we also found different genes affect different components of the same behavior (e.g. some genes affect only response duration but not probability and vice versa). These results support the hypothesis that habituation is controlled by several dissociable genetic mechanisms (McDiarmid et al., 2019), and underscore the need to assess multiple complex phenotypes to understand the functions of ASD-associated genes.

## 2.2.3 Phenotypic profiles of strains with mutations in ASD-associated genes define shared and unique functions

In addition to visualizing scores of all strains on each phenotype, one can also visualize the scores of each strain on all phenotypes, or the 'phenotypic profile' for each strain (**Fig. 2.5** and **Supplemental dataset 2.4**). Classical uncoordinated "unc" mutants, such as the calcium channel subunit *CACNA1E•unc-2(gk366)*, displayed the most severe phenotypic profiles, consistent with their documented roles in neuronal development and function (**Fig. 2.5A**). Note that throughout the manuscript the "•" symbol is used to represent the human-to-*C. elegans* orthology relationship of interest, e.g. *GAPVD1•rme-6(b1014)*.

Phenotypic profiles also revealed previously unknown phenotypes even in wellcharacterized mutants. For example,  $\beta$ -catenin has well-known roles in development, so it follows that *CTNNB1•bar-1(ga80)* mutants are smaller than wild-type animals. But, here we identify a previously unknown role for  $\beta$ -catenin in habituation learning, evidenced by a profound deficit in habituation of response probability in *CTNNB1•bar-1(ga80)* mutants (**Fig.**  **2.5B**). Animals carrying mutations in *KAT2B•pcaf-1(tm3318)*, involved in CREB related transcription co-activation and acetyltransferase activity, have relatively normal baseline locomotion but profound alterations in sensory and learning behaviors that are only revealed through stimulation (**Fig. 2.5C**). Conversely and more surprisingly, animals carrying mutations in genes such as *KCNJ10•irk-3(tm7416)*, an inward rectifying potassium channel, have relatively normal sensory and learning behaviors despite profoundly abnormal morphology and baseline locomotion (**Fig. 2.5D**). Finally, several strains were not significantly different from wild-type on almost all phenotypic features, such as the membrane palmitoylated protein encoded by *MPP6•C50F2.8(ok533)* (**Fig. 2.6** and **Supplemental dataset 2.4**). Together, these results indicate a remarkable degree of phenotypic modularity and provide a catalog of the unique phenotypic profiles of ASD-associated genes.



Feature class: Morphology Baseline locomotion Initial sensitivity Habituation learning Spontaneous recovery

#### Figure 2.5 Phenotypic profiles of strains with mutations in ASD-associated genes define shared and unique

#### functions.

(A) Phenoytpic profile for *CACNA1E•unc-2(gk366)*. (B) Phenotypic profile for *CTNNB1•bar-1(ga80)*. (C)

Phenotypic profile for KAT2B•pcaf-1(tm3318). (D) Phenotypic profile for KCNJ10•irk-3(tm7416). Bars represent

directional t-statistics for each phenotypic feature listed across the y-axis. Color coding reflects feature

classification.





### Figure 2.6 Phenotypic profiles of strains with mutations in ASD-associated genes define shared and unique functions and phenotypic modularity.

(A) Phenotypic profile for *MPP6*•*C50F2.8(ok533)* exhibiting relatively minor phenotypic alterations. Bars represent directional t-statistics for each phenotypic feature listed across the y-axis. Color coding reflects feature classification.

#### 2.2.4 A phenomic database of strains with mutations in ASD-associated genes

**Fig. 2.7A** summarizes the scores of ~27,000 animals (~200 animals/genotype) across 135 genotypes and 26 phenotypes for a total of 3,510 *in vivo* genotype-to-phenotype assessments (**Fig. 2.7A, Supplemental dataset 2.5**, and **Supplemental dataset 2.6**). We included several annotations on this heat map and reverse genetic screen plots to denote current estimates of ASD-association confidence and specificity (**Fig. 2.7A, Supplemental dataset 2.5**, and **Supplemental dataset 2.6**). For the phenotypes quantified and strains tested, no single phenotype is affected by all putative ASD-associated gene orthologs (**Fig. 2.7A**). However, all

strains were significantly different than wild-type on at least one phenotypic metric (**Fig. 2.7B**). A larger number of strains displayed altered morphological and baseline locomotion phenotypes compared to sensory or learning phenotypes (**Fig. 2.7B**). Finally, our results indicate that distinct, partially overlapping sets of genes influence the different classes of phenotypes (e.g. naïve sensitivity and habituation learning are influenced by different sets of ASD-associated genes) (**Fig. 2.7A and B**).



Figure 2.7 A phenomic database of strains with mutations in ASD-associated genes.

(A) Heat map summarizing the phenotypic profiles of 135 strains harboring a mutation in an ortholog of an ASDassociated gene. Cells represent directional t-statistics from comparisons to wild-type controls. T-statistics are clipped at 20 and only cells significant at FDR < 0.1 are colored for ease of interpretation. Interactive heat maps illustrating the full range of t-statistics can be found in Markdown 3. ASD<sub>p</sub> = ASD predominant, ASD<sub>NDD</sub> = ASD & Neurodevelopmental disorders (Satterstrom et al., 2019). (B) The number of significantly different features for each strain tested. Stacked bars are color coded according to feature sub-classes.

While there are many potential downstream uses for this database, we endeavored to illustrate a few of the most promising applications. First, we clustered the mutant strains based on their similarity in overall phenotypic profiles with the hypothesis, supported by recent largescale model organism phenotyping efforts, that phenotypic similarity would enrich for functional genetic interactions among ASD-associated genes. However, while it has been successful for C. elegans morphology and baseline locomotion profiles in the past, (Brown and de Bivort, 2018; Geng et al., 2003; Green et al., 2011; San-Miguel et al., 2016; Sieburth et al., 2005; Yemini et al., 2013; Yu et al., 2013) a large-scale phenotypic clustering approach has not been attempted in combination with complex sensory and learning phenotypes. Before clustering, we confirmed the sensitivity and consistency of our phenotypic measures by examining the correlation of independently generated alleles targeting the same gene within the set of strains we tested. Our sample contained 33 genes represented by multiple independently generated alleles (29 genes with secondary alleles and 4 with both secondary and tertiary alleles targeting the same gene). Analysis of the distribution of the overall phenomic correlations between same gene independent allele pairs revealed the average correlation was indeed higher than all other possible gene pairs (n = 9004) and that the same gene allele pair distribution was skewed towards highly positive correlations (Fig. 2.8A). We also included heat maps illustrating only genotype-phenotype relationships seen in all independent alleles of a given gene (Supplemental dataset 2.6).



Figure 2.8 Phenotypic correlations and t-SNE clustering of 135 strains harboring mutations in ASDassociated gene orthologs.

(A) Density plot illustrating the distribution of overall phenotypic profile Pearson correlation coefficients between independent alleles in the same gene (blue) and all other possible gene pairs (gray). (B) 2D (left) and 3D (right) t-SNE plots illustrating the distance between the 135 strains harboring mutations in ASD-associated gene orthologs (blue) and wild-type (black).

We then used several clustering methods to group genes based on phenotypic similarity to predict genetic interactions. Hierarchical clustering accurately identified several well-known interactions, such as those between voltage-gated calcium channels *CACNA1E•unc-2*, Syntaxin bindings proteins *STXBP1•unc-18*, and *Rab3* binding proteins *RIMS1•unc-10* (**Fig. 2.7A**, **Supplemental dataset 2.5**, and **Supplemental dataset 2.9**) (Kaeser et al., 2011; Tong et al., 2017). Our analysis also confirmed recently discovered interactions, such as those between the dual specificity kinase *DYRK1A•mbk-1* and the histone acetyltransferase *CREBBP•cbp-1*, and predicted several new interactions (**Fig. 2.7A**, **Supplemental dataset 2.5**, and **Supplemental dataset 2.9**) (Li et al., 2018). We then investigated the overall phenotypic architecture of ASDassociated genes. t-distributed Stochastic Neighbor Embedding (t-SNE) and multiple other clustering methods revealed that ASD-associated genes were largely continuously distributed in phenotypic space (**Fig. 2.8B**). We found no evidence for a small number of highly discrete clusters that would suggest distinct 'phenotypic classes' of ASD-associated genes (**Fig. 2.7A** and **Fig. 2.8B**).

Next, we used epistasis analysis to test some of our interactions predicted by phenotypic proximity in vivo. Using sensory and habituation learning features for hierarchical clustering revealed two high-confidence clusters with members that display impaired habituation of response probability and hyper-responsivity to mechanosensory stimuli (increased initial reversal response duration) (Fig. 2.9A and B, Fig. 2.10A and B and Fig. 2.7A, and Supplemental dataset 2.9). Genes within these clusters were selected for epistasis analysis based on confidence of ASD-association and confirmation of genotype-to-phenotype relationships using analysis of a second allele or transformation rescue. Crossing strains within the same cluster revealed a functional interaction between the Chromodomain Helicase DNA Binding Protein CHD8•chd-7(gk306) and the GTPase-activating protein GAPVD1•rme-6(b1014); the impairment in habituation of response probability of double mutants was not significantly different from single mutants, suggesting they function in the same genetic pathway to mediate short-term behavioral plasticity (Fig. 2.9C). A second allele of GAPVD1•rme-6(tm6649), and an additional allele of CHD8•chd-7(gk209) tested after the large-scale characterization also displayed the same phenotypic profile, confirming genotype-to-phenotype relationships (Fig. 2.10). CHD8 is a highconfidence ASD-associated gene whereas GAPVD1 is relatively low confidence, yet when they are inactivated in a model organism they cause strikingly similar phenotypic profiles and function together to promote normal habituation learning.

Crossing between clusters revealed that *CHD8•chd-7(gk306)* and the sole *C. elegans* ortholog of vertebrate neuroligins *NLGN1/2/3/4X•nlg-1(ok259)* function in parallel genetic pathways, with double mutants exhibiting additive impairments in habituation (**Fig. 2.9C**).
Interestingly, we also discovered a synthetic lethal interaction between *CHD8•chd-7(gk306)* and *CTNNB1•bar-1(ga80)*, suggesting *CHD8* can function independently of its canonical role in mediating the Wnt/ $\beta$ -catenin signaling pathway (Sakamoto et al., 2000). These results are consistent with the observation that Wnt/ $\beta$ -catenin targets are neither upregulated nor the cause of lethality in *CHD8* homozygous knockout mice (Nishiyama et al., 2009), suggesting Wnt-independent functions of *CHD8* are present in multiple species (Barnard et al., 2015). We combined our results with the recent observation that Wnt/ $\beta$ -catenin signaling increases expression and synaptic clustering of *NLGN3* (Medina et al., 2018) to draw parallel pathways underlying impaired habituation (**Fig. 2.9D**). Together, these results demonstrate that systematic phenotypic clustering and epistasis analysis of complex sensory and learning phenotypes present an *in vivo* approach to map functional genetic network interactions among ASD-associated genes and prioritize candidates that would be missed by focusing on currently high-confidence genes.





(A and B) Hierarchical clustering based on sensory and learning features identified two sets of genes with members who display impairments in habituation of response probability and hyper-responsivity to mechanosensory stimuli (increased initial reversal response duration). Rectangles outline the largest clusters with AU p-values >95%. (C) Final reversal probability across genotypes (average of the 28th-30th responses). Dots represent individual plate means, horizontal lines represent median of plate replicates. \*\*\*p<0.001, binomial logistic regression followed by Tukey's Honestly Significant Difference (HSD) criterion was used to determine significance of the habituated level (proportion reversing at tap 30) for each pair of strains. (D) Parallel genetic pathways of ASD-associated genes underlie impaired habituation learning.



Figure 2.10 Sensory and learning phenotypic profiles of cluster members.

(A) Sensory and learning phenotypic profile for  $NLGN1/2/3/4X \cdot nlg - 1(ok259)$  and  $CTNNB1 \cdot bar - 1(ga80)$  mutants. Each point represents a reversal response. Data are shown as mean±s.e.m using the number of plates as n. (B) Sensory and learning phenotypic profile for  $CHD8 \cdot chd - 7(gk306)$  and  $GAPVD1 \cdot rme - 6(b1014)$  mutants. Each point represents a reversal response. Data are shown as mean±s.e.m using the number of plates as n.



Figure 2.11 Second alleles of *GAPVD1•rme-6(tm6649)* and *CHD8•chd-7(gk209)* also display increased initial reversal response duration and impaired habituation of response probability.

(A) Sensory and learning phenotypic profile for *GAPVD1•rme-6(tm6659)* and (B) *CHD8•chd-7(gk209)* mutants. Data are shown as mean±s.e.m using the number of plates as n. \*\*\*p<0.001, binomial logistic regression followed by Tukey's HSD criterion was used to determine significance of the habituated level (proportion reversing at tap 30) for each pair of strains, n.s., not significant.

#### 2.2.5 Phenomic profiles can be leveraged for *in vivo* variant functional assays

Another challenge facing ASD is the inability to interpret the functional consequences of the large number of rare variants of uncertain significance (Starita et al., 2017). Many variants found in ASD are so rare they preclude traditional genetic approaches to infer pathogenicity (Karczewski et al., 2019; Starita et al., 2017). Experimental variant functional assays, where the variant is introduced into a model system, can be combined with computational approaches to guide clinical assessment, but there is a paucity of such assays for most ASD-associated genes due to a lack of understanding of their normal biological functions (Starita et al., 2017; Sun et al., 2016; Wong et al., 2019). Here, we show how the phenomic functional data we generated by studying inactivating mutations in ASD-associated gene orthologs can be combined with the genetic tractability of *C. elegans* to develop *in vivo* variant functional assays using complex sensory and learning behaviors as a readout.

In our large-scale characterization we discovered that mutations in the *C. elegans* ortholog of vertebrate neuroligins cause impairments in habituation of response probability. We generated a wild-type *nlg-1::YFP* fusion transgene and found that expression of this construct in *nlg-1(ok259)* deletion mutants was sufficient to restore normal habituation learning (**Fig. 2.12A-C**). We then used this transgenic rescue to assess the functional consequences of mutations equivalent to four conserved ASD-associated neuroligin missense variants by testing their ability to rescue, revealing varied partial loss-of-function (**Fig. 2.12D-G**). We found similar results for these variants on three additional behavioral functional assays: octanol avoidance chemotaxis, thermotaxis, and sensory integration (**Fig. 2.13**). Each of these functional assays involve distinct neural circuits suggesting a general partial loss-of-function mechanism due to ASD-associated

neuroligin variants in whole organism behavior. All ASD-associated variants were expressed at similar levels and properly localized to synapses *in vivo*, likely ruling out a simple pathogenic mechanism of improper trafficking or severely reduced cellular abundance (**Fig. 2.12B**). We also characterized two additional neuroligin variant lines generated with a precise CRISPR-Cas9 genome engineering method (Wong et al., 2019) and again observed varied partial loss-of-function (**Fig. 2.12H**). Finally, we assessed five additional missense variants in *chd-7* (orthologous to *CHD8* and *CHD7*) also generated with CRISPR-Cas9 and observed that P165L•P253L and R2627Q•R2624Q displayed wild-type function and habituated normally, while L834P•L1220P, G996S•G1225S and L1257R•L1487R displayed impairments in habituation learning (**Fig. 2.12I-K**). Importantly, all ASD-associated gene orthologs, including those with no prior functional annotation, were significantly different from wild-type animals on at least one phenotypic metric in our characterization (**Fig. 2.7B**), suggesting this will be a broadly applicable approach to decipher variants of uncertain significance.



Figure 2.12 Functional assessment of ASD-associated missense variants.

(A) Schematic of the *nlg-1(ok259)* deletion allele and NLG-1::YFP transgene. Lollipops indicate the approximate locations of the equivalent ASD-associated variants assessed. Note the R430C variant in *C. elegans* NLG-1 corresponds to a mutation in human *NLGN3* whereas the R62W, V397M, and R714C NLG-1 variants correspond to mutations in human *NLGN4X*. (B) All neuroligin variants were expressed at similar levels and localized properly to synapses in the nerve ring and nerve cords. A = anterior, P = posterior, D = dorsal, V = ventral, NR = nerve ring, scale bar = 0.02 mm. (C) Expression of wild-type NLG-1::YFP rescued *nlg-1(ok259)* deletion mutant impaired habituation of response probability. (D-G) Each neuroligin variant was scored for its ability to rescue impaired habituation of response probability, revealing varied partial loss-of-function. (H) Missense variants in *nlg-1* 

generated with CRISPR-Cas9 also displayed varied partial loss-of-function with corresponding impairments in habituation. Note the G59R and A268T variants in *C. elegans* NLG-1 correspond to variants in human *NLGN4X*. (I-K) P165L•P253L and R2627Q•R2624Q missense variants in *C. elegans* CHD-7 did not display habituation impairments, while L834P•L1220P, G996S•G1225S and L1257R•L1487R displayed impairments in habituation. Note that the G1225S, L1487R, and R2624Q missense variants in *C. elegans* CHD-7 correspond to variants in human *CHD7*, while L1220P and P253L correspond to variants in human *CHD8*. (C-K) Data shown as mean±s.e.m using plates as n. \*\*\*p<0.001, binomial logistic regression followed by Tukey's HSD criterion was used to determine significance of the habituated level (proportion reversing at tap 30) for each pair of strains, n.s., not significant.



Figure 2.13 Additional functional assays for ASD-associated neuroligin missense variants.

(A-D) Each ASD-associated neuroligin variant was scored for its ability to rescue spontaneous reversal frequency (A), thermotaxis (B), sensory integration (C), and octanol avoidance chemotaxis (D) in the *nlg-1(ok259)* deletion mutant, again revealing varied partial loss-of-function. Dots represent individual worm reversal frequencies (A) or plate means (B-D), horizontal lines represent median of replicates. (A-D) \*\*\*\*p<0.0001, one-way ANOVA followed by Tukey's HSD criterion was used to determine significance, n.s., not significant.

# 2.2.6 CRISPR-Cas9 auxin inducible degradation reveals *nlg-1* phenotypes are reversible by adult-specific re-expression

Historically it was believed that the neurodevelopmental insults caused by monogenic risk factors for ASD were so severe that they would not be reversible in adulthood, and thus any treatment would have to be administered early in development to be effective. The seminal discoveries that phenotypes caused by mutations in MECP2 could be reversed by transgenic expression of the protein in adulthood, after a presumed critical developmental period had passed, offered tremendous hope for families suffering from Rett syndrome (Guy et al., 2007). Conversely, the observation that inactivation of MECP2, neuroligins, and several other ASDassociated genes in adulthood could cause severe electrophysiological and behavioral impairments demonstrated that they were not simply neurodevelopmental genes, and instead that they function throughout the lifespan to promote normal sensory and learning behaviors (Jiang et al., 2017; Liang et al., 2015; McGraw et al., 2011). Such conditional rescue and inactivation experiments are valuable to ASD research, but they are expensive, technically difficult, and time consuming in mammalian model systems and are thus not yet practical for many genes. With C. elegans, however, conditional and reversible protein depletion is precise, rapid, and straightforward owing to their genetic tractability and the recent advent of CRISPR-Cas9 auxin inducible degradation (AID) (Zhang et al., 2015). AID relies on transgenic expression of TIR1, an inducible E3 ubiquitin ligase that targets any protein with an AID degron peptide tag for degradation by the proteasome only when in the presence of its activating hormone auxin (Fig. 2.14). Moreover, upon removal from auxin TIR1 is inactivated, allowing rapid reconstitution of protein expression in large populations of freely behaving and intact animals without the need for surgery or vector delivery (Zhang et al., 2015).

Despite extensive study, adult transgenic rescue tests of behavioral phenotype reversibility have not been reported for any member of the vertebrate neuroligin family (Baudouin et al., 2012; Tong et al., 2017). We used the CRISPR-Cas9 Dual Marker Selection (DMS) cassette genome editing strategy to insert GFP and a short AID degron peptide tag into the endogenous C. elegans neuroligin locus 13 residues before the stop codon (Fig. 2.14A, see Materials and Methods for details). This fusion protein localized properly to synapses and is fully functional – there were no habituation impairments or gross neuroanatomical abnormalities in genome-edited animals (Fig. 2.14C and D and Fig. 2.15). We then crossed this strain into animals expressing TIR1 under a ubiquitous promoter and observed that treatment with 0.025 mM auxin for <10.0 hours was sufficient for complete degradation of the functional NLG-1::AID::GFP fusion protein at multiple life stages (Fig. 2.14D). Moreover NLG-1::AID::GFP was partially recovered 48 hours after removal from auxin (Fig. 2.14D). We used this approach to test several conditional rescue and inactivation groups simultaneously (Fig. 2.14E). Animals reared on auxin for continuous degradation displayed habituation impairments equivalent to those of *nlg-1(ok259)* null mutants, confirming effective degradation (Fig. 2.14F). Importantly, wild-type animals continuously exposed to auxin displayed no morphological or behavioral abnormalities (Fig. 2.14G). Strikingly, we observed that for animals reared on auxin that degraded NLG-1 throughout development, adult-specific expression of neuroligin was sufficient to partially rescue the habituation impairment phenotype (Fig. 2.14H). Surprisingly, adultspecific degradation of neuroligin did not lead to impaired habituation (Fig. 2.14I). These results indicate a critical role for neuroligin in generating a circuit properly tuned for normal mechanosensory processing and short-term behavioral plasticity. Further, they suggest

behavioral disruptions caused by developmental loss of neuroligins might be at least partially reversible by adult expression. Given the speed of machine vision phenotyping, and relative ease of CRISPR-Cas9 genome editing in *C. elegans*, AID represents a scalable approach that can be applied to diverse ASD-associated genes.



Figure 2.14 CRISPR-Cas9 auxin inducible degradation reveals phenotypes caused by developmental loss of neuroligin can be rescued by adult re-expression.

(A) The modified DMS cassette genome editing strategy used to insert GFP and a short degron tag into the endogenous neuroligin locus. The maroon line in the repair template indicates the location of an engineered silent

mutation in the protospacer adjacent motif to prevent cleavage of exogenous DNA. (B) Schematic of the NLG-1::AID::GFP transgene. In the presence of auxin, TIR1 is activated targeting the fusion protein for degradation. Following auxin treatment TIR1 is inactivated allowing conditional degradation and re-expression of the fusion protein. (C) The fusion protein is fully functional; *NLG-1::AID::GFP* animals did not display habituation impairments before (top) or after (bottom) DMS cassette excision. (D) The fusion protein localizes properly to synapses in the nerve ring and nerve cords (top). Treatment with 0.025 mM auxin is sufficient for complete degradation of the fusion protein (middle) that is reversible 48 hours after removal from auxin (bottom). A = anterior, P = posterior, D = dorsal, V = ventral, NR = nerve ring, scale bar = 0.02 mm. (E) Period of auxin administration for each group. (F) Continuous auxin administration recapitulated impairments in habituation. (H) Adult-specific re-expression of neuroligin partially rescued impaired habituation. Solid blue circles represent continuous auxin administration group (as in F). (I) Adult-specific degradation of neuroligin did not induce habituation impairments (Solid blue circles as in F). (C, F-I) \*\*\*p<0.001, binomial logistic regression followed by Tukey's HSD criterion was used to determine significance of the habituated level (proportion reversing at tap 30) for each pair of strains, n.s., not significant. Data shown as mean±s.e.m using plates as n. Α



#### Figure 2.15 Synaptic localization of neuroligin transgenes.

(A) Immunofluorescent staining of *nlg-1(ok259)* animals expressing a NLG-1::YFP transgene driven by a musclespecific promoter, illustrating apposition of the post-synaptic NLG-1::YFP with the presynaptic vesicular GABA transporter (VGAT, a marker of GABAergic synapses). Shown is the dorsal nerve cord of a young adult stained with antibodies to GFP, which also recognize YFP (top, green), VGAT (middle, blue), and the merged image (bottom), scale bar = 1  $\mu$ m. (B) The NLG-1::AID::GFP endogenous transgene is apposed to a RAB-3::mCherry synaptic vesicle marker driven by an *flp-13* promoter in DD class GABAergic neurons (DD5 output region is shown), scale bar = 0.02 mm.

# 2.3 Discussion

We quantified 26 phenotypes for >27,000 animals across 135 genotypes and identified hundreds of genotype-phenotype relationships. When combined with a recent aggregation of all other machine vision behavioral datasets collected in this organism to date, our database more than doubles the number of tracked animals reported in the literature from ~15,000 to >30,000 (Javer et al., 2018). By making the raw and processed data available online, we have created open and shareable phenotypic atlas of *C. elegans* strains carrying mutations in orthologs of ASD-associated genes. We have shown how this database can be used to identify shared functions, map genetic networks, gauge missense variant effect, and determine reversibility of phenotypic disruptions.

We found that the vast majority of strains with mutations in ASD-associated genes displayed delayed development or growth impediments, a finding consistent with a recent largescale analysis of inactivating mutations in constrained genes in humans (Ganna et al., 2018). We also found that a large number of ASD-associated genes impair habituation, a finding consistent with a recently reported large-scale analysis of ASD- and Intellectual Disability/Developmental Delay-associated genes in *Drosophila* and humans (Fenckova et al., 2019). Interestingly, our data show ASD-associated genes specifically impair habituation of response probability. These results provide a genotype-to-phenotype list (**Fig. 2.3C**) that may hint at a common pathological mechanism that impairs plasticity of a neural circuit's decision to respond without altering response vigor. Indeed, these *Drosophila* and *C. elegans* results may reflect a behavioral outcome of circuit-level hyperexcitability recently discovered in several human iPSC derived neuronal culture models of a number of monogenic ASD risk factors (Deneault et al., 2018,

2019). The results also provide a potentially plausible explanation for inconsistent reports of impaired habituation in humans, which variably employ diverse response metrics most often without genetic stratification of patient populations (McDiarmid et al., 2017; Stessman et al., 2014). While these shared phenotypes are exciting, our results and those of several other model systems (Amin and Paşca, 2018; Deneault et al., 2018; Ellegood et al., 2015; Fenckova et al., 2019; Stessman et al., 2014; Zoghbi and Bear, 2012) reveal a remarkable diversity in phenotypic disruptions, suggesting single phenotype functional validation and characterization efforts will be insufficient to capture the complex multi-facetted phenotypic disruptions that stem from mutations in ASD-associated genes.

