Title

Whole Exome Sequencing Analysis of Individuals with Autism Spectrum Disorder

Permalink

https://escholarship.org/uc/item/2p13j7d3

Author

Procko, Andrea Lynn

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, IRVINE

Whole Exome Sequencing Analysis of Individuals with Autism Spectrum Disorder

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Genetic Counseling

by

Andrea Lynn Procko

Thesis Committee:
Professor John Jay Gargus, MD, PhD, Chair
Professor Pamela Flodman. MSc, MS
Professor Moyra Smith, MD, PhD

2016

DEDICATION

For the families who participated in this study and for all the families who let me participate in their care, my sincerest gratitude.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I would like to thank many people for their assistance in this thesis, and I'll probably remember a few more after I recover from turning it in.

My committee members have been incredibly supportive in this work, and I can't thank them enough. Pam has given great feedback on this project and beyond, read every draft thoughtfully, and been an amazing role model for me as I become a genetic counselor who wants to continue to participate in research. Her turns of phrase and moral compass are something I will continually strive to emulate. Moyra has given me so much insight into how someone can delve into bioinformatics as well as into the complexity of genetics and autism. She has also graciously shared her computer and office with me. Jay has given constant encouragement and enthusiasm. I have also learned an immense amount by working with him in the clinic and will be forever grateful for that experience. Katie Singh also served as unofficial committee member, advising on early drafts and encouraging me when I needed it most. After reading many drafts including while on much-deserved vacations abroad and sharing in my enthusiasm for discovery, I couldn't have asked for a better group to support and challenge me throughout this process.

I had the opportunity to work on the CART research project before beginning this thesis, and the experience led to me enthusiastically jumping into uncharted territory. Thanks to all of the research team for training me in the ins and outs of clinical research coordination. The Order Sons of Italy in America, Renaissance Lodge, located in Orange, CA contributed funds, some of which were used to purchase the VarSeq software that was critical for my analysis, so they have my sincere gratitude. Also, Fabio Macciardi was willing to help a novice in the world of bioinformatics and encouraged me to think of the question first.

The faculty of the UCI GC program has been so incredibly amazing throughout my training through their recognition of my strengths and support while overcoming my weaknesses has set me on the path to a career that I love. And, to Lisa and the GCs at UCSD, who started me on the path and continue to advise me.

A special thank you to my classmates, who, instead of making me feel like an oddball, embraced me and made me a better person. Two years is a short time to develop such amazing relationships. I feel very fortunate. Through all the frustrations, you kept me laughing. I look forward to reuniting every year at NSGC to hear about the amazing GCs and people you will continue to become. And, hey, we didn't die!

To my friends who have understood that my nomadic lifestyle the past two years has decreased my time with them, but not my love for them. To my family, especially my parents, who have not seriously questioned how many degrees I need before I will grow up. To Carl and Rocky, I'm coming home!

ABSTRACT OF THE THESIS

Whole Exome Sequencing Analysis of Individuals with Autism Spectrum Disorder

By

Andrea Lynn Procko

Master of Science in Genetic Counseling

University of California, Irvine, 2016

Professor John Jay Gargus, MD, PhD, Chair

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and interaction and the presence of repetitive behaviors and or restricted interests. ASD has heterogeneous genetic and environmental etiologies, but for most individuals with ASD the genetic cause is unknown. Determining the genetic factors contributing to an individual's ASD phenotype is a goal in the clinic setting, as well as an important mechanism for understanding the broader pathology of ASD. As part of a larger research study, individuals with ASD and some family members had whole exome sequencing. For six probands with samples from both parents available, potential *de novo* variants were identified and assessed for their ability to contribute to the proband's ASD phenotype. Some variants were in genes previously associated with non-syndromic ASD, and variants in common pathways across different probands were identified, including those involved in axon guidance. For an additional proband an analysis of the chromosomal region identified as a small deletion by a clinical microarray was completed to assess if sequence level information

could contribute to the interpretation of this result. The analysis lessened the suspicion that this copy number variant contributed to the proband's ASD phenotype. Future analysis of this data set, especially inherited variants, will continue to identify biological processes that affect the complex ASD phenotype. As an emerging technology in the genetics clinic and as a research tool, whole exome sequencing will contribute to our understanding of ASD and identify genetic targets for future discovery, diagnostics, and therapeutics.

INTRODUCTION

AUTISM SPECTRUM DISORDER

The first descriptions of "autism" came in the 1940s from Hans Asperger, a pediatrician at University Children's Hospital in Vienna (Asperger, 1944), and Leo Kanner, a psychiatrist and founder of the first academic pediatric psychiatry department at Johns Hopkins University (Kanner, 1943). Even then, the two presented divergent views of this disorder, with Asperger describing a relatively common condition with individuals along a continuum of severity and Kanner describing a rare and more severe form of pediatric psychosis (Green, 2015).

The American Psychiatric Association (APA) first recognized "autism" as a distinct diagnosis and outlined diagnostic criteria in the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM) (APA, 1980). The current clinical definition of autism is outlined by the DSM-V [APPENDIX A]. The new definition combines four previously delineated conditions (including autism, Aspergers disorder, childhood disintegrative disorder, and pervasive developmental disorder not otherwise specified (PDD-NOS)). Instead, a single condition, Autism Spectrum Disorder (ASD), is described with varying presence and severity of symptoms for individuals on the spectrum (APA, 2013). Despite the range, common deficits in two core domains persist: (1) social communication and interaction and (2) restricted, repetitive behavior, interests, or activities (APA, 2013), and the diagnosis of ASD requires that both symptoms be present in the early developmental period. Individuals can present with or

without intellectual and language impairments. In addition, there are other medical comorbidities that can be associated with ASD, discussed further below.

In the United States of America, the prevalence of ASD is estimated at 1 in 68 (ADDMN, 2014). This estimate comes from eight-year olds at 11 sites across the US using a combination of evaluations performed by medical providers in the community and records of special education services in public schools. There is an increasing trend in the number of diagnoses in the US, summarized in Figure 1. Possible explanations for this trend include improved recognition and diagnosis as well as changes to the classification schema; however, it is not clear if there is also a true increase in the number of people affected (Newschaffer *et al.*, 2007). ASD occurs in all racial, ethnic, and socioeconomic groups, and it is approximately five times more common in males than females (ADDMN, 2014).
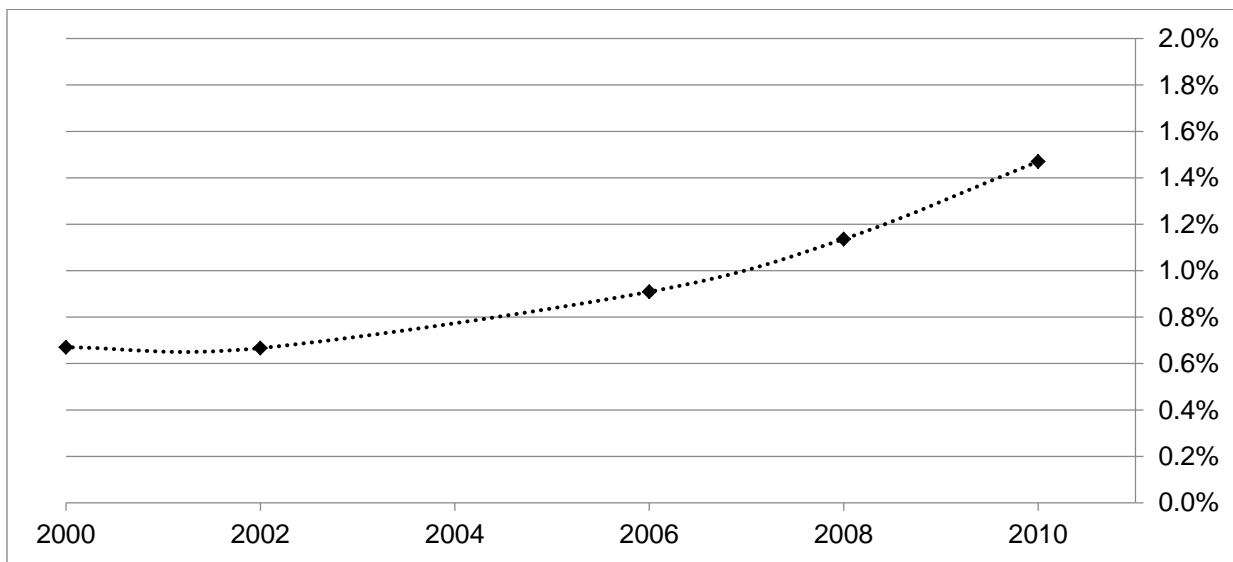
Figure 1: Prevalence of ASD in the United States. The prevalence of ASD has increased according to the CDC's ADDM Network (CDC, 2015).

2

Currently, there is no cure for ASD, and there is no pharmacological therapy for the core domains. There are limited interventions and symptomatic treatments available for individuals diagnosed with ASD. Fortunately, applied behavior analysis (ABA) therapy has been shown to improve communication and social skills as well as manage problem behavior for many individuals with ASD, but ABA is most effective when implemented early in development (Vismara, 2010).

CLINICAL PHENOTYPE OF ASD

CORE DOMAINS

The diagnosis of ASD requires deficits in the two core domains, as introduced above and described in the DSM-V (APA, 2013). While there is a cutoff for the severity of these symptoms to be diagnosed with ASD, it is set at a threshold for "clinically significant impairment", and the severity of symptoms for individuals can range significantly (APA, 2013).

Firstly, social communication and interaction are impacted, potentially including deficits in social-emotional reciprocity; nonverbal communicative behaviors; and developing, maintaining, and understanding relationships (APA, 2013). Examples of behaviors range within each category and can include reduced sharing of interests, abnormal eye contact, and difficulties making friends. More severely affected individuals may demonstrate a failure to initiate social interactions, an absence of facial expressions and non-verbal communication, and a lack of interest in peers.

The second core domain consists of restricted, repetitive patterns of behavior, interests, or activities (RRBs). An individual with ASD may display behaviors including lining up of objects, echolalia, rigid routines or patterns (with notable distress when not adhered to or during transitions), intense and focused interest, and hyper- or hyposensitivity to sensory input. Similarly, these can range in their severity and impact on an individual's ability to function in social and occupational realms.

COMORBIDITIES

In addition, there are characteristic comorbidities that affect some individuals with ASD. Regression of developmental milestones can be a red flag for additional screening; however, it is only noted in approximately 30% of individuals (Werner and Dawson, 2005). Many, but not all, have intellectual disability with an IQ less than 70 (Newschaffer *et al.*, 2007). Up to 60% of individuals with ASD also have intellectual disability (Vivanti *et al.*, 2013). Using the previous DSM subcategories, the prevalence of intellectual disability was highest in individuals with autistic disorder (66.7%) compared with individuals with PDD-NOS (12%) and Asperger disorder (0%), and the severity of intellectual disability seen also correlated with the severity of ASD subcategory (Chakrabarti and Fombonne, 2005). Some individuals have severe tantrums and/or self-injurious or outwardly aggressive behaviors (Baghdadli *et al.*, 2003; McClintock *et al.*, 2003). Sleep disorders are also common (Polimeni *et al.*, 2005; Richdale, 1999). Neurological findings can include seizures (in approximately 25% of individuals) (Spence and Schneider, 2009; Tuchman, 2006). Microcephaly or

macrocephaly is also seen in many cases (Fombonne *et al.*, 1999; Lainhart *et al.*, 1997; Miles *et al.*, 2000).

DIAGNOSIS OF ASD

The diagnosis of ASD is exclusively clinical and requires a detailed assessment of an individual's behavior and development. The process of diagnosis is divided into two steps: (1) developmental screening and (2) comprehensive diagnostic evaluation (Council on Children With Disabilities, 2006). Developmental screening is a normal part of any child's assessment during well-child visits with a pediatrician, including interviewing parents/caregivers or interacting with the child to assess if there are delays in developmental milestones. The American Academy of Pediatrics recommends that developmental surveillance happen at all visits in addition to standardized screening at least during the 9, 18 and 30 month check-ups (Council on Children With Disabilities, 2006). If concerns are raised by this screen, diagnostic tools can be used to further evaluate the child's behavior and development and are often administered by a specialist: potentially a developmental pediatrician, pediatric neurologist, child psychologist or psychiatrist, geneticist, or early-intervention program.

Different diagnostic tools, models, and behavioral scales exist to help standardize the assessment. They can include observation of patients or interviews with their parent(s). One example, the Autism Diagnostic Observation Schedule (ADOS), consists of a semi-structured, standardized assessment evaluating social interaction, communication, play, and imaginative use of materials (Lord *et al.*, 2000; Lord *et al.*, 1989). There have been

multiple versions of the ADOS over time, with improvements designed to address difficulties assessing individuals with different abilities in expressive language. The ADOS-II, the current version and the one used in this study, has four modules designed to match the verbal and communication abilities of the testee (Lord *et al.*, 2012). A quantitative score is assigned to the assessment, as well as sub-scores for social communication and restricted, repetitive behaviors. Classification of "Autism" or "Autism Spectrum" is given if the score exceeds the set thresholds and "Non-ASD" if it does not.

There may be evidence of ASD noted by parents or medical professionals in a child 18 months old or younger. Diagnosis at age 2 years old can be considered robustly reliable; however, if discordant, later evaluations are more likely to show a decrease in the severity of symptoms than increase and potentially remove the diagnosis (Kleinman *et al.*, 2008; Lord *et al.*, 2006). Still, many children don't get a final diagnosis until they are significantly older. This delay means children with ASD may not get the early intervention they need to achieve their personal best (Filipek *et al.*, 2000).

ETIOLOGY OF ASD

Despite the successful effort to standardize the diagnosis of ASD, an immense amount of clinical heterogeneity exists within the diagnosis, reflecting the variety of possible etiologies. It is well documented that there is a genetic contribution to the etiology of ASD. Even Kanner noted that some of the traits seen in his observations were noticeable in the parents as well (1943); however, this was initially misinterpreted for a strictly environmental influence (Green, 2015). The inheritance pattern of ASD,

however, is typically not a Mendelian, single-gene mode. Instead, a complex inheritance pattern is hypothesized for most individuals with contributions from multiple genes, as well as some environmental influences. In fact as of December 2015, the Human Gene Module of SFARI Gene, a publicly available database of human genes with published associations with ASD, lists 791 genes, although different genes and their associated polymorphisms vary widely in the strength of their evidence and penetrance (Banerjee-Basu and Packer, 2010; SFARI, 2015).

Evidence for genetic links to ASD includes the relative risk in families seen in twin studies and sibling recurrence risks. In British twin studies, monozygotic twins were found to be concordant for their ASD diagnosis (60-92%) significantly more frequently than dizygotic pairs (up to 10%) (Bailey *et al.*, 1995; Colvert *et al.*, 2015; Folstein and Rutter, 1977). Recent estimates for the heritability of ASD are approximately 50% (Gaugler *et al.*, 2014). For ASD with no specific known etiology, the empiric recurrence risk for a future sibling is elevated, but the specific risk ranges from approximately 5-20% in different studies (Constantino *et al.*, 2010; Gronborg *et al.*, 2013; Ozonoff *et al.*, 2011) and depends on factors including the sex of the proband (Sumi *et al.*).

In addition, individuals with specific genetic syndromes are at increased risk for having ASD. The most common single-gene disorder associated with ASD is fragile X syndrome (FXS) (Coffee *et al.*, 2009). FXS is caused by expansions of an unstable CGG-repeat sequence (≥200 repeats) in the promoter region of the *FMR1* gene on the X chromosome which causes gene expression to be silenced. Individuals with FXS can

have intellectual disability, characteristic facial features, and behavioral issues. Both males and females can be affected; however, the phenotype is less penetrant and often less severe in women, due to a combination of their heterozygous state and X-chromosome inactivation. Approximately 20-25% of individuals with FXS also have a diagnosis of ASD, with a sex-ratio (4 males:1 female) similar to that seen in all individuals with ASD (Hatton *et al.*, 2006).

Individuals with Tuberous Sclerosis (TSC) have non-malignant growths in different organs of their body, including the brain, eyes, skin, and heart, as well as others. Most individuals with TSC will have seizures, and many have intellectual disability, developmental delay, and sometimes ASD (a large range reflecting the clinical variability seen in TSC itself, up to approximately 50-60%) (Baker *et al.*, 1998; Jeste *et al.*, 2008). Mutations in two genes, *TSC1* and *TSC2* are causative of TSC; however, mutations in *TSC2* are more likely to be associated with ASD (Numis *et al.*, 2011). Interestingly, males and females with TSC are equally likely to have ASD (Curatolo *et al.*, 2004; Vignoli *et al.*, 2015).

In addition to FXS and TSC, there are many other well-characterized genetic conditions which confer an increased risk of ASD compared to the general population, including Down syndrome, Rett syndrome, and *PTEN* Hamartoma Tumor syndrome (Cohen *et al.*, 2005; Muhle *et al.*, 2004). Some microdeletion and microduplication syndromes, including 15q11-13 duplication of the maternal allele in Prader Willi syndrome and microdeletion/duplication of 16p11.2, have an increased association with ASD as well

(Baker *et al.*, 1994; Fernandez *et al.*, 2010; Weiss *et al.*, 2008). ASD can also occur as a consequence of inborn errors of metabolism, including phenylketonuria, Smith-Lemli-Opitz syndrome, and mitochondrial disease (Gargus and Imtiaz, 2008; Lombard, 1998; Palmieri and Persico, 2010; Pons *et al.*, 2004; Zecavati and Spence, 2009). While these well-studied disorders are incredibly powerful for understanding some cases of ASD, such syndromic forms combined account for less than 30% of ASD cases, and the penetrance of the ASD phenotype varies widely among them (Cohen *et al.*, 2005; Muhle *et al.*, 2004; Schaaf and Zoghbi, 2011).

By contrast, non-syndromic ASD, sometimes called idiopathic, describes cases where ASD is the primary diagnosis, not secondary to an existing condition or diagnosis with a well-known genetic etiology as described above. While traditionally ASD is described as not having defining physical features, within this subgroup of non-syndromic ASD approximately 20% of individuals have significant dysmorphology, sometimes distinguished as "complex autism" in contrast to "essential autism" (Miles *et al.*, 2005). However, there is still significant heterogeneity within these two groups.

For individuals with non-syndromic forms of ASD, there are a growing number of identified genomic regions and specific genes for which there are known variants associated with susceptibility to developing ASD. These have been identified through many studies, especially through studying copy number variations (CNV) as well as single nucleotide variations in large samples of individuals with ASD. Overall, these mutations can be categorized along a spectrum from common variants with relatively

small effect sizes to rare variants with larger effect (Anney *et al.*, 2012; De Rubeis and Buxbaum, 2015; Gaugler *et al.*, 2014; Veenstra-Vanderweele *et al.*, 2004). Rare variants can be inherited from unaffected parents or arise *de novo* in an offspring (De Rubeis *et al.*, 2014; Iossifov *et al.*, 2014; Krumm *et al.*, 2014). Many of the CNVs identified are seen more frequently in individuals with ASD but also present in controls (Sebat *et al.*, 2007; Zhao *et al.*, 2007). Studies looking at changes in sequence-level variation (single nucleotide substitutions or small insertions and deletions) show that individuals with ASD have significantly more *de novo* mutations than their unaffected siblings, especially those predicted to cause a loss of function (0.13-0.21 per individual with ASD compared to 0.09-0.12 for their siblings) (Iossifov *et al.*, 2014; Samocha *et al.*, 2014). This wide array of etiologies which are not mutually exclusive for any given individual with ASD and the lack of penetrance in unaffected individuals make it a challenge to advance past idiopathic to a molecular diagnosis, despite the well documented genetic contribution to ASD.

ASD IN THE GENETICS CLINIC

ASD remains a common referral indication in a pediatric genetics clinic, as well as a discussion point in many genetic counseling sessions when reviewing family history information. A complete clinical genetics evaluation is recommended to assist in determining the diagnostic etiology, including three-generation pedigree and clinical examination for dysmorphic features, and if indicated, metabolic/mitochondrial testing, brain magnetic resonance imaging (MRI) or other testing (Schaefer and Mendelsohn, 2013). If indicated, testing for specific syndromic etiologies can be recommended. For

example, Fragile X testing should be considered in any individual with intellectual disability, developmental delay or ASD that is consistent with X-linked inheritance (Sherman *et al.*, 2005). Importantly, this requires an assay that can distinguish the number of trinucleotide repeats, which is not a capability of most genome-wide genetic testing. *MECP2* sequencing for Rett syndrome is indicated in all females with ASD, and duplication testing in males, if they have a suggestive phenotype (Schaefer and Mendelsohn, 2013). However, since the large majority of individuals with ASD will fall into the non-syndromic category, less targeted, genome-wide investigations (e.g. chromosome analysis via karyotype, chromosome microarray, whole exome sequencing) may discover genetic etiologies for some of the remaining individuals.

Cytogenetic chromosome analysis, also referred to as a karyotype, involves looking at the chromosomes underneath a microscope to identify large scale deletions, duplications, or rearrangements. Diagnostic yields have increased with continuously improving cytogenetic technology and resulting increasing resolution, but the yield is still approximately 3% (Reddy, 2005; Shen *et al.*, 2010; Shevell *et al.*, 2001).

Chromosomal microarrays (CMA), first array-comparative genomic hybridization and now including single-nucleotide polymorphism (SNP) arrays, allow submicroscopic resolution for deletions and duplications (also called copy number variations). SNP arrays also detect runs of homozygosity that are longer than expected, which can be helpful for identifying regions of interest for recessive conditions or imprinting disorders. The diagnostic yield for CMA varies in different studies, partially because of the

continuing improvement of array resolution. Amalgamating recent studies estimates a 10% diagnostic yield for an unselected population of individuals with ASD and almost 30% for individuals who also have seizures, physical findings (e.g. microcephaly, congenital anomalies, or other dysmorphic features), or family history of developmental disabilities or psychiatric problems (Schaefer and Mendelsohn, 2013).

After studies showed the improved diagnostic yield for individuals with non-syndromic ASD when using CMA over karyotype (Shen *et al.*, 2010), the American College of Medical Genetics recommended offering CMA as first-tier testing for individuals with ASD, as well as those with developmental delay, intellectual impairment, and multiple congenital anomalies (Miller *et al.*, 2010). Karyotype is now primarily reserved for individuals for whom there is a high suspicion for aneuploidy or an inherited chromosomal rearrangement (e.g. Turner syndrome or a translocation).

