LANGUAGE PHENOTYPES OF AUTISM: TOWARDS AN UNDERSTANDING OF THE COMMUNICATION DOMAIN OF AUTISM

by

ABBY ELISE HARE

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Microbiology and Molecular Genetics

Written under the direction of

Dr. Linda Brzustowicz

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2013

© 2013

Abby Elise Hare

ALL RIGHTS RESERVED

ABSTRACT OF THE DISSERTATION

Language Phenotypes of Autism: Towards an Understanding of the Communication Domain of Autism

By ABBY ELISE HARE

Dissertation Director:

Dr. Linda Brzustowicz

Autism spectrum disorders (ASD) are complex neurodevelopmental disorders that are characterized by deficits in communication, social impairment, and the presence of restricted and repetitive behaviors. The work presented in this dissertation aims to reduce the genetic heterogeneity of samples ascertained for ASD by developing communication phenotypes for use in two genetics studies. Communication impairments in ASD can include impairments in speech or language and, like all traits in ASD, can range in severity from person to person. The first study involved a genome-wide linkage analysis in a sample of multiplex autism families for two non-verbal motor speech (NVMSD) phenotypes: NVMSD:ALL including nonverbal and minimally verbal subjects and NVMSD:C where there is behavioral evidence that language comprehension is relatively intact. Evidence for linkage was identified on several chromosomes: 1q24.2, 3q25.31, 4q22.3, 5p12, 5q33.1, 17p12, 17q11.2, and 17q22 for NVMSD:ALL and 4p15.2 and 21q22.2 for NVMSD:C. Genome-wide analysis and fine mapping of candidate genes did not produce strong evidence for association. The second study identified language (LI) and reading (RI) impairment phenotypes in a dataset ascertained for autism and specific language

impairment (SLI) in the same family. These families were extensively phenotyped with a comprehensive testing battery where all language measures were found to be heritable. In addition to LI and RI, social impairment and obsessive-compulsive behavioral phenotypes were identified in these families using well-respected assessments (SRS and Y-BOCS, respectively). Genome-wide linkage analysis yielded evidence for linkage on 13q21.2 (YBOCS), 14q32.31 (SRS), 15q25.1 (LI), 15q26.2 (SRS), and 16p12.3 (RI). Genome-wide analysis and fine mapping of candidate genes did not produce strong evidence for association. The identification of non-overlapping loci for each phenotype supports the hypothesis that these phenotypes successfully identify unique communication and social impairment loci in ASD. Furthermore, as the second study was conducted in families ascertained for autism and for SLI, these results support the hypothesis that some individuals with ASD and those with SLI without ASD may have some shared genetic etiology. The lack strong evidence for association suggests that rare and/or multiple variants may play a role in the etiology of ASD.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge the families that participated in the studies presented in this dissertation. These families have selflessly volunteered in the hopes of contributing to the knowledge of autism etiology. Without these families, genetic studies such as these would not be possible. Thank you to all who have donated their time, patience, and DNA to both the Autism Genetics Resource Exchange and the New Jersey Language and Autism Genetics Study (NJLAGS).

Scientific research is a collaborative field where many individuals come together from various fields to contribute their knowledge. I would like to thank the following collaborators for their technical and scientific contributions: Dr. Christopher Bartlett and his laboratory for his major contributions to the NJLAGS analyses and expertise in Specific Language Impairment, Dr. Steven Buyske for conducting factor analysis as well as other statistical analyses, Drs. Barbie Zimmerman-Bier and Charles Cartwright for their medical expertise and contributions to the medical evaluations of NJLAGS families, Dr. Derek Gordon for his consultation on simulation analyses of our genetic data, the many testers who have collected data for NJLAGS, and Dr. Veronica Vieland and her colleagues for developing and supporting the statistical methodologies used for family-based linkage and association analyses. I would also like to acknowledge the Rutgers University Cell and DNA Repository Infinite Biologics for DNA storage and genotyping of the NJLAGS samples.

I would like to extend my gratitude to the members of the Brzustowicz laboratory, past and present, who have I had the pleasure of working with: Bill Manley, Ray Zimmerman, Jaime Messenger, Jared Hayter, Carmen Ramirez, Gillian Silver, Zena Fermano, Josh Pennino, Anthony Marcketta, Brenda Patel, Pilar Garavito, Neda Gharani, Mike Moreau, Ariane Seto, and Karen

Law. I would especially like to thank Drs. Marco Azaro and Judy Flax, not only for their countless hours of guidance and collaboration, but also for their friendship and support.

Lastly, I would also like to thank my committee members Drs. Linda Brzustowicz, Jim Millonig, Tara Matise, Bonnie Firestein, and Gerard Costa for their guidance and consultation throughout my graduate school experience. I am deeply grateful to my advisor and mentor, Linda Brzustowicz, for her tremendous support and encouragement. Linda helped me to grow as a scientist by allowing me to pursue my own ideas while keeping me focused. I am eternally grateful to her for the many opportunities I have been given for those that are still to come.

Funding for the AGRE work was provided by NIMH R01MH76433 (Veronica Vieland), R01MH086117 (Veronica Vieland), R01MH76435 (Linda Brzustowicz) and R01MH070366 (Linda Brzustowicz). Funding for the NJLAGS work was provided by NIMH R01MH70366 (Linda Brzustowicz), Challenge Grant RC1MH088288-01 (Linda Brzustowicz), and the New Jersey Governor's Council for Medical Research and Treatment of Autism.

DEDICATION

This dissertation is dedicated to the following people:

My parents, Jeffrey and Cathryn Hare, who taught me the importance of hard work, courage, and perseverance.

My longtime friends, Robin Pierce and Lauren Rees, for their support and comic relief.

And to Matthew Harris, my soulmate and partner in life.

Thank you to all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Autism – General Background

Over the past few decades, there has been an increased interest in understanding the role of genetics in the characteristics, etiology, and treatment of autism. Autism is a complex neurodevelopmental disorder that is characterized by communication deficits, social impairments, and the presence of repetitive and stereotyped behaviors. The incidence of autism has been estimated to be 1 in 88 children in the United States [CDC, 2012]. The rise in prevalence of autism over the past decade has made autism research a top priority in the scientific community. Significant progress has been made in this field, leading to known etiologies of 10-20% of cases of autism. However, due to the combination of environmental interactions, phenotypic heterogeneity, and complex genetic factors, the etiology for the majority of cases of autism remains elusive.

The defining characteristics of autism range in severity, creating a spectrum of symptoms. Autism Spectrum Disorders (ASDs) encompass this range of symptoms and include Autistic Disorder, Asperger Syndrome, and Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS). The Diagnostic and Statistical Manual of Mental Disorders – IV (DSM-IV) provides the current gold standard definition of Autistic Disorder. To qualify for a diagnosis of Autistic Disorder, an individual must display a total of at least six items from the following three categories of impairments with onset prior to the age of three years:

A. At least two qualitative impairments in social interaction must be manifested by marked impairment in the use of nonverbal behaviors, failure to develop peer relationships, lack of spontaneous seeking to share interests with other individuals, or lack of social reciprocity.

B. At least one qualitative impairment in communication including delay in, or lack of, development of spoken language, marked impairment in the ability to initiate/sustain conversation, idiosyncratic language, or lack of social imitative play.

C. At least one restricted repetitive and stereotyped pattern of behavior and interests including preoccupation with restricted patterns of interest, inflexible adherence to routines, repetitive motor mannerisms, or persistent preoccupation with parts of objects.

Additionally, these symptoms cannot be better described by Childhood Disintegrative Disorder or Rett's Syndrome, which are characterized by a loss of acquired skills after age 2. Individuals with Autistic Disorder reside in the severe end of the autism spectrum in all three domains. Asperger Syndrome is distinct from Autistic Disorder due to a lack of clinically significant language delay, while individuals with PDD-NOS exhibit sub-threshold symptomatology or a later age onset than Autistic Disorder.

The diagnostic criteria of ASD have changed with the release of DSM-5 at the American Psychiatric Association Annual Meeting in May 2013. There is a new category called 'Autism Spectrum Disorder' that encompasses Autistic Disorder, Asperger's Disorder, Childhood Disintegrative Disorder, and PDD-NOS [Herold and Connors, 2012]. All of these disorders were defined individually in the DSM-IV. In addition to combining these disorders into one category, the DSM-5 will combine the communication and social domains (Criteria A and B listed above) into one category and will designate restricted and repetitive behaviors as the other category of symptoms. Unusual sensory response has been added to restricted and repetitive behaviors [Grzadzinski et al., 2013]. The DSM-IV requirement of delayed or lack of development of expressive language has been removed from the communication domain due to the high

variability of language ability in autism. The intention of these changes is to provide more accurate diagnoses [Herold and Connors, 2012]. However, these changes have raised concern over the potential for diagnoses to change, which could complicate an individual's qualification for human services and has implications for research cohorts. There are only a few studies that assess the stability of diagnoses from the DSM-IV to the DSM-5. Using the data collected for DSM-IV diagnoses, these studies have shown that 23-40% of ASD diagnoses are changed to unaffected using the DSM-5 criteria [Gibbs et al., 2012; Mattila et al., 2011; McPartland et al., 2012; Worley and Matson, 2012] but more a more recent study found that 91% of DSM-IV diagnosed PDDs were identified using the DSM-5 criteria [Huerta et al., 2012]. However, these studies are extremely preliminary due to the qualitative differences in the information collected for a DSM-IV diagnosis and the information necessitated by the changes in the DSM-5 criteria [Gibbs et al., 2012].

Environmental Factors and Autism

There is a growing body of evidence supporting a role of environmental factors in increased ASD risk. The Childhood Autism Risks from Genetics and Environment (CHARGE) study is the most comprehensive population based case-control study to investigate the role of environmental factors in ASD. The CHARGE study, which is based in California, has investigated three population groups: children with autism, children with developmental delays (not autism), and the general population. This study focuses on the interaction of genetics and a variety of environmental factors including, but not limited to, prenatal factors, maternal metabolic conditions, pesticide exposure, and socioeconomic status. Exposure to pesticides, air pollution, and overall socioeconomic status are included as part of the CHARGE questionnaire. The amount of pesticide and air pollution exposure is difficult to assess directly and has resulted in conflicting reports in regard to pesticide exposure [Shelton et al., 2012]. However, Volk et al.,

2011, reported a positive association between autism risk and residential proximity to freeways during the third trimester and at the date of birth. This suggests that an increased exposure to air pollution may be a contributing environmental factor to ASD. While overall socioeconomic status is not correlated with autism risk, increased toxin exposures (lacquer, solvents, and xylene) at the workplace have been reported at a higher in the parents of individuals with autism [McCanlies et al., 2012].

The CHARGE study also has a strong emphasis on the study of prenatal factors in the development of autism. To date, increased risk for developing ASD has been associated with several prenatal factors, such as maternal diabetes during pregnancy [Krakowiak et al., 2012], maternal fever during pregnancy [Zerbo et al., 2013], and reduced prenatal vitamin intake [Schmidt et al., 2011]. Genetic association to *MTHFR, CBS,* and *COMT* was identified in mothers who did not report taking prenatal vitamins [Schmidt et al., 2011]. The precise role of these prenatal factors in the development of autism is still unclear; however the correlation between ASD and a higher incidence of these factors provides a strong basis for future study. The association between prenatal vitamin intake and several genes supports the gene x environment interaction hypothesis of autism etiology.

Genetics of Autism

Twin and Family Studies

Twin studies have provided strong support for a genetic component of autism. Studies comparing monozygotic (MZ) twins and dizygotic (DZ) twins are used to control for environmental interactions since both twins experience the same *in utero* environment. Early twin studies indicated concordance rates of 60% in MZ twins and concordance rates of 3-5% in DZ twins [Bailey et al., 1995]. A more recent study by Hallmayer et al., 2011, of twin pairs

collected from 1987 to 2004 estimated concordance for a strict definition of autism in male twins at 58% for MZ pairs and 21% for DZ pairs. The estimate concordance for female twins was 60% for MZ pairs and 27% for DZ pairs. Using a broader ASD phenotype, the concordance rate for male twins rose to 77% for MZ pairs and 31% for DZ pairs. The concordance for female twins was 50% for MZ pairs and 36% for DZ pairs. The higher rate of co-occurrence of ASD in MZ twins supports an underlying genetic component of ASD but does not rule out other mitigating factors such as environment.

Family studies also provide support for a genetic component of ASD. Categorical estimates of sibling recurrence risk indicate that a sibling of an individual with autistic disorder has a 22-fold relative risk of developing autistic disorder, the most severe form of ASD [Lauritsen et al., 2005] and the recurrence rate has been estimated to be between 3% and 10% [Ozonoff et al., 2011]. In a longitudinal study conducted by Ozonoff et al., 2011, 664 infants with an older sibling diagnosed on the spectrum, 18.7% of the infants developed an ASD with an three-fold risk of development for male subjects. There was an additional two-fold increase of risk if the male proband had more than one affected sibling. Overall, the incidence of autism in males is four times higher than in females [CDC, 2012].

Each individual with a diagnosis of autism is unique and symptoms range in severity across all three domains of autism. This range of symptoms can manifest at a subclinical level in family members of individuals with autism and are known as the Broad Autism Phenotype (BAP). Characteristics of the BAP are qualitatively similar to those observed in probands and can manifest as mild impairments that are not disruptive, but are evident in an individual's daily life. The BAP can be observed in family members as having a lack of friendships, rigid personalities, impairment in pragmatic language, social cognition impairments, and overall social aloofness [Losh et al., 2008, 2009; Murphy et al., 2000]. The BAP is more likely to appear in family

members of individuals with autism than families without an autism proband [Abrahams and Geschwind, 2008]. Both simplex and multiplex ASD families have significantly higher instances of BAP characteristics than control families [Losh et al., 2008]. Families with multiple instances of autism have a higher incidence of BAP in parents of the proband than families with a single instance of autism and there is an increase of quantitative autistic traits in the unaffected siblings in multiple incidence families, suggesting an increased genetic loading in these families [Constantino et al., 2010]. There is also a higher prevalence of BAP characteristics in first- and second-degree relatives of individuals with ASD than control families. There is an increased incidence of reading and writing impairment, learning impairments, and anxiety in first-degree relatives and an increased incidence of language impairment and obsessive compulsive behavior in second-degree relatives [Micali et al., 2004].

Known Genetic Causes of ASD

It is estimated that 10% of ASDs can be attributed to an identifiable Mendelian condition or genetic syndrome [Devlin and Scherer, 2012]. Approximately 5% of ASD cases can be attributed to rare genetic syndromes such as Rett Syndrome, Fragile X syndrome, Smith-Lemli-Opitz syndrome, and Tuberous Sclerosis [Geschwind, 2011; Carter and Scherer, 2013; Devlin and Scherer, 2012]. Both Fragile X and Rett syndromes involve defects in genes located on the X chromosome. Fragile X syndrome (≈1-2% of ASD cases) is caused by the silencing of the fragile X mental retardation gene, *FMR1,* which leads to ASD characteristics in 15-33% of individuals with this silencing [National Fragile X Foundation, 2013]. Rett syndrome (≈0.5% of ASD cases) involves defects in methyl-CpG-binding protein 2, *MeCP2*. This syndrome typically affects females and is characterized by a regression in social engagement and psychomotor function [PubMed Health, 2012]. Smith-Lemli-Opitz syndrome (rare, <1% of ASD cases) is caused by mutations in sterol delta-7-reductase, *DHCR7*, which controls cholesterol synthesis. Mental

retardation, cleft palate, and polydactyly are common in Smith-Lemli-Opitz syndrome and it is estimated that 75% affected individuals can exhibit behaviors that meet the criteria for autism [Sikora et al., 2006]. An estimated 25-50% of individuals with Tuberous Sclerosis exhibit features of ASD. Tuberous Sclerosis (≈1% of ASDs) is caused by mutations in the tuberous sclerosis-1, *TSC1,* or tuberous sclerosis-2, *TSC2,* genes and is multisystem disorder that is characterized by hamartomas in the brain, skin, heart, kidneys and lungs [OMIM, 2013]. These syndromic forms are considered to be exceptional cases and each accounts for less than 1% of ASD cases [Geschwind, 2011; Devlin and Scherer, 2012].

Genetic Studies of Autism

The majority of ASD cases have unknown genetic etiology. Linkage analyses, genome-wide association analyses, copy number variation identification, murine models of candidate mutations, and rare variant identification through next-generation sequencing have produced a vast growing number of candidate regions for ASD, often with conflicting results. The following sections will review the strongest findings to date from these analyses.

Copy Number Variation (CNV)

Copy number variants (CNVs) account for up to 10% of idiopathic ASDs [Geschwind, 2011]. The majority of CNVs found in autism are *de novo* mutations that occur in the germline and have a pleiotropic effect. However, CNVs at 15q11-13, 16p11.2 and 22q11.2 have been reported to be inherited from an unaffected parent. About 1-3% of individuals with ASD report interstitial duplications (of the maternal allele) at 15q11-13 which encompasses 12 genes over a 5Mb region of the Prader-Willi/Angelman syndrome region [Abrahams and Geschwind, 2008; Carter and Scherer, 2013; Devlin and Scherer, 2012]. CNVs at 16p11.2 encompass 30 genes over a 700kb region and occurs in a higher frequency in ASD (1% of ASD cases) than in the general

population [Devlin and Scherer, 2012]. Deletions in this region tend to be *de novo* mutations and result in intellectual disability and difficulty with expressive language. Duplications are more commonly inherited and result in abnormally small head circumference [Carter and Scherer, 2013]. Deletions at 22q11.2 are associated with DiGeorge Syndrome, which is a syndrome that presents with physical abnormalities, such as congenital heart disease and cleft palate, as well as learning disabilities and psychiatric disorders which can include characteristics of ASD. While up to 50% of individuals who are carriers of a deletion at 22q11.2 have features of ASD [Vorstman et al., 2006], only 0.2% of cases of ASD can be attributed to this deletion [Devlin and Scherer, 2012]. A duplication at 7q11.23, which accounts for 0.2% of ASD cases, results in reduced verbal and social skills consistent with those seen in ASD [Carter and Scherer, 2013; Abrahams and Geschwind, 2008]. Both deletions and duplications at 1q21.1 account for 0.2% of ASD cases. Deletions in this region result in microcephaly, mild dysmorphic facial features, and mild intellectual disability while duplications result in macrocephaly, developmental delay and intellectual disability [Carter and Scherer, 2013]. The CNVs reported at 1q21.1, 7q11.23, 15q11-13 and 22q11.2 do not occur at a statistically higher rate in the ASD population than the general population [Devlin and Scherer, 2012], but each has a large effect size and are often included in clinical screenings for ASD [Carter and Scherer, 2013]. In addition to the identification of these rare variations, there has been a higher occurrence of *de novo* variations in multiplex and simplex ASD families when compared to controls [Abrahams and Geschwind, 2008]. However, most of reported CNVs are single occurrences and are rarely replicated in unrelated individuals. This could be due to the lack of power in small disease-specific cohorts. By combining CNV information from large clinical case-control cohorts, Moreno-De-Luca et al., 2012, were able to add statistical support for several deleterious CNVs in

ASD cohorts that may not have reached significance in smaller studies. This study predicts that rare CNVs may have a stronger effect on ASD risk than previous studies have indicated.

Linkage Studies

Over the past several decades, genome-wide linkage scans have been used as an approach to identify autism susceptibility loci. Several dozen family studies of ASD have identified genetic risk factors on almost every chromosome, however replication of these studies is hindered by the clinical and genetic heterogeneity among the samples being studied [Abrahams and Geschwind, 2008]. Linkage to ASD diagnosis has been identified by at least two independent studies in the following regions: 2q, 3q25-27, 3p25, 6p14-21, 7q31-36 and 17q11-21 (reviewed in: Freitag, 2007). The majority of linkage findings result in non-overlapping regions of interest that are unable to be replicated across samples, suggesting the existence of considerable genetic heterogeneity of samples. Two notable exceptions are the replicated loci on chromosomes 7 and 17.

The locus on chromosome 7 has been linked to ASD and other endophenotypes of ASD, including language phenotypes [Alarcón et al., 2005; Alarcon et al., 2002] and developmental regression [Molloy et al., 2005]. A regional meta-analysis of four linkage scans for a strict definition of autistic disorder [IMGSAC, 1998, 2001; CLSA et al., 1999; Risch et al., 1999; Philippe et al., 1999] revealed genome-wide significance on chromosome 7 [Badner and Gershon, 2002]. A separate heterogeneity-based genome search meta-analysis of nine linkage scans that included IMGSAC 2001, Risch et al. 1999, and Philippe et al. 1999, supported genome-wide significance for this region and revealed low between-scan heterogeneity for this region of chromosome 7 [Trikalinos et al., 2006]. The linkage region on Chromosome 17 is primarily seen in subsets of families from the Autism Genetics Resource Exchange (AGRE) database, which is a collection of genotypic and phenotypic data for multiplex autism families [Stone et al., 2004;

Cantor et al., 2005; Yonan et al., 2003; Bartlett et al., 2005; Flax et al., 2010; McCauley et al., 2005]. It is important to note that studies using the AGRE families inevitably have a considerable amount of sample overlap, which contributes to the replication of linkage findings in this dataset. The heterogeneity-based genome search meta-analysis performed by Trikalinos et al., 2006 included the studies conducted by Yonan et al., 2003 and Cantor et al., 2005. Suggestive signficance was reached for 17p11.2-q12 with significantly high between-scan hetergeneity, further supporting the evidence for high genetic heterogeneity in autism samples.

Association Analyses

Association analyses are used for identifying common variants within a population that may confer risk to a complex disorder. Several loci have been implicated in ASD using association analyses both on a selective candidate gene level and genome-wide scale, however, as with linkage analyses, the results of association analyses often produce conflicting results in non-overlapping regions. There have been three large genome-wide association studies (GWAS) that have investigated the role of common variants in ASD [Wang et al., 2010; Anney et al., 2010; Weiss et al., 2010]. Wang et al., 2010, identified association to a locus at Chromosome 5p14.1 in a sample selected from the AGRE and the Autism Case-Control (ACC) cohorts. This finding was replicated using the Collaborative Autism Project (CAP) and Center for Autism Research and Treatment (CART) cohorts [Wang et al., 2010]. A study by Weiss et al., 2010 also found association using the AGRE cohort in combination with National Institute of Mental Health (NIMH) samples, however their findings on Chromosome 5p15.2 do not overlap with the findings on 5p14.1. The third large scale GWAS identified association to Chromosome 20p12.1 using an initial sample set from the Autism Genome Project (AGP) that was supplemented with families from the AGRE database [Anney et al., 2010]. While some of the samples from AGRE overlapped in all three studies, the markers on the genotyping arrays utilized by each group did

not. This may partly explain the lack of replication in these regions; however a more likely explanation is the heterogeneous nature of these datasets.

The most consistent association findings have been the result of smaller investigations of candidate genes that have been identified as ASD susceptibility genes. One of the most replicated findings is association to *SLC6A4,* which is a serotonin transporter gene located on chromosome 17q11.2 that has been implicated in elevated blood serotonin levels in children with autism and their first-degree relatives [Anderson et al., 1987; Abramson et al., 1989; Leboyer et al., 1999]. A number of independent investigators have identified association to the short allele of 5HTTLPR region and a variable number of tandem repeats in intron 2 of *SLC6A4* [Tordjman et al., 2001; Sutcliffe et al., 2005; Mulder et al., 2005; McCauley et al., 2004; Conroy et al., 2004; Coutinho et al., 2004; Devlin et al., 2005; Kim et al., 2002; Klauck et al., 1997]. Despite the large body of work supporting evidence for association to *SLC6A4*, several other studies have shown evidence against association in various populations [Betancur et al., 2002; Koishi et al., 2006; Persico et al., 2000; Cross et al., 2008].

The *engrailed 2* (*EN2)* gene, which is located on Chromosome 7q36, has been shown to be associated in a family study using samples from the AGRE and NIMH databases [Benayed et al., 2005; Gharani et al., 2004] and in a case-control study [Petit et al., 1995]. The common alleles of rs1861972 (A) and rs1861973 (C) are over-transmitted to individuals affected for autism and this haplotype has been shown to have a functional effect on the expression of *EN2* in human and rat cell lines and mouse neuronal cultures [Benayed et al., 2009]. *EN2* encodes a homeobox transcription factor that plays a role in cerebellum and brainstem development. The expression of *EN2* is increased in post-mortem cerebellar samples from individuals with autism when compared to controls [James et al., 2013]. This overexpression in individuals with ASD

suggests that *EN2* may not be properly down-regulated as normally occurs in unaffected individuals to mediate Purkinje cells maturation. A knock-out mouse model of *EN2* displayed cerebellar abnormalities consistent with ASD as well as deficits in social behavior, spatial learning and memory tasks [Cheh et al., 2006; Brielmaier et al., 2012].

Another candidate gene that has been associated with autism is the *Reelin* gene*, RELN,* which is located on Chromosome 7q22. Reduced plasma levels of Reelin have been reported in individuals with autism and their first-degree relatives [Fatemi et al., 2002]. Reelin signaling, which is required for neuronal migration and the formation of cortical layers [Yamashita et al., 2006], has been shown to be impaired in post-mortem cortices of individuals with autism [Fatemi et al., 2005]. Association analyses of variants in *RELN* have been inconsistent with five studies reporting evidence for association [Serajee et al., 2006; Persico et al., 2001; Skaar et al., 2005; Sharma et al., 2013; Li et al., 2013] and several studies reporting evidence against association particularly in Chinese populations [He et al., 2001; Dutta et al., 2008; Li et al., 2008; Zhang et al., 2002; Bonora et al., 2003; Devlin et al., 2004; Krebs et al., 2002; Li et al., 2004].

Rare Variants identified by Next-Generation Sequencing

As seen in the association studies discussed above, there is conflicting evidence for a role of common variants in ASD. The lack of consistent replication of common variants has lead the field to investigate the role of rare variants, point mutations, and *de novo* mutations in ASD through the use of next-generation sequencing (NGS) at the whole genome (WGS) and, more commonly, the whole exome (WES) levels. Through the studies conducted to date, it is estimated that 15-20% of ASD cases may be accounted for by rare *de novo* events [Yu et al., 2013]. WGS of 10 sets of MZ twins concordant for ASD and their parents showed that *de novo* mutations identified in the twin pairs were distributed in a nonrandom pattern and were

observed to be closely spaced within an individual's genome [Michaelson et al., 2012]. This study also found that the genome-wide rate of mutations in individuals with ASD is not higher than the estimated rate of mutations in the general population, suggesting that the distribution of the mutations may incur a higher effect than the number of mutations [Michaelson et al., 2012]. In fact, the mutations they identified that fell within exonic regions overlapped with findings of WES in several independent studies of ASD [Neale et al., 2012; O'Roak et al., 2011, 2012; Sanders et al., 2012].

As the cost of deep WGS is prohibitive for many studies, research groups have focused on the use of WES to identify rare variants in families ascertained for ASD. To date, four major studies have completed WES in a combination of case-control and family studies. The first of these studies sequenced 20 individuals with sporadic ASD and their parents with the hypothesis that these families would be enriched for *de novo* mutations. A total of 201 *de novo* mutations were identified and 11 of these were protein altering within the following genes: *FOXP1, GRIN2B, SCN1A,* and *LAMC3* [O'Roak et al., 2011]. The overall rate of *de novo* events in protein-coding regions was slightly higher than expected. An independent study of 175 ASD probands and their parents found that the overall mutation rate in ASD probands was not significantly higher than expected, however the number of nonsense mutations identified was significantly high [Neale et al., 2012]. Additionally, they observed a correlation between the number of *de novo* events per offspring with paternal and maternal age [Neale et al., 2012]. The increased rate with paternal age was also seen in a study of WES in 238 families from the Simons Simplex Collection by Sanders et al., 2012. Of the 279 identified *de novo* coding mutations, three genes (*SCN2A, KATNAL2,* and *CHD8*) harbored disruptive mutations in two individuals. Like other sequencing analyses, this study observed that the overall mutation rate was not higher in probands [Sanders et al., 2012].

These studies demonstrate the presence of rare variants in ASD, however the effect size of these variants is still unknown. Overall, the rate and distribution of *de novo* mutations in ASD probands does not deviate from that of the general population. A single observation of a mutation in a sample is not sufficient evidence to implicate a gene as a risk factor for ASD due to its small effect on the overall sample. However, with the acquisition of larger datasets and meta-analyses, the effect size of these rare mutations may increase as statistical power increases and multiple instances are discovered [Allen et al., 2010; Ke, 2012; Maher et al., 2010]. Neale et al., 2012 observed that genetic models that fit the data seen in these sequencing analyses support a polygenic model suggesting that the combination of variants, not the number of variants, contribute to ASD risk.

Targeted Candidate Genes

The culmination of linkage, association, CNV, and sequence analyses has created an ever growing list of candidate genes for ASD susceptibility. One of the most studied sets of candidate genes is the Neurexin and Neuroligin gene families. Neurexins are presynaptic membrane molecules that act as trans-synaptic cell-adhesion molecules by forming a complex with neuroligins, which are located at the postsynaptic membrane [Boucard et al., 2005; Ichtchenko et al., 1995]. These complexes play a key role in proper synaptic formation and function [Missler et al., 2003] (reviewed in [Reichelt et al., 2012]). Neither of these gene groups have been consistently implicated in association analyses, however numerous investigators have identified mutations that could lead to improper synaptic functioning (reviewed in [Reichelt et al., 2012; Li et al., 2012; Abrahams and Geschwind, 2008]). Etherton et al., 2009, suggests that as many as 0.5% of ASD cases may be due to partial *NRXN1*α deletions and several studies have identified individuals with autism or autistic-like features who have hemizygous partial deletions of *NRXN1*α and *NRXN1*β. Of the five *NLGN* genes that have been identified, two of them,

NLGN3 and *NLGN4*, have been identified as ASD susceptibility genes (reviewed in [Li et al., 2012]). Rare mutations including an inherited non-synonymous point mutation, a *de novo* nonsense mutation, one missense variant, two base pair substitutions, and splice variants in both genes have been identified in independent autism samples [Zhang et al., 2009; Pampanos et al., 2009; Yan et al., 2008; Jamain et al., 2003]. The collection of variants identified in these genes lends support to the hypothesis that these rare variants represent rare causes of autism. While individually these variants do not confer a high risk for autism, the high number of variants found within these gene families makes the Neurexin and Neuroligin gene families prime autism susceptibility candidates.

NRXN1 and *NLGN3* have also been studied using mouse models to investigate the effect of these mutations on brain structure and function as well as behavioral (social, learning, grooming, memory, etc.) effects. A knock-in (KI) mouse of the *Nlgn3* non-synonymous single nucleotide polymorphism (SNP), R451C, alters inhibitory postsynaptic currents in the somatosensory cortex [Etherton et al., 2011] and excitatory postsynaptic currents in hippocampal region of the brain [Tabuchi et al., 2007]. This KI mouse model also exhibits impaired social interaction, while a knock-out (KO) of *Nlgn3* did not exhibit abnormalities, suggesting that this point mutation acts as a gain-of-function mutation [Shinoda et al., 2013; Etherton et al., 2009; Tabuchi et al., 2007; Etherton et al., 2011]. Another line of *Nlgn3* KI mice developed by Chadman et al., 2008, however, this line did not exhibit abnormal behavior. More consistent results have been produced using a KO of *Nrxn1* and a triple knock-out of all three α-Neurexin variants. Both models exhibit defective synaptic phenotypes and the single KO mice exhibit increased grooming behavior and decreased nest building behavior [Etherton et al., 2009]. These mouse models support the role of defective synaptic functioning in individuals with autism who harbor mutations within this complex.

In addition to the Neurexin-Neuroligin models, several other candidate genes with strong support for implications in ASD have been studied in the murine system (for a comprehensive review see [Shinoda et al., 2013; Ey et al., 2011]). These studies have resulted in several categories of phenotypes including social interactions, anxiety, motor learning and function, vocalizations, seizures, and altered response to stressful stimuli [Ey et al., 2011]. In addition to the Neurexin and Neuroligin mouse lines described above, lower levels of social interactions were observed for *Shank1-KO* and *Pten-KO* mice. Mutations in all three members of the Shank gene family (*SHANK1, SHANK2, SHANK3*) have been reported in individuals with ASD [Shinoda et al., 2013]. The Shank proteins are scaffolding proteins that tether neuroligins and NMDA receptor complexes at the excitatory postsynaptic density. The *Shank1-KO* line also exhibited increase anxiety, decreased locomotion, decreased long-term memory, and enhanced working memory. Mutations in *PTEN,* which is a gene that encodes a lipid and protein phosphatase that plays a role in brain morphology and synaptic function, are estimated to be present in 5-10% of individuals with ASD resulting in morphological abnormalities, spontaneous seizures and deficits in social and cognitive behaviors [Takeuchi et al., 2013]. Takeuchi et al., 2013 recently produced the conditional KO of *PTEN* that presented with decreased social interaction and locomotor deficits as well as increased anxiety and seizures [Takeuchi et al., 2013].