# 2.3.1 Phenomic clustering and epistasis to map genetic networks among ASD-associated genes *in vivo*

We combined phenotypic clustering with epistasis analysis to map functional genetic networks among ASD-associated genes and identified parallel genetic networks centered on *CHD8•chd-7* and *NLGN1/2/3/4X•nlg-1* that underlie hyper-responsivity and impaired habituation. These results provide *in vivo* functional support to the proposed broad categorization of ASD-associated genes into those involved in synaptic neuronal communication (neuroligin) and gene expression regulation (*CHD8*) (Sanders et al., 2015; Satterstrom et al., 2019). An exciting question for future research will be to determine how well the phenomic functional interactions delineated here map on to biochemical interactions. Indeed, *in vivo* genetic networks will serve as an important benchmark from which to compare and extend existing networks of ASD-associated genes. Phenomic clustering will be particularly useful for capturing long-range functional interactions between proteins expressed in different cells or even different points in

development, which cannot be detected by measures of direct protein-protein interactions or coexpression.

The strains we characterized were generally continuously distributed in phenotypic space, and we did not detect highly separated discrete phenotypic clusters. Even in this carefully controlled genetic background and environment we found no evidence for distinct 'phenotypic classes' of ASD-associated genes. This is in contrast to the molecular level, where multiple studies and our work suggest there are functionally distinct classes of ASD-associated genes (de la Torre-Ubieta et al., 2016; Satterstrom et al., 2019; Winden et al., 2018). While phenotypic profiles were continuously distributed, certain genes were closer to each other than others in a manner that reflected underlying molecular interactions. The observation that ASDs are a group of etiologically distinct and variably phenotypically similar disorders provides further motivation for tailored treatments designed on the bases of shared molecular pathway disruptions (de la Torre-Ubieta et al., 2016; Winden et al., 2018).

For the majority of genes studied in this work there is a clear *C. elegans* ortholog with a loss-of-function mutation available from stock centers, often times in an outcrossed strain to minimize the effects of background mutations. For some genes orthology relationships are less clear and there was a paucity of outcrossed deletion alleles making the functional annotations added on to these genes provisional at this point. This limitation becomes less important when the same phenotypes were observed in two strains carrying independently generated alleles of the same gene, when the phenotypes were rescued by expression of the wild-type allele, or when the phenotypes were observed in a strain generated with CRISPR-Cas9 genome editing. As such,

the reader should gauge the confidence of the genotype-to-phenotype relationships reported for individual strains accordingly when designing follow up studies – an approach that has proven to be straightforward and extremely valuable when applied to previous large-scale phenotyping of baseline locomotion in similarly generated *C. elegans* knockout libraries (Brown et al., 2013; Javer et al., 2018; Yemini et al., 2013; Yu et al., 2013).

It is also important to consider whether the functions of ASD-associated genes discovered here will be conserved to their human orthologs, and if they are conserved, whether they display specificity to ASD. Regardless of the qualitative diagnostic label applied, understanding the functions of these poorly characterized neurodevelopmental disorder-associated genes will be valuable for understanding pathology and designing treatments. Further, understanding gene functions that may not be specific to ASD pathology (e.g. seizures) can still be useful in diagnosis or treating comorbidities. Even in cases where a given gene may be implicated in multiple disorders, or where a deletion allele may not be the best model of genetic etiology, understanding the consequences of loss-of-function in that gene can offer valuable insights that aid in the interpretation of the functional impact of particular risk variants, that may be specific to a given disease pathology. Regarding conservation of function, several of the shared functions for ASD-associated gene orthologs identified here, such as promoting normal development or habituation of response probability, have concurrently been discovered in higher model organisms and human iPSCs, suggesting many of the other functions reported here will also be conserved. Indeed, our analysis and previous work in other model systems suggest that ASDassociated genes are highly conserved throughout evolution (Shpigler et al., 2017). It is also worth reiterating that many human genes have been shown to functionally replace their C.

*elegans* orthologs. For example, human *NLGN1* and *NLGN4* have both been shown to rescue multiple sensory abnormalities caused by loss of *C. elegans nlg-1* (Calahorro and Ruiz-Rubio, 2012; Hunter, 2011). We have also recently shown that directly replacing *daf-18* with only a single copy of its human ASD-associated ortholog *PTEN* at the endogenous locus using CRISPR-Cas9 is able to rescue multiple sensory abnormalities caused by complete deletion of *daf-18* (McDiarmid et al., 2018a). All model systems have their relative strengths and weaknesses, and the fastest and most generalizable insights in ASD research will undoubtedly come from synthesis of large amounts of information derived from diverse model systems combined with information from studies of individuals with ASD.

#### 2.3.2 Phenomic profiles can be leveraged for *in vivo* assays of missense variant effect

We used neuroligin as a proof-of-principle to show how our phenomic profiles can be leveraged to establish *in vivo* variant functional assays. We found that all neuroligin variants tested displayed variable partial loss-of-function despite effective expression and proper subcellular localization. Further, we observed that variant functional results were consistent across multiple behavioral functional assays involving distinct neural circuits. While the neurobiological mechanisms underlying the behavioral effects of ASD-associated neuroligin variants currently remain controversial (different studies have implicated altered receptor trafficking, circuit specific electrophysiological imbalances, etc.) (Bemben et al., 2015; Chadman et al., 2008; Chanda et al., 2016; Etherton et al., 2011; Nakanishi et al., 2017; Tabuchi et al., 2007) our results suggest a treatment should be designed to increase neuroligin function while taking into account the presence of an existing dysfunctional protein. We also assessed several missense variants in *chd-7* and observed habituation impairments in a subset, further prioritizing

these candidates for future study (Wong et al., 2019). Importantly, all characterized strains were significantly different from wild-type on at least one metric, allowing for many diverse *in vivo* variant functional assessments moving forward. Many genes were significantly different on multiple metrics, giving researchers with interests in particular biological processes the ability to choose a phenotypic functional assay most suited to their needs. A strength of *C. elegans* will continue to be the ability to rapidly assess multiple variants in complex sensory and learning behaviors *in vivo*.

# 2.3.3 Harnessing phenotypic profiles for tests of adult reversibility

We found adult expression of neuroligin could partially reverse the impaired habituation phenotypes of animals that developed without neuroligin. Interestingly, we also found adult-specific inactivation did not produce phenotypic disruptions. These results are surprising, as inactivation of *NLGN1*, *NLGN2* and *NLGN3* in mature vertebrates produce abnormalities in several complex behaviors (Bariselli et al., 2018; Jiang et al., 2017; Liang et al., 2015). However, a plausible explanation could be that neuroligin is only necessary in adulthood for forms of learning and memory that require structural plasticity. Indeed, all studies reporting an adult requirement for neuroligin examined different forms of long-term memory, which in contrast to the short-term memories studied here require *de novo* activity-dependent synapse growth and maturation. Further, the phenotypes observed often did not manifest until weeks after neuroligins were inactivated in adult animals (Liang et al., 2015). These results lead to a model in which neuroligin is sufficient to build a circuit capable of normal sensitivity and short-term habituation at any point throughout the lifespan, but once that circuit has been built its function is no longer required for short-term learning. Continued function of neuroligin would then remain necessary

only for more complex forms of long-term learning and memory that require *de novo* activitydependent growth and maturation of synaptic connections. The *C. elegans* nervous system is thought to have largely invariant connections across individuals, and there is no clear evidence yet of new synapse formation underlying long-term learning in *C. elegans*, making this model currently difficult to test. However, an adult form of experience/activity-dependent neural circuit remodeling where neural activity driving male mating behavior rewires synaptic connections has recently been discovered in *C. elegans*, and neuroligin was identified as a critical component in this process (Hart and Hobert, 2018). These results suggest that some phenotypic alterations due to developmental loss of neuroligin may be reversible in adulthood. More broadly, they provide a rapid and inexpensive strategy where AID can be used to test reversibility of phenotypic disruptions caused by diverse ASD-associated genes and thereby prioritize candidates for further study in less evolutionarily distant model systems with greater translational potential.

# 2.3.4 Conclusions

We have completed a systematic phenomics analysis of ASD-associated genes and identified shared delays in development, hyperactivity, and impairments in habituation learning. Our data adds to the rapidly expanding use of model organism phenomics to discover the functions of poorly characterized genes identified through genomic sequencing (Kochinke et al., 2016; McDiarmid et al., 2018a; Stessman et al., 2017; Thyme et al., 2019; Wangler et al., 2017). We have shown how such data can be used to identify genetic interactions, establish variant functional assays, and develop tests of phenotypic reversibility. This work sets the stage for numerous future experiments using cell-specific and inducible genetic manipulations to delineate each of the specific mechanisms through which mutations in these genes produce their emergent

phenotypic profiles. There is substantial evidence that an insufficient understanding of the biology of many disease-associated genes has prevented the successful development of therapies and that preclinical research is biased towards experimentally well-accessible genes (Stoeger et al., 2018). It is ideal to be systematic and unbiased whenever possible, an opportunity that high-throughput model organisms such as *C. elegans* afford and thus one we should continue to exploit. As we continue to chart the phenotypic landscape of ASD-associated genes, the complicated paths to understanding mechanisms and developing personalized treatments become simpler to navigate.

# 2.4 Materials and Methods

#### 2.4.1 Ortholog identification and strain selection

*C. elegans* orthologs of human ASD-associated genes were identified by querying OrthoList using Ensembl gene IDs, as previously described (Shaye and Greenwald, 2011). **Supplemental dataset 2.1** and **Supplemental dataset 2.2** describe all Orthology relationships used in this study. During the course of this project The Alliance of Genome Resources (https://www.alliancegenome.org/) created a web tool that allows for identification of the "best" matched Ortholog, defined as the ortholog predicted in the queried species by the highest number of gold-standard algorithms. The vast majority of the *C. elegans* orthologs predicted by OrthoList (85%) are also defined as the best ortholog by this new tool (**Dataset S1**). OrthoList was also updated to OrthoList2 during the course of this work (Kim et al., 2018). Again there is large agreement between our orthology predictions and those made using the recently updated OrthoList2 (95% of OrthoList relationships are supported by OrthoList2; **Supplemental dataset** 

**2.1**, note that 48 of the 135 strains were predicted later than the rest using the updated OrthoList2 and Alliance of Genome Resources tools). In situations where multiple human ASD-associated genes share a single *C. elegans* ortholog predicted by OrthoList the single *C. elegans* gene was used to study the larger vertebrate family, as has been done previously (Hu et al., 2012; Pym et al., 2017; Tong et al., 2015). For example, *nlg-1* and *shn-1* are the sole *C. elegans* orthologs of all vertebrate neuroligin and shank family proteins, respectively (**Supplemental dataset 2.1** and **Supplemental dataset 2.2**). Note that throughout the manuscript the "•" symbol is used to represent the human-to-*C. elegans* orthology relationship of interest, e.g. *GAPVD1•rme-6(b1014)*.

Strains harboring mutations in orthologs of ASD-associated genes were identified using WormBase and ordered from the *Caenorhabditis* Genetics Center (CGC), the National BioResource Project of Japan (NBRP), or received from a collaborator following a formal request. Strains carrying putative null and loss-of-function alleles were prioritized (C. elegans Deletion Mutant Consortium, 2012). In many cases multiple strains harboring distinct loss-of-function alleles were characterized. Where such alleles were not available, or where null alleles are known to result in lethality or fecundity defects severe enough to impede high-throughput characterization in our assay a strain carrying a known or predicted deleterious missense mutation was characterized instead. The complete list of all strains and alleles assessed in the initial large-scale characterization are described in **Supplemental dataset 2.1**. It is important to note that in *C. elegans* a single, heavily characterized wild-type strain (N2) is used across laboratories.

#### 2.4.2 Animals

Strains were maintained on NGM (nematode growth medium) plates seeded with *Escherichia coli* strain OP50 according to standard experimental procedures. 96h post-hatch hermaphrodite animals were used for all experiments.

#### 2.4.3 Microbe strains

The E. coli OP50 strain was used as a food source for C. elegans.

# 2.4.4 Behavioral assays

For the mechanosensory habituation paradigm animals were cultured on NGM seeded with *E. coli* (OP50) and age synchronized for behavioral tracking as described previously (Ardiel et al., 2018; Brenner, 1974; Loucks et al., 2019). Animals were synchronized for behavioral testing on Petri plates containing NGM seeded with 50  $\mu$ l of OP50 liquid culture 12-24 hours before use. Five gravid adults were picked to plates and allowed to lay eggs for 3-4 hours before removal. For all Multi-Worm Tracker experiments 3-6 plates (20-100 worms/plate) were run for each strain on each testing day. The animals were maintained in a 20°C incubator for 96 hours.

Our behavioral paradigm (**Fig. 2.1E**) consisted of a 5-minute period to recover from being placed on the tracker followed by a 5 min baseline period from which we computed multiple measures of morphology and baseline locomotion. Beginning at 10 minutes we administered 30 mechanosensory stimuli to the Petri plate holding the animals at a 10 second interstimulus interval (ISI) using an automated push solenoid (**Fig. 2.1A and E**). Animals respond to a mechanosensory stimulus by emitting a reversal response (crawling backwards) allowing us to assess multiple measures of naïve sensitivity (e.g. reversal likelihood, duration, etc.) (**Fig. 2.1E-I**). With repeated stimulation there is a decrease in the likelihood of a reversal, as well as the duration, speed, and distance of reversals (habituation). Following habituation training, we allowed a 5-minute recovery period after which we administered a 31<sup>st</sup> stimulus to gauge spontaneous recovery from short-term habituation - an assay of short-term memory retention (**Fig. 2.1E-I**).

Standard sensory integration, spontaneous reversal, octanol avoidance, and thermotaxis assays were conducted as previously described (Hunter, 2011; Hunter et al., 2010). Briefly, in order to assess thermotaxis, thermal gradients were established by placing a vial of frozen glacial acetic acid (16.7°C) in the center of inverted 10 cm unseeded plates kept in a 25°C incubator (Hedgecock and Russell, 1975). Approximately 50 worms were transferred to each thermal gradient plate and allowed to move freely on the gradient for 30 minutes, after which their positions were scored on an overlay of concentric circles demarking eight equal areas on the plate. Octanol chemotaxis plates were prepared by placing 2 µl of 0.1% octanol on one side of a 10 cm plate (side 'A') and 2 µl of diluent on the other side of the plate (side 'B'). Approximately 50 worms were transferred to the center of each assay plate and allowed to move freely for 20 minutes and then scored for location. Only worms within 2 cm of the test or control spots were scored, and therefore worms in neutral areas were disregarded. A chemotaxis index (C.I.) was calculated using the formula C.I.=(A-B)/(A+B). A positive C.I. value (up to 1.00) indicates attraction, a negative value (down to -1.00) indicates repulsion, and a value near 0.00 indicates neutrality. Spontaneous reversals were measured by transferring worms to unseeded plates, waiting two minutes for equilibration, and visually scoring for changes in direction for 10

minutes. To test the integration of two conflicting (attractive and repulsive) sensory stimuli a repellent barrier consisting of 50  $\mu$ l of 0.5 mM cupric acetate was pipetted in a line across the middle of a 10 cm plate (Ishihara et al., 2002). After approximately 16 hours, a chemo-attractant (2  $\mu$ l of 0.1% diacetyl in ethanol) was placed on one side of the barrier, and approximately 75 worms were transferred to the plate on the opposite side of the barrier. The worms were allowed to move freely on the plate for 30 minutes, at which point the fraction of worms that crossed the barrier was scored.

#### 2.4.5 Multi-Worm Tracker behavioral analysis and statistics

Multi-Worm Tracker software (version 1.2.0.2) was used for stimulus delivery and image acquisition (Swierczek et al., 2011). Phenotypic quantification with Choreography software (version  $1.3.0\_r103552$ ) used "--shadowless", "--minimum-move-body 2", and "--minimum-time 20" filters to restrict the analysis to animals that moved at least 2 body lengths and were tracked for at least 20 s. Standard choreography output commands were used to output morphology and baseline locomotion features (Swierczek et al., 2011). A complete description of the morphology, baseline locomotion, sensory, and habituation learning features can be found in the Multi-Worm Tracker user guide (https://sourceforge.net/projects/mwt/) (Swierczek et al., 2011) and **Table 2.1**. The MeasureReversal plugin was used to identify reversals occurring within 1 s (dt = 1) of the mechanosensory stimulus onset. Comparisons of "final response" comprised the average of the final three stimuli. Custom R scripts organized and summarized Choreography output files. No blinding was necessary because the Multi-Worm Tracker scores behavior objectively. For the initial large-scale characterization (**Figs. 2.1-2.7**), features were pooled across plate replicates for each mutant strain and means were compared to the mean of

the wild-type distribution with an unpaired t-test and Benjamini-Hochberg control of the false discovery rate at 0.1. Final figures were generated using the ggplot2 package in R (Wickham, 2009). For targeted confirmation and follow-up analyses (**Figs. 2.9-2.12**) responses were pooled across plates and compared across strains using binomial logistic regression for habituation or one-way ANOVA for additional behavioral assays (**Fig. 2.13**) with Tukey's honestly significant difference (HSD) criterion as previously described (Ardiel et al., 2017; Loucks et al., 2019). Each targeted confirmation and follow up experiment was independently replicated at least twice. Alpha values of 0.001 or 0.0001 were used to determine significance for Logistic regression and one-way ANOVA statistical tests respectively. Final figures were generated using the ggplot2 package in R. For all Multi-Worm Tracker experiments 3-6 plates (20-100 worms/plate) were run for each strain on each testing day. Sample sizes for each behavioral assay were chosen to be either equal to or greater than sample sizes reported in the literature that were sufficient to detect biologically relevant differences.

All raw and processed data, and the results of all statistical tests can be found at (<u>https://doi.org/10.5683/SP2/FJWIL8</u>), all analysis code is freely available at (<u>https://github.com/PavlidisLab/McDiarmid-etal-2019\_Multi-Worm-Tracker-analysis</u>).

# 2.4.6 Clustering analyses

T-statistic's were used as numerical scores to represent the difference between wild-type and mutant animals for each phenotypic feature; this created a numerical profile of phenotypic features for further analysis. All clustering analyses were performed in R. Correlation distributions were visualized using ggplot2 (Wickham, 2009). Average-linkage hierarchical clustering was performed with pvclust (Suzuki and Shimodaira, 2006) using correlation as the distance measure, and 50,000 rounds of bootstrapping. t-SNE clustering was performed using the Rtsne package (Maaten and Krijthe, 2018) with hyperparameters: initial principal component analysis = TRUE, perplexity = 10, and theta = 0 for the final 2D and 3D visualizations. The dendrogram and heat maps were visualized with pheatmap (Kolde, 2019) for the correlation matrix heat map, iheatmapr (Schep and Kummerfeld) for the phenomic profile heat maps, and pvclust for the dendrograms.

# 2.4.7 CRISPR-Cas9 missense variant lines

The *chd-7* and *nlg-1* missense variant lines were generated using CRISPR-Cas9 genome editing as previously described (Wong et al., 2019). The following missense variants strains were characterized:

PS7293 chd-7(sy861[G1225S]) I PS7317 chd-7(sy855[L1487R]) I PS7316 chd-7(sy854[R2624Q]) I PS7318 chd-7(sy859[L1220P]) I PS7267 chd-7(sy1049[P253L]) I PS7711 nlg-1(sy959[G59R]) X PS7713 nlg-1(sy961[A268T]) X

# 2.4.8 List of strains generated in this work

The following strains were created for this work either via standard genetic crosses for double mutants, or via microinjection of plasmid DNA for the generation of extrachromosomal array or CRISPR-Cas9 genome engineered transgenic lines:

VG870-871 chd-7(gk306) I; nlg-1(ok259) X

VG872-873 chd-7(gk306) I; rme-6(b1014) X

RM3707 nlg-1(ok259) X - 6x outcrossed

RM3249 pha-1(e2123ts) III; nlg-1(ok259) X

RM3540 pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1035[Pnlg-1::NLG-1(R62W)::YFP; pBX]

RM3516-17 pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1016-1017[Pnlg-1::NLG-

1(V397M)::YFP; pBX]

RM3536 pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1033[Pnlg-1::NLG-1(R430C)::YFP; pBX]

RM3537 pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1034[Pnlg-1::NLG-1(R714C)::YFP; pBX]

VG880 nlg-1(yv15[nlg-1p::nlg-1::AID::GFP:: + LoxP pmyo-2::GFP::unc-54 UTR prps-

27::NeoR::unc-54 UTR LoxP + nlg-1 3' UTR])

VG881 *nlg-1(yv16[nlg-1p::nlg-1::AID::GFP:: + LoxP + nlg-1 3' UTR]*)

VG890-891 nlg-1(yv16[nlg-1p::nlg-1::AID::GFP:: + LoxP + nlg-1 3' UTR]); ieSi57[eft-

3p::TIR1::mRuby::unc-54 3' UTR + cbr-unc-119(+)] II; unc-119(ed3) III

### 2.4.9 Strain and plasmid generation

The neuroligin missense variant plasmids were constructed by performing standard sitedirected mutagenesis on our previously described *nlg-1::YFP* functional fusion protein construct derived from the yk497a9 cDNA (Hunter et al., 2010).

The Moerman lab guide selection tool (http://genome.sfu.ca/crispr/) was used to identify the *nlg-1* targeting sgRNA (Au et al., 2018). The *nlg-1* sgRNA sequence: TCACCAACGTGTCCACGTCA was cloned into the *pU6::klp-12* sgRNA vector (obtained from Calarco lab) using site-directed mutagenesis and used for all editing experiments. The *nlg-1::AID::GFP::nlg-1* upstream and *nlg-1 3' UTR* downstream homology arms were synthesized by IDT and cloned into the loxP\_myo2\_neoR repair construct (obtained from Calarco lab) using Gibson Assembly.

*C. elegans* wild-type N2 strain was used for all CRISPR-Cas9 editing experiments. Genome edits were created as previously described (Norris et al., 2015). In brief, plasmids encoding sgRNA, Cas9 co-transformation markers pCFJ90 and pCFJ104 (Jorgensen lab, Addgene) and the selection cassette flanked by homology arms (~500 bp) containing *AID::GFP* were injected into wild-type animals. Animals containing the desired insertions were identified by G418 resistance, loss of extrachromosomal array markers, and uniform dim fluorescence of the inserted GFP.

#### 2.4.10 Genotype confirmation

Correct insertion of the *GFP::AID sequence* was confirmed by amplifying the two regions spanning the upstream and downstream insertion borders using PCR followed by Sanger sequencing (primer binding locations depicted in **Fig. 2.14A**). The genotyping strategy is essentially as described for deletion allele generation via DMS cassette insertion in Norris et al. (2015) (Dickinson and Goldstein, 2016; Norris et al., 2015).

The forward and reverse primers used to amplify the upstream insertion region were GAAGTTTCCAAATGGTCGTAGAAC (located within the *nlg-1* genomic locus) and CGAGAAGCATTGAACACCATAAC (located within GFP in the selection cassette) respectively.

The forward and reverse primers used to amplify the downstream insertion region were TTCCTCGTGCTTTACGGTATCG (located within the Neomycin resistance gene) and GGTAGCTTGATTCGCCTTCTAT (located downstream of the *nlg-1* genomic locus) respectively.

Following cassette excision via injection of cre-recombinase the *nlg-1* genomic locus (GAAGTTTCCAAATGGTCGTAGAAC) and *nlg-1* downstream (GGTAGCTTGATTCGCCTTCTAT) primers were used to amplify and confirm error free insertion of the *AID::GFP sequence* at the *nlg-1* locus via Sanger sequencing (**Fig. 2.14A**).

### 2.4.11 Auxin administration

Auxin administration was performed as previously described (Zhang et al., 2015). Auxin treatment was performed by transferring animals to bacteria-seeded plates containing auxin. Auxin indole-3-acetic acid (IAA) (Thermo Fisher, Alfa Aesar<sup>TM</sup> #A1055614) was dissolved in ethanol to create a 400 mM stock solution. The stock solution was stored at 4°C for up to one month. Auxin was diluted into the molten NGM agar (cooled to ~50°C before Auxin addition) before pouring plates. Auxin plates were seeded with 50  $\mu$ l of OP50 liquid culture 12-24 hours before use. For continuous exposure groups, animals were age synchronized (as described above) for behavioral testing on auxin plates and tested at 96 hours old. For developmental exposure animals were age synchronized and reared on auxin plates for 48 hours before being transferred to standard OP50 seeded NGM plates, and then tested 48 hours later (96 hours old). For adult auxin treatment groups animals were age synchronized on standard OP50 seeded NGM plates and reared to auxin plates and tested 48 hours later (96 hours old).

# 2.4.12 Confocal imaging

filter. Final figures were generated using ImageJ (National Institutes of Health, Bethesda, MD). For Fig. 2.15 NLG-1::AID::GFP fusion proteins were captured in live *C. elegans* using a Zeiss LSM800 confocal microscope (Carl Zeiss, Germany). Worms were immobilized on 2% agarose pad using a mixture of 7.5 mM levamisole (Sigma-Aldrich) and 0.225M BDM (2,3-butanedione monoxime) (Sigma-Aldrich). Images were analyzed with Zen software (Carl Zeiss). The Dorsal nerve cord was highlighted using 'Segmented Line', Edit, Selection, Straighten. The dorsal nerve cord of each individual channel (Pflp-13::mCherry::rab-3 and NLG-1::AID::GFP) were straightened and cropped using ImageJ (NIH, Bethesda, MD).

# 2.5 Supplemental information

We have included portions of our supplemental material as R Markdown .html documents or csv files to allow for interactive exploration that is necessary for readers' full understanding and to facilitate future use. These can be downloaded from the article page at the *PNAS* journal website, opened, and viewed by any Internet browser, similar to a supplemental dataset or video file (<u>https://www.pnas.org/content/117/1/656/tab-figures-data</u>). The captions provided below also appear in the R Markdown documents.

**Supplemental dataset 2.1. Genes phenotyped.** Contains the complete list of ASD risk gene orthologs phenotyped, orthology prediction software results, the specific alleles phenotyped, the outcross status of phenotyped strains, predicted molecular function of the genes from Satterstrom et al., and measures of ASD-association confidence from SFARI Gene and Satterstrom et al.