EMERGING GENETIC TESTING FOR ASD

As new genetic testing technology becomes clinically available, its utility in clinical testing for ASD diagnosis and susceptibility will inevitably be assessed. This has recently occurred with the development of massively-parallel sequencing (also known as next-generation sequencing) and the availability of clinical exome sequencing (or whole-exome sequencing, abbreviated WES). Massively-parallel sequencing involves taking an individual's DNA to generate a library of DNA fragments by clonal amplification, sequencing the fragments using chemistry that can produce a fluorescent signal (e.g. pyrosequencing or reversible dye terminators), and assembling the

generated reads bioinformatically based on a reference sequence to identify changes in the base pair sequence (including substitutions and depending on the technology, certain sizes of insertions and deletions). The regions interrogated by the analysis are determined by an initial targeted-capture step, which can allow for the analysis of genes of interest (e.g. clinically available panel testing for conditions where multiple genes may be clinically relevant) or a wider analysis of all the protein-coding regions using WES.

Since 2011, clinical WES has been available from multiple clinical testing laboratories (Farwell *et al.*, 2015; Lee *et al.*; Yang *et al.*). One demonstrated a detection rate of approximately 20% for the first 49 samples submitted with a primary indication of ASD (Farwell *et al.*, 2015); however, this may reflect a bias in the first cases submitted after the technology became available being more likely to get a diagnosis. Because of the large number of variants that are identified using WES, providing samples from other family members (most commonly a "trio" with the proband and both biological parents) is encouraged and does increase the diagnostic yield significantly, from 21-22% to 31-37% across all referral indications in two independent studies (Farwell *et al.*, 2015; Lee *et al.*, 2014).

A recent study from the Canadian province of Newfoundland and Labrador took a consecutive population of individuals diagnosed with ASD and compared the diagnostic yields of CMA and WES (Tammimies *et al.*, 2015). For the 258 probands in the study, all had CMA, while a randomly-selected subset of 95 proband-parent trios also had WES. The overall diagnostic yield for CMA was 9.3%, consistent with other studies,

and 8.4% for WES. Interestingly, of the 95 participants who received both, 2% received diagnoses from both methodologies, reiterating their complementary nature. Since the yields were similar, some argue that if future studies support this finding, a recommendation for WES as an additional first-line test will be logical (Miles, 2015), especially for children with more complex morphological phenotypes who had the highest yield (24.5% for CMA and 16.7% for WES) (Tammimies *et al.*, 2015). However, there remain questions about the relative causative nature of many of the genetic changes identified by these genome-wide genetic testing technologies as well as their clinical utility.

Eventually, the menu of testing options from clinical laboratories will expand. Clinical whole genome sequencing (using massively parallel sequencing techniques) is already available, though not widely utilized in genetic clinics. Some experts predict that regulatory regions that affect transcription quantitatively, like promoters and enhancers, will be important to understanding ASD (Stein *et al.*, 2013). Other technologies will also likely emerge to address the limitations of current technologies. In order for these tests to be utilized efficiently, accurately, and ethically in the future, a concerted effort to understand the etiologies for individuals with ASD must be made.

THE IMPORTANCE OF DETERMINING ETIOLOGY

While the criteria and diagnostic tools allow for a systematic diagnosis of ASD in affected individuals and genetic testing can help identify the etiology for some, the underlying cause of ASD is not determined for a large proportion of individuals

diagnosed (Schaefer and Mendelsohn, 2013). There are several important reasons why determining the genetic and molecular etiology of an individual's autistic traits may be valuable.

For the patient and patient's family, the benefits of early diagnosis and the ability to prevent or screen for any associated medical risks can be a significant impetus for identifying the cause. Early detection, of course, allows for early intervention, which can improve outcomes (Dawson and Bernier, 2013). Even in the absence of this potential for improved outcomes, families may benefit from the "end to the diagnostic odyssey" that comes with an answer (Lenhard *et al.*, 2005; Makela *et al.*, 2009), and there are certainly benefits, including medical, psychological, and financial, to ending additional diagnostic testing once an explanation is found. For some individuals, their diagnosis may also raise additional medical concerns and associated changes to medical management. For others, the information will not impact management, but may still provide benefit. In a study looking at the perceptions of parents receiving CMA results, twelve of eighteen parents whose child with ASD received a negative result (with no informative changes identified for their child) found the results at least "moderately helpful" (Reiff *et al.*, 2015).

A genetic diagnosis also allows for a refinement of the genetic counseling offered to the patient and patient's family. This includes a recurrence risk beyond a general multifactorial explanation with a generic empiric recurrence risk. A molecular diagnosis will often allow for targeted testing for family members as well, including prenatal

testing. However, the reduced penetrance and potential range of severity make any predictive testing complicated. Still when surveyed, most parents of a child with ASD (80%) would test younger siblings pre-symptomatically to identify potential risk (Narcisa *et al.*, 2013).

Ethical concerns for clinical genetic testing for ASD exist, especially presymptomatically (including prenatally) given incomplete penetrance and variable expressivity of most genetic contributions. Also, the delay between identification of such risk factors using genetic testing and the development of clinical utility of those results through proven treatments or interventions is challenging (McMahon, 2006). However, clinical genetic testing laboratories are already offering clinical tests for genes associated with increased risks of non-syndromic forms of ASD, so preparing for the counseling that families need to understand this complexity and ambiguity will be a clinical challenge until the benefits of these discoveries are translated into treatment.

From a research standpoint, further understanding of the genetic causes of ASD may provide opportunities to develop tools for additional investigation (including cell lines and model organisms) and targets for diagnostic tests and treatments. A refined diagnosis also allows for any clinical investigations of interventions and treatments to be assessed in more homogenous populations, eliminating some confounding variables. An analogous trajectory has been followed in the understanding of cancer which was initially thought of as a single, monolithic condition before being divided by the source tissue of the tumor (e.g. breast or colon cancer), and is now thought of with increasing

molecular resolution in terms of diagnosis, prognosis, and treatment (e.g. PARP inhibitors used for chemotherapy for individuals with germline mutations in *BRCA* pathway genes) (O'Connor, 2015). While the intricate neurodevelopmental processes involved in ASD will make this analysis complex, eventually improved understanding will enable many individuals with ASD to have their condition categorized by its molecular underpinnings.

SUMMARY OF RESEARCH PROGRAM

This thesis is part of an already existing research program at University of California, Irvine, entitled "Neurobiology and Genetics of Autism", approved by the Institutional Review Board (HS#1996-616).

In its current format, the study enrolls participants who are recruited by UCI clinical faculty from their clinics, referred to the study by other clinicians, or self-referred. Most of these participants have a clinical diagnosis of ASD and some have a previously identified genetic diagnosis or abnormal genetic test result (e.g. CNV identified by clinical microarray). To broadly characterize the phenotype of the individuals who choose to participate, several assessments are completed and biological samples are obtained during their visit(s). Assessments include an ADOS and IQ test, an EEG, and an assessment of sleep quality. Biological samples include saliva and breath samples as well as blood and urine for routine metabolic testing. The blood is also used to extract DNA for whole exome/genome sequencing, and parents or siblings who consent to the protocol provide a blood sample as well in order to improve the interpretation.

After the participant's visit, a medical record abstraction is completed and a family history is obtained. In total, this allows for a thorough description of each participant's ASD phenotype to be correlated with the genotype findings.

AIMS OF THIS STUDY

Specifically for this thesis, an analysis of the initial whole exome sequences completed so far will be the focus:

- The first aim of this study is an unbiased analysis of sequence data from probands for whom a trio is available to identify *de novo* changes (e.g. predicted heterozygous dominant autosomal changes in males and females and hemizygous X-linked changes in males). These are then assessed to identify mutations with potential impact on participants' ASD phenotypes.

- The second aim is an analysis of a specific participant, AU0237-0201, for whom there is a deletion of a region on chromosome 3 that was identified by a clinical microarray ordered for the proband.

This combination of unbiased and targeted analyses will build a foundation for a dataset that will continue to be a resource as more of the phenotypic information is compiled and further analyses of this extensive genomic data are endeavored.

METHODS

NEUROBIOLOGY AND GENETICS OF AUTISM STUDY

Between the dates of July 2014 and March 2015, twenty-five probands and their available parents were consented and enrolled into the University of California, Irvine's Autism Research Project (UC Irvine IRB# 1996-616). The families were referred from various sources, including the original cohort of previous participants and new referrals by faculty physicians. For those families who were willing, the study was described to them in person or over the phone, and consent documents were sent in advance for review. If the family showed continued interest, they were scheduled for a visit at the Institute for Clinical and Translational Science on UC Irvine's campus.

A unique identification (ID) number was assigned to each participating individual. This consists of a unique family ID, followed by a number for each individual within the family. For example, in family AU0002, the proband is assigned AU0002-0201 as the first born child of father (AU0002-0101) and mother (AU0002-0102). These anonymous IDs were used to identify all samples collected.

For most participants, a daylong visit was organized to allow for an in-person consent process [APPENDIX B], collection of biological samples (blood from proband and available parents, urine, breath, and saliva), and phenotype assessments including ADOS, IQ test, electroencephalography (EEG), and sleep assessment (including questionnaire and actigraph readings taken at home). IQ was assessed using either the

Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II) or Mullen Scales of Early Learning (MSEL). WASI-II is an abbreviated, standard IQ measurement tool which provides a normalized IQ score as well as subscores in verbal comprehension and perceptual reasoning (Wechsler, 2011) and was used for most probands ages 6 and above. MSEL focuses on early development (including visual reception, fine motor skills, receptive language and expressive language) to calculate a normalized Early Learning Composite score (Mullen, 1995). MSEL was used for younger probands and those with speech delays for whom the WASI-II was not appropriate. Both normalized scores have a mean of 100 and a standard deviation of 15; score below 70 indicates intellectual disability (50-69 mild, 35-49 moderate, 20-34 severe, and below 20 profound) (Tylenda et al., 2007).

The families selected to have whole exome sequencing (WES) in the first batch were chosen by the research team based on criteria including preferences for trios and absence of previous WES done for proband. For WES analysis, DNA was extracted from blood samples using standard techniques and sent to the Broad Institute for exome sequencing. Standard protocols for exon capture, amplification, and sequencing were used (as in Samocha et al., 2014). Sequence reads (mean length of 76bp across all samples) were processed and mapped to the reference genome (hg19/GRCh37) using Picard and Burrows-Wheeler Aligner (DePristo et al., 2011; Li and Durbin, 2010). Single nucleotide polymorphisms and small insertions and deletions were called using Genome Analysis Tool Kit (GATK), including calculating GATK Variant Quality Score Recalibration (VQSR) (DePristo et al., 2011; McKenna et al., 2010).

The Broad Institute provided individual Binary Alignment/Map (BAM) files and a variant call format (VCF) file with all of the samples combined. Individual VCF files were created using Galaxy software (usegalaxy.org; Afgan *et al.*, 2016). The variants identified were then filtered in VarSeq software (Golden Helix, 2016) as described below.

AIM I: FILTERING AND CHARACTERIZATION OF POTENTIAL *DE NOVO* VARIANTS

The first aim of this study is an analysis of all the variants identified for probands for which a trio is available, looking for *de novo* changes (e.g. predicted heterozygous dominant autosomal changes in males and females and hemizygous X-linked changes in males) for mutations with potential impact on participants' ASD phenotypes.

Using VarSeq software, a workflow for trios was developed to identify plausible *de novo* variants (Golden Helix, 2016). The quality threshold of the variant required more than 10 reads and a genotype quality (GQ) score of more than twenty. The minor allele frequency (MAF) in the population was required to be less than 0.01; although, if the MAF was not known, then the variant was still included. The effect of the variant had to be either loss of function (including frameshift mutations and mutations affecting splicing) or missense. Finally, the variant had to be present in the proband's genotype but absent from both parents' determined genotypes.

Variants were further filtered by criteria to prioritize variants with a higher likelihood of being accurate as well as potentially interesting *de novo* variants. Variants in the mucin family genes (*MUC*) were excluded because their evolutionary history of recent gene duplication (Chaturvedi *et al.*, 2008) can increase the chance of false-positives when reads are erroneously mapped to a paralog. By contrast, variants in the human leukocyte antigen (*HLA*) gene family on chromosome 6 were excluded because variation at this locus is very common and inherent to the immune system function of these genes. These exclusions were therefore practical, and not meant to assert that these genes have no role in ASD phenotypes. In fact, non-antigen presenting portions of the HLA region have been proposed as contributing to ASD for some individuals (Torres *et al.*, 2012), and a subsequent analysis of these variants could be a future research goal.

Because a germline heterozygous mutation would be expected to have an allele frequency of fifty percent, those variants found at an alternate allele frequency of more than twenty-five percent were prioritized and those with less were excluded. Variants with lower allele frequencies may represent errors or mosaicism in the individual, possibly due to somatic mutations. These may contribute to the ASD phenotype of an individual; however, their effect is more difficult to predict, especially without determining the level of mosaicism in different tissues.

The Variant Quality Score Recalibration (VQSR) filter included in the VCF file processing was used to identify variants that passed the filter ("PASS") (McKenna *et al.*,

2010). This filter gives a comprehensive score to several quality metrics including the number of reads covering a variant, the sequence surrounding the variant, and the number of reads in forward vs. reverse orientation. Restricting this initial analysis to only "PASS" variants maximized specificity.

While the VarSeq software's *de novo* workflow had a restriction of minor allele frequency less than 0.01, many of the variants were not assigned a frequency by the software, and upon inspection, some were not consistent with available databases. Therefore, the MAF for remaining variants was confirmed by consulting databases, including the Exome Aggregation Consortium (ExAC) database, which compiles exome data from 60,706 unrelated individuals from various large scale studies, including population genetic studies and disease-specific databases (Lek *et al.*, 2016). Pediatric disease subjects are excluded. Those variants with MAFs above 0.05 were excluded assuming that most *de novo* variants would be rarer in the population than that threshold.

Finally, the remaining variants were assessed to determine the quality of the variant as well as the likelihood that it influences the ASD phenotype of the proband. The predicted effect of the variant on the protein was estimated using available tools, including (i) modeling estimates of the effect of the variant on the protein, (ii) the location of the variant in relation to the functional domains of the protein, (iii) the expression pattern of the protein within human tissues, and (iv) the mutation constraint score for the gene.

(i) The Condel and Provean algorithms model the severity of the effect of a specific variant on the protein combining various other tools that factor in conservation of the affected amino acid in homologs and known protein domains and assigning either a "neutral" or "deleterious" effect (Choi *et al.*, 2012; Gonzalez-Perez and Lopez-Bigas, 2011). Condel was used for all missense mutations, and Provean was used for all deletions or insertions of amino acids, since this is not a function of Condel.

(ii) The location of the variant within the known or predicted functional domains of the protein was assessed using the database Uniprot (Uniprot Consortium, 2015) or other primary literature about the protein structure.

(iii) The expression pattern of mRNA and protein for the genes was also assessed (i.e. presence in brain tissues, especially embryonically) (Uhlen *et al.*, 2015).

(iv) The mutation constraint assigns a z-score representing the difference between the observed and expected number of mutations for synonymous, missense, and loss-of-function mutations for a given gene (Lek *et al.*, 2016).

The biological function, if predicted or known, of the gene affected was researched using databases, including PubMed (ncbi.nlm.nih.gov/pubmed), GeneCards (genecards.org), and Online Mendelian Inheritance in Man (OMIM; omim.org).

Some additional variants were removed for quality concerns or lack of evidence for their ability to contribute to an individual's ASD phenotype. For example for genes with no known or predicted function, the inclusion of the variant in the priority list depended on if

the protein was seen in brain tissues in the Human Protein Atlas database (Uhlen *et al.*, 2015). However, this assessment can be revisited in the future as more is known about those genes' functions because some may influence brain development and ASD phenotypes from afar. The variants removed by this filter ("Other") as well as the prioritized variants for each proband are detailed more in the "Results" section.

AIM II: TARGETED ANALYSIS OF PROBAND BASED ON CLINICAL FINDINGS

The second aim of this study is a targeted analysis looking for variants for proband AU0237-0201. Samples for this proband and her mother (AU0237-0102) were submitted for analysis as a "duo". A clinical microarray was previously ordered for this proband by her clinician and showed an approximately 371 kb loss on chromosome 3 at q25.32 (arr [hg19] 3q25.32 (158,405,216-158,776,705) x1). Given this result, any mutations identified in this region by the research whole exome sequencing would be hemizygous. Therefore, the variants identified in this region were further characterized for the possibility that they contribute to this proband's ASD phenotype.

The coordinates from the clinical microarray were used as a filter in the VarSeq program to identify variants for the proband and her mother in the region of this deletion (Golden Helix, 2016). Those variants that were not present in the proband or her mother were removed. The remaining variants are described in the Results section.

RESULTS

AIM I: FILTERING AND CHARACTERIZATION OF POTENTIAL *DE NOVO* VARIANTS

DEMOGRAPHICS OF PARTICIPANTS

For the *de novo* analysis, the six probands that were submitted for sequencing as part of a trio including samples from their parents were included. The demographic information, as well autism and IQ assessments done as part of the study, of these probands is summarized in **Table 1**.

PROBAND ID #	SEX	AGE	ADOS module	ADOS classification	IQ score (and subscores) [1]WASI-II or [2]Mullen
AU0002-0201	Male	26y 6m	1	Autism	79[1] (PR: 101, VC: 63)
AU0005-0202	Female	25y 1m	2	Autism	67[1] (PR: 86, VC: 49)
AU0238-0201	Male	5y 5m	1	Autism	≤49[2] (VR: 20, FM: <20, RL: 21, EL: <20)
AU0243-0201	Male	2y 10m	1	Autism	70[2] (VR: 31, FM: 38, RL: 35, EL: 31)
AU0245-0201	Female	20y 10m	4	Autism	119[1] (PR: 121, VC: 112)
AU0245-0202	Female	18y 4m	4	Autism spectrum	111[1] (PR: 102, VC:116)

Table 1: Demographic information of probands included in *de novo* analysis. [1]WASI-II, Full-Scale 4 Score, which incorporates all four subtests of the assessment. Subscores for Perceptual Reasoning (PR) and Verbal Comprehension (VC) are given below (Wechsler, 2011). [2]Mullen Scales of Early Learning, Early Learning Composite Score (Mullen, 1995). The T-score categories for Visual Reception (VR), Fine Motor (FM), Receptive Language (RL), and Expressive language (EL) are given below. All scores are normalized to have a mean of 100 and standard deviation of 15 in the general population, except the T-scores of the Mullen, which have a mean of 50 and a standard deviation of 10.

VARIANT FILTERING

For all of the probands that were submitted as part of a trio, the variants were filtered to prioritize those that were more likely to be accurately *de novo* for the proband as well as those that were more likely to contribute to the proband's ASD phenotype. **Figure 2**

summarizes the workflow used to prioritize variants, and **Table 2** shows the number of variants that were excluded at each step across all six probands. For each proband, the software assesses the variants that are identified in any of the samples. Therefore, the filters up until VarSeq's filter for *de novo* inheritance (present in proband genotype, absent in parent genotypes) includes variants that were not identified for that specific individual.
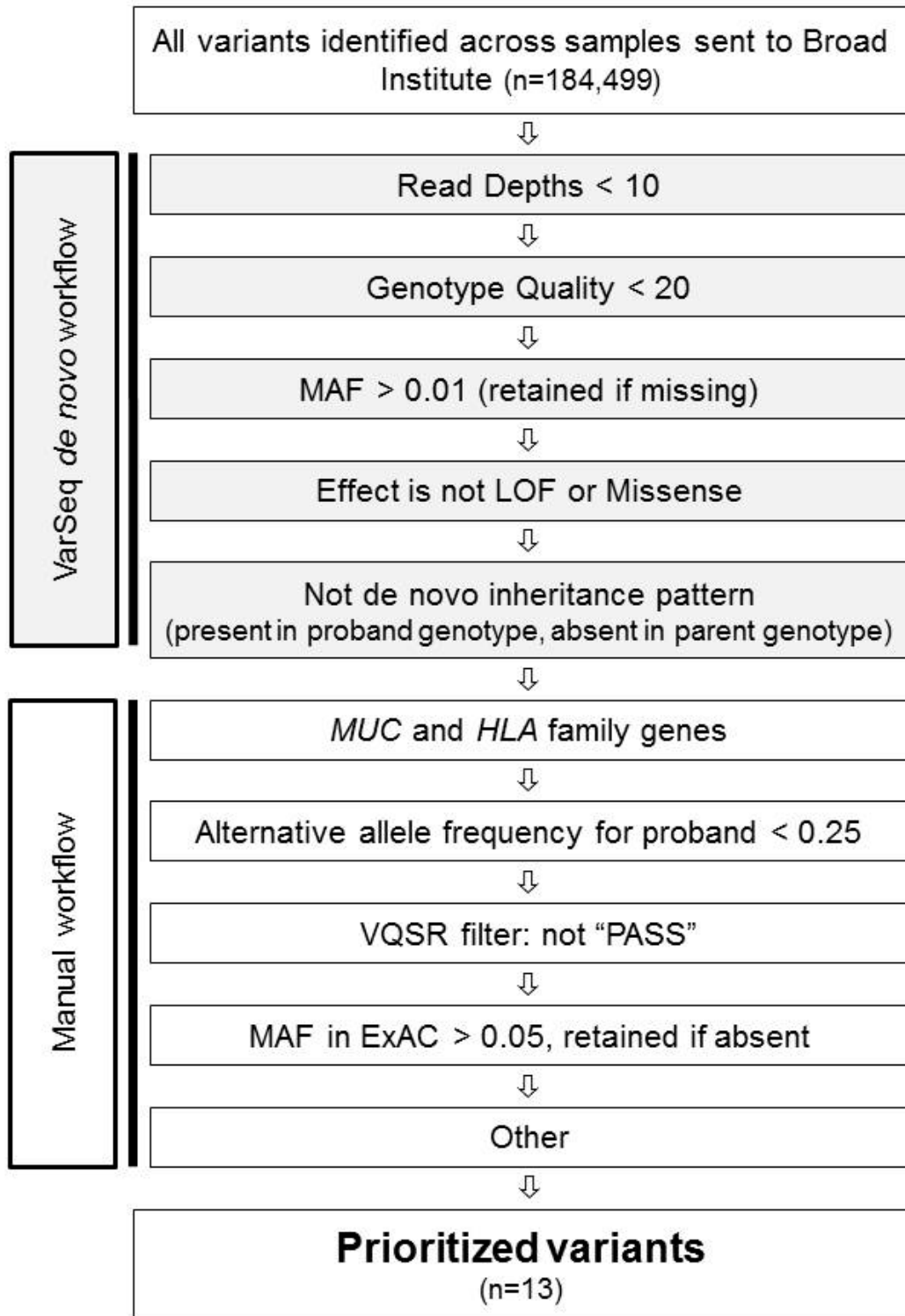
FIGURE 2: The exclusion criteria for filtering variants for *de novo* analysis. The number of variants eliminated by each filter for individual probands is listed in **Table 2**. MAF = minor allele frequency, LOF=loss of function, VQSR = Variant Quality Score Recalibration.