Methods to Reduce Heterogeneity

There is compelling evidence supporting a genetic risk for autism; however, genetic studies have produced inconsistent and often conflicting results as to the specific genetic etiology. Despite these inconsistencies, one conclusion that can be drawn from these studies is that there is an incredible amount of locus and allelic heterogeneity in ASD. This genetic heterogeneity is accompanied by phenotypic heterogeneity throughout all three domains of

autism. Susceptibility to ASD may be the result of a collection of variants that contribute to the specific phenotypic domains of ASD. This hypothesis is supported by the fact that the individual domains of ASD are independently inherited in the general population. Researchers have used this range of phenotypic characteristics of ASDs to develop endophenotypes in order to reduce this heterogeneity through several strategies. To qualify as an endophenotype a trait should ideally 1) be associated with the disease in the population, 2) be heritable, 3) be state-independent, 4) co-segregate with the disease within a family, and 5) for complex disorders, be found in affected families (both affected and non-affected family members) at a higher rate than the general population [Gottesman et al., 2003]. Endophenotypes can be either categorical or quantitative and can be utilized to reduce heterogeneity by a few methods. One method that has been successful is to separate a study sample into clinically more similar groups based on more specific characteristics of ASD. Several studies have stratified samples based on proband sex, phrase speech delay (PSD), and IQ. A second method is to use quantitative measures of ASD traits, such as age of first word, social responsiveness, regression, repetitive behaviors, and IQ, to map specific quantitative trait loci (QTL). Of the three domains of autism, language and communication phenotypes have been the most extensively studied.

Studies that focus on a stratified ASD sample have successfully identified linkage by reducing the heterogeneity of the selected sample subset and subsequently increasing the likelihood of identifying causative loci. A study by Vieland et al., 2011, stratified the Autism Genome Project (AGP) dataset by dividing the sample into three IQ groups: low IQ (LIQ), missing IQ (MIQ), and normal IQ (NIQ). Linkage analysis of the entire dataset identified evidence for linkage on Chromosomes 4, 11, and 16. Analysis of the individual IQ groups identified which groups contributed to each linkage signal and revealed evidence for linkage that was exclusive to the LIQ group. The LIQ group contributes strongly to the finding on Chromosome 16 and

partially to Chromosome 11. The NIQ group also contributes to the finding on Chromosome 11. The MIQ group does not appear to increase the evidence for linkage in any of these regions. The finding on Chromosome 4 was not seen in the analyses of the individual groups, suggesting that the entire sample contributes to this signal. In addition to the linkage regions identified by all groups, linkage was identified on Chromosomes 1 and 13 for the LIQ group only. Since these signals are only seen in the LIQ group, the MIQ and NIQ groups must provide strong evidence against linkage to eliminate the signal in these regions for the entire group. This finding supports the hypothesis that reduction of sample heterogeneity helps to identify causative loci to a specific group of individuals with ASD.

Sample stratification was also used to reduce the heterogeneity of the AGRE database by separating the sample by the presence or absence of PSD in autism probands. PSD was defined as the development of phrase speech after 36 months. Buxbaum et al., 2001 identified 95 families from the AGRE database that contained two individuals with ASD and PSD. Evidence for linkage in these families was identified in a region on Chromosome 2q24-32. This finding was replicated by Shao et al., 2002 using a set of 45 sib-pair families affected for PSD from the Collaborative Autism Team (CAT) dataset. Use of PSD to stratify ASD samples does not always replicate this finding on Chromosome 2. An independent study of a subset of the Collaborative Linkage Study of Autism (CLSA) cohort identified linkage to ASD on chromosomes 7q22-32 and 13q21-22 using a subset of sib-pair families (n = 50) where both probands met the criteria for PSD [Bradford et al., 2001]. While linkage to these regions has not been replicated using PSD, the finding on Chromosome 13q has also been linked to Specific Language Impairment (SLI) [Bartlett et al., 2002, 2004; Simmons et al., 2010].

The second method for reducing sample heterogeneity utilizes QTL mapping by analyzing the degree of genotype sharing and the similarity of the quantitative trait associated with ASD. In this method, the sample is not subdivided. Rather, the genotype and trait data of the entire sample is analyzed. Quantitative traits for ASD have been defined primarily using specific variables from the Autism Diagnostic Interview-Revised (ADI-R), which is a 93-item questionnaire administered to the parents or caregivers of the probands and is one of the gold-standard diagnostic algorithms for autism. Individual ADI-R variables, such as 'age of first word', can be used as reported or combined with several other ADI-R variables. Using the individual ADI-R variables 'onset of age of first word' and the 'onset of phrase speech in months', linkage was identified to a region on Chromosome 7q36 in 152 multiplex families from AGRE dataset [Alarcon et al., 2002]. This finding was replicated by supplementing the original 152 families with 139 additional multiplex families [Alarcón et al., 2005]. Another study identified linkage to Chromosome 7 using a combination of ADI-R variables that measure nonverbal communication in 284 sibling pairs from the AGRE database [Chen et al., 2006]. This study identified suggestive linkage to 1p13-q12 and possible linkage to 4q21-25, 7q35, 8q23-24, and 16p12-13. While the linkage regions on Chromosome 7 do not overlap, they both reside within regions that have been implicated in language impairment [Alarcon et al., 2002; Etherton et al., 2011; Vernes et al., 2008; Newbury et al., 2011].

In addition to the communication domain, the ADI-R has also been used to derive quantitative traits for the other two domains of ASD. Liu et al., 2008, developed two quantitative traits from the total scores for the reciprocal social interaction (SI) domain and the restricted and repetitive behavior (RRB) domain of the ADI-R. Evidence for linkage to these phenotypes was suggestive on Chromosome 12q13.11 (SI) and 14q22.1 (RRB). This finding was not replicated in a study by Duvall et al., 2007. However, they developed a more powerful

quantitative measure of SI based on the Social Responsiveness Scale (SRS) to identify linkage on Chromosomes 11p and 17p in 100 multiplex AGRE families [Duvall et al., 2007]. Linkage to additional loci on Chromosomes 4, 8 and 10 were identified using only affected males in this dataset.

A more recent study by Connolly et al., 2013 used all of the variables from the ADI-R and SRS as well as the Autism Diagnostic Observation Schedule (ADOS) as quantitative traits in a genome-wide association analysis of the AGRE database. Association was identified using the ADI-R variables in the following genes: *NOS2A* ('loss of motor skills'), *FER* ('loss of motor skills'), *NELL1* ('faints, fits or blackouts'), and *BIN1* ('general loss of skills'). *MPN2* was associated with the ADOS variable 'functional play with objects'. Additionally, several variables from the SRS were found to be associated with the following genes: *KCND2* ('has overly serious facial expressions'), *SDK1* ('is too tense in social situations'), and *C8ORFK32* ('concentrates too much on parts of things'). When this study was replicated in the AGP sample, only association to *KCND2* remained significant on the genome-wide level [Connolly et al., 2013].

Language and Autism

Analyses of language phenotypes in ASD have identified a link between genes involved in ASD and language impairment. Both ASD and language impairment have been shown to be heritable independent of each other. As language impairment is part of the communication domain of ASD, it is conceivable that ASD and language impairment may share some common etiology. This hypothesis is supported by the heritability of phonological processing in autism and language impaired families. Non-word repetition (NWR) tasks, which are a measure of phonological processing, have been shown to be heritable in twin studies of language impairment [Bishop et al., 1996, 1999, 2004a]. Poor performance on NWR tasks has also been

identified in ASD, suggesting a potential overlapping language phenotype in ASD and language impairment [Kjelgaard and Tager-Flusberg, 2001]. As is the case with many traits of ASD, the language impairment seen in ASD probands is highly variable with some individuals having normal language and others who are severely impaired. The performance profile of ASD probands with language impairment is similar to that seen in Specific Language Impairment (SLI). SLI is a neurodevelopmental disorder that is defined as the failure to acquire and/or use language normally in the absence of speech-motor or sensory deficits and mental retardation. SLI and autism are mutually exclusive disorders, however both included deficits in structural language such as difficulties with grammar and semantics, and low performance phonological tasks such as on non-word repetition. Leyfer et al., 2008, assessed two groups of individuals with SLI and ASD with both the ADI-R and ADOS. While the ASD group scored significantly higher (i.e. more severely affected) than the SLI group on both measures, there was overlap in domain scores for both groups. Using the ADI-R, 14% of the SLI sample scored above the ASD cutoff for the social domain and 11% was above the cutoff for the communication domain. Using the ADOS, 18% of the SLI sample met the ADOS cutoff for the social domain and 25% met the cutoff for the communication domain. The presence of these difficulties in each disorder may be due to some overlap in etiology.

In addition to qualitative similarities in language impairment between SLI and ASD, there is strong evidence for a genetic link between the two disorders. The linkage region on Chromosome 7q35-36 that was identified in linkage scans for the 'onset of age of first word' and the 'onset of phrase speech in months' variables in the ADI-R using the AGRE dataset is considered to be a strong link between language impairment and ASD [Alarcon et al., 2002]. This region on Chromosome 7q contains *CNTNAP2*, which is a member of the neurexin family that functions in nervous system development and differentiation of axons and is down-

regulated by FOXP2, a transcription factor that has been implicated in speech and language disorders [Vernes et al., 2008]. Alarcon et al., 2008 identified association of 'age of first word' in ASD to the C allele of rs2710102, a single nucleotide polymorphism (SNP) located within *CNTNAP2*. The T allele of rs2710102 has been associated 'nonsense-word repetition' in developmental dyslexia [Peter et al., 2011]. Other SNPs in this gene have been associated with SLI and early language development in the general populations [Vernes et al., 2008; Newbury et al., 2011] and individuals who were homozygous for the T allele of rs779745 showed significant cerebral morphological variation [Zeeland et al., 2011]. In addition to language impairment, variants in *CNTNAP2* have been associated with social behavior endophenotypes in autism populations [Leyfer et al., 2008; Steer et al., 2010]. For a current review of additional studies of *CNTNAP2,* see Peñagarikano and Geschwind, 2012. The identification of linkage to *CNTNAP2* in the AGRE dataset indicates the effectiveness of the use of language phenotypes to identify genetic loci that play a role in the language impairment aspect of ASDs. It is important to note that while these studies used multiplex autism families, the affection status utilized was the language phenotype, not autism diagnosis.

While many individuals on the autism spectrum develop language skills, it is estimated that 20-50% of individuals with autism do not acquire functional language [Lord et al., 2004; Bailey et al., 1996]. While some individuals with ASD may not develop adequate functional expressive language due to an underlying social impairment, others may have language processing impairments or motor speech deficiencies making speech extremely effortful or unintelligible, similar in character to deficits seen in severe cases of Childhood Apraxia of Speech (CAS). CAS is a neurological childhood speech sound disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits that affects 1 to 2 children per 1000 [Shriberg et al., 1997]. There has been limited study

of CAS in terms of its genetic origins; however, in their family and genetic studies of speech sound disorders, Lewis et al., 2004 looked at a small sample of children with a diagnosis of CAS and reported that 59% had at least one parent with some type of speech sound disorder. Moreover, in 86% of the families, at least one nuclear family member reported either a speech sound disorder or a language disorder. In a recent related study of speech sound disorders Lewis et al., 2007 report that 36 of 147 (24%) of parents of children with speech sound disorders also report similar problems as children. The presence of speech-sound disorders and language impairments in first-degree relatives of CAS probands suggests a genetic component. The degree of motor-speech impairment in non-verbal and minimally verbal children with autism reflects behaviors seen in the most impaired end of the CAS spectrum. It is not unreasonable to use CAS symptoms to define motor speech phenotypes of ASD much like genetic studies have done with SLI and language phenotypes of autism.

Rationale for Thesis Experiments

Autism is characterized as a range of symptom severity in the three domains: communication impairment, social impairment, and the presence of restricted/repetitive behaviors. Susceptibility to autism may be the result of a collection of variants contributing to the specific phenotypic domains of autism and reflected in a spectral range of patterns and behaviors. In particular, impairments in communication can be present in several different linguistic levels. Communication impairment involves the structure of speech and language and how phonology, grammar, and word meaning are used to convey information to others. In autism, communication impairments can also include the social aspects of verbal and non-verbal language; the use of language to communicate desires, and a general need to communicate with the world. All of these domains of speech and language have also been implicated in subgroups of individuals with autism. Furthermore, these impairments are distinct from each

other and have been shown to be heritable in the general population. The purpose of the studies in this thesis is to examine the genetics underlying specific communication phenotypes in two autism cohorts.

Nonverbal Motor Speech Disorder in AGRE

The first study presented in this thesis investigates the genetics of a non-verbal motor speech phenotype identified in the AGRE database that is similar to the characteristics of CAS. Using information from the ADI-R, 46% of individuals with ASD in AGRE are nonverbal or are unintelligible to others. While there is not a direct measure of CAS-like characteristics in the AGRE database, several ADI-R variables asses the level of language and articulation. The variables 'overall level of language', 'current articulation', 'articulation at age 5', 'comprehension of simple language', and 'comprehension of simple language at age 5' were used to define two nonverbal motor speech disorder (NVMSD) phenotypes in the AGRE database. The first phenotype, NVMSD:ALL, includes individuals who have fewer than five words or are unintelligible to others. The second phenotyped, NVMSD:C, is a subset of individuals from the NVMSD:ALL group who have comprehension of simple language, modeling CAS-like characteristics in an ASD cohort.

The study in Chapter 2 conducted linkage and association analysis of these two NVMSD phenotypes in families in the AGRE database. This study used only families who were multiplex for the NVMSD phenotypes and individuals who had existing genotype data available from AGRE. Following linkage analysis, Ingenuity Pathway Analysis (IPA) was used to select candidate genes for further analysis.

Chapter 3 presents a follow-up study to Chapter 2. The initial linkage scan for NVMSD:ALL was updated with genetic information from AGRE families who met criteria for

NVMSD:ALL but were not included in the initial study. The candidate genes selected in Chapter 2 were fine mapped for association analysis. This study also determined the power of the sample to detect association.

New Jersey Language and Autism Genetics Study

The connection between autism and SLI contributes to the rationale for the New Jersey Language and Autism Genetics Study (NJLAGS). Over the past decade, NJLAGS has collected a cohort of families that contain at least one ADIR/ADOS-confirmed autism proband and another proband with testing-confirmed SLI. This study design allows for the extensive characterization of the genetic connection between ASD and SLI by increasing the genetic loading for language impairment in these families. Each family contains at least five individuals who have been extensively phenotyped using a comprehensive neuropsychological testing battery that measures characteristics all three domains of autism with an emphasis on the communication domain. The phenotypic information collected from the testing battery measures language skills, reading ability, social impairments and obsessive-compulsive behavior in the family members and can be used to develop categorical and quantitative endophenotypes of ASD.

Chapter 4 describes the analysis of the phenotypic data collected from the testing battery in the NJLAGS sample. This study analyzes the correlation of testing scores between groups affected for language impairment, reading impairment, language and reading impairment, and the autism probands.

The study in Chapter 5 analyzed the heritability of each language subtest in the NJLAGS families while controlling for the influence of the autism and SLI proband on lower testing scores. The heritability was used in hierarchical clustering analysis of the language subtests to reduce the amount of quantitative variables.

Chapter 6 presents genome-wide linkage and association analysis of the 53 NJLAGS families who met the strictest definition of at least one narrowly defined autism proband and another SLI proband. These analyses investigated linkage and association for each of the phenotypes defined in Chapter 4 and the quantitative factors defined in Chapter 5.

Chapter 7 is a fine mapping association analysis of candidate genes selected from the linkage regions identified in Chapter 6. The candidate genes were selected using a scoring method derived to prioritize information from Ingenuity Pathway Analysis (IPA).

The over-arching goal of this thesis is to use communication impairments to create endophenotypes of autism in order to reduce the heterogeneity of our genetic analyses. The studies presented here approach communication impairments from two perspectives: speech and language. Speech is the verbal means of communicating that includes articulation, voice, and fluency. Language, on the other hand, refers to understanding what words mean, how to put words together, and the use of the appropriate word combinations in different situations. The communication domain of autism can manifest as either a speech (nonverbal) or language (receptive/expressive) impairment and can range in severity. The NVMSD endophenotypes focus on speech (or lack thereof in subjects who are non-verbal), while the NJLAGS phenotypes place emphasis on language impairments. NJLAGS takes this a step further by investigating oral language impairments (syntax, semantics, phonology, and pragmatics) and written language impairments (decoding, comprehension, reading fluency, and spelling), as both represent discreet forms of language impairments. By identifying phenotypes for each of these areas, we hope to also identify distinct genetic etiologies for these traits in families ascertained for autism.

Chapter 2: Combined linkage and linkage disequilibrium analysis of a motor speech phenotype within families ascertained for autism risk loci

Judy F. Flax, Abby E. Hare, Marco A. Azaro, Linda M. Brzustowicz

Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ

Veronica J. Vieland

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH

Keywords: Autism, Linkage Analysis, Association Analysis, Motor Speech Disorder, Autism genes, Ingenuity Pathway Analysis (IPA), PPL/PPLD

Published in The Journal of Neurodevelopmental Disorders

Abstract

Using behavioral and genetic information from the Autism Genetics Resource Exchange (AGRE) data set we developed phenotypes and investigated linkage and association for individuals with and without Autism Spectrum Disorders (ASD) who exhibit expressive language behaviors consistent with a motor speech disorder. Speech and language variables from Autism Diagnostic Interview-Revised (ADI-R) were used to develop a motor speech phenotype associated with non-verbal or unintelligible verbal behaviors (NVMSD:ALL) and a related phenotype restricted to individuals without significant comprehension difficulties (NVMSD:C). Using Affymetrix 5.0 data, the PPL framework was employed to assess the strength of evidence for or against trait-marker linkage and linkage disequilibrium (LD) across the genome. Ingenuity Pathway Analysis (IPA) was then utilized to identify potential genes for further investigation. We identified several linkage peaks based on two related language- speech phenotypes consistent with a potential motor speech disorder: chromosomes 1q24.2, 3q25.31, 4q22.3, 5p12, 5q33.1, 17p12, 17q11.2, and 17q22 for NVMSD:ALL and 4p15.2 and 21q22.2 for NVMSD:C. While no compelling evidence of association was obtained under those peaks, we identified several potential genes of interest using IPA. Conclusion: Several linkage peaks were identified based on two motor speech phenotypes. In the absence of evidence of association under these peaks, we suggest genes for further investigation based on their biological functions. Given that autism spectrum disorders are complex with a wide range of behaviors and a large number of underlying genes, these speech phenotypes may belong to a group of several that should be considered when developing narrow, well-defined, phenotypes in the attempt to reduce genetic heterogeneity.

Introduction

Over the past 10 years there has been compelling evidence supporting a genetic basis for autism using a combination of behavioral family studies and genetic linkage and association studies. However, these studies have produced results that are often inconsistent and sometimes contradictory [Newbury et al., 2002]. Some linkage studies have identified peaks based on the presence or absence of autism or autism spectrum disorders (ASD), while other studies have concentrated on more specific phenotypic and clinical characterizations such as onset age of first words, family language history, sex of proband, obsessive compulsive and ritualistic behaviors, and social skills ([Alarcon et al., 2002; Auranen et al., 2003; Bradford et al., 2001; Buxbaum et al., 2004; Liu et al., 2008; Shao et al., 2003]; See [Abrahams and Geschwind, 2012] for a current linkage review).

Of particular interest for several research groups has been the attempt to define and then replicate significant linkage signals using language-based phenotypes in ASD probands with the objective of finding genes that are associated with a specific language-related phenotype. An area on chromosome 7 (q34–36) has been linked to both autism and expressive language impairments. A gene for a contactin associated protein, CNTNAP2, that is down regulated by FOXP2 and is known to influence early brain development in humans, has been associated with both ASD and language [Alarcon et al., 2008; Arking et al., 2008; Vernes et al., 2008]. While chromosome 7q continues to be an area of intense interest for both autism and language, other linkage signals have been reported that are also based on language phenotypes in the ASD population. Alarcón et al., 2005 reported linkage on chromosomes 3q and 17q using onset of first words and phrases as the behavioral phenotype while linkage on chromosome 13q21 was reported by Bradford et al., 2001 for ASD probands and family members with a history of language-related problems. Bartlett et al., 2004 identified linkage in the same region for a

sample of families with a history of Specific Language Impairment (SLI) without ASD. SLI is a failure to develop language normally without explanatory factors such as low IQ, gross neurological impairment, or inadequate environment. They suggest that although SLI and ASD are distinctly different disorders, both are genetically complex and may share specific susceptibility genes or variants of genes. Spence et al., 2006 stratified expressive language characteristics into word and phrase speech delay in ASD probands and family members in an attempt to better define the language endophenotype and reduce phenotypic heterogeneity. They found evidence for link- age in several already identified locations supporting the idea that more discretely defined characteristics of ASD, specifically language endophenotypes, may improve localization of linkage signals and strengthen existing findings.

Speech and language in ASD

Speech and language impairments constitute a broadly defined area. In their mildest forms they may be characterized by a minor phonological or speech impairment that can affect speech production and possibly reading ability. On the more severe end of the language and speech continuum, a person might be unable to comprehend or process spoken language and/or be non-verbal or unintelligible. This vast scope of speech and language disabilities seen in the ASD population has been documented in detail [Rapin and Dunn, 2003; Tager-Flusberg et al., 2005]. While some research supports the notion that there may be multiple relations among the language problems seen in SLI and autism, others feel that there is not enough evidence to support a genetic link [Lindgren et al., 2009] and that ASD and SLI are distinctly different disorders that do not share the same genes.

Previous reports indicate that approximately 50% of all children with autism never acquire functional language by middle childhood [Bailey et al., 1996] while more current

estimates place this value as closer to 20% [Lord et al., 2004]. Yet little is known about why some individuals, despite years of intervention, never develop language while many others develop enough spoken language to communicate at least minimally. Often the underlying cause is not clear and may be presumed to be a social/interaction issue. But what if language processing problems make incoming verbal information difficult or impossible to understand and severely limit verbal output? Conversely, what if problems with speech output make speech very effortful, resulting in vocalizations that include only vowels sounds or verbalizations that are unintelligible to those around them, as in the case of childhood apraxia of speech (CAS)?

CAS is a motor speech disorder that involves poor motor planning and results in speech output with compromised intelligibility ranging from its most severe form of expressive language production, which is characterized by very limited consonant production, to full phrase production with multiple omissions, substitutions, distortions, and reversals of speech sounds. While good epidemiologic data on the prevalence of CAS is lacking, population estimates derived from referral data suggest that approximately one to two children per 1,000 are affected with CAS [Shriberg et al., 1997]. There has been limited study of CAS in terms of its genetic origins, however, in their family and genetic studies of speech sound disorders, Lewis et al., 2004 looked at a small sample of children with a diagnosis of CAS and reported that 59% had at least one parent with some type of speech sound disorder. Moreover, in 86% of the families, at least one nuclear family member reported either a speech sound disorder or a language disorder. In a recent related study of speech sound disorders Lewis et al., 2007 report that 36 of 147 (24%) of parents of children with speech sound disorders also report similar problems as children.

Very little has been reported about individuals with autism whose vocalizations are

effortful, unintelligible, or non-existent. One of the few studies [Gernsbacher et al., 2008] was a retrospective study of children's oral-motor skills that compared toddlers with autism to matched controls. Using videotapes and a detailed questionnaire, they determined that the quality of oral motor skills during the early years was associated with the level of speech intelligibility of the individuals with autism in later years. Minimally verbal older children had poorer oral motor skills as toddlers.

In the current Autism Genetic Resource Exchange (AGRE) dataset, approximately 16% of the individuals who were evaluated with the Autism Diagnostic Interview- Revised (ADI-R) [Lord et al., 1994; Rutter et al., 2003] are non-verbal or minimally verbal at the time of their evaluations. Another 16% of the individuals in the dataset have speech that is unintelligible to most people.

Based on our review of the speech and language characteristics of the subjects in the AGRE database, we suggest that there is a subset of individuals with and without ASD who exhibit an expressive language problem that ranges from being non-verbal to having expressive language that is unintelligible to others and may actually be described as a severe motor speech disorder such as verbal apraxia. As these speech and language behaviors are seen in only a subset of individuals with autism but also seen in individuals who do not meet ASD criteria, we investigated linkage and association for this behavior as part of a broader phenotype.

Methods

Participants

Subjects in this study were obtained from families who are part of the AGRE database. Unlike many of the other studies that use the AGRE data for linkage analysis and require at least two affected siblings with an ASD diagnosis, we targeted AGRE families who had at least two

individuals who were either non-verbal, minimally verbal, or who had speech that was unintelligible to others, regardless of ASD diagnosis.

Responses to specific questions from the ADI-R for all families with available Affymetrix 5.0 data were used (N= 723). Motor-speech phenotypes were then developed to explore linkage and association based on the hypothesis that a subset of individuals on the autism spectrum with little or no expressive language may be part of a distinct phenotype common to autism but also potentially common to other speech and language disorders.

Motor speech phenotypes

Responses by parents and caregivers to the ADI-R were used as variables to develop the phenotypes used for the current analyses. The ADI-R is a semi-structured clinical review for caregivers of children and adults who are suspected of being on the autism spectrum. The ADI-R focuses on three areas of behavior: (a) social interaction; (b) communication; and (c) interests and behaviors that are stereotyped or restricted and repetitive. Variables from the Communication Scale were used to develop the current motor-speech phenotypes and are available in Supplemental Table A1.1; Appendix 1.

Family members who received the ADI-R (irrespective of their diagnosis of autism, ASD, or not ASD) were included in the evaluation of phenotype status. Any family member who was at least 2 years old and non-verbal or minimally verbal or was at least 4 years old, verbal, but very difficult to understand due to poor sound production, was considered affected for the NVMSD:ALL phenotype (Non-Verbal Motor Speech Disorder:All). The NVMSD:C phenotype (Non-Verbal Motor Speech Disorder:Comprehension) represented a subset of the NVMSD:ALL phenotype and included subjects who were non-verbal or unintelligible but had at least minimal language comprehension (i.e. could at least follow simple directions) as reported in the ADI-R.

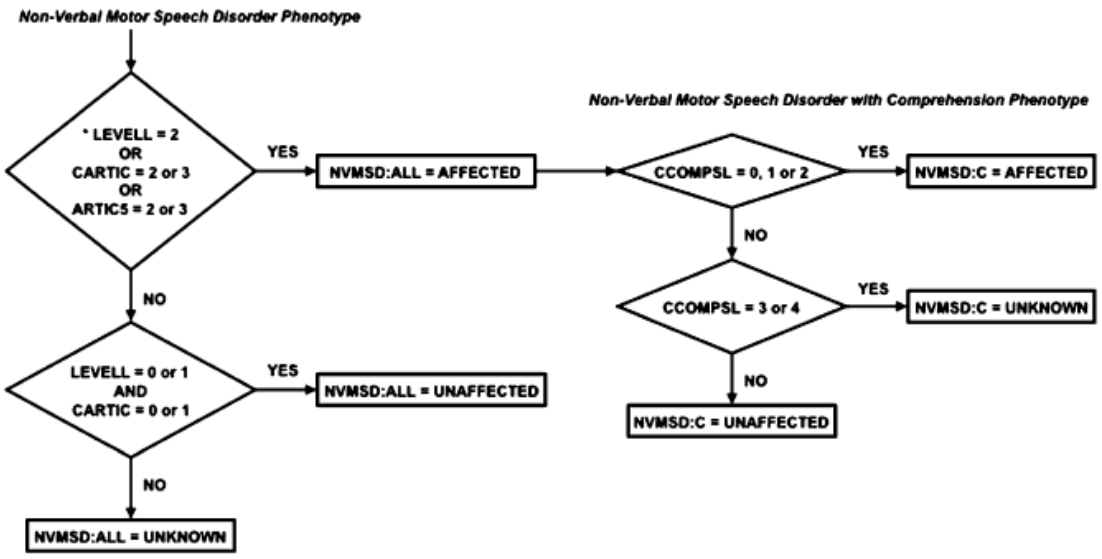


Figure 2. 1: Diagnostic Criteria for NVMSD:ALL and NVMSD:C using variables from the ADI-R

Individuals who lacked functional language (LEVELL = 2) or had poor speech intelligibility (CARTIC or ARTIC5 = 2 or 3) were AFFECTED for the NVMSD:ALL phenotype. If the language level was fair/good (LEVELL=0 or 1) and current intelligibility was fair/ good (CARTIC = 0 or 1) they were labeled UNAFFECTED, otherwise they were labeled UNKNOWN. Note the CARTIC ADI-R score took precedence over the ARTIC5 score (as long as ARTIC5 was not 2 or 3 which was established by the previous test) then its contribution was not considered for this distinction. Only those individuals that were affected according to the NVMSD:ALL phenotype were evaluated on their level of spoken language comprehension (NVMSD:C pheno- type). Individuals with some language comprehension (CCOMPSL= 0, 1 or 2) were labeled AFFECTED, those with very little to no comprehension (CCOMPSL = 3 or 4) were labeled UNKNOWN, and the remainder as UNAFFECTED. * Note that individuals had to be at least 2 years old to have a LEVELL score and had to be at least 4 years old to have a CARTIC score but could potentially be older. Also, if an individual scored 2 for the LEVELL variable, they did not receive a score for CARTIC or ARTIC5 since they did not produce enough language to be evaluated.

Figure 2.1 reflects the decision process used to assign affection status for the two motor speech phenotypes: Non-Verbal, Motor Speech Disorder (NVMSD:ALL) and Non-Verbal, Motor Speech with Comprehension (NVMSD:C).

There were 203 families (1,146 individuals) from the AGRE dataset who had both ADI-R data and Affymetrix 5.0 genotyping data and contained at least two individuals that met our criteria for the NVMSD:ALL phenotype. Among the 427 affected individuals for NVMSD: ALL

(79% male), 383 met criteria for the narrow definition of autistic disorder (AD) based on the ADI-R and came from 202 families. The mean age at ADI-R assessment was 98.83 months (s.d. 63.18 months). Of these 203 NVMSD:ALL families, 135 families (778 individuals) contained at least two family members who, irrespective of their final autism diagnosis, met criteria for the NVMSD:C phenotype. Among the 281 affected individuals (80% male), 249 met criteria for the narrow definition of AD based on the ADI-R and came from 133 families. For this phenotype the mean age at ADI-R assessment was 107.02 months (s.d. 52.40 months) (See Supplemental Table A1.2; Appendix 1). While all families were used in the Linkage Disequilibrium (LD) analyses, 35 NVMSD:ALL families were uninformative for linkage (19 in the case of NVMSD:C). This was due, in part, to affected sib-pair families in which the sibs turned out to be MZ twins.

Genotypes

Genotype data were downloaded from the AGRE site for all AGRE families with Affymetrix 5.0 data. Data on 443,106 SNPs were available for download. In preparation for linkage analysis, genotypic data were cleaned for missingness by marker (≤5% missing retained) and by individual (≤15% missing retained) (the average missingness rate was 1.5%, while the highest observed rate of missingness was 11%) and for relationship issues using RelCheck [Broman and Weber, 1998] (no families were dropped based on RelCheck identified problems). Data were then screened for Mendel errors and any SNP showing a Mendel error in a particular family was zeroed out for the entire family (the average number of Mendel errors per family was 2,605). However, there were no families excluded due to excessive Mendelian errors.

In preparation for linkage analysis, markers were dropped if the minor allele frequency was <5% or if they showed any signs of departure from Hardy-Weinberg equilibrium ($p < 0.05$). A subset of these markers was selected at 0.3 cM intervals resulting in a marker map comprising

11,100 SNPs.

For LD analyses, SNPs were dropped if the minor allele frequency was <1%, or if they failed a test of Hardy-Weinberg equilibrium at the $p=10^{-10}$ level. This left a total of 263,334 SNPs in the analysis.

Statistical analyses

Analyses were conducted using the software package KELVIN that implements the PPL (posterior probability of linkage) class of models for measuring the strength of genetic evidence [Huang et al., 2006; Vieland, 1998, 2006]. Below we report the PPL, the PPLD (posterior probability of trait-marker linkage disequilibrium (LD) and linkage) [Yang et al., 2005] and the PPLD|L (posterior probability of LD given linkage) [Wratten et al., 2009b]. We report genome-wide PPLD results, and use the PPLD|L to evaluate the evidence for LD under linkage peaks only.