**Supplemental dataset 2.2. Genes phenotyped**. Satterstrom et al. ASD risk gene to Caenorhabditis elegans orthology. Contains predicted C. elegans orthologs for ASD risk genes reported by Satterstrom et al., whether there

is an assay-suitable ortholog deletion or severe missense allele available, and whether the ortholog has been characterized in the present work.

**Supplemental dataset 2.3. Reverse genetic screens.** All plots illustrate the sample mean distance of each genotype group from wild-type. Strains outside the 95% confidence interval of the wild-type distribution are labeled and colored blue. Only a maximum of ten strains are labeled in each direction per feature to prevent overplotting.

**Supplemental dataset 2.4. Phenotypic profiles.** For all plots bars represent directional t-statistics from an unpaired t-test comparing the indicated mutant to wild-type for each phenotypic feature listed across the x-axis. Color coding reflects feature classification.

**Supplemental dataset 2.5. Phenomic heat maps.** Phenomic heat maps summarizing the phenotypic profiles of 135 strains harboring a mutation in an ortholog of an ASD-associated gene. Cells represent directional t-statistics from comparisons to wild-type controls. T-statistics are shown unclipped and at various clippings (t clipped at  $\pm 10, \pm 20$ , etc.). On select indicated heat maps, only cells significant at FDR < 0.1 are colored for ease of interpretation. The heat maps are interactive allowing for more detailed inspection of selected observations. Absolute t-statistics values are clipped at 3.0, 10.0 and 20.0 in the last three figures.

Supplemental dataset 2.6. Multiple independent allele confirmation phenomic heat maps. Phenomic heat maps summarizing the phenotypic profiles of multiple strains harboring independent mutations in an ortholog of an ASD-associated gene. Cells represent directional t-statistics from comparisons to wild-type controls. T-statistics are shown unclipped and at various clippings (t clipped at  $\pm 10, \pm 20$ , etc.). On select indicated heat maps, only cells significant at FDR < 0.1 and with directionally selective genotype-phenotype relationships observed in both strains are colored for ease of interpretation. The heat maps are interactive allowing for more detailed inspection of selected observations. Absolute t-statistics values are clipped at 3.0, 10.0 and 20.0 in the last three figures.

Supplemental dataset 2.7. Reverse genetic screens colored by ASD-association confidence. All plots illustrate the sample mean distance of each genotype group from wild-type. Individual genotype groups are colored by the current evidence for ASD-association of the corresponding gene according to Satterstrom et al. (2019) (see main text for citation). The Satterstrom et al. categories denote evidence for ASD-association from strongest to weakest:  $FWER \le 0.05$ ,  $FDR \le 0.05$ ,  $FDR \le 0.1$ , None (no categorization). Strains outside the 95% confidence interval of the wild-type distribution are labeled. Only a maximum of five strains are labeled in each direction per feature to prevent overplotting.

#### Supplemental dataset 2.8. Reverse genetic screens colored by ASD & NDD or ASD predominant

**categorization.** All plots illustrate the sample mean distance of each genotype group from wild-type. Individual genotype groups are colored by ASD & NDD (neurodevelopmental disorders) or ASD predominant categorization of the corresponding gene according to Satterstrom et al. (2019) (see main text for citation). None = no categorization. Strains outside the 95% confidence interval of the wild-type distribution are labeled. Only a maximum of five strains are labeled in each direction per feature to prevent overplotting.

**Supplemental dataset 2.9. Pvclust dendrograms.** Dendrograms depict hierarchical clustering of strains based on similarity in their phenotypic profiles. T-statistics were used as a numerical score to represent the difference between wildtype and mutant animals for each phenotypic feature; this created a numerical profile of phenotypic features for further analysis. Average-linkage hierarchical clustering was performed with pvclust using correlation as the distance measure, and 50,000 rounds of bootstrapping. Clustering was performed on all features as well as the morphology features only and sensory and learning features only.

# Chapter 3: 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

Our ability to sequence genomes has vastly surpassed our ability to interpret the genetic variation we discover. This presents a major challenge in the clinical setting, where the recent application of whole exome and whole genome sequencing has uncovered thousands of genetic variants of uncertain significance. Here, we present a strategy for targeted human gene replacement and phenomic characterization based on CRISPR-Cas9 genome engineering in the genetic model organism Caenorhabditis elegans that will facilitate assessment of the functional conservation of human genes and structure-function analysis of disease-associated variants with unprecedented precision. We validate our strategy by demonstrating that direct single-copy replacement of the C. elegans ortholog (daf-18) with the critical human disease-associated gene Phosphatase and Tensin Homolog (PTEN) is sufficient to rescue multiple phenotypic abnormalities caused by complete deletion of *daf-18*, including complex chemosensory and mechanosensory impairments. In addition, we used our strategy to generate animals harboring a single copy of the known pathogenic lipid phosphatase inactive PTEN variant (PTEN-G129E) and showed that our automated *in vivo* phenotypic assays could accurately and efficiently classify this missense variant as loss-of-function. Finally, we used our strategy to investigate the functional impact of two missense variants of uncertain significance implicated in ASD, PTEN-Y167C, which we found functioned similar to wild-type, and PTEN-H123Q, which displayed complete loss-of-function equivalent to null and known pathogenic alleles. The integrated nature
of the human transgenes allows for analysis of both homozygous and heterozygous variants and facilitates high-throughput precision medicine drug screens. By combining genome engineering with rapid and automated phenotypic characterization, our strategy streamlines identification of novel conserved gene functions in complex sensory and learning phenotypes that can be used as *in vivo* functional assays to decipher variants of uncertain significance.

#### 3.1 Introduction

The rapid development and application of whole exome and whole genome sequencing technology has dramatically increased the pace at which we associate genetic variation with a particular disease (Auton et al., 2015; Bamshad et al., 2011; Gonzaga-Jauregui et al., 2012; Karczewski et al., 2020; Lek et al., 2016; Metzker, 2010; Need et al., 2012; Ng et al., 2010). However, our ability to sequence genomes has vastly surpassed our ability to interpret the clinical implications of the genetic variants we discover. The majority of genetic variants identified in clinical populations are currently classified as "variants of uncertain significance" meaning their potential role as a causative agent in the disease in question, or their pathogenicity, is unknown (Richards et al., 2015). Many variants are exceedingly rare, making it extremely difficult to designate them as pathogenic using classical genetic methods such as segregation within a pedigree or by identifying multiple carriers of the variant. As such, it often remains challenging to predict clinical outcomes and make informed treatment decisions based on genetic data alone.

In an attempt to address this problem several computational tools have been developed that estimate the functional consequences and pathogenicity of disease-associated variants (Richards et al., 2015). These tools use a variety of predictive features such as evolutionary

sequence conservation, protein structural and functional information, the prevalence of a variant in large putatively healthy control populations, or a combination of annotations (Karczewski et al., 2020; Kircher et al., 2014; Lek et al., 2016; Rentzsch et al., 2019; Richards et al., 2015). Despite extensive efforts none of these tools used in isolation or combination can faithfully report on the functional effects of a large portion of disease-associated variation and their accuracy is intrinsically limited to existing experimental training data (Grimm et al., 2015; Miosge et al., 2015; Starita et al., 2017). These limitations were clearly demonstrated in a recent study that showed *in vivo* functional assays of 21 human genes in yeast identified pathogenic variants with significantly higher precision and specificity than computational methods (Sun et al., 2016). This means that even for genes with well-characterized biological functions there are often hundreds of variants of uncertain functional significance (Landrum et al., 2014; Starita et al., 2017). This creates a challenging situation that requires direct assessment of the functional effects of disease-associated variants *in vivo* (Starita et al., 2017).

Genetically tractable model organisms are critical for discovering novel gene functions and the functional consequences of disease-associated genetic variants (Boycott et al., 2020; Dunham and Fowler, 2013; Kepler et al., 2020; Lehner, 2013; Manolio et al., 2017; Wangler et al., 2017). Governmental and private funding agencies are increasingly commissioning largescale collaborative programs to use diverse genetic model organisms to decipher variants of uncertain significance (Boycott et al., 2020; Chong et al., 2015; Gahl et al., 2016; Wangler et al., 2017). Among genetic model organisms, the nematode *Caenorhabditis elegans* has proven to be a particularly powerful animal model for the functional characterization of human genes *in vivo* (Kaletta and Hengartner, 2006). *C. elegans* combines the throughput and tractability of a singlecelled organism with the complex morphology and behavioral repertoire of a multicellular

animal. In addition, >50% of human genes have an ortholog in the C. elegans genome (Kaletta and Hengartner, 2006; Kim et al., 2018; Lai et al., 2000; Shaye and Greenwald, 2011) and this proportion is generally higher for disease-associated genes (Aerts et al., 2006; López-Bigas and Ouzounis, 2004; McDiarmid et al., 2020a). Transgenic expression of human genes is routinely done to confirm functional conservation and to observe the effects of disease-associated mutations. Notable examples of the utility of C. elegans to determine conserved human gene functions relevant to disease include the identification of presenilins as part of the gamma secretase complex, the mechanism of action of selective serotonin reuptake inhibitors, and the role of the insulin signaling pathway in normal and pathological ageing (Kaletta and Hengartner, 2006; Levitan and Greenwald, 1995; Levitan et al., 1996; Murphy et al., 2003; Ranganathan et al., 2001). However, traditional methods for expression of human genes in C. elegans rely on mosaic and variable over-expression of transgenes harbored as extrachromosomal arrays or specialized genetic backgrounds that can confound phenotypic analysis. This presents several challenges that inhibit precise analysis of the often critical but subtle effects of missense variants and impede the use of these transgenic strains in large-scale drug screens.

The recent advent of CRISPR-Cas9-mediated genome editing has revolutionized structure-function analyses across model organisms (Cong et al., 2013; DiCarlo et al., 2013; Dickinson et al., 2013; Doudna and Charpentier, 2014; Friedland et al., 2013; Gratz et al., 2013; Hwang et al., 2013; Jinek et al., 2012, 2013; Li et al., 2013). This system uses a single guide RNA (sgRNA) to precisely target a nuclease (most often Cas9) to induce a DNA double strand break at a defined location (Doudna and Charpentier, 2014). The double strand break can then be repaired via the error-prone non-homologous end-joining pathway (often resulting in damaging frameshift mutations) or a homology repair pathway, e.g. homology directed repair (HDR) or

microhomology-mediated end joining (Nakade et al., 2014). Following a double strand break, exogenous DNA repair templates can be used as substrates for HDR, allowing virtually any desired sequence to be inserted anywhere in the genome. Importantly, CRISPR-Cas9 genome engineering is remarkably efficient and robust in *C. elegans* (Au et al., 2018; Chiu et al., 2013; Dickinson and Goldstein, 2016; Norris et al., 2015).

Here, we present a broadly applicable strategy that adapts CRISPR-Cas9 genome engineering for targeted replacement of *C. elegans* genes with human genes. We illustrate how the library of knockout and humanized transgenics generated with this approach can be efficiently combined with automated machine vision phenotyping to rapidly discover novel gene functions, assess the functional conservation of human genes, and how this will allow for analysis of the effects of variants of uncertain significance with unprecedented precision. It is our hope that the human gene replacement and phenomic characterization strategy delineated in this article will serve both basic and health researchers alike, by serving as an open and shareable resource that will aid any genome engineer interested in understanding the functional conservation of human genes, and the functional consequences of their variants.

#### 3.2 Results

# **3.2.1** A general genome editing strategy for direct replacement of a *C. elegans* gene with a single copy of its human ortholog

To replace the Open Reading Frame (ORF) of an orthologous gene with a human gene our strategy first directs an sgRNA to induce a Cas9 mediated DNA double strand break immediately downstream of the ortholog start codon (**Fig. 3.1A**). A co-injected repair template containing ~500 bp homology arms targeted to the regions immediately upstream and downstream of the ortholog ORF serve as a substrate for homology directed repair. By fusing the coding DNA sequence (CDS) of a human gene of interest to the upstream homology arm homology directed repair integrates the human gene in place of the ortholog at a single copy in frame (**Fig. 3.1A**).





(A) A schematic of the genome editing strategy. (top left) A sgRNA targets Cas9 to induce a DNA double strand break immediately downstream of the orthologs start codon. A co-injected repair template containing ~500BP homology arms targeted to the regions immediately upstream and downstream of the ortholog ORF serve as a substrate for homology directed repair. By fusing the CDS of a human gene of interest to the upstream homology arm homology directed repair integrates the human gene in place of the ortholog at a single copy in frame. A cointegrated Dual-Marker selection cassette consisting of an antibiotic resistance gene (Prps-27::neoR::unc-54 UTR) and a fluorescent marker (Pmyo-2::GFP::unc-54 UTR) facilitates the identification transgenic animals without inducing morphological or phenotypic abnormalities. (middle left) initial integration deletes the entire open reading frame of the C. elegans ortholog while separating the human gene from the orthologs transcriptional terminator to inhibit expression, creating an ortholog deletion allele for phenotypic analysis (Note: cassette is not shown to scale for most human gene CDS). (bottom left) Subsequent injection of Cre Recombinase excises the selection cassette and connects the human gene to the orthologous transcriptional termination sequence such that a single copy of the human gene will now be expressed under the control of all of the orthologs 5' and 3' cis- and trans-regulatory machinery. Validation of the desired edit is preformed using standard amplification and Sanger sequencing of the target region (primer binding locations represented by half arrows in the schematic). For analysis of a human gene variant of uncertain significance (VUS) the variant of interest is incorporated into the HDR plasmid using standard in vitro methods such as site-directed mutagenesis and the same genome editing process is repeated using the same validated sgRNA and homology arms. (B) Human gene replacement allows for straight-forward interpretation of variant functional effect. This process allows for: 1) initial generation and phenotypic analysis of a complete null allele in the C. elegans orthologous gene 2) Direct integration of the human gene to determine if the human gene can compensate for loss of the orthologous gene, measuring functional conservation, 3) structure-function analysis of the effects of variants of uncertain significance on WT gene function. (C) This strategy allows for straightforward assessment of heterozygous alleles using standard genetic crosses.

To streamline genome editing we have based our method on a recently described Dual-Marker Selection (DMS) cassette screening protocol (Norris et al., 2015). The DMS Cassette consists of an antibiotic resistance gene (*Prps-27::neoR::unc-54 UTR*) and a fluorescent marker (*Pmyo-2::GFP::unc-54 UTR*) that facilitates the identification transgenic animals (Norris et al., 2015). We chose this cassette over similar methods as it: 1) can be used in any wild-type or mutant strain amenable to transgenesis and does not require any specialized genetic backgrounds, and 2) avoids the use of morphology and/or behavior altering selection markers

that necessitate cassette excision prior to phenotypic analysis (Bend et al., 2016; Dickinson et al., 2013, 2015). In our strategy, the DMS cassette is placed between the human gene of interest and the downstream homology arm (Fig. 3.1A). This deletes the entire ORF of the ortholog and separates the human gene from the orthologs transcriptional terminator upon initial integration. In many cases this efficiently creates a useful deletion allele of the ortholog with no human gene expression (this can be confirmed via inclusion of an epitope tag where immunological reagents are not available). In the unlikely event human gene expression does occur without a transcription terminator a second repair template using the same validated homology arms and sgRNA but missing the human gene can be integrated to create an ortholog null. A deletion allele can also be ordered from the Caenorhabditis Genetics Center where available. Importantly, the DMS cassette is flanked by two LoxP sites housed within synthetic introns that allow subsequent excision of the selection cassette via transient expression of Cre Recombinase (Norris et al., 2015). DMS cassette excision connects the human gene to the endogenous C. elegans orthologs transcriptional termination sequence such that a single copy of the human gene will now be expressed under the control of all of the orthologs 5' and 3' cis- and trans-regulatory machinery (Fig. 3.1B). Validation of the desired edit is performed using standard PCR amplification and Sanger sequencing of both the 5' and 3' junctions of the target region (Fig. 3.1A; Au et al., 2018).

For structure-function analysis of a human gene variant of uncertain significance the variant of interest is incorporated into the HDR plasmid using standard *in vitro* methods such as site-directed mutagenesis and the same genome editing process is repeated using the same validated sgRNA and homology arms. This process allows for: 1) initial generation and phenotypic analysis of a complete deletion allele in the *C. elegans* orthologous gene, 2) direct

integration of the human gene to determine if the human gene can compensate for loss of the orthologous gene, measuring functional conservation, and 3) structure-function analysis of the effects of variants of uncertain significance on wild-type gene function (**Fig. 3.1B**). Importantly, the vast majority of variants of uncertain significance identified in patients are heterozygous and the integrated nature of the transgenes generated with this strategy allow for straightforward assessment of heterozygous alleles using standard genetic crosses (**Fig. 3.1C**).

#### 3.2.2 *PTEN* as a prototypic disease-associated gene for targeted human gene replacement

As a proof of principle, we demonstrated the utility of our strategy by focusing on the critical disease-associated gene phosphatase and tensin homolog (PTEN). PTEN is a lipid phosphatase that antagonizes the phosphoinositide 3-kinase (PI3K) signaling pathway by dephosphorylating phosphatidylinositol (3,4,5)-trisphosphate (PIP3) (Li et al., 1997; Maehama and Dixon, 1998). Heterozygous germline PTEN variants are associated with diverse clinical outcomes including several tumor predisposition phenotypes (collectively called *PTEN* hamartoma tumor syndrome), intellectual disability, and Autism Spectrum Disorders (Hobert et al., 2014; Li et al., 1997; Liaw et al., 1997; McBride et al., 2010; O'Roak et al., 2012; Orrico et al., 2009; Sanders et al., 2015; Varga et al., 2009). While the situation is improving rapidly following several recent reports of large-scale experimental functional assessments of PTEN variants, it is currently impossible to predict with certainty the clinical outcome of a *PTEN* mutation carrier using sequence data alone (Matreyek et al., 2018; Mighell et al., 2018, 2020; Post et al., 2020). PTEN also has several technical advantages that make it an ideal test case: 1) *PTEN* functions in the highly conserved insulin signaling pathway that is well characterized in C. elegans (Ozes et al. 2001; Ogg & Ruvkun 1998; Mihaylova et al. 1999, and Fig. 3.2B). 2) C.

*elegans* has a single *PTEN* ortholog called *daf-18* (**Fig. 3.2 A-D**) and transgenic overexpression of human *PTEN* using extrachromosomal arrays has been shown to rescue reduced longevity and dauer defective phenotypes induced by mutations in *daf-18* (Liu and Chin-Sang, 2015; Solari et al., 2005). 3) *C. elegans* harboring homozygous *daf-18* null alleles are viable and display superficially normal morphology and spontaneous locomotor behavior (Mihaylova et al. 1999 and **Fig. 3.3**).



Figure 3.2 Functional and structural similarity of C. elegans DAF-18 and human PTEN.

(A) Protein domain annotations for DAF-18 and PTEN. The canonical DAF-18 amino acid (AA) sequence is more than twice as long as the canonical PTEN AA sequence primarily due to elongation of the C2 membrane targeting and C-tail domains. (B) Both DAF-18 and PTEN function as lipid and protein phosphatases to antagonize the highly conserved canonical insulin signaling pathway. (C) Clustal alignment of DAF-18 and PTEN. DAF-18 and PTEN share a highly conserved phosphatase domain (46% identity) and fully conserved catalytic site (residues highlighted in grey). DAF-18 has a markedly longer and less conserved C-terminal region than human PTEN resulting in low

overall amino acid similarity (20%) and identity (13.6%). Note that although the C-terminal region is much longer there are small conserved motifs spread throughout that are not illustrated by this alignment (see Liu et al., 2014). (D) Comparison of DAF-18 and PTEN structural models. (D) (top left) Solved crystal structure of human PTEN (1D5R reference structure; Lee et al. 1999). (top right) Predicted structure of DAF-18 AA 53-506 indicating similar 3D structure to human PTEN. (bottom left) Predicted structure of full-length PTEN and (bottom right) DAF-18 illustrating the increased size of DAF-18. Note the full-length DAF-18 model is likely to be inaccurate due to poor homology-based modeling of the non-conserved C-terminal region. Domain colour mapping matches panel 2A except that the catalytic site within the phosphatase domain is shaded light gray.





(A) Baseline movement speed, (B) midline length, and (C) width are not significantly different across genotypes. Circles represent plate replicates run on the same day. Error bars represent standard deviation using the number of plates as n (n = 5 or 10). n.s. not significant, one-way ANOVA and Tukey's post-hoc test.

# 3.2.3 An automated chemotaxis paradigm reveals a conserved nervous system role for PTEN in controlling NaCl preference

Structure-function analyses of PTEN variants necessitate an in vivo phenotypic assay that reports on PTEN function. When wild-type worms are grown in the presence of NaCl and their food source (Escherichia coli) they display naïve attractive navigation behavior toward NaCl. This behavior is called NaCl chemotaxis and can be quantified by measuring navigation behavior of a population of animals up a controlled NaCl concentration gradient (Ward, 1973). Previous work has shown that *daf-18* is required for attractive navigation behavior up a concentration gradient of NaCl such that animals with deletion or reduction-of-function mutations in daf-18 display innate aversion to NaCl (Tomioka et al., 2006). Interestingly, Insulin/PI3K signaling normally actively promotes salt avoidance under naive conditions and daf-18 functions in a single chemosensory neuron (ASER) to antagonize the Insulin/PI3K pathway and promote salt attraction (Adachi et al., 2010; Tomioka et al., 2006). Using our machine vision system, the Multi-Worm Tracker, we developed an automated high-throughput NaCl chemotaxis paradigm (Fig. 3.4A and B) and replicated the finding that *daf-18(e1375)* reduction-of-function mutants display strong aversion to NaCl (Swierczek et al. 2011 and Fig. 3.4C and D). We then generated a transgenic line using traditional extrachromomosal array technology that directed pan neuronal expression of wild-type human PTEN using the aex-3 promoter (Kuroyanagi et al., 2010). Pan neuronal expression of human PTEN was able to rescue the daf-18 reduction of function phenotype and restore attractive NaCl chemotaxis to wild-type levels (Fig. 3.4C and D). This work establishes NaCl chemotaxis as an in vivo behavioral assay of conserved nervous system PTEN functions.





(A) (top) All phenotypic analysis is conducted using our machine vision system, the Multi-Worm Tracker. The Multi-Worm Tracker delivers stimuli and preforms image acquisition, object selection, and parameter selection in real time while choreography software extracts detailed phenotypic information offline (A bottom panels) I) petri plate of C. elegans II) A petri plate of C. elegans selected for analysis by the Multi-Worm Tracker III) A Multi-worm tracker digital representation showing the degree of phenotypic detail. An example behaviour scored by the Multi-Worm Tracker: the C. elegans response to a mechanosensory tap to the side of the Petri plate is brief backwards locomotion (from III to IV). (B) Behavioural track tracing of a plate of worms from a novel Automated Multi-Worm Tracker NaCl chemotaxis paradigm illustrating attractive navigation behaviour of wild-type animals toward a point source of NaCl. (B) (bottom left to right) circles and arrows and (C) (left to right) worm tracks

represent navigation trajectories of wild-type attraction to a point source of NaCl, a *daf-18(e1375)* reduction-offunction decrease in NaCl chemotaxis, and a transgenic rescue of NaCl preference via pan neuronal overexpression of WT human *PTEN* in *daf-18(e1375)* reduction of function mutants. (D) Quantitative chemotaxis index scores across genotypes. Pan neuronal expression of human PTEN rescues the reversed NaCl preference of *daf-18(e1375)* mutants to wild-type levels. Circles represent plate replicates run on the same day and inset number represent the number of individual animals registered by the tracker and located outside the center starting region (i.e. included in preference score) across the three plate replicates for each genotype. Error bars represent standard deviation using the number of plates as n (n = 3). (\*\*\*\*) P < 0.0001, n.s. not significant, one-way ANOVA and Tukey's post-hoc test.

# 3.2.4 Complete deletion of the *daf-18* ORF causes strong NaCl avoidance that is not rescued by direct single-copy replacement with the canonical human *PTEN* CDS

Next we used our strategy to create a *daf-18* deletion allele and single-copy replacement *daf-18* with the 1212 bp canonical human *PTEN* CDS. Complete deletion of the 4723 bp *daf-18* open reading frame resulted in strong aversion to NaCl and chemotaxis down the salt gradient (**Fig. 3.5A**). We achieved genome-editing efficiency for human gene replacement similar to what has been previously reported for plasmid-based CRISPR-Cas9-mediated DMS cassette integration alone (Au et al., 2018; Norris et al., 2015, 2017). Chemotaxis avoidance of worms harboring the *daf-18* complete deletion allele generated using CRISPR-Cas9 was not significantly different from that of worms carrying the previously characterized large deletion allele *daf-18*(**Fig. 3.5B**). DMS cassette excision and expression of a single copy of human *PTEN* was unable to substitute for *daf-18* and did not rescue attractive NaCl chemotaxis behavior (**Fig. 3.5C**). This result was observed in two independent single-copy human *PTEN* knock in lines on several independent experimental

replications (Fig. 3.5C). Transcription of human *PTEN* was confirmed using Reverse transcription PCR in both knock in lines (Fig. 3.5D). Sanger sequencing confirmed error free insertion of transgenes at base pair resolution before and after cassette excision. There are several potential reasons why expression of human *PTEN* using extrachromasomal arrays rescued *daf-18* mutant phenotypes (Solari et al. 2005 and Fig. 3.5C and D) while targeted single-copy replacement with *PTEN* did not (Fig. 3.5C). Two prominent differences between these two technologies are the expression level of the transgenes and the use of the endogenous 3' UTR in the CRISPR-Cas9 knock in versus the *unc-54* myosin 3' UTR used in most *C. elegans* transgenes to ensure proper processing of transcripts in all tissues, including all constructs previously shown to rescue *daf-18* phenotypes with human *PTEN* (Merritt and Seydoux, 2010; Solari et al., 2005, and Fig. 3.4C and D). Another potential reason this single copy PTEN transgene did not rescue is it did not contain introns, which boost transgene expression in *C. elegans* (Nance and Frøkjær-Jensen, 2019).



Figure 3.5 Complete *daf-18* ORF deletion causes strong NaCl avoidance that is not rescued by direct replacement with human *PTEN*.