Exclusion criteria:		AU0002-0201	AU0005-0202	AU0238-0201	AU0243-0201	AU0245-0201	AU0245-0202
VarSeq *de novo* workflow:	**Read Depths < 10**	26639	22003	29760	20720	15961	28826
	Genotype Quality < 20	4067	4140	4092	4238	4570	3997
	MAF > 0.01 (retained if missing)	87680	89910	85754	90970	93383	86118
	Effect is not LOF or Missense	48321	50386	47328	50513	52189	47787
	Not *de novo* inheritance pattern (present in proband genotype, absent in parent genotypes)	17650	17880	17412	17922	18223	17668
Manual workflow:	***MUC* and *HLA* family genes**	41	20	30	23	35	22
	Alt allele freq for proband < 0.25	81	118	85	83	108	61
	VQSR filter, not PASS	14	30	28	21	18	15
	MAF in ExAC database > 0.05 (retained if missing)	4	3	5	4	3	3
	Other (see Table### for more info)	1	6	3	3	5	1
Prioritized variants:		**1**	**3**	**2**	**2**	**4**	**1**

Table 2: Number of variants excluded by filters for *de novo* analysis. For each proband listed, the number of variants that were eliminated by each filter is listed, until the remaining prioritized variants (n=13). As noted in the text, variants that were absent in the proband may not have been excluded until the final filter of the VarSeq workflow requiring *de novo* inheritance (present in proband genotype, absent in parent genotypes).

While most of the filters were applied with clearly objective criteria, there were 19 variants that were eliminated across all the probands using other criteria. Major categories for exclusion in this "Other" filter include concerns about the *de novo* nature of the variant or the ability of the variant to impact protein function (e.g. non-coding variants), known function or expression of the gene without overlap with ASD-related

phenotypes, known polymorphisms that are common in the population, and genes with lack of evidence (unknown function and no protein expression in brain tissues). A brief summary of the rationales for their exclusions is summarized in **Table 3**. These variants were not prioritized in this analysis; however, they may still be of interest. For example, variants in the lack of evidence category may be reclassified as the genes' functions are better understood or non-coding variants can be assessed to see if they still alter the gene regulation.

Table 3: Variants removed by "Other" filter and rationale for not prioritizing them. (continued on next page)

	Proband	Gene	Variant	Rationale for not prioritizing	Citation(s)
Quality/effect concerns	AU0005-0202	TMEM254	NM_001270367.1:c.160-175delA	Non-coding, located in intron.	
	AU0238-0201	RBMX	NM_002139.3:c.642T>A	X-linked, called het. in male proband; alternate allele present in mother (1/25).	
	AU0245-0201	THEGL	NM_001256475.1:c.706C>T	Identified in sister, AU0245-0202	
	AU0245-0202	THEGL	NM_001256475.1:c.706C>T	Identified in sister, AU0245-0201	
	AU0245-0201	RNPEPL1	NM_018226.4:c.281_283delCCG	Non-coding. Located in 5'UTR.	
Other known functions	AU0002-0201	EPM2A	NM_005670.3:c.136G>C	Associated with AR progressive myoclonic epilepsy. In ClinVar as likely benign/benign.	1; 2
	AU0005-0202	TGFBR1	NM_004612.2:c.51_53delGGC	Associated with AD Loeys-Dietz syndrome. In ClinVar as benign.	1; 2
	AU0238-0201	CCDC40	NM_017950.3:c.2832+434_2832+462delGG GACGCGCGCGAGGCACGTGCACGAACAA	Associated with AR ciliary dyskinesia.	1
	AU0243-0201	KRTAP1-3	NM_030966.1:c.90_119dupTGAGACCAGC TGCTGCCAGCCAAGCTGCTG	Known role/expression in hair and skin.	1; 3
	AU0243-0201	MYO9B	NM_004145.3:c.6097A>C	Known role in ulcerative colitis and celiac disease.	1; 3
	AU0243-0201	PDLIM5	NM_001256426.1:c.96+1067G>A	Known role/expression in muscle.	1; 3
	AU0245-0201	PELI1	NM_020651.3:c.201+1G>A	Associated with Kawasaki disease suceptibility.	1; 3

	Proband	Gene	Variant	Rationale for not prioritizing	Citation(s)
Poly.	AU0005-0202	ATN1	NM_001940.3:c.1462_1482del(CAG)x7	Known polymorphic triplet-repeat region; expansion responsible for Dentatorubral pallidoluysian atrophy.	Koide et al., 1994
Lack of evidence	AU0238-0201	PRDM9	NM_020227.2:c.2497_2580del	Deletion of zinc finger; common polymorphic number of ZFDs observed.	1; 5
	AU0005-0202	PLEKHH3	NM_024927.4:c.1816C>T	Function unknown, and protein not seen in brain tissues.	3; 4
	AU0005-0202	FAM157A	NM_001145248.1:c.210_218del(GCA)x3	Function unknown, and protein not seen in brain tissues.	4
	AU0005-0202	DEF8	NM_207514.2:c.1288C>T	Function unknown, and protein not seen in brain tissues.	3; 4
	AU0245-0201	CCDC73	NM_001008391.3:c.3212C>T	Function unknown, and protein not seen in brain tissues.	1; 3
	AU0245-0201	ANKLE1	NM_152363.5:c.*125_*140delTGTGTGTGTGTGTT	Function unknown, and protein not seen in brain tissues.	3; 4

Table 3: Variants removed by "Other" filter and rationale for not prioritizing them. The general reasons for not prioritizing these variants were concerns about the quality of the variant (e.g. evidence against its de novo nature or ability to affect protein function); known functions without overlap with ASD phenotypes; (3) known common polymorphic sites (Poly.), and (4) lack of evidence for, namely genes whose function is not well understood and for which the protein is not expressed in the brain. For citations, 1: omim.org, 2: ClinVar, ncbi.nlm.nih.gov/clinvar (Landrum et al., 2016), 3: humanproteinatlas.org (Uhlen et al., 2015), 4: genecards.org, 5: uniprot.org (Uniprot Consortium, 2015).

After these filters were applied, thirteen variants remained across the six probands. Some are more compelling than others because of what is known about the function of the gene and what effect the specific variant is predicted to have. They are described in more detail below for each proband.

PRIORITIZED VARIANTS

AU0002-0201

For proband AU0002-0201, there is one variant that was prioritized by the described workflow.

I. UNC5B

The variant is in the *UNC5B* gene (NM_170744.4: c.2841_2490+41delinsT) and results in the deletion of fifty nucleotides that includes the splice junction of an exon near the terminal end of the transcript. **Figure 3** shows the region that is deleted, in detail and in the architecture of the gene. For the proband, there were 49 reference reads and 17 alternate reads (alternate allele frequency = 0.26); both parents were homozygous for the reference sequence. This change is found with an allele frequency of 0.004227 (505/119,478) in the ExAC database (exac.broadinstitute.org; Lek *et al.*, 2016). Seven homozygotes were also found in the database. The exact functional effect of this nucleotide change on the protein is uncertain; however, the ExAC database predicts that it will be a loss-of-function with high confidence.

(a)
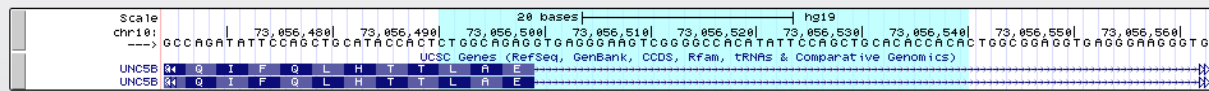
(b)

Figure 3: Deletion in *UNC5B*. Images from the UCSC Genome browser of *UNC5B* transcripts with the region deleted in the variant highlighted in light blue: (a) zoomed into the splice junction affected (at the 3' end of the 15th of 17 exons for the canonical transcript), including the nucleotide sequence above and the amino acid sequences of two different transcripts below and (b) zoomed out to see the entire gene, including two different transcripts.
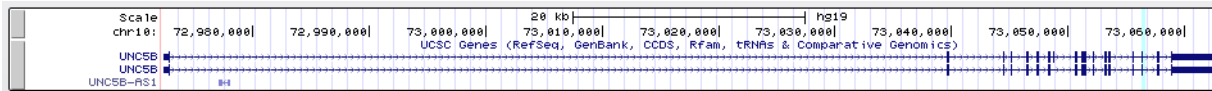
The *UNC5B* gene is a member of the netrin family of receptors (Leonardo *et al.*, 1997). It is expressed in brain tissue, but also widely expressed in other tissues, including hematopoietic, immune, and lung tissues (proteinatlas.org; Komatsuzaki *et al.*, 2002; Uhlen *et al.*, 2015). Netrin is secreted by cells to guide the growth of neuronal axons, as well as during other processes including angiogenesis (Duman-Scheel, 2009; Forcet *et al.*, 2002; Serafini *et al.*, 1996; Serafini *et al.*, 1994). The *UNC5B* receptor specifically mediates axon repulsion (Hong *et al.*, 1999; Keleman and Dickson, 2001; Lu *et al.*, 2004; Muramatsu *et al.*, 2010) and induces apoptosis when not bound to netrin (Guenebeaud *et al.*, 2010; Thiebault *et al.*, 2003).

Differences in brain function can be associated with differences in the connections set during fetal development between neurons, determined by the attractive and repulsive signaling between netrin (and other secreted molecules) and receptors like Unc5b that bind them. In fact, a balanced translocation that affects expression of the netrin gene

was identified in a female patient with symptoms of Rett syndrome (Borg *et al.*, 2005), and *de novo* mutations in netrin were identified in individuals with ASD from two different families (O'Roak *et al.*, 2012).

The variant in *UNC5B* identified for this proband is upstream of the death domain of the protein (uniprot.org; Uniprot Consortium, 2015) and could form a truncated protein without this protein interaction domain that allows it to induce apoptosis. However, many frameshifted RNAs are regulated by nonsense-mediated decay. Therefore, this variant could alternatively result in decreased levels of mRNA produced from this allele. In mouse models, haploinsufficiency of a paralog (*UNC5C*) causes expression changes and a behavioral phenotype (reduced amphetamine-induced locomotion) (Auger *et al.*, 2013). Therefore, either mechanism could cause this variant to influence the formation of neural networks in the brain.

AU0005-0202

For proband AU005-0202, there are three variants that were prioritized by the described workflow.

I. ASTN2

There is a variant in the *ASTN2* gene (NM_014010.4:c.661_663delCTG). This is an in-frame deletion of three nucleotides. For the proband, there were 12 reference reads and 5 alternate reads (alternate allele frequency = 0.29); both parents were homozygous for the reference sequence. This change is found with an allele frequency

of 0.01450 (1537/106,034) in the ExAC database (exac.broadinstitute.org; Lek *et al.*, 2016). No homozygotes were identified in the database. This variant in the nucleotide sequence is predicted to result in a deletion of a leucine that is found in a sequence of six leucines in the reference protein sequence. This occurs within a transmembrane domain (uniprot.org; Uniprot Consortium, 2015) and is predicted to be deleterious by modeling software Provean (provean.jcvi.org; Choi *et al.*, 2012).

The *ASTN2* gene codes for a protein, astrotactin 2, which regulates neuronal migration during brain development (Wilson *et al.*, 2010). *ASTN2* expression peaks in the early human cerebral cortex at the end of the first trimester and in the cerebellar cortex shortly after birth (Lionel *et al.*, 2014). Neurons expressing astrotactin 2 form and release cell adhesions as they migrate along glial cells in the forming brain, starting at approximately 12 weeks of gestation (Giedd *et al.*, 1999; Tau and Peterson, 2010).

Inhibition of neuronal migration has been proposed as an important pathophysiology in ASD (Reiner *et al.*, 2016). *ASTN2* is expressed in Purkinje cells of the cerebellum (Wilson *et al.*, 2010) which can be decreased in number by up to 50% in individuals with ASD (Fatemi *et al.*, 2012). Two studies have identified copy number variations including deletions of *ASTN2* as significantly enriched in individuals with ASD (Glessner *et al.*, 2009; Lionel *et al.*, 2014), supporting that haploinsufficiency could influence an individual's ASD phenotype. These CNVs were often inherited and also seen in controls; therefore, they exhibit reduced penetrance in their effect on ASD (Lionel *et al.*, 2014). Of note, the significant enrichment of deletions of the 3' terminus affecting all

known transcripts of *ASTN2* was only seen in males (Lionel *et al.*, 2014), and this proband is female.

II. RNF31

There is a variant in the *RNF31* gene (NM_024927.4:c.1816C>T). For the proband, there were 7 reference reads and 4 alternate reads (alternate allele frequency = 0.36); both parents were homozygous for the reference sequence. This change was found with a frequency of 0.01839 (2033/110,572) in the ExAC database (exac.broadinstitute.org; Lek *et al.*, 2016). One hundred and seventeen homozygotes were identified in the database. This is a single nucleotide substitution and is predicted to result in a substitution of an isoleucine for a valine in the protein produced. However, that change is predicted to be neutral by modeling software Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The *RNF31* gene produces a protein, with three RING finger domains, that forms a linear ubiquitin chain assembly complex, called LUBAC, with other proteins, for which it serves as the E3 ubiquitin-protein ligase (Ikeda *et al.*, 2011). Mutations in this complex have been associated with inflammation (through NF-kappa-B activation) and immune deficiencies (Boisson *et al.*, 2015; Ikeda *et al.*, 2011). RNF31 protein is found in many tissues but is present in brain tissues (proteinatlas.org; Uhlen *et al.*, 2015), and NF-kappa-B activation also contributes to signaling in synapses and learning behaviors (Meffert *et al.*, 2003).

The amino acid altered by this variant is near the C-terminal end of the RNF31 protein in a domain identified as important for the transfer of the ubiquitin to the chain (Smit *et al.*, 2012). This domain is unique to RNF31, so the effect of this variant is difficult to predict. A splice donor variant of some *RNF31* transcripts (but not all, including the canonical transcript) was identified in a study of high-risk ASD families (Matsunami *et al.*, 2014). That variant (c.42+1G>A) is located far upstream of the variant found in this proband and more convincing for a loss of at least some of RNF31 function than the one identified for the proband; however, it argues for the potential of haploinsufficiency of *RNF31* affecting ASD phenotypes.

III.   GET4

There is a variant in the *GET4* gene (NM_015949.2:c.3_5delGGC). For the proband, there were 12 reference reads and 8 alternate reads (alternate allele frequency = 0.40). Both parents were called homozygous for the reference sequence. This change is found with an allele frequency of 0.002252 (1/444) in the ExAC database; however, this site was covered in fewer than 80% of individuals in the database, which may indicate it is a low quality site (exac.broadinstitute.org; Lek *et al.*, 2016). No homozygotes were found in the database. This in-frame deletion results in the deletion of one alanine from the protein sequence, from a sequence of five alanines, and is considered to be neutral by modeling software Provean (provean.jcvi.org; Choi *et al.*, 2012).

GET4 codes for a protein that is a component of the BAT3 complex which transports tail-anchored membrane proteins to the endoplasmic reticulum (Mariappan *et al.*, 2010).

It can also enter the nucleus and influence DNA damage repair signaling and damage-induced cell death (Krenciute *et al.*, 2013). The GET4 protein is expressed broadly in many tissues; however, it is seen at high levels in brain tissue, including neuronal cells of the hippocampus and cerebral cortex and multiple cell types in the cerebellum (proteinatlas.org; Uhlen *et al.*, 2015). However, no specific phenotypes have been associated with mutations in the *GET4* gene, and the variant itself does not convincingly affect protein function. Therefore, more information is needed to understand if *GET4*, and specifically the variant identified in the proband, have any influence on ASD phenotypes.

AU0238-0201

For proband AU0238-0201, there are two variants that were prioritized by the described workflow.

I. ST6GAL2

There is a variant in the *ST6GAL2* gene (NM_032528.2:c.1319-6207T>C). For the proband, there were 6 reference reads and 8 alternate reads (alternate allele frequency = 0.57). Both parents were called homozygous for the reference sequence. This change is found with an allele frequency of 0.006675 (51/7640) in the ExAC database; however, this site was covered in fewer than 80% of individuals in the database, which may indicate it is a low quality site (exac.broadinstitute.org; Lek *et al.*, 2016). No homozygotes were found in the database. In the canonical transcript, this change falls within an intron. However, there is one transcript with a coding region overlapping this

variant. For that transcript, this single nucleotide substitution is predicted to result in a substitution of an arginine for a tryptophan in the protein produced. However, this change is predicted to be neutral by modeling software Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The *ST6GAL2* gene encodes an oligosaccharide-specific sialyltransferase (Takashima *et al.*, 2002). The protein is located in the Golgi membrane, allowing it to modify oligosaccharides. *ST6GAL2* is expressed in the fetal brain, as well as the small intestine and colon (Krzewinski-Recchi *et al.*, 2003; Takashima *et al.*, 2002).

The variant identified would alter the amino acid sequence of the luminal domain of the protein; however, the exact effect of this variant on protein function is difficult to predict. Interestingly, *ST6GAL2* expression is altered in the prefrontal cortex of mice by risperidone treatment, and a different polymorphism (rs1448110) located outside the coding region of *ST6GAL2* was associated with risperidone treatment response, albeit not strongly (Ikeda *et al.*, 2010). Risperidone and other antipsychotic drugs are often prescribed to individuals with ASD to treat symptoms, including aggressive behavior, hyperactivity, and repetitive behaviors (Mandell *et al.*, 2008). However, risperidone has undesirable side effects and is not effective for all individuals (Lemmon *et al.*, 2011), and previous studies have shown a genetic basis to these responses (Correia *et al.*, 2010). It is possible that *ST6GAL2* variations play a role in this pharmacogenetic variation and could help determine which patients would respond best to the drug.

II. SUFU

There is a variant in the *SUFU* gene (NM_016169.3:c.37A>C). For the proband, there were 35 reference reads and 16 alternate reads (alternate allele frequency = 0.31). Both parents were called homozygous for the reference sequence. This change was not present in the ExAC database, and while the *SUFU* gene had a mean coverage depth of 73.40, this variant is located in a region with particularly low coverage, see **Figure 4** (exac.broadinstitute.org; Lek *et al.*, 2016).   This is a single nucleotide substitution and is predicted to result in a substitution of a proline for a threonine in the protein produced. This change is predicted to be neutral by modeling software Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).
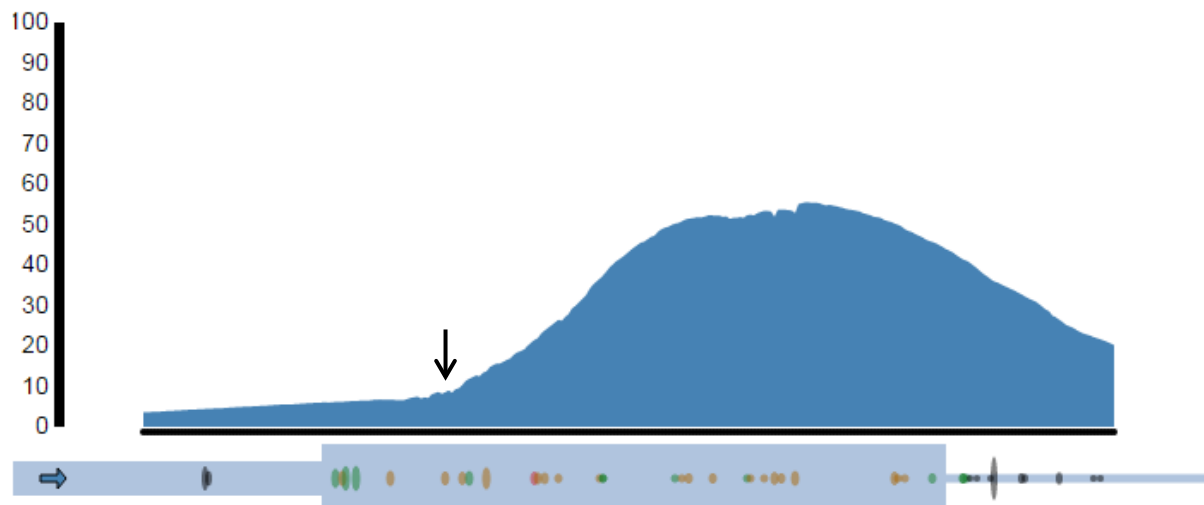
Figure 4: Mean coverage in ExAC database in the area surrounding the variant found in *SUFU*. The mean coverage across the first exon of the *SUFU* gene in the ExAC database on the top with the gene architecture represented below (exac.broadinstitute.org; Lek *et al.*, 2016). The direction of transcription is left to right in this image. The thicker blue box represents the exon, and the ovals are variants found in the ExAC database (with size representing frequency in the database). Importantly, the variant in proband AU0238-0201 (g. 104263946), which was not seen in individuals in the ExAC database, is immediately adjacent to the variant at g.104263947 (seen in 1/20,564 alleles), indicated by the arrow.

The *SUFU* gene encodes a negative regulator of the Hedgehog signaling pathway, which is important during embryogenesis, including the formation of the neural tube, as well as carcinogenesis (Svard *et al.*, 2006). SUFU is also a tumor suppressor and loss of function mutations cause an autosomal dominant hereditary predisposition to developing medulloblastoma, albeit with incomplete penetrance (Brugieres *et al.*, 2010; Taylor *et al.*, 2002).

The effect of this variant on SUFU protein function, if any, is difficult to predict. Sonic hedgehog (SHH) protein is found in higher levels in the serum of individuals with ASD

and positively correlated with an assessment of the severity of an individual's ASD (Al-Ayadhi, 2012). Mutations in another member of the signaling pathway (Patched) have also been associated with ASD and intellectual disability (Noor *et al.*, 2010). Therefore, the possibility that this repressor of the SHH pathway could affect ASD phenotypes as well is potentially interesting.

AU0243-0201

For proband AU0243-0201, there are two variants that were prioritized by the described workflow.