The PPL is parameterized in terms of a general approximating likelihood, and all parameters of the trait model are then integrated out permitting the use of Bayes' theorem to compute the posterior probability of the hypothesis of interest. Hardy-Weinberg equilibrium has been assumed throughout. The genetic map is based on the Rutgers Combined Linkage-Physical Map, http://compgen. rutgers.edu; [Matise et al., 2007] release 10/09/06. Because KELVIN is at present Elston-Stewart based [Elston and Stewart, 1971] , the (multipoint) linkage analyses utilized LOD scores computed in Merlin [Abecasis et al., 2002; Lander and Green, 1987] as input to PPL calculations [Vieland, 1998] using Merlin's SNP clustering (with $r^2 \geq 0.2$) to further reduce potential inflation due to residual LD in the marker map. (We experimented with marker effects by varying the density of the map, the particular markers included in the maps, and the r^2 threshold for clustering SNPs, and the results were virtually identical in all cases; results not shown.)

All analyses shown here utilize a simple dichotomous trait model, with parameters α (the standard admixture parameter of Smith, 1963 representing the proportion of 'linked' pedigrees), p (the disease allele frequency), and the penetrance vector fi (representing the probability that an individual with genotype i develops disease, for i = 1..3). All trait parameters are integrated out of the final statistic; while the gene frequency is integrated over its full range, an ordering constraint is imposed on the penetrances such that f1 ≥ f2 ≥ f3. This model provides a robust approximation for mapping complex traits in terms of the marginal model at each locus, and because the parameters are integrated over, no specific assumptions regarding their values are required. Uniform prior distributions are used for all trait parameters (with adjustment for the ordering constraint). This model implicitly allows for dominant, recessive, and additive models, along with an explicit allowance for heterogeneity. In secondary analyses, we additionally allowed for imprinting or other parent-of-origin effects by allowing the penetrances to depend on the sex of the transmitting parent.

The PPL is on the probability scale. For instance, a PPL of 40% means that there is a 40% probability of a trait gene at the given location based on these data. For biological reasons, the prior probability of linkage at each location is set to 2% [Elston and Lange, 1975] so that PPLs >2% indicate (some degree) of evidence in favor of the location as the site of a trait gene, while PPLs <2% represent evidence against the location. The prior probability of LD|L is also set to 2%, so that the prior probability of LD and L is 0.04%.

The PPL and PPLD are measures of statistical evidence, not decision making procedures; therefore there are no "significance levels" associated with them and they are not interpreted in terms of associated error probabilities [Vieland, 1998; Royall, 1997; Taper and Lele, 2004]. Similarly, no multiple testing corrections are applied to the PPL or the PPLD, just as one would

not “correct” a measure of the temperature made in one location for readings taken at different locations [Vieland, 1998]. Nevertheless, it may assist readers to have some sense of scale relative to more familiar frequentist test statistics. In a simulation of 10,000 replicates of 200 affected sib pairs per replicate under the null hypothesis (no trait gene at the location being tested) allowing for the observed pattern of missing data, PPLs of 5%, 15%, 25%, 50%, and 80% were associated with Type 1 error probabilities of 0.031, 0.0018, 0.0001, 0.00005, and 0.00001, respectively.

The “null” behavior of the PPLD is moot given the results of the analysis of the experimental data; however, we note that in these same 10,000 replicates no PPLD >5% was observed. Given the sample size, we did not expect to detect LD at unlinked locations in this small set of families assuming low genotypic relative risks (RR’s). However, RR’s under linkage peaks might be expected to be considerably higher in which case power to detect LD under a linkage peak could actually be quite good. But power is entirely a function of the underlying generating model, which remains unknown. For example, fixing the RR at 2.5 and assuming D’=0.7 between the trait allele and marker allele, we simulated data under two different models: (a) we assumed locus homogeneity and dominant inheritance; (b) we assumed that only 20% of families carried the associated disease variant and that the mode of inheritance was recessive. In the first case, 96% of replicates showed PPLD ≥20%, 88% showed PPLD ≥50%, and 78% showed PPLD ≥80%. Thus for a model like this, “power” is excellent in this sample and failure to find LD under the linkage peaks is an interesting finding (assuming relatively good marker coverage). In the second case, however, only 2% of replicates showed PPLDs ≥20%, and <1% of replicates showed PPLDs ≥50%. Hence if this latter model is closer to the truth, our failure to detect LD under the linkage peaks may simply reflect the fact that the sample is still quite small.

Identifying susceptibility genes

Following the linkage and LD analyses, we used the Ingenuity Pathway Analysis—IPA software (Ingenuity® Systems, www.ingenuity.com) to identify potential autism susceptibility genes that might fall within our linkage regions. In order to characterize the peaks in our linkage analyses, we used three definitions of peak endpoints. The genome-wide PPL values were ranked (based on calculations done every 1 cM) in ascending order, and the highest 1%, 2.5%, and 5% PPL scores were used to define the narrow, intermediate, and broad regions, respectively.

The narrow regions consisted of PPL values greater than 20%, the intermediate regions were greater than 15%, and the broad regions were greater than 5%. The genes within these regions were identified using the UCSC Genome Browser (NCBI Build 36.1,[Kent et al., 2002]) and were analyzed using the Core Analysis in IPA. The Core Analysis identified the biological functions and/or diseases that were most significant to each linkage analysis. A right-tailed Fisher's exact test was used to calculate a p-value determining the probability that each biological function and/or disease assigned to that linkage analysis was due to chance alone. We selected genes with functions related to Nervous System Development and Function, Neurological Disorders, Genetic Disorders, and Psychological Disorders from the list of functions with significant p-values as possible candidate genes for our phenotypes.

As a control experiment to assess the uniqueness of our significant findings, we conducted IPA analyses on randomly selected sets of 645 genes, to model the number of genes obtained in our intermediate-peak definition analysis of NVMSD:ALL. We first identified regions of the genome centered about 2% PPL (evidence neither for nor against linkage) under the intermediate NVMSD:ALL scan and randomly selected 645 genes from these regions (Control:

Gene Number—C:GN). However, as a total of only 3,549 genes were present in the areas with approximately 2% PPL values, this frequently led to partially overlapping sets of genes. In order to create more independent samples, a second set of control analyses were also conducted by selecting 645 genes at random from the entire genome (Control: Gene Number Genome—C:GNG). We conducted core analyses on 10 C:GN and 10 C:GNG control datasets and compared the results to the gene set defined by our intermediate linkage analysis results of NVMSD:ALL.

Results

Linkage and association

Figure 2.2 shows genome-wide PPL results for the NVMSD: ALL phenotype. As can be

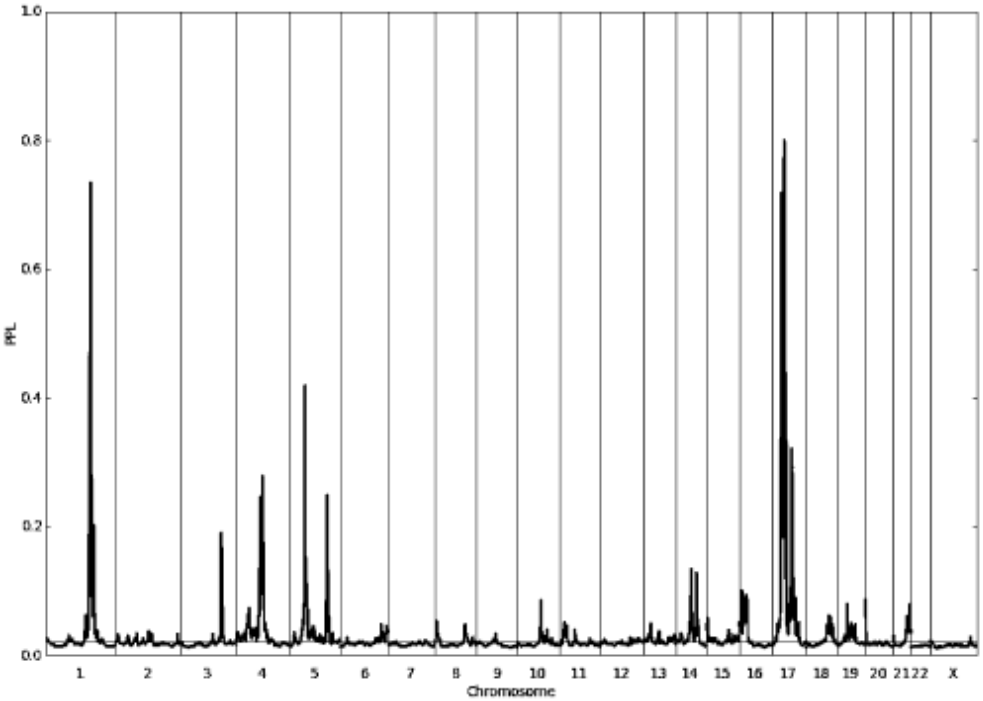


Figure 2. 2: Genome-wide PPL results for the NVMSD:ALL phenotype.

The PPL is on the probability scale. Values >2% represent evidence in favor of linkage, while values <2% represent evidence against linkage.

seen, while most of the genome shows evidence against linkage (PPL <2%) or very close to baseline (2%), there are several salient peaks. Table 2.1 shows all PPL peaks >15% for the NVMSD:ALL phenotype. On chromosome 17 there appear to be multiple peaks (Figure 2.3).

Table 2. 1: Phenotype NVMSD:ALL linkage peaks with PPL >15%

Phenotype NVMSD:ALL (nonverbal motor speech)			
Chr	cM	PPL	Band name
chr1	180	0.74	q24.2
chr3	166	0.19	q25.31
chr4	109	0.28	q22.3
chr5	66	0.42	p12
chr5	156	0.25	q33.1
chr17	36	0.51	p12
chr17	39	0.72	p12
chr17	48	0.80	p12
chr17	55	0.44	q11.2
chr17	80	0.32	q22

Figure 2.4 shows results by individual chromosomes for the NVMSD:C phenotype as well *as the NVMSD:ALL phenotype.* Because the families that are multiplex for NVMSD:C are a subset of those that are multiplex for the NVMSD:ALL phenotype, and because the two phenotypes themselves overlap by design, we expect correlation in the genome scans for the two phenotypes in these families. Moreover, because the NVMSD:C sample is smaller, we would expect to see smaller linkage signals in this group. As Figure 2.4 shows, across almost all of the genome, this is exactly what we see: peaks in the same places as in Figure 2.2, but lower. One notable exception to this is on 4p15.2, appeared to be silenced. The exceptions for NVMSD:C were on chromosome 4

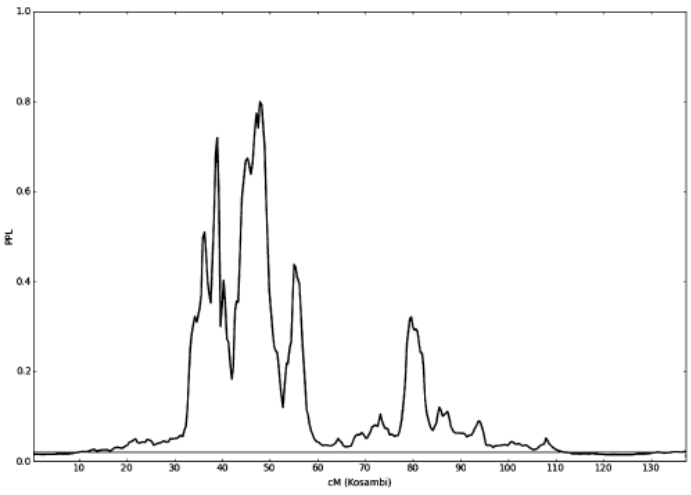


Figure 2. 3: PPL across chromosome 17 for the NVMSD:ALL phenotype.

(imprinting PPL=97% at 45 cM), and chromosome 14 (imprinting PPL=31% at 65 cM); in both of these cases penetrances appeared somewhat higher for paternal alleles but there was no indication of imprinting (full silencing) per se.

The PPLD accumulates evidence against LD as well as in favor of LD. Hence at a SNP that is not in LD with the trait, the larger the sample size the smaller the PPLD will become. For this

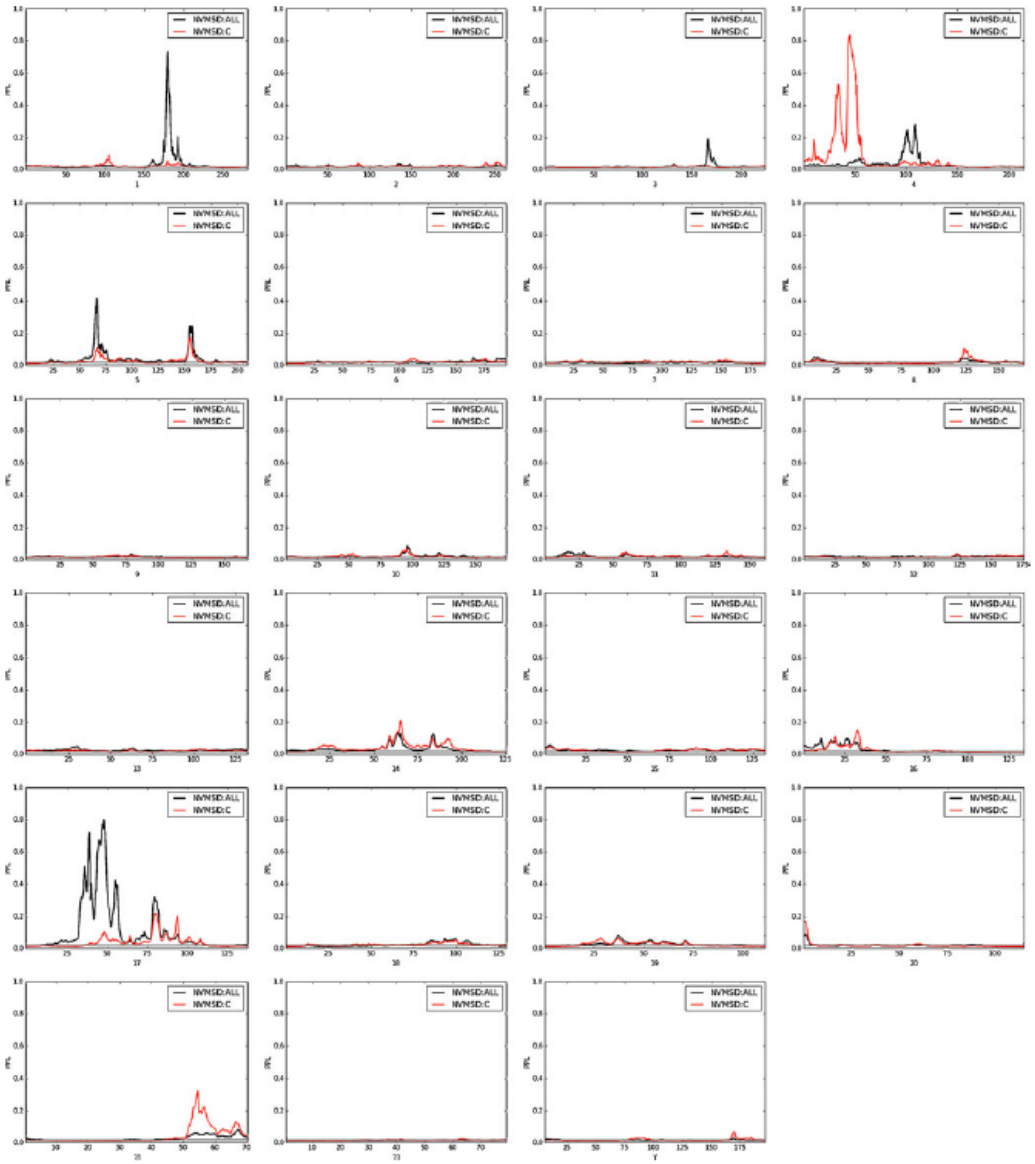


Figure 2. 4: Individual chromosome PPL results for NVMSD:ALL (in black) and NVMSD:C (in red).

reason, the smaller NVMSD:C data set yields a noisier GWAS plot around baseline. As discussed above, we were not expecting to see large PPLDs in a sample this size, and indeed, we do not see any. (Supplemental Figure A1.1; Appendix 1)

Of greater interest than the genome-wide PPLDs, however, are the PPLD|L results under the linkage peaks. However, for NVMSD:ALL we did not find any evidence of LD under the linkage peaks. Considering any genomic locations where the PPL was ≥20%, the largest PPLD|L was less than 5%. While the small sample size may make it difficult to detect LD under the peaks, assembling a very large sample of families with this phenotype is difficult. Thus, whether for underlying biological reasons or simply due to practicalities, it does not appear that fine mapping via LD analysis under these linkage peaks is likely to uncover the underlying genes. For NVMSD:C, there were 13 SNPs under the NVMSD:C-specific peak on chromo- some 4 (considering all locations where the PPL >20%) which yielded PPLD|L >5%; the maximum PPLD|L was where the large peak in the NVMSD:C analysis (PPL=84% at 45 cM) is quite far from the NVMSD:ALL peak, as shown in Figure 2.5. Also of possible interest are the NVMSD: C peaks on 21q22.2 (PPL = 32% at 55 cM) and 14q24.2 (PPL=20% at 65 cM).

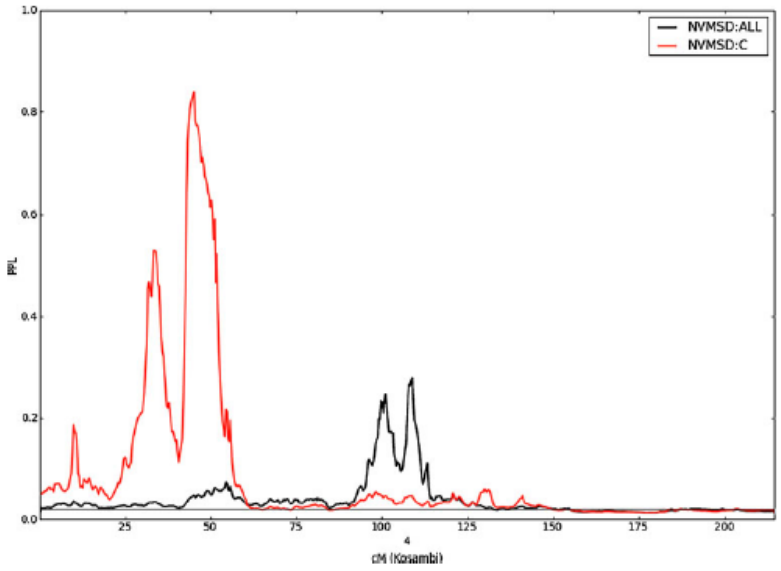


Figure 2. 5: PPL results across chromosome 4 for NVMSD: ALL (in black) and NVMSD:C (in red).

Notably, an allowance for imprinting did not

produce any new peaks or substantially change results at most loci seen under the non-imprinting analyses. In most cases, allowance for imprinting slightly depressed peaks. The exceptions for NVMSD:ALL were on chromosome 17, where the peaks rise to 86% at 48 cM, 64% at 55 cM, and 47% at 80 cM; in all three cases inheritance from the father 15%. However, these were distributed across a 20 cM region, complicating the interpretation for this (even smaller) sample. Two of these SNPs fall in genes, and one of these genes (KDNIP4, PPLD|L = 6% at rs1763197, located at physical location 20,692,185bp or approximately 36.6 cM) is of potential interest for its possible role in regulation of neuronal excitability and interactions between its protein product and presenilin.

IPA core analysis under linkage peaks

Using the output from the core analysis, we selectively identified genes with functions related to Nervous System Development and Function, Neurological Disorders, Genetic Disorders, and Psychological Disorders. For NVMSD:ALL, a total of 25 genes were selected from 261 genes input to IPA for the narrow definition (highest 1% of PPLs), 52 of 645 for the intermediate definition (2.5% highest PPLs), and 62 of 1371 for the broad definition (highest 5% of PPLs). The functions that were most represented overall were neuronal development, myelination, and axonal guidance. Likewise for NVMSD:C, 23 of 111 genes were selected for the narrow definition, 51 of 388 for the intermediate definition, and 147 of 954 for the broad definition. There was an increase of molecules involved in axonal guidance in NVMSD:C. Analyses for both phenotypes reported genes involved in motor function and various psychiatric and neurological disorders including Schizophrenia, Bipolar Disorder, and Alzheimer's disease (Supplemental Table A1.3a, b; Appendix 1).

IPA core analysis in control regions

The IPA core analysis of the C:GN (control based on gene number) datasets resulted in a different distribution of relevant significant functions than did the analysis of the linkage data. In contrast to the NVMSD:ALL intermediate analysis, each C:GN analysis resulted in several relevant functions with at least 20 contributing genes (Supplemental Table A1.4b; Appendix 1). Likewise, the number of multi-function control candidate genes is different in the C:GN analyses than in our linkage analyses. In our linkage analysis, there was one gene that contributed to 21 functions and the rest of the genes contributed 10 functions or less (Supplemental Table A1.4a; Appendix 1). In the C:GN analyses, there were multiple genes that contributed to more than 15 functions and the rest contributed to 5 or less functions. The most common functions identified in the C:GN analyses were synaptic transmission, neurotransmission, and various psychological disorders (Supplemental Table A1.4b; Appendix 1).

To overcome the potential bias introduced by restricting our control analysis to the relatively small percentage of the genome with PPL values tightly centered around 2%, we repeated this analysis using genes selected from the entire genome (C:GNG). A similar overall pattern of significant functions and multi-function candidate genes to the C:GN results was seen in the C:GNG analyses (see Supplemental Table A1.4c; Appendix 1). Like the C:GN results, the most common functions identified in the C:GNG analyses were various psychological disorders and synaptic functionality. It should be noted that genes identified by our linkage analyses were not excluded from the C:GNG datasets (Supplemental Table A1.4b and c; Appendix 1).

Overall, the most common diseases seen in both control analyses were neuropsychiatric disorders, such as Huntington's Disease and Schizophrenia, and the most common functions were neurotransmission/synaptic transmission and development of neurons and neurites. The

presence of these diseases and functions in our control analyses suggest that while they may be related to our phenotypes of interest, these functions are not unique to the core analysis of our linkage study, and may be an artifact of the extensive published research in these areas. Interestingly, only eight specific functions identified by the control analyses over- lapped with those identified from the genes from our linkage analysis. Each of these functions (cell death of neuroglia and oligodendrocytes, learning by mice, plasticity of synapse, survival of cortical neurons, development of dentate gyrus, motor neurons, and peripheral nervous system) appeared only once in the control analyses. While there is some commonality in functions, it is important to note that the candidate genes described below are not implicated by functions identified in the control analyses.

Discussion

Linkage and association

We have identified several peaks that represent strong evidence of linkage using two novel and relatively narrow behavioral phenotypes for non-verbal language and motor speech problems; characteristics that are associated with autism but not exclusive to the autism spectrum. While some of the peaks overlap with previously reported linkage locations [Alarcón et al., 2005; Bartlett et al., 2005; Cantor et al., 2005; McCauley et al., 2005; Schellenberg et al., 2006; Yonan et al., 2003], others are novel. In some cases, where results overlap, different behavioral phenotypes have been reported for those peaks. This is not surprising since, by definition, an individual with autism might share behaviors and belong to several phenotypic subgroups within the spectrum as well as share behaviors with individuals who do not meet criteria for ASD. Additionally, there have been multiple studies looking at language and autism using the AGRE sample and the potential for overlapping subjects is impossible to avoid. However, the strength of the peaks and the specificity of the phenotype lend support to the

idea that there could be genes of interest under these peaks that warrant further investigation.

We included all individuals from the AGRE data set who had ADI-R diagnostic information and Affymetrix 5.0 genetic data regardless of their final ASD diagnosis. A percentage of those identified as meeting criteria for one or both phenotypes did not meet the ASD cut-off criteria. Yet at some point, they must have demonstrated behaviors compatible with a potential ASD diagnosis or they would not have received the AGRE ASD study battery in the first place. This lends support to the notion that in families with ASD probands, there may be other family members that share behavioral characteristics and genes and fall into some kind of broader autism phenotype.

Evidence of linkage based on PPL values greater than 15% for the NVMSD:ALL phenotype was identified on chromosomes 1q24.2, 3q25.31, 4q22.3, 5p12, 5q33.1, 17p12, 17q11.2, and 17q22. Linkage on 3q25 and 17p supports previous findings in the same area where loci linked to word and phrase speech delays were identified by Alarcón et al., 2005. In that study, the authors identified a region on chromosome 3 (126–170 cM with a peak at 147 cM) that overlaps one of our peaks and was identified with an onset of first words phenotype. In addition, they identified a region on chromosome 17 (13–96 cM) that coincides with one of our peaks and is suggestive of linkage for first words and phrases. Since delayed first words, and/or delayed first phrases might also apply to a number of our probands, because they are nonverbal, one could speculate that there was overlap across our samples. In fact, our finding at chromosome 1q24 (PPL=74%) was located in the same region as previously reported by our group [Bartlett et al., 2005]. In the current study, 88 AGRE families also satisfied the diagnostic criteria for the Bartlett et al. study (i.e., the affected phenotype was based on delayed speech onset in two affected siblings).

For many previous autism linkage studies a more formal diagnostic phenotype was used that ranged from a narrow definition of autistic disorder to an autism spectrum disorder that included Asperger's disorders and PDD-NOS. Since we already know that a significant number of individuals with autism can be non-verbal or have speech that is significantly unintelligible, it is not surprising that overlapping linkage peaks were observed. McCauley et al., 2005 identified linkage on 17q11 for sib pairs consisting of at least one proband with autistic disorder and one on the spectrum. Using a broader definition of autism, Yonan et al., 2003 identified linkage in the same regions of 17q and 5p as our study. Studying male-only families has been another approach [Cantor et al., 2005; Schellenberg et al., 2006] resulting in linkage peaks on chromosomes 4 and 17. When Schellenberg et al., 2006 stratified their families by male-only, they identified linkage at 4q22 and Cantor et al., 2005 identified the 17q21 region (67 cM) when doing fine mapping of the area. Similarly, our families were enriched for affected male subjects; we had 128 male-only families with the remaining 75 families having at least one female affected for our phenotype, bringing our rate of affected males to approximately 80% in our 203 families. Buxbaum et al., 2004 identified a peak on 1q24 as well and another peak suggestive of linkage on chromosome 4 for an obsessive-compulsive phenotype, thus another example of overlapping linkage based on different phenotypes but behaviors that are part of the ASD profile. In summary, even though samples varied and descriptions of phenotypes differed, the fact that most of our linkage peaks have been previously identified in ASD populations, lends support to the idea that these particular locations are a source for continued investigation.

The NVMSD:C phenotype was created to narrow and better define the motor-speech characteristics that are found in a subset of individuals with ASD as well as other individuals with speech and language impairments. It was based on the premise that probands who have some language comprehension but display minimal or unintelligible expressive language, might belong

to a phenotypic group specifically characterized by a motor speech impairment that is seen in apraxia of speech. Using these criteria we identified similar, but weaker, linkage signals to the NVMSD:ALL phenotype, which is not surprising given the smaller sample. Moreover, we hypothesize that our stronger findings with NVMSD:C on chromosomes 4p15.2, 14p15.2, and 21q22.2 might actually be better capturing those individuals who have a more well-defined motor-speech disorder like apraxia.

Notably, we did not find a linkage signal on 7q, a location that has been strongly implicated in linkage and association with both ASD and speech and language impairments [Alarcon et al., 2008; Arking et al., 2008; Vernes et al., 2008; Lai et al., 2001; Feuk et al., 2006]. This was true even allowing for imprinting, which has been suggested for FOXP2 on chromosome 7 [Feuk et al., 2006] (Supplemental Figure A1.2; Appendix 1). However, imprinting gene candidates have been reported in regions where we did see evidence for linkage with imprinting. Luedi et al., 2007 report maternal expression of TRIM16 (17p12), TIAF1 (17q11.2), HOXB2 (17q21.32), and HOXB8 (17q21.32). All of these genes match the pattern of imprinting supported by our linkage results on chromosome 17 and so they represent higher priority positional candidates; the homeobox genes are of particular interest due to their role in development patterns in the brain [Fanarraga et al., 1997; Matis et al., 2007; Grados et al., 2003].

Potential genes of interest using IPA

NVMSD:ALL phenotype

When we used IPA to identify potential autism susceptibility genes that might fall within our linkage regions, we identified several genes associated with translation and transcription factors, brain development, nervous system development, and multiple psychiatric disorders.

We also took into consideration the overlap of functions between the control and linkage analyses. The candidate genes de- scribed below meet our criteria for the IPA core analysis of the linkage regions and have been filtered for overlap with the control analysis functions.

Our linkage region on 4q22.3 contains NKX6-1, which encodes a transcription factor that binds to AT-rich regions in the promoters of its target genes. NKX6-1 plays an important role in differentiation of motor neurons and the regulation of muscle nerve formation [Lee and Pfaff, 2001; Bohl et al., 2008; De Marco Garcia and Jessel, 2008]. Six gene targets of NKX6-1 (ATOX1, GPX2, HIF1A, HMOX2, IGFBP4, and PHB) fall within the broad linkage regions for the NVMSD:ALL analysis and one gene (ANAPC4) falls within the broad linkage regions for the NVMSD:C analysis.

The linkage peak on 5p (66 cM) contains GHR, which encodes a growth hormone receptor shown to be involved in brain development and neuronal differentiation [Harvey, 2001; Harvey et al., 2002; Ransome et al., 2004; Baudet et al., 2007]. A second linkage peak on 5q (154– 156 cM) contains two candidate genes: DPYSL3, which is involved in neurite outgrowth and guidance and shows a decreased expression in individuals with Down syndrome [Weitzdoerfer et al., 2001] and HTR4, which is a serotonin receptor that has been associated with Schizophrenia, Attention-Deficit Hyperactivity Disorder, and Bipolar Disorder [Hirata et al., 2010; Suzuki et al., 2003; Hayden and Nurnberger, 2006; Elia et al., 2009]. The regions on 17p and 17q contain NCOR1 and NOS2, which are involved in NOTCH signaling and the NOS pathway, respectively. NOS2, in particular, has been implicated in various neurological disorders such as Alzheimer's disease, and Amyotrophic Lateral Sclerosis [Colton et al., 2009; Chen et al., 2010]. The peak at 17p12 (PPL = 77%) also contains PMP22, which encodes a protein that comprises 2–5% of peripheral nervous system myelin. Most recently, Pinto et al., 2010 identified a rare maternally inherited copy number variation (CNV) that contains PMP22 in an

individual with ASD, however this CNV was not experimentally validated.

NVMSD:C phenotype

Our linkage analysis of NVMSD:C resulted in two novel regions, which we further investigated with the core analysis in IPA. Overall, there was a notable increase in genes involved in Down syndrome, which is primarily due to the inclusion of 21q22.2. IPA identified two genes on chromosome 4 involved in axonal guidance: SLIT2 [Hammond et al., 2005] and CRMP1 [Yamashita et al., 2006]. SLIT2 is of particular interest as both the SLIT1 and SLIT2 proteins have been identified as selective inhibitors and repellents for dorsally projecting cranial motor axons [Colton et al., 2009]. In addition to these axonal guidance genes, the core analysis also identified STIM2 on chromosome 4 (45 cM, PPL= 88%), which regulates calcium entry into neurons [Berna-Erro et al., 2009]. Like the NVMSD:ALL analysis, molecules involved in psychiatric disorders such as Schizophrenia, Panic Disorder, and Social Impairment were also identified as genes involved in the NOTCH Signaling and NOS pathways. Both analyses identified molecules that are involved in motor function, however IPA analysis of NVMSD:C did not produce significant findings within regions of our strongest linkage signals.

IPA control analyses

Our control analyses (C:GN and C:GNG) served as a test of the reliability of our IPA analysis which we will use to guide our future investigations of autism genes. As seen in Supplemental Table A1.4b and c; Appendix 1, the numbers of functions and candidate genes obtained for the C:GN and C:GNG analyses were comparable to those obtained in our linkage regions. The functions identified by IPA are not presented in a hierarchical order, which leads to the identification of several similar functions that are, in fact, subsets of one overall function. This is seen commonly in our control analyses (as demonstrated in analysis 9, Supplemental

Table A1.4b; Appendix 1) and also occurs in our linkage analyses. While there was some overlap in general function categories, there were only eight specific functions identified in the control analyses that were also identified in the linkage analysis. Despite this overlap in functions, viable candidate genes were identified from the core analysis of our linkage regions after filtering for these common functions. This filtering of the linkage analysis helps to ensure that the functions and genes/molecules identified in the core analysis of the linkage regions are unique to that analysis. Overall, the IPA core analysis used in this study functions primarily as a data reduction tool and was effective in identifying genes in our linkage analysis that require further investigation.