(A) NaCl chemotaxis preference scores of wild-type, daf-18(e1375) reduction of function, and CRISPR daf-18 complete deletion mutants. daf-18 complete deletion mutants show significantly stronger NaCl avoidance than daf-18(e1375) reduction of function mutants. (B) CRSPR daf-18 complete deletion mutants are not significantly different from animals harboring the putative null daf-18(ok480) deletion allele. (C) Expression of a single copy *PTEN* transgene from the native daf-18 locus is insufficient to significantly rescue NaCl chemotaxis towards wild-type levels. (D) RT-PCR confirming expression of full length PTEN mRNA in the two independent knock-in lines used for behavioural analysis. Previously validated primers that target *cmk-1* intronic regions of genomic DNA do not produce products following DNase treatment, confirming purity of the cDNA. Circles represent plate replicates run on the same day and inset number represent the number of individual animals registered by the tracker and located outside the center starting region (i.e. included in preference score) across the three plate replicates for each genotype. Error bars represent standard deviation using the number of plates as n (n = 3). (\*\*\*\*) P < 0.0001, (\*) P < 0.05 n.s. not significant, one-way ANOVA and Tukey's post-hoc test.

## 3.2.5 A streamlined human gene replacement strategy functionally replaces *daf-18* with human *PTEN*

In order to address these limitations and increase the speed of transgenesis we created an alternate repair template strategy that includes a transcriptional termination sequence so that expression of the human gene begins immediately upon integration (Fig. 3.6A). We included the validated unc-54 3' UTR sequence fused to the 3' end of human PTEN (Fig. 3.6A). Expression of wild-type human *PTEN* using this genome editing strategy significantly rescued NaCl chemotaxis, indicating production of functional PTEN immediately upon genomic integration prior to cassette excision (Fig. 3.6B). Again for this editing strategy we achieved efficiency similar to what has been previously reported for DMS cassette integration alone (Au et al., 2018; Norris et al., 2015, 2017). This alternate approach also offers the added benefits of increased throughput (as a second injection step to excise the cassette was not required for human gene expression as it is in the first strategy) and the option for retained visual transgenic markers (either within the selection cassette or by adding a 2A sequence to drive reporter expression from the same promoter), which allows for inclusion of introns in the mRNA to boost expression to presumably more physiologic levels and simplifies the generation and phenotypic analysis of heterozygotes and double mutants (Fig. 3.6A, Ahier and Jarriault, 2014; Calarco and Norris, 2018; Norris et al., 2017). Given the demonstrated versatility of CRISPR-Cas9-mediated genome editing in *C. elegans*, these results suggest our strategy should be broadly applicable for *in vivo* analysis of diverse human disease-associated genes.



**Figure 3.6 A streamlined human gene replacement strategy functionally replaces** *daf-18* with human *PTEN*. (A) Streamlined CRISPR-Cas9 gene replacement strategy. Inclusion of the validated *unc-54* 3' UTR in the upstream homology arm increases the speed of transgenesis by removing the need for cassette excision to induce transgene expression. This alternate approach also offers the option for retained visual transgenic markers harbouring introns, which boosts expression and simplifies the generation and phenotypic analysis of heterozygotes and double mutants. The inclusion of a GFP::T2A cassette is an optional addition to allow for confirmation of transgene expression without altering human gene function (Ahier and Jarriault, 2014). (B) Expression of wild-type human *PTEN* using this genome editing strategy significantly rescued NaCl chemotaxis toward wild-type levels. Circles represent plate replicates run on the same day and inset number represent the number of individual animals registered by the tracker and located outside the center starting region (i.e. included in preference score) across the three plate replicates for each genotype. Error bars represent standard deviation using the number of plates as n (n = 3). (\*\*\*\*) P < 0.0001, (\*\*\*) P < 0.001, (\*) P < 0.05, n.s. not significant, one-way ANOVA and Tukey's post-hoc test.

# 3.2.6 Peel-1 negative selection promotes screening-free CRISPR-Cas9 genome editing in *Caenorhabditis elegans*

In the standard DMS CRISPR-Cas9 genome editing strategy (which our human gene replacement method is built on), application of the antibiotic Neomycin (G418) for drug selection 24h after injection kills virtually all non-transgenic F<sub>1</sub> progeny (Fig. 3.7B). However, since all animals harboring transgenes will be resistant to drug selection, this scheme normally results in far more animals harbouring transgenes as extrachromosomal arrays (referred to hereafter as 'arrays') than desired genome-edited integrants due to limitations in Cas9 nuclease activity and HDR. These two transgenic populations can only be differentiated by the presence of body wall and pharyngeal muscle mCherry markers and/or the brightness/consistency of GFP fluorophore expression in the pharynx (Fig. 3.7B and see Materials and Methods and Norris et al., 2015 for details). While conceptually appealing, in practice the mCherry transgenes are often too dim to confidently visualize, leaving GFP as the only way to identify integrants, effectively reducing the process to searching for a dim single-copy integrant GFP needle in a variably bright extrachromosomal array GFP haystack (Fig. 3.7B). This presents an important limitation that needs to be overcome if CRISPR-Cas9 human gene replacement is to be used to functionally assess large allelic series of variants.

We hypothesized that the addition of *peel-1* negative selection to the optimized DMS pipeline (referred to hereafter as peel-1-DMS) delivered on the  $5^{\text{th}}$  day following injection via heatshock induction would kill arrays without killing genome edited integrants, which by that point would have lost the toxic extrachromosomal array. PEEL-1 is a naturally occurring *C*. *elegans* sperm-derived toxin that is normally counteracted in the embryo by its antidote, ZEEL-

1. Importantly, ectopic expression of *peel-1* at later life stages causes cell death and lethality allowing it to be used for array negative selection(Frøkjær-Jensen et al., 2012).

To test our hypothesis, we used an optimized guide selection tool (http://genome.sfu.ca/crispr/) to design programmed deletions at two separate loci, the uncharacterized genes *F53B6.7* and *F10E9.2* (**Fig. 3.7A**) (Au et al., 2018; Norris et al., 2015). These genes were selected to validate our peel-1-DMS selection approach while contributing to an ongoing effort to create null alleles in all genes in *C.elegans*. Indeed, we observed that while there were several integrants that had survived selection on days 7 and 8 post-injection using standard DMS selection, heat shock induction of *peel-1* killed arrays (**Fig. 3.7C**).





(A) Schematic of the peel-1-DMS CRISPR-Cas9 genome editing method. Dual crRNAs targeting genes of interest are injected as RNPs in complex with Cas9 to induce double strand breaks. A homology-directed repair template is used to integrate a *myo-2::GFP* pharyngeal visual marker and a Neomycin resistance gene at the cut site for integrant positive selection. Co-injected extrachromosomal mCherry markers provide visual selection against arrays while *peel-1* negative selection kills animals harboring arrays. In the standard DMS method arrays are manually distinguished from arrays based on dimness/consistency of GFP expression in the pharynx. (B) Injection and selection protocols/experimental design to test the efficacy of peel-1-DMS selection compared to our previously reported DMS method. (C) Peel-1-DMS selection kills arrays while sparing genome-edited integrants. Images from the F53B6.7 experiment 7 days post-injection.

An additional limitation of the standard DMS protocol is that worms carrying arrays crowd the culture plates, rapidly exhausting the food source and starving the population (**Fig. 3.7C**, top panel). This prevents the rare integrants from surviving and reproducing, making screening more difficult. A single heat shock to induce *peel-1* negative selection decreased

overpopulation-induced plate starvation approximately two-fold across two injectors each targeting the two F53B6.7 and F10E9.2 loci (Fig. 3.8A). Although a potential concern might be that *peel-1* would also kill array-carrying integrants, effectively decreasing the editing efficiency/recovery of genome edited animals, we observed no differences in the efficiency (or total number of integrant animals retrieved) of CRISPR-Cas9 genome editing following peel-1 treatment (Fig. 3.8A). We also did not observe an increase in male progeny following peel-1 induction (potentially because most animals that could be males would be arrays and have been killed off and/or a single 2h 34°C heat shock is insufficient to cause nondisjunction of the X chromosome and increase male proportions to a noticeable degree at a population level). Most importantly, the combined selection against arrays and reduced plate starvation allowed the integrants to win out, resulting in "pure" integrant plates and removing the need for any screening. We observed robust enrichment for pure integrant plates that was consistent across genomic loci (Fig. 3.8C). Thus, through the addition of peel-1 negative selection to an optimized DMS cassette method, C. elegans genome engineers can recover integrants (including human gene replacement lines) from diverse loci without visual screening.



Figure 3.8 Peel-1-DMS attenuates array-driven overpopulation/starvation and promotes screening-free genome editing at diverse loci.

(A) Proportion of injected plates showing signs of starvation with or without *peel-1* negative selection on the 7<sup>th</sup> day post-injection. Peel-1-DMS selection resulted in robust reductions in starvation across two injectors each targeting two distinct genomic loci. Note that "0%" indicates that no plates in that group showed signs of starvation. For the *F10E9.2* target, n = 6 plates for *peel-1* (+) and 6 plates for *peel-1* (-) for injector 1, and n = 8 plates for *peel-1* (+) and 16 plates for *peel-1* (-) for injector 2. For the *F53B6.7* target, n = 5 plates for *peel-1* (+) and 5 plates for *peel-1* (-) for injector 1, and n = 8 plates for *peel-1* (-) for injector 1, and n = 8 plates for *peel-1* (+) and 9 plates for *peel-1* (-) for injector 2. Note that each independent plate consists of 4 injected P<sub>0</sub> worms. (B) Peel-1-DMS selection did not alter the proportion of plates from which integrants were recovered (the number of plates per condition is the same as in panel A). (C) Proportion of plates for *peel-1* (+) and 11 plates for *peel-1* (-). For injector 2, n = 25 plates for *peel-1* (+) and 16 plates for *peel-1* (-). Note

that each independent plate consists of 4 injected P<sub>0</sub> worms. Peel-1-DMS selection robustly increased the proportion of integrant enriched plates.

## *3.2.7 daf-18* deletion causes mechanosensory hyporesponsivity that is rescued by targeted replacement with human *PTEN*

A long-standing goal has been to understand how human disease-associated variants alter normal gene function to produce sensory and learning abnormalities characteristic of diverse neurogenetic disorders. The massive number variants of uncertain significance recently implicated in the etiology neurogenetic disorders necessitates a dramatic increase in throughput of both transgenic construction and behavioral phenoptying if this goal is to be achieved (Ben-Shalom et al., 2017; Geschwind and Flint, 2015; Lim et al., 2017; Sanders et al., 2015; Starita et al., 2017; Wangler et al., 2017). By combing streamlined human gene integration with rapid machine vision phenotypic analysis of *C. elegans* our strategy greatly simplifies the identification of novel conserved gene functions in complex sensory and learning phenotypes.

We explored whether *daf-18* mutants displayed behavioral deficits in mechanosensory responding and/or habituation, a conserved form of non-associative learning that is altered in several neurodevelopmental and neuropsychiatric disorders (McDiarmid et al., 2017, 2018b; Rankin et al., 2009; Stessman et al., 2017; Timbers et al., 2017; van der Voet et al., 2014). When a non-localized mechanosensory stimulus is delivered to the side of the petri plate *C. elegans* are cultured on they respond by eliciting a brief reversal before resuming forward locomotion. Wild-type *C. elegans* habituate to repeated stimuli by learning to decrease the probability of eliciting a reversal (Rankin et al., 1990). To determine if *daf-18* is important for mechanoresponding and/or non-associative learning we examined habituation of the *daf-18(e1375)* and *daf-18* complete

deletion mutants. Compared to wild-type animals both daf-18(e1375) reduction of function and daf-18 complete deletion mutants exhibited significantly reduced probability of eliciting a reversal response throughout the habituation training session, indicating mechanosensory hyporesponsivity (Fig. 3.9A and B). Despite this hyporesponsivity, the plasticity of responses, or the pattern of the gradual decrement in the probability of emitting of a reversal response throughout the training session was not significantly altered in *daf-18* mutants (Fig. 3.9 A-B, and E). Importantly, targeted single-copy replacement of *daf-18* with human *PTEN* was sufficient to rescue the mechanosensory hyporesponsivity phenotype across the training session towards wild-type levels (Fig. 3.9D). These results identify a novel conserved role for PTEN in mechanosensory responding, a fundamental biological process disrupted in diverse genetic disorders (Badr et al., 1987; McDiarmid et al., 2017; Orefice et al., 2016). More broadly, they illustrate how the library of transgenic animals generated with our strategy can be used to rapidly characterize the role of diverse human genes in complex sensory and learning behaviors. These novel phenotypes can then be used to investigate the functional consequences of diseaseassociated variants in intact and freely behaving animals and to screen for therapeutics that reverse the effects of a particular pathogenic missense variant.





(A-D) Average probability of eliciting a reversal response to 30 consecutive tap stimuli delivered at a 10s ISI. (A) daf-18(e1375) reduction of function and (B) daf-18 complete deletion mutants exhibit significantly reduced probability of eliciting a reversal response throughout the habituation training session, indicating mechanosensory hyporesponsivity. (C) Replacement of daf-18 with human *PTEN* using the original strategy does not rescue mechanosensory hyporesponsivity. (D) Expression of human *PTEN* using the streamlined replacement strategy rescues mechanosensory responding to wild-type levels. Error bars represent standard error of the mean. (E) Habituation, or the ability to learn to decrease the probability of eliciting of a reversal response throughout the training session was not significantly altered in daf-18 mutants. Circles represent plate replicates run on the same day. Error bars represent standard deviation of the mean using the number of plates as n (n = 5 or 10). (\*) P < 0.05 n.s. not significant, one-way ANOVA and Tukey's post-hoc test.

# 3.2.8 Assessment of *PTEN* missense variants using CRISPR-Cas9 human gene replacement

To demonstrate the feasibility of our human gene replacement strategy for assessing variants of uncertain significance we set out to determine whether our *in vivo* functional assays could discern known pathogenic variants. Early studies characterizing the role of PTEN as a tumor suppressor suggested impaired protein phosphatase activity was key to the etiology of PTEN disorders (Tamura et al., 1998). However, this notion was challenged by the identification of cancer patients harboring a missense mutation that changes a glycine residue in the catalytic signature motif to a glutamate, which was predicted to abolish the lipid phosphatase activity of PTEN while leaving the protein phosphatase activity intact (Liaw et al., 1997; Myers et al., 1997). Subsequent biochemical analyses supported the now widely accepted view that the lipid phosphatase activity of PTEN is critical to its tumor suppressor activity (Myers et al., 1998).

We used site-directed mutagenesis to incorporate the PTEN-G129E (*PTEN*, c386G>A) missense variant into our repair template and used our human gene replacement strategy to replace *daf-18* with a single copy of human PTEN-G129E (**Fig. 3.10A**). Animals harboring the PTEN-G129E variant displayed strong NaCl avoidance equivalent to animals carrying the complete *daf-18* deletion allele, indicating loss-of-function (**Fig. 3.10B**). Similarly, PTEN-G129E mutants also displayed mechanosensory hyporesponsivity that was not significantly different from *daf-18* deletion carriers (**Fig. 3.10C**). These *in vivo* phenotypic results accurately classify the pathogenic PTEN-G129E as a strong loss-of-function variant. In addition, by taking advantage of a pathogenic variant with well-characterized biochemical effects these results identify a necessary role for PTEN lipid phosphatase activity in both chemotaxis and mechanosensory responding, providing novel insight into the molecular mechanisms underlying

these forms of sensory processing. Finally, we used our human gene replacement strategy to investigate the functional impact of two missense variants of uncertain significance implicated in ASD, PTEN-Y167C, which we found functioned similar to wild-type, and PTEN-H123Q, which displayed complete loss-of-function equivalent to null and known pathogenic alleles (McBride et al., 2010; Orrico et al., 2009; Post et al., 2020; **Figure 3.11**). Taken together, these results demonstrate the potential of human gene replacement and phenomic characterization to rapidly identify the functional consequences variants of uncertain significance.



Figure 3.10 Human gene replacement and *in vivo* phenotypic assessment accurately identifies functional consequences of the pathogenic PTEN-G129E variant.

(A) Location (top) and conserved amino acid sequence change (bottom) of the pathogenic lipid phosphatase inactive PTEN-G129E variant within the PTEN phosphatase domain. (B) Animals harboring the PTEN G129E variant

displayed strong NaCl avoidance equivalent to animals carrying the complete *daf-18* deletion allele, indicating lossof-function. Circles represent plate replicates run on the same day and inset number represent the number of individual animals registered by the tracker and located outside the center starting region (i.e. included in preference score) across the three plate replicates for each genotype. Error bars represent standard deviation using the number of plates as n (n = 3). (C) Similarly, PTEN-G129E mutants also displayed mechanosensory hyporesponsivity that was not significantly different from *daf-18* deletion mutants. Error bars represent standard error of the mean. (\*\*\*\*) P < 0.0001, (\*\*\*) P < 0.001, (\*) P < 0.05, n.s. not significant, one-way ANOVA and Tukey's post-hoc test.



### Figure 3.11 Functional assessment of ASD risk variants of uncertain significance using CRISPR-Cas9 human gene replacement.

(A) Location of the PTEN-H123Q (top) and PTEN-Y176C variants within PTEN. (B) Animals harboring the PTEN-H123Q variant displayed strong NaCl avoidance equivalent to animals carrying the complete *daf-18* deletion allele, indicating loss-of-function. Expression of the human PTEN-Y176C variant resulted in rescue that was not significantly different from wild-type human PTEN rescue, suggesting the variant functions similar to wild-type in

this sensory behavior. Circles represent plate replicates run on the same day and inset number represent the number of individual animals registered by the tracker and located outside the center starting region (i.e. included in preference score) across the three plate replicates for each genotype. Error bars represent standard deviation using the number of plates as n (n = 3 or 6). (\*\*\*\*) P < 0.0001, (\*\*\*) P < 0.001, (\*\*) P < 0.01, (\*) P < 0.05, n.s. not significant, one-way ANOVA and Tukey's post-hoc test.

### 3.3 Discussion

We have developed and validated a broadly applicable strategy for targeted human gene replacement and phenomic characterization in C. elegans that will facilitate assessment of the functional conservation of human genes and structure-function analysis of variants of uncertain significance with unprecedented precision. We established an automated NaCl chemotaxis paradigm and demonstrate that pan neuronal overexpression or direct replacement of daf-18 with its human ortholog PTEN using CRISPR-Cas9 is sufficient to rescue reversed NaCl chemotaxis preference induced by complete *daf-18* deletion. We further identified a novel mechanosensory hyporesponsive phenotype for *daf-18* mutants that could also be rescued by targeted replacement with human PTEN. In vivo characterization of mutants harboring a single copy of the known lipid phosphatase inactive G129E variant accurately classified this variant as pathogenic and revealed a critical role for PTEN lipid phosphatase activity in NaCl chemotaxis and mechanosensory responding. Functional analysis of two missense variants of uncertain significance implicated in ASD revealed that the Y176C variant was fully functional in our assay whereas the H123Q variant displayed complete loss-of-function equivalent to null and known pathogenic alleles. We provide novel high-throughput in vivo functional assays for PTEN, as well as validated strains, reagents, sgRNA and repair template constructs to catalyze further

analysis of this critical human disease-associated gene. More broadly, we provide a conceptual framework that illustrates how genome engineering and automated machine vision phenotyping can be combined to efficiently generate and characterize a library of knockout and humanized transgenic strains that will allow for straightforward and precise analysis of human genes and disease-associated variants *in vivo* (**Fig. 3.12**).



### Figure 3.12 A conceptual framework for *in vivo* functional analysis of human genetic variation using *C*. *elegans*.

(A) (top working clockwise) A human gene and/or variant of uncertain significance is implicated in disease etiology through clinical sequencing. Targeted CRISPR-Cas9 human gene replacement or analogous methods are used to generate a library of knock-out, human wild-type and variant transgenic strains. Large isogenic synchronous colonies of these transgenic worms are grown and their morphology, baseline locomotion, and sensory phenotypes are rapidly characterized using machine vision to establish novel functional assays and interpret variant effects. In vivo functional data can be used to probe epistatic network disruptions and cluster variants based on multiparametric phenotypic profiles. The integrated humanized transgenic lines and functional assays facilitate downstream applications including precision medicine drug screens designed to identify compounds that reverse the effects of a particular patient's missense variant. (B) Advantages of targeted human gene replacement using *C*. *elegans*.

# 3.3.1 Comparing CRISPR-Cas9 targeted human gene replacement with orthology-based variant assessment methods

To date, the most widely used genome editing-based human disease variant assessment strategy in C. elegans uses sequence alignments to identify and engineer the corresponding amino acid change into the orthologous C. elegans gene (Bend et al., 2016; Bulger et al., 2017; Canning et al., 2018; Pierce et al., 2018; Prior et al., 2017; Sorkaç et al., 2016; Troulinaki et al., 2018). A major advantage of this approach is that by using the C. elegans gene, intronic regulation, protein-protein interactions, subcellular localization, and biochemical activity of the protein of interest are, by design, perfectly modeled by C. elegans. Even when expressed at physiologically relevant levels directly from the orthologs native loci, and with evidence of phenotypic rescue, it is not guaranteed a transgenic human protein will recapitulate all functions and interactions of the orthologous C. elegans protein. This presents an important consideration when attempting to replace C. elegans proteins that must interact in extremely precise heteromeric complexes to perform their molecular functions (e.g. certain ion channel subunits) (Bend et al., 2016; Prior et al., 2017). The most obvious limitation of this approach is that it can only be used to study orthologous amino acids. Amino acids that have been conserved throughout evolution are, by definition, the least tolerant to mutation and are thus far more likely to be detrimental to protein function when mutated (Starita et al., 2017; Weile et al., 2017).

Indeed, many variant effect prediction algorithms rely on sequence conservation as the main predictor that a variant will be deleterious (Richards et al., 2015; Starita et al., 2017). It is also important to note that a large portion of amino acids will not be conserved to humans (e.g. >50% of amino acids are not conserved from DAF-18 to PTEN, **Fig. 3.2C**) and alignment algorithms that identify orthologous amino acids are imperfect. Many current implementations of orthologous amino acid engineering also require constant generation of completely new sgRNAs and repair templates for subsequent edits. Human gene replacement, in contrast, allows any coding variant of uncertain significance to be studied *in vivo* using the same validated sgRNA and homology arms.

One potential limitation of human gene replacement is that it relies on a *C. elegans* ortholog to replace. Estimates suggest there are *C. elegans* orthologs for >50% of human genes. (Kaletta and Hengartner, 2006; Lai et al., 2000; Shaye and Greenwald, 2011). This means there will not be an ortholog available for a minority of human genes. This is less likely to be a problem for disease variant modelling, as disease-associated genes are more likely to be highly conserved (Aerts et al., 2006; López-Bigas and Ouzounis, 2004; McDiarmid et al., 2020a). A recent study also showed that ~10% of human disease-associated genes are able to functionally substitute for their yeast paralogs (in addition to orthologs), further increasing the number of human genes that can be studied by replacement (Yang et al., 2017). Still, in the event there is no suitable *C. elegans* ortholog or paralog available for a human gene of interest the human gene can simply be integrated at a putatively neutral genomic location using CRISPR-Cas9 or transposon mobilization and expressed from a heterologous promoter (Frøkjær-Jensen et al., 2008, 2012, 2014). This approach can be used to screen for phenotypes induced by expression of

any human gene and to determine whether a variant exacerbates or eliminates these effects (Baruah et al., 2017).

With the rapidly expanding set of precise genome editing techniques available to *C*. *elegans*, researchers interested in variants of uncertain significance now have the freedom to choose the approach that best suits their particular needs and interests. The approaches described here provide a diverse collection of methods that can be sequentially tested in a pragmatic hierarchy of precision, beginning with direct replacement and working down until phenotypic rescue is achieved. Regardless, it will always be ideal to have corroborating evidence of variant effect from multiple techniques, indeed multiple model systems, to best inform clinical decisions.

# **3.3.2** Combining human gene replacement and automated phenomic characterization to discover conserved gene functions and establish variant functional assays

A necessary step in the establishment of human gene functional assays is the identification of phenotypes that are rescued by or induced upon expression of the reference (wild-type) human gene, as we have done here for human *PTEN* and NaCl chemotaxis. Indeed, the establishment of functional assays remains a major bottleneck for variant assessment across species (Starita et al., 2017; Weile et al., 2017). While traditional extrachromosomal array transgenes offer a quick way to establish such assays several aspects of these transgenes can severely impede this process. These include but are not limited to: 1) variable overexpression of transgenes which can lead to silencing of transgenes in certain tissues, complicating phenotypic analysis (e.g. multi-copy transgenes are expressed in the soma but silenced in the germline while low- and single-copy transgenes are expressed in both) (Kelly et al., 1997; Merritt and Seydoux,

2010), and 2) variably mosaic expression which can make it extremely difficult to assess rescue of partially penetrant and subtle complex phenotypes, as an animal must simultaneously carry the extrachromosomal array and be one of the members of the population that displays the partially penetrant phenotype that can often be difficult to score.

The human gene replacement approach described here allows for generation of ortholog deletion alleles directly in any wild-type or mutant strain amenable to transgenesis using the same reagents designed for replacement, thereby reducing the confounding effects of background mutations on phenotype discovery. The use of excisable selectable markers that do not severely alter morphology, baseline locomotion, and several evoked sensory behaviors further simplifies phenotypic analysis and removes the need for any specialized genetic backgrounds (Norris et al. 2015 and Fig. 3.1, Fig. 3.3, Fig. 3.6, and Fig. 3.7). Using this approach in combination with machine vision we provide two in vivo functional assays for human PTEN, NaCl chemotaxis and mechanosensory responsivity. In particular, NaCl chemotaxis possesses several characteristics that make it an ideal functional assay: 1) a large functional range between deletion and human gene rescue to discern potentially subtle functional differences among missense variants; 2) scalability as many plates can be run simultaneously; 3) straightforward analysis using an automatically calculated preference index (alternatively many labs score chemotaxis manually by simple blinded counting). Importantly, these reagents and functional assays can now be used in precision medicine drug screens aimed at identifying compounds that counteract the effect of a particular patient's missense variant. Further broad-scale phenomic characterization of targeted knockouts and mutant libraries (such as those described in Chapter 2) combined with new databases that curate ortholog functional annotation across model organisms should expedite the

process of *in vivo* functional assay development (Lee et al., 2018b; McDiarmid et al., 2018b, 2020a; Thompson et al., 2013; Wang et al., 2017a; Yemini et al., 2013).