I. BMPR1B

There is a variant in the *BMPR1B* gene (NM_001256793.1:c.761G>A). For the proband, there were 38 reference reads and 49 alternate reads (alternate allele frequency = 0.56); both parents were homozygous for the reference sequence. This change was found with a frequency of 0.001081 (131/121,222) in the ExAC database (exac.broadinstitute.org; Lek *et al.*, 2016). No homozygotes were identified in the database. This is a single nucleotide substitution and is predicted to result in a substitution of a histidine for an arginine in the protein produced, and that change is predicted to be deleterious by modeling software Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The *BMPR1B* gene is a member of the bone morphogenetic protein family of transmembrane serine/threonine kinases, and the ligands of these receptors are

members of the TGF-beta family (Astrom *et al.*, 1999). These signaling pathways are involved in many developmental processes including the proliferation and differentiation of cells during embryonic development and antagonize the SHH pathway (Panchision and McKay, 2002), described above. *BMPR1B* protein is expressed in late neural stem cells as they differentiate into astrocytes (Panchision and McKay, 2002) and may be involved in glioblastoma malignancy (Lee *et al.*, 2008; Liu *et al.*, 2012).

The variant identified in this proband would result in an amino acid change in the cytoplasmic protein kinase domain of the protein (uniprot.org; Uniprot Consortium, 2015; Demirhan *et al.*, 2005). A change in this region could negatively affect the ability to signal downstream, or potentially result in a constitutively active signal. Other conditions, notably brachydactyly, have been associated with a heterozygous dominant-negative mutation (Lehmann *et al.*, 2003). Therefore, additional phenotypic information about this proband might give evidence for or against this potential mechanism.

II. TMEM8A

There is a variant in the *TMEM8A* gene (NM_021259.2:c.10delG). For the proband, there were 18 reference reads and 10 alternate reads (alternate allele frequency = 0.36); both parents were called homozygous for the reference sequence. This change was not present in the ExAC database, and the coverage of the first exon where this variant is located was well below the mean coverage depth for the *TMEM8A* gene of 42.91 (exac.broadinstitute.org; Lek *et al.*, 2016). This is a single nucleotide deletion and is predicted to result in a frameshift in the first exon of the transcript.

The *TMEM8A* gene is expressed in placenta, pancreas, and lymphohematopoetic tissues and produces a putative five-span transmembrane protein (Motohashi *et al.*, 2000). The function of the protein is not well understood; however, it is expressed highly in resting T lymphocytes and is downregulated by cell activation, indicating a possible role in the resting status of T cells (Motohashi *et al.*, 2000). The protein also seen in brain tissues, especially neuronal cells of the lateral ventricle; however, the protein is expressed broadly across many tissues and at higher levels in immune system and gastrointestinal tract tissues (proteinatlas.org; Uhlen *et al.*, 2015). More information is necessary about the function of this gene to determine if this predicted loss-of-function variant could have an effect on the proband's ASD phenotype through haploinsufficiency.

AU0245-0201

For proband AU0245-0201, there are four variants that were prioritized by the described workflow.

I. DPF3

There is a variant in the *DPF3* gene (NM_001280544.1:c.599T>G). For the proband, there were 24 reference reads and 28 alternate reads (alternate allele frequency = 0.54); both parents were homozygous for the reference sequence. This change was not present in the ExAC database, although the *DPF3* gene had a mean coverage depth of

53.21 (exac.broadinstitute.org; Lek *et al.*, 2016).   This is a single nucleotide substitution and is predicted to result in a substitution of a glycine for a valine in the protein produced, and that change is predicted to be deleterious by modeling software Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

DPF3 is a member of a neuron-specific chromatin remodeling complex called nBAF (Son and Crabtree, 2014). The protein is found in the cerebellum (high in the molecular layer and medium in Purkinje cells) as well as other nervous system tissues (proteinatlas.org; Uhlen *et al.*, 2015). The protein binds acetylated lysines on histones 3 and 4 through interactions with its C-terminal plant homeodomain zinc fingers and regulates transcription (Zeng *et al.*, 2010). The switch from npBAF to nBAF regulated transcription coordinates a switch from neural progenitors to differentiated neurons (Son and Crabtree, 2014).

The variant identified for this proband in *DPF3* results in a change in the amino acid sequence of the first of three zinc finger domains, which is a C2H2 Kruppel-like zinc finger (uniprot.org; Uniprot Consortium, 2015). This domain could be important for other interactions in the nBAF complex. Mutations in other components of the nBAF and npBAF complexes have been identified in individuals with ASD (Neale et al., 2012; Nord et al., 2011; O'Roak et al., 2012), demonstrating that haploinsufficiency may be a potential mechanism.

II. EPHB2

There is a variant in the *EPHB2* gene (NM_004442.6:c.27_29delGCT). For the proband, there were 48 reference reads and 37 alternate reads (alternate allele frequency = 0.44); both parents were homozygous for the reference sequence. This change was not present in the ExAC database, although the variant was only covered in 34 individuals in the database indicating a potentially low-quality site (exac.broadinstitute.org; Lek *et al.*, 2016). This is a deletion of three nucleotides and is predicted to cause a deletion of a leucine in the amino acid sequence. This change is predicted to be neutral by modeling software Provean (provean.jcvi.org; Choi *et al.*, 2012); however, the *EPHB2* gene is predicted to be significantly intolerant of missense (z=3.45) and loss of function (pLI=1.00) mutations (Lek *et al.*, 2016).

EPHB2 is a member of a family of receptor tyrosine kinase transmembrane glycoproteins. EphB2 protein is found enriched in brain tissues; however, it is also broadly expressed in many other tissues (proteinatlas.org; Uhlen *et al.*, 2015). In mice, the *EPHB2* homolog is important for axon guidance (Henkemeyer *et al.*, 1996). When bound to ligands, Ephb2 receptors transduce signals through modulators of the actin cytoskeleton to regulate dendritic spine development (Irie and Yamaguchi, 2002; Kayser *et al.*, 2011). Studies have shown an increase in the density of dendritic spines in certain neurons of individuals with ASD (Hutsler and Zhang, 2010). Previous studies of *de novo* mutations in individuals with ASD have identified rare missense and nonsense variants in *EPHB2* (Kong *et al.*, 2012; Sanders *et al.*, 2012).

The variant in *EPHB2* identified for this proband is in the signal peptide domain (uniprot.org; Uniprot Consortium, 2015). These domains are not sequence specific, and prediction software indicated that the deleted lysine would not affect the cleavage site (cbs.dtu.dk/services/SignalP; Petersen *et al.*, 2011). Despite the functional connection of *EPHB2* that suggests a role in processes connected to ASD phenotype, it is uncertain if this particular variant would affect EPHB2 protein function.

III. TYRO3

There is a variant in the *TYRO3* gene (NM_006293.3:c.85_86insAGTC). For the proband, there were 47 reference reads and 36 alternate reads (alternate allele frequency = 0.43); both parents were homozygous for the reference sequence. This change was not present in the ExAC database, and the coverage of the first exon where this variant is located was well below the mean coverage depth for the *TYRO3* gene of 58.91 (exac.broadinstitute.org; Lek *et al.*, 2016). This is an insertion of four nucleotides and will result in a frameshift in the first exon of this gene. Other frameshifts in *TYRO3* found in the ExAC database were rare (n=5, frequency range = 1.702e-5 - 8.245e-6); however, the gene was not predicted to be significantly constrained for LOF mutations (exac.broadinstitute.org; Lek *et al.*, 2016).

TYRO3 is a member of a subfamily of receptor protein tyrosine kinases that play a role in neurogenesis in the adult hippocampus (Ji *et al.*, 2015). Tyro3 protein is found

enriched in brain tissues, especially the neuronal cells of the cerebral cortex and hippocampus; however, it is also broadly expressed in many other tissues (proteinatlas.org; Uhlen *et al.*, 2015). When activated by ligands, the receptor kinase forms a dimer and activates intracellular signaling pathways involved in cell survival and proliferation as well as inflammation responses and phagocytosis (Binder and Kilpatrick, 2009). In the central nervous system, signaling through TAM receptors suppresses inflammation that could otherwise impair neuronal stem cell proliferation and differentiation (Ji *et al.*, 2015). While the involvement of *TYRO3* in these important developmental processes in the brain and the potential loss of function caused by the early frameshift with this variant argue for the potential for it to affect this proband's ASD phenotype, TAM signaling has not been specifically implicated in ASD.

IV.   POU4F1

There is a variant in the *POU4F1* gene (NM_006237.3:c.741_743delGGC). For the proband, there were 31 reference reads and 19 alternate reads (alternate allele frequency = 0.38); both parents were homozygous for the reference sequence. This change was not present in the ExAC database, and while the *POU4F1* gene had a mean coverage depth of 28.96, this variant is located in a region with particularly low coverage, see **Figure 5** (exac.broadinstitute.org; Lek *et al.*, 2016). This is a deletion of three nucleotides predicted to result in the deletion of an alanine from the amino acid sequence.
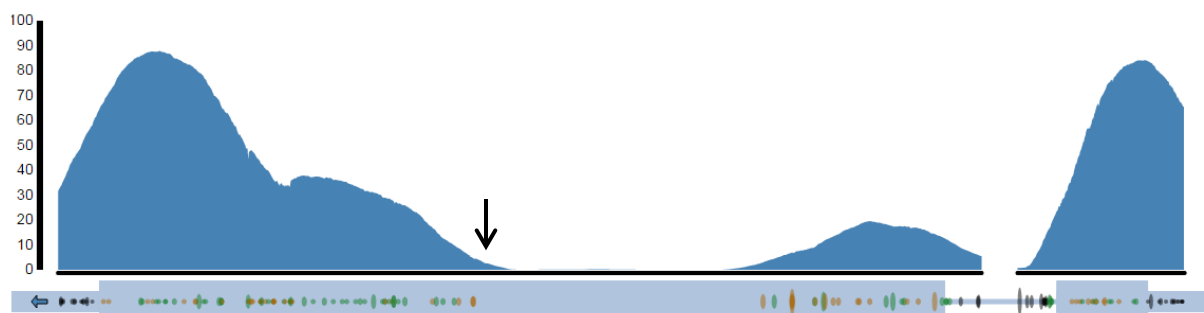
Figure 5: Mean coverage in ExAC database in the area surrounding the variant found in *POUF41*. The mean coverage across the *POU4F1* gene in the ExAC database on the top, with the gene architecture represented below (exac.broadinstitute.org; Lek *et al.*, 2016). The direction of transcription is right to left in this image. The thicker blue boxes represent the exons (n=2), and the ovals are variants found in the ExAC database (with size representing frequency in the database). Importantly, the variant in proband AU0245-0201 (g.79176067), which was not seen in individuals in the ExAC database, is found in the region of Exon 2 with low coverage (bounded by variants at g.79176056 and g.79176444), approximated by the arrow.

POU4F1 codes for a transcription factor that affects cell survival and neural differentiation, expressed especially in sensory neurons (Berwick *et al.*, 2009; Budhram-Mahadeo *et al.*, 1998; Latchman, 1998). A deletion of a 153.5kb region upstream of the *POUF41* gene was identified as tracking with the ASD phenotype in a family with multiple affected members (Salyakina *et al.*, 2011). While not directly affecting the coding region, this finding supports that altering the regulation of the *POUF41* gene could contribute to an individual's ASD phenotype, potentially in a haploinsufficient manner. However, the variant identified in this proband is not in a defined protein domain (uniprot.org; Uniprot Consortium, 2015) and is predicted to be neutral by modeling software Provean (provean.jcvi.org; Choi *et al.*, 2012). Therefore, it is difficult

51

to predict if the in-frame deletion will have any effect on the protein function for this individual.

AU0245-0202

For proband AU0245-0202, there is one variant that was prioritized by the described workflow.

I. *SLIT2*

There is a variant in the *SLIT2* gene (NM_004787.1:c.4357T>C). For the proband, there were 6 reference reads and 7 alternate reads (alternate allele frequency = 0.54); both parents were homozygous for the reference sequence. This change was not present in the ExAC database, although the *SLIT2* gene had a mean coverage depth of 63.55 and the region including this variant was covered at approximately that depth (exac.broadinstitute.org; Lek *et al.*, 2016). This is a single nucleotide substitution and is predicted to result in a substitution of an arginine for a cysteine in the protein produced, and that change is predicted to be deleterious by modeling software Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The *SLIT2* gene codes for a ligand involved in repulsive axon guidance (Bagri *et al.*, 2002; Brose *et al.*, 1999; Li *et al.*, 1999; Shu and Richards, 2001). Through interacting with Robo receptors, Slit proteins control cytoskeletal remodeling as a major regulator of axon pathfinding (Hammond *et al.*, 2005). These pathways are essential for proper

brain development, and mutations, including deletions, missense, and synonymous changes, in another Slit protein and Robo receptors have been associated with ASD (Cukier *et al.*, 2014; Iossifov *et al.*, 2014; Prasad *et al.*, 2012) and major depressive disorder (Glessner *et al.*, 2010).

The variant identified for this proband is located in the C-terminal cysteine knot (CTCK) domain and affects one of the cysteine residues involved in a disulfide bond important for the structure of this domain that is found in other signaling ligands (uniprot.org; Uniprot Consortium, 2015; Iyer and Acharya, 2011). This supports the potential for the variant to affect protein function, most likely negatively, and contribute to ASD phenotype through haploinsufficiency.

SUMMARY

For each of the probands that was submitted with a trio, the described workflow was performed to identify variants that were most likely to be accurate *de novo* variants with the potential to contribute to the ASD phenotype of the proband. Only a few were in genes that have been associated with ASD before, but many were involved in pathways with potential to affect brain development, including molecularly (e.g. SHH signaling) and pathologically (e.g. axon guidance). The prioritized variants are summarized in **Table 4**.

Proband	Gene	Variant	Association with ASD: Variant?	Gene?	Other
AU0002-0201	*UNC5B*	NM_170744.4: c.2841_2490+41delinsT	n	n	Receptor for netrin (axon guidance)[1]
AU0005-0202	*ASTN2*	NM_014010.4:c.661_663delCTG	n	y[2]	
AU0005-0202	*RNF31*	NM_017999.4:c.3181G>A	n	y[3]	
AU0005-0202	*GET4*	NM_015949.2:c.3_5delGGC	n	n	
AU0238-0201	*ST6GAL2*	NM_032528.2:c.1319-6207T>C	n	n	
AU0238-0201	*SUFU*	NM_016169.3:c.37A>C	n	n	Inhibitor of SHH[4]
AU0243-0201	*BMPR1B*	NM_001256793.1:c.761G>A	n	n	Antagonist of SHH[5]
AU0243-0201	*TMEM8A*	NM_021259.2:c.10delG	n	n	
AU0245-0201	*DPF3*	NM_001280544.1:c.599T>G	n	n	Other complex (nBAF and npBAF) members[6]
AU0245-0201	*EPHB2*	NM_044442.6:c.27_29delGCT	n	y[7]	
AU0245-0201	*TYRO3*	NM_006293.3:c.85_86insAGTC	n	n	
AU0245-0201	*POU4F1*	NM_006237.3:c.741_743delGGC	n	y[8]	
AU0245-0202	*SLIT2*	NM_004787.1:c.4357T>C	n	n	Other Slit/Robo pathway components (axon guidance)[9]

Table 4: Summary table of the prioritized variants from *de novo* analysis. None of the prioritized variants have been identified in studies of individuals with ASD; however, several of the genes had previous associations. Additionally, some of the genes are involved in cellular signaling pathways or developmental processes that have previous associations with ASD, as described above in more detail. Selected sources in support of ASD association: [1](Borg *et al.*, 2005; O'Roak *et al.*, 2012) [2](Glessner *et al.*, 2009; Lionel *et al.*, 2014) [3](Matsunami *et al.*, 2014) [4](Al-Ayadhi, 2012; Noor *et al.*, 2010) [5](Panchision and McKay, 2002) [6](Neale *et al.*, 2012; Nord *et al.*, 2011; O'Roak *et al.*, 2012) [7](Kong *et al.*, 2012; Sanders *et al.*, 2012) [8](Salyakina *et al.*, 2011) [9] (Cukier *et al.*, 2014; Iossifov *et al.*, 2014; Prasad *et al.*, 2012).

AIM II: TARGETED ANALYSIS OF PROBAND BASED ON CLINICAL FINDINGS

DEMOGRAPHIC INFORMATION FOR AU0237-0201

A sample for the proband AU0237-0201 was submitted for sequencing along with a sample from her mother. The proband is a 13 year 10 month old female. She was administered Module 3 of the ADOS and given a classification of Autism. She was administered the WASI, and her Full-Scale 4 composite score was 58. Her verbal and perceptual reasoning subscores were 64 and 56, respectively.

ANALYSIS OF VARIANTS

A chromosomal microarray, including 2.67 million probes, was ordered clinically for this proband prior to enrollment in the research project, and showed a deletion of approximately 371 kilobases (kb) on chromosome 3. The loss is located at 3q25.32 with coordinates 158,405,216 - 158,776,705 (hg19). The clinical laboratory identified this as a variation of unclear significance. Parental testing was done demonstrating paternal inheritance of this deletion, but documentation of this was not available for review. Because the proband would be hemizygous for any variants in this region, this region may be of particular interest. Therefore, the variants identified are discussed below to determine if they contribute to this proband's ASD phenotype.

Genes overlapping with this region include *RARRES1*, *MFSD1,* and part of *GFM1*, as well as a large intergenic region. For the proband and her mother there were nine total

variants identified by the research sequencing within the coordinates given by the clinical microarray. The variants identified in each gene are shown in **Figures 6-8**.
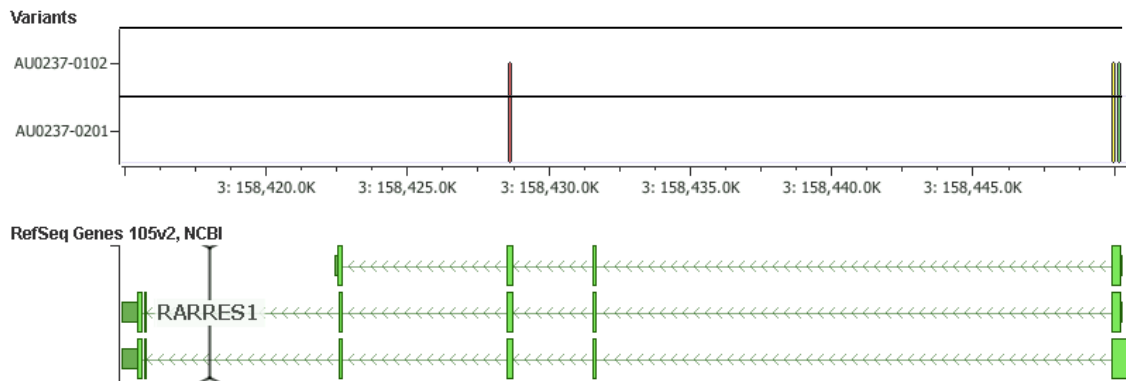
Figure 6: Variants in *RARRES1*. Image showing the variants identified in *RARRES1* (from left to right: rs2307064, rs6441223, and rs6786423). The top half of the image indicates the location of the variants, and shows their presence as heterozygous (half-height bar) or apparently homozygous (full-height bar) state for the proband (bottom) or her mother (top). The bottom half of the image shows the genomic structure including the introns (lines with arrows indicating direction of transcription) and exons (solid boxes) for multiple transcripts of the gene. Image from VarSeq v1.3.4 GenomeBrowse® visualization tool, by Golden Helix, Inc.

Figure 7: Variants in *MFSD1*. Image showing the variants identified in *MFSD1* (from left to right: rs28364680 and rs3765083). The top half of the image indicates the location of the variants, and shows their presence in the heterozygous (half-height bar) or apparently homozygous (full height bar) state for the proband (bottom) or her mother (top). The bottom half of the image shows the genomic structure including the introns (lines with arrows indicating direction of transcription) and exons (solid boxes) for multiple transcripts of the gene. Image from VarSeq v1.3.4 GenomeBrowse[®] visualization tool, by Golden Helix, Inc.
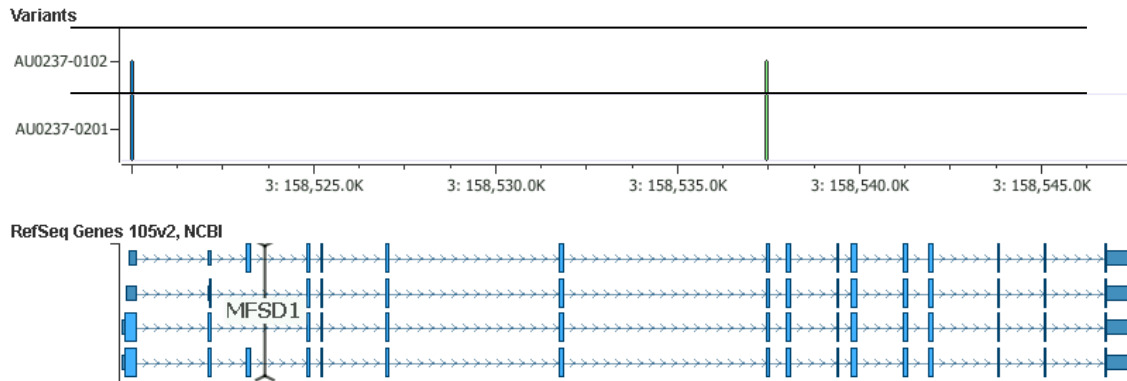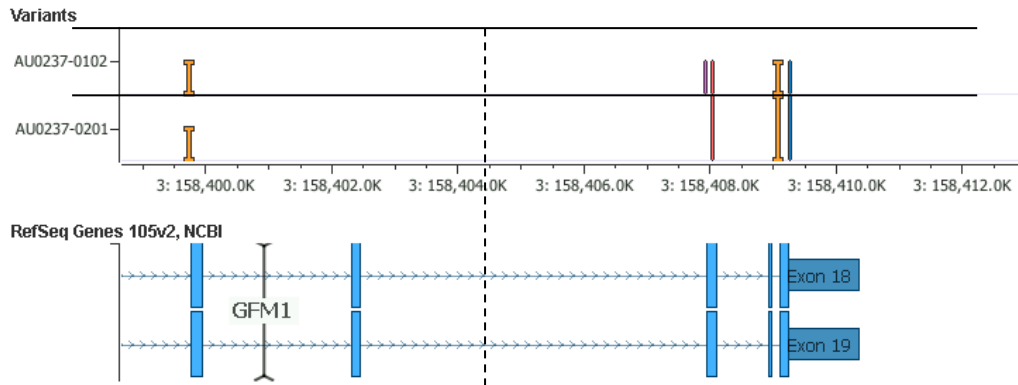
Figure 8: Variants in *GFM1*. Image showing the variants identified in *GFM1* (from left to right: rs11436468, rs34920045, rs62288347, rs145803239, rs1047355). The top half of the image indicates the location of the variants, and shows their presence in the heterozygous (half-height bar) or apparently homozygous (full height bar) state for the proband (bottom) or her mother (top). rs34920045 is absent in the proband, but present in her mother. The bottom half of the image shows the genomic structure including the introns (lines with arrows indicating direction of transcription) and exons (solid boxes) for multiple transcripts of the gene. The dashed line approximates the boundary of the deletion given by the clinical microarray report (=158,405,216); therefore, the variant to the farthest left is outside the deleted region (and the proband is heterozygous at this location). Image from VarSeq v1.3.4 GenomeBrowse® visualization tool, by Golden Helix, Inc.