One limitation of this study concerns the variables available to us from the AGRE data set to define a motor speech disorder. To make a clinical diagnosis of such a disorder, a speech language pathologist would use converging information including an extensive language history, a complete oral motor exam, and a comprehensive speech and language assessment that would include specific information about the phonological abilities of each proband. Yet, our strong linkage findings suggest that at this stage of investigation we have defined this disorder well enough to continuing pursuing the genes, gene interactions, and gene variants under the peaks.

Conclusion

We have identified several unique loci, based on two specific motor speech phenotypes, that are present in, but not exclusive to, a subset of individuals within families with autism spectrum disorders. In addition, we have identified several loci that had been previously isolated on the basis of somewhat different diagnostic criteria. Family members who are non-verbal or verbal but have speech that is unintelligible may or may not meet criteria for ASD but may share genes and behaviors that are also seen in other speech and language disorders. Although we

found no compelling evidence of association under our linkage peaks, we were able to suggest genes for further investigation based on their biological functions using IPA. It is well recognized that autism spectrum disorders are complex with a wide range of behaviors and, potentially, a large number of underlying genes, so that these particular sets of behaviors might fall into a broader phenotype and further emphasize the need to develop narrow well-defined phenotypes to reduce genetic heterogeneity.

Chapter 3: Follow-up Linkage and Association Analyses of a Nonverbal Motor Speech Phenotype Identified in the AGRE Data Set

Abby E. Hare, Marco A. Azaro, Raymond A. Zimmerman, Judy F. Flax, Linda M. Brzustowicz

Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ

John Burian, Veronica J. Vieland

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH

Keywords: Autism, linkage analysis, association analysis, PPL/cPPLD, motor speech disorder

Abstract:

Background: Using behavioral and genetic information from the Autism Genetics Resource Exchange data set we investigated linkage and association for individuals with and without Autism Spectrum Disorders who exhibit expressive language behaviors consistent with a motor speech disorder. Speech and language variables from the Autism Diagnostic Interview-Revised were used to develop a motor speech phenotype associated with non-verbal or unintelligible verbal behaviors (NVMSD:ALL). We previously identified several linkage peaks using the PPL framework to assess the strength of evidence for or against trait-marker linkage across the genome using Affymetrix 5.0 genotype data. As evidence for linkage disequilibrium was limited, Ingenuity Pathway Analysis was then utilized to identify candidate genes for further investigation.

Methods: 1) The SNPstream assay was used to genotype 450 SNPs from our linkage regions in 87 families that met criteria for NVMSD:ALL but did not have Affymetrix 5.0 data available at the time of ascertainment. The initial PPL analysis was sequentially updated to include evidence for or against linkage in the extended sample. 2) Tag SNPs were selected from our genes of interest and were genotyped using an Oligonucleotide Ligation Assay. Genotype information was analyzed for association using an extension of the PPL that detects linkage disequilibrium (cPPLD). Prior to analysis, all genotype information was checked for missingness, Mendelian inconsistencies, unlikely double recombination events, and departures from Hardy-Weinberg Equilibrium.

Results: The sequential update of our linkage analysis increased evidence for linkage in the following regions: 1q24.2, 3q25.31, 4q22.3, 5p12, and 17q22. Evidence for linkage decreased in

the following regions: 5q33.1, 17p12, and 17q11.2. Overall, evidence for linkage disequilibrium was limited in our candidate genes of interest, with the highest signals in TRPV2 (cPPLD = 7%).

Conclusions: While evidence for linkage disequilibrium remained limited, the additional 87 families added to the power of our PPL analysis. These families increased evidence for linkage in several regions, while the evidence decreased in other areas. The strong replication of our findings indicates that NVMSD:ALL is a stable phenotype for genetic analyses and is an effective endophenotype of autism for the reduction of genetic heterogeneity.

Background:

Autism is a complex neurodevelopmental disorder characterized by communication deficits, social impairment, and the presence of restrictive and repetitive behaviors. While there has been compelling evidence supporting a genetic basis for autism, behavioral family studies, genetic linkage analyses, and association studies have produced inconsistent and often conflicting results [Geschwind, 2011; Spence et al., 2006]. Due to the heterogeneity of autism spectrum disorders (ASD), linkage studies, based on the presence or absence of ASD, are difficult to replicate. Researchers have employed several strategies for addressing this issue. One method is to concentrate on more specific phenotypic and clinical characteristics of ASD by separating the study sample into more clinically similar groups. Another method is to derive phenotypes that may be etiologically more related to ASD from phenotypic measures used in ASD diagnosis, such as the Autism Diagnostic Interview-Revised (ADI-R). The study presented in this paper utilizes a hybrid of both methods by analyzing a subset of the Autism Genetic Resource Exchange (AGRE) database that meets criteria for a nonverbal motor speech phenotype derived from several variables in the ADI-R. Both aforementioned strategies have had success in identifying linkage to language phenotypes developed from quantitative data available in autism cohorts [Alarcón et al., 2002; Auranen et al., 2003; Bradford et al., 2001; Buxbaum et al., 2004; Liu et al., 2008; Shao et al., 2002; Abrahams and Geschwind, 2008].

The first method has allowed investigators to reduce the phenotypic heterogeneity within the AGRE database with the intent of genetically differentiating subgroups of ASD based on a language delay phenotype. This method was particularly successful in separating AGRE families based on the presence or absence of phrase speech delay (PSD) in autism probands, which is defined as the development of phrase speech after 36 months [Buxbaum et al., 2001]. This study identified linkage to ASD on chromosome 2q24-32 using 49 families that contained

two individuals with ASD who also met the criteria for PSD [Buxbaum et al., 2001]. This finding was later replicated using 45 sib-pair families affected for PSD from the Collaborative Autism Team dataset [Shao et al., 2002]. Linkage to ASD has also been identified on chromosomes 7q22-32 and 13q21-22 using a subset of sib-pair families (n = 50) from the Collaborative Linkage Study of Autism cohort where both probands met the criteria for PSD [Bradford et al., 2001]. While linkage to these regions has not been replicated using PSD, the finding on Chromosome 13q has also been linked to Specific Language Impairment (SLI) [Bartlett et al., 2002, 2004; Simmons et al., 2010].

Investigators have also been successful using the second method of reducing phenotypic heterogeneity by developing language phenotypes using variables from the ADI-R. Linkage to a region on Chromosome 7q36 was identified in linkage scans for the 'onset of age of first word' and the 'onset of phrase speech in months' variables in the ADI-R using the AGRE dataset [Alarcón et al., 2005, 2002]. This region on Chromosome 7q contains a gene encoding a contactin associated protein, *CNTNAP2*, which has been implicated in language impairment in non-ASD individuals [Alarcon et al., 2008; Arking et al., 2008; Vernes et al., 2008] and is down-regulated by FOXP2, a transcription factor that has been implicated in speech and language disorders [Vernes et al., 2008]. For a current review of additional studies of *CNTNAP2,* see Peñagarikano and Geschwind, 2012. The identification of linkage to *CNTNAP2* in the AGRE dataset indicates the effectiveness of the use of language phenotypes to identify genetic loci that play a role in the language impairment aspect of ASDs. It is important to note that while these studies used multiplex autism families, the affection status utilized was the language phenotype, not autism diagnosis.

Although many individuals on the autism spectrum develop language skills, it is estimated that 20-50% of individuals with autism do not acquire functional language [Lord et al., 2004; Bailey et al., 1996]. While some individuals with ASD may not develop adequate functional expressive language due to an underlying social impairment, others may have language processing impairments or motor speech deficiencies making speech extremely effortful or unintelligible, similar in character to deficits seen in Childhood Apraxia of Speech (CAS). In the AGRE database, 11.6% of individuals with ADI-R data are considered to be non-verbal or unintelligible to others. There is limited information about the genetics of CAS, but Lewis et al., 2004, reported that 86% of children with CAS had at least one nuclear family member with a speech sound disorder or a language disorder. Additionally, 59% of children with CAS had at least one parent with a speech sound disorder, suggesting a genetic component.

While there have been few studies involving the presence of CAS-like behaviors in ASD populations, there has been a clinical focus on the effectiveness of early motor skill deficiencies as a predictor of language acquisition and intelligible communication [Maski et al., 2011] and most recently a study on articulatory features and function being a possible early marker for ASD [Sullivan et al., 2013]. A retrospective study that compared toddlers with autism to matched controls showed that minimally verbal older children with ASD had poorer oral motor skills as toddlers [Gernsbacher et al., 2008]. Similarly, MacNeil and Mostofsky, 2012, have shown that impairments in perceptual-motor action consistent with dyspraxia appears to be specific to individuals with ASD when compared to individuals with Attention-Deficit Hyperactivity Disorder. With behavioral information supporting a potential link between CAS and ASD, it is not unreasonable to use CAS symptoms to define motor speech phenotypes of ASD with the objective of finding genes that are correlated with these subphenotypes.

Shriberg et al., 2011, investigated the presence of CAS in a sample of 46 verbal children aged 4-7 years diagnosed with an ASD. Individuals with ASD did have a higher prevalence of speech errors when compared to population estimates; however, they did not display the core features of apraxia of speech. These individuals did display increased vocal repetition and inappropriate pitch, stress and volume which are typical in ASD. This study does not support the concomitance of the classical definitions of CAS and ASD; however, the participants selected for the study were verbal individuals with ASD with relatively fluent language production. The sample did not include those individuals with autism whose lack of expressive language was characterized by effortful attempts to verbalize, resulting in unintelligible vocalizations comprised of either: (a) only vowel sounds with or without minimal consonant production or (b) those with severely unintelligible expressive language.

The study presented in this paper is a follow-up to the linkage and association analyses reported in Flax et. al., 2010, where a nonverbal motor speech phenotype was defined in a sample of families from the AGRE dataset. The motor speech phenotype NVMSD:ALL was developed using responses to three specific questions on the ADI-R assessing the overall level of expressive language and articulation [Lord et al., 1994; Rutter et al., 2003]. All families were required to have at least two individuals, who were either non-verbal, minimally verbal, or who had speech that was unintelligible to others, regardless of their ASD diagnosis. The original sample identified in AGRE included 203 families (1147 individuals, 427 affected) and contained only individuals with Affymetrix Genome-wide Human SNP Array 5.0 genotype data available at the time of the initial study. A subset of 11,100 SNPs was analyzed for linkage using the Posterior Probability of Linkage (PPL). Linkage peaks were identified on chromosomes 1q24.2, 3q25.31, 4q22.3, 5p12, 5q33.1, 17p12, 17q11.2, and 17q22 for NVMSD:ALL. Genome-wide association analysis of 263,334 SNPs using the Posterior Probability of Linkage Disequilibrium

(PPLD) did not yield strong evidence for association. However, the absence of evidence for linkage disequilibrium (LD) could have been the result of a modest sample size or from insufficient SNP coverage of common haplotypes in our regions of interest. We hypothesized that by sequentially updating the original sample with new subjects, linkage signals would increase due to increased genetic evidence. We also hypothesized that analysis of a higher density marker map under the linkage peaks could result in stronger evidence for association. This paper presents the results of both of the above hypotheses, sequentially updating our original linkage regions and the subsequent fine mapping and association analysis of genes selected from our initial regions of linkage.

Methods:

Participants

Subjects were selected from families who are part of the AGRE database and met the original study criteria for NVMSD:ALL but were not included in the original study because no Affymetrix 5.0 genotype data were available at the time of this study (87 families, 421 individuals). To be included in this study, each family was required to have at least two individuals who met criteria for NVMSD:ALL. The NVMSD:ALL phenotype includes individuals who are non-verbal, minimally verbal, or have speech that was unintelligible to others, regardless of their ASD diagnosis. For more details on the NVMSD:ALL criteria see the Supplemental Methods (Appendix 2) and Flax et al., 2010. A total of 216 individuals were affected for NVMSD:ALL (82% male) and 234 individuals meet criteria for ASD. The overall sample includes the 203 families described in Flax et al., 2010 yielding a total of 290 families (1567 individuals, 643 affected for NVMSD:ALL) in the final linkage analysis. Fine mapping of candidate genes for association analysis was conducted on the original 203 NVMSD:ALL families.

Genotypes

New genotype data was generated for family members who did not have Affymetrix 5.0 data available. SNPs were selected from linkage regions that were identified in the original genome scan and genotyped by the Rutgers University Cell and DNA Repository (RUCDR) Infinite Biologics at Rutgers University, Piscataway, NJ using SNPstream single base pair extension genotyping technology [Bell et al., 2002]. Initially, 450 SNPs were selected from these linkage regions. However, 170 of these SNPs failed to have genotyping probes designed by the dedicated Autoprimer application (Beckman Coulter, http://www.autoprimer.com), primarily due to extensive masking of repeats in their flanking regions. To select replacements, candidate SNPs in high LD (r^2 >= 0.8) with an original SNP were grouped into bins and a representative SNP was chosen from each based on filtering criteria recommended by Autoprimer's template preparation protocol. Specifically, the 301 bp sequence centered on each SNP was extracted from the current build of the human genome (hg19, Feb. 2009) and was submitted to a local RepeatMasker executable [Smit et al., 2010] that employed the WU-BLAST alignment engine [Gish, 2003] and default parameter settings. The SNP from each bin that had a minimal degree of proximal masking and whose flanking GC percentage was closest to 50% was chosen. In all, 173 replacement SNPs were found, bringing the total number genotyped by SNPstream to 453 SNPs.

In preparation for linkage analysis, the genotype data were cleaned for missingness by marker and by individual. A total of 28 SNPs were removed due to total assay failure and an additional 17 SNPs were removed for manifesting more than 5% missingness. The average missingness for partial assay failure was 11%. A total of 11 markers that showed departures from Hardy-Weinberg Equilibrium ($p < 0.01$) were dropped from further analysis. The highest rate of missingness for an individual after removal of these SNPs was 4.53%. As the missingness

cutoff for an individual was 15%, no individuals were excluded from analysis for this reason. PedCheck [O'Connell and Weeks, 1998] was used to identify Mendelian errors and Merlin [Abecasis et al., 2002] was used to identify unlikely double recombination events yielding an overall error rate of 1.2%. While any SNP showing an error was removed for the entire family, no families were eliminated due to excessive Mendelian or recombination errors. Following this cleaning procedure, genotypes for 397 SNPs were included for linkage analysis.

Fine mapping genes of interest

Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, www.ingenuity.com) was used to categorize and prioritize candidate genes identified in our linkage regions by their function [Flax et al., 2010] (See Supplement for more details; Appendix 2). Candidate genes were further prioritized by the strength of their PPL values, resulting in a total of 11 genes selected for further analysis from the original 645 genes located within our linkage regions (Table 3.1). HapMap was used to select tag SNPs (HapMap Data Rel 24 Phase 1 & 2 – full dataset, NCBI B36 assembly, dbSNP b126) using the CEU population with a minor allele frequency (MAF) ≥ 0.05 and $r^2 < 0.80$ for each gene region including 10kb upstream and downstream flanking regions. Tag SNPs were supplemented with markers from SNPbrowser (version 4.0, Applied Biosystems) and the 1000 genomes database [The 1000 Genomes Project Consortium, 2010] for a total of 201 SNPs (Table 3.1).

Multiplex PCR design templates were prepared for each SNP by extracting 500 bp of flanking upstream and downstream sequences from the current human genome build (hg19, Feb. 2009) and substituting all known neighboring SNPs obtained from dbSNP (Database of Single Nucleotide Polymorphisms (dbSNP), Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 131)) with their respective IUB

Table 3. 1: Candidate genes selected for fine mapping.

Gene	Linkage Region	PPL (%)	# tag SNPs	IPA Functions
TRPV2	17p11.2	79	14	Motor Neuron Function Axon Outgrowth
UBB	17p11.2	79	3	Huntington Disease Model
NCOR1	17p12	79	37	Neural Stem Cell Differentiation
PMP22	17p12	77	19	Peripheral Nervous System Myelin Neuromuscular Junctions
GHR	5p12	41	62	Brain Development
CDK5R1	17p11.2	39	12	Regulates Neuronal Mitochondria
NF1/OMG	17p11.2	25-34	13	Implicated in Autism
EIF4E	4q23	18	5	Cerebellar Growth Implicated in Autism
NKX6-1	4q22.3	15	3	Transcription Factor Differentiation of Motor Neurons Regulation of Muscle Nerve Formation
LMX1A	1q23.3	15	33	Cerebellar Growth

Genes are listed by PPL evidence and their respective functions identified by IPA are listed. SNPs were selected using HapMap, SNPbrowser and 1000 genomes.

codes. Candidate primers were extracted from these flanks based on several criteria: (1) they did not overlap with neighboring SNPs; (2) they lacked simple tandem and inverted repeat elements; (3) their predicted melting temperatures fell within a 2°C range; and (4) they yielded a minimal number of secondary alignments when subjected to local BLAST [Altschul et al., 1997] analysis. These filtered candidates were used to generate multiplex PCR primer sets for panels that ranged in size from 14 to 36 SNPs. For each multiplex panel, primer compatibility was enforced by requiring that no two PCR primers shared 3'-to-3'-end homology greater than 2 bp and that no given PCR primer had greater than 6 bp of homology between its 3' terminus and the internal region of any other primer. Amplicon sizes ranged from 98 bp to 434 bp.

SNPs were genotyped using the oligonucleotide ligation detection assay described in Bruse et al. 2008, that has been optimized to work in the 384 microtitre format. The 30 µL PCR

reaction was scaled down to 20 µL to accommodate the smaller well volume of the 384 microtitre plate. The smaller reaction mixture contained 40 ng of genomic DNA, 1U AmpliTaq Gold® DNA polymerase, 10x reaction buffer (provided with AmpliTaq Gold®), 2.5 mM MgCl (Applied Biosystems), 200 µM dNTPs, and 3.33 pmol of each PCR primer, though thermocycling conditions remain unchanged. The OLA reaction volume was scaled up to 20 µL from 15 µL to allow for the use of a Perkin Elmer – Janus® Automated Workstation, again using the original thermocycling conditions [Bruse et al., 2008]. The increased OLA reaction mixture contains 2 µL of the PCR product, 3 U of Taq DNA Ligase (New England Biolabs), 10 nM of each OLA probe, and a 10X reaction buffer (provided with the Taq DNA Ligase). The bead hybridization and fluorescent labeling steps remain as described in Bruse et al., 2008.

Ten SNPs were removed from the analysis due to total assay failure. After removal of these SNPs, this method resulted in an overall fail rate of 1.1% for the remaining SNPs. As above, genotypes were cleaned using PedCheck and Merlin. Two families were removed from the dataset due to excessive error rates resulting in an overall error rate of 0.3% for the remaining dataset. Four SNPs were removed due to departures from Hardy-Weinberg Equilibrium (p–value <0.01). With an overall missingness rate of 1.5%, no individuals or markers were removed due to an excessive amount of missing data.

Statistical methods

All analyses were conducted using the software package KELVIN. Evidence for linkage was calculated using the PPL [Huang et al., 2006; Vieland, 2006, 1998]; the PPLD was used to calculate evidence for association. Genetic distances were obtained from the Rutgers genetic map (http://compgen.rutgers.edu/mapopmat; release 10/09/06; [Matise et al., 2007]).

The PPL is a Bayesian statistical measure of linkage that is designed to accumulate evidence for or against linkage for complex traits [Vieland, 1998]. The PPL is parameterized in terms of a general approximating likelihood and utilizes a dichotomous trait model with an admixture parameter (α) representing the proportion of 'linked' pedigrees, the disease allele frequency (p), and the penetrance vector (f_i). The parameters of the trait model are integrated out of the algorithm, making this method of linkage analysis essentially model-free. The likelihood also contains the recombination fraction (θ) and the standardized LD parameter (D') as two additional parameters. The PPL incorporates a prior probability of linkage of 2% [Elston and Lange, 1975], therefore, PPL values greater than 2% indicate evidence in favor of linkage, while PPL values less than 2% indicate evidence against linkage for that particular locus. As this measure can be interpreted directly as a probability, corrections for multiple testing are not required. Furthermore, the PPL is a measure of evidence without inherent significance levels. A PPL of 40%, therefore, can be directly interpreted as evidence that there is a 40% probability of a trait locus at the given marker[Vieland, 2006].

The PPL is designed to accumulate evidence over multiple datasets through sequential updating. This approach was used to update the original multipoint PPL scan of the Affymetrix 5.0 data from the AGRE dataset [Flax et al., 2010] with the SNPstream genotype data for the 87 additional families. Multipoint linkage analysis was performed on each the original family set and the additional 87 family set independently as described in Flax et al 2010. For each linkage region identified with the NVMSD:ALL phenotype, the original PPL distribution was used as the prior probability of linkage. This prior distribution was updated with evidence obtained from the Bayes ratio calculated for the additional 87 families in each linkage region. The increased power obtained by expanding the sample size to a total of 290 families provides stronger evidence for or against linkage for the combined NVMSD:ALL phenotype.

As all markers selected reside within an identified region of linkage, evidence for association was calculated using Combined Posterior Probability of Linkage Disequilibrium (cPPLD) [Huang, 2011]. This function uses the PPL values identified in the NVMSD:ALL linkage analysis as the prior probability of linkage. Additionally, a cPPLD analysis was conducted using autism diagnosis as a liability class [Huang et al., 2007; Vieland and Huang, 2003]. KELVIN treats the liability class traits (affected or unaffected) individually to allow for separate penetrance calculations for each category. For a given marker, an increase in evidence for association while using the liability class indicates an interaction between the NVMSD:ALL phenotype and autism affection status. The cPPLD was also calculated using autism diagnosis as the affection status to rule out an increase solely driven by the autism diagnosis as opposed to its interaction with the NVMSD:ALL phenotype.

Determination of Sample Power

To assess the power of the sample to detect association, KELVIN was run using genotypes simulated by SLINK [Weeks et al., 1990]. Haplotypes that included one SNP disease marker and a SNP in LD with the disease marker (D' = 0.80) were generated for the NVMSD:ALL family set. The disease gene frequency (DGF) of the disease marker was simulated at percentages of 1, 2, 3, 4, 5, 10, 15, 20 and 30. The disease allele risk was varied for both recessive and dominant models (Supplemental Table A2.1; Appendix 2). The percentage of unlinked families was also varied from 0% to 100% in increments of 10% to model the heterogeneity of the NVMSD:ALL dataset (average α = 0.62; range 0.1 to 1.0). Each set of parameters was replicated 100 times and was analyzed for association using the KELVIN framework described above. All replicates were conducted using two estimations of the prevalence of NVMSD in the population. The first prevalence, 11.4%, was calculated as the percentage of individuals in the AGRE database who met the criteria for NVMSD:ALL phenotype,

regardless of the study criteria of multiplex families for NVMSD:ALL. The second prevalence is an estimation of the NVMSD:ALL phenotype in the general population. This estimation was derived from the percentage of individuals with an autism diagnosis in AGRE who met the criteria for NVMSD:ALL (46%) and the prevalence of ASD in the general population (1%,[CDC, 2012]), resulting in a prevalence of 0.46%. This is a reasonable estimation based upon the literature, which estimates the rate of nonverbal individuals with autism as between 25% and 50% [Lord et al., 2004; Bailey et al., 1996]. See the Supplemental Methods and Supplemental Table A2.1 (Appendix 2) for more details.

Results:

Sequential Updating

Our initial multipoint linkage analysis was sequentially updated with genotype data from the 87 additional families identified from the AGRE database who met our criteria for the NVMSD:ALL phenotype. There was increased evidence for linkage at loci 1q24.2, 3q25.31, 4q22.3, 5p12 and 17q22. The PPL signal at locus 4q22.3 remained stable while evidence for linkage decreased at loci 5q33.1, 17p12, and 17q11.2. Results of the updated NVMSD:ALL genome-wide linkage scan are shown in Figure 3.1A-3.1E and summarized in Table 3.2.

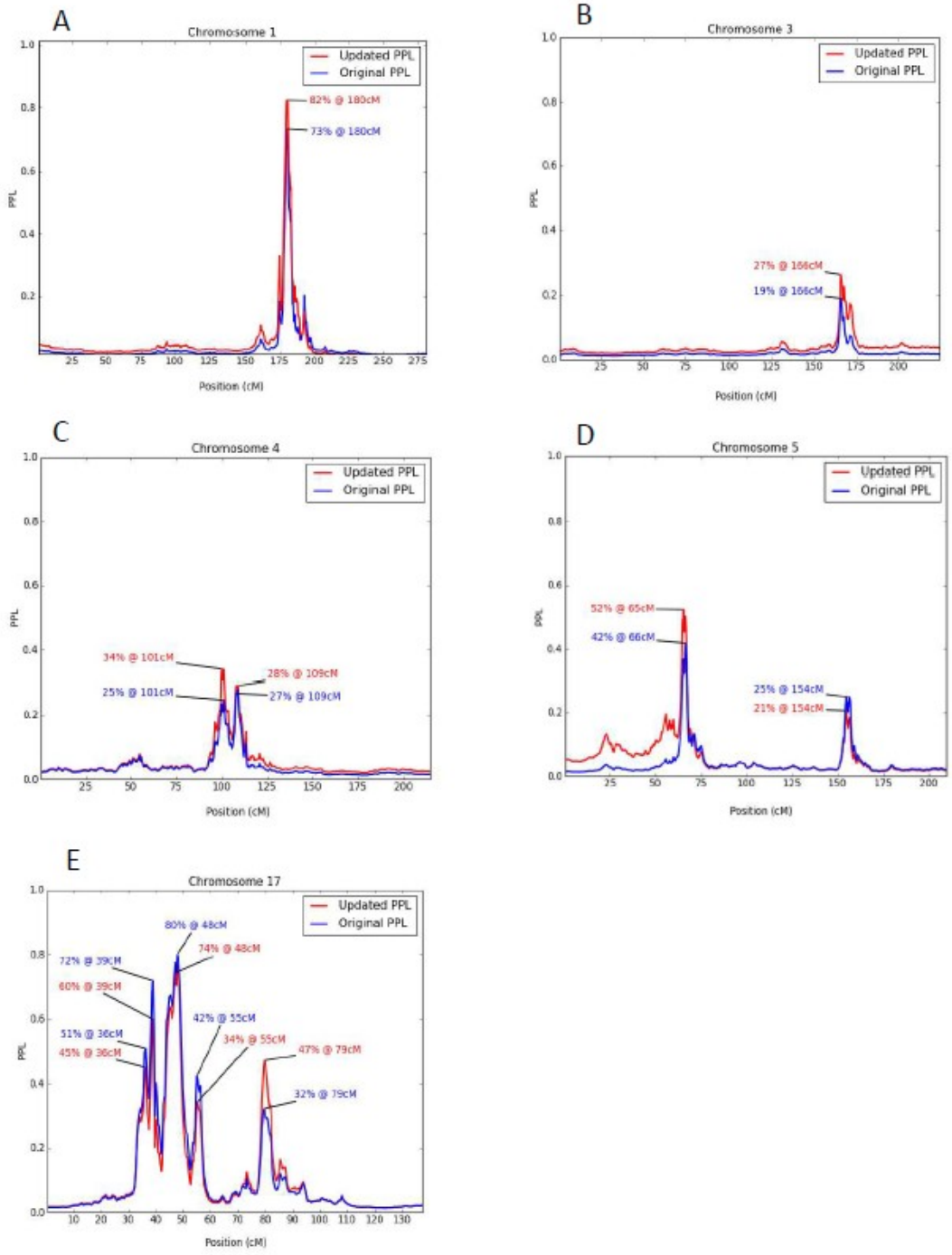


Figure 3. 1: Sequential Update of NVMSD:ALL families.

The Original PPL analysis (blue line) and Updated PPL analysis (red line) is plotted for each linkage region: Figure 1A) Chromosome 1, Figure 1B) Chromosome 3, Figure 1C) Chromosome 4, Figure 1D) Chromosome 5, and Figure 1E) Chromosome 17.

Table 3. 2: Sequential update of linkage scan.

Increased PPL Signal			
Locus	Position (cM)	Original PPL (%)	Updated PPL (%)
1q24.2	180	72	82
3q25.31	166	19	27
4q22.3	101	25	34
5p12	65-66	42	52
17q22	79-80	32	47

Stable PPL Signal			
Locus	Position (cM)	Original PPL (%)	Updated PPL (%)
4q22.3	109	27	28

Decreased PPL Signal			
Locus	Position (cM)	Original PPL (%)	Updated PPL (%)
5q33.1	154	25	21
17p12	36	51	45
17p12	39	72	60
17p12	48	80	76
17q11.2	55	42	34

The original linkage regions were updated with genetic information from an additional 87 families (NVMSD:ALL). The original PPL values and updated PPL values are listed for each linkage peak.

Fine Mapping Results

The cPPLD results are listed in Table 3.3. The marker with the highest cPPLD value was rs12938762 (cPPLD = 7%), located within the region on chromosome 17 containing the gene

Table 3. 3: Association Analysis Results for NVMSD:ALL

Chromosome	Marker	Position (cM)	cPPLD (%)	cPPLD-LC (%)	cPPLD-Aut (%)
17q12	rs16650682	48.100	3.00	3.00	4.00
17q12	rs12938762	48.351	7.00	7.00	4.00
17q12	rs35612940	48.356	4.00	4.00	2.42

Association results are listed for each SNP with a cPPLD > 2%. The cPPLD is for NVMSD:ALL as the affection status. The cPPLD-LC uses NVMSD:ALL as the affection status and autism diagnosis as the liability class. The cPPLD-Aut represents the use of autism diagnosis as the affection status.

TRPV2 at 48.35 cM. The two addition SNPs that showed modest evidence for linkage disequilibrium in TRPV2 were rs35612940 (cPPLD = 4%, 48.35cM) and rs11650682 (cPPLD = 3%, 48.10cM). All other SNPs analyzed did not show evidence for linkage disequilibrium (cPPLD ≤ 2%). When the cPPLD was calculated using autism as a liability class for NVMSD:ALL, evidence for linkage disequilibrium was unchanged for these three SNPs. Using autism diagnosis as the affection status resulted in decreased evidence of linkage for rs12938762 (cPPLD = 4%) and rs35612940 (cPPLD = 2.42%) but slightly increased it for rs11650682 (cPPLD = 4%). While these results indicate that there is not a strong interaction between NVMSD:ALL phenotype and autism diagnosis, they show that NVMSD:ALL is the driving force of the evidence seen for association.

Determination of Sample Power

Simulations were conducted to test the power of our sample to detect association. Using the prevalence of NVMSD:ALL in the AGRE database, association can be detected using either a semi-recessive model or a semi-dominant model. For the semi-recessive model, association cannot be detected for disease markers with DGFs less than 5% (Supplemental Figure A2.2; Appendix 2). For disease markers with DGF = 5%, models with high disease allele risks had 80% power to detect association when 0-10% of the families were unlinked. The power to detect association increased as the gene frequency increased. For disease markers with DGF = 10%, models with disease allele risks above 5 had 80% power to detect association. For models where the disease allele risk was 8 or higher, association was detected with as high as 80% unlinked families. This trend persisted and at a DGF = 30%, models with disease allele risks ≥ 3 had 80% power to detect association. For disease allele risks above 8, association was detected for models with up to 80% unlinked families.

These effects were less pronounced using a dominant model (Supplemental Figure A2.3; Appendix 2). Association was detected with DGFs as low as 1% and the percentage of unlinked families only had an effect at a low disease risk. When the DGF = 5%, models with disease allele ≥ 3 had 80% power to detect association. When the risk was above 7, association was detected for models with up to 70% unlinked families. For a DGF = 10%, models had 80% power when the disease allele risk was above 3. Models with a disease risk above 8 had 80% power to detect association with up to 80% unlinked families. This trend was echoed for models with a DGF = 30%, however a lower percentage of unlinked families was tolerated (disease risk ≥ 6 yielded 80% power with up to 60% unlinked families). Similar results were seen for both recessive and dominant models using the estimation of NVMSD:ALL prevalence in the general population (data not shown). Since this prevalence is much smaller, the presence of the disease allele muted the effects of varied DGFs. This set of simulations was most affected by the percentage of unlinked families and extreme relative risks.

Discussion:

In this follow up study to Flax et al., 2010, the inclusion of the 87 additional families that met criteria for NVMSD:ALL increased the statistical power of our linkage analysis and increased evidence of linkage to the NVMSD:ALL phenotype identified within the AGRE database. Fine mapping of candidate genes did not result in strong evidence of association. This could be the result of several possibilities including selection of a small number of candidate genes, low sample power to detect association, or the presence of rare variants that could dampen association signals and be contributing factors to autism spectrum disorders at these loci.