### 3.3.3 Further applications of targeted human gene replacement

One of the goals of this work is to illustrate how Cas9-triggered homologous recombination can be adapted to directly replace *C. elegans* genes with human genes. An exciting adaptation of this approach will be to combine targeted human gene insertions with bipartite systems for precise spatial-temporal control of human transgene expression, as has recently been done to overexpress human genes as UAS-cDNA constructs in *Drosophila* (Lee et al., 2018a; Luo et al., 2017; Wangler et al., 2017). The recent and long-awaited development of the cGAL4-UAS system should allow a similar approach to be developed for *C. elegans*, currently the only organism where the complete cell lineage and neuronal wiring diagram is known (Sulston and Horvitz, 1977; Sulston et al., 1983; Wang et al., 2016; White et al., 1986).

While one of the clearest uses for targeted replacement is precise structure-function analysis of variants of uncertain significance there are several exciting applications beyond modeling disease-associated alleles. Targeted human gene replacement is also particularly well suited for investigation of the evolutionary principles that determine the replaceability of genes. By allowing for human gene expression immediately upon genomic integration this approach should be also adaptable to essential genes. A homozygous integrant will only be obtained if the human gene can substitute for the orthologous essential gene, creating a complementation test out of the transgenesis process. This will allow for systematic and precise interrogation of the sequence characteristics and functional properties required for successful human gene

replacement (Kachroo et al., 2015). A library of humanized worms would also open the door to the rich resources of tools available to visualize and manipulate human genes and proteins that are often unavailable for *C. elegans* researchers (e.g. high-quality antibodies, biochemically characterized or known pathogenic control variants, and experimentally determined crystal structures; **Fig. 3.2** and **Fig. 3.10**; Berman et al. 2000). Given the throughput that has been achieved for reporter gene analysis (several thousand genes) and genome editing in *C. elegans* it should be possible to generate a humanized *C. elegans* library of similar size to those recently created in yeast (Dickinson and Goldstein, 2016; Dupuy et al., 2007; Hamza et al., 2015; Hunt-Newbury et al., 2007; Kachroo et al., 2015; Norris et al., 2017; Sun et al., 2016; Yang et al., 2017). Integrated transgenes would offer the possibility of humanizing entire cellular processes for detailed *in vivo* analysis in a relatively complex yet tractable metazoan with increasingly powerful tools for spatial-temporal control of transgene expression and protein degradation (Armenti et al., 2014; Wang et al., 2016, 2017b; Zhang et al., 2015).

Deep mutational scanning and related technologies have recently made it feasible to characterize the functional effects of virtually every possible amino acid change of a protein on a particular cellular phenotype (Fowler and Fields, 2014; Fowler et al., 2010). Several such exhaustive sequence-function maps have been recently been generated in yeast and human cell culture systems (Findlay et al., 2014, 2018; Majithia et al., 2016; Matreyek et al., 2018; Mighell et al., 2018; Weile et al., 2017). These tools offer amazing resources that serve as 'lookup tables' of functional missense variation in human genes, to enable experimentally confirmed variant interpretation immediately upon first clinical presentation (Starita et al., 2017; Weile et al., 2017). An ambitious and exciting goal for the *C. elegans* community will be to further streamline
genome engineering and high-throughput phenotyping to achieve the first comprehensive sequence-function map in a metazoan.

## 3.4 Materials and Methods

## 3.4.1 Strains and culture

Worms were cultured on Nematode Growth Medium (NGM) seeded with Escherichia coli (OP50) as described previously (Brenner, 1974). Strains were maintained at 20°C unless otherwise noted. N2 Bristol, and CB1375 daf-18(e1375) strains were obtained from the Caenorhabditis Genetics Center (University of Minnesota, USA). daf-18(e1375) harbors a 30– base pair insertion in the fourth exon and is predicted to insert six amino acids before introducing an early stop codon that truncates the C-terminal half of the protein while leaving the phosphatase domain intact (Ogg and Ruvkun, 1998).

The following strains were created for this work:

VG674 daf-18(e1375); yvEx674[paex-3::PTEN::unc-54; pmyo-2::mCherry::unc-54 UTR] VG810-813 daf-18(e1375); yvEx810-813[paex-3::PTEN::unc-54 UTR; pmyo-2::mCherry::unc-54 UTR]

VG712 daf-18(yv3[daf-18p::PTEN + LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP + daf-18 UTR]) VG713 daf-18(yv4[daf-18p::PTEN + LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP + daf-18 UTR]) VG714 daf-18(yv5[daf-18p::PTEN + LoxP + daf-18 UTR ]) VG715 daf-18(yv5[daf-18p::PTEN + LoxP + daf-18 UTR])

VG817 daf-18(yv7[daf-18p::GFP::T2A::PTEN::unc-54 UTR + LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP + daf-18 UTR]) VG818 daf-18(yv8[daf-18p::GFP::T2A::PTEN::unc-54 UTR + LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP + daf-18 UTR])

VG867 daf-18(yv14[daf-18p::GFP::T2A::PTEN-G129E::unc-54 UTR + LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP + daf-18 UTR])

VC4544 F53B6.7(gk5615[+LoxP Pmyo-2::GFP::unc-54 UTR Prps-27::NeoR::unc-54 UTR LoxP+]) IV

VC4352 F10E9.2(gk5435[+LoxP Pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP+]) I

VC4353 F10E9.2(gk5436[+LoxP Pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP+]) I

VC4603 pbs-1(gk5673[+LoxP Pmyo-2::GFP::unc-54 UTR Prps-27::NeoR::unc-54 UTR

LoxP+]) IV

VC4599 K04F10.3(gk5669[+LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP+]) I

### 3.4.2 Strain and plasmid generation

The reference *PTEN* CDS (UniProt consensus, identifier: P60484-1) was obtained from a pCMV-PTEN plasmid (Addgene plasmid #28298) and cloned into a TOPO gateway entry clone (Invitrogen) according to manufacturer's instructions. The *PTEN* entry clone was recombined with an pDEST-*aex-3p* destination vector (obtained from Dr. Hidehito Kuroyanagi; Kuroyanagi et al., 2010) to generate the *aex-3p::PTEN::unc-54 UTR* rescue construct using gateway cloning (Invitrogen), according to manufacturer's instructions.

The Moerman lab guide selection tool (http://genome.sfu.ca/crispr/) was used to identify the *daf-18* targeting sgRNA. The *daf-18* sgRNA sequence: GGAGGAGGAGGAGTAACCATTGG was cloned into the *pU6::klp-12* sgRNA vector (obtained from Calarco lab) using site-directed mutagenesis and used for all human gene replacement editing experiments. The *daf-18p::PTEN* CDS and *daf-18p::GFP::T2A::PTEN::unc-54 UTR* upstream homology arms were synthesized by IDT and cloned into the loxP\_myo2\_neoR repair construct (obtained from Calarco lab) using Gibson Assembly.

*C. elegans* wild-type N2 strain was used for all CRISPR-Cas9 editing experiments. Genome edits were created as previously described (Norris et al., 2015). In brief, plasmids encoding sgRNA, Cas9, co-transformation markers pCFJ90 (*Pmyo-2::mCherry*), 5 ng/µl pCFJ104 (*Pmyo-3::mCherry*) (Jorgensen lab, Addgene) and the selection cassette flanked by homology arms (~500 bp) containing *PTEN* were injected into wild-type worms. Animals containing the desired insertions were identified by G418 resistance, loss of extrachromosomal array markers, and uniform dim fluorescence of the inserted GFP.

For the peel-1-DMS experiments (**Fig. 3.7** and **Fig. 3.8**), the *C. elegans specific* guide RNA selection tool (<u>http://genome.sfu.ca/crispr/</u>) was used to identify the F53B6.7 and F10E9.2 targeting crRNAs (dual guides for each target). The complete list of crRNAs can be found in **Table 3.1**.

Gene	F10E9.2	F53B6.7
crRNA #1	TGAGAGGACTTGCTTGGATG	AATCTTAGGCGTGGATGAAT
crRNA #2	GAAGAAAGATGGAAGTGGTG	TCTTTCCCCCTCCTGTCTTC
Sequence Validation Forward Primer	agtegecagteagaatgeaa	attttgtcgcaaagccatgta
Sequence Validation Reverse Primer	cgttggtatcagggagatccg	gageetgaaaeggateacea
Wild-type Forward Primer	GCTCGACTCTACACGCTTCT	GTGGTCTGTTGAATGCCGGA
Wild-type Forward Primer	GCTCGACTCTACACGCTTCT	GTGGTCTGTTGAATGCCGGA

Table 3.1 CRISPR reagent sequences used to generate deletion alleles for Peel-1-DMS validation.

Gene-specific crRNAs and universal tracrRNAs, both ordered from Integrated DNA Technologies (IDT), were duplexed according to manufacturer's instructions then incubated with purified Cas9 protein (kindly provided by the lab of Dr. Geraldine Seydoux, Johns Hopkins University) to create RNPs for injection. Homology directed repair constructs were designed and constructed according to the optimized DMS protocol as previously described (Au et al., 2018; Norris et al., 2015). Briefly, homology arms flanking the region to be deleted (450 bp homology with 50 bp adapter sequences for Gibson assembly) were ordered as 500 bp gBlocks from Integrated DNA Technologies (IDT). Repair template plasmids were assembled using the NEBuilder Hifi DNA Assembly Kit (New England BioLabs) to incorporate homology arms into the *loxP* + *Pmyo-2::GFP::unc-54 3'UTR* + *Prps-27::neoR::unc-54 3'UTR* + *loxP* dual-marker selection cassette vector (provided by Dr. John Calarco).

Standard DMS injection mixes consisted of 2.5 ng/µl pCFJ90 (*Pmyo-2::mCherry*), 5 ng/µl pCFJ104 (*Pmyo-3::mCherry*), 50 ng/µl gene-specific repair templates, and 0.5 µM gene-specific Cas9 RNPs. Peel-1-DMS injection mixes were prepared the same except that they included pMA122 (*Phsp16.41::peel-1*) at 10ng/µl for *peel-1* negative selection (**Fig. 3.7A**).

Adult  $P_0$  hermaphrodites were microinjected and then transferred in groups of 4 to standard culture plates to recover (Au et al., 2018; Mello and Fire, 1995). 24h following microinjection 500µl of 25mg/ml G418 was added to the culture plates for antibiotic selection (**Fig. 3.7**).

### 3.4.3 *Peel-1* induction

Five days following microinjection, plates were transferred from 20°C incubation to a 34°C incubator for a 2h heat shock to induce *peel-1*.

## 3.4.4 CRISPR Screening and quantification for Peel-1-DMS

Genome edited animals were identified by *peel-1* and/or G418 resistance, loss of extrachromosomal array markers, and uniform dim fluorescence of the inserted GFP. Experimenters blinded to condition scored plates for signs of starvation (exhausted OP50 food source) at the indicated time points. Plates where virtually all animals (>95%) were putative integrants based on *peel-1* and/or antibiotic resistance and visual markers were counted as integrant enriched.

## **3.4.5** Genotype confirmation

Correct replacement of the *daf-18* ORF with human *PTEN* was confirmed by amplifying the two regions spanning the upstream and downstream insertion borders using PCR followed by Sanger sequencing (primer binding locations depicted in **Fig. 3.1**). The genotyping strategy is essentially as described for deletion allele generation via DMS cassette insertion in (Norris et al. 2015).

The forward and reverse primers used to amplify the upstream insertion region were TGCCGTTTGAATTAGCGTGC (located within the *daf-18* genomic promoter region) and CCCTCAATGTCTCTACTTGT (located within the *myo-2* promoter of the selection cassette) respectively.

The forward and reverse primers used to amplify the downstream insertion region were TTCCTCGTGCTTTACGGTATCG (located within the Neomycin resistance gene) and CTCAACACGTTCGGAGGGTAAA (located downstream of the *daf-18* genomic coding region) respectively.

Following cassette excision via injection of cre-recombinase the *daf-18* promoter (TGCCGTTTGAATTAGCGTGC) and *daf-18* downstream (CTCAACACGTTCGGAGGGTAAA) primers were used to amplify human *PTEN* and confirm error free insertion at the *daf-18* locus via Sanger sequencing (**Fig. 3.1**).

For peel-1-DMS validation targeting *F10E9.2* and *F53B6.7*, correct insertion of the DMS cassette sequence was confirmed by amplifying the two regions spanning the upstream and downstream insertion borders using PCR. Again, the genotyping strategy is essentially as described for deletion allele generation via DMS cassette insertion in (Au et al., 2018; Norris et al., 2015).

Gene-specific forward primers were used with a universal reverse primer located within the GFP coding region of the DMS cassette: CGAGAAGCATTGAACACCATAAC to amplify the upstream insertion region for sequence confirmation.

Gene-specific reverse primers were used with a universal forward primer located within the Neomycin resistance gene of the DMS cassette: CGAGAAGCATTGAACACCATAAC to amplify the downstream insertion region for sequence confirmation.

Gene-specific wild-type primers were used in conjunction with either the forward or reverse gene-specific primer to detect partial/imperfect edits or gene duplications.

The complete list of all gene-specific forward and reverse sequence confirmation primers can be found in **Table 3.1**.

### 3.4.6 RNA extraction, library preparation, and cDNA amplification

Total RNA was isolated from mixed stage VG714 and VG715 *PTEN* knock in animals using a GeneJET RNA Purification Kit (ThermoFisher) according to manufacturer's instructions.

Total RNA was treated with DNase (New England Biolabs) and purified with a RNeasy MinElute spin column (Qiagen) according to manufacturer's instructions. cDNA libraries were prepared from crude and purified total RNA using Superscript III (Invitrogen). All genes were amplified from cDNA libraries with Platinum Taq DNA Polymerase (ThermoFisher) and genespecific primer sets.

The forward and reverse primers used to amplify the *PTEN* CDS were ATGACAGCCATCATCAAAGA and TCAGACTTTTGTAATTTGTG respectively.

The forward and reverse *cmk-1* intronic control primers (Ardiel et al., 2018) were AGGGTAGGCTAGAGTCTGGGGATAGAT and ACGACTCCGTTGTCGTGCATAAAC respectively.

### 3.4.7 Protein structure modeling and visualization

The PTEN 1D5R reference structure (Berman et al., 2000; Lee et al., 1999) was visualized using PyMOL software (DeLano, 2002). Structural models for full-length human PTEN, DAF-18 53-506, and full-length DAF-18 were predicted using Phyre2 (Kelley et al., 2015) and visualized using PyMOL.

## 3.4.8 NaCl chemotaxis behavioral assays

The chemotaxis behavioral assay was conducted on a 6 cm assay plate (2% agar), where a salt gradient was formed overnight by inserting a 2% agar plug containing 50mM of NaCl (approximately 5 mm in diameter) 1 cm from the edge of the plate. A control 2% agar plug without NaCl was inserted 1 cm from the opposite edge of the plate. Strains were grown on NGM plates seeded with *E. coli* (OP50) for 3 or 4 days. Worms on the plates were collected and washed three times using M9 buffer before being pipetted onto an unseeded NGM plate to remove excess buffer and select animals carrying transformation markers. Adult worms were transferred and placed at the centre of the assay plates and were tracked for 40 minutes on the Multi-Worm Tracker (Swierczek et al., 2011). After the tracking period, the chemotaxis index was calculated as (A - B)/(A + B), where A was the number of animals that were located in a 1.5 cm-wide region on the side of the assay plate containing the 2% agar plug without NaCl (**Fig. 3.4B**). Animals not located in either region (i.e. the middle section of the assay plate) were not counted towards the chemotaxis index. One hundred to two hundred animals were used per plate, and two or three plate replicates were used for each line in each experiment. Any statistical comparisons were carried out on plates assayed concurrently (i.e. on the same day).

### 3.4.9 Mechanosensory habituation behavioral assays

Worms were synchronized for behavioral testing on Petri plates containing Nematode Growth Media (NGM) seeded with 50 µl of OP50 liquid culture 12-24 hours before use. Five gravid adults were picked to plates and allowed to lay eggs for 3-4 hours before removal. The animals were maintained in a 20°C incubator for 96 hours. Plates of worms were placed into the tapping apparatus and after a 100s acclimatization period, 30 taps were administered at a 10s ISI. Comparisons of "final response" comprised the average of the final three stimuli. Any statistical comparisons were carried out on plates assayed concurrently (i.e. on the same day).

### 3.4.10 Multi-worm tracker behavioral analysis and statistics

Multi-Worm Tracker software (version 1.2.0.2) was used for stimulus delivery and image acquisition (Swierczek et al., 2011). Behavioral quantification with Choreography software (version  $1.3.0\_r103552$ ) used "--shadowless", "--minimum-move-body 2", and "--minimum-time 20" filters to restrict the analysis to animals that moved at least 2 body lengths and were tracked for at least 20 s. The MeasureReversal plugin was used to identify reversals occurring within 1 s (dt = 1) of the mechanosensory stimulus onset. Custom R scripts organized and summarized Choreography output files. Final figures were generated using GraphPad Prism version 7.00 for Mac OS X or ggplot2 in R (Wickham, 2009). Each experiment was independently replicated at least twice. No blinding was necessary because the Multi-Worm Tracker scores behavior objectively. Morphology metrics, baseline locomotion metrics, initial and final reversal responses, habituation difference scores, or chemotaxis indices from all plates were pooled and metrics were compared across strains with ANOVA and Tukey honestly significant difference (HSD) tests. For all statistical tests an alpha value of 0.05 was used to determine significance.

Chapter 4: Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction

## 4.1 Introduction

Much of my work on PTEN was carried out as part of a multi-lab UBC collaboration funded by the Simons Foundation Autism Research Initiative. The goal of this collaboration is to combine diverse functional assays across model systems to gain a more comprehensive understanding of the functional impact of missense variants in multi-functional proteins like PTEN. Functional assays included measures of genetic interaction perturbations in yeast, developmental timing in *Drosophila*, biochemical assays of protein stability and phosphatase activity in human cells, measures of synapse morphology and axon outgrowth in mammalian cell culture, and the automated machine vision NaCl chemotaxis behavioral assay I developed. In total, the collaboration profiled the functional effects of 106 PTEN missense variants. This set included PTEN variants implicated in somatic cancer, PTEN hamartoma syndrome, and ASD, as well as variants found among the general population. We also included several variants with well-characterized impacts on PTEN function, which we termed Biochemical Variants. These include, for example, PTEN-C124S (Maehama and Dixon, 1998), which is both lipid- and protein-phosphatase catalytically inactive, PTEN-G129E (Myers et al., 1997), a lipid phosphatase-dead variant; and PTEN-Y138L (Davidson et al., 2010), a protein phosphatase-dead variant. This variant collection was designed to enable functional characterization of PTEN variants of uncertain significance relative to variants with well-characterized impacts on PTEN

function (biochemical and known pathogenic variants) and to variants found in the general population.

All 106 PTEN variants were tested for functional impact in high-throughput model systems. A subset of those variants was then chosen for analysis in mammalian and behavioral assays. Due to the highly collaborative nature of this published work (Post et al., 2020), and for the sake of brevity, this work is not reproduced in its entirety here. Instead, the findings from this collaboration that my dissertation research contributed to are presented and briefly discussed below.

## 4.2 **Results and discussion**

### 4.2.1 PTEN variants impact sensory processing and behavior in *C. elegans*

To determine the functional impact of *PTEN* variants of uncertain significance I generated *C. elegans* strains expressing 19 PTEN variants and tested them for their ability to rescue chemotaxis deficits in *daf-18* mutant worms. I found that most disease/disorder implicated PTEN variants displayed varied loss-of-function: 14/19 variants including the known catalytically inactive variant C124S showed complete loss-of-function, 3/19 variants showed partial loss-of-function, and A79T and P354Q retained WT-like function (**Fig 3h**). The ASD variant G132D exhibited stronger negative chemotaxis than *daf-18* mutants.



### Figure 4.1 Functional impact of PTEN variants on sensory behavior in C. elegans.

(A) Schematic of NaCl Chemotaxis variant functional assay (see Materials and Methods for detailed description). (B) Impact of 20 PTEN variants on NaCl chemotaxis in *C. elegans* is shown by scoring salt preference as (A-B)/(A+B). Data are expressed as means  $\pm$  SEM and are normalized to cePTEN(rf) + hPTEN = 1 and cePTEN(rf) = 0. Variant data represented in all histograms are color-coded by variant category depicted in the color table. PHTS: PTEN hamartoma syndrome. PTEN-C124S is a known catalytically inactive variant. \*p < 0.05 compared to WT, #p < 0.05 compared to cePTEN(rf) by two-tailed Satterthwaite approximation.

# 4.3 Cross-model comparisons reveal multiple molecular mechanisms of protein dysfunction stemming from missense variants in *PTEN*

Several recent studies have reported that many *PTEN* variants disrupt protein stability, revealing a mechanism driving loss-of-function effects (Matreyek et al., 2018; Mighell et al., 2020). A key general finding from the collaboration was that there are multiple molecular mechanisms of protein dysfunction stemming from missense variants in *PTEN*. Many variants

disrupted protein stability, and destabilized variants were found to exhibit highly correlated results across functional assays (**Fig. 4.2**). However, other variants displayed complete loss-of-function or dominant negativity that was independent of their effects on stability (**Fig. 4.2**). These variants displayed poor correlations across model systems and functional assays, suggesting differential sensitivity to protein dysfunction mechanisms across model systems (**Fig. 4.2**).





(A) High correlations are found for variant impact on multiple assays for variants in which dysfunction is associated with instability. Color scale at right. Note that measures of protein stability/abundance were determined using multiple methods (e.g. fluorophore-based flow cytometry methods, western blots, etc.) in yeast and HEK293 cellular models with high correlations across cellular contexts (for full details on variant abundance & stability measurements see Post et al., 2020). (B) Weaker cross-assay correlations are seen for variants in catalytic domains,

which exhibit greater dysfunction then explained by instability. Color scale at right. Individual variant means, error, n and nominal p-values for plot are provided in the Source Data file.

These results allowed us to map regions of PTEN where the effects of variants could be predicted by their effects on stability, and regions were functional impact was poorly predicted by stability, such as the PIP3-binding and N-terminal region of the phosphatase domain, and the catalytic pocket encompassing the WPD- and P-loops (**Fig. 4.3**).



#### Figure 4.3 Functional and stability data identifies distinct mechanisms of molecular dysfunction.

(A) A normalized plot displaying stability values subtracted from function values for each assay with variants displayed according to their amino acid position below a schematic of PTEN structural domains. Normalized function – normalized stability scores are depicted as a heat map in which a score of 0 (white) indicates variants whose function matches their stability. A positive score (orange) indicates higher function than predicted from instability, while negative scores (magenta) indicates greater dysfunctional than instability (color scale at top right).

We then used the resulting datasets to reclassify variants of uncertain significance according to their predicted pathogenicity (**Fig. 4.4**). This work provided an important demonstration of the value that can be gained from a multi-model system approach to variant functional assessment. More broadly, it represents a general strategy that is already being applied to additional ASD risk genes.



Figure 4.4 PTEN variant classification based on 9 variant functional assays.

(A) PTEN variant predicted impact based on a binary notation of either LoF (<50% of WT effect) or WT-like</li>
 (≥50% of WT effect) and classified accordingly based on their frequency of LoF (pathogenic) or WT (likely benign)
 in 9 assays. Color-coding defined in color table at right. Source data are provided as a Source Data file.

### 4.4 Materials and Methods

## 4.4.1 PTEN Variant selection and annotation

All assays were performed using the full-coding sequence for the most abundant PTEN isoform (RefSeq NP 000305.3, Uniprot P60484, GenBank AAB66902.1). Missense and nonsense base substitution variants were selected from VariCarta (Belmadani et al., 2019), SFARI Gene (Abrahams et al., 2013), ClinVar (Landrum et al., 2016), COSMIC (Tate et al., 2019), and ExAC (Lek et al., 2016) databases and the literature. Site directed mutagenesis was used to generate these variants. We used VariCarta to harmonize and annotate the ASDassociated variants. Biochemical Variants were selected from the literature. Population Variants were selected which have relatively high allele frequency in ExAC and/or gnomAD (Karczewski et al., 2020) and an absence of disease-association with CADD < 25. Predicted High Impact or Predicted Low Impact variants were selected based on the CADD phred version 1.0 (Kircher et al., 2014) or SNAP2 scores. SNAP2 scores were obtained by correspondence with the authors (Hecht et al., 2015). Variants were obtained in a Human Genome Variation Society (HGVS) protein mutation format and back-converted to genomic coordinates using TransVar (Zhou et al., 2015), using Reverse Annotation: Protein mode, GRCH37/hg19 as the reference genome and RefSeq for the annotation database. Resulting genomic coordinates were then annotated using both a local Annovar instance (Wang et al., 2010) and the wAnnovar web server (ran on 24 June 2019) for additional annotations. The complete list of variants can be found in **Supplemental** dataset 4.1. Variants were generated using site-directed mutagenesis using Agilent Pfu Polymerase and moved to expression vectors either using Gateway LR Clonase (Invitrogen) or restriction cloning. Primers used for each variant are listed in **Supplemental dataset 4.2**.

