

Computational Framework for the Identification of
Neural Circuits Underlying Psychiatric Disorders

Jonathan Chang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

© 2020

Jonathan Chang

All Rights Reserved

Abstract

Computational Framework for the Identification of Neural Circuits Underlying Psychiatric Disorders

Jonathan Chang

Autism spectrum disorders (ASDs) are characterized by phenotypic and genetic heterogeneity. Our analysis of functional networks perturbed in ASD suggests that both truncating and non-truncating *de novo* mutations contribute to autism. Moreover, we find that truncating mutations affecting the same exon lead to strikingly similar intellectual phenotypes in unrelated ASD probands and propose that exons, rather than genes, represent a unit of effective phenotypic impact for truncating mutations in autism. The phenotypic effects are likely mediated by nonsense-mediated decay of splicing isoforms and similar patterns may be observed in other genetic disorders. While multiple cell types and brain areas are affected, the impact of ASD mutations converge on a strongly interconnected system of neural structures that involve basal ganglia loops and the limbic system. We observe that distant projections constitute a disproportionately large fraction of the network composition, suggesting that the integration of diverse brain regions is a key property of the neural circuit. We demonstrate that individual *de novo* mutations impact several disparate components of the network and may further explain the phenotypic variability. Overall, our study presents a method that, to our knowledge, is the first unbiased approach using genetic variants to comprehensively discover and identify the neural circuitry affected in a psychiatric disorder.

Table of Contents

List of Figures	iii
Chapter 1: Introduction	1
1.1 Research Questions	1
1.2 Limitations	3
Chapter 2: Background	6
2.1 Core Symptoms.....	6
2.2 Epidemiological Features.....	6
2.3 Genetic Architecture	7
2.4 Molecular Properties and Biological Pathways	12
2.5 Brain Regions and Circuits	14
2.6 Intellectual and Behavioral Phenotypes	18
Chapter 3: Convergence on Functional Gene Networks.....	20
3.1 Functional Gene Networks Affected By <i>De Novo</i> Mutations	20
3.2 Association of the Implicated Genes with Specific Functional Subsets.....	24
3.3 Temporal and Spatial Brain Expression Patterns of Implicated Genes	27
3.4 Functional Properties of Implicated Genes and Disease Phenotypes	33
3.5 Implications for Neural Circuits and Phenotype Stratification.....	35
Chapter 4: Mutations Target Neural Circuits.....	38
4.1 ASD-associated Brain Structures are Highly Connected.....	38
4.2 Mutations Specifically Target Strongly Interconnected Structures	46

4.3 Mutations Target Cortical and Subcortical Neurons	47
4.4 Functionally Ascertained Structures Maintain Interconnectivity	49
4.5 Mutations Converge on Neural Circuits	52
4.6 ASD Circuit Integrates Distant and Functionally Diverse Regions.....	55
4.7 Small and Diverse Fraction of ASD Circuit is Perturbed in Each Proband.....	57
4.8 Relation to Known Neural Systems and Implications for Medicine	58
Chapter 5: Sensitivity of Phenotypes to Gene Dosage	62
5.1 Exons as Units of Phenotypic Impact for Mutations	62
5.2 Variation in Sensitivity of Phenotypes to Gene Dosage.....	66
5.3 Implications for Precision Medicine.....	69
References.....	73
Appendix A.....	92

List of Figures

- Figure 1.1 Structural levels of organization in the nervous system.
- Figure 2.3.1 Genetic architecture.
- Figure 2.3.2 Distinct Acquisition Mechanisms.
- Figure 2.3.3 Rates of *De Novo* Events by Mutational Type.
- Figure 2.5 Topography of Cortico-Striatal and Cortico-Thalamic Projections.
- Figure 3.1.1 Mutations Implicate Functional Gene Network.
- Figure 3.1.2 Genes Located at the Synapse or Associated with the Morphogenesis of Dendritic Spines.
- Figure 3.3 Cell-type Expression Biases for ASD Mutations.
- Figure 4.1.1: Flowchart of Expression and Circuit Analysis.
- Figure 4.1.2: Expression and Connectivity Analysis Diagram.
- Figure 4.1.3: ASD Structures are Strongly Interconnected.
- Figure 4.1.4: Proportion of Brain Regions Affected by ASD Mutations.
- Figure 4.3: Expression Bias Correlations.
- Figure 4.4: Expression Biases of the Most Strongly Biased Structures.
- Figure 4.5.1: Search Algorithm Diagram.
- Figure 4.5.2: Implicated ASD Structures Converge on Neural Circuits.
- Figure 4.5.3: Search Algorithm Finds Maximally Cohesive Circuits.
- Figure 4.6: ASD Circuit Integrates Functionally Diverse Neural Structures.
- Figure 4.8: ASD Circuit Involves Cortico-Striatal-Thalamo-Cortical (CSTC) Loop and Limbic System.
- Figure 5.1 Average Difference in IQ Between SSC Probands.
- Figure 5.2 Between the Relative Expression of Exons Harboring LGD Mutations and the Corresponding Decrease in Probands' Intellectual Phenotypes.

Chapter 1: Introduction

1.1 Research Questions

Psychiatric disorders such as autism spectrum disorders (ASD) are associated with a wide range of cognitive and behavioral abnormalities. However, for ASD and many other psychiatric disorders alike, the genotype to phenotype mechanisms and in particular, the underlying neural circuits and systems, are not fully understood and remain an active area of research.

We aim to understand the processes by which disease mutations manifest as and contribute to the associated phenotypes of psychiatric disorders by examining the effects of genetic perturbations at successive levels of biological organization (Figure 1.2), as is often described (Churchland and Sejnowski, 1988). In this manner, diseases may be viewed as an information flow problem, and dysfunction at the genetic level leads to downstream effects that ultimately result in divergent phenotypes. We start our investigation with mutations harbored in individuals with psychiatric disorders that have a strong likelihood of pathogenicity. In particular, we are interested in the levels of abstraction where available data makes a compelling case for exploration. We first investigate the effects of mutations at the level cellular pathways, then investigate the effects at the level of neural cell types and mesoscale structures of the brain. Next, we analyze how these effects propagate to the level of neural circuits in the brain. Finally, we explore the intellectual and behavioral phenotypes among a psychiatric disorder for which detailed phenotypic data is available, and the molecular and genic properties that influence them.

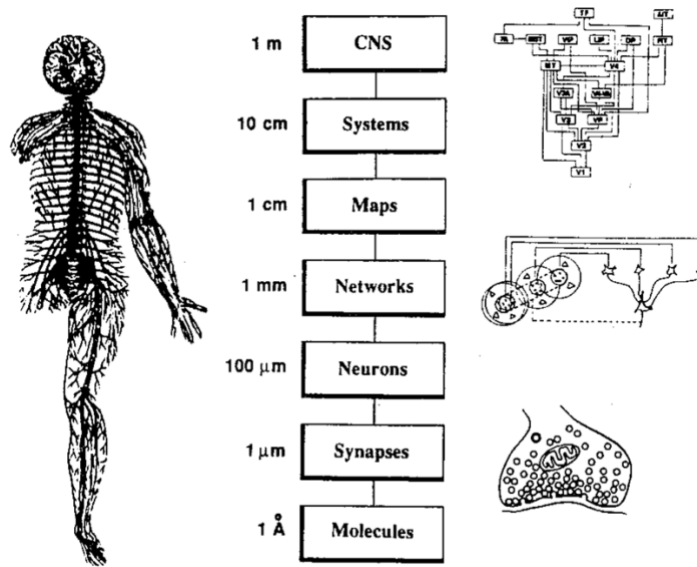


Figure 1.1: Structural levels of organization in the nervous system (Churchland and Sejnowski, 1988). (Left) Drawing by Vesalius of the human brain, the spinal column, and the peripheral nerves. Schematic diagrams illustrating (Top Right) a processing hierarchy of visual areas in monkey visual cortex; (Center Right) a small network model for the synthesis of oriented receptive fields of simple cells in visual cortex; and (Bottom Right) the structure of a chemical synapse.

To investigate the underlying biological pathways, we used a previously developed computational approach that searches for cohesive biological networks among a diverse collection of disease-associated genetic variants. Our network-based approach determines whether genetic variations associated with ASDs converge on biological networks. We subsequently explore the functional and transcriptional properties of network component genes.

We then ask which regions of the brain are involved in ASD using an unbiased analysis, as many regions of the brain are suspected to be involved in ASD. Others have observed abnormalities such as increased cerebral volumes and amygdalar enlargement (Sparks et al., 2002); however, hypothesis-driven studies investigating the same brain regions often do not provide concordant results. Rather than examine regions individually, we instead probe the brain in its entirety in an unbiased manner. We reason that investigating the brain at the scale of

regions might be intractable, as each brain region will involve numerous neural structures and neuronal cell types that behave and perform differently. Rather, it may be more opportune to examine the brain at a scale closer to the functional units of the brain. We hypothesize that ASD-associated mutations disproportionately impact mesoscale neural structures.

Much in the same way proteins do not act in isolation at the cellular level, we propose that ASD-associated mutations should converge on strongly interconnected neural systems in the brain and can ask which neuronal circuits are involved in ASD. Studies that use functional and diffusion MRI methods to implicate regions and circuits often do not consider genetic causes, and therefore, it is difficult to distinguish whether differential observations contribute to the observed behavioral phenotype or are pathophysiological outcomes of the disorder. By starting from changes at the level of the genome, any biases in neural circuitry that we observe should, in theory, be contributory. More specifically, we hypothesize that disease-associated mutations pathologically target particular neural circuits in the brain.

Finally, we investigate the phenotypic properties of the disease and explore the relationship between intellectual and behavioral phenotypes observed in probands and the various properties of the respective genes harboring mutations, thereby furthering our understanding of the mechanisms by which genotype contributes to phenotype. While the associated phenotypes are heterogenous and varied, we hypothesize that ASD mutations precipitate changes in phenotypes in a discriminatory manner.

1.2 Limitations

The complex genetic architecture of many psychiatric disorders is a stark reality, and many have yet to be fully resolved and understood. Assuming the genotype to phenotype relationship is a deterministic process, unraveling the complexity is of keen scientific interest.

Elucidating neural circuits may be a plausible route to understanding the mechanistic processes underlying ASD; however, several technological challenges exist. One limitation of the proposed work is the fraction of ASD risk genes that are still unknown. Of the total set of risk genes, others have estimated that approximately a quarter of the ASD genes have been discovered. However, it is possible that most of the involved biological pathways have been identified, and thus, it may not be necessary to uncover the remaining genes in order to successfully implicate neural cell types and circuits. In one of the first studies using network-based methods to investigate ASD-associated mutation events (Gilman et al., 2011a), several major biological processes were identified and include synapse development, axon targeting, and neuron motility. Several papers have since confirmed these findings and have not introduced additional processes that are significantly different from those already found.

Another limitation is the lack of complete human brain data. There are several technological limitations associated with generating human brain expression and projection data, including but not limited to low spatial resolution of MRI imaging techniques and difficulty in acquiring neurotypical post-mortem brain samples. Brain-wide human data at the resolution currently available for mice is unlikely to be available anytime soon. While comprehensive projection data is currently available for the mouse model organism, murine brain data is not ideal as there are several notable differences in neuroanatomy such as the size of the prefrontal cortex and the rates of neurogenesis in various structures of the brain (Ernst and Frisé, 2015). Moreover, it is not currently known how different the neural cell types and circuits may be between species. However, the murine brain has been previously studied for many neurological disorders, including Rett syndrome, Alzheimer's, Huntington, and Parkinson's. Additionally, there are numerous mouse models of ASD, and behavioral profiles of mice exhibiting ASD-like

symptoms are well characterized and documented (Silverman et al., 2010). Mouse studies are therefore scientifically relevant, and our findings using murine data may be a useful first look at the neural systems underpinning ASD and may provide a first look that may translate to the human brain.

Chapter 2: Background

2.1 Core Symptoms

ASD is a neuropsychiatric disorder characterized by two core symptoms, as stipulated by DSM-V: (1) social and communication impairments and (2) repetitive behaviors. Social deficits often include an inability to recognize and respond to social stimuli or subtle cues, difficulty maintaining eye contact, and problems with recognizing emotions and faces. Language impairments include late onset of speech and gesturing, diminished responsiveness, and difficulty with complex language patterns. Repetitive behaviors include stereotypy, but also more general behaviors such as compulsion towards certain activities, resistance to changes or interruptions, and a restrictive set of interests and activities. Individuals diagnosed with autism exhibit a wide range of severity among these traits. Asperger syndrome and pervasive developmental disorder, not otherwise specified (PDD-NOS) are included in the spectrum.

2.2 Epidemiological Features

ASD has a high prevalence, estimated at around 1% of the population, and the current estimate provided by the CDC is 1 in 68. As ASD is also a neurodevelopmental disorder, the onset of symptoms occurs early and ASD can be reliably diagnosed by age two. There are several interesting epidemiological features of ASD. The first is that there is a predominant gender bias, as evidenced by the incidence of ASD is 4-fold higher in boys than in girls. Secondly, a large fraction of individuals with ASD also have comorbid conditions. It is estimated that 70% of autistic patients show comorbidities, of which intellectual disability, epilepsy, ADHD, bipolar, and gastrointestinal problems are the most common.

2.3 Genetic Architecture

ASD has a large genetic component with heritability estimated at 40-90% (Bourgeron, 2015; Gaugler et al., 2014). Earlier twin studies estimated heritability based on the difference in concordance between monozygotic and dizygotic twins while more recent SNP-based studies estimated the variance in genetic liability explained by genome-wide SNPs; however, both methods arrived at similar heritability estimates. Primarily, the types of variants studied in the context of ASD include mutations with Mendelian/monogenic forms of inheritance, common variants that confer risk *en masse*, and highly-penetrant *de novo* copy-number variations (CNVs) and single-nucleotide variants (SNVs). While these variant types explain the majority of liability in ASD, approximately 41% is still unaccounted for, as shown in Figure 2.3.1 (Gaugler et al., 2014). The observed propensity of associated variants suggests that there are primarily two distinct acquisition mechanisms, as shown in Figure 2.3.2. In the first, a statistically unlikely but highly-penetrant *de novo* mutation manifests in a simplex case of ASD. This acquisition mechanism follows Mendelian genetics, where susceptibility due to inherited common variants is minimal and risk is largely conferred through the *de novo* mutation. In the second, inherited common variants lead to multiple cases in multiplex families. Here, the risk profile is reversed, and *de novo* mutations may not occur and bear little relevance while common variants confer significant contributions, thereby exemplifying quantitative genetics or Fisher's geometric/infinitesimal model. It is suspected that most cases of ASD are a result of a combination of *de novo*, rare, and common variants, and in such a way that both acquisition mechanisms are relevant (Sanders et al., 2015).

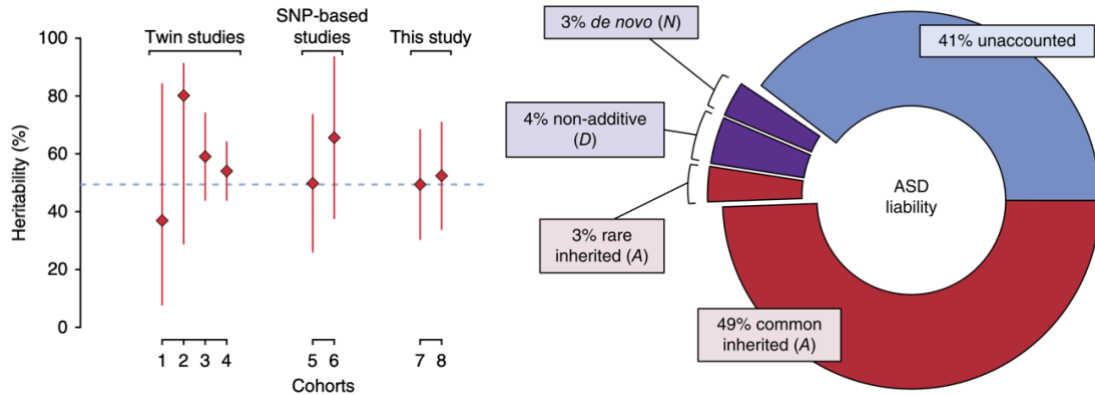


Figure 2.3.1: Genetic architecture (Gaugler et al., 2014). (Left) Heritability estimates compared across study designs and analytical methods. The error bars represent 95% confidence interval. (Right) Genetic architecture of ASD, whereby variance in liability is determined by genetic and environmental factors. The genetic factors include additive effects (A), non-additive effects (D; dominant, recessive, epistatic) and *de novo* mutations (N). Environmental factors are split between common or shared environment and stochastic or unique environment.

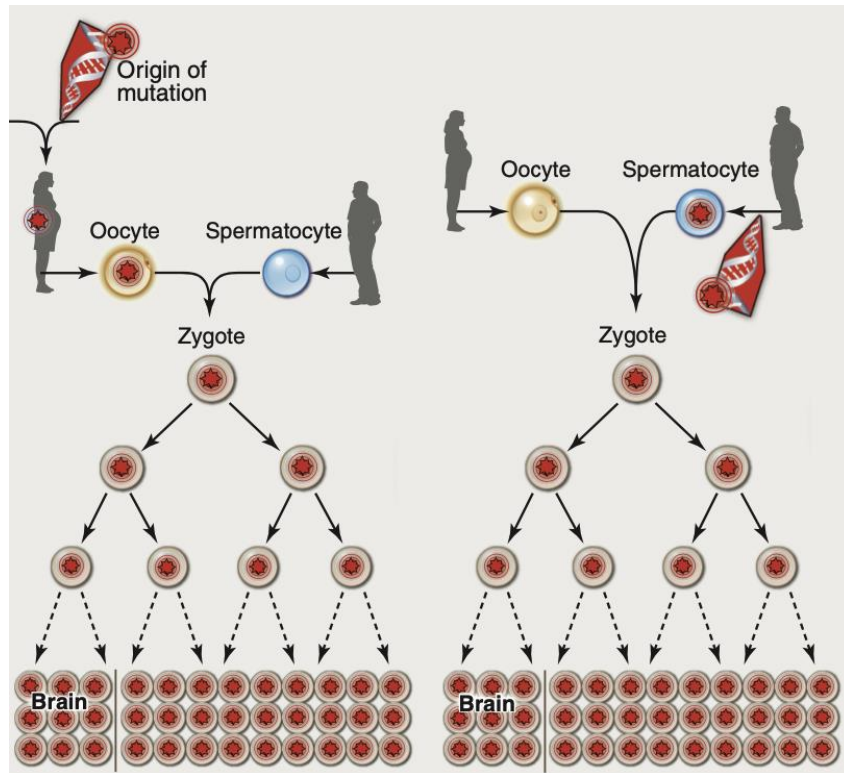


Figure 2.3.2: Distinct Acquisition Mechanisms (Poduri et al., 2013). (Left) A heterozygous mutation is inherited from one parent. This mechanism is typical of autosomal

dominant epilepsy. In this example, the mutation originally presented in the mother, whose oocytes in turn carry the mutation. It is present in the zygote and thus all cells of the affected child. (Right) A *de novo* mutation may arise sporadically during gametogenesis, in this case spermatogenesis. Even though every cell in the individual carries the mutation, the predominant effects of the mutation depend on the distribution of gene expression; in these examples, the brain is primarily affected.

The earliest studied cases of ASD were Mendelian diseases that greatly increased the risk of ASD. Monogenic diseases such as fragile X syndrome, Rett syndrome, tuberous sclerosis complex, and Timothy syndrome (with insults in the genes *FMRI*, *MECP2*, *TSC1/TSC2*, and *CACNA1C*, respectively) are highly comorbid for ASD (Chen et al., 2015). Additionally, large chromosomal abnormalities were also observed in individuals with ASD from precursory cytogenetic studies, with the most frequent abnormalities being 15q11–13 duplications and 16p11.2 micro-deletions/duplications. In total, these cases account for approximately 10-20% of ASD cases (Chen et al., 2015).

In the absence of classical Mendelian inheritance, ASD was proposed to be a polygenic trait involving many common variants. However, among several genome-wide association studies (GWAS), only one significant locus (5p14.1) has been consistently found to be genome-wide significant (Chen et al., 2015; Wang et al., 2009). The shortfall can be partially attributed to small study sizes and heterogeneity of the disease, but more importantly, this *missing heritability* among common variants may be due to genetic architecture. Others have suggested that common variants *en masse* confer strong risk for ASD (Gaugler et al., 2014), but each variant confers a miniscule amount of risk individually, making it exceptionally difficult to implicate individual SNPs.

Among GWAS, core functional loci explain a small fraction of the expected genetic heritability. Conversely, association signal tends to be scattered throughout the whole genome and do not necessarily cluster in a small number of key functional pathways. In this context, one

paper proposed a conceptual “omnigenic” model (Boyle et al., 2017) to explain the aforementioned patterns of genetic contribution observed in common human diseases, and may also be pertinent to ASD. Briefly, the omnigenic model suggests that for a given genetic phenotype, the contribution from core functional loci is often supplanted by contributions from indirect, trans-regulatory effects of thousands of peripheral genes. Because there are considerably more peripheral effects, they may also explain up to 70-90% of heritability. Under this model, multiple variants in peripheral genes may influence genetic phenotypes by regulating the expression of core genes and their functions; however, understanding how many thousands of functionally unrelated genomic loci distributed across the genome regulate the expression of core genes remains a challenging task.

Comparative genomic hybridization (CGH) studies showed that *de novo* copy number variants (CNVs) were strongly enriched in individuals with ASD (Marshall et al., 2008). These *de novo* structural mutations occur at the parental germline or in early embryonic somatic cells and are often the only *de novo* mutation in sporadic cases affecting the exome. Interestingly, these mutations have a large relative risk and are highly penetrant. These mutations are 2-7x more abundant in patients and the harbored CNVs tend to be larger and encompass a greater number of genes. There are several cytobands where *de novo* CNV events commonly occur, including 16p11.2, 22q11.2, 1q21.1, and 7q11.3 (Levy et al., 2011a; Sanders et al., 2011b). Interestingly, 7q11 duplications are associated with ASD whereas 7q11 deletions lead to Williams-Beuren syndrome, a distinct disorder characterized by an aortic heart defect, distinctive facial features, mental retardation, and hyper-socialization.

One of the first large-scale sequencing study surveyed 2,600 probands for *de novo* single nucleotide variants (SNVs) and CNVs in sporadic cases of ASD (Iossifov et al., 2014a). *De novo*

SNVs in the general population are well characterized. Approximately 60 mutations accrue in any given individual, of which one mutation will on average occur in the exome (Conrad et al., 2011). In the Simons Simplex Collection (SSC), approximately 1,500 missense SNVs and 400 likely gene-disrupting (LGD) SNVs were observed in ASD probands. Because the criteria of the study restricted enrollment to families with unaffected relatives, ASD probands were strongly enriched for pathogenic *de novo* mutations. As with *de novo* CNVs, these mutations were significantly overabundant among probands compared to mutations observed in unaffected siblings, as shown in Figure 2.3.3 (Iossifov et al., 2014a). Moreover, the ascertainment bias for LGD SNVs is much greater than for missense SNVs (0.43 vs. 0.14). Based on the number of recurrent mutations occurring in the same gene, the study estimates that approximately 800 ASD risk genes exist. Among ASD *de novo* mutations, two out of every three SNVs have paternal origins, emphasizing the importance of the father's age with regards to disease propensity (Iossifov et al., 2015; Kong et al., 2012). Conversely, inherited LGD mutations have been shown to have a strong maternal transmission bias (Krumm et al., 2015a), possibly due to a female protective mechanism that is largely unexplained.