With an average increase of 10%, evidence for linkage was strengthened in novel linkage regions on 3q25.31, 4q22.3, 5p12 and 17q22. Evidence for linkage also increased on 1q24.2,

which is a region where a chromosomal deletion was identified in one AGRE family [Davis et al., 2009]. Evidence for linkage decreased in other linkage regions. While 5q33.1 was a novel linkage finding to this study, the inclusion of the 87 additional families decreased the PPL value from 25% to a 21% chance of linkage. The regions on 17p12 and 17q11.2 had previously been identified in several linkage analyses that used AGRE samples [Yonan et al., 2003; McCauley et al., 2005; Ylisaukko-oja et al., 2006]; however, linkage findings in this region have been inconsistent and vary with the inclusion of different subsets and phenotypes of the AGRE database[Yonan et al., 2003; McCauley et al., 2005; Ylisaukko-oja et al., 2006; Cantor et al., 2005; Allen-Brady et al., 2010; Alarcon et al., 2008; Benayed et al., 2005; Molloy et al., 2005; Liu et al., 2008]. The inclusion of the additional 87 families in this study reduced the PPL values in this region by an average of 6%. Taken together, these findings suggest that there is a considerable amount of genetic heterogeneity in the AGRE database and the identification of endophenotypes, such as NVMSD:ALL, can reduce the heterogeneity of the sample set and identify loci associated with more specific characteristics of ASD.

Unlike our linkage analysis, our initial association analysis did not produce strong evidence for association [Flax et al., 2010]. The Affymetrix 5.0 SNP array may not have a SNP density that is comprehensive enough to tag all common haplotypes. In order to investigate association within our linkage regions, candidate genes were fine mapped. Modest association signals were obtained in an intron of *TRPV2* (rs12938762, 48.35 cM, cPPLD = 7%). *TRPV2* encodes a temperature sensitive ion channel protein and is implicated in 'Nervous System Development and Function' by IPA's core analysis and has Gene Ontology annotations related to axonal growth and axon growth cones [Carbon et al., 2009]. Shibasaki et al., 2010, showed expression of TRPV2 in motor neurons and implicated their function in axon outgrowth regulation.

The overall lack of strong evidence for association in the fine mapping analysis could be accounted for by three potential hypotheses. The first possibility is that the candidate gene list identified using IPA needs to be expanded. Of the 645 genes located within our linkage regions, 11 were selected for fine mapping analysis. While IPA is effective in identifying molecules that contribute to known functions, the analysis is limited by the availability of published data for these molecules. The causative loci may lie within a gene that has not been studied extensively or within an intergenic region.

The second possibility is that while the sample size is of sufficient power to detect linkage, perhaps the power is not sufficient enough to detect association. Results from simulated genotype data in this dataset confirm the ability to detect association in this sample with a few caveats. Recessive models with DGF = 10% had 80% power to detect association models with a moderate to high disease allele risk. As the DGF increased, the disease allele risk required to detect association decreased. This sample had greater power to detect association with dominant models than recessive models at lower disease allele risks. As current methods that are used to detect association can have difficulty detecting association for disease risks below 2 [Nsengimana and Bishop, 2012], it is not surprising that our sample lacked power for low risk models. Overall, the simulation analysis showed that the NVMSD dataset has sufficient power to detect association for models with a moderate disease allele risk with a high degree of unlinked families.

Given the simulation results, there are two scenarios that could contribute to the lack of detectable association in this dataset. The first would be the presence of a high percentage of unlinked families for the given genomic region(s). Although, this could be applicable to the linkage regions containing *EIF4E* and *GHR* as the heterogeneity in these regions is high (α =

0.25), the maximized DGF for these regions is 10%, which has a high probability of detecting association with heterogeneity in this dataset according to our simulations. The second scenario is that the disease risk of the causative allele is too low to detect in this dataset. While the PPL is essentially model free, the algorithm calculates the maximum likelihood of each model parameter. The maximized models for each gene of interest are listed in Table 3.4. With the exception of *LMX1A* and *NKX6-1*, the maximized models correspond to low disease risk and have less than 80% power to detect association in the NVMSD:ALL dataset (see Supplemental Figure A2.4; Appendix 2). Low risk could be attributed to the low disease penetrance that is predicted by the maximized model or the interaction of multiple causative loci, both of which would cause a lack of evidence for association.

The maximized model for the linkage peak containing *LMX1A* is correlated with a high disease risk with all families linked and a DGF of 1% (Table 3.4). Our simulations indicate that association, if present, is detectable using dominant conditions. However, a strong dominant model with 99% penetrance is not likely in the NVMSD:ALL dataset. As NVMSD:ALL phenotype is based on variables from the ADI-R, which is a parental report on affected children, the parents in our sample are, by default, unaffected for the phenotype. The overwhelming lack of affected parents renders the strong dominant model unlikely, resulting in low power to detect association (see Supplemental Figure A2.4; Appendix 2). Why, then does the PPL maximize at this model? *LMX1A* resides along the edge of this linkage region (PPL = 15%) and its maximized model is unique to this section of the linkage region. The maximized model for the region directly under the linkage peak has a high allele risk (DGF = 1% and P(DD) = 0.8, P(Dd) = 0.8, P(dd) = 0.1); the NVMSD:ALL dataset has 80% power to detect association. Perhaps the unlikeliness of the *LMX1A* model also decreases the evidence for linkage.

The maximized disease model for *NKX6-1* has a high disease risk and the NVMSD:ALL dataset has sufficient power to detect association for this model. Three tag SNPs were identified for this candidate gene and, after genotype cleaning, only one was retained for association analysis. This SNP did not show evidence for association, indicating a true lack of association at this locus. Overall, the NVMSD:ALL dataset has enough power to detect association under most models. However, simulations indicated that the regions selected for fine mapping are the result of low disease risk that result in the inability to detect association in this dataset.

Table 3. 4: Maximized PPL models for each linkage region.

Chromosome	Gene	PPL (%)	α	DGF	P(dd)	P(Dd)	P(DD)
1	LMX1A	15	1	0.01	0.1	0.8	0.999
4	NKX6-1	15	1	0.3	0.1	0.1	0.6
4	EIF4E	18	0.25	0.1	0	0	0.1
5	GHR	41	0.25	0.1	0	0	0.2
17	PMP22	77	0.65	0.1	0	0.1	0.1
17	UBB	79	0.65	0.1	0	0.1	0.1
17	TRPV2	79	0.65	0.1	0	0.1	0.1
17	NCOR1	79	0.65	0.1	0	0.1	0.1
17	NF1/OMG	34	0.4	0.3	0	0	0.1
17	CDK5R1	39	0.4	0.3	0	0	0.1

The maximized model consists of the following parameters: α is a measure of heterogeneity, DGF is the disease gene frequency, P(dd) is the penetrance of a homozygous non-risk allele, P(Dd) is the penetrance for a heterozygous genotype, P(DD) is the penetrance for a homozygous risk allele.

Lastly, the lack of strong evidence for association seen in this study also supports the hypothesis that allelic heterogeneity and/or rare variants are major contributing factors to ASDs. Due to the general lack of positive and replicative findings in genome-wide association studies, this hypothesis has become increasingly more popular. While linkage is a family based statistic, association is a population statistic that is based upon the segregation of alleles that originate from a common ancestor. Allelic heterogeneity, where multiple variants within the same gene

cause the same phenotype, is not a limitation to linkage analyses as an individual variant would segregate within a family. However, if a different causative variant resides in each family there would not be a universal variant for the entire population. This would cause the gene to be linked, but not associated for the given population. This is also true for recent or rare mutations that appear in a small percentage of individuals, but not in the population as a whole. However, with the acquisition of larger datasets and meta-analyses, the effect size of these rare mutations and allelic heterogeneity may increase as statistical power increases and multiple instances are discovered [Allen et al., 2010; Ke, 2012; Maher et al., 2010]. Additionally, many recent investigations have identified rare *de novo* mutations by directly sequencing autism candidate genes (reviewed in [Devlin and Scherer, 2012]), including *CNTNAP2* [O'Roak et al., 2012] which has been implicated in both autism and language impairment. In addition to focusing on known candidate genes, many investigators have identified mutations in new candidate genes and protein networks through whole exome sequencing of autism probands [O'Roak et al., 2012; Sanders et al., 2012; Neale et al., 2012]. A study by Neale et al., 2012, showed that while the rate of variants in ASD cases was only moderately higher than expected, protein-protein interaction analyses showed that proteins encoded by genes with *de novo* missense or nonsense mutations had a high rate of interconnectivity. This is consistent with results seen by O'Roak et al., 2012, in the Simons Simplex Collection. Exome sequencing revealed an interconnected network of disrupted β-catenin/chromatin remodeling proteins as well as a strong paternal bias for *de novo* mutations. Both Sanders et al., 2012 and O'Roak et al., 2012, identified a correlation between increased parental age and the rate of *de novo* single nucleotide variants. Another study from the Simons Simplex Collection by Sanders et al., 2012, identified several genes with disruptive *de novo* mutations in multiple probands. When conditioned for genes involved in brain function, the effect size for these mutations increased significantly. Similarly, Ben-David

and Shifman, 2012, identified networks of neuronal genes that were enriched for known rare and common variants in three autism samples.

Conclusion:

The linkage peaks that were first identified in Flax et al., 2010, and updated in this paper support the hypothesis that there may be a speech/language phenotype associated with autism that is also seen in other developmental disabilities without autism although often with different degrees of severity. The NVMSD:ALL phenotype is a behavioral biomarker that can be used to reduce the heterogeneity of ASD samples, especially in the communication domains and to allow for the identification of candidate loci. Despite strong evidence for linkage, allelic heterogeneity and rare variants may explain the lack of evidence for association in these linkage regions. The combination of well-defined phenotypes and genome sequencing are promising techniques that may help to identify causative variants in autism that have been overlooked by current association studies.

Chapter 4: Characterizing Language Profiles in Families Ascertained for Autism and Specific Language Impairment

Judy F. Flax, Abby Hare, Zena Fermano, Linda M. Brzustowicz

Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ

Barbie Zimmerman-Bier

Department of Pediatrics, Saint Peter's University Hospital, New Brunswick, NJ

Charles Cartwright

YAI Network, New York, NY and New Jersey Medical School, Newark, NJ

Steven Buyske

Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, NJ

Christopher W. Bartlett

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH

Abstract

Background: Autism and Specific Language Impairment (SLI) are complex and distinctly disorders, yet some individuals with autism who are verbal share both oral and written language characteristics with individuals diagnosed with SLI. It remains a question whether these similarities arise from a shared genetic etiology or if they are just descriptively similar to each other. In this paper we describe language phenotypes characterized for a genetics study of autism that support the idea of partial shared genetic etiology.

Methods: Language phenotypes were developed from verbal subjects with autism and family members who met criteria for language impairments. Fifty-three (53) nuclear and extended families received a battery of cognitive, oral, and written language assessments. Each nuclear and extended family had at least one proband with a diagnosis of autism and another who met criteria for Specific Language Impairment (SLI). Other family members with language issues were further characterized as meeting criteria for oral language impairment (LI), and/or written language impairment (RI).

Results: Family members who met criteria for LI and those who met criteria for both LI and RI did not differ significantly from verbal autism spectrum probands on most oral and written language measures regardless of the specific language constructs compared. Those family members who met criteria for RI alone performed significantly different from the verbal ASD probands on several higher order language measures.

Conclusion: The language impairment profiles seen in a subset of family members with a diagnosis of ASD and those with oral and written language impairments support a hypothesis suggesting partial shared genetic etiology between the two disorders thus, supporting the development and application of language phenotypes in genetic studies of ASD.

Introduction

Autism Spectrum Disorders (ASDs) are complex neurodevelopmental disorders encompassing a large behavioral continuum that includes 1) social interaction deficits, 2) communication deficits, and 3) excessive restricted and repetitive behaviors. The characteristics of ASD can range in severity across all three domains. While 20-40% of individuals with ASD fail to develop functional verbal language [Bailey et al., 1995; Lord et al., 2004], verbal individuals with ASD can have difficulty with pragmatic language, structural language, and phonological short-term memory (PSTM) [Kjelgaard and Tager-Flusberg, 2001; Lewis et al., 2007; Joseph et al., 2002; Rapin et al., 2009]. Even though the social aspects of communication deficits associated with ASD are defining in terms of ASD diagnosis, 39-61% of verbal individuals on the autism spectrum may present with communication deficits on the segmental level of language similar to the profiles that are seen with individuals with a diagnosis of Specific Language Impairment (SLI) [Rapin and Dunn, 2003]. SLI is a neurodevelopmental disorder that is characterized by delayed language development in the absence of intellectual, sensory, or other neurological abnormalities. ASD and SLI are distinct disorders, but the qualitative overlap in segmental language patterns in some individuals on the autism spectrum support the hypothesis that there may be some shared etiology between the two disorders.

The literature is clear that SLI is not a milder form of ASD and that the same genetics do not universally underlie both disorders. The critical gap in the literature is the limited knowledge of what modulates language ability within the ASD population to induce these high rates of communication deficits. Characterizing the language impairments observed in ASD in relation to those observed in SLI can play a significant role in addressing the issues of shared or partial etiology between these two disorders which in turn, from a behavioral standpoint, can play a significant role in the development of behavioral biomarkers for genetic linkage and

association studies that can be then follow up by the search for common and rare variants associated with ASD. The study presented in this manuscript analyzes these traits in a cohort of families that were ascertained for the presence of at least one autism proband and another SLI proband.

Studies of Autism and SLI

The definitions of ASD and SLI draw distinct lines between the two disorders. The communication impairment seen in ASD is primarily social and pragmatic, while SLI is predominantly structural. If ASD and SLI are truly separate and non-overlapping then impairments in structural language and pragmatic difficulties would not be displayed in both disorders, which is not the case. SLI-like structural language impairments have been reported in several ASD cohorts [Kjelgaard and Tager-Flusberg, 2001; Lewis et al., 2007; Rapin et al., 2009] and subsets of children with SLI also meet the criteria for social and communication impairments using ASD diagnostic tools [Leyfer et al., 2008]. The increased prevalence in overlapping characteristics suggests that the two disorders are not completely independent and some hypothesize that there is an etiological overlap in subgroups of individuals with these disorders [Roberts et al., 2004; Rapin and Dunn, 2003]. Other groups speculate that the qualitative overlap of characteristics in both disorders is superficial and the fundamental causes of these characteristics are distinct [Whitehouse et al., 2008; Williams et al., 2008]. The most common approach to addressing these hypotheses has been to compare the language profiles of individuals with ASD to those of SLI.

During the 1980's and 1990's Rapin and colleagues [Rapin, 1996; Tuchman et al., 1991] studied the language behaviors of children with ASD with concomitant language issues and children with Developmental Language Disorders (DLD), which are developmental disorders

where individuals have no other obvious behavioral, cognitive or neurological issues other than significantly deficient language proficiency. Their findings supported the concept that subsets of children with ASD who were verbal and children with DLD were parallel in language behaviors that include severe problems in language comprehension (with emphasis on impoverished and unintelligible speech) and higher order processing of language. Kjelgaard and Tager-Flusberg, 2001, reported similar findings when comparing the structural aspects of language in a group of verbal children with ASD to the types of language profiles that might be seen in a group of children with SLI. Especially noteworthy was the finding that children with ASD who performed poorly on a comprehensive test of language and vocabulary also performed poorly on a non-word repetition task, a task that has since been replicated and identified as representing a potential biomarker for both SLI and autism [Vernes et al., 2008].

More recently, McGregor et al., 2012, looked at the semantic language abilities of a group of children with SLI and a group of verbal children with ASD who had syntactic deficits (ASDLI). Both groups performed poorly on tasks of syntax, vocabulary, word definitions and word association while there were differences in the group in terms of pragmatic abilities. Leyfer et al., 2008 explored pragmatic abilities in both groups by comparing the scores on the Communication and Social Scales of the ADI-R and the ADOS for children with ASD and children with SLI (without a diagnosis of autism/ASD) who were matched for non-verbal IQ. They reported that 41% of the children with SLI also met the ASD cut-offs in the Social and Communication domains of either assessment. This overlap also exists in adolescents and young adults [Howlin et al., 2000; Mawhood et al., 2000; Conti-Ramsden et al., 2006]. These studies support the notion that both disorders may share some etiological factors.

Conversely, there are studies suggesting that similarities found in the language behaviors observed in ASD and SLI may not be supportive of shared underlying genetic risk with

differences being reflected in the rates of impairment in families and behavioral differences primarily in social language and pragmatics skills. One of these, Lindgren et al., 2009, compared ASD and SLI probands as well as their nuclear family members on a series of cognitive, language and reading measures. The profiles of children with ASD with structural language impairments and those with SLI looked quite similar and 16-35% of the nuclear family members of the ASD group did poorly on the language measures. However, a majority of the family members of the SLI cohort did more poorly on most of the tasks than the nuclear family members of the ASD group. This finding supports a similar finding in a comparison of parents of ASD probands (Par-A), parents of SLI probands (Par-L), and parents of typically developing individuals (Par-T) [Whitehouse et al., 2007]. This study found that the Par-T group was most similar to the Par-A group for language tests and the Par-L group for pragmatic difficulties. Overall the Par-A group performed better on the language tests than the Par-L group, but presented with more pragmatic difficulties. Most recently, using questionnaire history data, Pickles et al., 2012, compared the first-degree relatives of children with ASD, SLI, and Down Syndrome (DS). Like the previous studies, they reported that nuclear family members of children with SLI presented with more communication deficits and fewer social deficits, while the families members of children with ASD had more social deficits. These studies support the hypothesis that any overlap in qualitative traits between ASD and SLI is superficial and is not the result of an underlying shared etiology.

All of these studies examined SLI and ASD using a more traditional case-control study design. By convention, researchers studying SLI and ASD define these to be mutually exclusive and therefore each group of the case-control design has distinctly different diagnostic entities in terms of behavior. However, both ASD and SLI are genetically complex, raising the possibility that some individuals within each disorder who share similar language characteristics may also

share specific susceptibility genes or variants of genes. Further investigation using alternative study designs is required to identify the role of shared genetics between ASD and SLI.

Relations Among Oral and Written Language Impairments and ASD

It is now widely accepted that oral and written language impairments may result from the same underlying deficits with both behavioral and genetic findings suggesting a shared etiological component [Newbury et al., 2011]. This evidence is supported by two sources. There is extensive research of longitudinal studies of children who are followed from the preschool years to adulthood. Typically, these studies report on children who were identified with oral language issues during the preschool years and who have difficulties learning phonics or have difficulty with sight word reading during the earlier grades. These individuals often have reading comprehension and spelling issues in later school years and adulthood, even after some oral language issues have resolved [Ek et al., 2012; Ricketts, 2011; Whitehouse et al., 2009]. The second source comes from family studies. In these studies, a subject is ascertained based on either a language or reading impairment. Direct testing and/or history report data yield rates and co-occurrence of reading and/or language-based learning disabilities in other family members that is significantly greater than what is found in the normal population or in a designated control group ([Flax et al., 2003]; see Stromswold, 2008 for a review). Further evidence of this link comes from genetic association studies. Newbury et al., 2011, reported trends toward association on several genes in SLI families using family members presenting with reading and/or language impairments. They concluded that at the very least there are some shared genetic effects across SLI and dyslexia.

There is great variability in behaviors and level of function of all characteristics of individuals on the autism spectrum. This variability is also evident in the reading abilities of

individuals on the spectrum. Reading ability in individuals with ASD can include non-readers, those with exceptional phonological or sight reading skills with poor comprehension (hyperlexia), and those who have overall exceptional reading abilities. As oral language and reading may share some underlying deficit and some individuals with ASD present with oral language deficits similar to those identified in SLI, the examination of written language behaviors in ASD should be included when exploring the genetics of language and autism.

Language Impairment as a Genetic Biomarker

Over the past ten years, attempts to use language phenotypes as genetic biomarkers in ASD have produced inconsistent findings. Family studies of linkage and association have been the primary model for study. An area on chromosome 7 (q34–36) and the gene *CNTNAP2* have been linked to both autism and expressive language impairments. *CNTNAP2* encodes a protein that is known to influence early brain development in humans [Alarcon et al., 2008; Arking et al., 2008; Vernes et al., 2008]. Using language phenotypes such as onset of first words, onset of first phrases, and unintelligible speech in ASD probands or delayed language in other family members, genetic linkage has been reported on several other chromosomes [Alarcón et al., 2005; Bartlett et al., 2004; Bradford et al., 2001; Flax et al., 2010]. In an attempt to better define the language endophenotype and reduce phenotypic heterogeneity, Spence et al., 2006 stratified expressive language characteristics into word and phrase speech delay in ASD probands and history of language delay or reading difficulties in their parents. They found nominal evidence of linkage in several already reported areas but not for locations on chromosomes 7 and 13 where several other research groups had identified significant linkage findings. Stratification by parental history did not strengthen linkage, resulting in the conclusion that more discretely defined characteristics of ASD, specifically language endophenotypes, may improve localization of linkage signals and strengthen existing findings.

One approach to the genetic links that may exist between autism and SLI was explored using the "common variant-multiple disease" hypothesis first suggested by Becker, 2004, and then related to autism and SLI by Stromswold, 2008. Becker's theory proposed that common alleles that contribute to a disease under certain genetic and environmental conditions may present as a completely different disease under different genetic and environmental conditions. Stromswold related this theory to autism and SLI where particular etiological factors and behaviors may be specific to a disorder (i.e. repetitive behaviors in autism) while other factors and behaviors may transcend disorders and be present in two or more neurodevelopmental disorders (i.e. language in this case). This could explain the current difficulties for both SLI and autism genetic research where a variety of specific language phenotypes may be present in a subgroup of individuals in both disorders (i.e. grammar or phonology) resulting in overlapping findings.

Rationale for NJLAGS design

Many groups have examined the role of language impairment as part of the broader autism phenotype, others have looked at the roles and rates of language-learning impairments in the families of children with ASD, and still others have compared these rates to rates in other developmental disabilities and rates in control families. While hypotheses and results differ, much may be attributed to variation in definitions of language impairment, the distinction between individuals with pure SLI versus individuals with autism who have an associated language impairment (ALI), and the size and description of the samples. To date, case-control studies of separate ASD and SLI cohorts have been used to investigate the clinical overlap in symptoms. However, by design these studies may miss the role of partial genetic etiology by selecting for specific subsets of individuals with ASD or SLI that do not overlap clinically. As there are subsets of individuals with ASD who have SLI-like communication impairments, and

vice versa, it is reasonable to hypothesize that these subsets of individuals may have some genetic variants in common. Similarly, linkage and association studies based on language phenotypes have resulted in few replications of results based on ASD probands and/or family member phenotypic characterization. We have taken all of these issues under consideration in this first study to intentionally ascertain families for ASD and SLI within the same family. In all families, the SLI probands have no evidence of ASD and the ASD probands may or may not present with a language impairment but do not meet strict criteria for SLI. We report the methods involved in the creation of communication phenotypes and the disproportionate rates of language issues in these families even after accounting for ascertainment bias. The results of the behavioral and molecular correlations and heritability data will support the argument that a subset of individuals on the autism spectrum may share some genetic and behavioral vulnerability to communication disorders also seen in individuals with SLI.

Methods

The New Jersey Language and Autism Genetics Study (NJLAGS) has as its over-arching goal advancement in the development of behavioral and genetic biomarkers for autism spectrum disorders. One of the primary aims of the study is to characterize language in families with at least one person with the diagnosis of Autism and at least one other person with a diagnosis of SLI.

Subject Selection

Criteria for family recruitment for the NJLAGS study included: 1) at least one individual with a diagnosis of Autism or Autistic Disorder with no known genetic cause (i.e. Fragile X or Rett's), 2) at least one individual who met criteria for SLI (and not on the autism spectrum), 3) at least three other family members willing to participate, and 4) English as the primary language of all individuals participating. All family members also agreed to submit a blood sample for

DNA extraction that was then processed and genotyped at the Rutgers University Cell and DNA Repository (RUCDR) Infinite Biologics, Rutgers University in Piscataway, NJ. All research for the NJLAGS study was approved by the Rutgers University Institutional Review Board with all eligible subjects consented or assented based on university guidelines.

To be identified as the *Autism Proband*, the individual was required to meet the cut-off for Autism or Autistic Disorder on at least two of the three following measures: 1) Autism Diagnostic Interview-Revised (ADI-R) (Lord et al., 1994), 2) Autism Diagnostic Observation Scale (ADOS) (Lord C, et al., 2000), and 3) Diagnostic and Statistical Manual-IV (DSM-IV) (American Psychiatric Association, 1994). An autism diagnosis was chosen as the proband criteria for induction into the study rather than the more general criteria of ASD in order to reduce phenotypic and genetic heterogeneity. For this study, 60 individuals met the diagnostic criteria for autism and were identified as autism probands. Ten other family members met the diagnostic cut-off for ASD (total ASD= 70). Thirty-two of the 70 subjects with autism and ASD (45.7%) had sufficient language skills to participate in a portion of the behavioral testing battery along with all other family members. The 38 individuals who met criteria for autism or ASD but did not participate in the testing battery were non-verbal, minimally verbal or unable to participate in testing due to associated behavioral issues.

An individual qualified as the *SLI Proband* if the individual obtained the following scores: 1) an age appropriate Comprehensive Evaluation of Language Fundamentals-4 (CELF-4) [Semel et al., 2003] or CELF Preschool-2 [Wiig et al., 2004] core standard score ≤ 85, or 2) at least 1 SD below the standard score on ≥ 60% of all language subtest scores with a significant history of language/reading difficulties and/or a significant history of language and reading difficulties (childhood diagnosis of language and/or reading impairment and/or at least 2 years of intervention). All SLI probands passed a hearing screening (20dB at 1000, 2000, 4000 Hz and

30dB at 500Hz) and an oral motor exam to rule out expressive language difficulties resulting from neuromuscular involvement. See Supplementary Table 3A.1 (Appendix 3) for a description of the full testing battery. Additionally, the proband must have had a non-verbal IQ ≥ 80 and always greater than the CELF-4 core standard; no individuals were excluded from the analyses presented here due to low IQ. Importantly, the SLI proband had no history or diagnosis of any other developmental or neurological disorders, including ASD. A total of 62 non-ASD individuals met criteria for the SLI proband.

Once the Autism and SLI probands had been identified, the entire family (including all verbal family members with ASD) received the neuropsychological battery that included multiple tests of spoken and written language, cognition, phonological processing, and scales of social functioning and associated ritualistic and rigid behaviors (Supplementary Table 3A.1; Appendix 3). Trained professionals administered assessments in clinical settings or in the family home. From a previous family study of SLI, there is evidence that behavioral testing in either the home or the clinic is equally reliable [Flax et al., 2003; Tallal et al., 2001].

Table 4. 1: Nuclear and extended family profiles for the 53 NJLAGS families with at least one ASD proband and one SLI proband.

	N	mean age (s.d.)
Fathers	50	42.6 (6.4)
Mothers	53	40.7 (8.5)
Brothers	39	9.7 (5.8)
Sisters	29	12.4 (6.5)
Grandparents of probands	15	68.2 (8.1)
Aunts and Uncles of probands	19	39.5 (4.6)
Cousins of probands	23	10.5 (5.4)
Other extended family members	20	32.4 (9.1)
Total	312	
Family members meeting criteria for Autism	60	10 (6.9)
Family members meeting criteria for ASD	10	7.0 (2.3)

In this paper we are reporting results from 53 families to date that have been genotyped and have participated in most of the behavioral testing program. Due to minimum age requirements of some assessment some family members were unable to complete the entire testing battery. Table 4.1 represents the breakdown of the nuclear and extended

family subjects in those 53 families and the numbers of individuals meeting study criteria.

Oral and Written Language Assessments

The comprehensive testing battery of language-based and cognitive assessments was used to develop behavioral phenotypes. Although tasks varied as a function of the age of family members, the same or similar language constructs were tested for all subjects capable of receiving the assessments. While it is recognized that, in some cases, parents and other adult family members fell outside the standardized scoring range of the tests used (age > 18-21 years), there was no evidence of ceiling effects in this sample when administered to ages beyond the normative data.

The language assessment battery targeted multiple areas and constructs associated with language development. Segmental Language is involved in the basic comprehension and expression of syntax, morphology and vocabulary at the sentence level. Phonological Awareness has been associated with difficulties in listening and reading. Higher Order and Pragmatic Language addresses issues in abstraction, metaphor, intention, ambiguity, and real world knowledge. Higher Order language measures were included to evaluate language abilities often not captured in traditional language tests but that may reflect undetected language issues present in older children and adults. Written Language included tests of single word reading, nonsense word reading, reading comprehension, rate, accuracy, and spelling.

All family members who were able to participate in the testing battery were classified as unaffected for oral and written language or affected for language using the following acronyms: oral language impaired (LI) or written language (reading) impaired (RI) Note: The acronyms of LI (for oral language impairment) and RI (for written language impairment) were used to better match the current literature in this area.

Oral Language Impairment Criteria (LI): defined as an age appropriate CELF-4 core standard score ≤ 85, or at least 1 SD below the standard score on ≥ 60% of all language subtest scores with a significant history of language/reading difficulties. The language subtests included age appropriate CELF-4 subtests, Comprehensive Assessment of Spoken Language (CASL) subtests (Ambiguous Sentences, Inferences, Nonliteral Language, Pragmatic Judgment, and Meaning from Context) [Carrow-Woolfolk, 1999], and Comprehensive Test of Phonological Processing (CTOPP) subtests (Elision and Nonword Repetition) [Wagner et al., 1999].

Written Impairment Criteria (RI): defined as a score of at least SD below the standard mean on at least 60% of the Woodcock Reading Mastery Tests (WRMT) subtests (Word Identification and Word Attack) [Woodcock, 1987], the Gray Oral Reading Test - 4 (GORT-4) [Wiederholt and Bryant, 2001] comprehension standard score, and overall reading quotient.

The breakdown of individuals affected for LI and/or RI is listed in Table 4.2. A total of 10 verbal autism probands also met criteria for LI only, 3 met criteria for RI only, and 11 met criteria for both LI and RI. All autism probands who met criteria for LI or LI+RI were included as affected for ALI (N = 21) where noted. As only 3 autism probands met the criteria for RI, ARI was not included as an independent comparison group.

Determination of Impairment Rates

The rates of language-based learning impairments were determined for the 53 NJLAGS families who met the strictest criteria of having at least one Autism proband and at least one SLI proband. First, the rates of language-based learning impairments in nuclear families were determined for individuals who met criteria for LI ONLY, RI ONLY, and those who met criteria for both impairments (LI+RI). Rates were then determined in both nuclear and extended families by including and then excluding the ASD probands who met criteria for ALI.

Comparison of Group Test Performance

SPSS [IBM Corp, 2012] was used to conduct independent t-tests to compare the mean language subtest scores of the LI group to the ASD group as well as the RI group to the ASD group. Individuals who were affected for LI+RI were included in both the LI group and the RI group in order to maximize our statistical power. Differences, confidence intervals, and nominal p-values were obtained for both comparisons. Bonferroni adjustments for multiple testing made separately for the ASD/LI and ASD/RI comparisons gave a threshold for significance of 0.002 for the nominal p-value.

Random-Effects Linear Model

Differences, confidence intervals, and nominal (i.e., unadjusted for multiple-testing) p-values were also obtained via a random-effects linear model. Individuals were considered as ASD, LI, or unaffected (for the variables on which we tested ASD/LI differences) or ASD, RI, or unaffected (for the variables on which we tested for ASD/RI differences). After blocking on family, considered as a random effect, the ASD versus LI or RI contrasts were estimated and tested. By blocking on families, we reduced confounding caused by variable numbers of LI or RI subjects per family. Bonferroni adjustments for multiple testing made separately for the ASD/LI and ASD/RI comparisons gave a threshold for significance of 0.002 for the nominal p-value.

Results

Rates of Oral and Written Language Impairment Among Family Members

We examined the actual rates of language-based learning impairments in study families looking specifically at LI ONLY, RI ONLY, and then those who met criteria for both impairments (LI+RI); keeping in mind that by design all 53 families had at least one individual with an SLI diagnosis (Table 4.2). The total number of individuals within the nuclear families who met

Table 4. 2: Rates of Oral Language Impairment only (LI only), Written Language Impairment only and Oral and Written Language Impairment (LLI).