## 4.4.2 *C. elegans* chemotaxis assay

Worms were cultured on Nematode Growth Medium (NGM) seeded with Escherichia coli (OP50). N2 Bristol, and CB1375 daf-18(e1375) strains were obtained from the Caenorhabditis Genetics Center (University of Minnesota, USA). All PTEN variants were expressed as pan-neuronal extrachromosomal arrays injected into the *daf-18(e1375)* reductionof-function background at 50 ng/µl final DNA concentration. daf-18(e1375) harbors a 30-base pair insertion in the fourth exon and is predicted to insert six amino acids before introducing an early stop codon that truncates the C-terminal half of the protein while leaving the phosphatase domain intact. The strains created for this work are listed in Supplementary Dataset 4.3. The PTEN entry clones were recombined with a pDEST-aex-3p destination vector for pan-neuronal expression (obtained from Dr. Hidehito Kuroyanagi, Tokyo Medical and Dental University) to generate the *aex-3p::PTEN::unc-54 UTR* rescue construct using gateway cloning (Invitrogen), according to manufacturer's instructions. Multiple transgenic strains were generated per genotype using standard microinjection and fluorescent screening for a pmyo-2::mCherry::unc-54 3' UTR (pCFJ90) co-transformation marker (3 ng/ul). PTEN variant sequences were confirmed by amplifying the entire CDS by PCR followed by Sanger sequencing. The forward and reverse primers used to amplify the PTEN CDS were ATGACAGCCATCATCAAAGA and TCAGACTTTTGTAATTTGTG respectively. The chemotaxis behavioral assay was conducted on a 6-cm assay plate (2% agar) where a salt gradient was formed overnight by inserting a 2% agar plug containing 50 mM of NaCl (~5 mm in diameter) 1 cm from the edge of the plate (Tomioka et al., 2006). A control 2% agar plug without NaCl was inserted 1 cm from the opposite edge of the plate. Strains were grown on NGM plates seeded with E. coli (OP50) for 3 or 4 days. Worms on the plates were collected and washed three times using M9 buffer before

being pipetted onto an unseeded NGM plate to remove excess buffer and select animals carrying transformation markers. Adult worms were transferred and placed at the center of the assay plates and tracked for 40 min on the Multi-Worm Tracker (Swierczek et al., 2011). After the tracking period, the chemotaxis index was calculated as (A - B)/(A + B), where A was the number of animals that were located in a 1.5-cm-wide region on the side of the assay plate containing the 2% agar plug with 50 mM NaCl and B was the number of animals that were located in a 1.5-cm-wide region on the side of the assay plate containing the 2% agar plug without NaCl. Animals not located in either region (i.e. the middle section of the assay plate) were not counted towards the chemotaxis index. One hundred to two hundred animals were used per plate, and two or three plate replicates were used for each line in each experiment. Multi-Worm Tracker software (version 1.2.0.2) was used for image acquisition (Swierczek et al., 2011). Behavioral quantification with Choreography software (version 1.3.0 r103552) used -shadowless, --minimum-move-body 2, and --minimum-time 20 filters to restrict the analysis to animals that moved at least 2 body lengths and were tracked for at least 20 s. Custom R scripts organized and summarized Choreography output files. Each experiment was independently replicated at least twice. No blinding was necessary because the Multi-Worm Tracker scores behavior objectively.

### 4.4.3 Data modeling and variant effects analysis

The quantitative phenotypes for the yeast, fly, worm, and rat assays were analyzed using a hierarchical (mixed effect) model approach to account for experimental batch effects and variability between replicates. For each assay, we treat the genotype as a fixed effect, which can be either the positive control, the negative control, or a variant, and treat blocking factors such as

experimental batch or day as random effects (stratified by variant and accompanying controls). All R scripts are provided at (<u>https://github.com/PavlidisLab/Post-PTEN</u>). For each assay, we fit models to the data for all tested variants and accompanying controls jointly. Given a genotype *i* and a sample *j*, we fit:

$$y_{i,j} = X_{i,j}\beta_{i,j} + Z_{i,j}u_{i,j} + \varepsilon$$

Where *y* is the observed quantitative phenotype (length *N* where *N* is the number of data points for the assay in total), *X* and *Z* are the model matrices for fixed and random effects respectively,  $\beta$  is a vector of parameters for the fixed effects (the genotypes of each variant), *u* is vector of parameters for random effects, and *e* is residual error. Models were fit using the lmer function of the lmerTest package (Kuznetsova et al., 2017) in the R programming environment (R Core Team 2019). To assess the significance of effect between the positive (wild-type PTEN) and negative (no PTEN, empty vector or GFP background) controls, we extracted the model fit *p*-values of each genotype under a comparison where each control is used as the contrast group, using the method of Satterthwaite as implemented in lmerTest (Kuznetsova et al., 2017). Both comparisons yield a vector of *p*-values for all other genotypes except the contrast variant. For visualization purposes, we plot data adjusted for the estimated random effects. Specifically, we compute adjusted values through a multiplication between the fixed-effects model matrix X with the  $\beta$  fixed-effect parameter estimate and a sum of the model residuals vector. In other words:

$$\hat{y}_{i,j} = X_{i,j}\hat{\beta}_{i,j} + \hat{\varepsilon}$$

We then take the *y* values for each genotype and rescale the data such that the mean positive control equals 1.0, and the mean negative control equals 0.0. Finally, we rescale the standard errors for each genotype from the adjusted data proportionally to the ratio of the adjusted data and the normalized data distributions, thus keeping the values consistent with the 0.0–1.0 re-scaling. All results are presented as means  $\pm$  SEM. For HEK293 experiments, data were analyzed with Pearson correlations tests and Student's *t* tests with WT unless otherwise indicated. All statistical analysis was done using the statistical programming software Graphpad for HEK293 experiments,  $p \le 0.05$  was considered as significant.

The datasets generated during and/or analyzed during the current study are available at (<u>https://doi.org/10.5683/SP2/DQOKPB</u>). The data underlying the figures in this chapter (**Figs. 4.1-4.3**) are also available as a Source Data file at (<u>https://www.nature.com/articles/s41467-020-15943-0#Sec25</u>).

## 4.5 Supplemental information

We have included portions of our supplemental material as csv files to allow for interactive exploration that is necessary for readers' full understanding and to facilitate future use. These can be downloaded from the article page at the *Nature Communications* journal website, opened, and viewed by any Internet browser, similar to a supplemental dataset or video file (<u>https://www.nature.com/articles/s41467-020-15943-0#Sec25</u>).

Supplemental dataset 4.1. Complete list of *PTEN* variants assessed and disease/disorder associations. Listed as "Supplementary Data 1" in (Post et al., 2020) at the above link.

Supplemental dataset 4.1. Site-directed mutagenesis primers used to generate each *PTEN* variant. Listed as "Supplementary Data 3" in (Post et al., 2020) at the above link.

**Supplemental dataset 4.3. List of** *PTEN* **variant** *C. elegans* **strains generate assessed.** Found in "Supplementary information" in (Post et al., 2020) at the above link.

# **Chapter 5: Conclusion**

The goals of my thesis were to combine emerging methods in CRISPR-Cas9 genome editing with machine vision phenomics to gain insight into the functions of ASD risk genes and the functional impact of specific genetic variants. I developed an experimental pipeline to discover functions of ASD-associated genes by inactivating each gene in the genetic model organism *Caenorhabditis elegans* and observing the phenotypic consequences using machine vision. I created digital representations summarizing the phenomic profiles of ~27,000 worms across 135 genotypes, revealing hundreds of novel genotype-phenotype relationships. Phenotypes were diverse but specific, including wide-spread developmental delays and impaired habituation learning. I then demonstrated how phenomic profiles can be used to map genetic network interactions among ASD risk genes, revealing parallel pathways of high-confidence ASD risk genes underlying deficits in habituation. Further, I illustrated how these novel phenotypes can be combined with the genetic tractability of C. elegans to determine the functional consequences of missense variants and whether phenotypic alterations are reversible. To enable more accurate functional assessment of human genetic variation, I developed a broadly applicable humanization strategy that allows for in vivo analysis of any human coding variant with unprecedented precision. Finally, I contributed to a multi-model system pipeline for variant functional assessment, allowing us to provide high confidence functional classification of PTEN variant pathogenicity and identify multiple molecular mechanisms underlying protein dysfunction. These findings represent advances in our understanding of the functions of ASD risk genes and functional impact of specific risk variants.

### 5.1 Insights into the convergent functions of ASD risk genes

Our in vivo phenotypic atlas provides clear experimental evidence of important roles for recently identified ASD risk genes in the development and function of the nervous system. For the phenotypes measured and genes assessed there were no universal phenotypic disruptions shared by all ASD risk genes, or even one that was observed in >60% of strains. This result held if we stratified genes by ASD association confidence, or by ASD versus ASD & neurodevelopmental disorders predominance. We did, however, identify several convergent functions shared by a large number of ASD risk genes which appear to be conserved, including delayed development and/or growth impediments and impaired habituation learning (McDiarmid et al., 2020a). Developmental delays have recently been reported in clinical cohorts of individuals with rare variants in several of the ASD risk genes studied here (Ganna et al., 2018). Further, a recent neuron-specific RNAi screen of 286 genes associated with intellectual disability (with or without cooccurring ASD) conducted in Drosophila also identified roles for >50% of ASD risk genes in short-term habituation (Fenckova et al., 2019). This study further reported an enrichment of habituation impairments among intellectual disability genes with cooccurring ASD compared to other intellectual disability genes (Fenckova et al., 2019). Identifying a biological correlate of the differences between ASD and intellectual disability would represent a major advance in the field, as it may offer a path to understanding the mechanistic differences between these related yet distinct diagnoses. We exclusively studied ASD risk genes, preventing us from conducting a similar enrichment analysis. However, we did use the ASD predominant vs. ASD & neurodevelopmental delay classifications proposed by Satterstrom et al. (2020) and did not observe enrichment of habituation impairments for one category over the other. So, while identifying a biological correlate of the proposed categorization of ASD predominant and ASD &

neurodevelopmental delay risk genes remains a challenge for the field moving forward, these results do suggest that habituation deficits are enriched among genes implicated in ASD, and that this enrichment is not driven by intellectual disability. By quantifying multiple measures of sensitivity, habituation, and short-term memory retention we were able to discover that the habituation impairments are largely specific to response probability. These results lead to a model where many ASD risk genes are important for influencing the plasticity of a neural circuits' decision to respond without influencing response vigor.

## 5.2 Parallel pathways among high-confidence ASD risk genes

We also identified two parallel pathways among high-confidence ASD risk genes underlying impaired habituation of response probability. These results indicate that ASD risk genes can function in parallel to mediate sensory and learning phenotypes *in vivo*. Finding a point of global convergence among ASD risk genes at the organismal, cellular, or molecular level, if one exists, represents a major challenge for the field moving forward. The clinical and model system data collected to date, including our work, may be more consistent with ASD representing a group of disorders stemming from a set of distinct pathophysiological mechanisms. At minimum, the diversity of phenotypic disruptions observed here provides a clear argument for the need for detailed phenotypic characterization to understand the functions of ASD risk genes. Importantly, our data do support the notion that certain ASD risk genes function together in specific pathways, providing support to the argument for tailored treatments based on shared pathway disruptions. Promisingly, our results, as well as recent results in additional model systems, suggest that phenotypic disruptions stemming from several ASD risk genes can be reversed later in life.

## 5.3 Deciphering variants of uncertain significance in PTEN

Our multi-model system pipeline allowed for high-confidence assessment of whether a given *PTEN* variant of uncertain significance was likely to be pathogenic (Post et al., 2020). However, it remains unclear why different PTEN variants lead to ASD while others lead to cancer (i.e. PTEN hamartoma syndrome with malignant growth or somatic cancer). Severity of variant functional impact has been proposed as potential mechanism to explain this, with ASD variants being less severe on average compared to cancer variants (Leslie and Longy, 2016; Mighell et al., 2018, 2020; Spinelli et al., 2015). This highly influential hypothesis aligns with our findings that variants more frequently observed in somatic cancer were more functionally severe, on average, across model systems (Mighell et al., 2020; Post et al., 2020). However, this model cannot fully explain the genotype-phenotype relationships observed in *PTEN* variant carriers. There are variants identified in individuals with ASD which are as severe as known pathogenic cancer variants across functional assays, and several "ASD variants" have also been observed in cancer (Post et al., 2020). In theory, this could be explained by ASD-ascertained individuals with severe PTEN variants being at high risk of developing cancer later in life, though there is not yet sufficient longitudinal evidence to support this claim (Mighell et al., 2020). Additional potential explanations for discrepant clinical annotations include differences in genetic background, environmental exposures, diagnostic challenges, or stochastic developmental process among variants carriers, all of which will require substantially larger clinical sample sizes and phenotypic detail to discern. This is a challenge in understanding many ASD risk genes as they have all been implicated in other disorders (Iakoucheva et al., 2019; Satterstrom et al., 2020; Schaaf et al., 2020). Ideally, with more sophisticated analysis methods more genes will turn out to be like SCN2A, where a clear molecular difference between missense

variants (i.e. more or less ions through the channel) is highly predictive of whether a given variant will cause ASD versus another disorder (e.g. in this case infantile-onset epilepsy) (Ben-Shalom et al., 2017; Sanders et al., 2018). While we were successful at identifying multiple molecular mechanisms of PTEN variant dysfunction, none of these mechanisms clearly differentiated between ASD and cancer predisposing variants. With the exception of protein stability and lipid phosphatase activity, studies to date have generally employed targeted approaches to assess alternative *PTEN* functions in search of a biological basis for the differentiation of ASD versus cancer variants (Matreyek et al., 2018; Mighell et al., 2018, 2020). For example, targeted studies focusing on a small set of variants have proposed nuclear exclusion as a potential molecular mechanism to differentiate between these variant classes, with ASD variants being associated with aberrant nuclear depletion (Frazier et al., 2015; Fricano-Kugler et al., 2017; Tilot et al., 2016). However, new methods are making it increasingly possible to engineer every possible variant and screen for their effects on diverse phenotypes (e.g. PTEN's protein-phosphatase activity, subcellular localization, nuclear function, proteinprotein interactions, and transcriptome influences). These datasets could reveal the mechanistic differences between ASD and cancer predisposing PTEN variants. It is now possible to determine with high accuracy whether a given PTEN variant is likely to be pathogenic, an achievement my dissertation research contributed to. The next challenge will be predicting exactly what the pathology will be and how to treat it.

### 5.4 Future directions

The conceptual and technical advances described in this dissertation present several exciting avenues for future research. First, there are numerous additional analyses that can be

conducted on the phenomic database of ASD risk genes generated here. Our set of 26 phenotypes spanned morphology, locomotion, sensory, and learning behaviors and were useful for discovering novel functions of ASD risk genes and establishing variant functional assays. However, these are user defined features that represent only a small fraction of what can be computationally extracted from the high-resolution tracker data (Brown and de Bivort, 2018; Mcdiarmid et al., 2018). Other labs have developed highly productive research programs by applying increasingly sophisticated analysis methods to datasets of similar size to the one generated here (e.g. by searching for higher-order patterns and motifs in ongoing behavior, or assessing penetrance and individuality) (Brown et al., 2013; Ghosh et al., 2012; Gomez-Marin et al., 2016; Mcdiarmid et al., 2018; Stephens et al., 2011; Stern et al., 2017; Yemini et al., 2013). In the spirit of the worm community, all digital representations have been made freely available for anyone to use and should catalyze diverse future investigations into the functions of ASD risk genes.

As discussed earlier, the CRISPR-Cas9 human gene replacement strategy described here offers several opportunities for future research. Other labs and biotech companies have already adopted the strategy and are working on humanizing entire pathways (e.g. the insulin signaling pathway and core synaptic release machinery) for *in vivo* variant analysis and drug discovery. Humanization of entire gene families has just been achieved in yeast (Garge et al., 2020; Kachroo et al., 2015; Laurent et al., 2020), and the reduced evolutionary distance between *C*. *elegans* and humans should enhance the number of genes and diversity of pathways that can be studied. Further, it should soon be possible to achieve saturation genome editing in a metazoan. Using the CRISPR-Cas9 methodology described here it is possible for a single technician to

generate >100 edits targeting independent loci per year (Au et al., 2018; McDiarmid et al., 2018a, 2020b). This is already sufficient throughput to achieve saturation genome editing of small metazoan exons through brute force genome editing. If multiplex pooling of homology repair templates is feasible in *C. elegans*, as it is in cell culture systems (Findlay et al., 2014, 2018), it may be possible to extend saturation genome editing to entire protein domains or even entire genes in the coming years. Finally, the CRISPR-Cas9 Auxin Inducible Degradation and machine vision phenotyping strategy delineated here should allow for systematic reversibility analysis of a large number of ASD risk genes. How many ASD risk genes can restore neural function if re-expressed later in life? And what aspects of their phenotypic disruptions are most plastic or least sensitive to critical periods? In addition to the biological insights discussed earlier, this thesis has provided a conceptual and technical framework for pursuing all of these future directions at rapid throughput. The phenotypic landscape of ASD risk genes and variants that we have started to uncover will likely hold the key to understanding the mechanisms that underly ASD.

## References

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36. Adachi, T., Kunitomo, H., Tomioka, M., Ohno, H., Okochi, Y., Mori, I., and Iino, Y. (2010). Reversal of salt preference is directed by the insulin/PI3K and Gq/PKC signaling in Caenorhabditis elegans. Genetics *186*, 1309–1319.

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.-C.,

De Moor, B., Marynen, P., Hassan, B., et al. (2006). Gene prioritization through genomic data fusion. Nat. Biotechnol. *24*, 537–544.

Ahier, A., and Jarriault, S. (2014). Simultaneous expression of multiple proteins under a single promoter in Caenorhabditis elegans via a versatile 2A-based toolkit. Genetics *196*, 605–613. American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th edition) (Arlington, VA: American Psychiatric Publishing).

Amin, N.D., and Paşca, S.P. (2018). Building Models of Brain Disorders with Three-Dimensional Organoids. Neuron *100*, 389–405.

Araujo, D.J., Anderson, A.G., Berto, S., Runnels, W., Harper, M., Ammanuel, S., Rieger, M.A., Huang, H.C., Rajkovich, K., Loerwald, K.W., et al. (2015). FoxP1 orchestration of ASD-relevant signaling pathways in the striatum. Genes Dev. *29*, 2081–2096.

Ardiel, E.L., Yu, A.J., Giles, A.C., and Rankin, C.H. (2017). Habituation as an adaptive shift in response strategy mediated by neuropeptides. Npj Sci. Learn. *2*.

Ardiel, E.L., McDiarmid, T.A., Timbers, T.A., Lee, K.C.Y., Safaei, J., Pelech, S.L., and Rankin,

C.H. (2018). Insights into the roles of CMK-1 and OGT-1 in interstimulus interval-dependent

habituation in Caenorhabditis elegans. Proc. R. Soc. B Biol. Sci. 285, 20182084.

Armenti, S.T., Lohmer, L.L., Sherwood, D.R., and Nance, J. (2014). Repurposing an endogenous degradation system for rapid and targeted depletion of C. elegans proteins. Development *141*, 4640–4647.

Arnett, A.B., Trinh, S., and Bernier, R.A. (2019). The state of research on the genetics of autism spectrum disorder: methodological, clinical and conceptual progress. Curr. Opin. Psychol. *27*, 1–5.

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

Rosell, F.L., Wang, S., et al. (2018). CRISPR/Cas9 Methodology for the Generation of Knockout Deletions in Caenorhabditis elegans. G3 g3.200778.2018.

Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R.,

Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Badr, G.G., Witt-Engerström, I., and Hagberg, B. (1987). Brain stem and spinal cord impairment in Rett syndrome: somatosensory and auditory evoked responses investigations. Brain Dev. 9, 517–522.

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

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.

Bariselli, S., Hörnberg, H., Prévost-Solié, C., Musardo, S., Hatstatt-Burklé, L., Scheiffele, P., and

Bellone, C. (2018). Role of VTA dopamine neurons and neuroligin 3 in sociability traits related to nonfamiliar conspecific interaction. Nat. Commun. *9*, 3173.

Barnard, R.A., Pomaville, M.B., and O'Roak, B.J. (2015). Mutations and Modeling of the

Chromatin Remodeler CHD8 Define an Emerging Autism Etiology. Front. Neurosci. 9, 477.

Baruah, P.S., Beauchemin, M., Parker, J.A., and Bertrand, R. (2017). Expression of human Bcl-

xL (Ser49) and (Ser62) mutants in Caenorhabditis elegans causes germline defects and aneuploidy. PLoS One *12*, e0177413.

Baudouin, S.J., Gaudias, J., Gerharz, S., Hatstatt, L., Zhou, K., Punnakkal, P., Tanaka, K.F.,

Spooren, W., Hen, R., De Zeeuw, C.I., et al. (2012). Shared synaptic pathophysiology in syndromic and nonsyndromic rodent models of autism. Science *338*, 128–132.

Beighley, J.S., Hudac, C.M., Arnett, A.B., Peterson, J.L., Gerdts, J., Wallace, A.S., Mefford,

H.C., Hoekzema, K., Turner, T.N., O'Roak, B.J., et al. (2019). Clinical phenotypes of carriers of mutations in CHD8 or its conserved target genes. Biol. Psychiatry *0*.

Bellmaine, S.F., Ovchinnikov, D.A., Manallack, D.T., Cuddy, C.E., Elefanty, A.G., Stanley,

E.G., Wolvetang, E.J., Williams, S.J., and Pera, M. (2017). Inhibition of DYRK1A disrupts neural lineage specification human pluripotent stem cells. Elife *6*.

Belmadani, M., Jacobson, M., Holmes, N., Phan, M., Pavlidis, P., and Rogic, S. (2019). VariCarta: a comprehensive database of harmonized genomic variants found in ASD sequencing studies. BioRxiv 608356.

Bemben, M.A., Nguyen, Q.-A., Wang, T., Li, Y., Nicoll, R.A., and Roche, K.W. (2015). Autismassociated mutation inhibits protein kinase C-mediated neuroligin-4X enhancement of excitatory synapses. Proc. Natl. Acad. Sci. U. S. A. *112*, 2551–2556.

Ben-Shalom, R., Keeshen, C.M., Berrios, K.N., An, J.Y., Sanders, S.J., and Bender, K.J. (2017).

Opposing Effects on NaV1.2 Function Underlie Differences Between SCN2A Variants Observed in Individuals With Autism Spectrum Disorder or Infantile Seizures. Biol. Psychiatry.

Bend, E.G., Si, Y., Stevenson, D.A., Bayrak-Toydemir, P., Newcomb, T.M., Jorgensen, E.M.,

and Swoboda, K.J. (2016). NALCN channelopathies. Neurology 87, 1131–1139.

Bergen, S.E., Ploner, A., Howrigan, D., O'Donovan, M.C., Smoller, J.W., Sullivan, P.F., Sebat,

J., Neale, B., and Kendler, K.S. (2019). Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. Am. J. Psychiatry *176*, 29–35.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N.,

and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235-242.

Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K.,

Gerdts, J., Baker, C., Vulto-van Silfhout, A.T., et al. (2014). Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. Cell *158*, 263–276.

Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Res. *1380*, 42–77.

de Bono, M., and Maricq, A.V. (2005). Neuronal substrates of complex behaviors in C. elegans. Annu. Rev. Neurosci. 28, 451–501.

Boycott, K.M., Campeau, P.M., Howley, H.E., Pavlidis, P., Rogic, S., Oriel, C., Berman, J.N., Hamilton, R.M., Hicks, G.G., Lipshitz, H.D., et al. (2020). The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting Understudied Genes to Model Organisms. Am. J. Hum. Genet. *106*, 143–152.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics 77, 71–94.

Brown, A.E.X., and de Bivort, B. (2018). Ethology as a physical science. Nat. Phys. 14, 653–657.

Brown, A.E.X., Yemini, E.I., Grundy, L.J., Jucikas, T., and Schafer, W.R. (2013). A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion. Proc. Natl. Acad. Sci. U. S. A. *110*, 791–796.

Bulger, D.A., Fukushige, T., Yun, S., Semple, R.K., Hanover, J.A., and Krause, M.W. (2017).Caenorhabditis elegans DAF-2 as a Model for Human Insulin Receptoropathies. G3 (Bethesda).7, 257–268.

Butler, M.G., Dasouki, M.J., Zhou, X.-P., Talebizadeh, Z., Brown, M., Takahashi, T.N., Miles, J.H., Wang, C.H., Stratton, R., Pilarski, R., et al. (2005). Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. J. Med. Genet. *42*, 318–321.

C. elegans Deletion Mutant Consortium (2012). Large-Scale Screening for Targeted Knockouts in the *Caenorhabditis elegans* Genome. G3: Genes|Genomes|Genetics 2, 1415–1425.
C. elegans Sequencing Consortium (1998). Genome sequence of the nematode C. elegans: A platform for investigating biology. Science (80-. ). 282, 2012–2018.

Calahorro, F., and Ruiz-Rubio, M. (2012). Functional Phenotypic Rescue of Caenorhabditis elegans Neuroligin-Deficient Mutants by the Human and Rat NLGN1 Genes. PLoS One *7*, e39277.

Calarco, J.A., and Norris, A.D. (2018). Synthetic Genetic Interaction (CRISPR-SGI) Profiling in Caenorhabditis elegans. Bio-Protocol 8.

Canning, P., Park, K., Gonçalves, J., Li, C., Howard, C.J., Sharpe, T.D., Holt, L.J., Pelletier, L., Bullock, A.N., and Leroux, M.R. (2018). CDKL Family Kinases Have Evolved Distinct Structural Features and Ciliary Function. Cell Rep. *22*, 885–894.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C.,

Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science (80-. ). *357*, 661–667.

Chadman, K.K., Gong, S., Scattoni, M.L., Boltuck, S.E., Gandhy, S.U., Heintz, N., and Crawley, J.N. (2008). Minimal aberrant behavioral phenotypes of neuroligin-3 R451C knockin mice. Autism Res. *1*, 147–158.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802–805.

Chanda, S., Aoto, J., Lee, S.-J., Wernig, M., and Südhof, T.C. (2016). Pathogenic mechanism of an autism-associated neuroligin mutation involves altered AMPA-receptor trafficking. Mol. Psychiatry *21*, 169–177.

Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J., and Vitkup, D. (2015). Genotype to phenotype relationships in autism spectrum disorders. Nat. Neurosci. *18*, 191–198.

Chiu, H., Schwartz, H.T., Antoshechkin, I., and Sternberg, P.W. (2013). Transgene-Free

Genome Editing in Caenorhabditis elegans Using CRISPR-Cas. Genetics 195, 1167–1171.

Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell,

T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., et al. (2015). The Genetic Basis of

Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am. J. Hum. Genet. 97,

199–215.

Chung, H. lok, Mao, X., Wang, H., Park, Y.J., Marcogliese, P.C., Rosenfeld, J.A., Burrage, L.C., Liu, P., Murdock, D.R., Yamamoto, S., et al. (2020). De Novo Variants in CDK19 Are Associated with a Syndrome Involving Intellectual Disability and Epileptic Encephalopathy. Am. J. Hum. Genet. *106*, 717–725.

Coll-Tané, M., Krebbers, A., Castells-Nobau, A., Zweier, C., and Schenck, A. (2019).
Intellectual disability and autism spectrum disorders 'on the fly': insights from *Drosophila*. Dis. Model. Mech. *12*, dmm039180.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-. ). *339*.