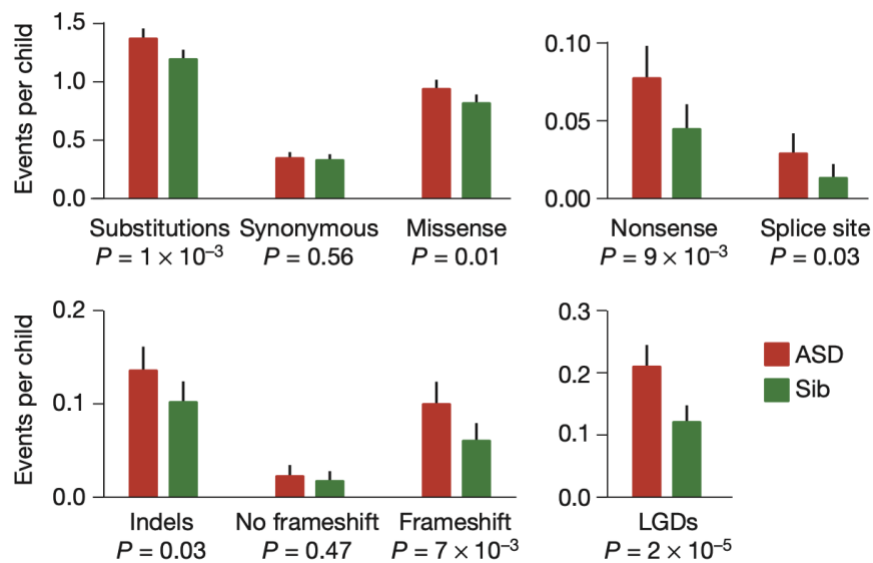


Figure 2.3.3: Rates of *De Novo* Events by Mutational Type (Iossifov et al., 2014a). Mutation types for probands in the SSC are displayed by class, and the combined rate for all LGDs is shown at the bottom right. For each event type, the significance between probands and unaffected siblings (Sib) is given. The errors bars represent 95% confidence interval for the mean rates.

Some initial studies have investigated other mutation classes. Early work has shown that incidents of ASD may be due to somatic mosaic mutations in the brain that occur in critical or regulatory genes such as *MTOR* (Bae et al., 2015). Mutations were also identified in non-coding regions such as human-accelerated regions (HAR) that contain regulatory sequences and are active in the brain, regulating various neurodevelopmental processes (Doan et al., 2016). However, another study (C Yuen et al., 2017) was not able to conclusively pinpoint any non-coding regions. It is possible that genetic mosaicism and non-coding mutations explain a significant fraction of the missing heritability, however, the degree to which cases of ASD originate from these modalities is not accurately known.

2.4 Molecular Properties and Biological Pathways

Several molecular properties were identified among ASD-associated genes. ASD mutations tend to occur in genes with strong genic intolerance to functional variation as well as in haploinsufficient genes (i.e., loss of function with a single functional allele). For instance, one study used a gene-based *de novo* mutational model to show that genes with ASD *de novo* mutations are under selective evolutionary constraint (Samocha et al., 2014). Based on a calibrated model of background variation, the authors observed a significant excess of *de novo* loss-of-function mutations among ASD probands. Another similar study used a likelihood model with per-gene mutation rates to identify genes with excess mutations (De Rubeis et al., 2014) and observed that, in addition to *de novo* loss-of-function mutations, the amount of *de novo*

damaging missense and inherited loss-of-function variants in an ASD cohort were also significantly in excess of the expected number of mutations.

Several studies successfully identified and mapped the cellular mechanisms involved in ASD. Many such pathway-based methods adhere to the *many genes, common pathways* hypothesis, in which mutations observed in individuals with a particular heritable disease are not randomly dispersed throughout the genome and are instead convergent on a number of biological pathways and processes. One of the first studies used a network-based method to identify a biological network of strongly related genes affected by rare *de novo* CNVs in ASD (Gilman et al., 2011a). After building a background network that encapsulates the *a priori* expectation that two genes participate in the same genetic phenotype, the authors searched for functionally connected clusters of genes with observed mutations. They identified a statistically significant gene network that was strongly enriched for functions primarily related to synapse development, axon targeting, and neuron motility, and involved in the WNT signaling, AKT/mTOR, and reelin pathways.

Another such study examined the protein-protein interactions among genes with *de novo* SNVs, and found that much of the ASD-associated gene network involves a highly interconnected chromatin remodeling protein network (O'Roak et al., 2012b). The network was also enriched for β -catenin and p53 signaling, chromatin remodeling, ubiquitination machinery, and neuronal development. A different study mapped various ASD-associated genes onto co-expression networks spanning the developmental range of the brain (Parikshak et al., 2013). These co-expression modules were enriched for biological functions that are active during human cortical development, namely transcriptional regulation and synaptic development. Another study combined *de novo* CNVs and loss-of-function SNVs and using various functional-

enrichment association tests and pathway analyses found that the affected genes converged on networks related to neuronal signaling and development, synapse function, and chromatin regulation (Pinto et al., 2010b). An aforementioned study found that genes with ASD mutations were significantly enriched for genes encoding messenger RNAs targeted by RNA-binding proteins FMRP and RBFOX, pathways which are involved in post-transcriptional splicing and translation (De Rubeis et al., 2014). Moreover, other studies have shown that ASD mutations were strongly enriched for targets of FMRP RNA-binding protein (Darnell et al., 2011; Parikshak et al., 2013; Samocha et al., 2014), which is disrupted in fragile X syndrome, a commonly comorbid condition.

2.5 Brain Regions and Circuits

Studies based on transcriptional profiling and MRI measurements have been used to implicate brain regions or structures in ASD. One such expression study compared ASD and neurotypical postmortem human brains using differential expression analysis (Voineagu et al., 2011). The authors found that patterns of gene expression were attenuated in the frontal and temporal cortex of the ASD brain, suggesting abnormal development of these regions as possible drivers of phenotype in individuals with ASD. With similar objectives, a co-expression study used expression data from BrainSpan, a brain-wide and development-focused data set, and cortical laminar/layer-level data that they generated (Parikshak et al., 2013). The authors used Weighted Gene Co-Expression Network (WGCNA) analysis to discover gene modules with shared expression patterns and differential analysis to determine genes that are expressed with high cortical layer specificity. These gene modules generated using neurotypical data were then compared against ASD risk genes for enrichment. The authors found that ASD genes were enriched in modules involved in cortical development, specifically during the periods of

neuronal migration and cortical lamination, and in genes specific to layers 2-4 of the cortex. They further speculated that disruption occurs at the level of cortical-cortical circuits. Another study used a set of high-confidence ASD genes to seed a co-expression network built from BrainSpan expression data (Willsey et al., 2013b). Here, the authors assembled multiple networks spanning both spatial and temporal dimensions and selected two that were enriched for held-out ASD genes. These two networks implicated the prefrontal and primary motor-somatosensory cortices during mid-fetal development (10-24 weeks post conception) as critical to the etiology of ASD. Moreover, the study used the enriched networks to implicate deep-layer neuronal cell types. However, as both co-expression studies acknowledge, extricating cell type or structure-specific expression information solely from brain region tissue homogenates is an intractable task. Both these studies use WGCNA to infer modules of genes, however, it is unclear whether such methods can capture the full heterogeneity of cell type or structure-specific expression. Furthermore, expression studies are limited by the spatial resolution of the data and often samples are broad swaths of the brain. Profiling gross tissue homogenates incurs information loss because expression is aggregated across several unrelated and distinct structures and cell types which have varying transcriptional profiles, electrophysiochemical properties, and functional connections. As such, studies that use higher spatial resolution expression data for neural structures that are a closer representation of the cardinal units of the brain may help to pinpoint dysfunction in the ASD brain.

Neuroanatomical differences in ASD probands have also been investigated through functional magnetic resonance imaging (fMRI) studies. fMRI measures oxygenated blood flow as a proxy for glucose metabolism and neuronal activity. In general, these studies are structured such that a set of ASD individuals and an appropriate set of unaffected controls perform a

defined task, and the observed fMRI activations are compared. However, fMRI studies appear to provide numerous contradictions, although the most common observations are hypoactivation of the prefrontal cortex and widespread reduction in brain connectivity. Furthermore, these studies are biased towards high-functioning adults due to the demanding nature of the fMRI scanning procedure. Researchers also speculate that ASD individuals may have developed compensatory processing methods to cope with symptoms, thereby confounding behavioral studies such as task-based fMRI studies. If so, it would be intractable to disentangle causal dysfunction from secondary adaptations among the observed connectivity differences. Furthermore, current fMRI technology has limited resolution and implication of finer brain structures is technologically not feasible.

Studies have mapped brain-wide connectomes at the macroscale (Glasser et al., 2016; Toga et al., 2012) and at the mesoscale (Oh et al., 2014). The connectome denotes a comprehensive reconstruction of the neuronal circuits in the brain, often described as a wiring diagram that maps all connections throughout the entire brain. Functionally, neuronal fiber tracts serve to transmit information between one region or structure to another. Macroscale connectomes have found that brain connectomes appear to be either small-world and broad-scale (i.e., power law degree distribution with a sharp cutoff) networks (Iturria-Medina et al., 2008). Using the same network formalisms to understand whole brain architecture, connectomes appear to have evolved to maintain greater efficiency for local communication than for global communication. Mesoscale connectomes, as shown in Figure 2.5, also have found that brain connectomes have a high clustering coefficient and appear to be small-world and scale-free networks. Connectivity strengths, on the other hand, appear to follow a lognormal distribution and span a 10^5 -fold range (Oh et al., 2014). Furthermore, projections at the mesoscale appear to

be strongly correlated across certain structures. Complete microscale connectomes have only been completed for the nematode *C. elegans*, which only has 302 neurons and 7,000 synapses, and individual neurons and neuronal circuits have been successfully mapped to behavior (Szigeti et al., 2014). However, partial microscale connectomes have been completed for segments of brain tissue for higher-order organisms.

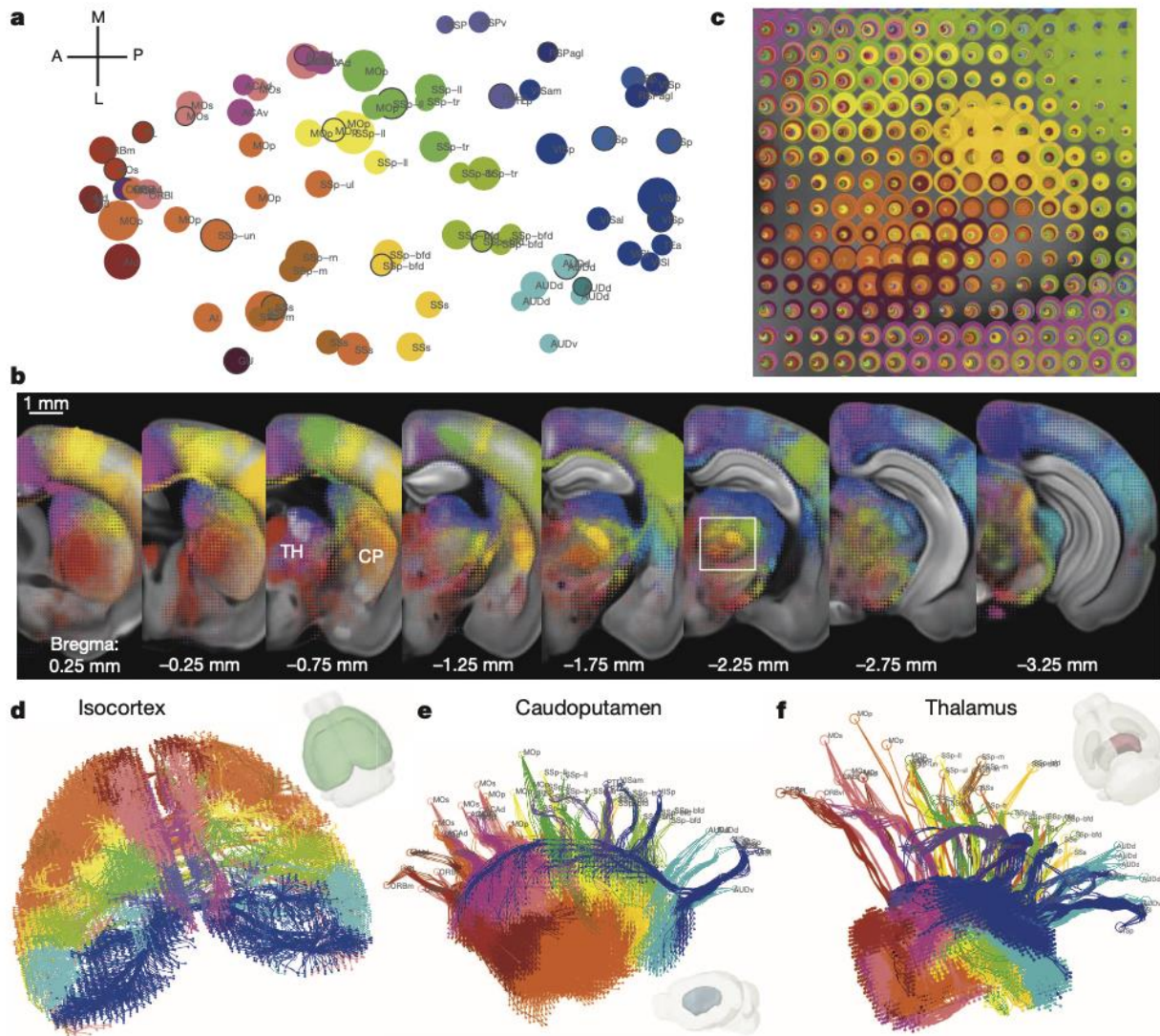


Figure 2.5: Topography of Cortico-Striatal and Cortico-Thalamic Projections (Oh et al., 2014). (A) Cortical domains in the cortex flat-map. Each circle represents one of 80 cortical injection experiments, whose location is obtained via multidimensional scaling from 3D to allow visualization of all the sites in one 2D plane. The size of the circle is proportional to the injection volume. The selected injections for b are marked with a black outline. (B) For

co-visualization, voxel densities from the 21 selected injections from (A) are overlaid as dotograms at 8 coronal levels for ipsilateral hemisphere. For the dotogram, one circle, whose size is proportional to the projection strength, is drawn for each injection in each voxel; the circles are sorted so that the largest is at the back and the smallest at the front, and are partially offset as a spiral. (C) Enlarged view of the dotogram from the area outlined by a white box in b. (D) 3D tractography paths in both cortical hemispheres. (E) A medial view of 3D tractography paths into the ipsilateral caudoputamen. Voxel starting points are represented as filled circles and injection site end points as open circles. (F) A top-down view of 3D tractography paths into the ipsilateral thalamus.

Few generalizable models of circuit dysfunction in ASD exist, but one study examined the maturation of cortico-striatal circuits in a mouse model of autism with deletions in *Shank3B* and observed imbalances in activity and connectivity (Peixoto et al., 2016). However, it has been shown that disrupted circuitry underlies other neurological diseases. Huntington's and Parkinson's diseases manifest due to dysregulation of a basal ganglia circuit. Using optogenetic activation methods, several groups successfully emulated symptoms and induced rescue by activating key components of the circuit (Tye and Deisseroth, 2012).

2.6 Intellectual and Behavioral Phenotypes

The phenotypes associated with ASD are remarkably heterogeneous. Phenotypes include both intellectual and behavioral impairments and occupy a wide range of associated severities. While ASD phenotypes have been studied in relation to underlying genetics, it has not been studied in the context of circuits or cell types. Primarily, syndromic subtypes of ASD have been described based on common genetic contributors, among either CNVs and SNVs, and certain molecular properties of affected genes are associated to phenotypic outcome.

One study examined a large cohort of ASD individuals with 16p11.2 duplications, which is liable for 1% of ASD cases (D'Angelo et al., 2016b). The authors were interested in characterizing the phenotypic effects when specifically incurring this mutation. In comparison to other ASD cases, they observed an association with significantly lower IQ scores, smaller head

circumferences, and lower BMI measurements. In comparison with 16p11.2 deletions, IQ scores were significantly less varied and individuals fell within a narrower phenotypic range.

Syndromic ASD subtypes were identified for other *de novo* CNVs, such as 7q11, 15q11.2–13.3, and 16p11.2 (Huguet et al., 2013). Similarly, different clinical outcomes were observed for deletions and duplications that occur at the same loci. For instance, 7q11 deletions are associated with Williams syndrome, which typically involves heart defects and unusual facial features, and are overly sociable. 7q11 duplications, on the other hand, are associated with ADHD, intellectual disability, and epilepsy (Sanders et al., 2011b). Likewise, clinical outcomes of 15q11.2-13.3 and 16p11.2 CNV events differ based on whether the mutational event is a deletion or a duplication.

Sequencing studies have also discovered syndromic forms of ASD that are associated with SNVs in certain genes, such as *CHD8*, *DYRK1A*, and *MECP2*. A recent study resequenced the *CHD8* gene in 3,700 children and found enrichment for certain clinical characteristics including macrocephaly and gastrointestinal problems (Bernier et al., 2014b). Another study resequenced the *DYRK1A* gene in 7,100 patients and observed an higher prevalence of microcephaly, stereotypy, febrile seizures, and impaired speech (van Bon et al., 2016b). Another study observed that *MECP2* transgenic monkeys, in comparison to wild-type monkeys, showed dramatic repetitive circular locomotion and progeny of the transgenic monkeys showed significant social impairment (Liu et al., 2016).

Chapter 3: Convergence on Functional Gene Networks

3.1 Functional Gene Networks Affected By *De Novo* Mutations

To explore the underlying biological pathways, we previously developed a computational approach (NETBAG+) that searches for cohesive biological networks using a diverse collection of disease-associated genetic variants (Gilman et al., 2012b; Gilman et al., 2011a). Using network-based approaches, we and others recently found that genetic variations associated with ASDs and other psychiatric disorders converge on several biological networks that are involved in neurogenesis and synaptic function (Gilman et al., 2012b; Gilman et al., 2011a; Kelleher and Bear, 2008; Zoghbi and Bear, 2012). In parallel with the identification of disease-associated genetic variations, complementary data sets of brain-related functional and phenotypic resources are rapidly being accumulated. These include a comprehensive database of gene expression across different cell types, distinct anatomical brain regions and developmental stages (Doyle et al., 2008; Kang et al., 2011). In addition, resources such as the Simons Simplex Collection (SSC) (Fischbach and Lord, 2010) have assembled a large compendium of ASD-related phenotypic data, including intelligence and social phenotypic scores. We focused our analyses on a set of genes that were implicated by our network-based computational approach and on all *de novo* truncating mutations from several recent studies. These two independent approaches provided us with complementary genes sets enriched in causal ASD mutations. We investigated the temporal, spatial and cell-specific expression profiles of implicated genes. We also explored how expression and functional properties of autism-associated genes affect ASD phenotypes.

To elucidate functional networks perturbed in ASD, we applied NETBAG+ to a set of genes affected by *de novo* CNVs and SNVs observed in autistic patients from the SSC (Iossifov et al., 2012; Levy et al., 2011b; O'Roak et al., 2012b; Sanders et al., 2012). All of the mutations

used as the input for our analyses were obtained using genome-wide methodologies and therefore are not biased by any preexisting hypotheses of ASD etiology. The combined input data contained a total of 991 unique genes from 624 independent genomic loci, including 580 unique genes with *de novo* SNVs and 434 genes within *de novo* CNVs; we note that the number of genomic loci used was considerably larger than the 47 loci considered in our previous analysis of *de novo* CNV events in autism¹³. We used NETBAG+ to identify a subset of the input genes that are strongly connected in the underlying phenotypic network. The NETBAG+ search revealed a functional network containing 159 genes ($P = 0.036$; Figure 3.1.1), of which 131 genes were affected by *de novo* SNVs and 31 by *de novo* CNVs. The network's significance was estimated using random input sets that matched the real data in terms of protein length and network connectivity. Notably, no significant networks were detected using genes associated with the 368 non-synonymous *de novo* mutations identified in siblings.

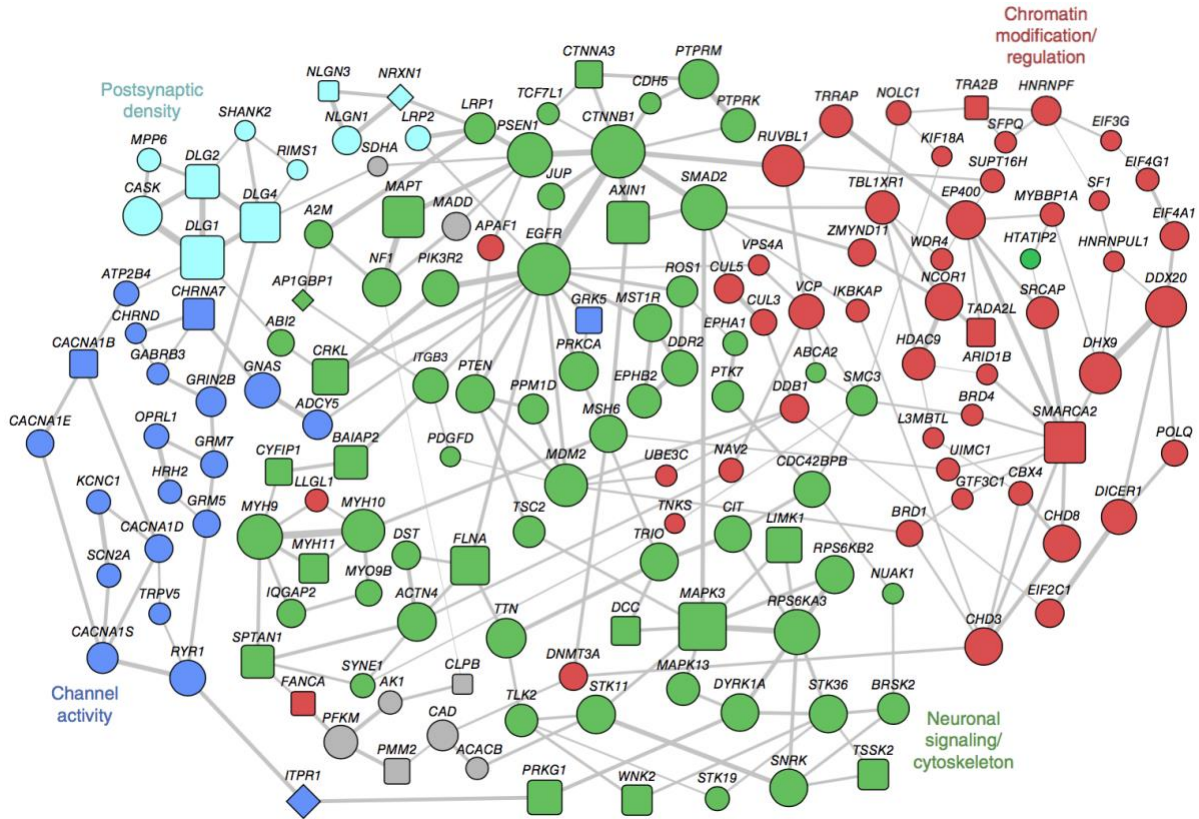


Figure 3.1.1: Mutations Implicate Functional Gene Network (Chang et al., 2015). The network implicated by NETBAG+ based on ASD-associated *de novo* SNVs and CNVs from recent studies (network is comprised of 159 genes, $P = 0.036$). Node sizes are proportional to the contributions of each gene to the overall network score and edge widths are proportional to the likelihood that the corresponding gene pair contributes to the same genetic phenotype. For clarity, only the two strongest edges for each gene are shown. Node shapes indicate types of the corresponding mutations: circles represent genes from SNVs, squares represent genes from CNVs, and diamonds represent genes affected by both mutation types. The network was divided into four cohesive functional clusters (indicated by node colors) using hierarchical clustering; general biological functions of these clusters, determined using DAVID, are shown. Gray nodes represent genes that are not members of the network clusters.

To explore the biological functions associated with the implicated network, we employed enrichment analysis, which identified a diverse set of functions associated with the implicated network, including synaptic function, chromatin modification and calcium channel activity. To better understand the functional relationships between genes in the network, we performed

hierarchical clustering of the network genes using the strength of interactions in the phenotypic network as a metric.