Nuclear Family (excluding ASD probands)	Unaffected LI or RI	LI only	RI only	LI+RI	TOTAL LLI	Total tested
Fathers	32 (68%)	1(2%)	8 (17%)	6 (13%)	15 (32%)	47
Mothers	38 (78%)	1 (2%)	7 (14%)	2 (4%)	11 (22%)	49
Brothers	14(38%)	10(27%)	9(24%)	4(11%)	23(62%)	37
Sisters	13(45%)	2(7%)	8(28%)	6(21%)	16(55%)	29
TOTAL nuclear family	97 (60%)	14 (9%)	32 (20%)	18 (11%)	64 (40%)	161
Extended Family	41 (71%)	4 (7%)	8 (14%)	5 (9%)	17 (29%)	58
Verbal ASD	8 (25%)	10 (31%)	3 (9%)	11 (34%)	24 (75%)	32
Total Tested (nuclear and extended families including probands with ASD)	146 (58%)	28 (11%)	43 (17%)	34 (14%)	105 (42%)	251
Nonverbal or not testable=38						

criteria for LI and/or RI was 64, meaning that there was an increase of 11 non-designated probands with language issues in the nuclear families. When both nuclear and extended family members were included in the calculations but ASD probands who met criteria for ALI were not included, the numbers of LI and/or RI increased to 81, representing 28 non-designated probands (increase of 53%). Finally, when nuclear and extended family members as well as the ALI probands were included the numbers almost doubled to 105, or 52 non-designated probands (increase of 98%).

Comparison of Mean Performance on Language Tests

Supplementary Table 3A.2 (Appendix 3) shows group mean scores and standard deviations on all standardized measures for all family members including those who were 1) unaffected for ASD, LI, or RI, 2) affected for ASD, 3) affected for LI only, or 4) affected for RI only. Scores were broken down by segmental language, phonological processing, higher order language, and written language. It is important to note that for the following comparisons, the LI and RI groups were not mutually exclusive since there were impaired family members who met criteria for both oral and written language impairments.

Table 4.3 represents the results of comparing those family members who met criteria for LI and ASD on the subtest scores of segmental language, higher order language, phonological processing, and written language (For full table see Appendix 3; Supplementary Table 3A.3). Based on nominal p-values, the LI and ASD groups did not differ on any measures of oral

Table 4. 3: T-test comparisons of oral and written language scores for individuals with ASD (who are verbal) and individuals with LI or RI.

Segmental Language						
				LI vs. ASD	RI vs. ASD	
Subtest	Mean ASD (s.d.)	Mean LI (s.d.)	Mean RI (s.d.)	Nom. p- value	Nom. p- value	Comparison
CLFCFD	5.54 (4.45)	4.11 (2.65)	5.47 (3.70)	0.316	0.963	
CLFFS	5.70 (3.92)	5.61 (2.81)	9.17 (3.54)	0.932	<.001	RI > ASD
CLFWS	5.12 (4.26)	4.07 (3.12)	4.50 (4.07)	0.522	0.769	
CLFRS	4.64 (3.39)	4.31 (2.25)	7.02 (3.49)	0.679	0.005	RI > ASD
CLFWCT	6.31 (2.89)	5.95 (2.34)	9.24 (2.91)	0.683	0.001	RI > ASD
CLFWD	7.73 (3.17)	7.94 (3.13)	10.51 (3.27)	0.866	0.013	RI > ASD
Higher Order Language						
CASLAS	77.91 (12.49)	74.00 (7.99)	86.21 (14.13)	0.368	0.079	
CASLIN	70.10 (14.95)	74.50 (16.20)	88.18 (20.12)	0.536	0.021	RI > ASD
CASLMC	71.09 (16.31)	79.06 (8.73)	91.04 (12.30)	0.104	<.001	RI > ASD
CASLNL	70.80 (20.99)	79.07 (15.55)	88.76 (17.15)	0.127	<.001	RI > ASD
CASLPJ	70.32 (20.25)	78.94 (14.90)	88.95 (16.86)	0.078	<.001	RI > ASD
Phonological Processes						
C-TOPPEL	8.14 (3.89)	5.33 (3.10)	7.07 (3.50)	0.008	0.242	ASD > LI
C-TOPPNR	8.30 (2.72)	6.81 (2.84)	8.02 (3.13)	0.066	0.701	
Written Language						
GRTACC	6.94 (3.77)	6.45 (3.54)	7.91 (3.95)	0.681	0.372	
GRTCOMP	5.62 (3.61)	5.75 (3.11)	6.28 (2.62)	0.912	0.509	
GRTFLU	6.59 (3.57)	5.50 (3.60)	7.35 (4.84)	0.353	0.484	
GRTRQ	77.12 (19.00)	74.35 (15.56)	80.98 (15.31)	0.349	0.403	
GRTRATE	7.35 (3.00)	6.41 (3.14)	7.79 (3.50)	0.633	0.643	
WRMWA	98.67 (18.83)	89.93 (14.04)	95.05 (15.60)	0.056	0.363	
WRMWID	99.12 (20.98)	88.12 (18.73)	91.88 (15.87)	0.042	0.082	ASD > LI
WRAT	95.78 (17.74)	88.97 (15.86)	93.00 (17.01)	0.147	0.511	
WASI PIQ	93.89 (15.45)	89.81 (12.59)	99.19 (14.84)	0.295	0.078	

Comparisons that are significant after a Bonferroni correction are shown in red.

CELF-4 Subtests: CLFCFD = Concepts and Following Directions, CLFWS = Word Structure, CLFRS = Recalling Sentences, CLFWCT=Word Class Total, CLFWD = Word Definitions; CASL Subtests: CASLAS = Ambiguous Sentences, CASLIN = Inferences, CASLMC = Meaning from Context, CASLNL = Non-Literal Language, CASLPJ = Pragmatic Judgment; C-TOPP Subtests: C-TOPPEL = Ellison, C-TOPPNR = Non-word Repetition; GORT 4 Subtests: GRTACC = Accuracy, GRTCOMP = Comprehension, GRTFLU = Fluency, GRTORQ = Reading Quotient, GRTRATE = Rate; WASI PIQ=Performance IQ; WRAT = Spelling; Woodcock Reading Mastery: WRMWA = Word Attack, WRMWID= Word Identification.

language (segmental measures) except for the Elision (deletion) task of the CTOPP with the ASD group performing significantly better than the LI group on this task (ASD = 8.14, LI = 5.25, p = 0.008). However, since several of the same measures used to categorize family members as LI are included in this table, we chose to look more closely at the measures that were not used to define the LI. The LI and ASD groups did not differ on most written language measures except for the WRMT- Word Identification task, with the ASD group performing significantly better than the LI group (ASD = 99.12, LI = 88.12, p= 0.042).

Table 4.3 also represents the results of comparing the RI and ASD groups on all language measures. The RI group did not differ from the ASD group on any measures of written language. There were differences in oral language performance between the ASD and RI groups, with the RI group performing better on several segmental and higher order language measures. Based on nominal p-values, differences were seen on the following CELF-4 subtests: Formulated Sentences (RI = 9.09, ASD = 5.70, $p < 0.001$), Recalling Sentences (RI = 6.9, ASD = 4.6, p = 0.005), Word Classes (RI = 9.15, ASD = 6.31, p = 0.001 and Word Definitions (RI = 10.5, ASD = 7.7, p = 0.013). Differences were also seen on the Pragmatic Language task of the CASL (RI = 88.95, ASD = 70.32, $p < 0.001$) and the following higher order language tasks: CASL-Inference (RI = 88.2, ASD = 70.1, p = 0.021) Meaning from Context (RI = 91.07, ASD = 71.09, $p < 0.001$), and Non-literal Language (RI = 88.70, ASD = 70.80, $p < 0.001$).

In order to account for the confounding caused by variable numbers of LI or RI subjects per family, a random-effects linear model blocked by families was used to further compare these means. Table 4.4 represents the results of comparing those family members who met criteria for LI, RI and ASD on the subtest scores of segmental, phonological processing higher order language, and written language assessments. Like the results of the t-test analyses, there

was no evidence of statistical difference on any measures of oral language (segmental, higher order language and phonological processing) except for the Elision (deletion) task of the CTOPP with the ASD group performing significantly better than the LI group on this task (ASD = 8.14, LI

Table 4. 4: Comparisons of oral and written language scores for individuals with ASD (who are verbal) and individuals with LI or RI with blocking for families.

Segmental Language						
				LI vs. ASD	RI vs. ASD	
Subtest	Mean ASD (s.d.)	Mean LI (s.d.)	Mean RI (s.d.)	Nom. p- value	Nom. p- value	Comparison
CLFCFD	5.54 (4.45)	4.11 (2.65)	5.47 (3.70)	0.127	0.975	
CLFFS	**5.70 (3.92)**	5.61 (2.81)	**9.17 (3.54)**	0.989	**<.001**	**RI > ASD**
CLFWS	5.12 (4.26)	4.07 (3.12)	4.50 (4.07)	0.449	0.981	
CLFRS	**4.64 (3.39)**	4.31 (2.25)	**7.02 (3.49)**	0.642	**<.001**	**RI > ASD**
CLFWCT	**6.31 (2.89)**	5.95 (2.34)	**9.24 (2.91)**	0.604	**<.001**	**RI > ASD**
CLFWD	**7.73 (3.17)**	7.94 (3.13)	**10.51 (3.27)**	0.790	**0.001**	**RI > ASD**
Higher Order Language						
CASLAS	77.91 (12.49)	74.00 (7.99)	86.21 (14.13)	0.548	0.035	
CASLIN	**70.10 (14.95)**	74.50 (16.20)	**88.18 (20.12)**	0.490	**<.001**	**RI > ASD**
CASLMC	**71.09 (16.31)**	79.06 (8.73)	**91.04 (12.30)**	0.075	**<.001**	**RI > ASD**
CASLNL	**70.80 (20.99)**	79.07 (15.55)	**88.76 (17.15)**	0.070	**<.001**	**RI > ASD**
CASLPJ	**70.32 (20.25)**	78.94 (14.90)	**88.95 (16.86)**	0.008	**<.001**	**RI > ASD**
Phonological Processes						
C-TOPPEL	**8.14 (3.89)**	**5.33 (3.10)**	7.07 (3.50)	**0.001**	0.189	**ASD > LI**
C-TOPPNR	8.30 (2.72)	6.81 (2.84)	8.02 (3.13)	0.077	0.925	
Written Language						
GRTACC	6.94 (3.77)	6.45 (3.54)	7.91 (3.95)	0.662	0.263	
GRTCOMP	5.62 (3.61)	5.75 (3.11)	6.28 (2.62)	0.896	0.333	
GRTFLU	6.59 (3.57)	5.50 (3.60)	7.35 (4.84)	0.446	0.450	
GRTRQ	77.12 (19.00)	74.35 (15.56)	80.98 (15.31)	0.637	0.320	
GRTRATE	7.35 (3.00)	6.41 (3.14)	7.79 (3.50)	0.286	0.538	
WRMWA	**98.67 (18.83)**	**89.93 (14.04)**	95.05 (15.60)	**0.020**	0.261	**ASD > LI**
WRMWID	**99.12 (20.98)**	**88.12 (18.73)**	91.88 (15.87)	**0.011**	0.051	**ASD > LI**
WRAT	95.78 (17.74)	88.97 (15.86)	93.00 (17.01)	0.125	0.417	
WASI PIQ	93.89 (15.45)	89.81 (12.59)	99.19 (14.84)	0.295	0.078	

Comparisons incorporate blocking for families. Nominal p-values are shown; comparisons significant after a Bonferroni correction are shown in red.

CELF-4 Subtests: CLFCFD = Concepts and Following Directions, CLFWS = Word Structure, CLFRS = Recalling Sentences, CLFWCT=Word Class Total, CLFWD = Word Definitions; CASL Subtests: CASLAS = Ambiguous Sentences, CASLIN = Inferences, CASLMC = Meaning from Context, CASLNL = Non-Literal Language, CASLPJ = Pragmatic Judgment; C-TOPP Subtests: C-TOPPEL = Ellison, C-TOPPNR = Non-word Repetition; GORT 4 Subtests: GRTACC = Accuracy, GRTCOMP = Comprehension, GRTFLU = Fluency, GRTORQ = Reading Quotient, GRTRATE = Rate; WASI PIQ=Performance IQ; WRAT = Spelling; Woodcock Reading Mastery: WRMWA = Word Attack, WRMWID= Word Identification.

= 5.33, nominal p = 0.001). There was still no evidence of a statistical difference on any of the written language measures between the ASD and LI groups.

There was no evidence of statistical differences for RI and ASD on written language measures. However, as with the LI and ASD groups, written language measures were used for impairment classification so that oral language performance was of more interest. There were substantial subtest differences when oral language performance between the ASD and RI groups was examined. The RI group performed better on several segmental, phonological, and higher order language measures. Differences were seen on the following CELF-4 subtests: Formulated Sentences (RI = 9.17, ASD = 5.70, $p < 0.001$), Recalling Sentences (RI = 7.02, ASD = 4.64, $p < 0.001$), Word Classes- Total (RI = 9.24, ASD = 6.31, $p < 0.001$ and Word Definitions (RI = 10.51, ASD = 7.73, p = 0.001). Differences were also seen on the Pragmatic Judgment task of the CASL (RI = 88.95, ASD = 70.32, $p < 0.001$) and the following higher order language tasks: CASL-Inference (RI = 88.18, ASD = 70.10, $p < 0.001$), Meaning from Context (RI = 91.04, ASD = 71.09, $p < 0.001$), and Non-literal Language (RI = 88.76, ASD = 70.80, $p < 0.001$).

The LI and RI groups were not mutually exclusive since there were impaired family members who met criteria for both oral and written language impairments. The RI group was then divided into those family members who met criteria for just written language impairment (RI ONLY, n = 43) and those who met criteria for both oral and written language, (LI+RI, n = 34). When RI ONLY family members were compared to family member with ASD, the RI ONLY group performed significantly better on several segmental language measures and most of the higher order language measures. However, when the ASD probands were compared to those language probands who were affected for LI+RI, results were strikingly different. The ASD group and the

LI+RI groups only differed on the single word reading task (Word Identification) with the ASD group performing significantly better than the LI+RI group (See Appendix 3 for more details).

Since approximately two thirds of the family members with ASD who received the test battery met criteria for LI, we examined those who were affected for ASD with a language impairment (ALI) to see how closely their language performance resembled the family members with LI without autism and whether the result was any different than when the entire ASD cohort was analyzed. After correcting for multiple testing, the two groups did not differ on any measures of segmental, higher order language, or phonological processing (all used to classify them as LI), or on any of the written language measures (see Appendix 3 for more details).

Discussion

The study described in this paper was designed specifically to explore the possible behavioral, etiological, and genetic relations that exist among language behaviors seen in individuals with ASD and those who meet criteria for oral or written language impairment by intentionally ascertaining for ASD and SLI within the same family. In all families, the SLI probands have no evidence of ASD but the ASD probands may or may not present with language impairment but do, by definition, not meet the strict criteria for SLI. As there are subsets of individuals with ASD who have SLI-like communication impairments, and vice versa, it is reasonable to hypothesize that these subsets of individuals may have some genetic variants in common. By creating well-characterized language phenotypes in families where both ASD and language impairment were present, reliable language biomarkers for genetic analyses could be generated while questions concerning shared genetic etiology in language and autism could be further elucidated.

If there is any shared etiology in language and autism, the qualitative nature of these language phenotypes should be similar in both the ASD proband and other affected family members affected for language but not ASD. In fact, few differences were found in any language domains for family members who met criteria for LI when they were compared with verbal family members with ASD. Moreover, as a group the subset of ASD probands with language impairment (ALI) did not significantly differ from the LI group on any segmental, phonological, higher order, or written language measures. Specifically, for the language profiles of family members diagnosed with ASD and those who met criteria for LI, the only task in which they differed was the Elision task of the C-TOPP, which is a phonological task that requires verbal and mental manipulation of sounds. The ASD subjects performed significantly better than the LI subjects. When the family members who were identified as having a RI were compared to those with ASD, there were several significant group differences in oral language measures with the RI group always performing significantly better than the ASD group. The RI group was sub-categorized into two groups: those who met criteria for RI ONLY and those who met criteria for LI+RI. The RI ONLY group differed from the ASD probands on several measures, while the LI+RI group's performance was not different than ASD group in both oral and written language suggesting that the RI ONLY group is distinct from the other probands in the sample. This finding indicates that we might be observing two different types of written language impairments. One type of impairment was associated with a strong oral language base including a history of oral language problems in the past and potentially another type of language impairment associated with a more traditional orthographic base; important information to consider when developing behavioral phenotypes for genetic analyses.

Interestingly, scores on the C-TOPP Non-word Repetition task were not significantly different for our ASD, LI, or RI family members. This assessment has been identified as a marker

that has been strongly associated with SLI on both the behavioral and genetic levels [Bishop et al., 1996, 1999, 2004]. The similarity in scores between the ASD probands and the individuals affected for LI or RI in the NJLAGS sample suggests that this test may be a useful marker for shared etiology between ASD and SLI. These findings provide support for the notion that identifying discrete language phenotypes within families and using probands with different overall clinical diagnoses but similar language patterns, is a reasonable approach to address issues of shared etiology of the two disorders as suggested by [Williams et al., 2008].

Additional support for shared etiology of ASD and SLI is the increased rate of language impairment within the NJLAGS dataset. Since these families were deliberately ascertained for language impairment, rates of LI and RI were naturally higher than those reported in the literature. However, even when the designated probands were excluded from the rates of language impairments, there was an increased rate considerably higher than what would be expected in the general population [Tomblin et al., 1997] but consistent with the rates of SLI in family members of SLI probands [Tallal et al., 2001; Tomblin, 1989]. The increased rates are indicative of an increased loading for language impairment in these families, supporting a role of some shared etiology for ASD and SLI.

Two genetic studies of the NJLAGS families provide additional support for our hypothesis that ASD and SLI may share some genetic etiology. The first study used heritability calculations to examine the familial associations of segmental, higher order language, pragmatic language, and written language constructs within the family sample described in this paper (Bartlett et al., 2012; Chapter 5). They calculated heritability estimates of individual language and communication tasks with and without the ASD probands. All measures included in this study were shown to be heritable ($h^2 > 0$) and Non-word Repetition and Word Identification had

the highest heritability ($h^2 > 0.8$) (see Chapter 5 and Appendix 3; Supplemental Table 3A.4). They reported that when ASD probands were removed from the analysis the heritability increased for the higher order and pragmatic language constructs but not for the individual segmental and written language constructs. This suggests that there is evidence for some shared language abilities between the two classifications of developmental disabilities but also some constructs that are unique to each disorder. This is not surprising as the literature generally reports that pragmatic deficits are near universal in autism but occur close to the population rate in subjects with SLI. The second study identified novel linkage regions using the LI and RI phenotypes described in this study (Bartlett et al., 2013; Chapter 6). In order to investigate the shared etiology for language and ASD, both the language probands and ASD probands were coded as "affected" for phenotype. Evidence for linkage was identified on 15q23-26 (LI) and 16p12 (RI). Taken together, these findings support our hypothesis of some shared etiology between SLI and ASD in a subset of individuals with ASD who display segmental language problems.

Considerations

Although other investigators have reported similar findings to this study [Bishop et al., 2004; Kjelgaard and Tager-Flusberg, 2001], more recent studies [Whitehouse et al., 2008] argue that any similarities are superficial and that careful examination of error patterns suggest that the kinds of errors made for their SLI and ASD subjects were qualitatively different. NJLAGS differs significantly from previous studies that have compared language behaviors among individuals with SLI and those with ASD. It is the first study of entire nuclear families (and in some cases extended family members) where both the SLI and autism probands were identified within the same family. Perhaps the studies that most parallel this study are Lindgren et al., 2009 and Pickles et al., 2012. These two studies compared ASD and SLI probands as well as their

nuclear family members but not within the same families. Lindgren et al., 2009, compared nuclear family members on a series of cognitive, language and reading measures for three family types: 1) those with an ASD proband with no language impairment (ALN), 2) those with an ASD proband with language impairment (ALI), and 3) those with an SLI proband. The profiles of children with ASD with structural language impairments and those with SLI looked quite similar. However, the rates of impairment on most language tasks for family members of the SLI cohort (42%- 60%) were greater than the nuclear family members of the ASD group with language impairment (16%- 35%), suggesting that their findings do not support similar genetic loading for ASD and SLI. The NJLAGS study, on the other hand, found that the rate of language impairment was consistent with the rates of SLI in family members of SLI probands [Tallal et al., 2001; Tomblin, 1989].

Pickles et al., 2012, compared communication behaviors in three family groupings: 1) those with an SLI proband, 2) those with an ASD probands, and 3) those with Down Syndrome (DS). Their study differed from both the current study and the Lindgren et al., study in that questionnaire data, rather than direct testing, were used to evaluate communication status of relatives of the ASD and SLI probands. They conclude that "communication and social deficits breed true in SLI and ASD," meaning that the increased rates of structural language issues in SLI families and increased rates of social language deficits in the ASD families support the idea of distinct etiologies. However, Pickles et al., 2012, found no differences in the communication domains of the SLI only and the SLI+ASD relatives, thus leading them to combine their data for subsequent analyses. Just combining their data for these two groups suggests that the possibility of shared genetics could exist.

As the study design of NJLAGS is unique from previous studies of SLI and ASD, we hypothesize that increasing the genetic loading for language impairments within families will identify any shared variants that contribute to the language phenotypes defined in this study. While the SLI proband of the NJLAGS sample excludes an ASD diagnosis, the ASD proband may present with language and or reading impairments. There are, however, confounding effects that are inherent to family studies such as this one. One confounding effect could be the increased shared genetics by the virtue of being related. We investigated the confounding effect of familiality in pairs of related SLI-ASD probands (N = 55 pairs). A mixed effects linear model was used to regress the absolute differences of assessment scores in a pair against estimate of the proportion of SNPs that identical by descent (see Appendix 3). While the sample size for this analysis may limit our statistical power, familiality did not have a significant effect on the differences in scores for any subtest (Supplemental Table 3A.5). A more problematic confounding factor is the role of shared environment in language acquisition for individuals within the same household. This is an affect that cannot be directly measured given the constraints of this sample and should be included as an important caveat to all analyses.

Another potential confounding effect is the stability of language phenotypes over time. While the initial SLI proband in the NJLAGS sample was identified by more conservative, standardized language measures, other family members were then identified as having oral or written language impairments based on combinations of test scores, subtest scores, and family history (a method which is particularly useful when addressing the history of older family members). A battery of higher order language measures and a detailed family and personal history were included to capture the more subtle and challenging aspects of language where segmental tests are insensitive. Because older children and adults may have learned to compensate for their language-based learning difficulties, identifying them as affected for

genetics studies with traditional standardized measures can be complicated. We hypothesized that taxing the language system with tasks that include language abstractions, inferences, and double meanings would give us a better picture of innate language issues that cannot be compensated for over the years as is the case for more traditional language and reading tests. The ASD probands and the LI family members did not differ on any of the higher order measures, nor did they differ on the measure of pragmatic language on the CASL. While we would like to attribute this lack of significance to the language similarities between the two groups, we realize that difficulties in these tasks may emanate from different causes. One might hypothesize that the ASD probands had difficulty with the comprehension, conceptualization and abstraction of the tasks, while the LI group may have had more difficulty with the verbal short-term memory, verbal expression and formulation skills required for a correct response to the items. Item analysis of responses would be required for this question to be addressed and is beyond the scope of this paper.

What needs to be taken into consideration by our group or by any groups studying the genetics of language and autism is the status of those individuals with ASD who are non-verbal or minimally verbal and are unable to receive the language batteries in our study and similar studies. In this study, 54% of family members with ASD (n=38) were non-verbal or minimally verbal. This group may constitute yet another speech/ language phenotype that has not been extensively explored. So the question still lingers. How are non-verbal individuals with autism categorized when attempting to address questions of etiology or genetics? One might conclude that they are affected for language since their language and reading abilities are untestable. Yet there have been several instances of older, functionally non-verbal children, who found “their voices” through voice and text technology and consequently demonstrating sophisticated language comprehension and reading abilities [Mukhopadhyay, 2013]. Other questions may

then arise and need to be addressed related to the source of the non-verbal status such as seizure activity, regressive disorder, or even the potential of another language phenotype associated specifically with the inability to acquire intelligible spoken language because of a motor speech-related disorder [Flax et al., 2010].

Conclusion

There is agreement that both SLI and ASD are complex disorders with multiple genetic and environmental risk factors, so it is not unreasonable to suggest that individuals with different clinical diagnoses may still share some genetic load for specific language disabilities. This approach fits comfortably into the "common variant-multiple disease" hypothesis addressed in the introduction of this paper [Becker, 2004; Stromswold, 2008] where common alleles that contribute to a disease under certain genetic and environmental conditions may present as completely different disorders under different genetic and environmental conditions. It is clear that ASD and SLI do not have identical genetic etiologies and therefore the focus of research on shared genetics cannot be viewed from an all-or-nothing perspective. As described above, Bartlett et al., 2012, used heritability calculations to examine the familial associations of segmental, higher order language, pragmatic language, and written language constructs within the family sample described in this paper. Their findings suggest that there is evidence for some shared language abilities between the two classifications of developmental disabilities but also some constructs that are unique to each disorder.

Both Stromwold's and Bartlett's views reflect the recent framework proposed by Tomblin, 2011, where the co-morbidity of autism and SLI is dissected. He concludes the paper by stating:

> "They [children with SLI or ASD], like all children, develop from very complex developmental systems. Some of these systems when perturbed in particular ways may tip the developmental trajectories of a child towards ASD or SLI, thus there are likely to be some unique factors associated with each, but also there are likely to be shared or common aetiological factors as well" (p.9).

Taken to another logical level, if language profiles observed in family members with SLI and ASD are consistent with those observed in families with other neurodevelopmental disorders, then perhaps the genetic mechanisms involved in impaired language development may very well supersede the clinical labels of autism and SLI while remaining valuable in studying the genetic architecture of both disorders.

In conclusion, in this sample of families specifically ascertained for both SLI and autism, verbal individuals with autism performed similarly to other nuclear and extended family members who were classified as LI, RI and to a subset of family members with LI+RI. These results lend strong support to the notion that there is, at the very at least, evidence of shared, rather than identical, genetic load for language as characterized in SLI and language in autism. The issue of shared genetic etiology cannot be properly addressed with only behavioral information and requires that behavioral information be combined with family genetic studies as we are beginning to do. Only then can questions of shared etiology of autism and language disorders; two complex disorders be answered.

Chapter 5: Gene x Gene Interaction in Shared Etiology of Autism and Specific Language Impairment?

Christopher W. Bartlett, Liping Hou

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH

Judy F. Flax, Zena Fermano, Abby Hare, Linda M. Brzustowicz

Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ

Stephen A. Petrill

The Battelle Center for Mathematical Medicine (CWB, LH, SAP), The Research Institute at Nationwide Children's Hospital and Department of Pediatrics; Department of Human Development and Family Science

Steven Buyske

Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, NJ

Keywords: Autism, epistasis, gene-gene interaction, heritability, shared etiology, specific language impairment

Published in Biological Psychiatry.

Abstract

Background: To examine the relationship between autism spectrum disorders (ASD) and specific language impairment (SLI), family studies typically take a comparative approach where families with one disease are examined for traits of the other disease. In contrast, the present report is the first study with both disorders required to be present in each family to provide a more direct test of the hypothesis of shared genetic etiology.

Methods: We behaviorally assessed 51 families including at least one person with ASD and at least one person with SLI (without ASD). Pedigree members were tested with 22 standardized measures of language and intelligence. Because these extended families include a nonshared environmental contrast, we calculated heritability, not just familiality, for each measure twice: 1) baseline heritability analysis, compared with; 2) heritability estimates after statistically removing ASD subjects from pedigrees.

Results: Significant increases in heritability on four supra-linguistic measures (including Pragmatic Judgment) and a composite language score but not on any other measures were observed when removing ASD subjects from the analysis, indicating differential genetic effects that are unique to ASD. Nongenetic explanations such as effects ofASDseverity or measurement error or low score variability in ASDsubjects were systematically ruled out, leaving the hypothesis of non-additive genetics effects as the potential source of the heritability change caused by ASD.

Conclusions: Although the data suggest genetic risk factors common to both SLI and ASD, there are effects that seem unique to ASD, possibly caused by nonadditive gene-gene interactions of shared risk loci.

Introduction

Autism spectrum disorders (ASD) are a group of complex developmental disabilities that might include problems with: 1) social interaction, 2) communication, and 3) restricted interests and/or stereotypies. When there is profound impairment in all three areas, the individual is classified with autistic disorder, whereas deficits in one area paired with lesser impairments in two and/or three are considered as part of the autism spectrum. Specific language impairment (SLI) is a neurodevelopmental disorder that is characterized by significant limitations in language abilities occurring in the absence of any other frank neurological disorder or environmental cause. Both disorders share variable limitations of communication, making etiological overlap between the disorders possible. Initial studies indicated similar patterns of impairment in SLI and ASD, including in structural language (semantics and syntax) and phonological short-term memory (PSTM) in a subgroup of children with autism [Kjelgaard and Tager-Flusberg, 2001; Joseph et al., 2002]. Since that initial work, numerous studies have examined what might be grossly considered as three competing hypotheses explored in the following paragraphs: 1) incidental etiological overlap, where overlap is induced by the definitions of the two disorders; 2) familiality of the two disorders but not necessarily due to shared genetics and; 3) shared genetics.

Incidental Etiological Overlap

A recent review proposed that the SLI and ASD might have incidental overlap without meaningful shared etiology [Tomblin, 2011]. This conclusion assumes that structural language—such vocabulary and grammar affected in SLI—and social use of language—such as pragmatics and supralinguistic tasks known to be impaired in ASD—form two quantitative dimensions for language competence. Therefore, in this system, the hallmark deficits of SLI and ASD are on different axes. Persons with SLI (low structural language) will display the full range of social

language ability, and persons with ASD (low social language ability) will display the full range of structural language. Some overlap of SLI and ASD is inevitable in this bivariate system but not productively considered shared etiology, because the two dimensions might be independent. If SLI and ASD were unrelated, the prevalence of one estimated from a sample ascertained for the other would match the population estimate. Two such studies have been reported where families ascertained for SLI were examined for cases of ASD. The prevalence of autism was estimated to be 2.5 and 4.3 times higher than the prevalence in the general population [Conti-Ramsden et al., 2006]. These studies, although limited in sample size, indicate that SLI and ASD are not independent.

Familiality

Most family studies recruit nuclear families ascertained through autism probands and then compare relatives of the autism proband with relatives in control families. Assuming that autism liability might not manifest in relatives as a distinct clinical entity, researchers assess differences in quantitative assessments of autism symptomatology. Quantitative differences relative to control families are defined as the broader autism phenotype (BAP). Several autism domains have well-replicated BAP features, including social and communication behaviors [Bishop et al., 2004b; Losh et al., 2009; Murphy et al., 2000; Ruser et al., 2007] and restricted interests and rigidity [Losh et al., 2008]. Pragmatics, or how language context contributes to meaning, is the most replicated language domain within the BAP where parents of children with autism have lower mean scores than control subjects [Ruser et al., 2007; Whitehouse et al., 2007, 2010]. Because SLI is defined by structural language deficits and not pragmatic deficits, it is not unexpected that, whereas pragmatics is consistently shown to be part of the BAP, results from structural language deficit studies have not generally been supportive of common etiology. Phonological short term memory deficit, a common marker for SLI, has not been consistently

found in ASD families [Whitehouse et al., 2007; Lindgren et al., 2009], regardless of the structural language impairment status of the autism proband (impaired vs. unimpaired). Comparisons of SLI relatives with relatives of typically developing probands or with mental retardation probands yield similar negative findings [Bishop et al., 2004b; Pilowsky et al., 2003]. To date, there is no compelling evidence that ASD and SLI are jointly familial.

Shared Genetic Markers

In 2008, a series of three articles provided converging evidence from independent study designs/methods that implicate *CNTNAP2* in autism susceptibility [Alarcon et al., 2008; Arking et al., 2008; Bakkaloglu et al., 2008]. This striking result has since been followed up in SLI proband families with significant results for association of SLI with *CNTNAP2* [Vernes et al., 2008]. It is not known whether the risk mutations are identical in the two disorders. *CNTNAP2* is also associated with communication ability at age 2 in a general population sample [Whitehouse et al., 2011]. Further studies are needed to fully assess whether alleles within *CNTNAP2* affect pragmatic aspects of language. It is unclear how much of the shared genetic liability between ASD and SLI is accounted for by *CNTNAP2*; these data provide evidence that shared genetic etiology is possible and remains to be fully explained. *CNTNAP2* has also been implicated in other cognitive and neurological disorders, from intellectual disability to stuttering [Peter et al., 2011; Newbury et al., 2011; Gregor et al., 2011; Mefford et al., 2010; Petrin et al., 2010; Zweier et al., 2009; Elia et al., 2010; Friedman et al., 2008], making the ASD–SLI connection harder to disentangle.