One variant in *GFM1* (rs34920045) was identified as heterozygous for the mother and not found in the proband. The remaining eight variants were identified as heterozygous for the mother and found in the proband, summarized in **Table 5**. These were most likely inherited from the mother; however, since the father was not included in the analysis (and his deletion was reported but not confirmed), this cannot be determined absolutely. All of the variants identified for the proband in this region passed the VQSR filter. For the mother, the alternate allele frequencies ranged from ~0.23-0.70, while the alternate allele frequencies for the proband ranged from 0.95-1. This is consistent with

the proband being hemizygous with this region, though the proband was called as homozygous for these eight variants by the sequencing analysis.

	Gene	Variant	Identifier	Freq in ExAC	Alternate Allele Freq:	
					Proband	Mother
MISSENSE MUTATIONS	GFM1	NM_024996.5:c.1990G>A	rs62288347	0.018	1.00	0.51
	MFSD1	NM_022736.2:c.217C>T	rs28364680	0.135	1.00	0.46
	MFSD1	NM_022736.2:c.805A>G	rs3765083	0.561	1.00	0.70
NON CODING CHANGES	GFM1	NM_024996.5:c.2125-32_2125-31insA	rs145803239	0.363	0.95	0.36
	GFM1	NM_024996.5:c.*6C>T	rs1047355	0.366	1.00	0.50
	RARRES1	NM_206963.1:c.420C>T	rs2307064	0.194	1.00	0.43
SYNONYMOUS MUTATIONS	RARRES1	NM_206963.1:c.240A>G	rs6441223	0.529	1.00	0.23
	RARRES1	NM_206963.1:c.27T>C	rs6786423	0.389	1.00	0.32

Table 5: Variants within chromosome 3 deletion for AU0237-0201. Summary of the eight variants found in the proband within the deleted region identified on a clinical microarray. The frequency of each variant from the ExAC database is given (exac.broadinstitute.org; Lek *et al.*, 2016). The frequencies of the alternate allele sequencing reads for the samples of the proband and her mother are given. The mother was called heterozygous for all of these variants while the proband was called homozygous for the variants; however, given the clinical microarray showing a deletion in this region, she is presumed to be hemizygous.

Of these eight variants, all but one are common, with minor allele frequencies in the ExAC database above ten percent (range= 0.135-0.529). Two (rs145803239 and rs1047355) are in non-coding regions of *GFM1*, so their effect is difficult to predict. Of the remaining variants in the coding regions, three (rs2307064, rs6441223, and rs6786423) are synonymous changes in *RARRES1,* which would not be predicted to result in an amino acid substitution in the protein produced. The remaining three missense mutations include two in *MFSD1* and one in *GFM1*. They are described in more detail below.

MFSD1

There were two variants in *MFSD1* predicted to cause missense mutations (rs28364680 and rs3765083) identified in both the proband and her mother. The mother was heterozygous for both, and the proband is predicted to be hemizygous due to the deletion identified in the clinical microarray.

The first (rs28364680; NM_022736.2:c.217C>T) is found at a frequency of 0.135 (13,634/101,376) in the ExAC database. Approximately nine hundred homozygotes were identified in the database. It is predicted to result in the substitution of a serine for a proline in the amino acid sequence of the protein. This is predicted to be a neutral change by Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The second (rs3765083; NM_022736.2:c.805A>G) is found at a frequency of 0.561 (66,592/118,780) in the ExAC database. Approximately nineteen thousand homozygotes were identified in the database. It is predicted to result in the substitution of a valine for an isoleucine in the amino acid sequence of the protein. This change is located in one of the transmembrane domains of the protein (uniprot.org; Choi *et al.*, 2012) but is predicted to be a neutral change by Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The role of *MFSD1* is not well understood, and while the RNA is found in low levels in brain tissues, it is broadly expressed in many others (proteinatlas.org; Uhlen *et al.*, 2015). The potential for these variants to contribute to this proband's ASD phenotype is uncertain, but not very convincing due to their relatively high frequency in the population and predicted lack of effect on protein structure.

GFM1

There was one variant in *GFM1* predicted to cause missense mutations (rs62288347) identified in both the proband and her mother. The mother was heterozygous, and the proband is predicted to be hemizygous due to the deletion identified in the clinical microarray.

The variant (rs62288347; NM_024996.5:c.1990G>A) is found at a frequency of 0.018 (2,148/121,398) in the ExAC database. Thirty-six homozygotes were identified in the database. It is predicted to result in the substitution of an isoleucine for a valine in the amino acid sequence of the protein. This is predicted to be a neutral change by Condel (bg.upf.edu/fannsdb; Gonzalez-Perez and Lopez-Bigas, 2011).

The *GFM1* gene codes for a GTPase that is involved in the elongation phase of translation in the mitochondria (Hammarsund *et al.*, 2001). The protein is found in brain tissues, but is also found broadly expressed in many others (proteinatlas.org; Uhlen *et al.*, 2015). Mutations in *GFM1* have been associated with a rare autosomal recessive oxidative phosphorylation deficiency (Coenen *et al.*, 2004; Ravn *et al.*, 2015; Valente *et al.*, 2007). Mutations in the peripheral regions of the protein, similar to the one identified in the proband, are more likely to be associated with brain findings, including encephalopathy compared to those in the central part associated with hepatic failure (Galmiche *et al.*, 2012). However, the specific variant identified in this proband and her mother is listed by a clinical testing laboratory in the public database, ClinVar, as benign (SCV000168668.5; Landrum *et al.*, 2016). It is possible that this variant contributes to a much more subtle phenotype than that seen in the autosomal recessive condition described; however, this would require additional studies.

DISCUSSION

The Neurobiology and Genetics of Autism study at University of California, Irvine is collecting an array of phenotype information about individuals with ASD that will eventually make a powerful dataset for investigating the genetic etiologies of a diverse and deeply phenotyped cohort. An initial analysis of the first batch of whole exome sequencing sent from this cohort served as the foundation for this thesis. The two concentrations of that analysis are (i) identifying potential *de novo* variants in probands who were submitted with a trio and (ii) assessing variants within a chromosomal region identified as a deletion on a clinical microarray for one proband. These two highlight some of the benefits of broad genomic testing for individuals with ASD, as well as many of the limitations.

AIM I: FILTERING AND CHARACTERIZATION OF POTENTIAL *DE NOVO* VARIANTS

For the six probands submitted as part of a trio, a workflow was designed to identify those variants which are likely accurate, truly *de novo*, and potentially able to contribute to ASD phenotypes. However at every filter, there is a possibility of eliminating accurate variants of interest.

For example, filters for quality of the variants were set relatively stringently (e.g. VQSR filter PASS). As noted above, variants in *MUC* family genes were removed due their high potential for false positives, and variants in *HLA* genes were removed because of

high expected variation. There are likely variants of interest that were not prioritized out of caution using these filters. Further analysis could be done to look at the variants removed by these filters, possibly by accepting a lower quality score and filtering for very rare variants or particularly damaging variants instead. Despite these measures, one variant (*THEGL*, NM_001256475.1:c.706C>T) was filtered at the "Other" stage because it was found in two probands who are sisters. This may be due to mosaicism or an erroneous genotype call for one of the parents. This demonstrates that having additional family samples can improve the quality of the analysis, and that the prioritized variants identified require additional confirmatory testing (e.g. Sanger sequencing).

De novo variants were the focus of this analysis, based on research showing the importance of rare *de novo* variants as significant contributors to ASD, especially in simplex families (De Rubeis *et al.*, 2014; Iossifov *et al.*, 2014; Neale *et al.*, 2012; O'Roak *et al.*, 2012; Sanders *et al.*, 2012). However, the contribution of common variants is also seen with individually small effects that can have large combined impacts (Anney *et al.*, 2012; Gaugler *et al.*, 2014; Oikonomakis *et al.*, 2016), and likely to be more important in multiplex families. Recent studies estimate that the contribution of common variants to heritability is at least 41%, while rare variants contribute at most 17% (Gaugler *et al.*, 2014), and inherited mutations significantly contribute to ASD susceptibility as well (Krumm *et al.*, 2015). However, these mechanisms are not mutually exclusive and are affected by epigenetic and environmental factors as well (Abrahams and Geschwind, 2008; Geschwind, 2011). Overall, *de novo* mutations have

the ability to contribute to ASD phenotypes, but they are only one factor in a complicated disorder.

Identifying *de novo* variants requires having a sample from both the proband and parents for a trio. This allows for a more unbiased analysis; notably, *de novo* mutations in novel genes can be investigated. Trios have a significantly higher yield overall for clinical WES and have added benefits of reduced analytic costs and turnaround time (Farwell *et al.*, 2015; Lee *et al.*, 2014). However, as a practical matter in the clinic, trios are not always a possibility due to parents being unavailable or unwilling to submit a sample. There are also additional ethical concerns that come along with submitting a trio, including identifying incidental findings in unaffected individuals, non-paternity, non-maternity, or consanguinity. When available, trio analysis is preferable to improve diagnostic yield but requires proper consenting.

COMMON PATHWAYS

The resulting prioritized presumably *de novo* variants ranged in their ability to affect protein function and their potential for contributing to ASD phenotypes. None of the variants described meet the criteria put forward by the American College of Medical Genetics and Genomics (ACMG) for interpreting and reporting sequence variants as pathogenic in clinical laboratories (Richards *et al.*, 2015). Still, there was some overlap in common pathways amongst even this small dataset. For example, variants in genes involved in classical axon guidance cues (e.g. *UNC5B*, *EPHB2*, *SLIT2*), as well as morphogen signaling pathways that can affect axon guidance (e.g. *SUFU*, *BMPR1B*),

were identified in several probands. With the extraordinary etiological heterogeneity that is seen in ASD, identifying commonalities including molecular pathways and developmental processes can unify the etiology of some of that complexity (Berg and Geschwind, 2012). Previous studies have identified some common pathways in ASD etiology, including abnormal neuronal migration, axon pathfinding, synaptic function, and dendritic maturation (Geschwind, 2011). Another study analyzing groups of interacting genes that could contribute to ASD identified a circuit of genes with enhanced expression in the corpus callosum (Li *et al.*, 2014). That region of the brain is important for mediating signals between the two hemispheres and is reduced in size in individuals with ASD (Egaas *et al.*, 1995). Variants in different genes that impact the connectivity of critical brain circuits within the brain can be put into context with phenotype information, including brain pathology and behavioral assessments.

AIM II: TARGETED ANALYSIS OF PROBAND BASED ON CLINICAL FINDINGS

The sequencing results for proband AU0237-0201 were investigated to determine the potential for the deletion on chromosome three from her clinical microarray to contribute to her ASD phenotype. Chromosomal microarrays are regularly ordered in the clinic for individuals with ASD (Miller *et al.*, 2010), and approximately 10-20 percent of the time they identify a copy number variant (Miller *et al.*, 2010; Schaefer and Mendelsohn, 2013; Shen *et al.*, 2010). Some of these copy number variations will focus clinical attention on regions that may indicate an etiology for that individual's ASD phenotype. However, specific sequence information is sometimes needed to interpret the potential effect.

Genomic sequencing, like the WES for this research project, is available clinically but is not yet standard of care for individuals with ASD like chromosomal microarray. The two technologies can be complementary, but cannot replace each other. For example, the microarray for this proband suggested a region of interest by identifying the deletion. However, the phenotype associated with mutations in *GFM1* has been associated with autosomal recessive inheritance (Coenen *et al.*, 2004; Valente *et al.*, 2007) . Therefore, identifying if there were any mutations in the maternally-inherited copy of *GFM1* was an important prerequisite to understanding the potential of the deletion to affect the proband's ASD phenotype. While this deletion only included three genes, for other individuals a microarray will identify copy number variations in more genes or in multiple regions of the genome. When the regions involved are large or dispersed, genome-wide sequencing, like WES, may be the most efficient way to assess the effect of the copy number variation(s).

In this case of this proband and for many others, the copy number variation involves both genes and intergenic regions. Of the approximately 371kb deletion, 303kb are outside the three genes. Variation in intergenic regions as well as intronic regions that are not targeted by WES could still be involved in the regulation of these or other genes and contribute to an individual's ASD phenotype. In fact, studies have identified rare copy number variations of intergenic regions in individuals with ASD, especially near genes associated with ASD (Walker and Scherer, 2013). However, predicting those

effects *a priori* is more difficult given a general lack of knowledge about how CNVs in non-coding regions assert their effects.

Neither of these technologies, chromosomal microarray or WES, can define the breakpoint of the deletion precisely. The resolution of a single nucleotide polymorphism (SNP) microarray is defined by the density of the SNPs used, and specifically for the clinical microarray for this proband, the resolution is 1.15kb. The WES technology selectively analyzes the sequences of the exome portion of the genome; therefore, it does not provide additional information about the breakpoint of the deletion which is in an uncovered intronic region. The WES data did support the deletion, with the proband being apparently homozygous for the variants detected in the region and heterozygous for the first variant upstream of the estimated breakpoint. Whole genome sequencing would help refine the breakpoints by identifying and mapping the split reads.

Overall, the analysis of the variants within the region that was identified in the deletion decreased the suspicion that they contribute to the proband's ASD phenotype. The variants were all relatively common in the population with additional evidence supporting their lack of severity on protein structure and function. This analysis did highlight the complementary nature of the different testing technologies. Chromosomal microarrays identify CNVs. The clinical SNP microarray ordered for this proband could detect deletions larger than 50kb and duplications larger than 200kb. WES can detect nucleotide sequence level changes in the regions targeted, including insertions and deletions that can be seen within a mapped read. Algorithms to identify CNVs using

next generation sequencing read depth have been developed but are not yet as sensitive as a chromosomal microarray (de Ligt *et al.*, 2013; Samarakoon *et al.*, 2014). In addition, neither of these technologies would identify the trinucleotide repeats that are causative of fragile X syndrome. Recognizing the benefits and limitations of each genetic testing technology available remains an important contribution of genetics professionals in the clinic.

LIMITATIONS TO THE STUDY AND WES ANALYSIS

While some of the limitations of the specific aims are discussed above, there are additional concerns that affect the analysis as a whole. The number of individuals involved in the Neurobiology and Genetics of Autism study as a whole is limited, especially for a disorder as heterogeneous as ASD. Only a subset of those individuals was included in the WES. This makes internal comparisons unsatisfactory. Fortunately, there are large databases, both specific to ASD (like SFARI) and not (like ExAC), which make it possible to compare our results to the larger sample sizes, and large-scale studies are continuing to add to our understanding of ASD. In fact, Simons Foundation Powering Autism Research for Knowledge (SPARK) was recently launched to recruit tens of thousands of individuals with ASD and their families to share genetic information. However, there is also power in having smaller studies, like the one that this thesis contributes to, which collect detailed phenotypic information that can inform genetic analysis.

At this time, the genomic analysis done for this thesis did not incorporate much phenotypic information about the probands. Besides the sex of the proband and assessment scores (for ASD and IQ), the analysis was done without reference to the proband's phenotype. Eventually this information will be helpful for analyzing the genomic information. For example, the variant found for AU0243-0201 in *BMPR1B* could affect the proband's ASD phenotype, but it has also been associated with brachydactyly (Lehmann *et al.*, 2003). Other variants can be prioritized based on known overlap of mutations in genes with the proband's phenotypic presentation. Clinically-available whole exome testing benefits from detailed clinical observations (Lee *et al.*, 2014). While it was outside the scope of this thesis, incorporating the phenotypic information collected will improve the ability to understand the genetic information.

The samples were sent to the laboratory at the Broad Institute for whole exome sequencing. This technology uses next generation sequencing to target the coding portions of the genome. While coding regions make up the minority of the genome (~1%), most disease-causing mutations are found in these regions (Majewski *et al.*, 2011; Ng *et al.*, 2009). As described in the context of the deletion assessed in Aim II, the non-coding regions are likely to contribute to disease phenotypes, including ASD; however, the effects of mutations in these regions are more difficult to predict. As more data about variation in non-coding regions is accumulated, it will be easier to interpret its effects.

CONCLUSIONS

Despite these limitations, we were able to identify many interesting *de novo* variants within the probands that warrant further investigation. We were also able to decrease suspicion that the deletion identified in the clinical microarray for AU0237-0201 contributes to her ASD phenotype by analyzing the variants she inherited from her mother. The most interesting *de novo* variants identified can be confirmed and studied further through functional assays to understand their effects.

In addition to following up on these findings, there remain a multitude of ways to analyze the WES sequencing data to identify other variants of interest. For example, inherited variants that result in a compound heterozygote state for the proband as well as variants in the probands that are very rare in the population will yield additional variants of interest. This initial analysis will be incorporated into the larger body of knowledge about the etiology of ASD, and our improved understanding of this heterogeneous condition will allow us to serve individuals with ASD and their families better with earlier interventions, more effective treatments, and personalized recurrence risk estimates.

REFERENCES

Abrahams, B.S., and Geschwind, D.H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet *9*, 341-355.

ADDMN (2014). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill Summ *63*, 1-21.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C*., et al.* (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res.

Al-Ayadhi, L.Y. (2012). Relationship between Sonic hedgehog protein, brain-derived neurotrophic factor and oxidative stress in autism spectrum disorders. Neurochem Res *37*, 394-400.

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 disorderspi. Hum Mol Genet *21*, 4781-4792.

APA, D.-I. (1980). Diagnostic and statistical manual of mental disorders, 3rd edn (Washington, D.C., American Psychiatric Association).

APA, D.-V. (2013). Diagnostic and statistical manual of mental disorders, 5th edn (Washington, D.C., American Psychiatric Association).

Asperger, H. (1944). Die „Autistischen Psychopathen" im Kindesalter. Archiv für Psychiatrie und Nervenkrankheiten *117*, 76-136.

Astrom, A.K., Jin, D., Imamura, T., Roijer, E., Rosenzweig, B., Miyazono, K., ten Dijke, P., and Stenman, G. (1999). Chromosomal localization of three human genes encoding bone morphogenetic protein receptors. Mamm Genome *10*, 299-302.

Auger, M.L., Schmidt, E.R., Manitt, C., Dal-Bo, G., Pasterkamp, R.J., and Flores, C. (2013). unc5c haploinsufficient phenotype: striking similarities with the dcc haploinsufficiency model. Eur J Neurosci *38*, 2853-2863.

Baghdadli, A., Pascal, C., Grisi, S., and Aussilloux, C. (2003). Risk factors for self-injurious behaviours among 222 young children with autistic disorders. J Intellect Disabil Res *47*, 622-627.

Bagri, A., Marin, O., Plump, A.S., Mak, J., Pleasure, S.J., Rubenstein, J.L., and Tessier-Lavigne, M. (2002). Slit proteins prevent midline crossing and determine the dorsoventral position of major axonal pathways in the mammalian forebrain. Neuron *33*, 233-248.

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

Baker, P., Piven, J., and Sato, Y. (1998). Autism and tuberous sclerosis complex: prevalence and clinical features. J Autism Dev Disord *28*, 279-285.

Baker, P., Piven, J., Schwartz, S., and Patil, S. (1994). Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. J Autism Dev Disord *24*, 529-535.

Banerjee-Basu, S., and Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. Dis Model Mech *3*, 133-135.

Berg, J.M., and Geschwind, D.H. (2012). Autism genetics: searching for specificity and convergence. Genome Biol *13*, 247.

Berwick, D.C., Calissano, M., Corness, J.D., Cook, S.J., and Latchman, D.S. (2009). Regulation of Brn-3a N-terminal transcriptional activity by MEK1/2-ERK1/2 signalling in neural differentiation. Brain Res *1256*, 8-18.

Binder, M.D., and Kilpatrick, T.J. (2009). TAM receptor signalling and demyelination. Neurosignals *17*, 277-287.

Boisson, B., Laplantine, E., Dobbs, K., Cobat, A., Tarantino, N., Hazen, M., Lidov, H.G., Hopkins, G., Du, L., Belkadi, A.*, et al.* (2015). Human HOIP and LUBAC deficiency underlies autoinflammation, immunodeficiency, amylopectinosis, and lymphangiectasia. J Exp Med *212*, 939-951.

Borg, I., Freude, K., Kubart, S., Hoffmann, K., Menzel, C., Laccone, F., Firth, H., Ferguson-Smith, M.A., Tommerup, N., Ropers, H.H.*, et al.* (2005). Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. Eur J Hum Genet *13*, 921-927.

Brose, K., Bland, K.S., Wang, K.H., Arnott, D., Henzel, W., Goodman, C.S., Tessier-Lavigne, M., and Kidd, T. (1999). Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. Cell *96*, 795-806.

Brugieres, L., Pierron, G., Chompret, A., Paillerets, B.B., Di Rocco, F., Varlet, P., Pierre-Kahn, A., Caron, O., Grill, J., and Delattre, O. (2010). Incomplete penetrance of the predisposition to medulloblastoma associated with germ-line SUFU mutations. J Med Genet *47*, 142-144.

Budhram-Mahadeo, V., Parker, M., and Latchman, D.S. (1998). POU transcription factors Brn-3a and Brn-3b interact with the estrogen receptor and differentially regulate transcriptional activity via an estrogen response element. Mol Cell Biol *18*, 1029-1041.

Chakrabarti, S., and Fombonne, E. (2005). Pervasive developmental disorders in preschool children: confirmation of high prevalence. Am J Psychiatry *162*, 1133-1141.

Chaturvedi, P., Singh, A.P., and Batra, S.K. (2008). Structure, evolution, and biology of the MUC4 mucin. FASEB J *22*, 966-981.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS One *7*, e46688.

Coenen, M.J., Antonicka, H., Ugalde, C., Sasarman, F., Rossi, R., Heister, J.G., Newbold, R.F., Trijbels, F.J., van den Heuvel, L.P., Shoubridge, E.A.*, et al.* (2004). Mutant mitochondrial

elongation factor G1 and combined oxidative phosphorylation deficiency. N Engl J Med *351*, 2080-2086.