One cluster in our network contained genes responsible for synapse formation and function. This cluster contained neurexin (*NRXN*) and neuroligin (*NRLG*), as well as important components (*SHANK2* and *DLG2/DLG4*) of the postsynaptic density (PSD) of excitatory synapses, also shown in Figure 3.1.2. A related cluster included a diverse set of ion channels and receptors. This cluster contained genes (*CACNA1B*, *CACNA1D*, *CACNA1E*, *CACNA1S*) for subunits of several voltage-dependent calcium channels, which are important in learning and memory (Krey and Dolmetsch, 2007). The largest cluster of the implicated network contained genes that are associated with neuronal signaling and migration (*NF1*, *DCC*, *EPHA1/B2*), intracellular signaling (*MAPK3*, *EGFR*, *PTEN*, *MDM2*, *CTNNB1*, *PRKCA*, *LIMK1*), and the development of neuronal projections and actin cytoskeleton (*CYFIP1*, *TRIO*, *SPTAN1*, *FLNA*, *ACTN4*). ASD-associated mutations often perturb multiple signaling pathways that converge on the regulation of actin cytoskeleton and other structural processes that are required for neuron migration, cell-cell adhesion and the development of neuronal projections (Pinto et al., 2010a). Many genes in the clusters described above encode proteins related to the growth and function of dendritic spines. The final cluster was primarily related to functions associated with chromatin modifications, chromatin remodeling and transcriptional regulation. Notably, there is growing evidence that chromatin regulatory mechanisms crucially affect various stages of neural development, neuroplasticity and learning (Ronan et al., 2013). This cluster also contained genes involved in other processes that have been implicated in ASDs: RNA interference (*DICER1*, *AGO1/EIF2C1*), translation (*EIF4A1*, *EIF4G1*, *EIF3G*) and splicing (*SFPQ*, *TRA2B*) (Santini et

al., 2013). Many studies have reached similar conclusions through independent functional analyses (Bourgeron, 2015).

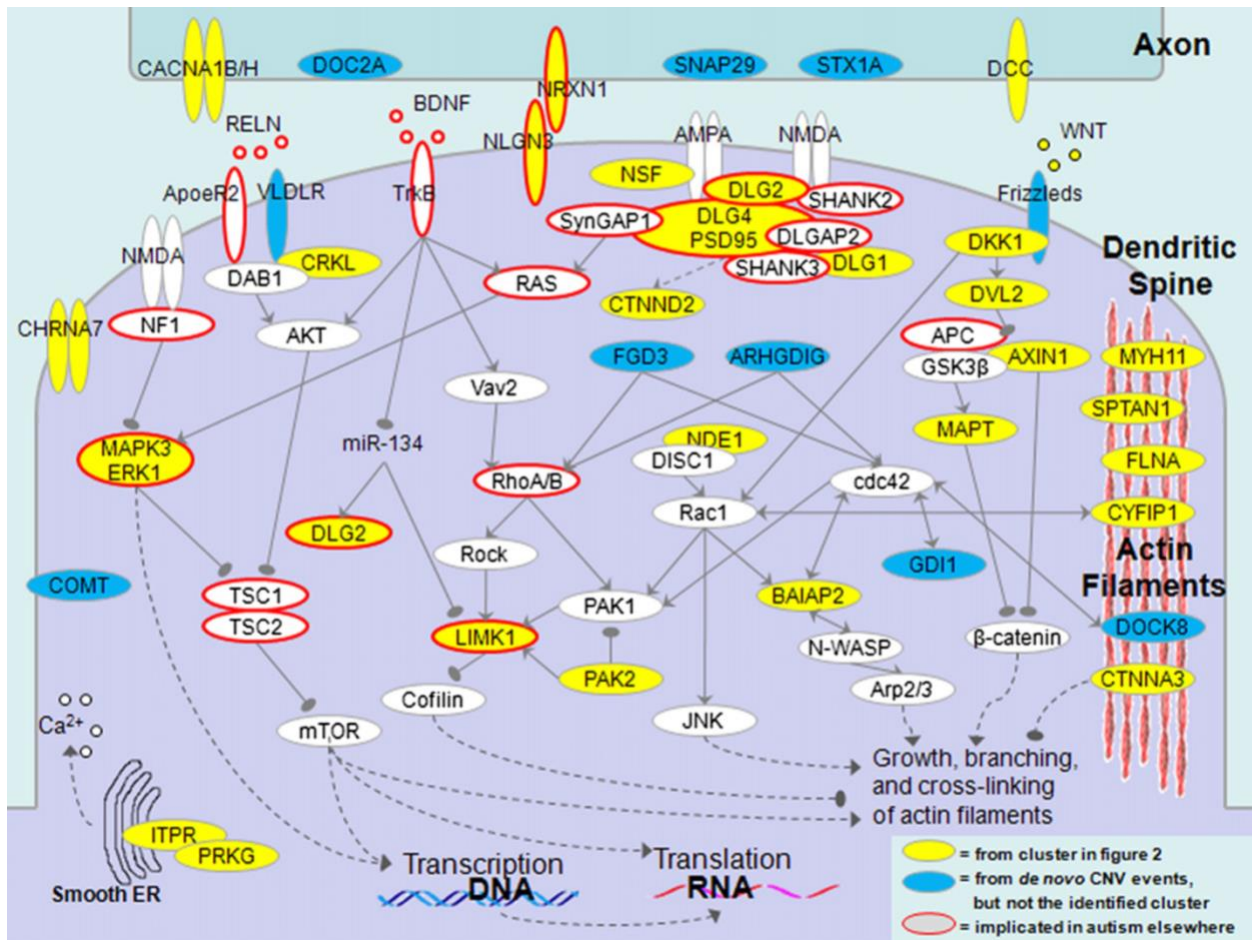


Figure 3.1.2: Genes Located at the Synapse or Associated with the Morphogenesis of Dendritic Spines (Gilman et al., 2012b). ER, endoplasmic reticulum; IP3, inositol-1,4,5-trisphosphate; PIP3, IP3 phosphatidylinositol-1,4,5-trisphosphate.

3.2 Association of the Implicated Genes with Specific Functional Subsets

In addition to the molecular and biological GO categories discussed above, we investigated whether there was an overlap between the network and specific gene subsets that may highlight the phenotypic and functional properties of the network genes. Recent exome-sequencing studies have revealed recurrent truncating mutations in several genes (Iossifov et al., 2012; O’Roak et al., 2012a; O’Roak et al., 2012b; Sanders et al., 2012; Ziff, 1997). Although

only 131 of 580 (~23%) of all genes harboring missense SNVs form the implicated network, 6 of 11 genes with recurrent truncating mutations are in the network (Fisher's exact one-tail test, $P = 0.02$). This suggests that there is a substantial enrichment of causal genes in the implicated network. Because ASD *de novo* mutations are predominantly heterozygous, it is likely that true causal mutations would preferentially target haploinsufficient genes. Indeed, using haploinsufficiency probabilities from a recent study (Huang et al., 2010), we found that genes in the implicated network were significantly more likely to be haploinsufficient than genes that were not selected by NETBAG+ (median probability for network and non-network genes are 0.57 and 0.32, respectively; Wilcoxon rank-sum two-tail test, $P < 10^{-14}$); a similar result was also observed for genes with recurrent truncating mutations (median probability for genes with recurrent truncating mutations and unaffected sibling SNV genes are 0.69 and 0.39, respectively; Wilcoxon rank-sum two tail, $P = 0.016$). We also considered the Genomic Evolutionary Rate Profiling (GERP) scores to characterize the severity of SNV mutations. Notably, the average GERP score of network genes was significantly higher than the average GERP score of genes not selected into the network (average network and non-network GERP scores are 4.0 and 3.3, respectively; Wilcoxon rank-sum one tail, $P = 0.001$). As higher GERP scores correspond to more damaging mutations, this analysis demonstrates that NETBAG predominantly selects genes harboring SNVs with potentially stronger functional effect.

Notably, a recent exome study showed a significant overlap between genes harboring truncating ASD *de novo* SNVs and targets of the fragile X mental retardation protein (FMRP) (Darnell et al., 2011; Iossifov et al., 2012). FMRP is an RNA-binding protein that is essential for a wide array of cognitive functions including synaptic plasticity and learning (Edbauer et al., 2010; Huber et al., 2002). Failure to properly express FMRP can cause fragile X syndrome, a

genetic disorder resulting in a spectrum of cognitive disabilities often accompanied by autistic symptoms. Following a previous analysis (Iossifov et al., 2012) of truncating *de novo* SNVs, we calculated the expected number of FMRP targets for ASD genes by inferring gene mutabilities from a large exome-sequencing study (Fu et al., 2013); the expected number of FMRP targets was then compared with the observed number in various ASD gene sets. This analysis revealed that, similar to truncating SNV genes, genes in the implicated network were significantly enriched for FMRP targets (truncating SNV gene enrichment 1:2.13, two-tail binomial test, $P = 3 \times 10^{-4}$; network gene enrichment 1:2.78, $P = 3 \times 10^{-11}$). The enrichment remained significant when considering only network genes harboring non-truncating SNVs (enrichment = 1:2.67, $P = 4 \times 10^{-9}$). Significant enrichment was also observed for FMRP targets from another recent study 31. In contrast, there was no significant enrichment for proband SNV genes that were not selected to the network ($P = 0.2$) nor for *de novo* SNV genes from unaffected siblings ($P = 0.15$). Furthermore, enrichment for FMRP targets remained significant when we separately considered each of the four functional clusters of the implicated network. Overall, these analyses confirm that FMRP targets are important for autism etiology and suggest a causal role for a substantial fraction of genes with non-truncating *de-novo* SNVs.

Genes forming the PSD are also likely to be important in ASDs and other psychiatric disorders (Bayes et al., 2011; Gilman et al., 2012a). The PSD is localized at the postsynaptic membrane of excitatory synapses and is crucial to synaptic communication and plasticity (Kennedy, 1997, 2000; Ziff, 1997). There was a significant enrichment of PSD-associated genes in the implicated network (enrichment 1:3.4, two-tail binomial test, $P = 7 \times 10^{-9}$) and marginal enrichment for truncating *de novo* SNVs in probands (enrichment 1:1.9, $P = 0.05$). No enrichment was observed for genes harboring SNVs in unaffected siblings (enrichment 1:1.05, P

= 0.8). For genes not selected to the network, PSD genes were significantly under-represented (enrichment 1:0.58, $P = 0.04$).

3.3 Temporal and Spatial Brain Expression Patterns of Implicated Genes

Gene expression patterns provide important clues for understanding physiological and developmental contexts of gene function. Thus, we next investigated brain expression of the ASD-implicated genes using the Human Brain Transcriptome database (Kang et al., 2011). Based on postmortem transcriptional analysis of tissue samples from healthy individuals, the HBT database provides a comprehensive map of mRNA expression across multiple brain regions and developmental stages. We determined the average expression levels across developmental periods for the various gene sets that formed the implicated network, as well as for genes with truncating *de novo* SNVs. We used a Wilcoxon rank-sum test (P_{WT}) to estimate the significance of the expression difference between sets of biological samples and a permutation-based test (P_{PT}) to estimate the probability of observing a greater bias in random probe sets of equal size (Online Methods). The brain expression of network genes was significantly higher than the expression of SNV-containing genes that were not selected to the network ($P_{WT} < 10^{-15}$; $P_{PT} < 10^{-4}$). This result confirms that the implicated network is significantly enriched in brain-related genes. Notably, high brain expression of network genes is not simply a consequence of their length or connectivity in the phenotypic network; randomly selected human genes with equivalent length or network connectivity are expressed in the brain at significantly lower levels ($P < 10^{-4}$) across all developmental periods.

Expression of the network genes was, on average, highest during the early fetal to early mid-fetal periods (8-19 post-conception weeks), possibly as a result of higher activity of ASD-related genes in brain cells or changes in brain cellular composition. To quantify the prenatal

expression bias, we calculated the difference between the expression during prenatal and postnatal developmental periods for every gene harboring *de novo* mutations. The prenatal bias was highly significant for network genes (bias 0.16, one-tail test, $P_{WT} < 10^{-15}/P_{PT} < 10^{-4}$) and for genes with truncating SNVs (bias 0.18, $P_{WT} < 10^{-15}/P_{PT} < 10^{-4}$). However, for ASD network genes with truncating *de novo* SNVs or *de novo* CNVs, embryonic expression (less than 8 post-conception weeks) was substantially lower as compared with that in later developmental periods (two-tail test, $P_{WT} = 6 \times 10^{-7}/P_{PT} < 10^{-4}$ and $P_{WT} = 10^{-14}/P_{PT} < 10^{-4}$, respectively). This effect is apparent when comparing the average expression profiles for network genes with truncating (CNVs and truncating SNVs) and non-truncating mutations. The bias against mutations in the embryonic period was absent for non-synonymous *de novo* SNV genes not selected to the network ($P_{WT} = 0.4/P_{PT} = 0.3$) and for genes with *de novo* SNVs in siblings ($P_{WT} = 0.2/P_{PT} = 0.2$). The observed bias against *de novo* truncating SNVs and CNVs suggests that there is strong selection against harmful mutations during embryonic development, as these mutations lead to more severe developmental consequences than mutations typically associated with autism.

We also examined the temporal expression profiles for each of the four clusters forming the implicated network. The cluster that was primarily associated with the postsynaptic density had the highest overall expression in the brain. Both the PSD cluster and the cluster associated with various channel activities showed a distinct rise during early fetal development, consistent with the start of synaptogenesis at the early mid-fetal developmental stage. In contrast, the cluster associated with chromatin modification and regulation showed high embryonic and fetal expression, consistent with a developmental peak of neuronal proliferation and differentiation, and then gradually decreased to a postnatal plateau. Sustained expression of chromatin modification genes after neurodevelopmental stages may be attributed to their involvement in

synaptic plasticity (Vogel-Ciernia et al., 2013). Finally, the cluster associated with neuronal signaling and cytoskeleton was relatively constant across all developmental stages. This is consistent with the important roles of signaling and structural genes across all developmental stages. Notably, there was substantial variability in the expression of the individual genes around the average cluster expression profiles. However, this variability is likely to be at least partially explained by different dosage requirements of individual genes. Normalized expression trajectories revealed that the temporal profiles of individual genes were generally consistent with the average expression profiles of corresponding functional clusters.

A high male-to-female incidence ratio, estimated at more than 4:1 for high-functioning individuals, is one of the most consistent and notable findings in ASDs (Fombonne, 2009; Newschaffer et al., 2007). Evidence suggests that this bias may be a result of a female protective effect that requires a higher threshold of genetic insults to trigger ASDs in females than in males (Zhao et al., 2007). We also observed that females in the SSC collection had a greater burden of truncating mutations than males (Fisher's exact one-tail test, $P = 0.07$), which is consistent with this hypothesis. This gender dimorphism was even stronger for the genes in the implicated network (30% of SNV mutations in females are truncating compared with 13% in males; Fisher's exact one-tail test, $P = 0.03$). Notably, the average brain expression was significantly higher for genes harboring truncating *de novo* mutations in females than in males (average \log_2 expression for genes: females, 7.98; males, 7.40; one-tail test, $P_{WT} < 10^{-15}/P_{PT} = 2 \times 10^{-3}$). These log-scale differences translate to absolute expression level differences of 50-100% across developmental stages in the average expression of genes harboring truncating SNVs in females and males. Thus, the expression data also suggest that relatively stronger genetic perturbations, that is, truncating mutations in genes with higher brain expression, are preferentially associated

with ASDs in females. Notably, the observed expression patterns cannot be explained by normal expression differences between males and females. For genes with truncating SNV mutations in female probands the average brain expression levels in females and males were 8.02 and 7.95 (<5% difference), respectively; for genes with truncating SNVs in male probands the average brain expression levels in females and males were 7.42 and 7.39 (<3% difference), respectively. We also note that the relative difference in brain expression for genes harboring mutations in females versus males was larger for genes with truncating SNVs than for implicated network genes. This result is likely a consequence of the severe effect of truncating mutations on protein function, which makes the expression of corresponding genes an important factor in explaining the variability of ASD phenotypes across patients.

To investigate whether ASD mutations preferentially affect specific brain regions, we analyzed spatial expression data available in the HBT database. For each brain region, we compared the average expression levels of network genes to the expression levels of genes with SNVs in unaffected siblings. Compared with controls, network genes showed significantly higher expression across all brain regions, which is consistent with the wide spectrum of phenotypic abnormalities observed in ASDs. Furthermore, we investigated the prenatal expression bias for network and truncating genes and compared these with the prenatal bias for genes with SNVs in siblings. We found that the prenatal bias was generally similar and statistically significant for all brain regions. We note, however, that the amygdala, which is known to have a crucial role in processing emotional stimuli (Adolphs, 2010), as well as cerebellum and striatum, showed a numerically, although not significantly, larger prenatal bias than other regions. Overall, however, the lack of regional specificity underscores the sharing of gene functions across brain areas.

The human brain contains a variety of neuronal and non-neuronal cell types, and we sought to investigate possible biases of implicated genes toward specific cell types. To explore this question, we used an independent data set of cell-specific gene expression generated using translating ribosome affinity purification (TRAP) (Doyle et al., 2008); the data set contains gene expression profiles for 25 distinct cell types from the mouse CNS. To assess cell-specific expression biases, we compared, for each cell type, the expression of mouse orthologs for the implicated network genes with the expression of orthologs for genes harboring mutations in unaffected siblings. This analysis revealed that multiple neuronal and non-neuronal cell types were likely affected by *de novo* ASD mutations in the network genes (Figure 3.3). The diversity of affected cells could be explained, at least partially, by shared usage of common signaling, structural and neural pathways across diverse cell types. Although multiple cell types were affected, particularly strong and significant expression biases for implicated network genes were observed in cortical neurons, especially cortical interneurons, pyramidal neurons and medium spiny neurons of the striatum (Figure 3.3). Notably, similar cell types were also independently implicated by considering expression biases for 11 genes with recurrent truncating *de novo* mutations (Figure 3.3). On the other hand, some other cell types, such as motor neurons and astroglial cells, were markedly less affected. Notably, deep layer cortical glutamatergic projection neurons were also identified recently as a point of spatiotemporal convergence in coexpression networks built around high confidence ASD genes (Willsey et al., 2013a).

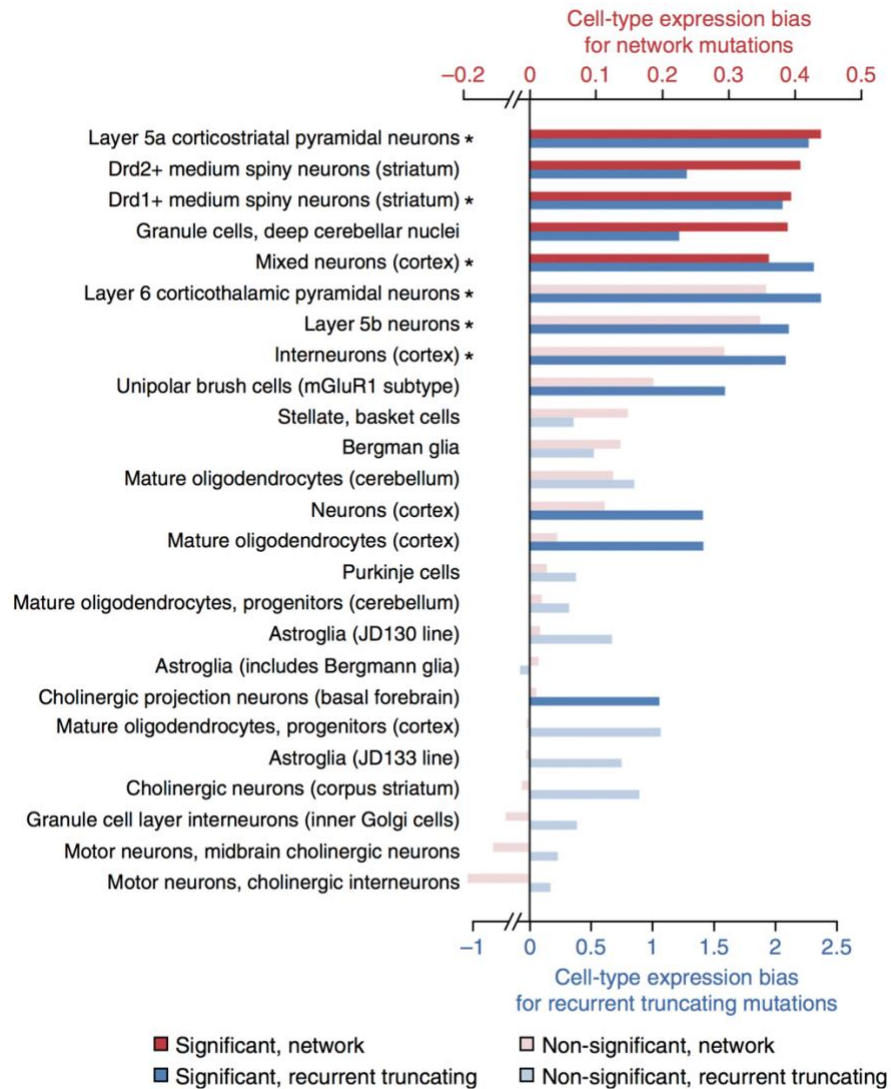


Figure 3.3: Cell-type Expression Biases for ASD Mutations (Chang et al., 2015). Proband versus unaffected sibling expression biases were computed across 25 cell types of the CNS for implicated network genes (shown in red) and for 11 genes with recurrent truncating SNVs (blue). The biases were calculated using *Mus musculus* expression data. To quantify the expression biases, we calculated for each cell type the difference between the average \log_2 expression of mouse orthologs for implicated human genes and the average \log_2 expression of mouse ortholog for human genes with *de novo* SNVs in unaffected siblings. Cell types are ordered by the magnitude of the cell type expression bias for network genes (red). The significance of the expression biases was evaluated using the Wilcoxon rank-sum one-tail test and corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure with a false discovery rate of 10%; significant cell types are shown in dark red and blue, and non-significant types are shown in light red and blue. In addition, the *P* values obtained from the two independent approaches, one based on network genes (red) and the other based on genes affected by recurrent truncating mutations (blue), were combined using

Fisher's and Stouffer's meta-analysis methods. The combined P values were corrected using the Bonferroni method, and the cell types passing the significance cutoff for both meta-analyses are indicated with an asterisk (*).

3.4 Functional Properties of Implicated Genes and Disease Phenotypes

Because ASDs manifest substantial pathophysiological differences across probands, it is important to understand how functional properties of implicated genes and associated mutations affect phenotypic characteristics of the disease. To investigate the effect of truncating *de novo* mutations on IQ, we calculated, separately for CNVs and truncating SNVs, the average number of truncating mutations per individual across the IQ spectrum. Notably, for high-functioning ASD probands, the average number of truncating mutations (CNVs and truncating SNVs) decreased and became similar to the average number in unaffected siblings. The average number of truncating SNVs for probands with IQ less than 100 was about twofold higher than probands with IQ greater than or equal to 100 (0.17 for $IQ < 100$, 0.09 for $IQ \geq 100$; Fisher's exact two-tail test, $P = 0.02$). In contrast, similar analyses for non-truncating mutations (synonymous and non-truncating SNVs) in probands showed that the average number of non-truncating mutations per individual was relatively constant across IQs, without a prominent decrease for high-functioning ASD probands (0.43 for $IQ < 100$, 0.46 for $IQ \geq 100$; Fisher's exact two-tail test, $P = 0.7$). Relative to synonymous mutations, non-synonymous non-truncating were significantly enriched in probands with IQs above 70 compared with those below 70: 160 non-truncating and 72 synonymous mutations (2.2:1 ratio) were observed in probands with $IQ \leq 70$, whereas 330 non-truncating and 110 synonymous mutations (3.3:1 ratio) in probands with $IQ > 70$ (Fisher's exact one-tail test, $P = 0.04$). Overall, these analyses suggest that truncating mutations are likely to play a less prominent role in high-functioning ASD cases. Another study (Samocha et al., 2014)

also demonstrated no excess of *de novo* loss-of-function mutations for high-functioning ASD probands.

Notably, there was a significant difference in IQs between probands with network genes affected by CNV deletions and duplications. The average IQs for probands with CNV deletions and duplications were 64.8 and 83.9, respectively (Wilcoxon rank-sum one-tail test, $P = 6 \times 10^{-3}$). This result suggests that a decrease in gene dosage has a substantially stronger functional effect than an increase in dosage.