Hypothesis

The present study presents a complementary study design that extends the ASD–SLI overlap literature in a novel direction. Our project was designed to address the question of shared genetic etiology in ASD and SLI by directly testing genetic overlap in pedigrees containing

both disorders. We have ascertained nuclear families with both ASD and SLI in mutually exclusive persons and then collected direct measurement of language in all family members, both affected and unaffected, including as many extended family members as possible. The extended family design is useful for quantitative genetic heritability studies, because extended families have relatives that share genetics but not environment (such as cousins), thus providing the key contrast that allows for dissociation of shared environment from additive genetics. We assessed the effect of ASD on heritability estimates by comparing an analysis using all subjects with one where ASD subjects are removed. We hypothesized that measures known to be associated with the language profile in autism, such as supralinguistic skills including pragmatics, would show differential heritability indicating genetic effects unique to autism. We were also specifically interested in nonword repetition and overall structural language ability, hypothesizing that both would be associated with differential heritability potentially indicating genetic effects unique to SLI.

Methods and Materials

Families

Fifty-one families were ascertained through a proband with autism and with the additional requirements that each family have at least one additional family member meeting the study criteria for SLI and no less than five participants (affected and unaffected)/family (mean = 6.9, SD = 2.8, range = 5–20). Families were recruited from the greater New Jersey area for a total of 234 subjects with at least some quantitative language phenotypic data, including 27 persons with ASD, 55 with SLI, and 152 unaffected. Subjects gave informed consent conforming to the guidelines for treatment of human subjects governed by the Institutional Review Board at Rutgers University.

Autism Proband Criteria

To be identified as an autism proband, the following criteria were met: a diagnosis of autistic disorder on at least two of the following three measures: 1) Autism Diagnostic Interview—Revised (ADI-R) [Lord et al., 1994, 1989] score of "autism"; 2) Autism Diagnostic Observation Schedule [Lord et al., 1989; Gotham et al., 2007] score of "autism"; and 3) DSM-IV, autistic disorder.

Specific Language Impairment Proband Criteria

To be identified as an SLI proband, the following inclusionary/exclusionary criteria were met:

1. A core standard score of ≤ 85 on the age-appropriate version of the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4) [Semel et al., 2003; Wiig et al., 2004] and ≤ Performance IQ.
2. Performance IQ ≥ 80 on the Wechsler Abbreviated Scale for Intelligence [Wechsler, 1999].
3. Hearing within normal limits (positive identification of 500 Hz at 30 dB [SPL]; and 1000, 2000, and 4000 Hz at 20 dB [SPL]).
4. No motor impairments or oral structural deviations affecting speech or nonspeech movement of the articulators.
5. No history of autism spectrum disorders or frank neurological disorders such as intellectual disability or brain injury, as determined by parental interview.
6. Native English speaker with English as the primary language spoken at home.

Measures

All SLI probands, non-ASD non-SLI family members, and higher functioning family members with ASD received age appropriate measures of language (means and SD by diagnostic status in Table 5.1). The standardized language battery included:

1. CELF-4 and CELF Preschool [Wiig et al., 2004; Semel et al., 2003]. Core standard scores were derived from 3–5 subtests scaled scores (age-depending) that addressed areas of language comprehension, expression, and structure, including Word Structure, Recalling Sentences, Formulating Sentences, Word Classes (Expressive and Receptive), and Word Definitions.
2. The Comprehensive Assessment of Spoken Language [Carrow-Woolfolk, 1999]. The supralinguistic core of subtests addressed metalinguistic language skills, including abstraction, inference, and a subtest on the pragmatic aspects of language. These areas are of great relevance to older children, adults, and higher-functioning individuals with autism who might be challenged by meaning that cannot be accessed directly through lexical and grammatical information. Subtests include Meaning from Context, Nonliteral Language, Ambiguous Sentences, and Pragmatic Judgment (which has been shown to correlate [Reichow et al., 2008] with the Vineland Adaptive Behavior Scales [Sparrow et al., 1984]).
3. The Comprehensive Test of Phonological Processing (CTOPP) [Wagner et al., 1999]. The Elision subtest was used to measure deletion and phonological manipulation of sounds in words, whereas the Non Word Repetition task

measured phonological short-term memory; both have a strong documented relationship with oral language abilities.

Descriptive data for subjects with ASD capable of taking the language battery, SLI subjects, and all other family members are included in Table 5.1. Tested ASD and SLI subjects only differed on the Elision subtest ($p < 0.05$), where the average of SLI subjects is 2.5 points lower than ASD. On the Elision subtest, ASD subjects did no differ from all other family members. Families had an average 1.1 persons with language impairment and 1.3 with ASD (0.43 for subjects with ASD that had at least some language data). The total sample was 58% male, as is consistent with an increased male risk of ASD and SLI. Average age of all subjects was 30 (range 5–80), with subgroups listed in Table 5.2. Although all ASD subjects were in the child

Table 5. 1: Summary Statistics for Affected and Unaffected Family Members

	ASD Subjects Mean (SD)	SLI Subjects Mean (SD)	Other Subjects Mean (SD)
Number	27[a]	55	152
Age (yrs)	13.9 (9.8)	24.4 (19.9)	32.8 (19.1)
Gender (% Male)	72%	71%	47%
CELF			
Core score	71.5 (26.5)	69.2 (18.2)	107 (12)
Formulating sentences	5.7 (3.8)	5.2 (3.1)	11.3 (2.5)
Repeating sentences	4.7 (3.4)	3.8 (2.3)	9.9 (2.6)
Word classes receptive	7.8 (2.8)	6.9 (2.9)	12 (1.6)
Word classes expressive	5.7 (3.1)	5.5 (2.5)	11.2 (2.5)
Word classes total	6.5 (2.6)	6 (2.4)	11.6 (2)
Word definitions	8.4 (2.4)	8.1 (2.5)	12.9 (2.4)
CASL			
Ambiguous sentences	76.6 (12.3)	74.8 (8.4)	98.3 (14.2)
Meaning from context	71.1 (16.3)	74.3 (11.4)	103.5 (11.1)
Nonliteral language	70.5 (21)	70.2 (15.1)	100 (14.7)
Pragmatic judgment	68.9 (21.2)	71.3 (16.9)	99.2 (10.9)
CTOPP			
Elision	8 (3.9)[b]	5.5 (3.4)[b]	9.6 (2.8)
Nonword repetition	8.6 (2.8)	7.4 (3.1)	9 (2.5)
WASI			
PIQ	93 (16.5)	90.9 (13.9)	108.9 (12.1)

ASD, autism spectrum disorder; CASL, Comprehensive Assessment of Spoken Language; CELF, Clinical Evaluation of Language Fundamentals; CTOPP, Comprehensive Test of Phonological Processing; PIQ, Performance IQ; SLI, specific language impairment; WASI, Wechsler Abbreviated Scale for Intelligence.

[a]There are 64 subjects with ASD in the sample, but only 27 had quantitative language data.

[b]ASD mean is different from SLI mean, $p < .05$ after Bonferroni correction.

generation, when interpreting the average age of SLI subjects, it should be noted that eight were parents, with the remainder in the child generation (bimodal distribution). Although our standard measures used normative samples not exceeding 18 years old, and in some cases 21 years old, there was no evidence of ceiling effects in this sample when administered to ages beyond the normative data.

Statistical Analysis

Additive genetic heritability was estimated with procedures described in detail previously [Logan et al., 2011]. Heritability was calculated by the SOLAR package v4.3.1 (http://solar.txbiomedgenetics.org/download.html) [Almasy and Blangero, 1998] through maximum likelihood procedures with information from the entire pedigree jointly (see Supplement 1; Appendix 4 for more methodological details). The 234 participants that had behavioral data comprised 498 nonmutually exclusive relative pairs, with 133 having shared genetics but no shared environment (Table 5.2), indexing the pedigree complexity relative power of the sample. This sample size is roughly comparable to our previous study of heritability in SLI pedigrees [Logan et al., 2011].

Tests of heritability changes based on the presence of ASD in the pedigree were conducted by comparing the likelihood of two models with the likelihood ratio test (LRT). For the baseline heritability model we tested age, gender, and performance IQ as covariates (kept in model if $p < 0.05$). Baseline analysis used all available pedigree data, which we

Table 5. 2: Relationships Used in Heritability Analysis by SOLAR

Relative Pairs	*n*
Parent-Offspring	261
Sibling	104
Grandparent-Grandchild	24
Avuncular	61
Half Sibling	4
Great Grandparent-Grandchild	2
First Cousin	39
First Cousin, Once Removed	3
Total	498

denote as ASD^{+}. The likelihood of the baseline model was compared with the likelihood of a model that included ASD status as a covariate, which we denote as ASD^{-}. The ASD status was coded as a binary variable, where persons with ASD were coded as 0 and unaffected persons were coded as 1. All pedigree members were coded in this way. This procedure statistically controls subjects for variation caused by ASD status, which has the net effect of removing ASD from the pedigree in a way that is less wasteful of valid information than artificially setting scores from ASD persons to missing values. Changes in heritability between the baseline model and comparison model are informative about the statistical fit of an additive genetics model for ASD. A parallel set of analyses controlling for SLI status was also conducted. We applied a Bonferroni correction for 14 LRTs, requiring a critical *p*-value < 0.0035 to reach overall ($p < 0.05$) significance.

Interpretation of the Model.

The statistical model (see Supplement 1; Appendix 4 for details) assumes only additive effects. Therefore data that are fully inconsistent with an additive model would therefore be expected to show no heritability. Data that are consistent with additive genetics would simply yield the heritability of the trait. In the present study, we tested for the presence of an in-between case. If heritability increases when excluding ASD subjects, this is an indication that subjects with ASD are reducing the additive heritability (i.e., an additive model is not the best fit for the data). To reject the additive genetic model in favor of a nonadditive model, complicating features of the data must be ruled out (outliers, non-normal data, possible mediator variables). Nonadditive models include dominance (where carriers of ASD do not have intermediate phenotypes) and gene–gene interactions, also called epistasis.

Results

Baseline

Results for language measures are presented in Table 5.3. Estimates using the entire pedigree are called ASD^+ and are used as a baseline reference for further comparisons. All measures showed significant evidence for an additive genetic component (h^2 = .24 to .86).

Table 5. 3: Heritability (SE) of Language and Reading Measures With/Without ASD

	ASD^+	ASD^-	Δh^2	Covariates
CELF				
Core score	.43 (.19)	.53 (.18)	.10[a]	Gender, PIQ
Formulating sentences	.24 (.18)	.24 (.19)	.00	Gender, PIQ
Repeating sentences	.62 (.17)	.65 (.16)	.03	PIQ
Word classes expressive	.34 (.17)	.37 (.17)	.03	Gender, PIQ
Word classes receptive	.47 (.20)	.48 (.19)	.01	Gender, PIQ
Word classes total	.36 (.20)	.40 (.19)	.04	Gender, PIQ
Word definitions	.64 (.16)	.65 (.15)	.01	Age, gender, PIQ
CASL				
Ambiguous sentences	.38 (.21)	.46 (.19)	.08[a]	PIQ
Meaning from context	.59 (.17)	.70 (.14)	.11[a]	PIQ
Nonliteral language	.39 (.16)	.45 (.16)	.06[a]	PIQ
Pragmatic judgment	.38 (.18)	.44 (.20)	.06[a]	PIQ
CTOPP				
Elision	.46 (.23)	.48 (.25)	.02	Age, PIQ
Nonword repetition	.86 (.13)	.86 (.13)	.00	Age, gender
WASI				
PIQ	.43 (.20)	.43 (.20)	.00	Gender

Abbreviations as in Table 1.
[a] ASD^+ h^2 different from ASD^-, $p < .05$ after Bonferroni correction.

Removing ASD Subjects

The effect of ASD status was statistically controlled to (operationally) remove all subjects with ASD (ASD^- in Table 5.3). In examination of the distribution of all heritability estimates, ASD^+ versus ASD^-, a paired *t* test was significant [t (13) = -3.94, $p < 0.05$], indicating differences in heritability estimates between conditions; because the changes in heritability cannot be thought of as truly independent observations, this result should be considered suggestive because the *t* test might be anti-conservative. Post hoc LRTs comparing the model likelihoods for ASD^+ versus ASD^- indicated that five heritabilities were significantly different ($p <$

0.05), including all four supralinguistic tasks and the overall language score from the CELF (Figure 5.1). In each case, heritability was higher when controlling for ASD (i.e., $ASD^- > ASD^+$). The largest change occurred in Meaning from Context, which increased 11.5 percentage points. The three other supralinguistic measures also showed significant increases in heritability of 5, 7, and 9 percentage points. There was also a significant increase in heritability for overall language ability as measured by the CELF (10 percentage points) but not on any individual subscales, although semantics (Word Classes Total, 4%) and syntax (Recalling Sentences, 3%) both trended toward significance.

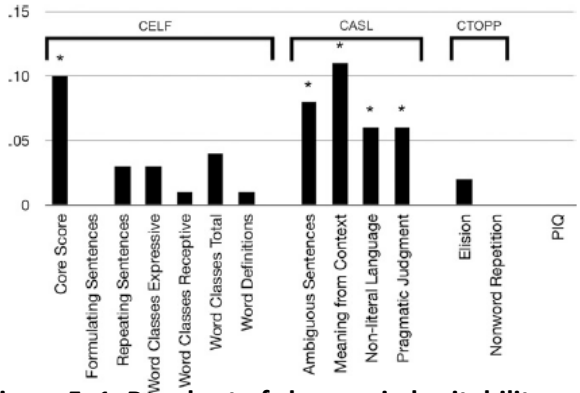


Figure 5. 1: Bar chart of changes in heritability caused by including autism spectrum disorder status as a covariate.

Although the correlation between heritability estimates from the baseline models versus the comparison models was high overall (p = 0.97, *p<0*.05), several measures clearly deviate from the overall trend that are statistically significant (denoted with asterisks). The bar chart demonstrates the signal-to-noise ratio of the results. CASL, Comprehensive Assessment of Spoken Language; CELF, Clinical Evaluation of Language Fundamentals; CTOPP, Comprehensive Test of Phonological Processing

Examination of Moderator Effects Potentially Causing Heritability Changes

If scores from persons with ASD are associated with greater measurement error, then controlling for ASD should reduce the standard error of the heritability estimates (noise reduction). The average of the standard errors for models with significant heritability differences were identical in ASD^+ and ASD^- to 3 decimal places (0.181). This was also true of models that did not show heritability changes. Significant ASD effects on heritability were not accompanied by

significant changes in standard errors for the models. The observed slight fluctuations in standard errors between ASD^+ versus ASD^- models seemed random (sign test, $p = 0.14$).

Although mean scores in the ASD subjects were lower than the corresponding population means, the distribution of scores did not seem to play a role in heritability changes (Figure 5.2). Inclusion of outliers could affect heritability estimation. Although the data from ASD subjects did not contain outliers, as defined by the absolute value of a score >3 SD from the population mean, we did examine whether moderately extreme scores >2 SD had an effect. A total of eight datapoints across all five measures were moderately extreme, according to this definition, and removal of these scores by treating them as missing data did not change the five heritabilities for the ASD^- models. Furthermore, all heritability changes remained significant (LRT, $p < 0.05$).

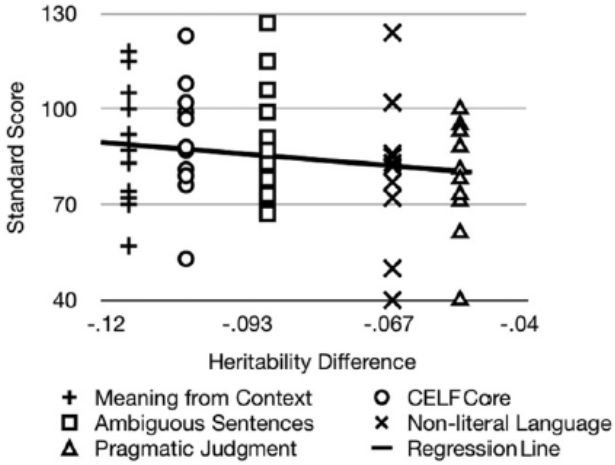


Figure 5. 2: Scatter plot of scores from autism spectrum disorder (ASD) subjects (y-axis) on changes in heritability (x-axis).

Although the average for each measure is below population levels, it is clear that changes in heritability are not driven strictly by universally low performance of ASD subjects. Regression of ASD proband scores on changes in heritability (line in figure) yielded nonsignificant effects of both trait and quantitative value of the heritability difference. CELF, Clinical Evaluation of Language Fundamentals.

We also examined ASD severity as a possible moderator of the heritability changes whereby increasing severity results in less reliable quantitative language scores. For this analysis, we applied Autism Diagnostic Observation Schedule calibrated severity scores for Modules 1–3 as covariates [Gotham et al., 2009]. Because Module 4 does not have calibrated

severity scores, we also incorporated information about that module as an additional binary covariate (not administered = 0, administered = 1) to account for the lack of quantitative severity scores. Estimates of heritability were highly similar when Autism Diagnostic Observation Schedule severity was included as a covariate in the model, failing to support severity of impairment as a moderator of heritability changes.

Removing Language-Impaired Subjects

To examine whether the genetic etiology of ASD was similar to SLI, we compared (baseline) SLI^{+} analysis with SLI^{-} analysis for all traits except the CELF core score, which by definition would be censored, in the analytical sense, because all SLI subjects scored below a fixed threshold, thus removing one tail of the distribution. Note that calculation of heritability on censored data in general pedigrees has not been developed in the literature. When the

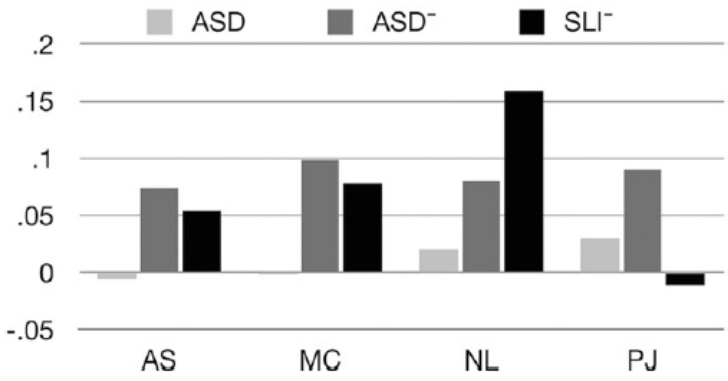


Figure 5. 3: Changes in supralinguistic measure heritability.

Changes in supralinguistic measure heritability, relative to a control condition, where unaffected individuals were randomly removed from the analysis with the same statistical procedure as for autism spectrum disorder (ASD) and specific language impairment. Our baseline analysis, ASD (all subjects), is equivalent to randomly removing individuals from the analysis as shown by the slight deviations from the control baseline (lightest gray). Both ASD^{-} and SLI^{-} show large deviations on the supralinguistic measures, relative to the control condition. The ASD^{-} and SLI^{-} induce similar changes in heritability for both ambiguous sentences (AS) and meaning from context (MC), whereas nonliteral language (NL) also shows changes in heritability, but SLI^{-} is greater than ASD^{-}. A notable difference is in pragmatic judgment (PJ), where SLI^{-} is consistent with the control condition (i.e., no effect of SLI on PJ heritability).

remaining 13 traits are considered, only three of the supralinguistic measures showed significant differential heritability (Ambiguous Sentences, Meaning from Context, Non-Literal Language), whereas Pragmatic Judgment did not (Figure 5.3).

Discussion

When ASD in families ascertained for both ASD and SLI subjects are controlled for, heritability increases are seen for 5 of 14 measures, by as much as 11 percentage points. Changes occurred in the global language measure and all four supralinguistic measures. These results cannot be accounted for by potentially misleading properties of the data such as outliers, reduced variability in ASD subjects, measurement error, or severity of the ASD phenotype. We conclude that the heritability increases caused by exclusion of ASD family members are not consistent with a simple additive-genetics model of heritability. The data, therefore, suggest that nonadditive genetic effects contribute to ASD. The SLI heritabilities also show similar nonadditive effects for three of the supralinguistic measures excluding pragmatics. Taken together, the results provide empirical support for a hypothesis initially articulated and validated through an extensive computer simulation study [Bishop, 2010], which stated that gene–gene interaction can account for molecular genetic findings shared between ASD and SLI while still allowing for phenotypic difference that give rise to the different diagnoses. The present dataset is not large enough to support direct estimation of the additional gene–gene interaction variance component parameters (additive-additive, additive-dominant, dominant-dominant), although this could be accomplished with a larger sample size in future studies.

The lack of nonadditive effects for pragmatics in SLI while present in ASD indicates that some genetic effects are unique to ASD, although only in this one domain of language. Pragmatic language has previously been shown to be heritable in twins ascertained for SLI with the Children's Communications Checklist self/teacher report [Bishop et al., 2006], but this is the

first study to estimate the genetic variance on a wider range of supralinguistic skills through direct quantitative assessment. Such higher order linguistic tasks are associated with the limited ability of individuals with ASD to understand language abstraction and also the well-appreciated pragmatic deficits in ASD. All four supralinguistic scales from the Comprehensive Assessment of Spoken Language showed significant heritability, although inclusion of ASD individuals in the analysis significantly diminishes that heritability, implying that some differing etiologies are influencing the performance on these tasks in ASD versus non-ASD family members. Although diminished, the heritabilities for these traits are still substantial when considering all family members (ASD and non-ASD), implying that there are also genetic loci in common across all family members that contribute to these abilities. It is interesting to note that, in our sample, mean scores from persons with ASD are not significantly different from those of SLI subjects, and both groups are lower than population means. Reduced supralinguistic skills are not a defining characteristic of SLI, so this observation is unexpected and novel relative to the SLI and ASD published data. This is the first study to ascertain families with both disorders, and it is possible that, because supralinguistic skills are heritable, selecting a family into the study where at least one person has ASD (and thus is expected to have supralinguistic deficits) essentially selects for risk loci that might produce such deficits in family members without ASD. Under those selection demands, the subjects with SLI might be at greatest risk for reduced performance.

The Core standard score of the CELF is less heritable when including ASD subjects in the analysis, whereas only a few individual subscales displayed that trend, and none were significant. It is possible that the nature of the composite CELF Core standard score, which weights data from multiple subscales, has greater variability and/or reliability than the individual subscales. If true, then the observed heritability results would be expected, because the

subscales have less reliability and/or variability for the analysis. Yet without strong results from the subscales, the standard score on the CELF is indeed quite broadly representative of language skills, making interpretation challenging. Further studies with greater sample sizes will be needed to identify which aspects of structural language are key to the observed decreased heritability.

We hypothesized that heritability differences would be observed in PSTM, because this is a cognitive domain where ASD and SLI have notably divergent presentations in the types of errors that tend to manifest with low performance [Riches et al., 2011; Whitehouse et al., 2008]. Furthermore, PSTM deficits are quite common in SLI but occur only in a subgroup of subjects with ASD. Our data indicate that the same genetics that influence PSTM in subjects with SLI (and the rest of the pedigree) also influence PSTM in ASD, leaving the heritability estimates essentially the same. The PSTM performance was previously shown to differentiate ASD and SLI—not in overall performance as scored in the CTOPP but only when errors by syllable length were considered [Riches et al., 2011; Whitehouse et al., 2008]. It is possible that the quantitative score studied here is simply not suitable for detection of such qualitative differences in performance, making genetics that are unique to ASD possible.

Overall, the presented heritability estimates are consistent with the literature with one qualified exception. Our nonword repetition heritability estimate from the CTOPP is higher than commonly reported for that measure [Raskind et al., 2000; Logan et al., 2011]. However, the Children's Test of Nonword Repetition [Gathercole et al., 1994], a similar measure of PSTM, was estimated to have heritability >1.0 in twins with SLI [Bishop et al., 1996]. This estimate is higher than genetically possible, due to the nature of their chosen statistical formulation, but does indicate genetic effects much larger than 0 in SLI subjects. It is therefore possible that our CTOPP nonword repetition heritability is reasonable for pedigrees selected for SLI and autism.

Several aspects of statistical modeling should be considered when interpreting the data. The nature of pedigrees precludes holding age constant across all subjects. And although we applied age as a covariate to capture age effects, it is still possible that age effects, such as changes in additive genetics throughout the lifespan remain. Recently, we applied these pedigree methods to a collection of SLI pedigrees without ASD and found the results to be quite comparable to the literature on twin children, presumably due to generally modest sizes of age-related changes in heritability [Logan et al., 2011]. Here too, the heritability estimates are very similar to previous literature on twin children.

Heritability of the ASD diagnosis as a categorical trait shows wide variability, with high heritability [Taniai et al., 2008; Rosenberg et al., 2009; Lichtenstein et al., 2010] generally >80 in previous studies, whereas one recent study indicates a more modest effect of genetics [Hallmayer et al., 2011]. Additionally, ASD symptoms show considerable variation, even when considering the differences in study designs and sample sizes [Lichtenstein et al., 2010; Hallmayer, 1999; Robinson et al., 2012, 2011]. In contrast, our quantitative phenotype heritability results are consistent with the language genetics literature, which generally show much less variability [Stromswold, 2001] than has been shown in ASD [Lichtenstein et al., 2010; Hallmayer, 1999; Robinson et al., 2012, 2011]. The magnitude of the heritability changes observed in our study should therefore be interpreted relative to the stability of the language genetics literature rather than the variability associated with a categorical diagnosis of autism. Due to the limited number of persons with ASD in our pedigrees (many families have only 1) as a consequence of our ascertainment protocol, analysis of ASD as a categorical trait is not powerful enough to inform the ongoing debate about the heritability of the categorical ASD diagnosis. Lastly, although epistasis seems to be a likely explanation for how ASD differs from SLI despite shared genetics, it was not possible to formally estimate the necessary additional parameters

with the current sample size. Additional studies with much larger samples are necessary to test an epistasis model for ASD and SLI.

Chapter 6: A Genome-scan for Loci Shared by Autism Spectrum Disorder and Language Impairment

Christopher W. Bartlett, Liping Hou, Soo Yeon Cheong

The Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and Department of Pediatrics, The Ohio State University, Columbus, OH

Judy F. Flax, Abby Hare, Zena Fermano, Marco A. Azaro, Linda M. Brzustowicz

Department of Genetics and the Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ

Barbie Zimmerman-Bier

Department of Pediatrics, Saint Peter's University Hospital, New Brunswick, NJ

Charles Cartwright

Department of Psychiatry, University of Medicine and Dentistry of New Jersey – New Jersey

Medical School, Newark, NJ

Steven Buyske

Department of Statistics and Biostatistics, Rutgers University, Rutgers University, Piscataway, NJ

Keywords: genome scan, gene mapping, complex traits, genomewide association, linkage analysis

In Press: American Journal of Psychiatry, 2013

Abstract

Objective: The authors conducted the first genetic linkage study of families that segregate both autism and specific language impairment (SLI) to find common communication impairment loci. The hypothesis was that these families have a high genetic loading for impairments in language ability, thus influencing the language and communication deficits of the subjects with autism. Comprehensive behavioral phenotyping of the families also enabled linkage analysis of quantitative measures, including normal, subclinical and disordered variation on all family members for the three general autism domains, social, communication and compulsive behaviors.

Method: The primary linkage analysis coded persons with either autism or SLI as "affected" with language impairment. The secondary linkage analysis consisted of quantitative metrics of autism-associated behaviors capturing normal to clinically severe variation, measured in all family members.

Results: Linkage to language phenotypes was established at two novel loci including 15q23-26 and 16p12. The secondary analysis of normal and disordered quantitative variation in social and compulsive behaviors established linkage to two loci for social behaviors (on chromosomes 14q and 15q) and one locus for repetitive behaviors (on 13q).

Conclusion: These data indicate shared etiology of autism and SLI at two novel loci. Additionally, non-language phenotypes based on social aloofness and rigid personality traits showed compelling evidence for linkage in this sample. Further mapping is warranted to elucidate these loci.

Introduction

Autism is a severe neurodevelopmental disorder characterized by altered functioning in three domains: 1) social interaction, 2) communication, and 3) stereotyped behavior and/or restricted interests and activities. Separately, there is a well-known classification in language disorders for children who have difficulties in acquiring language but are otherwise neurologically and psychologically normal, known as specific language impairment or SLI. Given that both disorders have a language component to their diagnosis, previous work has suggested that SLI and autism could have shared genetic contributors. This hypothesis is supported by a series of genetic mapping studies examining the relationship between autism and language impairment in complementary ways [Bartlett et al., 2002, 2004; Bradford et al., 2001; Buxbaum et al., 2001; Shao et al., 2002; Simmons et al., 2010; Spence et al., 2006; Lord et al., 1994; Alarcón et al., 2005, 2002; Alarcon et al., 2008; Warburton et al., 2000].

Autism linkage mapping studies have examined the relationship of language and autism with two paradigms. The first paradigm used language delay status of autism spectrum disorder (ASD) probands to stratify families into two groups, most often based on presence/absence of phrase speech by 36 months of age. Stratification on phrase speech delay postulates that there are subgroups of autism that can be genetically differentiated by language status. The language stratification paradigm has yielded autism findings on chromosomes 2q24-32 [Buxbaum et al., 2001; Shao et al., 2002], 7q22-32 [Bradford et al., 2001] and 13q21-22 [Bradford et al., 2001]; this 13q region is also linked to SLI in non-ASD samples [Bartlett et al., 2004, 2002; Simmons et al., 2010]. At all three locations, stratifying ASD families on phrase speech delay nearly perfectly separated families that were linked to the given locus versus families that were not, in an *a priori* fashion. Some loci were linked only in families including a person with autism and phrase speech delay and other loci were only linked in families without phrase speech delay. In

contrast, a replication study used both stratification of families and determination of affection status in non-ASD parents based on self-report of language impairment history in childhood [Spence et al., 2006]. This study found that the coding of parents as affected based on a history of language intervention or language delay minimally impacted the results, and while the group with phrase speech delay did have increased evidence of linkage, the chromosomes 2, 7 and 13 findings were not replicated.

The second paradigm used language phenotypes to directly map language quantitative trait loci or language impairment as a dichotomous trait within families with autism. Genomewide language quantitative trait locus mapping in ASD was performed using two items from the Autism Diagnostic Interview- Revised (ADI-R) [Lord et al., 1994], "age in months of onset of first word" and "age in months of onset of phrase speech"; both showed linkage to 7q36 [Alarcón et al., 2005, 2002]. *CNTNAP2,* a gene also associated with both SLI and normal language development, was later shown to be the most likely candidate responsible for this linkage [Alarcon et al., 2008], thus providing evidence that genetic variation relevant to both ASD and SLI can occur within the same gene.

This is the first linkage/association study to present a novel paradigm that complements both stratification and quantitative trait locus approaches, described above, in order to better understand the relevance of language variation in ASD. We selected families for the presence of both autism and SLI under the hypothesis that if autism and SLI are etiologically related then this sampling scheme will enrich our sample for loci related to language difficulties in autism and also reduce the genetic heterogeneity of ASD. Within these families, we collected state-of–the-art language phenotypes on relatives, and whenever possible, individuals with ASD. This study

thus represents the most comprehensive family-based language phenotyping in a molecular genetic study of ASD.

We tested the hypothesis of genetic overlap by performing our analysis assuming that ASD and language impairments have the same underlying genetic etiology (i.e., we considered phenotypes from both disorders as equally affected) and then performed genome-wide linkage scans. Our analyses coded individuals with oral or written language impairments without ASD, as well as individuals with similar types of oral and written language impairments and ASD, as "affected." Additionally, persons with ASD who could not be evaluated using quantitative language measures were incorporated into the analysis using a method for censored data (i.e., systematically missing data), which in this case assumed the censoring was due to low language ability [Hou et al., 2012] (see also Supplemental Methods; Appendix 5). Our study design is most effective for mapping loci that are etiologically relevant to both ASD and language impairment. If ASD and language impairment are genetically unrelated in these families, then coding both disorders as "affected" in the same analysis will reduce evidence for linkage/association. Given this, positive findings were formally hypothesis-tested to determine if autism or SLI or both jointly contributed to detected linkage signals.

Further, the phenotypic battery included seventeen quantitative population-normed language assessments, and quantitative measures of social and compulsive behaviors. As a secondary objective, we performed linkage analysis using these quantitative data to capture both normal variation and clinically severe variation on the same scale, an analysis generally considered to be more powerful than analysis of only affected/unaffected. These analyses included the first use of the Yale-Brown Obsessive Compulsive Scale in a molecular genetic study of autism.

Methods

Overview of Design

Our primary goal was to find genetic variation relevant to both language impairments and ASD, using a set of previously described families recruited for the presence of persons with autism and *separate individuals* with SLI [Bartlett et al., 2012]. To accomplish our goal, we created subgroups from 79 families (Supplementary Table 5A.1; Appendix 5) according to phenotypic characteristics. We established three groups of families that we denoted Tier I-III. Tier I (N=46) consisted of families with both an autism proband and a different proband with SLI, as defined in this sample previously [Bartlett et al., 2012], or in a few cases, one autism proband and one ASD proband with low language often called "autism language impaired" in the literature [Lindgren et al., 2009; De Fosse et al., 2004]. Autism language impaired individuals are contrasted with ASD probands who are language normal ("autism language normal"), and ASD probands who are nonverbal ("autism nonverbal"), with this last category often ignored in the literature since language cannot be quantitatively assessed.