Coffee, B., Keith, K., Albizua, I., Malone, T., Mowrey, J., Sherman, S.L., and Warren, S.T. (2009). Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA. Am J Hum Genet *85*, 503-514.

Cohen, D., Pichard, N., Tordjman, S., Baumann, C., Burglen, L., Excoffier, E., Lazar, G., Mazet, P., Pinquier, C., Verloes, A*., et al.* (2005). Specific genetic disorders and autism: clinical contribution towards their identification. J Autism Dev Disord *35*, 103-116.

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S.R., Woodhouse, E., Gillan, N., Hallett, V., Lietz, S., Garnett, T*., et al.* (2015). Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. JAMA Psychiatry *72*, 415-423.

Consortium, U. (2015). UniProt: a hub for protein information. Nucleic Acids Research *43*, D204-D212.

Constantino, J.N., Zhang, Y., Frazier, T., Abbacchi, A.M., and Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. Am J Psychiatry *167*, 1349-1356.

Correia, C.T., Almeida, J.P., Santos, P.E., Sequeira, A.F., Marques, C.E., Miguel, T.S., Abreu, R.L., Oliveira, G.G., and Vicente, A.M. (2010). Pharmacogenetics of risperidone therapy in autism: association analysis of eight candidate genes with drug efficacy and adverse drug reactions. Pharmacogenomics J *10*, 418-430.

Council on Children With Disabilities, S.o.D.B.P., Bright Futures Steering Committee and Medical Home Initiatives for Children With Special Needs Project Advisory Committee (2006). Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics *118*, 405-420.

Cukier, H.N., Dueker, N.D., Slifer, S.H., Lee, J.M., Whitehead, P.L., Lalanne, E., Leyva, N., Konidari, I., Gentry, R.C., Hulme, W.F*., et al.* (2014). Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. Mol Autism *5*, 1.

Curatolo, P., Porfirio, M.C., Manzi, B., and Seri, S. (2004). Autism in tuberous sclerosis. Eur J Paediatr Neurol *8*, 327-332.

Dawson, G., and Bernier, R. (2013). A quarter century of progress on the early detection and treatment of autism spectrum disorder. Dev Psychopathol *25*, 1455-1472.

de Ligt, J., Boone, P.M., Pfundt, R., Vissers, L.E., Richmond, T., Geoghegan, J., O'Moore, K., de Leeuw, N., Shaw, C., Brunner, H.G*., et al.* (2013). Detection of clinically relevant copy number variants with whole-exome sequencing. Hum Mutat *34*, 1439-1448.

De Rubeis, S., and Buxbaum, J.D. (2015). Genetics and genomics of autism spectrum disorder: embracing complexity. Hum Mol Genet *24*, R24-31.

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.

Demirhan, O., Turkmen, S., Schwabe, G.C., Soyupak, S., Akgul, E., Tastemir, D., Karahan, D., Mundlos, S., and Lehmann, K. (2005). A homozygous BMPR1B mutation causes a new subtype of acromesomelic chondrodysplasia with genital anomalies. J Med Genet *42*, 314-317.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M.*, et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491-498.

DiLavore, P.C., Lord, C., and Rutter, M. (1995). The pre-linguistic autism diagnostic observation schedule. J Autism Dev Disord *25*, 355-379.

DSM-V, A. (2013). Diagnostic and statistical manual of mental disorders: DSM-V, 5th edn (Washington, D.C., American Psychiatric Association).

Duman-Scheel, M. (2009). Netrin and DCC: axon guidance regulators at the intersection of nervous system development and cancer. Curr Drug Targets *10*, 602-610.

Egaas, B., Courchesne, E., and Saitoh, O. (1995). Reduced size of corpus callosum in autism. Arch Neurol *52*, 794-801.

Farwell, K.D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E.C., Tippin Davis, B., Baxter, R.M., Zeng, W., Mroske, C., Parra, M.C.*, et al.* (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med *17*, 578-586.

Fatemi, S.H., Aldinger, K.A., Ashwood, P., Bauman, M.L., Blaha, C.D., Blatt, G.J., Chauhan, A., Chauhan, V., Dager, S.R., Dickson, P.E.*, et al.* (2012). Consensus paper: pathological role of the cerebellum in autism. Cerebellum *11*, 777-807.

Fernandez, B.A., Roberts, W., Chung, B., Weksberg, R., Meyn, S., Szatmari, P., Joseph-George, A.M., Mackay, S., Whitten, K., Noble, B.*, et al.* (2010). Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. J Med Genet *47*, 195-203.

Filipek, P.A., Accardo, P.J., Ashwal, S., Baranek, G.T., Cook, E.H., Jr., Dawson, G., Gordon, B., Gravel, J.S., Johnson, C.P., Kallen, R.J.*, et al.* (2000). Practice parameter: screening and diagnosis of autism: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. Neurology *55*, 468-479.

Folstein, S., and Rutter, M. (1977). Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry *18*, 297-321.

Fombonne, E., Roge, B., Claverie, J., Courty, S., and Fremolle, J. (1999). Microcephaly and macrocephaly in autism. J Autism Dev Disord *29*, 113-119.

Forcet, C., Stein, E., Pays, L., Corset, V., Llambi, F., Tessier-Lavigne, M., and Mehlen, P. (2002). Netrin-1-mediated axon outgrowth requires deleted in colorectal cancer-dependent MAPK activation. Nature *417*, 443-447.

Galmiche, L., Serre, V., Beinat, M., Zossou, R., Assouline, Z., Lebre, A.S., Chretien, F., Shenhav, R., Zeharia, A., Saada, A*., et al.* (2012). Toward genotype phenotype correlations in GFM1 mutations. Mitochondrion *12*, 242-247.

Gargus, J., and Imtiaz, F. (2008). Mitochondrial Energy-Deficient Endophenotype in Autism. American Journal of Biochemistry and Biotechnology *4*, 198-207.

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

Geschwind, D.H. (2011). Genetics of autism spectrum disorders. Trends Cogn Sci *15*, 409-416.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., and Rapoport, J.L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. Nat Neurosci *2*, 861-863.

Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P*., et al.* (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature *459*, 569-573.

Glessner, J.T., Wang, K., Sleiman, P.M., Zhang, H., Kim, C.E., Flory, J.H., Bradfield, J.P., Imielinski, M., Frackelton, E.C., Qiu, H*., et al.* (2010). Duplication of the SLIT3 locus on 5q35.1 predisposes to major depressive disorder. PLoS One *5*, e15463.

Golden Helix, I. (2016). VarSeq[TM] (Bozeman, MT, Golden Helix, Inc.).

Gonzalez-Perez, A., and Lopez-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet *88*, 440-449.

Green, E. (2015). Rewriting Autism History. In HEALTH (The Atlantic).

Gronborg, T.K., Schendel, D.E., and Parner, E.T. (2013). Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. JAMA Pediatr *167*, 947-953.

Guenebeaud, C., Goldschneider, D., Castets, M., Guix, C., Chazot, G., Delloye-Bourgeois, C., Eisenberg-Lerner, A., Shohat, G., Zhang, M., Laudet, V*., et al.* (2010). The dependence receptor UNC5H2/B triggers apoptosis via PP2A-mediated dephosphorylation of DAP kinase. Mol Cell *40*, 863-876.

Hammarsund, M., Wilson, W., Corcoran, M., Merup, M., Einhorn, S., Grander, D., and Sangfelt, O. (2001). Identification and characterization of two novel human mitochondrial elongation factor genes, hEFG2 and hEFG1, phylogenetically conserved through evolution. Hum Genet *109*, 542-550.

Hammond, R., Vivancos, V., Naeem, A., Chilton, J., Mambetisaeva, E., Andrews, W., Sundaresan, V., and Guthrie, S. (2005). Slit-mediated repulsion is a key regulator of motor axon pathfinding in the hindbrain. Development *132*, 4483-4495.

Hatton, D.D., Sideris, J., Skinner, M., Mankowski, J., Bailey, D.B., Jr., Roberts, J., and Mirrett, P. (2006). Autistic behavior in children with fragile X syndrome: prevalence, stability, and the impact of FMRP. Am J Med Genet A *140A*, 1804-1813.

Henkemeyer, M., Orioli, D., Henderson, J.T., Saxton, T.M., Roder, J., Pawson, T., and Klein, R. (1996). Nuk controls pathfinding of commissural axons in the mammalian central nervous system. Cell *86*, 35-46.

Hong, K., Hinck, L., Nishiyama, M., Poo, M.M., Tessier-Lavigne, M., and Stein, E. (1999). A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. Cell *97*, 927-941.

Hutsler, J.J., and Zhang, H. (2010). Increased dendritic spine densities on cortical projection neurons in autism spectrum disorders. Brain Res *1309*, 83-94.

Ikeda, F., Deribe, Y.L., Skanland, S.S., Stieglitz, B., Grabbe, C., Franz-Wachtel, M., van Wijk, S.J., Goswami, P., Nagy, V., Terzic, J.*, et al.* (2011). SHARPIN forms a linear ubiquitin ligase complex regulating NF-kappaB activity and apoptosis. Nature *471*, 637-641.

Ikeda, M., Tomita, Y., Mouri, A., Koga, M., Okochi, T., Yoshimura, R., Yamanouchi, Y., Kinoshita, Y., Hashimoto, R., Williams, H.J.*, et al.* (2010). Identification of novel candidate genes for treatment response to risperidone and susceptibility for schizophrenia: integrated analysis among pharmacogenomics, mouse expression, and genetic case-control association approaches. Biol Psychiatry *67*, 263-269.

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.

Irie, F., and Yamaguchi, Y. (2002). EphB receptors regulate dendritic spine development via intersectin, Cdc42 and N-WASP. Nat Neurosci *5*, 1117-1118.

Iyer, S., and Acharya, K.R. (2011). Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. FEBS J *278*, 4304-4322.

Jeste, S.S., Sahin, M., Bolton, P., Ploubidis, G.B., and Humphrey, A. (2008). Characterization of autism in young children with tuberous sclerosis complex. J Child Neurol *23*, 520-525.

Ji, R., Meng, L., Li, Q., and Lu, Q. (2015). TAM receptor deficiency affects adult hippocampal neurogenesis. Metab Brain Dis *30*, 633-644.

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

Kayser, M.S., Lee, A.C., Hruska, M., and Dalva, M.B. (2011). Preferential control of basal dendritic protrusions by EphB2. PLoS One *6*, e17417.

Keleman, K., and Dickson, B.J. (2001). Short- and long-range repulsion by the Drosophila Unc5 netrin receptor. Neuron *32*, 605-617.

Kleinman, J.M., Ventola, P.E., Pandey, J., Verbalis, A.D., Barton, M., Hodgson, S., Green, J., Dumont-Mathieu, T., Robins, D.L., and Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. J Autism Dev Disord *38*, 606-615.

Komatsuzaki, K., Dalvin, S., and Kinane, T.B. (2002). Modulation of G(ialpha(2)) signaling by the axonal guidance molecule UNC5H2. Biochem Biophys Res Commun *297*, 898-905.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Wong, W.S*., et al.* (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature *488*, 471-475.

Krenciute, G., Liu, S., Yucer, N., Shi, Y., Ortiz, P., Liu, Q., Kim, B.J., Odejimi, A.O., Leng, M., Qin, J*., et al.* (2013). Nuclear BAG6-UBL4A-GET4 complex mediates DNA damage signaling and cell death. J Biol Chem *288*, 20547-20557.

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.* (2015). Excess of rare, inherited truncating mutations in autism. Nat Genet *47*, 582-588.

Krzewinski-Recchi, M.A., Julien, S., Juliant, S., Teintenier-Lelievre, M., Samyn-Petit, B., Montiel, M.D., Mir, A.M., Cerutti, M., Harduin-Lepers, A., and Delannoy, P. (2003). Identification and functional expression of a second human beta-galactoside alpha2,6-sialyltransferase, ST6Gal II. Eur J Biochem *270*, 950-961.

Lainhart, J.E., Piven, J., Wzorek, M., Landa, R., Santangelo, S.L., Coon, H., and Folstein, S.E. (1997). Macrocephaly in children and adults with autism. J Am Acad Child Adolesc Psychiatry *36*, 282-290.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J*., et al.* (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res *44*, D862-868.

Latchman, D.S. (1998). The Brn-3a transcription factor. Int J Biochem Cell Biol *30*, 1153-1157.

Lee, H., Deignan, J.L., Dorrani, N., Strom, S.P., Kantarci, S., Quintero-Rivera, F., Das, K., Toy, T., Harry, B., Yourshaw, M*., et al.* (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA *312*, 1880-1887.

Lee, J., Son, M.J., Woolard, K., Donin, N.M., Li, A., Cheng, C.H., Kotliarova, S., Kotliarov, Y., Walling, J., Ahn, S*., et al.* (2008). Epigenetic-mediated dysfunction of the bone morphogenetic protein pathway inhibits differentiation of glioblastoma-initiating cells. Cancer Cell *13*, 69-80.

Lehmann, K., Seemann, P., Stricker, S., Sammar, M., Meyer, B., Suring, K., Majewski, F., Tinschert, S., Grzeschik, K.H., Muller, D*., et al.* (2003). Mutations in bone morphogenetic protein receptor 1B cause brachydactyly type A2. Proc Natl Acad Sci U S A *100*, 12277-12282.

Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B.*, et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv.

Lemmon, M.E., Gregas, M., and Jeste, S.S. (2011). Risperidone use in autism spectrum disorders: a retrospective review of a clinic-referred patient population. J Child Neurol *26*, 428-432.

Lenhard, W., Breitenbach, E., Ebert, H., Schindelhauer-Deutscher, H.J., and Henn, W. (2005). Psychological benefit of diagnostic certainty for mothers of children with disabilities: lessons from Down syndrome. Am J Med Genet A *133A*, 170-175.

Leonardo, E.D., Hinck, L., Masu, M., Keino-Masu, K., Fazeli, A., Stoeckli, E.T., Ackerman, S.L., Weinberg, R.A., and Tessier-Lavigne, M. (1997). Guidance of developing axons by netrin-1 and its receptors. Cold Spring Harb Symp Quant Biol *62*, 467-478.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

Li, H.S., Chen, J.H., Wu, W., Fagaly, T., Zhou, L., Yuan, W., Dupuis, S., Jiang, Z.H., Nash, W., Gick, C.*, et al.* (1999). Vertebrate slit, a secreted ligand for the transmembrane protein roundabout, is a repellent for olfactory bulb axons. Cell *96*, 807-818.

Li, J., Shi, M., Ma, Z., Zhao, S., Euskirchen, G., Ziskin, J., Urban, A., Hallmayer, J., and Snyder, M. (2014). Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. Mol Syst Biol *10*, 774.

Lionel, A.C., Tammimies, K., Vaags, A.K., Rosenfeld, J.A., Ahn, J.W., Merico, D., Noor, A., Runke, C.K., Pillalamarri, V.K., Carter, M.T.*, et al.* (2014). Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. Hum Mol Genet *23*, 2752-2768.

Liu, S., Yin, F., Fan, W., Wang, S., Guo, X.R., Zhang, J.N., Tian, Z.M., and Fan, M. (2012). Over-expression of BMPR-IB reduces the malignancy of glioblastoma cells by upregulation of p21 and p27Kip1. J Exp Clin Cancer Res *31*, 52.

Lombard, J. (1998). Autism: a mitochondrial disorder? Med Hypotheses *50*, 497-500.

Lord, C., Cook, E.H., Leventhal, B.L., and Amaral, D.G. (2000). Autism spectrum disorders. Neuron *28*, 355-363.

Lord, C., Risi, S., DiLavore, P.S., Shulman, C., Thurm, A., and Pickles, A. (2006). Autism from 2 to 9 years of age. Arch Gen Psychiatry *63*, 694-701.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., and Schopler, E. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord *19*, 185-212.

Lu, X., Le Noble, F., Yuan, L., Jiang, Q., De Lafarge, B., Sugiyama, D., Breant, C., Claes, F., De Smet, F., Thomas, J.L.*, et al.* (2004). The netrin receptor UNC5B mediates guidance events controlling morphogenesis of the vascular system. Nature *432*, 179-186.

Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., and Jabado, N. (2011). What can exome sequencing do for you? J Med Genet *48*, 580-589.

Makela, N.L., Birch, P.H., Friedman, J.M., and Marra, C.A. (2009). Parental perceived value of a diagnosis for intellectual disability (ID): a qualitative comparison of families with and without a diagnosis for their child's ID. Am J Med Genet A *149A*, 2393-2402.

Mandell, D.S., Morales, K.H., Marcus, S.C., Stahmer, A.C., Doshi, J., and Polsky, D.E. (2008). Psychotropic medication use among Medicaid-enrolled children with autism spectrum disorders. Pediatrics *121*, e441-448.

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. Nature *466*, 1120-1124.

Matsunami, N., Hensel, C.H., Baird, L., Stevens, J., Otterud, B., Leppert, T., Varvil, T., Hadley, D., Glessner, J.T., Pellegrino, R.*, et al.* (2014). Identification of rare DNA sequence variants in high-risk autism families and their prevalence in a large case/control population. Mol Autism *5*, 5.

McClintock, K., Hall, S., and Oliver, C. (2003). Risk markers associated with challenging behaviours in people with intellectual disabilities: a meta-analytic study. J Intellect Disabil Res *47*, 405-416.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

Meffert, M.K., Chang, J.M., Wiltgen, B.J., Fanselow, M.S., and Baltimore, D. (2003). NF-kappa B functions in synaptic signaling and behavior. Nat Neurosci *6*, 1072-1078.

Miles, J.H. (2015). Complex Autism Spectrum Disorders and Cutting-Edge Molecular Diagnostic Tests. JAMA *314*, 879-880.

Miles, J.H., Hadden, L.L., Takahashi, T.N., and Hillman, R.E. (2000). Head circumference is an independent clinical finding associated with autism. Am J Med Genet *95*, 339-350.

Miles, J.H., Takahashi, T.N., Bagby, S., Sahota, P.K., Vaslow, D.F., Wang, C.H., Hillman, R.E., and Farmer, J.E. (2005). Essential versus complex autism: definition of fundamental prognostic subtypes. Am J Med Genet A *135*, 171-180.

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J.*, et al.* (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet *86*, 749-764.

Motohashi, T., Miyoshi, S., Osawa, M., Eyre, H.J., Sutherland, G.R., Matsuda, Y., Nakamura, Y., Shibuya, A., Iwama, A., and Nakauchi, H. (2000). Molecular cloning and chromosomal mapping of a novel five-span transmembrane protein gene, M83. Biochem Biophys Res Commun *276*, 244-250.

Muhle, R., Trentacoste, S.V., and Rapin, I. (2004). The genetics of autism. Pediatrics *113*, e472-486.

Mullen, E.M. (1995). Mullen Scales of Early Learning (Circle Pines, MN, American Guidance Services, Inc.).

Muramatsu, R., Nakahara, S., Ichikawa, J., Watanabe, K., Matsuki, N., and Koyama, R. (2010). The ratio of 'deleted in colorectal cancer' to 'uncoordinated-5A' netrin-1 receptors on the growth cone regulates mossy fibre directionality. Brain *133*, 60-75.

Narcisa, V., Discenza, M., Vaccari, E., Rosen-Sheidley, B., Hardan, A.Y., and Couchon, E. (2013). Parental interest in a genetic risk assessment test for autism spectrum disorders. Clin Pediatr (Phila) *52*, 139-146.

Neale, B.M., Kou, Y., Liu, L., Ma/'ayan, A., Samocha, K.E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L.-S., Makarov, V.*, et al.* (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242-245.

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.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E.*, et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature *461*, 272-276.

Noor, A., Whibley, A., Marshall, C.R., Gianakopoulos, P.J., Piton, A., Carson, A.R., Orlic-Milacic, M., Lionel, A.C., Sato, D., Pinto, D.*, et al.* (2010). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci Transl Med *2*, 49ra68.

Nord, A.S., Roeb, W., Dickel, D.E., Walsh, T., Kusenda, M., O'Connor, K.L., Malhotra, D., McCarthy, S.E., Stray, S.M., Taylor, S.M.*, et al.* (2011). Reduced transcript expression of genes affected by inherited and de novo CNVs in autism. Eur J Hum Genet *19*, 727-731.

Numis, A.L., Major, P., Montenegro, M.A., Muzykewicz, D.A., Pulsifer, M.B., and Thiele, E.A. (2011). Identification of risk factors for autism spectrum disorders in tuberous sclerosis complex. Neurology *76*, 981-987.

O'Connor, M.J. (2015). Targeting the DNA Damage Response in Cancer. Mol Cell *60*, 547-560.

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

Oikonomakis, V., Kosma, K., Mitrakos, A., Sofocleous, C., Pervanidou, P., Syrmou, A., Pampanos, A., Psoni, S., Fryssira, H., Kanavakis, E.*, et al.* (2016). Recurrent copy number variations as risk factors for autism spectrum disorders: analysis of the clinical implications. Clin Genet *89*, 708-718.

Ozonoff, S., Young, G.S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L.J., Constantino, J.N., Dobkins, K.*, et al.* (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics *128*, e488-495.

Palmieri, L., and Persico, A.M. (2010). Mitochondrial dysfunction in autism spectrum disorders: cause or effect? Biochim Biophys Acta *1797*, 1130-1137.

Panchision, D.M., and McKay, R.D. (2002). The control of neural stem cells by morphogenic signals. Curr Opin Genet Dev *12*, 478-487.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods *8*, 785-786.

Polimeni, M.A., Richdale, A.L., and Francis, A.J. (2005). A survey of sleep problems in autism, Asperger's disorder and typically developing children. J Intellect Disabil Res *49*, 260-268.

Pons, R., Andreu, A.L., Checcarelli, N., Vila, M.R., Engelstad, K., Sue, C.M., Shungu, D., Haggerty, R., de Vivo, D.C., and DiMauro, S. (2004). Mitochondrial DNA abnormalities and autistic spectrum disorders. J Pediatr *144*, 81-85.

Prasad, A., Merico, D., Thiruvahindrapuram, B., Wei, J., Lionel, A.C., Sato, D., Rickaby, J., Lu, C., Szatmari, P., Roberts, W.*, et al.* (2012). A discovery resource of rare copy number variations in individuals with autism spectrum disorder. G3 (Bethesda) *2*, 1665-1685.

Ravn, K., Schonewolf-Greulich, B., Hansen, R.M., Bohr, A.H., Duno, M., Wibrand, F., and Ostergaard, E. (2015). Neonatal mitochondrial hepatoencephalopathy caused by novel GFM1 mutations. Mol Genet Metab Rep *3*, 5-10.