Given that genes exhibited diverse expression patterns, we asked whether the level of brain expression for affected genes is associated, on average, with different phenotypic outcomes. For this analysis, we considered the full-scale IQ and Autism Diagnostic Interview-Revised (ADIR) social interaction and repetitive behavior scores. The ADIR scores are based on structured interviews with proband parents and reflect patterns in reciprocal social interactions (ADIR-S) and repetitive/restrictive behaviors (ADIR-R) (Lord et al., 1994). To explore the relationship between the average expression level of affected genes and corresponding phenotypes, we divided the ASD cases into low- and high-scoring phenotypic subsets relative to the corresponding median phenotype scores. We then calculated the average expression levels of implicated genes identified in each phenotype subset. This analysis revealed that affected genes associated with lower IQ scores or higher ADIR scores (both indicating more severe phenotypes) usually had significantly higher brain expression (network genes: one-tail test, $P_{WT} < 10^{-15}/P_{PT} = 0.01$ for IQ, $P_{WT} < 10^{-15}/P_{PT} = 0.3$ for ADIR-S, $P_{WT} < 10^{-15}/P_{PT} = 0.024$ for ADIR-R; genes with truncating mutations: one-tail test, $P_{WT} < 10^{-15}/P_{PT} < 10^{-4}$ for IQ, $P_{WT} < 10^{-15}/P_{PT} = 1.5 \times 10^{-3}$ for ADIR-S, $P_{WT} < 10^{-15}/P_{PT} = 0.08$ for ADIR-R).

3.5 Implications for Neural Circuits and Phenotype Stratification

Our results suggest that the pathophysiological heterogeneity of ASD is matched by the diversity of genetic and functional insults associated with the disorder. We found that affected genes had diverse developmental expression profiles and are therefore likely to be important in multiple stages of neurogenesis, neuron mobility, synaptogenesis and brain function. Although the implicated genes and processes are active across multiple cell types, some types, such as cortical interneurons, pyramidal neurons and medium spiny neuron of the striatum, seemed to be more strongly affected. Notably, layer 5 cortical pyramidal neurons often project to the striatum and layer 6 often project to the thalamus; the corresponding cortico-striatal-thalamic circuits are known to mediate diverse motor, emotional, cognitive and habit-forming behaviors that are often perturbed in ASD (Burguiere et al., 2013; Langen et al., 2011; Rothwell et al., 2014; Shepherd, 2013). Consequently, our unbiased genome-wide analysis implicates specific functional neural circuits that may mediate stereotypical and repetitive behaviors in ASD.

Although previous studies have primarily emphasized the role of truncating *de novo* mutations in ASD (Iossifov et al., 2012; O'Roak et al., 2012b; Sanders et al., 2012), our analysis suggests that a substantial fraction of non-truncating *de novo* missense mutations observed in probands also contributes to the disorder. Notably, we found that functional mutations were preferentially observed in haploinsufficient genes, which confirms that dosage effects are important for the disease mechanisms. We found that functional characteristics of affected genes, such as brain expression levels, are likely to influence the observed phenotypic consequences of *de novo* ASD mutations. Stronger functional insults lead, on average, to more severe ASD phenotypes. Thus, the distinction between intellectual disability and autism may lie primarily in

the degree of overall functional effect rather than specific genes and pathways affected in the two disorders.

Our analysis of brain expression provides further evidence for the hypothesis that stronger functional insults, such as perturbation of genes with substantially higher brain expression, are associated with female autistic phenotypes. Stronger functional perturbations in females were previously demonstrated through analyses of autism CNVs sizes (Levy et al., 2011b; Sanders et al., 2011a) and gene network properties (Gilman et al., 2011b). Thus, multiple independent sources of evidence suggest a protective effect in females, although the mechanisms of this effect remain to be elucidated.

We also find, in agreement with recently published studies (Iossifov et al., 2014b; Samocha et al., 2014), that truncating mutations play a smaller role in high-functioning ASD cases. Because truncating mutations are usually associated with a loss of function, this result suggests that high-functioning autism phenotypes are less likely to be mediated by a loss of normal gene function in the brain. Functional gain and other types of genetic variations, such as non-truncating *de novo* mutations, common polymorphisms or mutations in non-coding regulatory regions, may be the primary contributors to high-functioning ASD cases.

Taken together, our results suggest that various functional properties of mutations and target genes may be useful in predicting ASD phenotypic severity. In the future, it will be important to investigate the extent to which individual ASD patients can be stratified on the basis of affected pathways, biological functions and cell types. Such patient stratification, now a common practice in cancer, may lead to individualized diagnostic and prognostic predictions and, ultimately, to targeted ASD therapies.

While the network-based methods we developed provided insights into the biological processes and molecular pathways underlying ASD, the processes and pathways themselves do not fully inform the etiology of the disease. The defining features and impairments of ASD are specific and distinct; however, the pathways identified in our network involve general processes that are commonplace throughout the brain. For this reason, we next investigated the neural structures and circuits involved in ASD that may further explain the behavioral phenotypes associated with the disorder.

Chapter 4: Mutations Target Neural Circuits

4.1 ASD-associated Brain Structures are Highly Connected

Large-scale sequencing studies have made considerable progress in uncovering a number of mutations with a strong causal link to ASD (Iossifov 2014, Iossifov 2015, Feliciano 2018, Satterstrom 2020). Network analysis based on these studies have shown significant enrichment for genes encoding synaptic and postsynaptic density proteins involved in synaptic transmission and cell–cell communication as well as an overrepresentation of histone-modifying enzymes and chromatin remodelers (Chang 2015, Parikshak 2015). However, impairment of these genes and associated pathways do not fully explain the behavioral and cognitive phenotypes specific to ASD patients. Specifically, how neural physiology is altered in autistic individuals and the origin of the disease remains to be addressed. Many genes and pathways associated with ASD overlap with those involved in other disorders such as schizophrenia or bipolar disorder, suggesting that differentiation occurs at a higher level of functional organization. One prevailing hypothesis is the connectivity theory of autism, which suggests that communication between brain regions is atypical. If phenotypes do indeed arise from dysfunctional neural circuits rather than at the level of gene networks, then investigating the connectome landscape would be a logical next step. In our study, we investigate this theory and explore whether neural systems and circuits are involved in the pathology of ASD.

Connectivity studies have made several notable observations with regards to differences in the neurobiology of ASD patients. Brain imaging studies based primarily on fMRI methods highlight systematic differences in the functional connectivity between regions, such as hypo-connectivity among distal connections to the frontal cortex and hyper-connectivity among local connections in the temporal, occipital, prefrontal, and parietal cortex (Holiga 2019). However,

these studies were not able to distinguish causal from compensatory changes. While evidence from both transcriptional and brain imaging studies suggest that a number of cortical and subcortical regions are involved in ASD (Voineagu 2011, Willsey 2013), there are, to our knowledge, no such methods that comprehensively discover and identify the neural circuits affected by ASD mutations. In this work, we undertake an unbiased approach that establishes causality through the inclusion of mutational events and couples brain-wide expression data with connectivity data. Our analysis (Figure 4.1.1) attempts to unify approaches undertaken by connectivity and transcriptional studies to establish a neural circuit underlying ASD etiology.

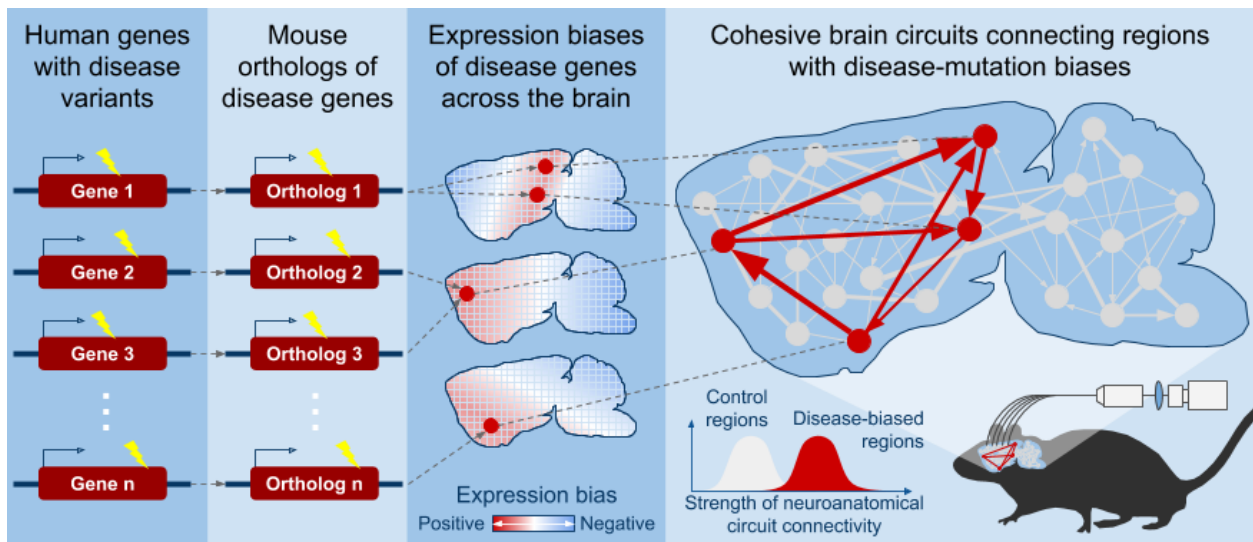


Figure 4.1.1: Flowchart of Expression and Circuit Analysis. Mutations are gathered from genome-wide sequencing studies and are used to predict high-confidence ASD genes, which are then mapped to mouse orthologs (1st and 2nd panel). Expression biases for these genes are calculated across the entire brain and provide an approximation of the mutational effect or impact on the brain (3rd panel). The neural structures with the strongest expression biases are then mapped to the brain connectome, and a search algorithm determines a cohesive neural circuit and further tests the strength of connections among the system of neural structures (4th panel).

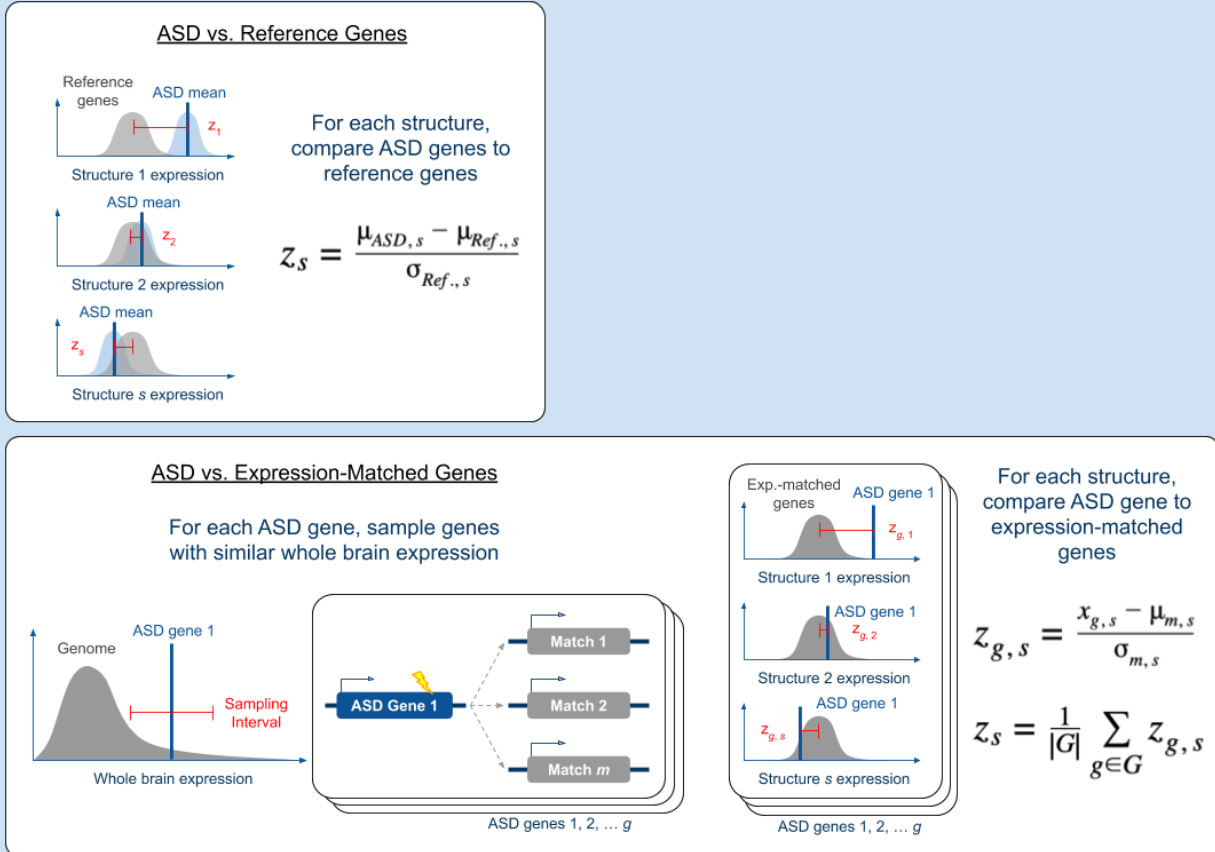
For our analysis, we used brain-wide projection data obtained through viral tracing of physical axonal tracts available from the Allen Mouse Brain Connectivity Atlas (Oh 2014). This data set uses the same reference model of neuroanatomical partitions as the Allen Mouse Brain

Atlas (Lein 2006), which provides brain-wide and genome-wide transcriptomic profiles obtained through high-resolution imaging of in situ hybridization (ISH) data and is significantly correlated with the human brain transcriptome. We investigate the effects of ASD mutations on the brain using these data sets, and we employ two independent collections of ASD mutations. Both mutational data sets derived *de novo* mutations from family exome sequencing, and computational and statistical methods were applied to determine sets of high-confidence disease genes. The first gene set is derived from mutations primarily collected by the Autism Sequencing Consortium and Simons Simplex Collection (ASC-SSC), and uses TADA, a method that combines data from *de novo* mutations, inherited variants, and standing variants in the population, to discover ASD genes (Satterstrom 2020, Iossifov 2014). The second gene set uses mutations collected by the Simons Foundation Powering Autism Research for Knowledge (SPARK) project, on which we applied an intolerance-based model to prioritize genes (Feliciano 2018, Iossifov 2015).

We first determined the neural structures, or mesoscale functional neuronal bundles, with the highest expression of ASD genes across the entire brain. For each structure, we computed the mean difference in expression between ASD genes and a reference set of genes derived from mutations observed in a neurotypical control population (Figure 4.1.2a). Interestingly, we observed that the strongest expression biases did not occur in highly connected structures in the connectome (ASC-SSC Mann-Whitney U test $P = 0.3$, SPARK $P = 0.8$). We then investigated the strength of connections between the neural structures showing the strongest expression biases. We tested structure connectivity by network permutation and by comparing the weights of connections that occur internally among biased structures to those that occur externally (Figure 4.1.2b). The network permutation test (P_{Perm}) quantifies the probability of observing the

specified structures with more or equal or more connections among randomized networks. Importantly, our permutation procedure maintains the total degree of each structure in the connectome. The second connectivity test compares, for each ASD structure, the strength of connections to other ASD structures to the strength of connections to structures external to the ASD set using a Wilcoxon signed-rank test (P_{WSR}). Because connectivity strengths are evaluated relatively for each individual structure, this test also accounts for differences in overall connectivity among structures in the connectome.

a Expression Biases



b Connectivity Tests

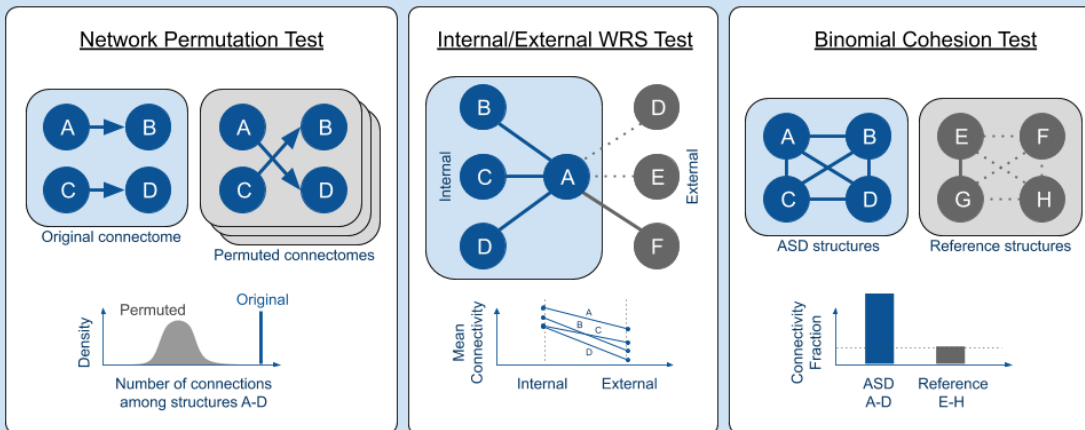


Figure 4.1.2: Expression and Connectivity Analysis Diagram. (A) Expression biases were calculated using either reference genes or ascertained genes. For reference genes, a z-score for each neural structure was calculated by standardizing the mean expression of ASD genes against the distribution of expressions for reference genes. For ascertained genes, a z-score for each neural structure was calculated by standardizing the expression of a single ASD gene against the distribution of expressions for matched genes with similar overall

brain expression. Z-scores were then averaged across all ASD genes to obtain an expression bias for each structure. For both expression biases, two measures of expression were used: expression level and expression specificity. (B) Connectivity significance was estimated using a network permutation test, a Wilcoxon rank-sum (WRS) test, and a binomial test. The network permutation test compares the observed number of connections among ASD neural structures in the non-permuted original connectome to the expected number of connections among ASD structures based on permuted connectomes. The WRS test compares, for each ASD neural structure, the internal connectivities with other ASD structures to the connectivities between the given ASD structures and structures external to the ASD set. The binomial test compares the probability of connectivity among a set of ASD structures to a base probability among a set of reference structures.

We observed that ASD genes are significantly expressed in a number of cortical and subcortical regions, and these structures with the largest expression biases were also strongly connected. For the ASC-SSC genes, the number of connections among the most biased structures is significantly greater than expected ($P_{\text{Perm}} < 10^{-4}$, 414 observed connections vs. 226.5 expected, Figure 4.1.3) and the connections are also significantly stronger than external connections ($P_{\text{WSR}} = 2.2 \times 10^{-9}$, 4.12 mean internal \log_{10} connectivity weight vs. 3.02 external). Similarly, for SPARK genes, we observed a significant number of connections and greater strength of internal connections among the most biased structures ($P_{\text{Perm}} < 10^{-4}$, 261 observed vs. 179.9 expected; $P_{\text{WSR}} = 0.0006$, 3.77 internal vs. 3.26 external). Moreover, we observe significant concordance among the most biased structures for ASC-SSC and SPARK (Fisher's exact test $P < 10^{-10}$). Notably, these strongly connected structures were distributed throughout the brain and include limbic cortical areas, thalamus, striatum, hippocampus, and amygdala. Interestingly, strongly connected structures were not uniformly distributed throughout the brain. We observe that a large fraction of the structures in the hippocampus formation, striatum, isocortex, and cortical subplate were strongly targeted by ASD mutations, whereas fewer structures in the thalamus and pallidum were targeted (Figure 4.1.4). Conversely, ASD mutations appear to avoid specific regions of the brain, namely the cerebellum, medulla, pons, midbrain, and hypothalamus. Among

the 213 structures defined in the Allen Mouse Brain Connectivity Atlas, we selected the top 50 structures with the strongest expression biases for our analysis, and we confirm that significant connectivity is maintained when the number of structures considered is varied.

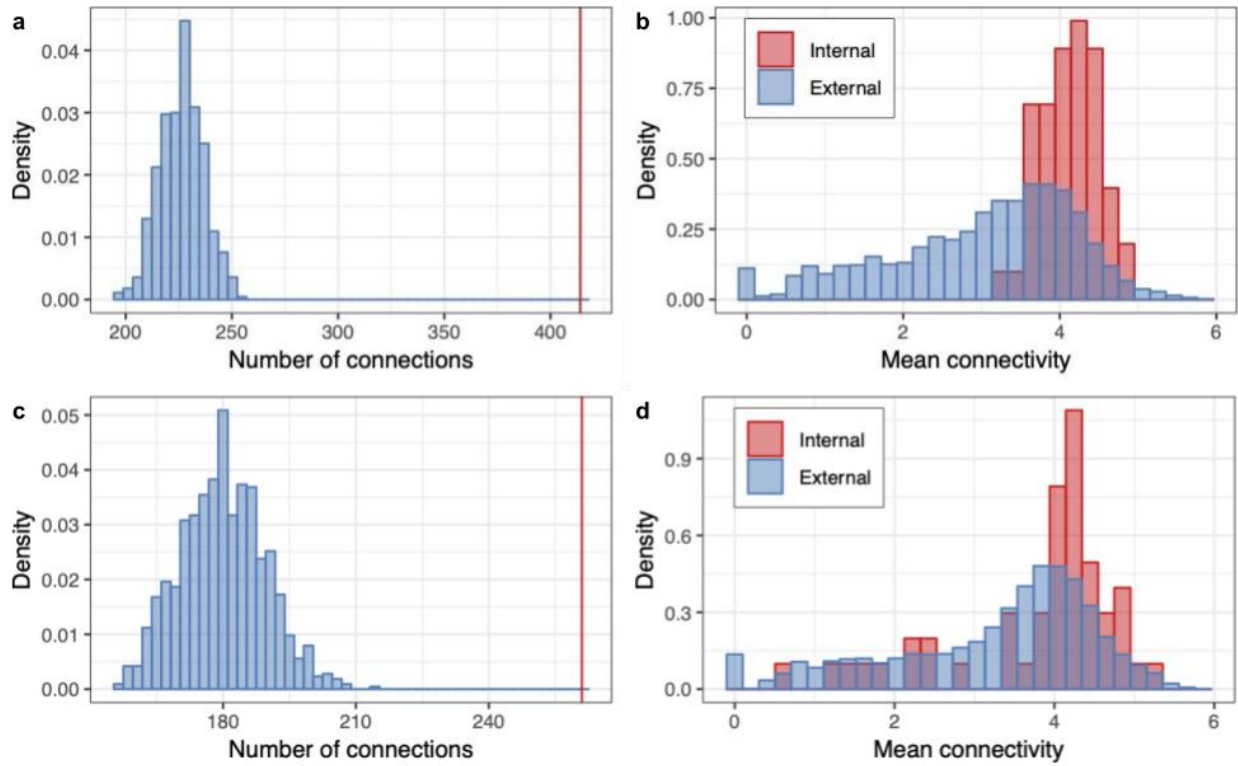


Figure 4.1.3: ASD Structures are Strongly Interconnected. (A) Distribution of the number of connections among selected ASC-SSC neural structures following network permutation. Red vertical line indicates the observed number of connections in the non-permuted connectome. Permutation test $P < 10^{-4}$. **(B)** Distributions of mean connectivity weights for selected ASC-SSC neural structures. Internal connectivity weights between ASD structures are shown in red and weights for connections with external structures are shown separately in blue. Wilcoxon signed-rank test $P < 10^{-10}$. **(C)** Distribution of the number of connections among selected SPARK neural structures following network permutation. Red vertical line indicates the observed number of connections in the non-permuted connectome. Permutation test $P < 10^{-4}$. **(D)** Distributions of mean connectivity weights for selected SPARK neural structures. Internal connectivity weights between ASD structures are shown in red and weights for connections with external structures are shown separately in blue. Wilcoxon signed-rank test $P < 10^{-10}$.