Corsi, A.K., Wightman, B., and Chalfie, M. (2015). A transparent window into biology: A primer on Caenorhabditis elegans. Genetics *200*, 387–407.

Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J.,

Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*.

Courcet, J.B., Faivre, L., Malzac, P., Masurel-Paulet, A., Lopez, E., Callier, P., Lambert, L.,

Lemesle, M., Thevenon, J., Gigot, N., et al. (2012). The DYRK1A gene is a cause of syndromic

intellectual disability with severe microcephaly and epilepsy. J. Med. Genet. 49, 731-736.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247–261.

Davidson, L., MacCario, H., Perera, N.M., Yang, X., Spinelli, L., Tibarewal, P., Glancy, B., Gray, A., Weijer, C.J., Downes, C.P., et al. (2010). Suppression of cellular proliferation and invasion by the concerted lipid and protein phosphatase activities of PTEN. Oncogene *29*, 687– 697.

DeLano, W.L. (2002). The PyMOL molecular graphics system.

Deneault, E., White, S.H., Rodrigues, D.C., Ross, P.J., Faheem, M., Zaslavsky, K., Wang, Z.,

Alexandrova, R., Pellecchia, G., Wei, W., et al. (2018). Complete Disruption of Autism-

Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. Stem Cell Reports *11*, 1211–1225.

Deneault, E., Faheem, M., White, S.H., Rodrigues, D.C., Sun, S., Wei, W., Piekna, A.,

Thompson, T., Howe, J.L., Chalil, L., et al. (2019). CNTN5-/+or EHMT2-/+human iPSC-

derived neurons from individuals with autism develop hyperactive neuronal networks. Elife 8.

Devlin, B., and Scherer, S.W. (2012). Genetic architecture in autism spectrum disorder. Curr. Opin. Genet. Dev. *22*, 229–237.

Devlin, B., Melhem, N., and Roeder, K. (2011). Do common variants play a role in risk for autism? Evidence and theoretical musings. Brain Res. *1380*, 78–84.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. *41*, 4336–4343.

Dickinson, D.J., and Goldstein, B. (2016). CRISPR-Based Methods for Caenorhabditis elegans Genome Engineering. Genetics *202*, 885–901.

Dickinson, D.J., Ward, J.D., Reiner, D.J., and Goldstein, B. (2013). Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. Nat. Methods *10*, 1028–1034.

Dickinson, D.J., Pani, A.M., Heppert, J.K., Higgins, C.D., and Goldstein, B. (2015). Streamlined Genome Engineering with a Self-Excising Drug Selection Cassette. Genetics *200*, 1035–1049. Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science (80-. ). *346*, 1258096–1258096.

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. PLoS Genet. *9*. Dunham, M.J., and Fowler, D.M. (2013). Contemporary, yeast-based approaches to

understanding human genetic variation. Curr. Opin. Genet. Dev. 23, 658-664.

Dupuy, D., Bertin, N., Hidalgo, C.A., Venkatesan, K., Tu, D., Lee, D., Rosenberg, J., Svrzikapa, N., Blanc, A., Carnec, A., et al. (2007). Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. Nat. Biotechnol. *25*, 663–668.

Duyzend, M.H., and Eichler, E.E. (2015). Genotype-first analysis of the 16p11.2 deletion defines a new type of "autism." Biol. Psychiatry *77*, 769–771.

Ellegood, J., Anagnostou, E., Babineau, B.A., Crawley, J.N., Lin, L., Genestine, M., DiCicco-Bloom, E., Lai, J.K.Y., Foster, J.A., Peñagarikano, O., et al. (2015). Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. Mol. Psychiatry *20*, 118–125.

Etherton, M., Földy, C., Sharma, M., Tabuchi, K., Liu, X., Shamloo, M., Malenka, R.C., and Südhof, T.C. (2011). Autism-linked neuroligin-3 R451C mutation differentially alters hippocampal and cortical synaptic function. Proc. Natl. Acad. Sci. U. S. A. *108*, 13764–13769. Fenckova, M., Blok, L.E.R., Asztalos, L., Goodman, D.P., Cizek, P., Singgih, E.L., Glennon, J.C., IntHout, J., Zweier, C., Eichler, E.E., et al. (2019). Habituation Learning Is a Widely Affected Mechanism in Drosophila Models of Intellectual Disability and Autism Spectrum Disorders. Biol. Psychiatry *86*, 294–305.

Fernandez, B.A., and Scherer, S.W. (2017). Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. Dialogues Clin. Neurosci. *19*, 353–371. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. Nature *513*, 120–123.

Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D.,

Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants

with saturation genome editing. Nature 562, 217–222.

Folstein, S., and Rutter, M. (1977). Genetic influences and infantile autism [13]. Nature 265, 726–728.

Fotaki, V., Dierssen, M., Alcántara, S., Martínez, S., Martí, E., Casas, C., Visa, J., Soriano, E., Estivill, X., and Arbonés, M.L. (2002). Dyrk1A Haploinsufficiency Affects Viability and Causes Developmental Delay and Abnormal Brain Morphology in Mice. Mol. Cell. Biol. *22*, 6636– 6647.

Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields,S. (2010). High-resolution mapping of protein sequence-function relationships. Nat. Methods 7, 741–746.

Frazier, T.W., Embacher, R., Tilot, A.K., Koenig, K., Mester, J., and Eng, C. (2015). Molecular and phenotypic abnormalities in individuals with germline heterozygous PTEN mutations and autism. Mol. Psychiatry *20*, 1132–1138.

Fricano-Kugler, C.J., Getz, S.A., Williams, M.R., Zurawel, A.A., DeSpenza, T., Frazel, P.W., Li,
M., O'Malley, A.J., Moen, E.L., and Luikart, B.W. (2017). Nuclear-Excluded AutismAssociated PTEN Mutations Dysregulate Neuronal Growth. Biol. Psychiatry.

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J.A. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods *10*, 741–743.

Frøkjær-Jensen, C., Wayne Davis, M., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen,S.-P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in

Caenorhabditis elegans. Nat. Genet. 40, 1375–1383.

Frøkjær-Jensen, C., Davis, M.W., Ailion, M., and Jorgensen, E.M. (2012). Improved Mos1mediated transgenesis in C. elegans. Nat. Methods *9*, 117–118.

Frøkjær-Jensen, C., Davis, M.W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., Pozniakovsky,

A., Moerman, D.G., and Jorgensen, E.M. (2014). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat. Methods *11*, 529–534.

J.J., Fenwick, R.G., and Warren, S.T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell *67*, 1047–1058.

Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden,

Gahl, W.A., Mulvihill, J.J., Toro, C., Markello, T.C., Wise, A.L., Ramoni, R.B., Adams, D.R., and Tifft, C.J. (2016). The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. Mol. Genet. Metab. *117*, 393–400.

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., et al. (2018). Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. Am. J. Hum. Genet. *102*, 1204– 1211.

Garge, R.K., Laurent, J.M., Kachroo, A.H., and Marcotte, E.M. (2020). Systematic Humanization of the Yeast Cytoskeleton Discerns Functionally Replaceable from Divergent Human Genes. Genetics genetics.303378.2020.

Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., et al. (2014). Most genetic risk for autism resides with common variation. Nat. Genet. *46*, 881–885.

Geng, W., Cosman, P., Baek, J.-H., Berry, C.C., and Schafer, W.R. (2003). Quantitative

classification and natural clustering of Caenorhabditis elegans behavioral phenotypes. Genetics *165*, 1117–1126.

Geschwind, D.H., and Flint, J. (2015). Genetics and genomics of psychiatric disease. Science *349*, 1489–1494.

Ghosh, R., Mohammadi, A., Kruglyak, L., and Ryu, W.S. (2012). Multiparameter behavioral profiling reveals distinct thermal response regimes in Caenorhabditis elegans. BMC Biol. *10*, 85.
Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C., et al. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat. Genet. *42*, 203–209.

Gomez-Marin, A., Stephens, G.J., and Brown, A.E.X. (2016). Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. J. R. Soc. Interface *13*, 20160466.

Gompers, A.L., Su-Feher, L., Ellegood, J., Copping, N.A., Riyadh, M.A., Stradleigh, T.W.,

Pride, M.C., Schaffler, M.D., Wade, A.A., Catta-Preta, R., et al. (2017). Germline Chd8

haploinsufficiency alters brain development in mouse. Nat. Neurosci. 20, 1062–1073.

Gonzaga-Jauregui, C., Lupski, J.R., and Gibbs, R.A. (2012). Human Genome Sequencing in Health and Disease. Annu. Rev. Med. *63*, 35–61.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2013). Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. Genetics *194*, 1029–1035.

Green, R.A., Kao, H.-L., Audhya, A., Arur, S., Mayers, J.R., Fridolfsson, H.N., Schulman, M., Schloissnig, S., Niessen, S., Laband, K., et al. (2011). A High-Resolution C. elegans Essential Gene Network Based on Phenotypic Profiling of a Complex Tissue. Cell *145*, 470–482. Griffin, A., Krasniak, C., and Baraban, S.C. (2016). Advancing epilepsy treatment through personalized genetic zebrafish models. In Progress in Brain Research, pp. 195–207.

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., et al. (2015). The Evaluation of Tools Used to Predict the Impact of Missense Variants Is Hindered by Two Types of Circularity. Hum. Mutat. *36*, 513–523.

Grissom, N.M., McKee, S.E., Schoch, H., Bowman, N., Havekes, R., O'Brien, W.T., Mahrt, E., Siegel, S., Commons, K., Portfors, C., et al. (2018). Male-specific deficits in natural reward learning in a mouse model of neurodevelopmental disorders. Mol. Psychiatry *23*, 544–555.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.

Gupta, S., Ellis, S.E., Ashar, F.N., Moes, A., Bader, J.S., Zhan, J., West, A.B., and Arking, D.E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat. Commun. *5*.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science *315*, 1143–1147.

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., et al. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. Arch. Gen. Psychiatry *68*, 1095–1102.

Hamza, A., Tammpere, E., Kofoed, M., Keong, C., Chiang, J., Giaever, G., Nislow, C., and Hieter, P. (2015). Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. Genetics *201*, 1263–1274. Hart, M.P., and Hobert, O. (2018). Neurexin controls plasticity of a mature, sexually dimorphic neuron. Nature *553*, 165–170.

Hecht, M., Bromberg, Y., and Rost, B. (2015). Better prediction of functional effects for sequence variants. BMC Genomics *16 Suppl 8*, S1.

Hedgecock, E.M., and Russell, R.L. (1975). Normal and mutant thermotaxis in the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. 72, 4061–4065.

Hippolyte, L., Maillard, A.M., Rodriguez-Herreros, B., Pain, A., Martin-Brevet, S., Ferrari, C.,

Conus, P., Macé, A., Hadjikhani, N., Metspalu, A., et al. (2016). The Number of Genomic

Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. Biol. Psychiatry *80*, 129–139.

Hmeljak, J., and Justice, M.J. (2019). From gene to treatment: supporting rare disease translational research through model systems. Dis. Model. Mech. *12*.

Hobert, J.A., Embacher, R., Mester, J.L., Frazier, T.W., and Eng, C. (2014). Biochemical screening and PTEN mutation analysis in individuals with autism spectrum disorders and macrocephaly. Eur. J. Hum. Genet. *22*, 273–276.

Hoffman, E.J., Turner, K.J., Fernandez, J.M., Cifuentes, D., Ghosh, M., Ijaz, S., Jain, R.A., Kubo, F., Bill, B.R., Baier, H., et al. (2016). Estrogens Suppress a Behavioral Phenotype in Zebrafish Mutants of the Autism Risk Gene, CNTNAP2. Neuron *89*, 725–733.

Hu, Z., Hom, S., Kudze, T., Tong, X.-J., Choi, S., Aramuni, G., Zhang, W., Kaplan, J.M., Simon, D.J., Futai, K., et al. (2012). Neurexin and neuroligin mediate retrograde synaptic inhibition in C. elegans. Science *337*, 980–984.

Hunt-Newbury, R., Viveiros, R., Johnsen, R., Mah, A., Anastas, D., Fang, L., Halfnight, E., Lee,D., Lin, J., Lorch, A., et al. (2007). High-Throughput In Vivo Analysis of Gene Expression in

Caenorhabditis elegans. PLoS Biol. 5, e237.

Hunter, J.W. (2011). Of Autism and Worms: Neuroligin mutants and synaptic function in C. elegans. University of Oklahoma, Oklahoma City.

Hunter, J.W., Mullen, G.P., McManus, J.R., Heatherly, J.M., Duke, A., and Rand, J.B. (2010).
Neuroligin-deficient mutants of C. elegans have sensory processing deficits and are
hypersensitive to oxidative stress and mercury toxicity. Dis. Model. Mech. *3*, 366–376.
Hutter, H., and Moerman, D. (2015). Big Data in Caenorhabditis elegans: quo vadis? Mol. Biol.
Cell *26*, 3909–3914.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Iakoucheva, L.M., Muotri, A.R., and Sebat, J. (2019). Getting to the Cores of Autism. Cell *178*, 1287–1298.

Ijaz, S., and Hoffman, E.J. (2016). Zebrafish: A Translational Model System for Studying Neuropsychiatric Disorders. J. Am. Acad. Child Adolesc. Psychiatry *55*, 746–748.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y., Narzisi, G., Leotta, A., et al. (2012). De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron *74*, 285–299.

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., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

Ishihara, T., Iino, Y., Mohri, A., Mori, I., Gengyo-Ando, K., Mitani, S., and Katsura, I. (2002). HEN-1, a Secretory Protein with an LDL Receptor Motif, Regulates Sensory Integration and Learning in Caenorhabditis elegans. Cell 109, 639-649.

Javer, A., Currie, M., Lee, C.W., Hokanson, J., Li, K., Martineau, C.N., Yemini, E., Grundy, L.J., Li, C., Ch'ng, Q., et al. (2018). An open source platform for analyzing and sharing worm behavior data. Nat. Methods *15*, 645.

Jiang, M., Polepalli, J., Chen, L.Y., Zhang, B., Südhof, T.C., and Malenka, R.C. (2017).

Conditional ablation of neuroligin-1 in CA1 pyramidal neurons blocks LTP by a cell-

autonomous NMDA receptor-independent mechanism. Mol. Psychiatry 22, 375-383.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A

programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., and Marcotte, E.M.

(2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science (80-. ). *348*, 921–925.

Kaeser, P.S., Deng, L., Wang, Y., Dulubova, I., Liu, X., Rizo, J., and Südhof, T.C. (2011). RIM proteins tether Ca2+ channels to presynaptic active zones via a direct PDZ-domain interaction. Cell *144*, 282–295.

Kaletta, T., and Hengartner, M.O. (2006). Finding function in novel targets: C. elegans as a model organism. Nat. Rev. Drug Discov. *5*, 387–399.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231–237.

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

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins,

R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv 531210.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

Kazdoba, T.M., Leach, P.T., Yang, M., Silverman, J.L., Solomon, M., and Crawley, J.N. (2015). Translational Mouse Models of Autism: Advancing Toward Pharmacological Therapeutics. In Current Topics in Behavioral Neurosciences, pp. 1–52.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics *146*, 227–238.

Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A C. elegans mutant that lives twice as long as wild type. Nature *366*, 461–464.

Kepler, L.D., McDiarmid, T.A., and Rankin, C.H. (2020). Habituation in high-throughput genetic model organisms as a tool to investigate the mechanisms of neurodevelopmental disorders. Neurobiol. Learn. Mem. *171*.

Kim, W., Underwood, R.S., Greenwald, I., and Shaye, D.D. (2018). OrthoList 2: A New Comparative Genomic Analysis of Human and *Caenorhabditis elegans* Genes. Genetics *210*,

445-461.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science (80-. ). *361*, 866–869.

Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S.,

Oortveld, M.A.W., Kleefstra, T., Kramer, J.M., et al. (2016). Systematic Phenomics Analysis

Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am.

J. Hum. Genet. 98, 149–164.

Kolde, R. (2019). Package "pheatmap."

Kuroyanagi, H., Ohno, G., Sakane, H., Maruoka, H., and Hagiwara, M. (2010). Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic Caenorhabditis elegans. Nat. Protoc. *5*, 1495–1517.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). ImerTest Package: Tests in Linear Mixed Effects Models . J. Stat. Softw. *82*, 1–26.

Kwon, C.-H., Luikart, B.W., Powell, C.M., Zhou, J., Matheny, S.A., Zhang, W., Li, Y., Baker, S.J., and Parada, L.F. (2006). Pten Regulates Neuronal Arborization and Social Interaction in Mice. Neuron *50*, 377–388.

de la Torre-Ubieta, L., Won, H., Stein, J.L., and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. Nat. Med. *22*, 345–361.

Lai, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S., and Lin, W. (2000). Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics.

Genome Res. 10, 703–713.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980-5.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res. *44*.

Laurent, J.M., Garge, R.K., Teufel, A.I., Wilke, C.O., Kachroo, A.H., and Marcotte, E.M. (2020). Humanization of yeast genes with multiple human orthologs reveals functional divergence between paralogs. PLoS Biol. *18*.

Lee, J.O., Yang, H., Georgescu, M.M., Di Cristofano, A., Maehama, T., Shi, Y., Dixon, J.E., Pandolfi, P., and Pavletich, N.P. (1999). Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. Cell *99*, 323–334.

Lee, P.-T., Zirin, J., Kanca, O., Lin, W.-W., Schulze, K.L., Li-Kroeger, D., Tao, R., Devereaux,
C., Hu, Y., Chung, V., et al. (2018a). A gene-specificT2A-GAL4library forDrosophila. Elife 7.
Lee, R.Y.N., Howe, K.L., Harris, T.W., Arnaboldi, V., Cain, S., Chan, J., Chen, W.J., Davis, P.,
Gao, S., Grove, C., et al. (2018b). WormBase 2017: molting into a new stage. Nucleic Acids
Res. 46, D869–D874.

Lehner, B. (2013). Genotype to phenotype: lessons from model organisms for human genetics. Nat. Rev. Genet. *14*, 168–178.

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., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

Leslie, N.R., and Longy, M. (2016). Inherited PTEN mutations and the prediction of phenotype. Semin. Cell Dev. Biol. *52*, 30–38.

Levitan, D., and Greenwald, I. (1995). Facilitation of lin-12-mediated signalling by sel-12, a Caenorhabditis elegans S182 Alzheimer's disease gene. Nature *377*, 351–354.

Levitan, D., Doyle, T.G., Brousseau, D., Lee, M.K., Thinakaran, G., Slunt, H.H., Sisodia, S.S., and Greenwald, I. (1996). Assessment of normal and mutant human presenilin function in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *93*, 14940–14944.

Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X., et al.

(2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat. Biotechnol. *31*, 681–683.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., et al. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science *275*, 1943–1947.

Li, S., Xu, C., Fu, Y., Lei, P.-J., Yao, Y., Yang, W., Zhang, Y., Washburn, M.P., Florens, L., Jaiswal, M., et al. (2018). DYRK1A interacts with histone acetyl transferase p300 and CBP and localizes to enhancers. Nucleic Acids Res. *46*, 11202–11213.

Liang, J., Xu, W., Hsu, Y.-T., Yee, A.X., Chen, L., and Südhof, T.C. (2015). Conditional neuroligin-2 knockout in adult medial prefrontal cortex links chronic changes in synaptic inhibition to cognitive impairments. Mol. Psychiatry *20*, 850–859.

Liaw, D., Marsh, D.J., Li, J., Dahia, P.L.M., Wang, S.I., Zheng, Z., Bose, S., Call, K.M., Tsou, H.C., Peacoke, M., et al. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat. Genet. *16*, 64–67.

Lichtenstein, P., Carlström, E., Råstam, M., Gillberg, C., and Anckarsäter, H. (2010). The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am. J. Psychiatry *167*, 1357–1363.

Lim, C.-S., Kang, X., Mirabella, V., Zhang, H., Bu, Q., Araki, Y., Hoang, E.T., Wang, S., Shen,
Y., Choi, S., et al. (2017). BRaf signaling principles unveiled by large-scale human mutation
analysis with a rapid lentivirus-based gene replacement method. Genes Dev. *31*, 537–552.
Lin, G.N., Corominas, R., Lemmens, I., Yang, X., Tavernier, J., Hill, D.E., Vidal, M., Sebat, J.,
and Iakoucheva, L.M. (2015). Spatiotemporal 16p11.2 Protein Network Implicates Cortical Late
Mid-Fetal Brain Development and KCTD13-Cul3-RhoA Pathway in Psychiatric Diseases.
Neuron *85*, 742–754.

Liu, J., and Chin-Sang, I.D. (2015). C. elegans as a model to study PTEN's regulation and function. Methods 77–78, 180–190.

Liu, J., Visser-Grieve, S., Boudreau, J., Yeung, B., Lo, S., Chamberlain, G., Yu, F., Sun, T., Papanicolaou, T., Lam, A., et al. (2014). Insulin activates the insulin receptor to downregulate the PTEN tumour suppressor. Oncogene *33*, 3878–3885.

López-Bigas, N., and Ouzounis, C.A. (2004). Genome-wide identification of genes likely to be involved in human genetic disease. Nucleic Acids Res. *32*, 3108–3114.

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., et al. (2020). Autism spectrum disorder. Nat. Rev. Dis. Prim. 6.

Loucks, C.M., Park, K., Walker, D.S., McEwan, A.H., Timbers, T.A., Ardiel, E.L., Grundy, L.J.,

Li, C., Johnson, J.-L., Kennedy, J., et al. (2019). EFHC1, implicated in juvenile myoclonic

epilepsy, functions at the cilium and synapse to modulate dopamine signaling. Elife 8.

Luo, X., Rosenfeld, J.A., Yamamoto, S., Harel, T., Zuo, Z., Hall, M., Wierenga, K.J., Pastore,

M.T., Bartholomew, D., Delgado, M.R., et al. (2017). Clinically severe CACNA1A alleles affect synaptic function and neurodegeneration differentially. PLoS Genet. *13*, e1006905.

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., et al. (2009). A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. Ann. Hum. Genet. *73*, 263–273.

Maaten, L. van der, and Krijthe, J. (2018). Package 'Rtsne'. T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation.

Maehama, T., and Dixon, J.E. (1998). The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J. Biol. Chem. *273*, 13375–13378.

Maenner, M.J., Shaw, K.A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen,

D.L., Wiggins, L.D., Pettygrove, S., Andrews, J.G., et al. (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, 1–12.

Majithia, A.R., Tsuda, B., Agostini, M., Gnanapradeepan, K., Rice, R., Peloso, G., Patel, K.A.,

Zhang, X., Broekema, M.F., Patterson, N., et al. (2016). Prospective functional classification of all possible missense variants in PPARG. Nat. Genet. *48*, 1570–1575.

Manolio, T.A., Fowler, D.M., Starita, L.M., Haendel, M.A., MacArthur, D.G., Biesecker, L.G.,

Worthey, E., Chisholm, R.L., Green, E.D., Jacob, H.J., et al. (2017). Bedside Back to Bench: Building Bridges between Basic and Clinical Genomic Research. Cell *169*, 6–12.

Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of rett syndrome using human induced pluripotent stem cells. Cell *143*, 527–539.

Matreyek, K.A., Starita, L.M., Stephany, J.J., Martin, B., Chiasson, M.A., Gray, V.E., Kircher,

M., Khechaduri, A., Dines, J.N., Hause, R.J., et al. (2018). Multiplex assessment of protein

variant abundance by massively parallel sequencing. Nat. Genet. 50, 874–882.

McBride, K.L., Varga, E.A., Pastore, M.T., Prior, T.W., Manickam, K., Atkin, J.F., and Herman,

G.E. (2010). Confirmation study of PTEN mutations among individuals with autism or

developmental delays/mental retardation and macrocephaly. Autism Res. 3, 137-141.

Mcdiarmid, T.A., Yu, A.J., and Rankin, C.H. (2018). Beyond the response-High throughput behavioral analyses to link genome to phenome in Caenorhabditis elegans. Genes, Brain Behav. McDiarmid, T.A., Bernardos, A.C., and Rankin, C.H. (2017). Habituation is altered in neuropsychiatric disorders - A comprehensive review with recommendations for experimental design and analysis. Neurosci. Biobehav. Rev. *80*, 286–305.

McDiarmid, T.A., Au, V., Loewen, A.D., Liang, J., Mizumoto, K., Moerman, D.G., and Rankin, C.H. (2018a). 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. Dis. Model. Mech. *11*, dmm036517.

McDiarmid, T.A., Yu, A.J., and Rankin, C.H. (2018b). Beyond the response-High throughput behavioral analyses to link genome to phenome in *Caenorhabditis elegans*. Genes, Brain Behav. e12437.

McDiarmid, T.A., Yu, A.J., and Rankin, C.H. (2019). Habituation Is More Than Learning to Ignore: Multiple Mechanisms Serve to Facilitate Shifts in Behavioral Strategy. Bioessays *41*, e1900077.

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., et al. (2020a). Systematic phenomics analysis of autismassociated genes reveals parallel networks underlying reversible impairments in habituation. Proc. Natl. Acad. Sci. U. S. A. *117*, 656–667.

McDiarmid, T.A., Au, V., Moerman, D.G., and Rankin, C.H. (2020b). Peel-1 negative selection promotes screening-free CRISPR-Cas9 genome editing in Caenorhabditis elegans. BioRxiv 2020.05.22.110353.

McGraw, C.M., Samaco, R.C., and Zoghbi, H.Y. (2011). Adult neural function requires MeCP2. Science *333*, 186.

Medina, M.A., Andrade, V.M., Caracci, M.O., Avila, M.E., Verdugo, D.A., Vargas, M.F., Ugarte, G.D., Reyes, A.E., Opazo, C., and De Ferrari, G. V (2018). Wnt/β-catenin signaling stimulates the expression and synaptic clustering of the autism-associated Neuroligin 3 gene. Transl. Psychiatry *8*, 45.

Mello, C., and Fire, A. (1995). DNA Transformation. Methods Cell Biol. 48, 451-482.

Merritt, C., and Seydoux, G. (2010). Transgenic solutions for the germline. WormBook 1–21. Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat. Rev. Genet. *11*, 31–46.