Reddy, K.S. (2005). Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder. BMC Med Genet *6*, 3.

Reiff, M., Giarelli, E., Bernhardt, B.A., Easley, E., Spinner, N.B., Sankar, P.L., and Mulchandani, S. (2015). Parents' perceptions of the usefulness of chromosomal microarray analysis for children with autism spectrum disorders. J Autism Dev Disord *45*, 3262-3275.

Reiner, O., Karzbrun, E., Kshirsagar, A., and Kaibuchi, K. (2016). Regulation of neuronal migration, an emerging topic in autism spectrum disorders. J Neurochem *136*, 440-456.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E.*, et al.* (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405-424.

Richdale, A.L. (1999). Sleep problems in autism: prevalence, cause, and intervention. Dev Med Child Neurol *41*, 60-66.

Salyakina, D., Cukier, H.N., Lee, J.M., Sacharow, S., Nations, L.D., Ma, D., Jaworski, J.M., Konidari, I., Whitehead, P.L., Wright, H.H.*, et al.* (2011). Copy number variants in extended autism spectrum disorder families reveal candidates potentially involved in autism risk. PLoS One *6*, e26049.

Samarakoon, P.S., Sorte, H.S., Kristiansen, B.E., Skodje, T., Sheng, Y., Tjonnfjord, G.E., Stadheim, B., Stray-Pedersen, A., Rodningen, O.K., and Lyle, R. (2014). Identification of copy number variants from exome sequence data. BMC Genomics *15*, 661.

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., 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.

Schaaf, C.P., and Zoghbi, H.Y. (2011). Solving the autism puzzle a few pieces at a time. Neuron *70*, 806-808.

Schaefer, G.B., and Mendelsohn, N.J. (2013). Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. Genet Med *15*, 399-407.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J.*, et al.* (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445-449.

Serafini, T., Colamarino, S.A., Leonardo, E.D., Wang, H., Beddington, R., Skarnes, W.C., and Tessier-Lavigne, M. (1996). Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. Cell *87*, 1001-1014.

Serafini, T., Kennedy, T.E., Galko, M.J., Mirzayan, C., Jessell, T.M., and Tessier-Lavigne, M. (1994). The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. Cell *78*, 409-424.

SFARI (2015). SFARI Gene - Human Gene Module, S.F.A.R. Foundation, ed. (Simons Foundation Autism Research Foundation).

Shen, Y., Dies, K.A., Holm, I.A., Bridgemohan, C., Sobeih, M.M., Caronna, E.B., Miller, K.J., Frazier, J.A., Silverstein, I., Picker, J.*, et al.* (2010). Clinical genetic testing for patients with autism spectrum disorders. Pediatrics *125*, e727-735.

Sherman, S., Pletcher, B.A., and Driscoll, D.A. (2005). Fragile X syndrome: diagnostic and carrier testing. Genet Med *7*, 584-587.

Shevell, M.I., Majnemer, A., Rosenbaum, P., and Abrahamowicz, M. (2001). Etiologic yield of autistic spectrum disorders: a prospective study. J Child Neurol *16*, 509-512.

Shu, T., and Richards, L.J. (2001). Cortical axon guidance by the glial wedge during the development of the corpus callosum. J Neurosci *21*, 2749-2758.

Smit, J.J., Monteferrario, D., Noordermeer, S.M., van Dijk, W.J., van der Reijden, B.A., and Sixma, T.K. (2012). The E3 ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension. EMBO J *31*, 3833-3844.

Son, E.Y., and Crabtree, G.R. (2014). The role of BAF (mSWI/SNF) complexes in mammalian neural development. Am J Med Genet C Semin Med Genet *166C*, 333-349.

Spence, S.J., and Schneider, M.T. (2009). The role of epilepsy and epileptiform EEGs in autism spectrum disorders. Pediatr Res *65*, 599-606.

Stein, J.L., Parikshak, N.N., and Geschwind, D.H. (2013). Rare inherited variation in autism: beginning to see the forest and a few trees. Neuron *77*, 209-211.

Sumi, S., Taniai, H., Miyachi, T., and Tanemura, M. (2006). Sibling risk of pervasive developmental disorder estimated by means of an epidemiologic survey in Nagoya, Japan. J Hum Genet *51*, 518-522.

Svard, J., Heby-Henricson, K., Persson-Lek, M., Rozell, B., Lauth, M., Bergstrom, A., Ericson, J., Toftgard, R., and Teglund, S. (2006). Genetic elimination of Suppressor of fused reveals an essential repressor function in the mammalian Hedgehog signaling pathway. Dev Cell *10*, 187-197.

Takashima, S., Tsuji, S., and Tsujimoto, M. (2002). Characterization of the second type of human beta-galactoside alpha 2,6-sialyltransferase (ST6Gal II), which sialylates Galbeta 1,4GlcNAc structures on oligosaccharides preferentially. Genomic analysis of human sialyltransferase genes. J Biol Chem *277*, 45719-45728.

Tammimies, K., Marshall, C.R., Walker, S., Kaur, G., Thiruvahindrapuram, B., Lionel, A.C., Yuen, R.K., Uddin, M., Roberts, W., Weksberg, R.*, et al.* (2015). Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. JAMA *314*, 895-903.

Tau, G.Z., and Peterson, B.S. (2010). Normal development of brain circuits. Neuropsychopharmacology *35*, 147-168.

Taylor, M.D., Liu, L., Raffel, C., Hui, C.C., Mainprize, T.G., Zhang, X., Agatep, R., Chiappa, S., Gao, L., Lowrance, A.*, et al.* (2002). Mutations in SUFU predispose to medulloblastoma. Nat Genet *31*, 306-310.

Thiebault, K., Mazelin, L., Pays, L., Llambi, F., Joly, M.O., Scoazec, J.Y., Saurin, J.C., Romeo, G., and Mehlen, P. (2003). The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment. Proc Natl Acad Sci U S A *100*, 4173-4178.

Torres, A.R., Westover, J.B., and Rosenspire, A.J. (2012). HLA Immune Function Genes in Autism. Autism Research and Treatment *2012*, 13.

Tuchman, R. (2006). Autism and epilepsy: what has regression got to do with it? Epilepsy Curr *6*, 107-111.

Tylenda, B., Beckett, J., and Barrett, R.P. (2007). Assessing Mental Retardation Using Standardized Intelligence Tests. In International Review of Research in Mental Retardation, L.M. Johnny, ed. (Academic Press), pp. 27-97.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A*., et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Valente, L., Tiranti, V., Marsano, R.M., Malfatti, E., Fernandez-Vizarra, E., Donnini, C., Mereghetti, P., De Gioia, L., Burlina, A., Castellan, C*., et al.* (2007). Infantile encephalopathy and defective mitochondrial DNA translation in patients with mutations of mitochondrial elongation factors EFG1 and EFTu. Am J Hum Genet *80*, 44-58.

Veenstra-Vanderweele, J., Christian, S.L., and Cook, E.H., Jr. (2004). Autism as a paradigmatic complex genetic disorder. Annu Rev Genomics Hum Genet *5*, 379-405.

Vignoli, A., La Briola, F., Peron, A., Turner, K., Vannicola, C., Saccani, M., Magnaghi, E., Scornavacca, G.F., and Canevini, M.P. (2015). Autism spectrum disorder in tuberous sclerosis complex: searching for risk markers. Orphanet J Rare Dis *10*, 154.

Vismara, L.A.a.R., S.J. (2010). Behavioral Treatments in Autism Spectrum Disprder: What Do We Know? Annu Rev Clin Psychol *6*, 447-468.

Vivanti, G., Barbaro, J., Hudry, K., Dissanayake, C., and Prior, M. (2013). Intellectual development in autism spectrum disorders: new insights from longitudinal studies. Front Hum Neurosci *7*, 354.

Walker, S., and Scherer, S.W. (2013). Identification of candidate intergenic risk loci in autism spectrum disorder. BMC Genomics *14*, 499.

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence, 2 edn (San Antonio, TX, The Psychological Corporation).

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T*., et al.* (2008). Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med *358*, 667-675.

Werner, E., and Dawson, G. (2005). Validation of the phenomenon of autistic regression using home videotapes. Arch Gen Psychiatry *62*, 889-895.

Wilson, P.M., Fryer, R.H., Fang, Y., and Hatten, M.E. (2010). Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. J Neurosci *30*, 8529-8540.

Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C*., et al.* (2014). Molecular findings among patients referred for clinical whole-exome sequencing. JAMA *312*, 1870-1879.

Zecavati, N., and Spence, S.J. (2009). Neurometabolic disorders and dysfunction in autism spectrum disorders. Curr Neurol Neurosci Rep *9*, 129-136.

Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature *466*, 258-262.

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.

APPENDIX A: Diagnostic Criteria for Autism Spectrum Disorder, 299.00 (F84.0)

(DSM-V, 2013)

A.   Persistent deficits in social communication and social interaction across multiple

contexts, as manifested by the following, currently or by history (examples are

illustrative, not exhaustive, see text):

1.   Deficits in social-emotional reciprocity, ranging, for example, from abnormal social

approach and failure of normal back-and-forth conversation; to reduced sharing of

interests, emotions, or affect; to failure to initiate or respond to social interactions.

2.   Deficits in nonverbal communicative behaviors used for social interaction,

ranging, for example, from poorly integrated verbal and nonverbal communication; to

abnormalities in eye contact and body language or deficits in understanding and use of

gestures; to a total lack of facial expressions and nonverbal communication.

3.   Deficits in developing, maintaining, and understanding relationships, ranging, for

example, from difficulties adjusting behavior to suit various social contexts; to difficulties

in sharing imaginative play or in making friends; to absence of interest in peers.

Specify current severity: Severity is based on social communication impairments and

restricted repetitive patterns of behavior (see Table 2).

B.   Restricted, repetitive patterns of behavior, interests, or activities, as manifested by

at least two of the following, currently or by history (examples are illustrative, not

exhaustive; see text):

1.   Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

2.   Insistence on sameness, inflexible adherence to routines, or ritualized patterns or verbal nonverbal behavior (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day).

3.   Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interest).

4.   Hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

Specify current severity: Severity is based on social communication impairments and restricted, repetitive patterns of behavior (see Table 2).

C.   Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D.   Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E.   These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level.

Note: Individuals with a well-established DSM-IV diagnosis of autistic disorder, Asperger's disorder, or pervasive developmental disorder not otherwise specified should be given the diagnosis of autism spectrum disorder. Individuals who have marked deficits in social communication, but whose symptoms do not otherwise meet criteria for autism spectrum disorder, should be evaluated for social (pragmatic) communication disorder.

Specify if:

With or without accompanying intellectual impairment

With or without accompanying language impairment

Associated with a known medical or genetic condition or environmental factor (Coding note: Use additional code to identify the associated medical or genetic condition.)

Associated with another neurodevelopmental, mental, or behavioral disorder (Coding note: Use additional code[s] to identify the associated neurodevelopmental, mental, or behavioral disorder[s].)

With catatonia (refer to the criteria for catatonia associated with another mental disorder, pp. 119-120, for definition) (Coding note: Use additional code 293.89 [F06.1] catatonia associated with autism spectrum disorder to indicate the presence of the comorbid catatonia.)

Table 2: Severity levels for autism spectrum disorder

Severity level	Social communication	Restricted, repetitive behaviors
Level 3 "Requiring very substantial support"	Severe deficits in verbal and nonverbal social communication skills cause severe impairments in functioning, very limited initiation of social interactions, and minimal response to social overtures from others. For example, a person with few words of intelligible speech who rarely initiates interaction and, when he or she does, makes unusual approaches to meet needs only and responds to only very direct social approaches	Inflexibility of behavior, extreme difficulty coping with change, or other restricted/repetitive behaviors markedly interfere with functioning in all spheres. Great distress/difficulty changing focus or action.
Level 2 "Requiring substantial support"	Marked deficits in verbal and nonverbal social communication skills; social impairments apparent even with supports in place; limited initiation of social interactions; and reduced or abnormal responses to social overtures from others. For example, a person who speaks simple sentences, whose interaction is limited to narrow special interests, and how has markedly odd nonverbal communication.	Inflexibility of behavior, difficulty coping with change, or other restricted/repetitive behaviors appear frequently enough to be obvious to the casual observer and interfere with functioning in a variety of contexts. Distress and/or difficulty changing focus or action.
Level 1 "Requiring support"	Without supports in place, deficits in social communication cause noticeable impairments. Difficulty initiating social interactions, and clear examples of atypical or unsuccessful	Inflexibility of behavior causes significant interference with functioning in one or more contexts. Difficulty switching between activities. Problems of organization and planning

	response to social overtures of others. May appear to have decreased interest in social interactions. For example, a person who is able to speak in full sentences and engages in communication but whose to-and-fro conversation with others fails, and whose attempts to make friends are odd and typically unsuccessful.	hamper independence.

APPENDIX B: Consent form for The Neurobiology and Genetics of Autism Assessment Core (UC Irvine IRB# 1996-616).

UNIVERSITY OF CALIFORNIA, IRVINE
CONSENT TO ACT AS A HUMAN RESEARCH SUBJECT

The Neurobiology and Genetics of Autism
Assessment Core

You are being asked to participate in a research study. Participation in this study is completely voluntary. Please read the information below and ask questions about anything that you do not understand before deciding if you want to participate. A researcher listed below will be available to answer your questions.

RESEARCH TEAM AND SPONSORS

Lead Researcher
Pamela Flodman, MSc, MS
Department of Pediatrics
714-456-5789
pflodman@uci.edu

Other Researchers:
Co-Investigators: Moyra Smith, MD; M. Anne Spence PhD;
Nicole Gage, PhD; Wendy Goldberg, PhD;
Maureen Bocian, MD; Jay Gargus, MD, PhD; Ira T. Lott, MD;
Kathy Osann, PhD; Gregory Hickok; PhD, Fan-Gang Zeng, PhD;
June-Anne Gold, MD, FACMG

Study Location:
University of California, Irvine
University of California, Irvine Medical Center, Orange, CA

Study Sponsor: National Institutes of Health (NIH)

NAME OF PARTICIPANT: _____

This participant is a: _____ Proband with autistic disorder
                                _____ Family member of a proband
                                _____ Typically developing control

PURPOSE OF STUDY
The purpose of this research study, The Neurobiology and Genetics of Autism, is to learn more about the cause of autistic disorder. In order to accomplish this goal several research projects have been designed to study 150 children with Autistic Disorder and their families. Project 1 will undertake family genetic studies, and Project 6 will study how children with Autistic Disorder process sound. Other projects which have been active in the past but are no longer actively collecting data include Project 2, which used magnetic resonance imaging (MRI) to test whether specific areas of the brain are different in children with autism, and Project 3, which looked at the development of children with autism to establish when the disorder first appears and whether or not the children lose milestones of development when they are autistic.

In order to participate in any or all of these studies the child with Autistic Disorder (proband), family members, and selected normal children (controls) will be evaluated using a set of standardized tests. Each of the studies has an informed consent which discusses the aspects of that study in some detail. This form is to provide consent for the testing which will be done under the direction of the Assessment Core.

The purpose of the Assessment Core is to provide detailed information about the development, cognitive skills, language use and understanding, and executive functioning of the children (probands, family members, and controls). Since the children differ by age and ability different tests will be used depending on the age of the child and their skills. On the list of tests in the attached Appendix, project personnel have checked the tests they think you/your child will have, but this may be changed as they start the testing. If changes are necessary these will be discussed with you/your child and the changes noted on this form. After these screening tests are completed, you/your child and other family members may choose to participate or not participate in the other active research project (Project 6: Sound Processing). You and your child are not required to do this additional research project after the screening tests are complete but you/your child cannot participate in the research unless the Assessment Core tests are completed.

WHY IS THIS RESEARCH?
This is a research study because we do not yet fully understand the causes of Autistic Disorder.

Autism is a complex disorder, and there are likely many different contributing factors. There is also great variation in the presentation of autism in different individuals. Therefore, the results of genetic studies can only be fully interpreted if detailed information is also available about the phenotype (or characteristics) of each study participant. The data from this study will be analyzed to improve our understanding of how autistic children differ from non-autistic children, how autistic children differ from each other, and the role of genetic and other factors in these differences.

Study Design
Subjects who agree to participate will be children with Autistic Disorder and their family members, and children who do not have Autistic Disorder. Participation will include the tests and assessments which are outlined below under "Procedures", and described in more detail in the Appendix. In addition, being in this study will require the collection of a blood sample (approximately 2 tablespoons) for genetic analysis.

SUBJECTS
Inclusion Requirements
You are eligible to participate in this study if you are either:
- An individual with Autistic Disorder,
- OR the parent, sibling or other relative of an individual with Autistic Disorder,
- OR an individual with typical development.

Number of Participants and Time Commitment
This study will include approximately 565 participants and participation in the Assessment Core will involve approximately 12 hours of your time over the next 12 months.

PROCEDURES
The investigator will first outline the purpose and enrollment process of the study, so that you can determine whether you or your child will be eligible to participate in this study. The investigator will then review this form with you in detail and answer any questions that you may have. If you consent to participate after reviewing this form, additional assessment tests and interviews will be given in four or

five sessions lasting between 1-2 hours each, depending on your/your child's tolerance. Based on your/your child's age and skill level, the appropriate tests will be selected from the list in the Appendix. These tests will include:

- questionnaires/ structured interviews;
- observational schedules;
- tests of intelligence;
- tests of academic achievement;
- tests of executive function, which refers to the ability to plan, adapt to new situations, and think abstractly;
- tests of language and communication function;
- a full neurological examination which tests the function of the nerves, strength, coordination, reflexes, and sensation;
- a sleep diary and actigraph (a watch-shaped electronic device that is worn on the wrist of the non-dominant arm) for a 7-night period to evaluate sleep quality. Daily sleep diary will be entered using pre-programmed smart phones provided by the study. You will receive training by the study team on how enter these data, and you will be instructed <u>not</u> to use the phone for personal use.
- an electroencephalogram (EEG) to record the electrical discharges produced by the brain;
- medical records will be collected, with your permission, to document prenatal, perinatal, and developmental history. Taking account of these factors during analysis may make it easier for the researchers to find the genes that may play a role in autism.
- routine analysis and biochemical tests of urine sample.
- micro-organic analysis of collected breath samples.
- If a DNA analysis for Fragile X has already been performed on you/your child, you agree to forward a copy of the results to the project, and these blood tests will not be repeated.
- a full genetics evaluation to include:
    - a clinical evaluation, including a physical exam to document physical characteristics and standard measurements such as height, weight, and head circumference.
    - an extensive family history called a pedigree.
    - collection of a blood sample (30cc, 2 tablespoons) for attempts to identify the genetic factors that may be involved in autism
        - The blood sample may be used to count the number of you/your child's chromosomes (called a karyotype), and/or to use a test of DNA analysis for Fragile X, a genetic condition often associated with autism.
        - The blood sample may also be used to make a cell line. The purpose of establishing the cell line is to have a source of additional DNA if the first DNA samples are used up before the study is completed. The cell lines and DNA will remain in the human genetics laboratory at UC Irvine. They will not be sold.
        - The blood sample may also be used for whole genome sequencing (WGS), the decoding of your entire genetic code, and providing serum for genetic and biochemical tests. It is highly unlikely that a significant and medically-actionable finding will be found through our analysis of the WGS data. However, if such an occasion should arise and if you and/or your guardians/legal representative express the desire to be informed of such findings, we will obtain a clinical laboratory report for the purpose of informing you of the finding. There will be no cost to you for any of these tests. Depending on the inheritance pattern of a particular condition, different associated inheritable risks may be involved. If you have questions involving your genetic risks, it is recommended that you pursue an appointment with a clinical geneticist and/or a genetic counselor as part of your standard medical care.

Please check <u>one</u> of the following options and initial:
Check / Initial

_____ _____ I do **NOT** wish to know genomic testing results of any kind.

_____ _____ I wish to know genomic testing results for medically-significant and actionable findings.

Data sharing through the National Database for Autism Research (NDAR):
You may choose to allow the researchers to share your or your child's de-identified research data and related findings with other researchers through the National Database for Autism Research (NDAR). NDAR is a computer system created by and located at the National Institutes of Health to help researchers to learn more about autism and other related disorders. During and after the study, the researchers will send information about health and behavior and genetic information to NDAR. However, before they send it to NDAR, they will remove all identifying information such as name, address, and phone number, and will replace that information with a code number. No names or other personally identifying information are ever shared with the data repository or anyone else outside of the research group.

Other researchers nationwide can file an application with the National Institutes of Health to obtain access to the NDAR data for research purposes. Experts at the National Institutes of Health will look at every request carefully to minimize risks to privacy. If you agree now to have your data shared with NDAR, you can change your mind later by contacting the researchers who conducted this study, and they will tell NDAR, which can stop sharing the research information. However, NDAR cannot take back information that was shared before you changed your mind. More information about NDAR is available on-line at http://ndar.nih.gov.

Please check one of the following options, and initial:

_____ _____ You want your or your child's research data and related findings to be shared with other researchers through the National Database for Autism Research (NDAR).

_____ _____ You do **not** want your or your child's research data and related findings to be shared with other researchers through the National Database for Autism Research (NDAR) or any other similar data repository.

Analysis of your research sample by other expert research groups:
Autism is a complex condition, and often the causes are unknown. Sometimes the research testing may identify variations which require further study in order to determine whether they are related to autism. Often, laboratories other than this UC Irvine research group are experts in the investigation of these genetic variations. By initialing below, you may indicate whether you agree that your biological sample may be shared with such expert groups. However, your name and other identifying information will not be disclosed to anyone outside of the UC Irvine research group. No costs will incur due to any such research testing to you or to your insurer.   If you agree now that your research sample may be shared, you can change your mind later by contacting the researchers.

Please check one of the following options, and initial:

____ ____ You agree that your or your child's research sample may be shared with other expert research groups.

____ ____ You do **not** want your or your child's research sample to be shared with other expert research groups.

Re-contact regarding future research projects:
Please check one of the following options, and initial:

____ ____ You wish to be contacted by us in the future regarding other research projects for which you or your child may be eligible. We will provide you with information about these research projects, as well as the contact information for the investigators. However, your name and identifying information will **not** be provided to the investigators, unless you choose to provide this information yourself.

____ ____ You do **not** want to be contacted by us regarding other research projects for which you or your child may be eligible.

RISKS AND DISCOMFORTS
The possible risks and/or discomforts associated with the procedures described in this study include:

Cognitive & Behavioral Testing: There are no physical risks to the intelligence testing, neurological or genetics evaluation. There is the risk that you/your child may become frustrated, especially if you/your child becomes bored, tired, or feels that they are not performing well. If you find the questionnaire or interview items to be too personal or objectionable in any way, you may omit any answer to any question.