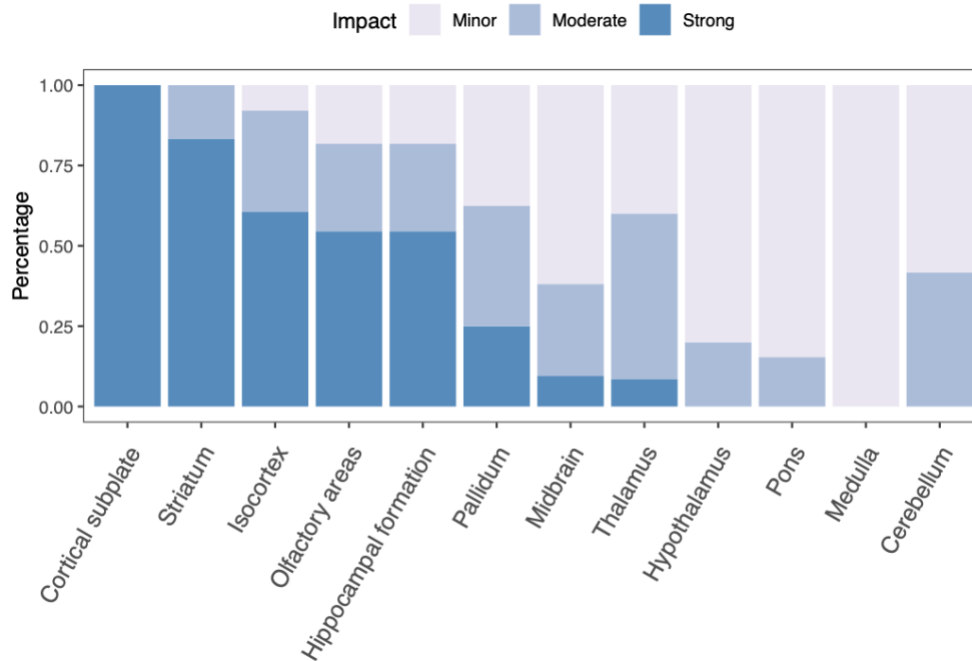
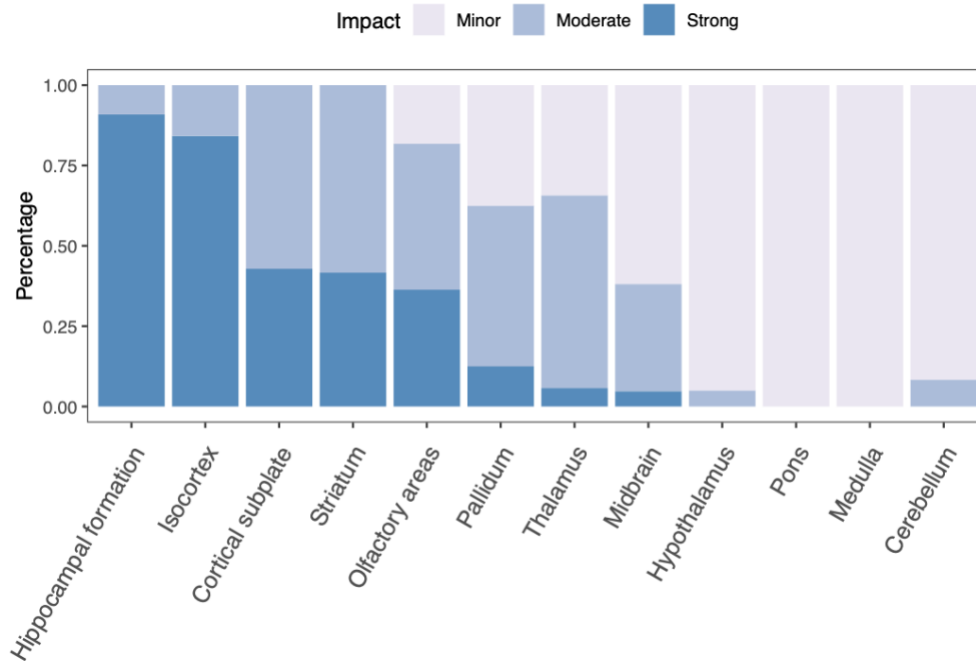


Figure 4.1.4: Proportion of Brain Regions Affected by ASD Mutations. (Top) Fraction of neural structures with strong, moderate, and weak expression biases for each high-level brain region. Impact groups are shown as different shades of blue. Expression biases are shown for expression specificity for ASC-SSC genes. (Bottom) Fraction of neural structures with strong, moderate, and weak expression biases for each high-level brain region. Impact groups are shown as different shades of blue. Expression biases are shown for expression specificity for SPARK genes.

4.2 Mutations Specifically Target Strongly Interconnected Structures

In addition to assessing structures by the level of ASD gene transcription, we also examined the specificity of expression of ASD genes among all neural structures in the brain, and investigated the strength of connectivity among the most biased structures. To that end, we calculated the specificity of gene expression for each gene, defined as the z -score (i.e. standard deviations from mean expression), in each structure relative to the rest of the brain. Importantly, by standardizing the overall expression level, structure expression biases cannot be attributed to genes that are highly and ubiquitously expressed across structures. We then computed, for each structure, the mean difference in specificity between ASD genes and the reference gene set, and as was previously done, we investigated the nature of connections among the most biased structures (Figure 4.1.2).

Consistent with our previous observations, ASD mutations appear to specifically target a number of key cortical and subcortical structures that are abundantly and tightly connected. For the ASC-SSC genes, the number of connections significantly exceeded expectation in a random network ($P_{\text{Perm}} < 10^{-4}$, 361 observed vs. 221.7 expected) and the structures were significantly more strongly connected to each other than to external structures ($P_{\text{WSR}} = 3.8 \times 10^{-8}$, 4.13 internal vs. 3.16 external). Moreover, we observe that the most biased structures by expression specificity were strongly concordant with the most biased structures by expression level (Fisher's exact test $P < 10^{-10}$). For SPARK genes, we similarly observed a significant number of connections and greater strength of internal connections among the most biased structures ($P_{\text{Perm}} < 10^{-4}$, 277 observed vs. 204.7 expected; $P_{\text{WSR}} = 2.5 \times 10^{-5}$, 3.97 internal vs. 3.32 external). Additionally, we used bootstrap methods to estimate the standard error of our connectivity statistics, and we

observed limited variability of our estimates (ASC-SSC number of connections SEM = 54.7, mean internal/external SEM = 0.10; SPARK = 44.7, 0.15).

Previous studies have established that ASD-associated genes are more active prenatally than postnatally (Willsey 2013, Uddin 2014, Chang 2015, Buxbaum 2020). Among the genes we considered in our analysis, 66% of ASC-SSC and 57% of SPARK genes, including many that are differentially expressed prenatally, are actively transcribed in the postnatal brain (BrainSpan expression intensity > 1). We investigated whether postnatal genes would lead to non-significant connectivities by limited our analysis to genes that are postnatally active, and confirmed that our connectivity results remain largely consistent ($P_{\text{Perm, ASC-SSC}} < 10^{-4}$, $P_{\text{WSR, ASC-SSC}} = 2.6 \times 10^{-9}$; $P_{\text{Perm, SPARK}} < 10^{-4}$, $P_{\text{WSR, SPARK}} = 1.2 \times 10^{-5}$, Figure 4.3). Moreover, a significant number of biased structures remain for both ASC-SSC and SPARK (Fisher's exact test $P < 10^{-10}$).

4.3 Mutations Target Cortical and Subcortical Neurons

Previous studies by us and others have shown that ASD mutations affect specific neuronal cell-types in the brain (Willsey 2013, Chang 2015). However, the ratio of neuronal to glial cells is known to vary considerably throughout the brain (Zhu 2018, Erö 2018). Therefore, we asked whether the expression biases in ASD reflect specific neuronal types endogenous to certain structures or a more general property corresponding to neural cell type composition. Specifically, we investigated whether the expression biases could be explained by differences in synaptic density throughout the brain (Zhu 2018). On the contrary, we observed that synaptic density and expression biases are only moderately correlated (Spearman $\rho = 0.47$). Moreover, we obtained similar structures when the expression of ASD genes was compared against the expression of genes annotated with synapse-related GO terms (Fisher's exact test $P < 10^{-10}$,

Figure 4.3). Overall, synaptic density may explain some of the difference in structures biases, but the contribution is limited to a minor fraction of the total variance.

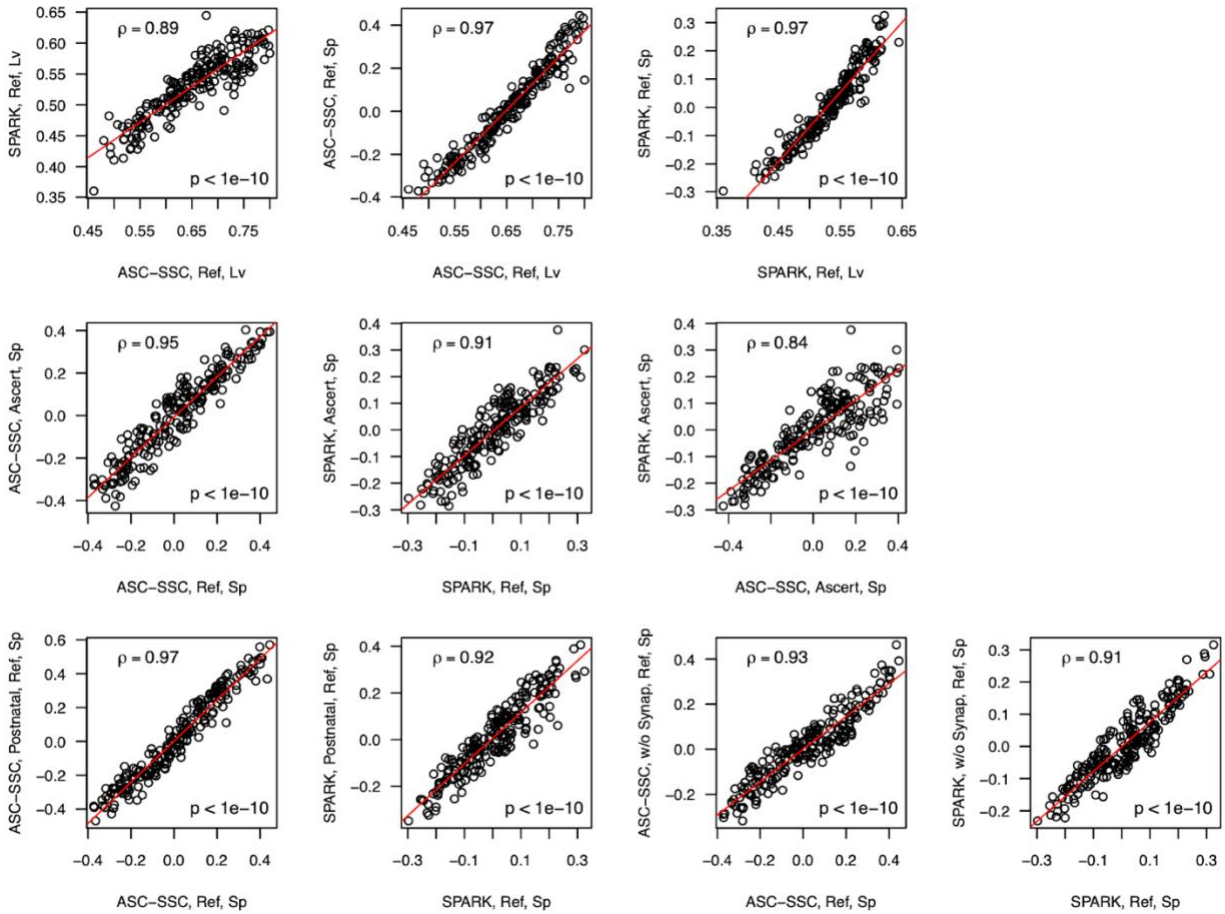


Figure 4.3: Expression Bias Correlations. (a) Correlation between neural structure expression level biases for ASC-SSC genes and biases for SPARK genes, computed against reference genes. Line indicates linear regression. Spearman correlation coefficients and p-values are provided in each panel. (b) Correlation between neural structure expression level biases and expression specificity biases for ASC-SSC genes, computed against reference genes. (c) Correlation between neural structure expression level biases and expression specificity biases for SPARK genes, computed against reference genes. (d) Correlation between neural structure expression specificity biases computed against reference genes and expression specificity biases computed against ascertained genes for ASC-SSC genes. (e) Correlation between neural structure expression specificity biases computed against reference genes and expression specificity biases computed against ascertained genes for SPARK genes. (f) Correlation between neural structure expression specificity biases for ASC-SSC genes and expression specificity biases for SPARK genes, computed against ascertained genes. (g) Correlation between neural structure expression specificity biases for ASC-SSC genes and expression specificity biases for postnatal ASC-SSC genes, computed against

reference genes. (h) Correlation between neural structure expression specificity biases for SPARK genes and expression specificity biases for postnatal SPARK genes, computed against reference genes. (i) Correlation between neural structure expression specificity biases for ASC-SSC genes and expression specificity biases for non-synaptic ASC-SSC genes, computed against reference genes. (j) Correlation between neural structure expression specificity biases for SPARK genes and expression specificity biases for non-synaptic SPARK genes, computed against reference genes.

We performed a similar analysis using single-cell expression data collected for 690,000 cells obtained from various regions of the brain in order to investigate the neuronal populations underlying the observed regional biases (Saunders 2018). These results confirm and extend our previous work using cell type-specific expression data which implicated striatal and cortical neuronal populations in ASD (Chang 2015). While the expression data is not available for all regions of the brain, we observed significant biases among neuronal populations in the cortex, striatum, amygdala, and hippocampus, whereas biases among other regions remained non-significant. ASC-SSC and SPARK genes specifically target pyramidal and interneuronal neurons from the cortex, medium spiny neurons from the striatum, as well as novel neurons from the amygdala, hippocampus, and thalamus. This further suggests that the identified neural structure biases cannot be attributed to differences in cell composition across brain areas. As in our previous study, we did not observe targeting of non-neuronal cell types among any of the available regions.

4.4 Functionally Ascertained Structures Maintain Interconnectivity

We use genes with mutations observed in an unaffected control population as a reference gene set in our calculations, specifically the sibling population from the SPARK cohort. It is known that such genes are expressed at low levels throughout the brain (Chang 2015), and consequently, these genes may lack functional relevance to neural structures and circuits. We therefore computed expression biases for ASD genes using sampled genes with similar levels of

transcriptional activity in the brain (Figure 4.1.2a). Indeed, we observed that for both ASC-SSC and SPARK genes, the most biased structures remain significantly interconnected in terms of the number of connections and the strength of connections ($P_{\text{Perm, ASC-SSC}} < 10^{-4}$, $P_{\text{WSR, ASC-SSC}} = 5.2 \times 10^{-8}$; $P_{\text{Perm, SPARK}} < 10^{-4}$, $P_{\text{WSR, SPARK}} = 0.02$). Moreover, we observe strong concordance among these structures and the structures calculated using the gene set derived from a control population (Fisher's exact test $P < 10^{-10}$).

We can likewise perform the analysis on the genes with non-synonymous mutations observed in the unaffected sibling population, which is the gene set previously used as our reference gene set. We expectedly observed that the expression biases, in comparison to those for ASD genes, are much weaker (Figure 4.4). Moreover, we utilize the structures with the largest expression biases as a reference set of structures. This permits us to perform an additional test of structure interconnectivity in which we compare the number of connections among ASD structures to a reference set of structures, and indeed, ASD structures are significantly more connected ($P_{\text{Binom, ASC-SSC, Level}} < 10^{-10}$, $P_{\text{Binom, ASC-SSC, Specific}} < 10^{-10}$, $P_{\text{Binom, SPARK, Level}} = 2.5 \times 10^{-6}$, $P_{\text{Binom, SPARK, Specific}} < 10^{-10}$). We also show the connectivity test remains significant as the threshold for connection binarization is varied. More importantly, a reference set of structures allows us to search for and identify a strongly interconnected subset of structures among the most biased, thereby implicating a neural circuit associated with the disease.

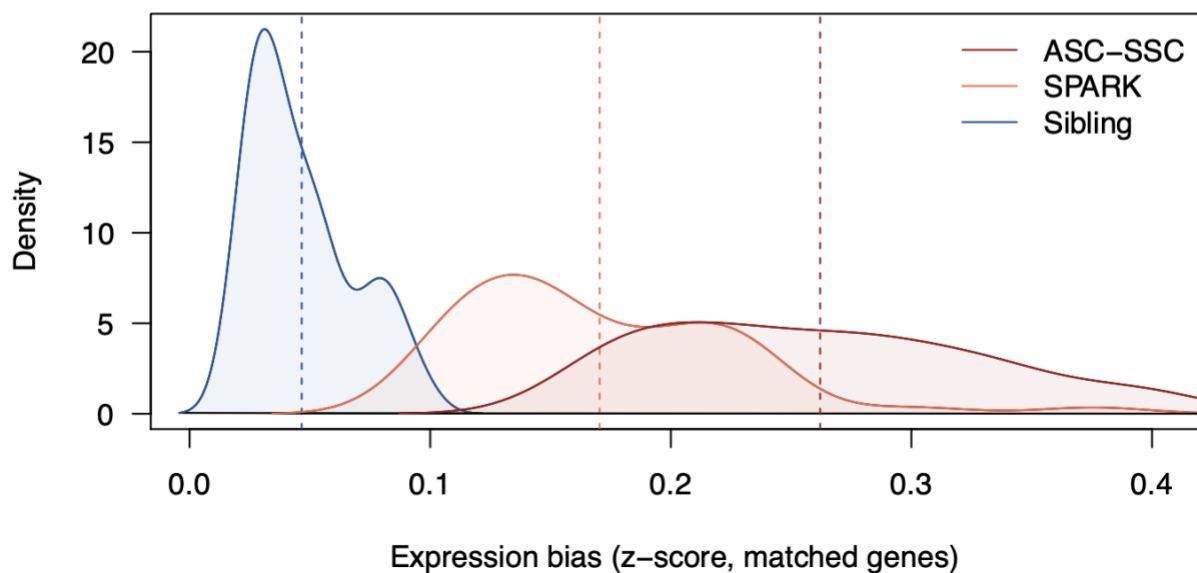
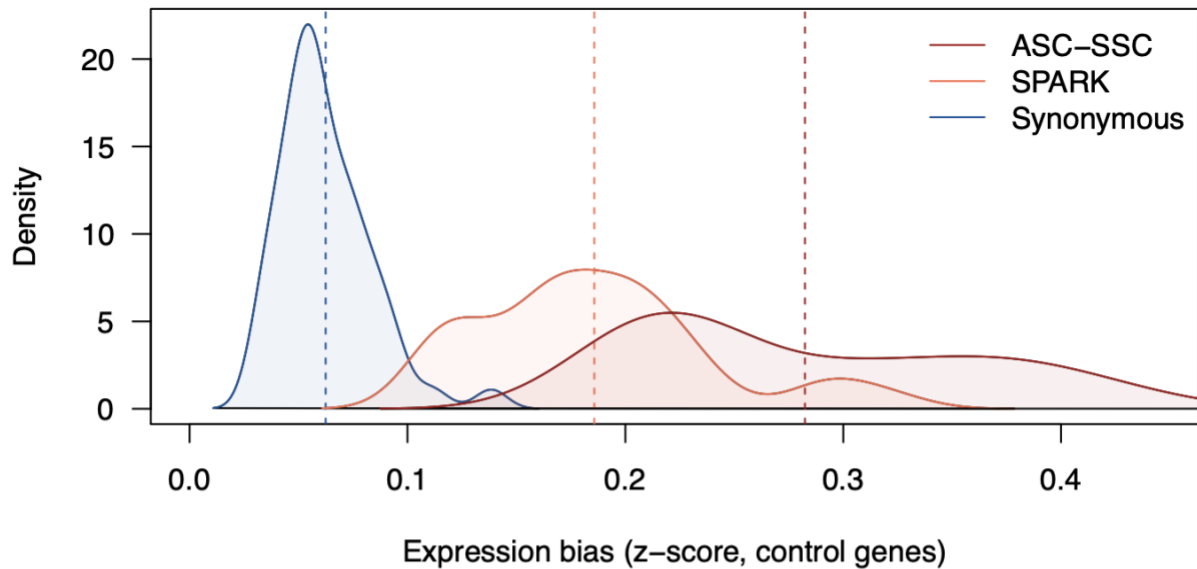


Figure 4.4: Expression Biases of the Most Strongly Biased Structures. (Top) Distributions of expression biases for the 50 most biased neural structures. Expression biases are calculated using control genes based on non-synonymous mutations observed in an unaffected sibling population from SPARK. Distributions are shown for ASD gene sets (ASC-SSC and SPARK) and for genes with synonymous mutations observed in SPARK probands. Vertical dotted lines indicate mean expression biases. **(Bottom)** Distributions of expression biases for the 50 most biased neural structures. Expression biases are calculated using expression-matched genes. Distributions are shown for ASD gene sets (ASC-SSC and SPARK) and for genes with non-synonymous mutations observed in an unaffected sibling population from SPARK. Vertical dotted lines indicate mean expression biases.

4.5 Mutations Converge on Neural Circuits

Having identified the neural structures that ASD mutations specifically target throughout the brain, we proceeded to find a functional circuit among these structures that reflects the core functional components of ASD. We define a circuit as the set of structures and the interconnecting projections within. While we have shown that the implicated structures are strongly connected, we devised an algorithm to search for a strongly connected subset of structures, thereby removing structures with a limited functional relationship to the circuit. Therefore, a circuit represents a system of interconnected components with localized and correlated functional activity such that perturbation of circuit components, namely the structures or the interconnected projections, may explain the behaviors associated with ASD. We implemented a search algorithm to find the most cohesive subset among the biased structures. Specifically, we searched for a subset of structures that are strongly connected to each other but weakly connected to the remainder of the connectome. The algorithm iteratively removes structures following a greedy strategy that optimizes an objective function proportional to the cohesiveness of the remaining structures (Figure 4.5.1). We performed the algorithm on the implicated ASD structures as well as the reference structures, and tested the connectivity of the maximally cohesive ASD subset against the maximal subset for the reference structures. Our method found cohesive networks of neural structures from the cortex, striatum, thalamus, hippocampus, and amygdala (Figure 4.5.2), and we observed that these circuits are significantly and strongly connected (Figure 4.5.3, $P_{\text{Binom, ASC-SSC, Level}} < 10^{-10}$, $P_{\text{Binom, ASC-SSC, Specific}} < 10^{-10}$, $P_{\text{Binom, SPARK, Level}} = 0.0002$, $P_{\text{Binom, SPARK, Specific}} < 10^{-10}$). Moreover, searches among structures determined from ASC-SSC and SPARK genes resulted in similar circuits, suggesting that independent ASD gene sets replicate the implicated ASD circuit (Fisher's exact test $P < 10^{-10}$).

Furthermore, the enrichment of common structures increases as structures are removed, suggesting that the algorithm converges to a common circuit when starting from variable sets of implicated structures.