Mighell, T.L., Evans-Dutson, S., and O'Roak, B.J. (2018). A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. Am. J. Hum. Genet. *102*, 943–955.

Mighell, T.L., Thacker, S., Fombonne, E., Eng, C., and O'Roak, B.J. (2020). An Integrated Deep-Mutational-Scanning Approach Provides Clinical Insights on PTEN Genotype-Phenotype Relationships. Am. J. Hum. Genet. *106*, 818–829.

Mihaylova, V.T., Borland, C.Z., Manjarrez, L., Stern, M.J., and Sun, H. (1999). The PTEN tumor suppressor homolog in Caenorhabditis elegans regulates longevity and dauer formation in

an insulin receptor-like signaling pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 7427–7432.

Miosge, L.A., Field, M.A., Sontani, Y., Cho, V., Johnson, S., Palkova, A., Balakishnan, B.,

- Liang, R., Zhang, Y., Lyon, S., et al. (2015). Comparison of predicted and actual consequences of missense mutations. Proc. Natl. Acad. Sci. *112*, E5189–E5198.
- Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H.,
- and Kenyon, C. (2003). Genes that act downstream of DAF-16 to influence the lifespan of
- Caenorhabditis elegans. Nature 424, 277–283.
- Myers, M.P., Stolarov, J.P., Eng, C., Li, J., Wang, S.I., Wigler, M.H., Parsons, R., and Tonks,

N.K. (1997). P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. Proc. Natl. Acad. Sci. U. S. A. *94*, 9052–9057.

Myers, M.P., Pass, I., Batty, I.H., Van der Kaay, J., Stolarov, J.P., Hemmings, B.A., Wigler,

M.H., Downes, C.P., and Tonks, N.K. (1998). The lipid phosphatase activity of PTEN is critical for its tumor supressor function. Proc. Natl. Acad. Sci. U. S. A. *95*, 13513–13518.

Myers, S.M., Challman, T.D., Bernier, R., Bourgeron, T., Chung, W.K., Constantino, J.N.,

Eichler, E.E., Jacquemont, S., Miller, D.T., Mitchell, K.J., et al. (2020). Insufficient Evidence for "Autism-Specific" Genes. Am. J. Hum. Genet. *106*, 587–595.

Nageshappa, S., Carromeu, C., Trujillo, C.A., Mesci, P., Espuny-Camacho, I., Pasciuto, E.,
Vanderhaeghen, P., Verfaillie, C.M., Raitano, S., Kumar, A., et al. (2016). Altered neuronal network and rescue in a human MECP2 duplication model. Mol. Psychiatry *21*, 178–188.
Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T., et al. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. Nat. Commun. *5*, 5560.

Nakanishi, M., Nomura, J., Ji, X., Tamada, K., Arai, T., Takahashi, E., Bućan, M., and Takumi, T. (2017). Functional significance of rare neuroligin 1 variants found in autism. PLOS Genet. *13*, e1006940.

Nance, J., and Frøkjær-Jensen, C. (2019). The caenorhabditis elegans transgenic toolbox. Genetics *212*, 959–990.

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

Need, A.C., Shashi, V., Hitomi, Y., Schoch, K., Shianna, K. V, McDonald, M.T., Meisler, M.H., and Goldstein, D.B. (2012). Clinical application of exome sequencing in undiagnosed genetic conditions. J. Med. Genet. *49*, 353–361.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D.,

Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. *42*, 30–35.

Nishiyama, M., Oshikawa, K., Tsukada, Y., Nakagawa, T., Iemura, S., Natsume, T., Fan, Y., Kikuchi, A., Skoultchi, A.I., and Nakayama, K.I. (2009). CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. Nat. Cell Biol. *11*, 172–182.

Norris, A.D., Kim, H.-M., Colaiácovo, M.P., and Calarco, J.A. (2015). Efficient Genome Editing in Caenorhabditis elegans with a Toolkit of Dual-Marker Selection Cassettes. Genetics *201*. Norris, A.D., Gracida, X., and Calarco, J.A. (2017). CRISPR-mediated genetic interaction profiling identifies RNA binding proteins controlling metazoan fitness. Elife *6*. Notwell, J.H., Heavner, W.E., Darbandi, S.F., Katzman, S., McKenna, W.L., Ortiz-Londono, C.F., Tastad, D., Eckler, M.J., Rubenstei, J.L.R., McConnell, S.K., et al. (2016). TBR1 regulates autism risk genes in the developing neocortex. Genome Res. *26*, 1013–1022.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246–250.

Ogg, S., and Ruvkun, G. (1998). The C. elegans PTEN homolog, DAF-18, acts in the insulin receptor-like metabolic signaling pathway. Mol. Cell *2*, 887–893.

Orefice, L.L., Zimmerman, A.L., Chirila, A.M., Sleboda, S.J., Head, J.P., and Ginty, D.D.

(2016). Peripheral Mechanosensory Neuron Dysfunction Underlies Tactile and Behavioral Deficits in Mouse Models of ASDs. Cell *166*, 299–313.

Orrico, A., Galli, L., Buoni, S., Orsi, A., Vonella, G., and Sorrentino, V. (2009). Novel PTEN mutations in neurodevelopmental disorders and macrocephaly. Clin. Genet. *75*, 195–198.

Ozes, O.N., Akca, H., Mayo, L.D., Gustin, J.A., Maehama, T., Dixon, J.E., and Donner, D.B.

(2001). A phosphatidylinositol 3-kinase/Akt/mTOR pathway mediates and PTEN antagonizes tumor necrosis factor inhibition of insulin signaling through insulin receptor substrate-1. Proc. Natl. Acad. Sci. *98*, 4640–4645.

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan,K., Trapnell, C., Kim, J., et al. (2019). A lineage-resolved molecular atlas of C. Elegansembryogenesis at single-cell resolution. Science (80-. ). *365*.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008.

Paşca, S.P., Portmann, T., Voineagu, I., Yazawa, M., Shcheglovitov, A., Paşca, A.M., Cord, B.,

Palmer, T.D., Chikahisa, S., Nishino, S., et al. (2011). Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. Nat. Med. *17*, 1657–1662.

Peñagarikano, O., Lázaro, M.T., Lu, X.-H., Gordon, A., Dong, H., Lam, H.A., Peles, E.,

Maidment, N.T., Murphy, N.P., Yang, X.W., et al. (2015). Exogenous and evoked oxytocin

restores social behavior in the Cntnap2 mouse model of autism. Sci. Transl. Med. 7, 271ra8.

Pettersson, E., Lichtenstein, P., Larsson, H., Song, J., Agrawal, A., Borglum, A.D., Bulik, C.M.,

Daly, M.J., Davis, L.K., Demontis, D., et al. (2019). Genetic influences on eight psychiatric

disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748

cases and controls. Psychol. Med. 49, 1166–1173.

Pierce, S.B., Stewart, M.D., Gulsuner, S., Walsh, T., Dhall, A., McClellan, J.M., Klevit, R.E., and King, M.-C. (2018). De novo mutation inRING1with epigenetic effects on neurodevelopment. Proc. Natl. Acad. Sci. U. S. A. *115*, 1558–1563.

Pizzo, L., Jensen, M., Polyak, A., Rosenfeld, J.A., Mannik, K., Krishnan, A., McCready, E.,
Pichon, O., Le Caignec, C., Van Dijck, A., et al. (2019). Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. Genet. Med. *21*, 816–825.

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., et al. (2020). Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. Nat. Commun. *11*.

Prior, H., Jawad, A.K., MacConnachie, L., and Beg, A.A. (2017). Highly Efficient, Rapid and Co-CRISPR Independent Genome Editing in *Caenorhabditis elegans*. G3: Genes|Genomes|Genetics g3.300216.2017.

Pym, E., Sasidharan, N., Thompson-Peer, K.L., Simon, D.J., Anselmo, A., Sadreyev, R., Hall, Q., Nurrish, S., and Kaplan, J.M. (2017). Shank is a dose-dependent regulator of Cav1 calcium current and CREB target expression. Elife *6*.

Randlett, O., Haesemeyer, M., Forkin, G., Shoenhard, H., Schier, A.F., Engert, F., and Granato,M. (2019). Distributed Plasticity Drives Visual Habituation Learning in Larval Zebrafish. Curr.Biol.

Ranganathan, R., Sawin, E.R., Trent, C., and Horvitz, H.R. (2001). Mutations in the Caenorhabditis elegans serotonin reuptake transporter MOD-5 reveal serotonin-dependent and - independent activities of fluoxetine. J. Neurosci. *21*, 5871–5884.

Rankin, C.H., Beck, C.D., and Chiba, C.M. (1990). Caenorhabditis elegans: a new model system for the study of learning and memory. Behav. Brain Res. *37*, 89–92.

Rankin, C.H., Abrams, T., Barry, R.J., Bhatnagar, S., Clayton, D.F., Colombo, J., Coppola, G., Geyer, M.A., Glanzman, D.L., Marsland, S., et al. (2009). Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. Neurobiol. Learn. Mem. *92*, 135–138.

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47*, D886–D894.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (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. Genet. Med. *17*, 405–423.

Robinson, E.B., Lichtenstein, P., Anckarsäter, H., Happé, F., and Ronald, A. (2013). Examining

and interpreting the female protective effect against autistic behavior. Proc. Natl. Acad. Sci. U. S. A. *110*, 5258–5262.

Ronald, A., and Hoekstra, R.A. (2011). Autism spectrum disorders and autistic traits: A decade of new twin studies. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. *156*, 255–274.

Ronemus, M., Iossifov, I., Levy, D., and Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. Nat. Rev. Genet. *15*, 133–141.

Rosenberg, R.E., Law, J.K., Yenokyan, G., McGready, J., Kaufmann, W.E., and Law, P.A.

(2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs.

Arch. Pediatr. Adolesc. Med. 163, 907–914.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Ercument Cicek, A., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Sakamoto, I., Kishida, S., Fukui, A., Kishida, M., Yamamoto, H., Hino, S., Michiue, T., Takada,
S., Asashima, M., and Kikuchi, A. (2000). A novel beta-catenin-binding protein inhibits betacatenin-dependent Tcf activation and axis formation. J. Biol. Chem. *275*, 32871–32878.
San-Miguel, A., Kurshan, P.T., Crane, M.M., Zhao, Y., McGrath, P.T., Shen, K., and Lu, H.
(2016). Deep phenotyping unveils hidden traits and genetic relations in subtle mutants. Nat.
Commun. *7*, 12990.

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., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237–241.
Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into Autism Spectrum Disorder

Genomic Architecture and Biology from 71 Risk Loci. Neuron 87, 1215–1233.

Sanders, S.J., Campbell, A.J., Cottrell, J.R., Moller, R.S., Wagner, F.F., Auldridge, A.L.,
Bernier, R.A., Catterall, W.A., Chung, W.K., Empfield, J.R., et al. (2018). Progress in
Understanding and Treating SCN2A-Mediated Disorders. Trends Neurosci. *41*, 442–456.
Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., Rubeis, S. De, An, J.-Y., Peng, M.,
Collins, R., Grove, J., Klei, L., et al. (2019). Large-scale exome sequencing study implicates
both developmental and functional changes in the neurobiology of autism. BioRxiv 484113.
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., et al. (2020). Large-Scale Exome Sequencing Study Implicates
Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568-584.e23.

Schaaf, C.P., Betancur, C., Yuen, R.K.C., Parr, J.R., Skuse, D.H., Gallagher, L., Bernier, R.A., Buchanan, J.A., Buxbaum, J.D., Chen, C.A., et al. (2020). A framework for an evidence-based gene list relevant to autism spectrum disorder. Nat. Rev. Genet. *21*, 367–376.

Schep, A.N., and Kummerfeld, S.K. iheatmapr: Interactive complex heatmaps in R.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon,

S., Krasnitz, A., Kendall, J., et al. (2007). Strong Association of De Novo Copy Number Mutations with Autism. Science (80-. ). *316*, 445–449.

Shaye, D.D., and Greenwald, I. (2011). OrthoList: A Compendium of C. elegans Genes with Human Orthologs. PLoS One *6*, e20085.

Shpigler, H.Y., Saul, M.C., Corona, F., Block, L., Cash Ahmed, A., Zhao, S.D., and Robinson,G.E. (2017). Deep evolutionary conservation of autism-related genes. Proc. Natl. Acad. Sci. U.S. A. *114*, 9653–9658.

Sieburth, D., Ch'ng, Q., Dybbs, M., Tavazoie, M., Kennedy, S., Wang, D., Dupuy, D., Rual, J.-F., Hill, D.E., Vidal, M., et al. (2005). Systematic analysis of genes required for synapse structure and function. Nature *436*, 510–517.

Silverman, J.L., Yang, M., Lord, C., and Crawley, J.N. (2010). Behavioural phenotyping assays for mouse models of autism. Nat. Rev. Neurosci. *11*, 490–502.

Sinclair, D., Oranje, B., Razak, K.A., Siegel, S.J., and Schmid, S. (2016). Sensory processing in autism spectrum disorders and Fragile X syndrome-From the clinic to animal models. Neurosci. Biobehav. Rev.

Solari, F., Bourbon-Piffaut, A., Masse, I., Payrastre, B., Chan, A.M.-L., and Billaud, M. (2005). The human tumour suppressor PTEN regulates longevity and dauer formation in Caenorhabditis elegans. Oncogene *24*, 20–27.

Sorkaç, A., Alcantara, I.C., and Hart, A.C. (2016). In Vivo Modelling of ATP1A3 G316S-Induced Ataxia in C. elegans Using CRISPR/Cas9-Mediated Homologous Recombination Reveals Dominant Loss of Function Defects. PLoS One *11*, e0167963.

Spinelli, L., Black, F.M., Berg, J.N., Eickholt, B.J., and Leslie, N.R. (2015). Functionally distinct groups of inherited PTEN mutations in autism and tumour syndromes. J. Med. Genet. *52*, 128–134.

Splawski, I., Timothy, K.W., Sharpe, L.M., Decher, N., Kumar, P., Bloise, R., Napolitano, C., Schwartz, P.J., Joseph, R.M., Condouris, K., et al. (2004). CaV1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell *119*, 19–31.

Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J.,

Fowler, D.M., Stratton, R., Pilarski, R., et al. (2017). Variant Interpretation: Functional Assays to the Rescue. Am. J. Hum. Genet. *101*, 315–325.

Stephens, G.J., Bueno de Mesquita, M., Ryu, W.S., and Bialek, W. (2011). Emergence of long timescales and stereotyped behaviors in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *108*, 7286–7289.

Stern, S., Kirst, C., and Bargmann, C.I. (2017). Neuromodulatory Control of Long-Term
Behavioral Patterns and Individuality across Development. Cell *171*, 1649-1662.e10.
Stessman, H.A., Bernier, R., and Eichler, E.E. (2014). A genotype-first approach to defining the subtypes of a complex disease. Cell *156*, 872–877.

Stessman, H.A.F., Xiong, B., Coe, B.P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat. Genet. *49*, 515–526.

Stoeger, T., Gerlach, M., Morimoto, R.I., and Nunes Amaral, L.A. (2018). Large-scale investigation of the reasons why potentially important genes are ignored. PLOS Biol. *16*, e2006643.

Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., Ragavendran,
A., Brand, H., Lucente, D., Miles, J., et al. (2014). CHD8 regulates neurodevelopmental
pathways associated with autism spectrum disorder in neural progenitors. Proc. Natl. Acad. Sci.
U. S. A. *111*, E4468–E4477.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev. Biol. *56*, 110–156.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Sun, S., Yang, F., Tan, G., Costanzo, M., Oughtred, R., Hirschman, J., Theesfeld, C.L., Bansal,

P., Sahni, N., Yi, S., et al. (2016). An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome Res. *26*, 670–680.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics *22*, 1540–1542.

Suzuki, H., Kerr, R., Bianchi, L., Frøkjaer-Jensen, C., Slone, D., Xue, J., Gerstbrein, B., Driscoll, M., and Schafer, W.R. (2003). In vivo imaging of C. elegans mechanosensory neurons demonstrates a specific role for the MEC-4 channel in the process of gentle touch sensation.
Neuron *39*, 1005–1017.

Swierczek, N.A., Giles, A.C., Rankin, C.H., and Kerr, R.A. (2011). High-throughput behavioral analysis in C. elegans. Nat. Methods *8*, 592–598.

Tabuchi, K., Blundell, J., Etherton, M.R., Hammer, R.E., Liu, X., Powell, C.M., and Südhof,

T.C. (2007). A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. Science *318*, 71–76.

Tamura, M., Gu, J., Matsumoto, K., Aota, S., Parsons, R., and Yamada, K.M. (1998). Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN. Science *280*, 1614–1617.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947.

Tejedor, F., Zhu, X.R., Kaltenbach, E., Ackermann, A., Baumann, A., Canal, I., Heisenberg, M., Fischbach, K.F., and Pongs, O. (1995). minibrain: A new protein kinase family involved in postembryonic neurogenesis in Drosophila. Neuron *14*, 287–301.

The Autism Spectrum Disorders Working Group of The Psychiatric Genomics (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. Mol. Autism 8, 21.

Thompson, O., Edgley, M., Strasbourger, P., Flibotte, S., Ewing, B., Adair, R., Au, V.,

Chaudhry, I., Fernando, L., Hutter, H., et al. (2013). The million mutation project: A new

approach to genetics in Caenorhabditis elegans. Genome Res. 23, 1749-1762.

Thyme, S.B., Pieper, L.M., Li, E.H., Pandey, S., Wang, Y., Morris, N.S., Sha, C., Choi, J.W.,

Herrera, K.J., Soucy, E.R., et al. (2019). Phenotypic Landscape of Schizophrenia-Associated Genes Defines Candidates and Their Shared Functions. Cell *177*, 478-491.e20.

Tick, B., Bolton, P., Happé, F., Rutter, M., and Rijsdijk, F. (2016). Heritability of autism spectrum disorders: A meta-analysis of twin studies. J. Child Psychol. Psychiatry Allied Discip. *57*, 585–595.

Tilot, A.K., Bebek, G., Niazi, F., Altemus, J.B., Romigh, T., Frazier, T.W., and Eng, C. (2016). Neural transcriptome of constitutional Pten dysfunction in mice and its relevance to human idiopathic autism spectrum disorder. Mol. Psychiatry *21*, 118–125.

Timbers, T.A., Ardiel, E.L., McDiarmid, T.A., Lee, K., Safaei, J., Pelech, S.L., and Rankin, C.H. (2017). CaMK (CMK-1) and O-GlcNAc transferase (OGT-1) modulate mechanosensory responding and habituation in an interstimulus interval-dependent manner in Caenorhabditis elegans. BioRxiv.

Tomioka, M., Adachi, T., Suzuki, H., Kunitomo, H., Schafer, W.R., and Iino, Y. (2006). The insulin/PI 3-kinase pathway regulates salt chemotaxis learning in Caenorhabditis elegans. Neuron *51*, 613–625.

Tong, X.-J., Hu, Z., Liu, Y., Anderson, D., Kaplan, J.M., Ashby, M., Maier, S., Nishimune, A., Henley, J., Atasoy, D., et al. (2015). A network of autism linked genes stabilizes two pools of synaptic GABA A receptors. Elife 4, e09648.

Tong, X.-J., López-Soto, E.J., Li, L., Liu, H., Nedelcu, D., Lipscombe, D., Hu, Z., and Kaplan, J.M. (2017). Retrograde Synaptic Inhibition Is Mediated by α-Neurexin Binding to the α2δ Subunits of N-Type Calcium Channels. Neuron *95*, 326-340.e5.

Troulinaki, K., Büttner, S., Marsal Cots, A., Maida, S., Meyer, K., Bertan, F., Gioran, A.,

Piazzesi, A., Fornarelli, A., Nicotera, P., et al. (2018). WAH-1/AIF regulates mitochondrial

oxidative phosphorylation in the nematodeCaenorhabditis elegans. Cell Death Discov. 4, 2.

Varga, E.A., Pastore, M., Prior, T., Herman, G.E., and McBride, K.L. (2009). The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. Genet. Med. *11*, 111–117.

Veenstra-Vanderweele, J., and Blakely, R.D. (2012). Networking in autism: Leveraging genetic, biomarker and model system findings in the search for new treatments.

Neuropsychopharmacology 37, 196–212.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., and Zhang, F.P. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905–914.

van der Voet, M., Nijhof, B., Oortveld, M.A.W., and Schenck, A. (2014). Drosophila models of early onset cognitive disorders and their clinical applications. Neurosci. Biobehav. Rev. 46.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M.,

Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–386.

Volkmar, F.R., and McPartland, J.C. (2014). From Kanner to DSM-5: Autism as an Evolving

Diagnostic Concept. Annu. Rev. Clin. Psychol. 10, 193–212.

Wagnon, J.L., Briese, M., Sun, W., Mahaffey, C.L., Curk, T., Rot, G., Ule, J., and Frankel, W.N.(2012). CELF4 Regulates Translation and Local Abundance of a Vast Set of mRNAs, IncludingGenes Associated with Regulation of Synaptic Function. PLoS Genet. 8.

Wang, H., Liu, J., Gharib, S., Chai, C.M., Schwarz, E.M., Pokala, N., and Sternberg, P.W.

(2016). cGAL, a temperature-robust GAL4–UAS system for Caenorhabditis elegans. Nat. Methods 14, 145–148.

Wang, J., Al-Ouran, R., Hu, Y., Kim, S.-Y., Wan, Y.-W., Wangler, M.F., Yamamoto, S., Chao,
H.-T., Comjean, A., Mohr, S.E., et al. (2017a). MARRVEL: Integration of Human and Model
Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. Am. J.
Hum. Genet. *100*, 843–853.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164–e164.

Wang, S., Tang, N.H., Lara-Gonzalez, P., Zhao, Z., Cheerambathur, D.K., Prevo, B., Chisholm, A.D., Desai, A., and Oegema, K. (2017b). A toolkit for GFP-mediated tissue-specific protein degradation in C. elegans. Development.

Wangler, M.F., Yamamoto, S., Chao, H.-T., Posey, J.E., Westerfield, M., Postlethwait, J., Hieter,
P., Boycott, K.M., Campeau, P.M., Bellen, H.J., et al. (2017). Model Organisms Facilitate Rare
Disease Diagnosis and Therapeutic Research. Genetics 207, 9–27.

Ward, S. (1973). Chemotaxis by the nematode Caenorhabditis elegans: identification of attractants and analysis of the response by use of mutants. Proc. Natl. Acad. Sci. U. S. A. 70, 817–821.

Weile, J., and Roth, F.P. (2018). Multiplexed assays of variant effects contribute to a growing

genotype-phenotype atlas. Hum. Genet. 137, 665-678.

Weile, J., Sun, S., Cote, A.G., Knapp, J., Verby, M., Mellor, J.C., Wu, Y., Pons, C., Wong, C., van Lieshout, N., et al. (2017). A framework for exhaustively mapping functional missense variants. Mol. Syst. Biol. *13*, 957.

Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., et al. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat. Genet. *49*, 978–985.

Werling, D.M., and Geschwind, D.H. (2013). Understanding sex bias in autism spectrum disorder. Proc. Natl. Acad. Sci. U. S. A. 110, 4868–4869.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *314*, 1–340.

Wickham, H. (2009). Introduction. In Ggplot2, (New York, NY: Springer New York), pp. 1–7.
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., et al. (2013). Coexpression networks implicate human midfetal
deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997.

Winden, K.D., Ebrahimi-Fakhari, D., and Sahin, M. (2018). Abnormal mTOR Activation in Autism. Annu. Rev. Neurosci. *41*, 1–23.

Wong, W.-R., Brugman, K.I., Maher, S., Oh, J.Y., Howe, K., Kato, M., and Sternberg, P.W. (2019). Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. Hum. Mol. Genet. *28*, 2271–2281.

Xu, X., Wells, A.B., O'Brien, D.R., Nehorai, A., and Dougherty, J.D. (2014). Cell type-specific

expression analysis to identify putative cellular mechanisms for neurogenetic disorders. J. Neurosci. *34*, 1420–1431.

Yang, F., Sun, S., Tan, G., Costanzo, M., Hill, D.E., Vidal, M., Andrews, B.J., Boone, C., and Roth, F.P. (2017). Identifying pathogenicity of human variants via paralog-based yeast complementation. PLoS Genet. *13*, e1006779.

Yemini, E., Jucikas, T., Grundy, L.J., Brown, A.E.X., and Schafer, W.R. (2013). A database of Caenorhabditis elegans behavioral phenotypes. Nat. Methods *10*, 877–879.

Yu, H., Aleman-Meza, B., Gharib, S., Labocha, M.K., Cronin, C.J., Sternberg, P.W., and Zhong,
W. (2013). Systematic profiling of Caenorhabditis elegans locomotive behaviors reveals
additional components in G-protein G q signaling. Proc. Natl. Acad. Sci. *110*, 11940–11945.
Zafeiriou, D.I., Ververi, A., Dafoulis, V., Kalyva, E., and Vargiami, E. (2013). Autism spectrum
disorders: The quest for genetic syndromes. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. *162*, 327–366.

Zaslavsky, K., Zhang, W.-B., McCready, F.P., Rodrigues, D.C., Deneault, E., Loo, C., Zhao, M., Ross, P.J., El Hajjar, J., Romm, A., et al. (2019). SHANK2 mutations associated with autism spectrum disorder cause hyperconnectivity of human neurons. Nat. Neurosci. 22, 556–564.
Zhang, L., Ward, J.D., Cheng, Z., and Dernburg, A.F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development *142*, 4374–4384.

Zhou, W., Chen, T., Chong, Z., Rohrdanz, M.A., Melott, J.M., Wakefield, C., Zeng, J., Weinstein, J.N., Meric-Bernstam, F., Mills, G.B., et al. (2015). TransVar: A multilevel variant annotator for precision genomics. Nat. Methods *12*, 1002–1003.

Zhou, Y., Sharma, J., Ke, Q., Landman, R., Yuan, J., Chen, H., Hayden, D.S., Fisher, J.W.,

Jiang, M., Menegas, W., et al. (2019). Atypical behaviour and connectivity in SHANK3-mutant macaques. Nature *570*, 326–331.

Zoghbi, H.Y., and Bear, M.F. (2012). Synaptic Dysfunction in Neurodevelopmental Disorders Associated with Autism and Intellectual Disabilities. Cold Spring Harb. Perspect. Biol. *4*, a009886–a009886.