Neurodevelopmental/Family Histories: The risks involved with the neurodevelopmental and family histories are minimal. Some families may feel uncomfortable providing detailed family history and pedigree information. The trained medical genetics staff are used to explaining the necessity of certain questions and helping the family to understand why the family information is important. If any questions do arise concerning a potential genetic disorder based on the pedigree, you/your family will be referred to the Genetics Clinic for evaluation.

Drawing Blood: When your/your child's blood is drawn, there could be minor discomfort or pain associated with brief pricking of the skin and, possibly, anxiety, dizziness and possible fainting while waiting for the blood to be drawn. Sometimes, a small bruise, swelling, and rarely, an infection may later occur where the blood was drawn. To decrease the pain and discomfort as much as possible, all blood samples will be drawn by staff who have a lot of experience drawing blood. The investigators will not allow the procedure to continue if your child does not cooperate sufficiently and may therefore pose an unreasonable risk to yourself/your child or to others. EMLA cream will not be routinely used for the blood draw, but may be available upon request.

Collection of Urine and Breath Samples: There are no risks involved in the collection of urine and breath samples.

Incidental Findings: In the unlikely event that an incidental genetic finding unrelated to autism research is found through whole genome sequencing, you may experience shock and emotional discomforts in learning about unexpected genetic risks. Only findings that have been well established to be significant and medically-actionable will be provided to you, following confirmation in a clinical laboratory; these

findings will only be disclosed if you have previously expressed the desire to receive incidental findings, at the time of consent. Qualified researchers who are trained to give sensitive genetic information will disclose the results and provide genetic counseling, to help you understand the implications of the genetic finding and to address any psychological issues that may come up. Alternatively, referrals can be made to other professionals not affiliated with the study. You will be responsible for the costs of visiting professionals outside of the study.

Through genetic studies and pedigree analysis, it may be determined that some members in your family are not genetically related (i.e. non-paternity, adoption, etc.). Such information will not be disclosed to you or your family members. Additionally, family members may learn private genetic information about the participant. The study team will make no attempts to contact other family member to inform them of any genetic information about a participant. However, all participants are encouraged to contact relevant family members to make sure they are aware of any significant genetic information that may affect them.

Actigraph: participant will be asked to wear an actigraph monitor for a 7-night period to evaluate sleep, daytime activity, and sleep activity rhythms. For the participants who dislike the sensation of wearing the actigraph on their wrist, they may choose to wear a long-sleeve shirt under the device.

EEG: EEG records the electrical discharges produced by the brain and no electrical current is put into the subject's body. Saline is used with the electrode cap and can cause mild irritation of the skin. This will be minimized by removing any excess saline from the cap before placing it on the participant's head. Some may experience discomfort with either wearing the cap and/or the need to sit still during the experiment. They may end the procedure at any time.

Breach of Confidentiality: As this study involves the use of identifiable information, there is a potential for a breach of confidentiality, although all efforts will be made to minimize this.

UNKNOWN RISKS
There may be risks to being in this study that we don't know about now. You will be informed of any changes in the way the study will be done and any additional identified risks to which you may be exposed.

BENEFITS
Subject Benefits
The main benefit to this study is that your child will undergo neurologic, genetic, neuropsychologic, language and behavioral evaluations, which may provide insight into your child's cognitive, language and behavioral strengths and weaknesses. Findings may be of clinical, medical or educational significance. You will receive, upon your request, a written report of the results of the clinical testing. Note: the results of the genetic tests included under this protocol are for research purposes only and cannot be used for medical treatment or diagnosis: if any questions arise regarding potential genetic disorder, the family will be referred to the Genetics clinic for medical treatment or diagnosis. The results of the neurological and language/ neuropsychological testing may have a direct benefit to your child, such that a report can be made available to your child's educators. If you choose to share this information, you can request in writing that a copy of the report be sent to your child's school.

Benefits to Others or Society
The practical and scientific benefit to this study is to provide information that will assist professionals to a greater understanding of the neurological and genetic basis of autism. This study will shed light on the relationships between genetic factors and development of the brain, and the relationship with the behaviors which are characteristic of autism.

ALTERNATIVES TO PARTICIPATION
This is not a treatment study, and there are no alternative treatments or procedures available. The only alternative is not to participate in this study.

COMPENSATION, COSTS AND REIMBURSEMENT
Compensation for Participation
Your family will be compensated $30 for completing each EEG recording and up to $125 for completing the Sleep Quality study as part of the Assessment Core.

Costs
There will be no cost to you or to your family if you participate in this study.

Reimbursement
You may incur travel expenses for commuting to UC Irvine or to the UCI Medical Center. You will not be directly reimbursed for these expenses.

Compensation for Injury
If you are injured as a direct result of your participation in this study, the University of California will provide reasonable and necessary medical care to treat the injury at no cost to you or to your insurer/third party payer. The University of California does not routinely provide any other form of compensation for injury. It is important that you report any suspected study-related injury to the research team listed at the top of this form immediately.

WITHDRAWAL OR TERMINATION FROM THE STUDY AND CONSEQUENCES
You are free to withdraw from this study at any time. **If you decide to withdraw from this study you should notify the research team immediately**. The research team may also end your participation in this study if you do not follow instructions, miss scheduled visits, your safety and welfare are at risk, or the study sponsor decides to stop the study.

If you experience any of the side effects listed in the <u>Risks and Discomforts</u> section or if you become ill during the research, you may need to be withdrawn from the study, even if you would like to continue. The research team will make the decision and let you know if it is not possible for you to continue. The decision may be made to protect your safety and welfare, or because it is part of the research plan that people who develop certain conditions may not continue to participate.

If you withdraw or are removed from the study, the researcher may ask you to return for a final close-out visit or complete an exit telephone interview.

CONFIDENTIALITY
Subject Identifiable Data
- All participants (children with Autistic Disorder, their family members, and controls) will receive a family ID number and a personal ID number.
- All identifiable information that will be collected about you will be removed and replaced with a code. A list linking the code and your identifiable information will be kept separate from the research data.
- All members of the same family will share the family ID number with their own personal ID numbers. Names, addresses, and phone numbers will be retained in locked file cabinets for the purpose of contacting to schedule research visits only.
- All material used by the experimenters will be restricted to the minimal amount of information necessary to insure that the correct person is being tested.

- **Genetics** A Federal law, called the Genetic Information Nondiscrimination Act (GINA) generally makes it illegal for health insurance companies, group health plans, and employers of 15 or more persons to discriminate against you based on your genetic information. This means that they may not use your genetic information when making decisions regarding insurability. GINA does not prohibit discrimination on the basis of an already manifest genetic disease or disorder. This means if you have a diagnosis and/or are being treated for a genetic condition, a health insurer may use the information to determine eligibility or rates. Also, GINA does not protect you against genetic discrimination by companies that sell life insurance, disability insurance, or long-term care insurance. If you would like more information about GINA go to: http://www.genome.gov/Pages/PolicyEthics/Genetic Discrimination/GINAInfoDoc.pdf

Data Storage
- All research data will be maintained in a secure location at UCI. Only authorized individuals will have access to it.
- De-identified research data will be maintained on a secure computer system at the National Institutes of Health if and only if you choose to allow the researchers to share these data and related findings with other researchers through the National Database for Autism Research (NDAR).
- All data stored on secure computer systems and will be identified by ID number only. A strict password policy (correct choice of words with frequent changes mandated) is in place for all computers associated with the project.
- The video recordings will also be stored in a secure location and transcribed. The recordings will be retained with the other research data.

Data Access
- All access to information collected in this study will be limited to project staff on a "need to know" basis and will not be generally available even to the research assistants.
- De-identified research data will be available through the National Database for Autism Research (NDAR) if and only if you choose to allow researchers to share these data through NDAR.
- The research team, authorized UCI personnel, the study sponsor (if applicable), and regulatory entities such as the Food and Drug Administration (FDA) and the Office of Human Research Protections (OHRP), may have access to your study records to protect your safety and welfare. Any information derived from this research project that personally identifies you will not be voluntarily released or disclosed by these entities without your separate consent, except as specifically required by law. Research records provided to authorized, non-UCI entities will not contain identifiable information about you. Publications and/or presentations that result from this study will not include identifiable information about you.

Data Retention
The researchers intend to keep the research data indefinitely.

If you have chosen to share your or your child's de-identified research data through the National Database for Autism Research (NDAR), the intention is that these data will be kept at NDAR indefinitely.

NEW FINDINGS
If, during the course of this study, significant new information becomes available that may relate to your willingness to continue to participate, this information will be provided to you by the research team listed at the top of the form.

OTHER CONSIDERATIONS

Use of Specimens

Any specimens (e.g., tissue, blood, urine) obtained for the purposes of this study will become the property of the University of California, Irvine (UCI). Once you provide the specimens you will not have access to them. The specimens will be used for research and such use may result in inventions or discoveries that could become the basis for new products or diagnostic or therapeutic agents. In some instances, these inventions and discoveries may be of potential commercial value and may be patented and licensed by the University. You will not receive any money or other benefits derived from any commercial or other products that may be developed from use of the specimens.

Investigator Financial Conflict of Interest

No one on the study team has a disclosable financial interest related to this research project.

IF YOU HAVE QUESTIONS

If you have any comments, concerns, or questions regarding the conduct of this research please contact the research team listed at the top of this form.

If you are unable to reach a member of the research team listed at the top of the form and have general questions, or you have concerns or complaints about the research study, research team, or questions about your rights as a research subject, please contact UCI's Office of Research by phone, (949) 824-6068 or (949) 824-2125, by e-mail at IRB@rgs.uci.edu or in person at 5171 California Ave., Suite 150, Irvine, CA 92697-7600.

VOLUNTARY PARTICIPATION STATEMENT
You should not sign this form unless you have read the attached "Experimental Subject's Bill of Rights" and have been given a copy of it and this consent form to keep. **Participation in this study is voluntary.** You may refuse to answer any question or discontinue your involvement at any time without penalty or loss of benefits to which you might otherwise be entitled. Your decision will not affect your future relationship with UCI or your quality of care at the UCI Medical Center. Your signature below indicates that you have read the information in this consent form and have had a chance to ask any questions that you have about the study.

I agree to participate in the study.

_____   _____
Subject Signature   **Date**

_____
Printed Name of Subject

_____   _____
Legally Authorized Representative/Guardian Signature   *Date*

_____
Printed Name of Legally Authorized Representative/Guardian

_____   _____
Researcher Signature   **Date**

_____
Printed Name of Researcher

_____   _____
Witness Signature   *Date*

_____
Printed Name of Witness

UNIVERSITY OF CALIFORNIA, IRVINE
Experimental Subject's Bill of Rights

The rights listed below are the right of every individual asked to participate in a research study. You have the right:

1. To be told about the nature and purpose of the study.

2. To be told about the procedures to be followed in the research study, and whether any of the drugs, devices, or procedures is different from what would be used in standard practice.

3. To receive a description of any side effects, discomforts, or risks that you can reasonably expect to occur during the study.

4. To be told of any benefits that you may reasonably expect from the participation in the study, if applicable.

5. To receive a description of any alternative procedures, drugs, or devices that might be helpful, and their risks and benefits compared to the proposed procedures, drugs or devices.

6. To be told of what sort of medical treatment, if any, will be available if any complications should arise.

7. To be given a chance to ask any questions concerning the research study both before agreeing to participate and at any time during the course of the study.

8. To refuse to participate in the research study. Participation is voluntary. You may refuse to answer any question or discontinue your involvement at any time without penalty or loss of benefits to which you might otherwise be entitled. Your decision will not affect your right to receive the care you would receive if you were not in the experiment.

9. To receive a copy of the signed and dated written consent form and a copy of this form.

10. To be given the opportunity to freely decide whether or not to consent to the research study without any force, coercion, or undue influence.

If you have any concerns or questions regarding the research study you should contact the research team listed at the top of the consent form.

If you are unable to reach a member of the research team and have general questions, or you have concerns or complaints about the research study, research team, or questions about your rights as a research subject, please contact the UCI's Human Research Protections Program in the Office of Research by calling (949) 824-6068 or (949) 824-2125 Monday – Friday, 8 am – 5 pm; or by e-mail at IRB@rgs.uci.edu; or by writing us at 5171 California Ave., Suite 150, Irvine, CA 92697-7600.

APPENDIX

A. Questionnaires/ Structured Interviews

1. Childhood Autism Rating Scale: a 15-item behavior rating scale to identify children with autism and distinguish them from other nonautistic developmentally disabled children. Each of the 15-items uses a 7-point rating scale to indicate the degree to which the child's behavior deviates from an age-appropriate norm; in addition, it distinguishes mild-to-moderate from severe autism. The CARS takes approximately 15 minutes to administer.

2. Autism Diagnostic Interview-Revised: a diagnostic parent interview with a comprehensive list of autistic symptoms in the spheres of social relatedness, communication, and ritualistic/ perseverative behaviors. The ADI-R takes approximately 60-90 minutes to administer, and is currently the □gold standard□ for the diagnosis of Autistic Disorder by DSM-IV criteria in research protocols.

3. Autism Diagnostic Observation Schedule: a semistructured observational assessment for autistic subjects ranging in age from preschool through adulthood. It requires approximately 45 minutes to administer, and includes investigator-directed activities to evaluate communication, reciprocal social interaction, play, stereotypic behavior or restricted interests, and other abnormal behaviors.

4. Vineland Adaptive Behavior Scales II: a parent interview to estimate development in interpersonal relationships, play and leisure time, and coping skills; personal, domestic and community daily living skills; and receptive, expressive and written communication, with developmentally ordered skills for each area. Standard scores, percentile ranks, adaptive levels, and age equivalents are available. The Vineland takes approximately 30 minutes to administer.

5. Child Behavior Checklist: a widely used parent-rated behavior problem checklist to screen for other psychiatric symptoms, such as anxiety, depression, delinquent behavior, etc. The CBCL takes approximately 15 minutes to fill out.

6. Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS): a diagnostic interview developed to allow a best estimate or summary rating based on the child's report, the parent's report, and other sources of information (school, medical chart, other) to diagnose comorbid psychiatric conditions in autistic individuals.

7. SNAP-IV-C Rating Scale: a rating scale checklist that lists 90 items in 2 parts. Part I includes all inattention and hyperactivity criteria for Attention Deficit Hyperactivity Disorder, as well as all criteria for Oppositional Defiant and Conduct Disorders. Part II screens for symptoms of other psychopathologic disorders, such as Tourette's, Obsessive-Compulsive, or Generalized Anxiety Disorders. Both parts of the SNAP-IV take approximately 20 minutes to fill out.

8. The Pittsburgh Sleep Quality Index: a self-report measure 19-item questionnaire that asks about sleep routines and problems going to sleep, staying asleep, and getting sufficient rest during the previous month.

9. The Cleveland Adolescent Sleepiness Questionnaire: is a brief 16-item self-report measures the target participant's likeliness to fall asleep in certain situations.

10. The Network of Relationships Inventory-Relationship Qualities Version: an adolescent and parent measure assessing supportive and discordant parent-adolescent relationship quality.

11. Autism Spectrum Quotient: a 50-item self-report measure of autistic traits that is appropriate for older adolescents and adults.

12. Center for Epidemiologic Studies Depression Scale: a 20-item scale with established validity and reliability to measure depressive symptomatology.

13. b: a 40-item measure appropriate for individuals 15 years of age and older to measure the intensity of state and trait anxiety.

14. School Engagement: a three-question survey using a 3-point scale to questions that assess whether the participant believes people at school care about him/her, whether he/she feels a part of his/her school, and how often teachers at his/her school treat students fairly.

15. Work Engagement: a four-question survey of perceived organizational support using a 5-point scale that assesses whether the participant believes that their place of work cares about his/her well-being and whether the participant feels that his/her supervisor supports the participant's opinions.

16. Social Communication Questionnaire: a brief parent-report instrument that evaluates the degree of impairments in social skills and communication, and the presence of restricted and repetitive behaviors that are associated with ASD. It is appropriate for the evaluation of individuals age four years and over.

B. Tests of Intelligence

1. Stanford-Binet Intelligence Test- 4th edition is standardized from age 2 years to adulthood, and is comprised of 15 subtests assessing cognitive abilities in four areas: Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning and Short-Term Memory. Some subtests are administered only at the preschool and elementary school ages (e.g., Absurdities and Copying), while others are administered only to subjects aged 12 years and older (e.g., Verbal Relations, Paper Folding & Cutting, Equation Building). Standard age scores are computed for each subtest and for the four areas on a separate scale. It takes approximately 75 minutes to administer.

2. Mullen Scales of Early Learning: a comprehensive scale of multiple developmental domains which assesses a child□s learning abilities and patterns from birth through 5 years of age. It assesses development in five areas: Gross and Fine Motor, Visual Reception, Expressive and Receptive Language. The Mullen takes approximately 15 minutes to administer, and will also be used for those children who have a mental age of less than 2.0 years. It takes approximately 30-40 minutes to administer.

3. Weschler Abbreviated Scale of Intelligence – Second Edition: a brief measure of verbal comprehension and perceptual reasoning abilities in individuals 6-90 years of age. It includes 4 subtests: Block design, vocabulary, matrix reasoning, and similarities.

C. Tests of Academic Achievement

The **Wechsler Individual Achievement Tests:** an achievement battery standardized with the WISC-III to screen for any discrepancy between cognition abilities and academic achievement in the domains of Basic Reading, Mathematics Reasoning, Spelling, Reading Comprehension, Numerical Operations, Listening Comprehension, and Oral and Written Expression. The WIAT takes approximately 20 minutes to administer.

D. Tests of Executive Function

Cambridge Neuropsychological Test Automated Battery (CANTAB): a computer administered series of the following subtests to measure executive function. The length of each subtest depends upon the child's response time.

1. Visual Memory Battery: *Pattern and Spatial Recognition* – A series of visual patterns appear on a screen. In the recognition phase, subjects are required to choose between a pattern they have already seen and a novel pattern. *Delayed Matching to Sample* – Subjects are shown a complex visual pattern (sample) and then, after a brief delay, four choice patterns, one of which is identical to the sample. *Paired Associates Learning* – Subject is required to remember patterns associated with different locations on the screen, and during the test phase, as each pattern is presented, point to the appropriate location.

2. Attention Battery: *Big/Little Circle* – A series of pairs of circles, one large and one small, are presented to the subject. The subject is instructed to point to the smaller of the two, and then after 20 trials, to point to the larger. *Intradimensional/Extradimensional Shift* – After learning a set of criterion to identify a set of stimuli (shapes or lines), the stimuli are shifted or reversed, and the subject is required to make the shift. *Reaction Time* – This task trains the subject in skills related to holding down a touch pad and touching the screen, requiring increasingly complex chains of responses. *Matching to Sample Visual Search* – An abstract pattern composed of four colored shapes is presented in the center of the screen. After a brief delay, a varying number of similar patterns is shown around the edge of the screen. The subject must choose the pattern that matches the one in the center of the screen by touching it.

3. Working Memory and Planning Battery: *Spatial Span* – A pattern of white squares is shown on the screen. Some change color, one by one, in a variable sequence. Subject touches the boxes colored by the computer in the same order as they were originally presented. *Spatial Working Memory* – A number of colored boxes are shown. The overall aim is that the subject should find a blue "counter" in each of the boxes and use them to fill up an empty column on the right side of the screen. *Tower of London* – The subject is shown two displays containing three colored balls. The subject must use the balls in the lower display to copy the pattern shown in the upper display. The time taken to complete the pattern and the number of moves required are taken as measures of the subject's planning ability. *Tower of Hanoi* - to measure executive functioning. *The Wisconsin Card Sort* – to measure perseveration, abstract reasoning, and inhibitory control.

4 Theory of Mind: *Second-Order False-Belief Task* and *Strange Stories Test* – stories are presented and questions are asked to measure the participant's understanding.

E. Tests of Language and Communication Function

1. *Preschool Language Scale-3:* a standardized assessment that includes two subscales of Auditory Comprehension and Expressive Communication to assess attention, vocal development, social communication, semantics, language structure and integrative thinking skills. It is designed to be used from birth through 6 years of age, and takes approximately 45-60 minutes to administer.

2. *Peabody Picture Vocabulary Test:* a widely used test of single-word language comprehension for use in children aged 2 to 18 years. Children are presented with a four-picture choice from which they are required to correctly choose the noun or verb spoken by the examiner. The PPVT takes approximately 20 minutes to administer.

3. *Expressive One Word Picture Vocabulary Test:* a measure of expressive language and vocabulary requiring confrontation naming of black and white line drawings, for use in children aged 2.0 to 15.11 years. The EOWPVT takes approximately 30 minutes to administer.

4. *Token Test for Children:* assesses language comprehension of more complex sentences in five Parts of increasing complexity. The TTFC takes approximately 20 minutes to administer.

5. *Clinical Evaluation of Language Fundamentals- Third Edition:* comprehensive assessment of receptive and expressive language skills in children aged 6-21 years. The CELF-3 takes approximately 60 minutes to administer.

6. *Test of Language Competence:* measures skills in children aged 5.0 to 18.0 years in the areas of vocabulary and meaningful language, sentence structure and complexity, and the communicative use of language. Concepts tested include multiple meanings, inferences, figurative usage, conversational sentence production, and memorization strategies. The TLC takes approximately 30 minutes to administer.

7. *Early Social Communication Scales:* a test developed and used in research protocols to assess the presence of nonverbal communication function. The tester systematically presents the child with each of six mechanical toys. The tester also claps, sings a brief song, gently tickles the child, presents two pointing trials and offers the child several turn-taking opportunities.

8. *Spontaneous Language Sample:* Either during the speech and language evaluation, or during another session, my child's spontaneous language will be audio taped and/or videotaped. At a later time, my child's language will be evaluated for quantitative data such as mean length of utterances (MLU), speech classes and verb forms. There will be no additional time required on my part for this portion of the analyses.

9. *Rossetti Infant-Toddler Language Scale:* assesses the preverbal and verbal aspects of communication and interaction in the young child. Administration time varies for this test.

10. *Test of Auditory Perceptual Skills-Revised (TAPS-R):* a test that measures a child's ability to perceive auditory matter, providing valuable information in diagnosing auditory difficulties, imperceptions of auditory modality, and/or language problems. The TAPS-R takes approximately 10-15 minutes to administer.

11. *Language Processing Test-Revised (LPT-R):* a test of verbal reasoning and spoken vocabulary to assess a child's ability to organize information