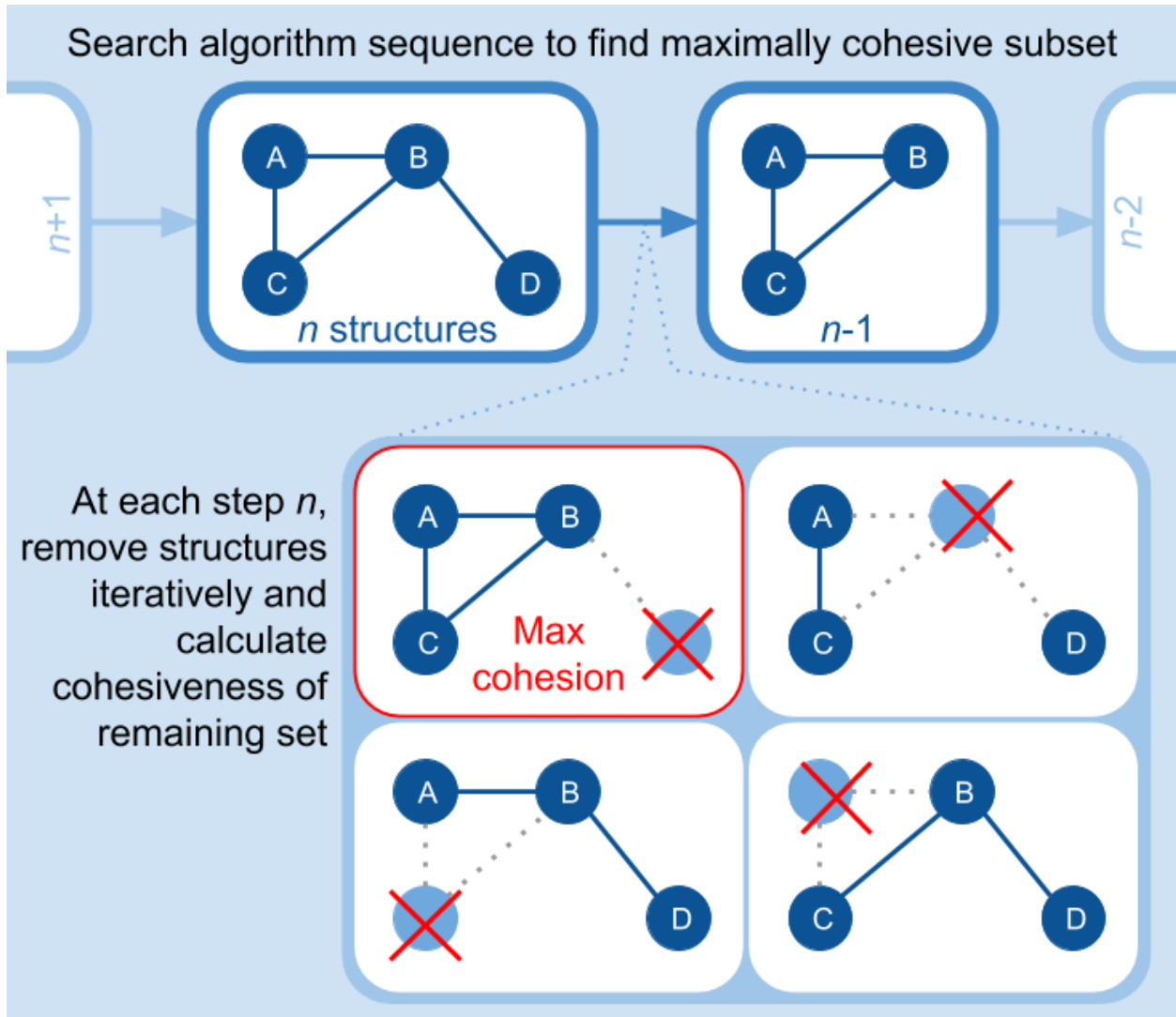


Figure 4.5.1: Search Algorithm Diagram. Diagram shows the greedy search algorithm that iteratively trims structures. The top panel shows the iterative greedy search, while the bottom panel shows the process of dropping out each of the four neural structures during one iterative step of the search. The subpanel in red shows the subset without node D that maximizes cohesiveness of the remaining nodes A, B, and C.

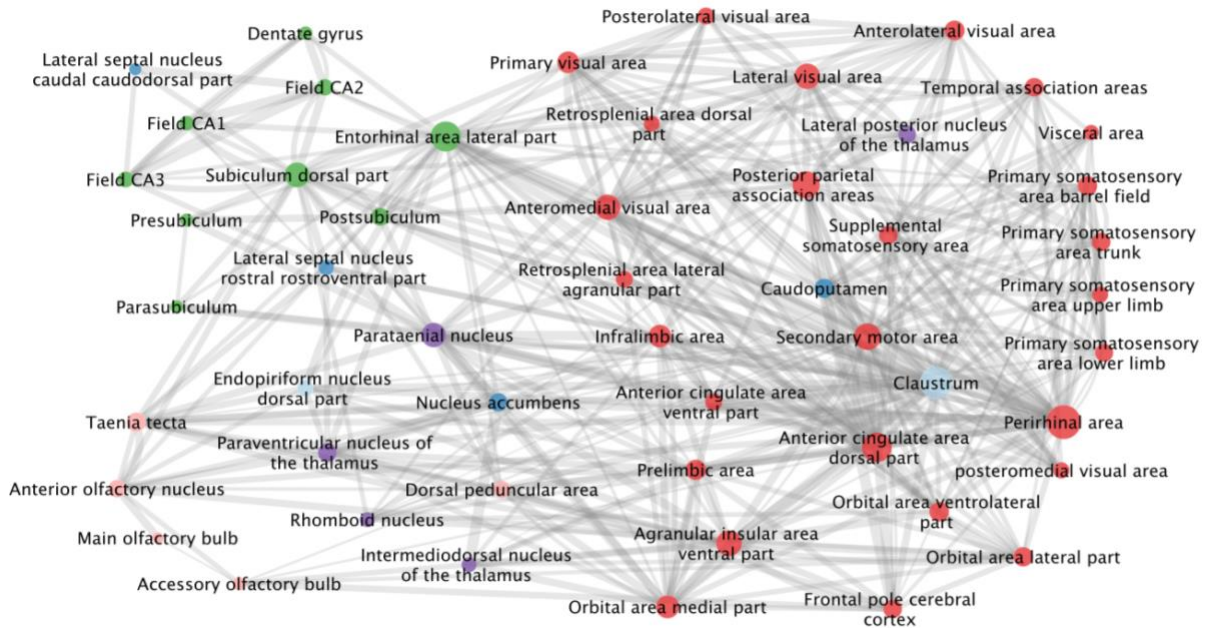


Figure 4.5.2: Implicated ASD Structures Converge on Neural Circuits. Network of component neural structures and the associated connections between structures. Network is a composite of the implicated circuits determined by our maximal-cohesion search algorithm. Nodes in the network indicate a neural structure with large expression biases and are colored by broad regions of the brain (isocortex in dark red, hippocampal formation in green, striatum in dark blue, thalamus in purple, cortical subplate in light blue, and olfactory areas in light red). Edges between nodes indicate directed connections between structures. Edge widths indicate strength of projection, and node sizes are proportional to the total strengths of a given structure's incoming and outgoing projections.

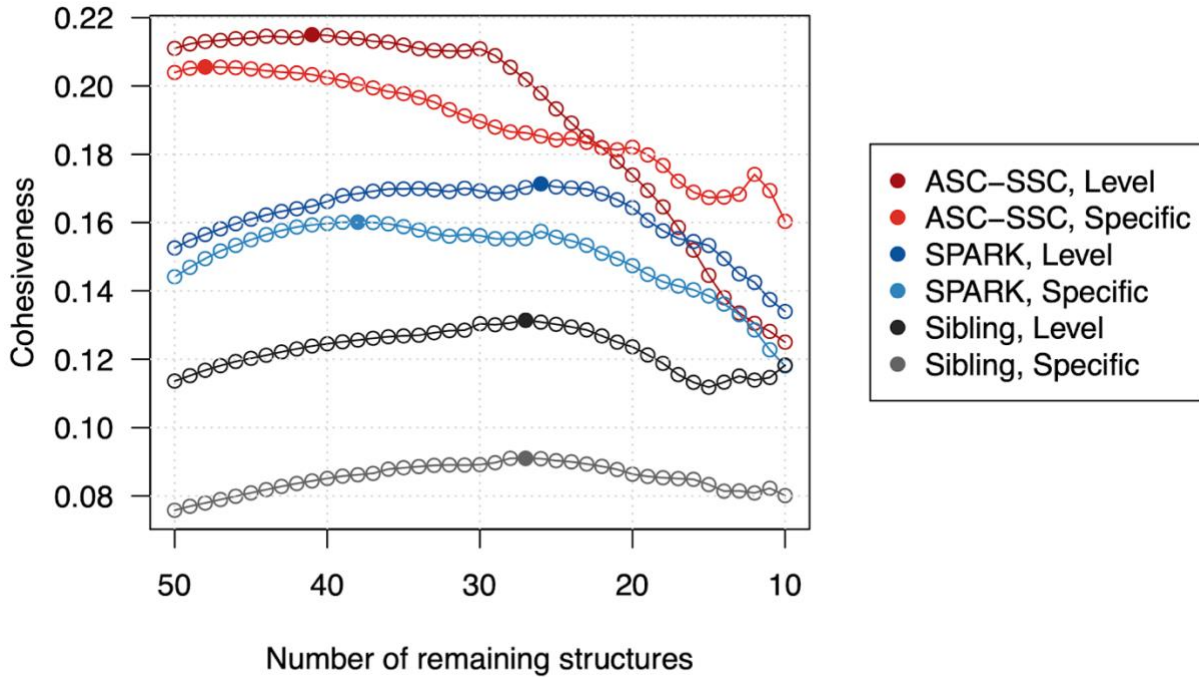


Figure 4.5.3: Search Algorithm Finds Maximally Cohesive Circuits. The greedy search algorithm was applied to sets of 50 structures showing the strongest expression biases. Neural structures are iteratively removed in a manner that maximizes a cohesion objection function. Mean binary cohesiveness is shown for the remaining structures as the search progresses and individual structures are removed, as indicated by the y-axis, whereas the x-axis indicates the number of remaining structures. Maximally cohesive structure subsets are indicated by solid points in each search profile. Search profiles are shown for various expression biases.

We explored the topological properties of the ASD circuit in the context of the broader connectome. As is the case with the connectome, connections among our circuit are sparse (8.4% versus 6.8% connectome density). On average, each ASD structure has approximately 8 projections from and to other ASD structures, and communicates with 6 distal structures and 2 local structures. Moreover, we estimate an exponential degree distribution among the circuit, which reflects the distribution of the wider connectome.

4.6 ASD Circuit Integrates Distant and Functionally Diverse Regions

We also investigated the proportion of distant and local projections that constitute the ASD circuit. We used gene expression correlation as an inverse proxy for neural structure

distance. Interestingly, following stratification of the connectome by distance, we observed that distal connections comprise an outsized fraction of the connections in the circuit (Figure 5), with the most distal connections accounting for more than 5x the fraction for the most proximal connections. This trend was witnessed across both ASC-SSC and SPARK circuits and may suggest that long-range projections or projections between transcriptionally diverse structures play a crucial role in the ASD circuit, possibly in the coordination of functional activity across different regions of the brain. Indeed, the most distal circuit projections largely consist of connections between cortical and thalamic structures, as well as connections among the hippocampal formation, pallidum, and cortex.

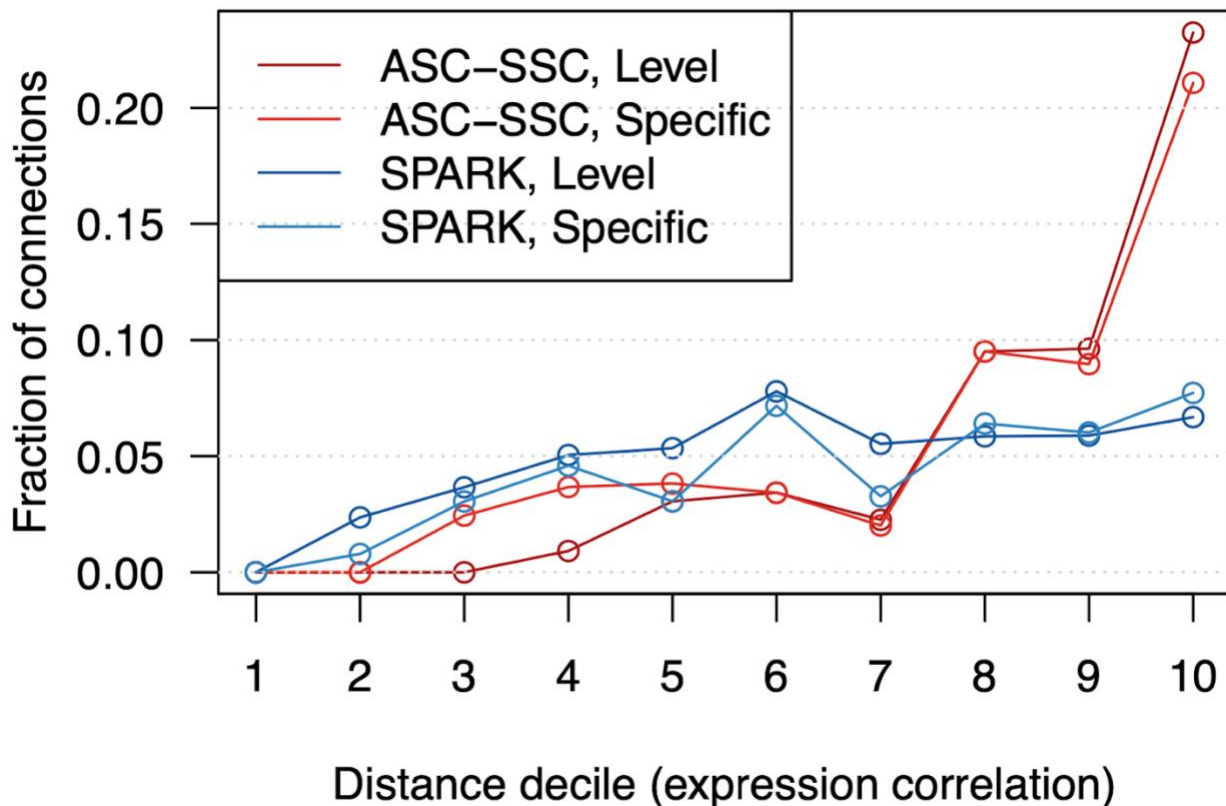


Figure 4.6: ASD Circuit Integrates Functionally Diverse Neural Structures. Fraction of connections among the connectome that constitute the ASD circuit stratified structure distance. Gene expression correlation is used as an inverse proxy for distance between structures. Connectome is stratified by distance decile, as indicated by x-axis. Y-axis shows

the fraction of connections for each decile that occurs between the neural structures in a circuit. Fractions are shown for the various expression biases.

4.7 Small and Diverse Fraction of ASD Circuit is Perturbed in Each Proband

On average, a single *de novo* exomic mutational event occurs in an autistic individual in the ASC, SSC, and SPARK cohorts. When we consider mutations individually, ASD mutations target significantly more structures that are more strongly interconnected than mutations occurring in the neurotypical sibling population ($P_{\text{MWU, Structures, ASC-SSC}} = 5.0 \times 10^{-7}$, $P_{\text{MWU, Cohesion, ASC-SSC}} = 1.7 \times 10^{-4}$, $P_{\text{MWU, Structures, SPARK}} = 7.0 \times 10^{-5}$, $P_{\text{MWU, Cohesion, SPARK}} = 0.012$). This suggests that individual mutations observed in ASD probands have a more profound impact on the brain. We also characterized the impact of a single ASD mutation on the brain. When we consider neural structures with gene expression specificity z-scores greater than 1, we observe that individual ASD mutations on average specifically impact 11.9 ASC-SSC structures and 9.9 SPARK structures among the most biased structures. At this specificity, affected structures are distributed among 4.2 regions and share few connections, suggesting that the mutational impact is dispersed throughout the circuit and not limited to similar or neighboring structures. Moreover, we observe that specificity of impact varies considerably depending on the harboring gene (ASC-SSC CV = 0.90, SPARK CV = 0.75), as genes may be specifically expressed in many structures or more ubiquitously expressed throughout the brain. Interestingly, mutations that target more structures also appear to affect structures that are more highly connected (ASC-SSC Spearman $\rho = 0.55$, SPARK $\rho = 0.56$). Thus, certain mutations have a much larger impact on the ASD circuit than others both in terms of the number of targeted structures and the proportion of associated connections as well.

4.8 Relation to Known Neural Systems and Implications for Medicine

One established hallmark of ASD is the broad phenotypic heterogeneity as evidenced by the “subclinical constellation of traits” observed in affected individuals (Losh 2009). This hallmark may be partly explained by the convergence of ASD mutations on functionally diverse cortical and subcortical regions, as we observed in our study. Moreover, we demonstrate that long-range connections, such as those connecting cortical and subcortical regions, appear to play a large role in the ASD circuit. Others have suggested that long-range connections enable efficient communication between distant brain regions that are functionally distinct (Sporns 2019). Furthermore, global integration and communication in the brain are shown to be risk factors for disease, and may explain why densely connected cortical and subcortical regions are involved in numerous disorders including ASD.

The ASD circuit resembles known neural systems with relevance to the disorder (Figure 4.8). Using a community detection algorithm, we identified clusters among the circuit in an analytical manner. We observed that one cluster primarily centers on the cortex and the other is localized to the hippocampus. However, most ASD genes (85%) impact structures in both clusters, and only a small fraction appear to target one cluster. The cortical cluster includes many components of known cortico-basal ganglia-thalamo-cortical regulatory loops. Cortico-basal ganglia circuits have been previously implicated not only in movement/kinetic disorders but also in neuropsychiatric disorders such as schizophrenia and obsessive-compulsive disorder (DeLong 2007, Utter 2008). Projections from diverse regions of the cortex are integrated in the striatum, which, via the thalamus, project back to the originating cortex, thereby forming cortico-basal ganglia loops (Shepherd 2013, Rothwell 2014, Goto 2008, Peters 2016, Price 2012, Halassa 2017, Downar 2016, Robbins 2006, Redgrave 2005). Basal ganglia circuits integrate sensory and

motor information, memories, and motivational states in order to direct attention and select specific responses (Wimmer 2015, Halassa 2017, Rikhye 2018). Multiple parallel loops have been proposed, and the limbic basal ganglia circuit is of particular interest as it may integrate information conveyed from the limbic system including hippocampal structures such as those in the hippocampal cluster, amygdalar nuclei, and limbic cortical areas (LeDoux 2000). Due to the integrative nature of basal ganglia circuits, one could reason that dysfunction in sensory selection and processing in individuals with ASD would manifest as restricted interests or as abnormal or limited reaction to social or environmental stimuli (Fuccillo 2016).

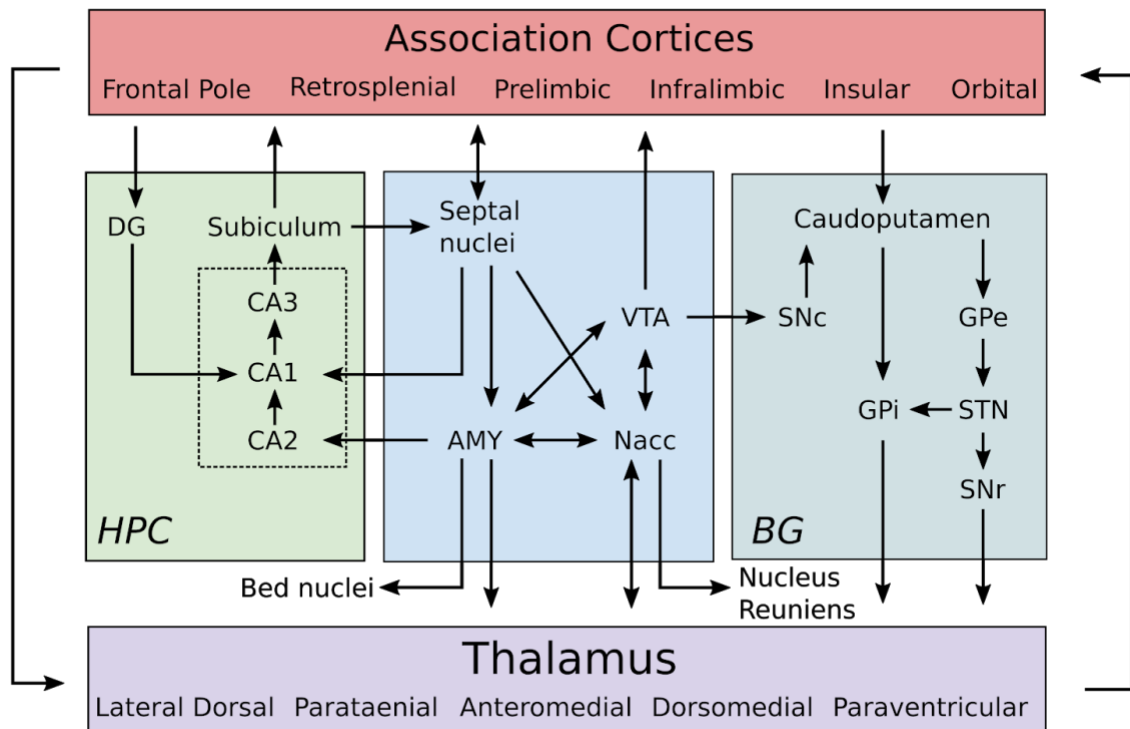


Figure 4.8: ASD Circuit Involves Cortico-Striatal-Thalamo-Cortical (CSTC) Loop and Limbic System. Diagram of the proposed neural circuits associated with ASD along with known systems with overlapping component structures, such as the CSTC loop and limbic system. Arrows indicate direction of information flow, and structures are grouped by broad regions of the brain, such as hippocampal formation (HPC), basal ganglia (BG), striatum, cortex, and thalamus. DG indicates dentate gyrus, AMY amygdala, VTA ventral tegmental area, Nacc nucleus accumbens, SNc substantia nigra pars compacta, GPe external globus

pallidus, GPI, internal globus pallidus, STN subthalamic nucleus, and SNr substantia nigra pars reticulata.

Basal ganglia circuits utilize functionally opposed direct and indirect pathways to initiate or inhibit activity, and are operated in concert to select intended actions while suppressing unintended activity. The pathways are modulated by differential responses to dopamine, whereby D1-class and D2-class dopamine receptors in striatal medium spiny neurons respectively excite and inhibit downstream cortical neurons (Utter 2008). These two pathways mediate action selection, and a balance of these opposing pathways leads to cognitive flexibility and behavioral variability. In ASD individuals, inefficiency in these pathways may explain phenotypes such as motor and speech stereotypy as well as cognitive rigidity (Fuccillo 2016).

Dopaminergic neurons, in modulating the striatum, also serve a critical role in learning, memory, and reward-processing by reinforcing associations between environmental cues, responses, and outcomes. For instance, dopaminergic modulation of the nucleus accumbens mediates motivational states and reward processing by calibrating expectation with outcomes. In individuals with ASD, abnormalities in basal ganglia reward processing may manifest as stifled or abnormal responses to social stimuli or reward (Fuccillo 2016). Moreover, ineffective reward processing when rewards change may facilitate the inflexibility in responses to stimuli that is characteristic of ASD. Dopaminergic modulation has also been proposed to facilitate and encode motor chunking, or simple movement programs that can be linked together in sequence to form more complex motor output. Stereotyped routines such as stimming observed in ASD individuals may reflect perturbed action selection of motor programs, whereby sequences are not correctly assembled and instead a single program is repeatedly initiated.

In our circuit, the striatum acts as a central hub through which distant regions communicate. We ranked structures based on graph theoretic properties such as node

betweenness and eigenvector centrality, and we observed that the caudoputamen and nucleus accumbens are two of the most critical structures. Other studies have observed that rich network hubs in the brain are vulnerable hot spots for disease due to the propensity for high disruption (Sporns 2019), and we observed a similar phenomenon for ASD circuits. In our previous work using cell type-specific expression data, we found that medium spiny neurons, which comprise the majority of the neuronal population in the striatum, showed prominent mutational biases (Chang 2015).

Our work is, to the best of our knowledge, the first analysis of a psychiatric disorder to use a combination of both genome-wide and brain-wide mutation, expression, and connectivity data for the unbiased detection of brain circuits that may explain disease etiology. While we specifically investigated the neural circuits involved in ASD, our method uses data collected from neurotypical brains and does not presuppose biases in the autistic brain. As such, our method is general and can be used to identify suboptimal networks for other psychiatric or neurological disorders with a large genetic component, such as major depression disorder or epilepsy. Our framework may also have the potential to disambiguate the relationship between dysconnectivity and phenotype, and may be used to identify neural pathways for optogenetic studies (Tye 2012) or to predict symptoms, severity, or long-term outcomes for ASD or for other disorders of the brain. As we and others continue to investigate the “connectome landscape of dysconnectivity”, connectotyping individual patients based on their genetic information may provide another avenue for precision medicine in psychiatric or neurological clinics and allow for interventions such as deep brain stimulation (Cagan 2019) and transcranial magnetic stimulation (Peters 2016) to be personalized.

Chapter 5: Sensitivity of Phenotypes to Gene Dosage

5.1 Exons as Units of Phenotypic Impact for Mutations

Phenotypes associated with ASD vary considerably (de la Torre-Ubieta et al., 2016; Krumm et al., 2014; Ronemus et al., 2014a; Sharma et al., 2015) across autism probands, and the nature of this phenotypic heterogeneity is not well understood (Jeste and Geschwind, 2014; Talkowski et al., 2014). Despite the complex genetic architecture of ASD (Gaugler et al., 2014) (Anney et al., 2012; Gratten et al., 2014; Krumm et al., 2015b; Robinson et al., 2014; Turner et al., 2017), a subset of cases from simplex families are known to be strongly affected by *de novo* mutations with severe deleterious effects (Ronemus et al., 2014b). Interestingly, despite having relatively simpler genetic architecture, simplex autism cohorts often display as much phenotypic heterogeneity as more general ASD cohorts (Berends et al., 2019; Dissanayake et al., 2019). This provides an opportunity for an in-depth exploration of the etiology of the autism phenotypic heterogeneity, and we performed such an analysis, focusing on severely damaging, likely gene-disrupting (LGD) mutations, using data collected in the Simons Simplex Collection (SSC) and the Simons Variation in Individuals Project (VIP) (Fischbach and Lord, 2010; Simons Vip, 2012).

We investigated phenotypes resulting from truncating mutations affecting the same exon in unrelated ASD probands, and we observed that unrelated ASD probands showed strikingly more similar phenotypes compared to probands with LGD mutations in the same gene (Figure 5.1; same exon FSIQ/NVIQ/VIQ average IQ difference 8.9, 8.3, 17.3 points; same gene average difference 28.3, 25.7, 34.9 points; Mann-Whitney *U* one-tail test $P = 0.003, 0.005, 0.016$). Because of well-known gender differences in autism susceptibility, we also compared IQ differences between probands of the same gender harboring truncating mutations in the same

exon to IQ differences between probands of different genders. Thus, stratification by gender further decreases the phenotypic differences between probands with LGD mutations in the same exon (same gender FSIQ/NVIQ/VIQ average difference 5.4, 7.2, 12.2; different gender average difference 14.7, 10, 25.7; MWU one-tail test $P = 0.04, 0.29, 0.07$). Notably, the patterns of phenotypic similarity between different probands only extended to mutations affecting the same exon. The average IQ differences between probands with LGD mutations in neighboring exons were not significantly different compared to mutations in non-neighboring exons (MWU one-tail test $P = 0.6, 0.18, 0.8$). The observed effects were also specific to LGD mutations; probands with either synonymous or missense mutations in the same exon were as phenotypically diverse as random pairs of ASD probands ($P = 0.93, 0.97, 0.95; P = 0.8, 0.5, 0.8$).

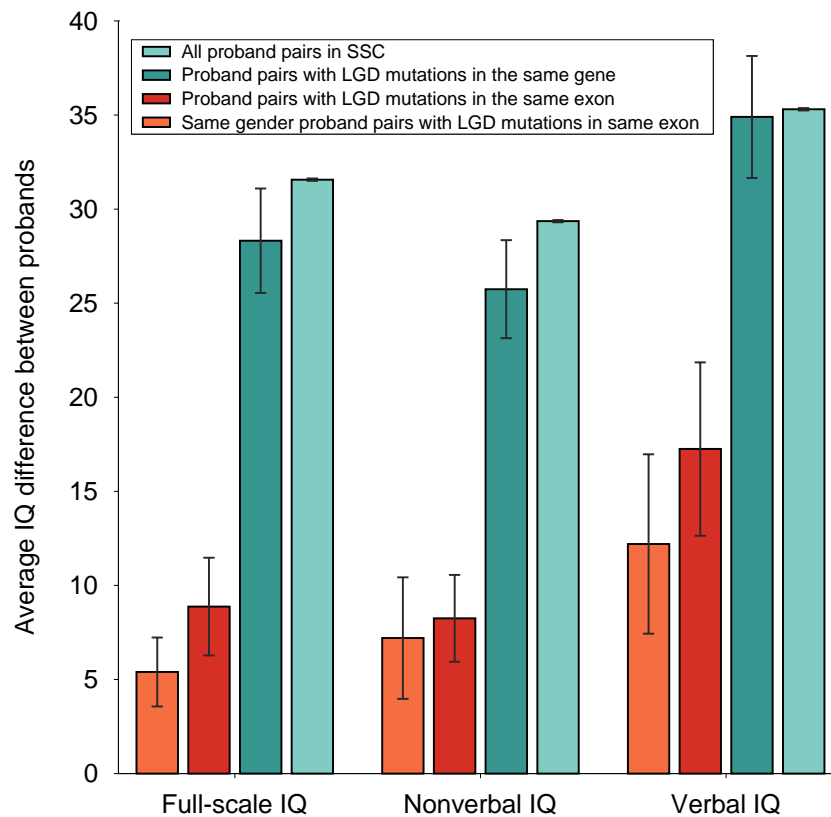


Figure 5.1: Average Difference in IQ Between SSC Probands (Chiang 2020). Each bar shows the average IQ difference between pairs of probands from different groups. From right to left, the bars represent the average IQ difference between pairs of probands in the

entire SSC population (light green), between probands with *de novo* LGD mutations in the same gene (dark green), between probands with *de novo* LGD mutations in the same exon (red), between probands of the same gender and with *de novo* LGD mutations in the same exon (orange). Error bars represent the SEM.

We next explored the relationship between phenotypic similarity and the proximity of truncating mutations in the corresponding protein sequences. This analysis revealed that probands with LGD mutations in the same exon often had similar IQs, despite being affected by truncating mutations separated by scores to hundreds of amino acids in protein sequence. Furthermore, we found probands with LGD mutations in the same exon to be more phenotypically similar than probands with LGD mutations separated by comparable amino acid distances in the same protein sequence but not necessarily in the same exon (NVIQ distance-matched permutation test $P = 0.002$). We also investigated whether *de novo* mutations truncating a larger fraction of protein sequences resulted in more severe intellectual phenotypes, and we observed no significant correlations between the fraction of truncated protein and the severity of intellectual phenotypes (NVIQ Pearson's $R = 0.05$). We also did not find any significant biases in the distribution of truncating *de novo* mutations across protein sequences compared with the distribution of synonymous *de novo* mutations (Kolmogorov-Smirnov two-tail test $P = 0.9$). It is possible that the lack of the correlation between phenotypic impact and the fraction of truncated sequence is due to the averaging of various effects across different proteins. Therefore, for genes with recurrent mutations, we used a paired test to investigate whether truncating a larger fraction of the same protein sequence led to more severe phenotypes. This analysis also showed no substantial phenotypic difference due to LGD mutations truncating different fractions of the same protein (average NVIQ difference 0.24 points; Wilcoxon signed-ranked one-tail test $P = 0.44$). We also investigated, using the Pfam database (El-Gebali et al., 2019), whether mutations that truncate the same protein domain lead to more similar phenotypic differences. The results

demonstrated that mutations in different exons, even when truncating the same protein domain, resulted, on average, in phenotypes as different as due to LGD mutations in the same protein (average NVIQ differences = 28.1).

Although we primarily analyzed the impact of autism mutations on intellectual phenotypes, similar dosage and isoform expression changes of affected genes lead to analogous results for other quantitative ASD phenotypes. Specifically, SSC probands with truncating mutations in the same exon exhibited more similar Vineland Adaptive Behavior Scales (VABS) scores compared to probands with mutations in the same gene (composite standard score difference of 4.7 versus 12.1 points, MWU one-tail test $P = 0.017$). In contrast, VABS differences between probands with truncating mutations in the same gene were not significantly different than for randomly paired probands (12.1 versus 13.7 points, MWU one-tail test $P = 0.23$). Furthermore, probands with truncating mutations in the same exon also displayed more similar fine motor skills based on the Purdue Pegboard Test (1.2 versus 3.0 for the average difference in normalized tasks completed with both hands, MWU one-tail test $P = 0.02$). Coordination scores in the Social Responsiveness Scale questionnaire were also more similar in probands with LGD mutations in the same exon compared to probands with mutations in the same gene (0.6 versus 1.1 for the average difference in normalized response, MWU one-tail test $P = 0.05$). We also sought to validate the observed phenotypic patterns using an independent cohort of ASD probands. To that end, we analyzed an independently collected dataset from the ongoing Simons Variation in Individuals Project (VIP). Reassuringly, probands from the VIP cohort with truncating *de novo* mutations in the same exon also exhibited strikingly more similar VABS phenotypic scores compared to probands with mutations in the same gene (VABS composite standard score difference 6.0 versus 12.4, MWU one-tail test $P = 0.014$).

These results suggest that the occurrence of *de novo* LGD mutations in the same exon, rather than simply the proximity of mutation sites in protein or nucleotide sequences, is primarily responsible for similar phenotypic consequences in unrelated probands. We hypothesized that truncating mutations in the same exon usually affect, due to nonsense-mediated decay (NMD) (Chang et al., 2007), the expression of exactly the same sets of splicing isoforms. Therefore, such mutations should lead to particularly similar phenotypes, both through similar decreases in overall gene dosage and similar perturbations to the mRNA expression of affected transcriptional isoforms. We next investigated the relationship between gene dosage and phenotype across different genes.

5.2 Variation in Sensitivity of Phenotypes to Gene Dosage

To quantify the sensitivity of IQ to changes in the expression of specific genes, we considered a simple linear dosage model where we assumed for genes with recurrent truncating mutations in SSC that decreases in probands' IQs are linearly proportional to the predicted relative decrease in overall gene dosage due to NMD. We further assumed that each human gene can be characterized by a parameter, which we call its phenotypic dosage sensitivity (PDS), characterizing the linear relationship between changes in gene dosage compared to wild type and the corresponding changes in a given human phenotype. Numerically, we defined IQ-associated PDS to be equal to the average change in IQ resulting from a 10% change in gene dosage. Notably, as we expected, PDS values varied substantially across 24 considered human genes (NVIQ CV = 0.57).

We used the aforementioned linear model to explore the relationship between the relative expression values of exons harboring LGD mutations and the corresponding decreases in probands' intellectual phenotypes. To account for differences in phenotypic sensitivity to dosage

changes across genes, we normalized the observed changes in IQ by the estimated PDS values of affected genes. This analysis revealed that mutation-induced gene dosage changes are indeed strongly correlated with the normalized phenotypic effects (NVIQ Pearson's $R = 0.63$, permutation test $P = 0.02$; Figure 5.2). Moreover, PDS values for the same gene vary across phenotypes, which suggests that PDS values are specific to phenotype-gene pairs. Although we evaluated PDS parameters using predicted NMD-induced dosage changes, it may be possible to infer these parameters using other mechanisms of dosage change, such as regulatory mutations. As genetic and phenotypic data accumulate, it will be interesting to estimate gene-specific PDS values for multiple phenotypes and for a substantial number of ASD risk genes. Furthermore, due to consistent patterns of gene expression changes across tissues, it may be possible to estimate PDS parameters for other genetic disorders and phenotypes.

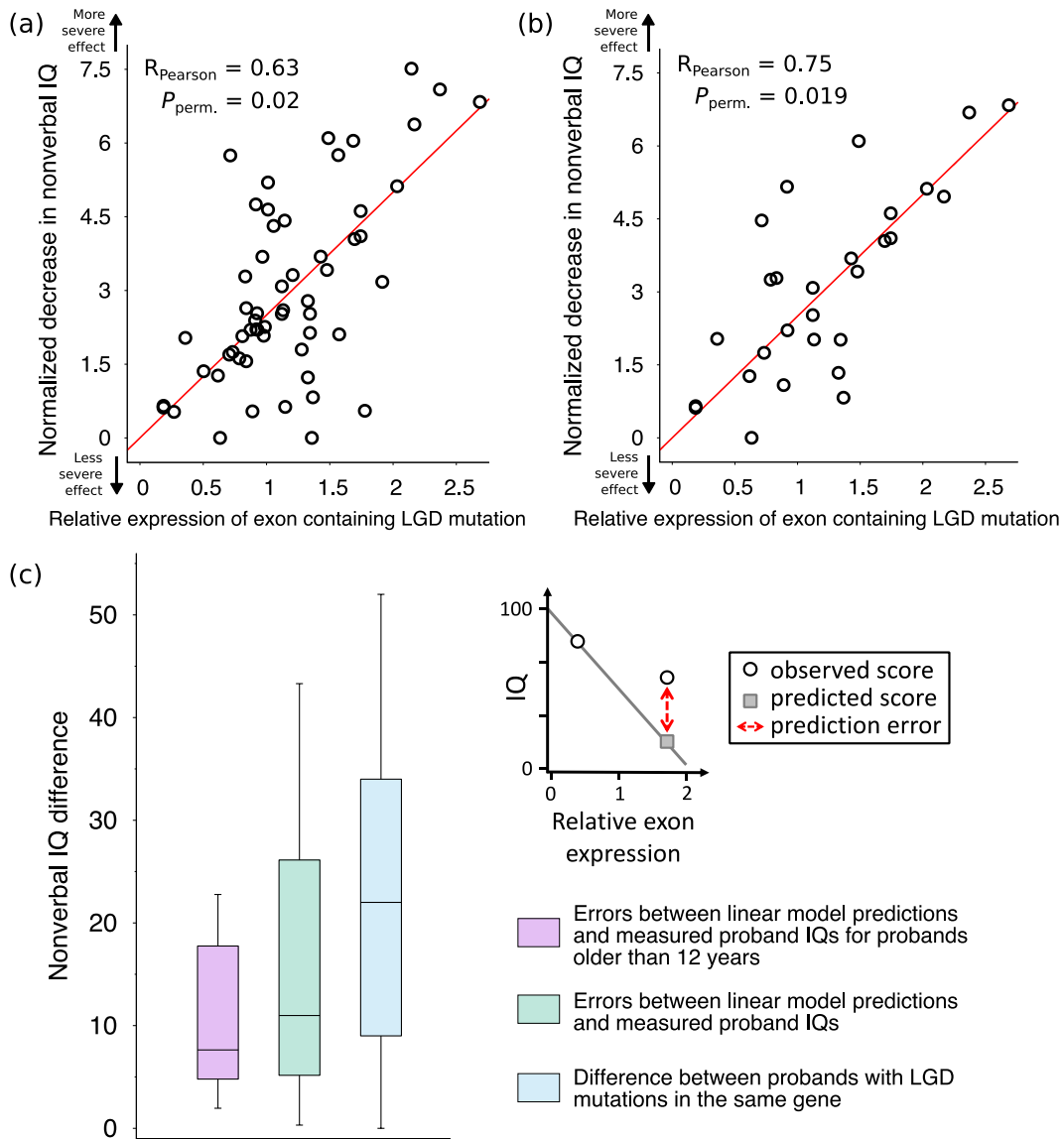


Figure 5.2: Relationship Between the Relative Expression of Exons Harboring LGD Mutations and the Corresponding Decrease in Probands' Intellectual Phenotypes (Chiang 2020). (A) Each point corresponds to a proband with an LGD mutation in a gene; only genes with multiple LGD mutations in SSC were considered. The x-axis represents the relative expression (exon expression normalized by the total gene expression) of the exon harboring the LGD mutation. The y-axis represents the effect of each mutation on the affected proband's nonverbal IQ, normalized by the sensitivity of nonverbal IQ to changes in gene dosage. The regression line across all points is shown in red. *P*-values were calculated based on randomly shuffled data. (B) Same as (B), but with the analysis restricted to the older half of probands in SSC (median age 8.35 years). (C) Boxplots represent the distribution of errors in predicting the effects of LGD mutations on nonverbal IQ compared to the differences in IQ scores between probands with LGD mutations in the same gene (blue); prediction errors

are shown for all probands (green) and for probands older than 12 years (purple). Only genes with multiple LGD mutations in SSC were considered. The ends of each solid box represent the upper and lower quartiles; the horizontal lines inside each box represent the medians; and the whiskers represent the 5th and 95th percentiles. The inset panel illustrates the linear regression model used to perform leave-one-out predictions of proband IQs. Round open points represent measured scores for probands with LGD mutations in the same gene, the grey square point represents the predicted phenotypic score, and the red dotted line represents the prediction error.

5.3 Implications for Precision Medicine

Previous studies explored phenotypic similarity in syndromic forms of ASD due to mutations in specific genes (Ben-Shalom et al., 2017; Bernier et al., 2014a; Helsmoortel et al., 2014; Sztainberg and Zoghbi, 2016; van Bon et al., 2016a). Nevertheless, across a large collection of contributing genes, the nature of the substantial phenotypic heterogeneity in ASD is not well understood. Interestingly, the diversity of intellectual and other important ASD phenotypes resulting from de novo LGD mutations in the same genes is usually only slightly (~10%) smaller than the phenotypic diversity across the entire ASD cohort (Figure 5.1). The presented results suggest that truncating mutations usually result in a range of relatively mild NMD-induced gene dosage changes, on average decreasing gene expression by ~15-30%. Our study further suggests a hierarchy of biological mechanisms contributing to phenotypic heterogeneity in simplex ASD cases triggered by LGD mutations in different genes and within the same gene.

Across LGD mutations, there is a significant but small correlation between a target gene's brain expression level and the resulting intellectual phenotype ($R^2 = 0.02$, $P = 0.03$). This correlation is small, at least in part, due to the significant variability of expression levels across different exons in a gene. Indeed, intellectual phenotypes correlate significantly better with the relative expression level of exons harboring LGD mutations ($R^2 = 0.10$, $P = 0.011$). In addition to effects associated with different expression levels of exons, there is also substantial variability

in the sensitivity of each specific phenotype to dosage changes of different genes. When we account for varying dosage sensitivities using gene-specific PDS values, the correlation between predicted dosage changes and normalized phenotypic effects becomes substantial ($R^2=0.4$, $P = 0.02$, Figure 5.2). As the heritability of IQ phenotypes usually increases with age, we observe even stronger dosage-phenotype correlations for older probands ($R^2 = 0.56$, $P = 0.019$, Figure 5.2). Furthermore, even perturbations leading to similar dosage changes in the same gene may result in diverse phenotypes in cases where different, functionally distinct splicing isoforms are truncated. However, when exactly the same sets of isoforms are perturbed, as for LGD mutations in the same exon, the resulting phenotypes, even in unrelated ASD probands, become especially similar (Figure 5.1). For LGD mutations affecting intellectual phenotypes, we found that same exon membership accounts for a larger fraction of phenotypic variance than multiple other genomic features, including expression, evolutionary conservation, pathway membership, and domain truncation. There are likely deviations from the aforementioned patterns for specific genes and specific truncating mutations. For example, truncated proteins that escape NMD may lead to partial buffering, due to remaining activity, or to further damaging effects, due to dominant negative interactions. Nevertheless, our results demonstrate that for de novo LGD mutations in ASD, exons, rather than genes, usually represent a unit of effective phenotypic impact.

Our results also suggest that changes in ASD phenotypes induced by LGD mutations may be characterized by a simple linear model quantifying the sensitivity of a phenotype to changes in gene dosage. We observe that PDS values for the same phenotype vary substantially across genes, and that PDS differences are a major source of phenotypic variability. Moreover, PDS values for the same gene vary across phenotypes (for example, correlation between PDS values

for IQ and VABS across 24 genes, $R^2 = 0.37$, $P = 0.001$), which suggests that PDS values are specific to phenotype-gene pairs. Although we evaluated PDS parameters using predicted NMD-induced dosage changes, it may be possible to infer these parameters using other mechanisms of dosage change, such as regulatory mutations. As genetic and phenotypic data accumulate, it will be interesting to estimate gene-specific PDS values for multiple phenotypes and for a substantial number of ASD risk genes. Furthermore, due to consistent patterns of gene expression changes across tissues, it may be possible to estimate PDS parameters for other genetic disorders and phenotypes. We note in this respect that quantitative gene-dosage relationships have been recently characterized for yeast fitness values in different environmental conditions (Keren et al., 2016).

In the present study, we focused specifically on simplex cases of ASD, in which *de novo* LGD mutations are highly penetrant and where the contribution of genetic background is minimized. It is likely that differences in genetic background and environment represent other important sources of phenotypic variability (Gandal et al., 2016; Robinson et al., 2014; Robinson et al., 2016). Therefore, in more diverse cohorts, individuals with LGD mutations in the same exon will likely display greater phenotypic heterogeneity. For example, the Simons Variation in Individuals Project identified broad spectra of phenotypes associated with specific variants in the general population (D'Angelo et al., 2016a; Hanson et al., 2015; Qureshi et al., 2014; Simons Vip, 2012). We also observed significantly larger phenotypic variability for probands from sequenced family trios, i.e. families without unaffected siblings. For these probands, the enrichment of *de novo* LGD mutations is substantially lower and the contribution from genetic background is likely to be larger (Zhao et al., 2007), resulting in more pronounced phenotypic variability.

Our study may have important implications for precision medicine (Collins and Varmus, 2015; Gandal et al., 2016; Geschwind and State, 2015). The presented results indicate that relatively mild decreases in affected gene dosage may account for a substantial fraction of adverse phenotypic consequences. Thus, from a therapeutic perspective, compensatory expression of intact alleles, as was recently demonstrated in mouse models of ASD and other diseases (Ehninger et al., 2008; Guy et al., 2007; Matharu et al., 2019; Mei et al., 2016), may provide an approach for alleviating phenotypic effects for at least a fraction of ASD cases. From a prognostic perspective, our results suggest that by sequencing and phenotyping sufficiently large patient cohorts with truncating mutations in different exons, it may be possible to understand likely phenotypic consequences originating from LGD mutations in specific exons. Furthermore, because we observe consistent patterns of expression changes across multiple human tissues, similar analyses may be also extended to other disorders affected by highly penetrant truncating mutations.

References

- Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Ann N Y Acad Sci* 1191, 42-61.
- Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 21, 4781-4792.
- Bae, B.-I., Jayaraman, D., and Walsh, C.A. (2015). Genetic Changes Shaping the Human Brain. *Developmental cell* 32, 423-434.
- Bayes, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci* 14, 19-21.
- 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* 82, 224-232.
- Berends, D., Dissanayake, C., and Lawson, L.P. (2019). Differences in Cognition and Behaviour in Multiplex and Simplex Autism: Does Prior Experience Raising a Child with Autism Matter? *J Autism Dev Disord* 49, 3401-3411.
- Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerds, J., Baker, C., Vulto-van Silfhout, A.T., et al. (2014a). Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 158, 263-276.

- 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. (2014b). Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. *Cell* 158, 263-276.
- Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature reviews Neuroscience* 16, 551-563.
- Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169, 1177-1186.
- Burguiere, E., Monteiro, P., Feng, G., and Graybiel, A.M. (2013). Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. *Science* 340, 1243-1246.
- C Yuen, R.K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience* 1, 160271-160611.
- Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J., and Vitkup, D. (2015). Genotype to phenotype relationships in autism spectrum disorders. *Nature Neuroscience* 18, 191-198.
- Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76, 51-74.
- Chen, J.A., Peñagarikano, O., Belgard, T.G., Swarup, V., and Geschwind, D.H. (2015). The emerging picture of autism spectrum disorder: genetics and pathology. *Annual review of pathology* 10, 111-144.
- Churchland, P.S., and Sejnowski, T.J. (1988). Perspectives on cognitive neuroscience. *Science* 242, 741-745.

- Collins, F.S., and Varmus, H. (2015). A new initiative on precision medicine. *N Engl J Med* 372, 793-795.
- Conrad, D.F., Keebler, J.E.M., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V., et al. (2011). Variation in genome-wide mutation rates within and between human families. *Nature Genetics* 43, 712-714.
- D'Angelo, D., Lebon, S., Chen, Q., Martin-Brevet, S., Snyder, L.G., Hippolyte, L., Hanson, E., Maillard, A.M., Faucett, W.A., Mace, A., et al. (2016a). Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA Psychiatry* 73, 20-30.
- D'Angelo, D., Lebon, S., Chen, Q., Martin-Brevet, S., Snyder, L.G., Hippolyte, L., Hanson, E., Maillard, A.M., Faucett, W.A., Macé, A., et al. (2016b). Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA psychiatry* 73, 20-30.
- Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., 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.
- 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.
- De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209-215.

Dissanayake, C., Searles, J., Barbaro, J., Sadka, N., and Lawson, L.P. (2019). Cognitive and behavioral differences in toddlers with autism spectrum disorder from multiplex and simplex families. *Autism Res* 12, 682-693.

Doan, R.N., Bae, B.-I., Cubelos, B., Chang, C., Hossain, A.A., Al-Saad, S., Mukaddes, N.M., Oner, O., Al-Saffar, M., Balkhy, S., et al. (2016). Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* 167, 341-354.e312.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* 135, 749-762.

Edbauer, D., Neilson, J.R., Foster, K.A., Wang, C.F., Seeburg, D.P., Batterton, M.N., Tada, T., Dolan, B.M., Sharp, P.A., and Sheng, M. (2010). Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron* 65, 373-384.

Ehninger, D., Han, S., Shilyansky, C., Zhou, Y., Li, W., Kwiatkowski, D.J., Ramesh, V., and Silva, A.J. (2008). Reversal of learning deficits in a *Tsc2*^{+/-} mouse model of tuberous sclerosis. *Nat Med* 14, 843-848.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res* 47, D427-D432.

Ernst, A., and Frisé, J. (2015). Adult neurogenesis in humans- common and unique traits in mammals. *PLoS biology* 13, e1002045.

Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* 68, 192-195.

- Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatr Res* 65, 591-598.
- Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216-220.
- Gandal, M.J., Leppa, V., Won, H., Parikshak, N.N., and Geschwind, D.H. (2016). The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci* 19, 1397-1407.
- 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. *Nature Genetics* 46, 881-885.
- Geschwind, D.H., and State, M.W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol* 14, 1109-1120.
- Gilman, S.R., Chang, J., Xu, B., Bawa, T.S., Gogos, J.A., Karayiorgou, M., and Vitkup, D. (2012a). Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci* 15, 1723-1728.
- Gilman, S.R., Chang, J., Xu, B., Bawa, T.S., Gogos, J.A., Karayiorgou, M., and Vitkup, D. (2012b). Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nature Neuroscience* 15, 1723-1728.
- Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011a). Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron* 70, 898-907.

- Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011b). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70, 898-907.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171-178.
- Gratten, J., Wray, N.R., Keller, M.C., and Visscher, P.M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci* 17, 782-790.
- 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.
- Hanson, E., Bernier, R., Porche, K., Jackson, F.I., Goin-Kochel, R.P., Snyder, L.G., Snow, A.V., Wallace, A.S., Campe, K.L., Zhang, Y., et al. (2015). The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. *Biol Psychiatry* 77, 785-793.
- Helsmoortel, C., Vulto-van Silfhout, A.T., Coe, B.P., Vandeweyer, G., Rooms, L., van den Ende, J., Schuurs-Hoeijmakers, J.H., Marcelis, C.L., Willemsen, M.H., Vissers, L.E., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet* 46, 380-384.
- Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 6, e1001154.
- Huber, K.M., Gallagher, S.M., Warren, S.T., and Bear, M.F. (2002). Altered synaptic plasticity in a mouse model of fragile X mental retardation. *Proc Natl Acad Sci U S A* 99, 7746-7750.

- Huguet, G., Ey, E., and Bourgeron, T. (2013). The Genetic Landscapes of Autism Spectrum Disorders. *dxdoiorgezproxyculcolumbiaedu* 14, 191-213.
- Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y.-H., Yamrom, B., and Wigler, M. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proceedings of the National Academy of Sciences of the United States of America* 112, E5600-5607.
- 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. (2014a). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216-221.
- 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. (2014b). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216-221.
- Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron* 74, 285-299.
- Iturria-Medina, Y., Sotero, R.C., Canales-Rodríguez, E.J., Alemán-Gómez, Y., and Melie-García, L. (2008). Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. *NeuroImage* 40, 1064-1076.
- Jeste, S.S., and Geschwind, D.H. (2014). Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* 10, 74-81.
- Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489.

- Kelleher, R.J., 3rd, and Bear, M.F. (2008). The autistic neuron: troubled translation? *Cell* 135, 401-406.
- Kennedy, M.B. (1997). The postsynaptic density at glutamatergic synapses. *Trends Neurosci* 20, 264-268.
- Kennedy, M.B. (2000). Signal-processing machines at the postsynaptic density. *Science* 290, 750-754.
- Keren, L., Hausser, J., Lotan-Pompan, M., Vainberg Slutskin, I., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2016). Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* 166, 1282-1294 e1218.
- Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488, 471-475.
- Krey, J.F., and Dolmetsch, R.E. (2007). Molecular mechanisms of autism: a possible role for Ca²⁺ signaling. *Curr Opin Neurobiol* 17, 112-119.
- Krumm, N., O'Roak, B.J., Shendure, J., and Eichler, E.E. (2014). A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 37, 95-105.
- Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.-X., et al. (2015a). Excess of rare, inherited truncating mutations in autism. *Nature Genetics* 47, 582-588.
- Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.X., et al. (2015b). Excess of rare, inherited truncating mutations in autism. *Nat Genet* 47, 582-588.

- Langen, M., Durston, S., Kas, M.J., van Engeland, H., and Staal, W.G. (2011). The neurobiology of repetitive behavior: ...and men. *Neurosci Biobehav Rev* 35, 356-365.
- Levy, D., Ronemus, M., Yamrom, B., Lee, Y.-H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011a). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 70, 886-897.
- Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011b). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 70, 886-897.
- Liu, Z., Li, X., Zhang, J.-T., Cai, Y.-J., Cheng, T.-L., Cheng, C., Wang, Y., Zhang, C.-C., Nie, Y.-H., Chen, Z.-F., et al. (2016). Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. *Nature* 530, 98-102.
- Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24, 659-685.
- Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics* 82, 477-488.
- Matharu, N., Rattanasopha, S., Tamura, S., Maliskova, L., Wang, Y., Bernard, A., Hardin, A., Eckalbar, W.L., Vaisse, C., and Ahituv, N. (2019). CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science* 363.
- Mei, Y., Monteiro, P., Zhou, Y., Kim, J.A., Gao, X., Fu, Z., and Feng, G. (2016). Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature* 530, 481-484.

Newschaffer, C.J., Croen, L.A., Daniels, J., Giarelli, E., Grether, J.K., Levy, S.E., Mandell, D.S., Miller, L.A., Pinto-Martin, J., Reaven, J., et al. (2007). The epidemiology of autism spectrum disorders. *Annu Rev Public Health* 28, 235-258.

O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012a). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619-1622.

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. (2012b). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246-250.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. *Nature* 508, 207-214.

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, 1008-1021.

Peixoto, R.T., Wang, W., Croney, D.M., Kozorovitskiy, Y., and Sabatini, B.L. (2016). Early hyperactivity and precocious maturation of corticostriatal circuits in Shank3B mice. *Nature Neuroscience* 19, 716-724.

Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010a). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368-372.

- Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010b). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368-372.
- Poduri, A., Evrony, G.D., Cai, X., and Walsh, C.A. (2013). Somatic mutation, genomic variation, and neurological disease. *Science* 341, 1237758-1237758.
- Qureshi, A.Y., Mueller, S., Snyder, A.Z., Mukherjee, P., Berman, J.I., Roberts, T.P., Nagarajan, S.S., Spiro, J.E., Chung, W.K., Sherr, E.H., et al. (2014). Opposing brain differences in 16p11.2 deletion and duplication carriers. *J Neurosci* 34, 11199-11211.
- Robinson, E.B., Samocha, K.E., Kosmicki, J.A., McGrath, L., Neale, B.M., Perlis, R.H., and Daly, M.J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci U S A* 111, 15161-15165.
- Robinson, E.B., St Pourcain, B., Anttila, V., Kosmicki, J.A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K.E., Sanders, S.J., Ripke, S., et al. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* 48, 552-555.
- Ronan, J.L., Wu, W., and Crabtree, G.R. (2013). From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet* 14, 347-359.
- Ronemus, M., Iossifov, I., Levy, D., and Wigler, M. (2014a). The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 15, 133-141.
- Ronemus, M., Iossifov, I., Levy, D., and Wigler, M. (2014b). The role of de novo mutations in the genetics of autism spectrum disorders. *Nature reviews Genetics* 15, 133-141.

Rothwell, P.E., Fuccillo, M.V., Maxeiner, S., Hayton, S.J., Gokce, O., Lim, B.K., Fowler, S.C., Malenka, R.C., and Sudhof, T.C. (2014). Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors. *Cell* 158, 198-212.

Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 46, 944-950.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011a). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863-885.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011b). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863-885.

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., 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.

- Santini, E., Huynh, T.N., MacAskill, A.F., Carter, A.G., Pierre, P., Ruggero, D., Kaphzan, H., and Klann, E. (2013). Exaggerated translation causes synaptic and behavioural aberrations associated with autism. *Nature* 493, 411-415.
- Sharma, N., Mishra, R., and Mishra, D. (2015). The fifth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-5): what is new for the pediatrician? *Indian Pediatr* 52, 141-143.
- Shepherd, G.M. (2013). Corticostriatal connectivity and its role in disease. *Nat Rev Neurosci* 14, 278-291.
- Silverman, J.L., Yang, M., Lord, C., and Crawley, J.N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nature reviews Neuroscience* 11, 490-502.
- Simons Vip, C. (2012). Simons Variation in Individuals Project (Simons VIP): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron* 73, 1063-1067.
- Sparks, B.F., Friedman, S.D., Shaw, D.W., Aylward, E.H., Echelard, D., Artru, A.A., Maravilla, K.R., Giedd, J.N., Munson, J., Dawson, G., et al. (2002). Brain structural abnormalities in young children with autism spectrum disorder. *Neurology* 59, 184-192.
- Szigeti, B., Gleeson, P., Vella, M., Khayrulin, S., Palyanov, A., Hokanson, J., Currie, M., Cantarelli, M., Idili, G., and Larson, S. (2014). OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. *Frontiers in computational neuroscience* 8, 137.
- Sztainberg, Y., and Zoghbi, H.Y. (2016). Lessons learned from studying syndromic autism spectrum disorders. *Nat Neurosci* 19, 1408-1417.
- Talkowski, M.E., Minikel, E.V., and Gusella, J.F. (2014). Autism spectrum disorder genetics: diverse genes with diverse clinical outcomes. *Harv Rev Psychiatry* 22, 65-75.

- Toga, A.W., Clark, K.A., Thompson, P.M., Shattuck, D.W., and Van Horn, J.D. (2012). Mapping the Human Connectome. *Neurosurgery* 71, 1-5.
- Turner, T.N., Coe, B.P., Dickel, D.E., Hoekzema, K., Nelson, B.J., Zody, M.C., Kronenberg, Z.N., Hormozdiari, F., Raja, A., Pennacchio, L.A., et al. (2017). Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell* 171, 710-722 e712.
- Tye, K.M., and Deisseroth, K. (2012). Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nature reviews Neuroscience* 13, 251-266.
- van Bon, B.W., Coe, B.P., Bernier, R., Green, C., Gerdts, J., Witherspoon, K., Kleefstra, T., Willemsen, M.H., Kumar, R., Bosco, P., et al. (2016a). Disruptive de novo mutations of DYRK1A lead to a syndromic form of autism and ID. *Mol Psychiatry* 21, 126-132.
- van Bon, B.W.M., Coe, B.P., Bernier, R., Green, C., Gerdts, J., Witherspoon, K., Kleefstra, T., Willemsen, M.H., Kumar, R., Bosco, P., et al. (2016b). Disruptive de novo mutations of DYRK1A lead to a syndromic form of autism and ID. *Molecular psychiatry* 21, 126-132.
- Vogel-Ciernia, A., Matheos, D.P., Barrett, R.M., Kramar, E.A., Azzawi, S., Chen, Y., Magnan, C.N., Zeller, M., Sylvain, A., Haettig, J., et al. (2013). The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat Neurosci* 16, 552-561.
- 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, 380-384.
- Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M.A., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528-533.

- 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. (2013a). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997-1007.
- 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. (2013b). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997-1007.
- Zhao, X., Leotta, A., Kustanovich, V., Lajonchere, C., Geschwind, D.H., Law, K., Law, P., Qiu, S., Lord, C., Sebat, J., et al. (2007). A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A* 104, 12831-12836.
- Ziff, E.B. (1997). Enlightening the postsynaptic density. *Neuron* 19, 1163-1174.
- 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.
- LeDoux, J.E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci* 23, 155-184.
- Everitt, B.J., and Robbins, T.W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature neuroscience* 8, 1481.
- McHaffie, J.G., Stanford, T.R., Stein, B.E., Coizet, V., and Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends in neurosciences* 28, 401-407.
- Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2006). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168.

- DeLong, M.R., and Wichmann, T. (2007). Circuits and circuit disorders of the basal ganglia. *Arch Neurol* 64, 20-24.
- Goto, Y., and Grace, A.A. (2008). Limbic and cortical information processing in the nucleus accumbens. *Trends Neurosci* 31, 552-558.
- Utter, A.A., and Basso, M.A. (2008). The basal ganglia: an overview of circuits and function. *Neurosci Biobehav Rev* 32, 333-342.
- Losh, M., Adolphs, R., Poe, M.D., Couture, S., Penn, D., Baranek, G.T., and Piven, J. (2009). Neuropsychological Profile of Autism and the Broad Autism Phenotype. *JAMA Psychiatry* 66, 518-526.
- DeLong, M., and Wichmann, T. (2010). Changing views of basal ganglia circuits and circuit disorders. *Clin EEG Neurosci* 41, 61-67.
- 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.
- Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., Van De Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391-399.
- Price, J.L., and Drevets, W.C. (2012). Neural circuits underlying the pathophysiology of mood disorders. *Trends Cogn Sci* 16, 61-71.
- Tye, K.M., and Deisseroth, K. (2012). Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nat Rev Neurosci* 13, 251-266.
- Shepherd, G.M. (2013). Corticostriatal connectivity and its role in disease. *Nat Rev Neurosci* 14, 278-291.

Willsey, A.J., Sanders, Stephan J., Li, M., Dong, S., Tebbenkamp, Andrew T., Muhle, Rebecca A., Reilly, Steven K., Lin, L., Fertuzinhos, S., Miller, Jeremy A., et al. (2013). Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* 155, 997-1007.

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.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. *Nature* 508, 199-206.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. *Nature* 508, 207-214.

Rothwell, P.E., Fuccillo, M.V., Maxeiner, S., Hayton, S.J., Gokce, O., Lim, B.K., Fowler, S.C., Malenka, R.C., and Sudhof, T.C. (2014). Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors. *Cell* 158, 198-212.

Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., et al. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* 83, 309-323.

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.

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y.-H., Yamrom, B., and Wigler, M. (2015). Low load for disruptive mutations in autism genes and their biased transmission.

- Proceedings of the National Academy of Sciences of the United States of America 112, E5600-5607.
- Parikshak, N.N., Gandal, M.J., and Geschwind, D.H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet* 16, 441-458.
- Wimmer, R.D., Schmitt, L.I., Davidson, T.J., Nakajima, M., Deisseroth, K., and Halassa, M.M. (2015). Thalamic control of sensory selection in divided attention. *Nature* 526, 705-709.
- Fuccillo, M.V. (2016). Striatal circuits as a common node for autism pathophysiology. *Frontiers in neuroscience* 10, 27.
- 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.
- Peters, S.K., Dunlop, K., and Downar, J. (2016). Cortico-Striatal-Thalamic Loop Circuits of the Salience Network: A Central Pathway in Psychiatric Disease and Treatment. *Front Syst Neurosci* 10, 104.
- Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature Genetics* 49, 1458.
- Halassa, M.M., and Kastner, S. (2017). Thalamic functions in distributed cognitive control. *Nat Neurosci* 20, 1669-1679.
- Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al. (2017). Rare and low-frequency coding variants alter human adult height. *Nature* 542, 186.

- Feliciano, P., Daniels, A.M., Snyder, L.G., Beaumont, A., Camba, A., Esler, A., Gulsrud, A.G., Mason, A., Gutierrez, A., and Nicholson, A. (2018). SPARK: a US cohort of 50,000 families to accelerate autism research. *Neuron* 97, 488-493.
- Rikhye, R.V., Wimmer, R.D., and Halassa, M.M. (2018). Toward an Integrative Theory of Thalamic Function. *Annu Rev Neurosci* 41, 163-183.
- Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* 174, 1015-1030 e1016.
- Holiga, Š., Hipp, J.F., Chatham, C.H., Garces, P., Spooren, W., D'Ardhuy, X.L., Bertolino, A., Bouquet, C., Buitelaar, J.K., Bours, C., et al. (2019). Patients with autism spectrum disorders display reproducible functional connectivity alterations. *Science Translational Medicine* 11, eaat9223.
- van den Heuvel, M.P., and Sporns, O. (2019). A cross-disorder connectome landscape of brain dysconnectivity. *Nature Reviews Neuroscience* 20, 435-446.

Appendix A

Allen Mouse Brain Atlas Data Set

The atlas used *in situ hybridization* (ISH) methods to measure the gene expression of approximately 20,000 genes across the entire brain (Lein 2007). The processed data was obtained from: *api.brain-map.org*. For our analysis, we used expression energy (product of expression density and expression intensity) as a measure of expression activity. Expression values were summarized for each neural structure, and we used summarized expression values for the 213 structures in the Allen Reference Atlas for which connectivity measures are also available. Expression values were \log_2 transformed and quantile normalized.

Gene Data Sets

We used two data sets derived from three independent ASD cohorts: Autism Sequencing Consortium (ASC), SFARI Simons Simplex Collection (SSC), and SFARI Simons Foundation Powering Autism Research for Knowledge (SPARK) (Satterstrom 2020, Feliciano 2018). *De novo* mutations were determined through trio and quad familial sequencing. One study applied TADA on an aggregated set of mutations from ASC and SSC, among others, to prioritize ASD genes. Likewise, we applied a genic intolerance-based model (Iossifov 2015) on the mutations from SPARK to determine ASD-associated genes. Genes with *de novo* non-synonymous mutations observed in unaffected/neurotypical SPARK siblings were used as a reference gene set.

For the analysis that uses functionally ascertained genes, we sampled genes with similar levels of overall brain expression in the Allen Mouse Brain Atlas. Mean expression values were calculated across the entire brain and then converted to percentiles. ASD genes were then individually

matched by uniformly sampling genes within a set interval centered at the expression percentile for a given ASD gene.

Calculation of Structure Expression Biases

Expression biases were calculated for each neural structure by comparing ASD genes to either a reference gene set or to sampled genes with similar overall expression levels in the brain. Biases were calculated using gene expression level or gene expression specificity. Expression specificity was used to measure the specificity of the expression of a gene in a given neural structure relative to the overall expression of the same gene throughout the entire brain. We used the z-score of gene expression in a given structure standardized by the expression in the entire brain as expression specificity, shown below.

$$z_{g,s} = \frac{x_{g,s} - \mu_g}{\sigma_g}$$

$z_{g,s}$ = expression specificity of gene g in structure s

$x_{g,s}$ = expression level of gene g in structure s

μ_g, σ_g = expression mean, s.d. of gene g in all structures

To calculate neural structure expression biases, we compared ASD expression values to the expression of genes in the reference set. We defined an expression bias statistic in which ASD gene expression in a given neural structure is standardized by expression in the same structure for the reference gene set, shown below. Expression biases were calculated for both gene expression level and gene expression specificity.

$$z_{D,s} = \frac{\mu_{D,s} - \mu_{R,s}}{\sigma_{R,s}}$$

D = disease gene set

R = reference gene set

Similarly, to calculate neural structure expression biases, we also compared ASD expression values to the expression of matched genes. We defined an expression bias statistic in which ASD gene expression in a given neural structure is standardized by expression in the same structure

for the matched genes with similar overall expression in the entire brain, shown below. For each neural structure, the expression bias statistic is calculated for each ASD gene individually, then the mean statistic is taken across all ASD genes. Expression biases were calculated for both gene expression level and gene expression specificity.

$$z_{g,s} = \frac{x_{g,s} - \mu_{M,s}}{\sigma_{M,s}}$$

$$z_{D,s} = \frac{1}{|G|} \sum_{g \in G} z_{g,s}$$

$z_{g,s}$ = statistic for gene g in structure s

$z_{D,s}$ = expression bias for disease gene set D of structure s

Allen Mouse Brain Connectivity Atlas Data Set

The atlas used enhanced green fluorescent protein-expressing adeno-associated viral vectors to trace measure axonal projections throughout the brain (Oh 2014). The processed connectivity data was obtained from Supplementary Table 3. For our analysis, we used normalized projection density measurements, which are normalized for source and target structure volumes. Ipsilateral and contralateral connectivity measurements were combined by taking the pairwise maximum weights. Connectivity values were summarized to the same 213 neural structures in the Allen Reference Atlas structures used in the expression analysis. Connectivity weights are \log_{10} transformed and shifted such that all values were non-negative and the minimum weight was anchored at 0. Edges where no measurable connection occurs between neural structures or where the edges do not surpass a significance threshold of 0.05 for the linear regression model were coded as 0. Neural structures were partitioned into 12 high-level regions: isocortex, olfactory areas, hippocampal formation, cortical subplate, striatum, pallidum, thalamus, hypothalamus, midbrain, pons, medulla, and cerebellum.

Evaluation of the Connectivity Among Structures

We evaluated the interconnectivity of ASD structures using three tests: (1) network permutation test of connectivity, (2) internal-external paired test of cohesion, and (3) binomial test of cohesion. For the network permutation test, all edges in the global connectome were binarized, then pairs of edges were swapped in a random order such that the number of edges per vertex was consistent with that in the non-permuted connectome. We then compared the observed number of connections in the non-permuted connectome among ASD structures to the expected number of connections in permuted connectomes to estimate the significance of ASD connectivity.

For the internal-external paired test of cohesion, we compared internal connectivity weights among ASD structures to the weights between ASD structures and random external structures in the connectome. For each structure in a given set, we calculated the mean of the five strongest internal connectivity weights and mean of the five strongest external connectivity weights. For external weights, we sampled structures from outside of the set while maintaining the same number of structures as within the set, and calculated mean external weights for each random set. We then took the mean external weight across the random sets. We then used a Wilcoxon signed-rank test to compare the mean internal weights to mean external weights to determine significance of cohesion.

For the binomial test of cohesion, we compared the fraction of connectivity among ASD structures to the fraction of connectivity among a reference set of structures. For a given set of structures, we can determine the fraction of connectivity by calculating the number of connections between structures in the set, then dividing by the total connectivity of those same structures in the global connectome (see connectivity statistic in next section). This fraction is

the probability that structures are connected given the overall connectivity of the structures. The fraction for the reference structures is then used as the base probability for the 1-tail binomial exact test used to estimate significance of cohesion among ASD structures.

Search for Strongly Connected Circuits

We defined a connectivity statistic that is a measure of the cohesion or cohesiveness of a set of structures, shown below.

$$C_S = \frac{\sum_{s \in S} n_S}{\sum_{s \in S} n_T}$$

C_S = cohesiveness of structures S

n_S = number of connections between structure s and other structures among set S

n_T = number of connections between structure s and other structures in total connectome T

We used a greedy search algorithm to find the maximally cohesive subset of neural structures, shown below as pseudo-code. At each step of the search, one structure is removed such that the cohesiveness of the remaining structure is maximized. The algorithm iteratively removes each structure and recomputes cohesiveness to determine the structure to remove at each step. This process is repeated until two structures remain. We then select the subset of structures with the maximum cohesiveness statistic.

Initiate global connectome $G = \{V, E\}$ with neural structures as vertices V and connections as edges E .

Define a starting set of structures $S_1 = \{s_1, s_2, \dots, s_n\}$.

While $|S_i| > 2$:

For each structure s_j in set S_i :

Calculate connectivity statistic C_S for $S_i \setminus \{s_j\}$.

Set $S_{i+1} = S_i \setminus \{s_j\}$ where $j = \text{argmax}(C_S)$.

Set $i = i + 1$

Analysis of Distant Projections

To assign distance between neural structures, we calculated the correlation of gene expression between all pairs of structures. The correlations were then stratified into deciles, and the fraction of connections that occur between ASD structures was computed for each decile.