Tier I had an internal contrast since any observed linkage could be further examined by excluding either the SLI proband or the autism proband to understand the connection between SLI and autism to the linkage signal on a locus-by-locus basis. The other two tiers included autism families that did not have an SLI or autism language impaired proband after direct testing. Tier II (N=15) consisted of multiplex ASD families without an SLI proband where neither ASD proband were operationally defined as language impaired (neither normal structural language deficits nor nonverbal). Tier III contained families (N=9) with an ASD proband and at least one relative who scored in the impaired range on either the Social Responsiveness Scale, a well-studied inventory of social functioning [Constantino and Todd, 2005; Constantino et al., 2003], or the Yale-Brown Obsessive Compulsiveness Scale, used to evaluate OCD in psychiatric

evaluations [Goodman et al., 1989b, 1989a; Scahill et al., 1997, 2006]. For association analysis, an additional nine autism trios (N=9) were added to Tier III.

Prior to behavioral testing all subjects gave informed consent conforming to the guidelines for treatment of human subjects at Rutgers University. All family members as well as higher functioning family members with ASD received age appropriate measures of language and reading (Supplementary Table 5A.2; Appendix 5). Descriptive statistics by diagnostic group, SLI, autism and other, have been previously published [Bartlett et al., 2012] (see Chapters 4 and 5). Observed correlations between measures are in the Supplement (Supplementary Table 5A.3; Appendix 5).

For the purpose of categorical phenotype linkage/association analysis, we define *oral language impairment*, called "LI" in our previous papers [Bartlett et al., 2004, 2002; Simmons et al., 2010], as either an age appropriate Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4) core standard score of ≤ 85, or at least 1 SD below peers on ≥ 60% of all oral language subtest scores and a significant history of language/reading difficulties, defined as >2 years of intervention and/or childhood diagnosis of language and/or reading impairment. For purpose of finding loci that jointly influence oral language impairment and autism, we define "LI*" as a phenotype that includes as affected persons affected with our definition of language impairment as well as persons affected with ASD (i.e., etiological equivalence).

In our previous studies of multiplex language impairment families, we observed many instances of semi-compensated adults with a childhood diagnosis of language problems and currently presenting with weak language skills who did not meet the cut-off for language impairment but did meet the cut-off for reading impairment [Bartlett et al., 2004, 2002; Simmons et al., 2010]. Based on our prior successful mapping of an SLI locus with that reading

impairment phenotype [Bartlett et al., 2004, 2002; Simmons et al., 2010], we defined *written language (reading) impairment*, called "RI" in our previous publications [Bartlett et al., 2004, 2002; Simmons et al., 2010; Bartlett et al., 2012], as >=1 SD below the population mean on 60% of all reading tests and subtests. For purpose of finding shared loci that jointly influence written language impairment and autism, we define "RI*" as a phenotype that includes as affected persons affected for our definition of reading impairment as well as persons affected with ASD (i.e., etiological equivalence). Throughout this paper, LI* and RI* refer to our specific diagnostic definitions of language impairment and/or autism and reading impairment and/or autism, respectively, while the term language impairments is meant in a more general sense to apply to oral and/or written language impairments, in context.

Genotyping

Affymetrix Axiom 1.0 arrays were used to generate 567,893 SNP genotypes on 440 individuals from the 79 families. Quality control on SNP genotypes was conducted as described previously [Simmons et al., 2010], with additional details included in the Supplementary Methods, based on individual/SNP genotype completion, relationship checking, Mendelian errors and ancestry. A subset of 8086 SNPs was chosen for linkage analysis to minimize marker-to-marker LD and retain high minor allele frequency to provide suitable genomic coverage of recombination events in the pedigrees. Association analysis used all SNPs that met quality control standards and had a minor allele frequency > 0.05, yielding 529,874 SNPs. Validation genotyping was conducted on a Luminex 200 machine using a custom oligonucleotide ligation assay [Bruse et al., 2008], with allele calling and quality control as described elsewhere [Simmons et al., 2010; Hou et al., 2011].

Statistical Analysis

Overall data analysis plan. We first conducted genomewide linkage scans with follow-up association analysis in the linkage regions. We also conducted a genomewide association analysis over the remainder of the genome. Given the depth of phenotyping, it was not considered reasonable to perform univariate analyses of all 21 cognitive measures on a genomewide basis due to the difficulty of interpreting results from analyses of many correlated traits. Instead, we opted for a mix of empirically and theoretically driven phenotypes. We used two categorical phenotypes, LI* for oral language impairment and RI* for written language (reading) impairment, the latter being a strong indicator of an unresolved oral language impairment in multiplex SLI pedigrees (reading deficits caused by an underlying language deficit) based on our previous studies [Bartlett et al., 2004, 2002; Simmons et al., 2010]. We also derived three quantitative traits using a factor analysis (see Supplemental Methods; Appendix 5) to reduce the phenotypic data, which we called factors one through three (F1, F2, and F3; factor loadings in Supplementary Table 5A.4; Appendix 5). To elucidate possible shared etiology in language phenotype linked regions, follow-up analyses were conducted to assess if ASD or language impairment or both were required to detect the linkage peak. The non-language traits were analyzed as quantitative traits, which included the Yale-Brown Obsessive Compulsiveness and the Social Responsiveness Scale (SRS-QT). The Social Responsiveness Scale was also analyzed as a dichotomous trait (SRS-DT) using a mild impairment threshold (see Appendix 5).

Linkage/association analysis methods. Linkage and association analyses were conducted with the KELVIN 2.3.3 package (http://kelvin.mathmed.org/;(26)). KELVIN implements the posterior probability of linkage (PPL) metric to measure the probability that a genetic location is linked with the trait of interest and the combined posterior probability of linkage disequilibrium (cPPLD) metric to measure the probability that a single nucleotide polymorphism

(SNP) is in linkage disequilibrium (LD) with the trait of interest conditional upon the evidence for linkage at a given locus. It is important to note that the PPLD uses a pedigree likelihood that explicitly accounts for family structure while assessing the evidence that LD is present between the SNP and the disease [Huang, 2011; Yang et al., 2005]. For quantitative trait analysis where some ASD subjects lacked data due to inability to participate in some cognitive tests, the measures for those individuals were treated as censored data, meaning the "true scores" were unknown but known to be below a threshold. These analyses were handled using the PPL model for censored data [Hou et al., 2012] as described in the Supplemental Methods (Appendix 5). The sex-averaged marker map for linkage was obtained from the Rutgers Combined Linkage-Physical Map of The Human Genome [Matise et al., 2007].

Primary linkage analysis of language phenotypes was conducted on each of the three tiers separately and the linkage evidence was sequentially updated across the three tiers to provide a single metric for linkage evidence. Follow-up family-based association analysis was conducted similarly; however, families that contained persons that were not of European ancestry (N=1 in Tier III) were dropped from the association analysis, since combining samples with different genetic ancestries can generate false positive results [Hou et al., 2011]. For non-language phenotypes, all families were run as a single dataset.

Statistical Correction for Multiple Phenotypes. In order to assess the effect of performing multiple genome-scans using correlated phenotypic traits, we simulated 3000 genomes without regard to phenotypes to create an empirical null distribution for estimating p-values. We simulated chromosomes using the same SNP allele frequency and genetic distances as our linkage dataset. For each simulated genome, we conducted analysis with the five phenotypes, saving the overall maximum PPL per replicated genome. This list of maximum PPLs

is the null distribution accounting for analysis with our five phenotypes, using our specific pedigree configuration and patterns of missing data etc. After correcting for multiple phenotypes, a PPL of 3.47% or greater retains a genome-wide error rate consistent with $p < 0.001$, a PPL of 2.69% corresponds to $p < 0.01$, and a PPL of 1.04% to $p < 0.05$. Note these threshold values are slightly higher than two previous studies of the false positive rate of the PPL [Bartlett et al., 2002; Logue et al., 2003] due to correcting for multiple phenotypes in the present study.

Results

Initial Linkage Results for Language. We performed genomewide linkage analysis using two categorical definitions of language impairment (affected/unaffected status) and three quantitative language scores from a factor analysis of all 21 language measures, which we denote F1, F2 and F3 (see Appendix 5). The first trait, abbreviated LI*, defined individuals as

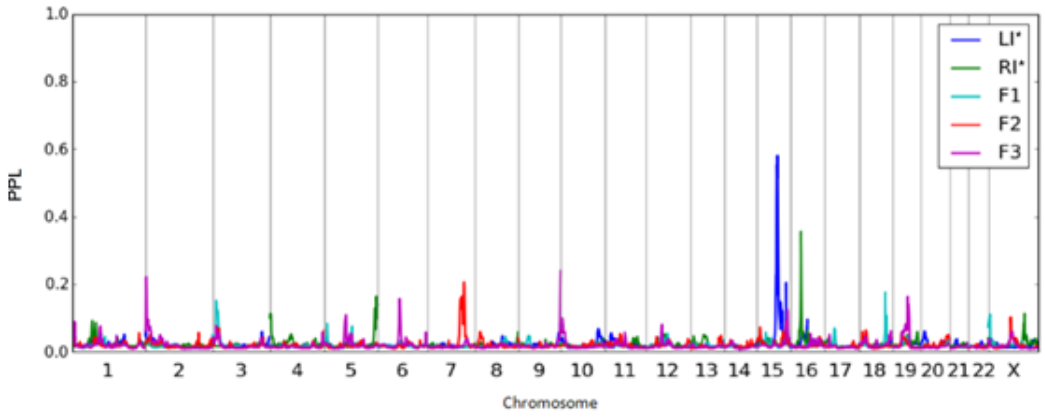


Figure 6. 1: Genomewide linkage analysis of the 5 language-related traits.

The posterior probability of linkage (PPL) is scaled such that PPL values < 2% represent evidence against linkage to that location while values > 2% represent evidence for linkage to that location. A PPL value of exactly 2% indicates that the data are not informative for linkage. The peaks on chromosomes 15 and 16, which represent oral language impairment and/or ASD (LI*) and written language (reading) impairment and/or ASD (RI*), respectively, clearly stand out from the rest of the genome, and overall the PPL displays a high signal to noise ratio for linkage mapping. While the three factor scores derived from 21 standardized measures of language (F1, F2 and F3) lack strong peaks, several regions of potential interest are identified.

affected if they possessed either an oral language impairment or an ASD. The second trait, abbreviated RI*, defined individuals as affected if they possessed either a written language impairment or an ASD. LI* produced clear evidence of linkage to chromosome 15 and RI* to chromosome 16, as shown in Figure 6.1 (with a summary of all large linkage peaks given in Table 6.1). The posterior probability of linkage (PPL) is scaled such that genomewide plots show very clear signal-to-noise ratios as seen in the figure. The magnitudes of both signals (> 35%) have Type I error rates appropriate for establishing linkage in a genome-wide scan, even after accounting for testing of multiple phenotypes (see Methods). Table 6.1 also includes the fully maximized LOD score (or MOD score) to allow comparison of the PPL to a statistic with a more commonly used scale; the estimated disease gene frequency and risk probabilities by genotype (penetrances) for the linkage peak are also included in the table including the estimate of the proportion of families linked to a given locus (α).

Table 6. 1: Linkage Peaks with PPL > 0.35 from All Analyses.

						------Maximizing Model------					
Phenotype	Chr	cM	PPL	Band range	Width (Mb)	LOD	-/-	+/-	+/+	DGF	α
YBOCS	13	58	0.36	q14.3-21.3	17.3	4.2	-3	0	3	0.3	0.7
SRS-DT	14	117	0.37	q32.2-32.23	7.6	3.5	0	0.1	0.8	0.1	1
LI*	15	83	0.57	q23-26.2	24.2	4.1	0	0.7	0.99	0.001	1
SRS-QT	15	120	0.52	q26.2-26.3	6.2	4.5	-2	1	2	0.1	0.9
RI*	16	43	0.36	p12.1-12.3	8.9	4.6	0	0	0.9	0.2	1

PPL is the posterior probability of linkage. LOD is the fully maximized LOD score, sometimes referred to as a MOD score. -/-, +/-, and +/+ are the estimated genotypic effects for the locus; for categorical analysis these quantities are penetrances and for quantitative traits they are genotypic means on a z-score scale. DGF is the disease gene frequency and α is the heterogeneity parameter in the admixture likelihood at the maximizing model.

Phenotypes are: YOBCS=Yale-Brown Obsessive Compulsive Scale, SRS-DT=Social Responsiveness Scale- Dichotomous trait, LI*=Oral Language Impairment and/or ASD, SRS-QT=Social Responsiveness Scale-Quantitative Trait, RI*= Written Language (Reading) Impairment and/or ASD.

Chromosome 15q23-26.2 was linked to LI* with a maximum PPL of 57% and implicating a region of 24.1 Mb from 73 cM to 106 cM. As can be seen in Figure 6.2A, the linked region has

a 15 cM high confidence linkage region with a much larger low confidence region (to the right) that accounts for about half of the total implicated region. Non-verbal IQ was not linked at this locus, indicating a dissociation of language and intelligence at this locus, as expected, since SLI does not include deficits in intelligence. We next wanted to assess how the level of language impairment in the ASD subjects (autism language impaired, autism language normal, and autism nonverbal) in each family modulates the linkage signal. We therefore defined a metric to quantify the relative contribution of the three language levels, which indicated homogeneous contributions from all three groups to the linkage at this peak (see *Assessing the relative contribution of the three proband types to the final PPL* in Appendix 5).

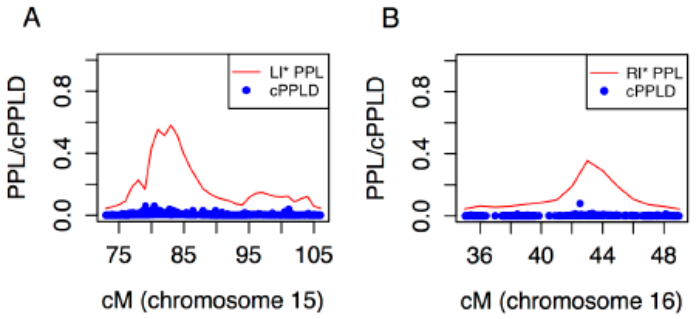


Figure 6. 2: Follow-up association analysis of SNPs under the language-related linkage peaks for oral language impairment and/or ASD, LI* (A) and written language (reading) impairment and/or ASD, RI* (B) phenotypes.

Family-based association analysis of all available SNPs with minor allele frequency > 0.05 in the linked region yielded only weak evidence of association, with a maximum combined posterior probability of linkage disequilibrium (cPPLD) of 6%. These results are not consistent with strong evidence for association that would account for the observed linkage. However, the Axiom array SNPs in the region successfully haplotype tag only 48% of the common variation (as described in the Supplemental Methods with results in Supplementary Table 5A.5; Appendix 5); thus, follow-up of the region based on this SNP genotyping platform should be considered incomplete.

Chromosome 16p12.1-12.3 was linked to RI* over 8.9 Mb from 35 cM to 49 cM, with a maximum PPL of 36% (Figure 6.2B). All three ASD language levels (autism language impaired, autism language normal, and autism nonverbal) contributed equally to the PPL (see *Assessing the relative contribution of the three proband types to the final PPL* in the Supplement). Similar to the chromosome 15 results, analysis of non-verbal IQ as a quantitative trait yielded evidence against linkage at this locus (PPL = 1.7%, or below the prior probability of linkage, which is 2%), indicating a dissociation of language and intelligence at this locus as well. Follow-up cPPLD analysis to find SNPs that accounted for the linkage signal yielded a maximum cPPLD of 8% at a single SNP (next highest cPPLD = 1.7%). Genotyped SNPs in this region only successfully tagged 55% of the common variation (Supplementary Table 5A.5; Appendix 5).

Additional language-related peaks of interest (PPL > 20%) were observed on chromosome 7 with F2 (the second trait defined from the factor analysis of all 21 language tests) and chromosomes 2 and 9 with F3 (the third trait defined from our factor analysis). The linkage to chromosome 7 (PPL = 21%), located over the region containing *CNTNAP2*, has been replicated in both ASD and SLI. This PPL may be considered appropriately large to replicate the *CNTNAP2* locus.

Further Characterizing the Role of Language and Autism at Linked Loci. We sought to define the relative contribution of autism versus SLI to the linkage results on 15q and 16p, i.e., to assess if each disorder contributes equally to those linkage findings as an indication of the equivalence of the disorders. We restricted these analyses to Tier I since only this subset of families contains both autism and SLI probands in each pedigree. Each locus was assessed separately. We assessed the specificity of chromosome 15 for language versus autism by removing the autism proband from each pedigree and repeating the linkage analysis, then doing

a separate and equivalent analysis removing the SLI proband while retaining the ASD proband. In both cases, the linkage signal was greatly reduced (PPLs dropped to 2% and 4%). Similarly, on chromosome 16 we removed the autism proband from each pedigree yielding a PPL of 3%, and removed one non-ASD RI* subject from the linkage analysis, giving a PPL of 6%. Therefore, both the 15q and 16p loci are sensitive to the presence of both autism and SLI. However, the reduction in PPL signals could have been the result of lower power from including fewer affected individuals in the analysis or due to loss of specific and relevant disease information. To rule out the former, we then used a permutation study to assess if removing language-impaired and/or ASD subjects from the analysis induced a greater average drop in the PPL than removing subjects randomly (see also *Assessing the relative contributions of language impairment and ASD within families to the final PPL* in the Supplement). For chromosome 15, the permutation test was significant for an effect of SLI and ASD having more dramatic effects on the PPL than other combinations ($p < 0.01$). This was not the case for chromosome 16, where the test was not significant, indicating that low power cannot be ruled out as a confound when interpreting the contribution of reading impairment and autism to this linkage peak.

To test whether our phenotypic definitions of LI* and RI* were too restricted, we also repeated the linkage analysis on chromosomes 15 and 16 using a combined phenotypic definition of both oral and written language impairment, where persons were defined as affected if they were either LI* or RI* (or both). For both chromosomes, the PPL was attenuated to less than half the original linkage signal (dropped to 19% and 12%) and the linkage region was greatly broadened (data not shown).

Linkage Analysis of Non-Language Phenotypes. All families were ascertained only for language phenotypes in the non-ASD individuals, not for any non-language characteristics of the

broad autism phenotype. However, we quantitatively assessed social responsiveness and aspects of obsessive compulsiveness in all pedigree members able to participate in an assessment and used the phenotypes for linkage analysis. Results are summarized in Figure 6.3. We analyzed the Yale–Brown Obsessive Compulsive Scale (YBOCS) as a quantitative trait yielding

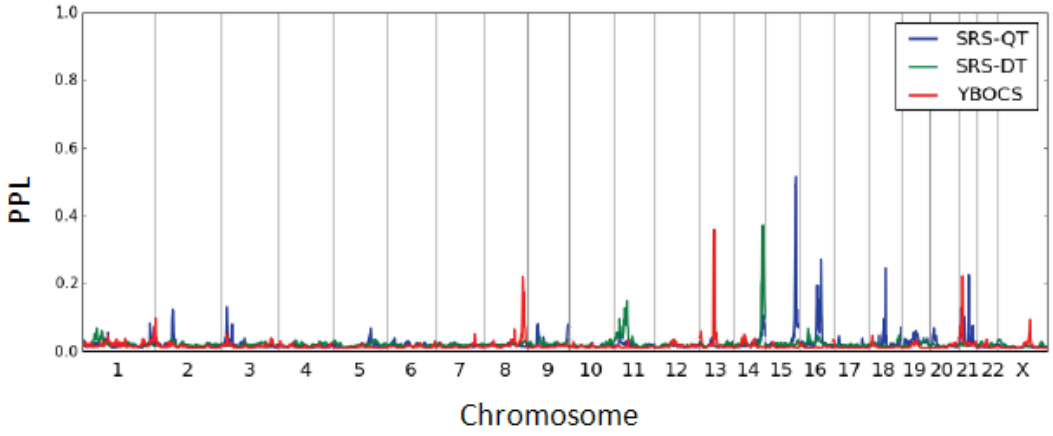


Figure 6. 3: Genomewide linkage analysis of the 3 non-language-related traits.

The largest signals with each of the three traits are on chromosomes 13, 14 and 15. No overlap was observed between the social assessment scale, Social Responsiveness Scale (SRS) for both quantitative (SRS-QT) and dichotomous (SRS-DT) trait analyses.

a 36% PPL on 13q14.3-21.33 (55cM to 63 cM) with the peak directly over *PCDH20*, close to previous linkage studies of ASD and studies of SLI [Bartlett et al., 2004, 2002; Bradford et al., 2001]. The linked region is 17.3 Mb in size. Analysis of all available SNPs with minor allele frequency > 0.05 did not yield any cPPLD results over 2% (Figure 6.4). Two smaller peaks of interest occurred on chromosomes 8 (PPL = 22%) and 21 (PPL = 22%) with no cPPLDs under those peaks greater than 3%.

Our analysis of the Social Responsiveness Scale was conducted two ways. The first was as a quantitative trait (SRS-QT), using the numerical scores generated by the test. This yielded a linkage peak PPL of 52% on chromosome 15q26.2-26.3 (113cM to 133 cM) that does not overlap with the LI* peak described above. We observed three SNPs spanning 3 Mb (rs12440787,

rs7170868, rs9672677) with cPPLD = 7% under this peak, but none are in linkage disequilibrium with each other and so they do not represent a single coherent signal. In addition, we observed two smaller peaks on chromosomes 16 (PPL = 27%) and 18 (PPL = 25%). The second was using the "mild" impairment threshold (see Appendix 5) to dichotomize the Social Responsiveness Scale for analysis (SRS-DT). We chose to create this simple mild impairment v. unimpaired distinction since this distinction is commonly used in the broader autism phenotype literature. We saw the peak PPL of 37% on chromosome 14q32.2-32.33 (110-126cM) encompassing 7.7 Mb. No cPPLD from available SNPs in these regions was greater than 2%.

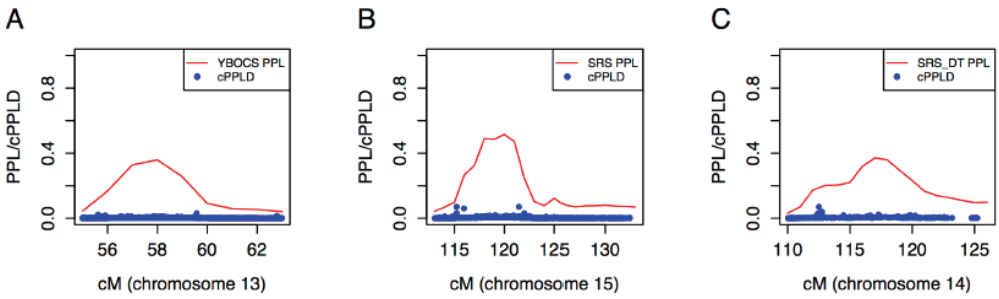


Figure 6. 4: Follow-up analysis of non-language-related linkage peaks.

No cPPLDs were observed that could account for the linkage signals with the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) (A) or the Social Responsiveness Scale (SRS) as a quantitative trait (B) or dichotomous trait (C).

Genomewide Association Analysis. Linkage analysis only requires a limited number of SNPs to attain essentially full information and thus utilized only 8,086 of the 529,874 SNPs that passed quality control filtering. We conducted association analyses on the 529,874 genomewide SNPs, yielding 19 SNPs with PPLD > 10% (see Supplementary Table 5A.6 for SNP names; Appendix 5). We then performed follow-up genotyping, using a different platform, of these SNPs or SNPs in strong LD ($r^2 > 0.95$) for validation, and cPPLD analysis using both linkage information from the families as well as LD with the trait. The highest cPPLD of 20% was located at rs3792495 for the F2 trait.

Discussion

We identified two linkage peaks for language impairment in families with both language impairment and autism (15q25.1 and 16p12.3) that do not overlap with previously discovered autism or language impairment loci. The two linkage signals showed specificity for oral language impairments for 15q and for written language impairment for 16p. This specificity was evidenced by attenuation of the linkage signals when a combined oral/written language impairment phenotype was applied suggesting there is a subset of individuals with reading problems that are not comorbid with oral language deficits. Additionally, there was no evidence that either locus is primarily related to SLI alone or autism alone, rather, it appears each locus is jointly related to both SLI and ASD. These findings are in keeping with the goal of the study to find genetic variation that is relevant to both disorders. This is the first molecular genetic study of families that segregate both SLI and autism, and we hypothesize that these families have a high genetic loading for impairments in language ability, further influencing the language and communication deficits of the autism probands. In our previous studies (See Chapters 4 and 5; [Flax et al., 2013; Bartlett et al., 2012]), we also observed that the average scores on a standardized test of the social use of language (pragmatics) were similar in the SLI and ASD subjects in these families. Pragmatic impairment is not part of the defined deficits in SLI but is commonly seen in ASD, suggesting that the two disorders, as presenting in the families selected by our ascertainment criteria and recruitment methods, may be on an etiological continuum.

There were also several compelling linkage peaks for non-language traits even though the sample ascertainment scheme did not include any requirements of non-language traits beyond the autism proband. The Social Responsiveness Scale peak on 15q26.3 was the second largest in the study (PPL = 52%) and had the narrowest linkage region (6.2 Mb). This region was implicated in a meta-analysis of ASD and schizophrenia [Chagnon, 2006]. An additional Social

Responsiveness Scale peak was noted when using a mild cut-off to create a categorical affection status (14q32; PPL = 37%). The 14q32 region has been associated with autism through cytogenetic abnormalities and copy number variation [Qiao et al., 2013; Merritt et al., 2005]. While social skills and communication are fundamentally related, it is unclear from our data if ascertainment for performance language assessments increased power to detect social behavior [Bolte et al., 2008; Constantino et al., 2004; Kamio et al., 2012; Wigham et al., 2012] due to that relationship. A conservative interpretation is that the deep phenotyping performed here was simply more likely to find multiple strong effects across phenotypic domains relative to other studies with less phenotypic data. The Social Responsiveness Scale is a good quantitative metric for mapping autism loci as it has yielded strong findings in other studies [Duvall et al., 2007; Coon et al., 2010]. However, this was the first gene mapping study to also use a mild cut-off with the Social Responsiveness Scale, which is more analogous to the broader autism phenotype literature where affected/unaffected distinctions are commonly applied. We also believe this is the first use of the YBOCS as a quantitative trait in genome-wide analysis of ASD. The empirical performance appears quite good based on our data, suggesting that wider use of this measure may be warranted in ASD research, especially in family genetic studies where specific behaviors may be apparent in some family members but not severe enough to impair activities of daily-living. Our specific finding on 13q does not coincide with OCD studies but does align with previous autism genetic studies [Bradford et al., 2001; CLSA et al., 1999; Steele et al., 2011].

While the linkage analysis showed several strong peaks with different language and non-language related traits, there were no strong association signals either under the peaks or across the remaining genome. As mentioned in the results section, the regions under the linkage peaks are not adequately tagged for comprehensive association analysis by the Axiom 1.0 array, as is the case for much of the genome, thus greatly decreasing the chance of

observing associations. The modest sample size also limits power to detect common variants of small effect. Additionally, recent evidence suggests autism has significant allelic heterogeneity, and association analyses have been generally not replicated except for variants with very small effects [Anney et al., 2012]. Rare variant studies indicate widespread heterogeneity [Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012; Iossifov et al., 2012], though it remains unclear what proportions of genetic mechanisms for ASD involve rare, infrequent and common variants.

The issue of allelic heterogeneity *across disorders* is still uncharacterized from both theoretical and empirical points of view but quite relevant to the debate on genetic overlap between disorders. Until reasonably inferred functional alleles are associated in each disorder, it is not possible to directly address the issue of whether the same genetic variants within the same gene are relevant to both SLI and ASD [Chapman et al., 2011; Alarcon et al., 2008]. Hence it is possible that SLI and ASD have some of the same key genes in pathogenesis but not the same underlying variants or molecular mechanisms. However, if functional variants are the same for both disorders, then it remains to be explained why some members of the family develop SLI and not ASD. It could be that individuals manifesting SLI have a smaller genetic load for such variants, or that SLI is a truly dissociable sub-component in at least some forms of ASD. We will continue susceptibility allele mapping in our autism-SLI pedigrees to untangle these complicated mechanisms.

Chapter 7: Fine Mapping and Association Analysis of Candidate Genes for Autism Spectrum Disorder and Language Impairment in the NJLAGS Sample

Abby Hare, Ariane Seto, Judy Flax, Marco Azaro, Linda Brzustowicz

Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ

Steven Buyske

Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, NJ

Christopher Bartlett

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH

Keywords: Autism, SLI, PPLD|L, Ingenuity Pathway Analysis, Candidate Gene, Association

Abstract:

Over the past decade, the New Jersey Language and Autism Study (NJLAGS) has collected detailed behavioral and genetic information on families that contain one individual with autism and another individual with Specific Language Impairment (SLI) but not autism. This is the first study of its kind to investigate the share genetics between autism and SLI. Using a comprehensive neuropsychological testing battery, three categorical phenotypes: language impairment (LI), reading impairment (RI), and social impairment (SRS-DT), and two quantitative phenotypes: social impairment (SRS-QT) and obsessive-compulsive behaviors (YBOCS) were developed. Autism proband scores were included in the quantitative phenotypes and all categorical phenotypes included autism diagnosis as impaired for language, reading, or social skills, respectively. A previous study identified linkage in these families to 13q21.2 (YBOCS), 14q32.31 (SRS-QT), 15q25.1 (LI), 15q26.2 (SRS-DT), and 16p12.3 (RI). As genome-wide association did not reveal strong evidence for association, Ingenuity Pathway Analysis (IPA) was used to select candidate genes for fine mapping analysis. Four groups were analyzed in IPA: communication impairment (LI + RI), social impairment (SRS-DT + SRS-QT), restricted/repetitive behaviors (YBOCS), and an overall autism model (LI + RI + SRS-DT + SRS-QT + YBOCS). Genes in each analysis group were given a score that corresponded to the number of relevant functions identified by IPA. Seven of the highest-ranking genes were selected for association analysis from the LI, RI, and SRS-DT linkage regions: *AKT1, JAG2, PTPN9A, SEMA7A, NTRK3, FES,* and *SCCN1B.* Each gene was fine mapped using an oligonucleotide ligation assay and was analyzed for association using the KELVIN framework. Each gene was analyzed for association to its respective phenotype with and without the inclusion of autism diagnosis and for autism diagnosis only. *JAG2* yielded the strongest evidence for association (PPLD|L = 5%) when autism

diagnosis was included. When autism diagnosis was excluded, *NTRK3* yielded the strongest evidence for association (PPLD|L =7%). For autism diagnosis only, no evidence for association was detected. The lack of strong evidence for association provides support for a role of rare variants in autism susceptibility and their role in the shared genetics between autism and SLI.

Introduction:

Autism spectrum disorders (ASD) are a group of complex neurodevelopmental disorders that are characterized by three domains: 1) communication impairments, 2) social impairments and 3) the presence of restricted and repetitive behaviors. The characteristics of each domain have a wide range of severity. An individual with severe impairment in all three domains is diagnosed with Autistic Disorder, however in individuals with a broader definition of autism each domain can range in severity independently. While 20-40% of individuals with ASD fail to develop functional verbal language [Bailey et al., 1996; Lord et al., 2004], verbal individuals with ASD can have difficulty with pragmatic language, structural language, and phonological short-term memory (PSTM) [Kjelgaard and Tager-Flusberg, 2001; Lewis et al., 2007; Joseph et al., 2002; Rapin et al., 2009]. The structural language and phonological impairments displayed in subgroups of ASD are qualitatively similar to those seen in Specific Language Impairment (SLI). SLI is a neurodevelopmental disorder that is characterized by delayed language development in the absence of intellectual, sensory, or other neurological abnormalities. Like ASD, impairments in SLI can range in severity. While SLI and ASD are distinct disorders, overlapping characteristics of each disorder has led some to hypothesize that there is an etiological overlap in subgroups of individuals with these disorders. Other groups speculate that the qualitative overlap of characteristics in both disorders is superficial and the fundamental causes of these characteristics are distinct [Whitehouse et al., 2008; Williams et al., 2008].

The definitions of ASD and SLI draw distinct lines between the two disorders. The communication impairment seen in ASD is primarily pragmatic and supralinguistic, while SLI is predominantly structural. If ASD and SLI are truly separate and non-overlapping then impairments in structural language and pragmatic difficulties would not be displayed in both disorders, which is not the case. The structural language impairments that are hallmarks of SLI

have also been reported in several ASD cohorts [Kjelgaard and Tager-Flusberg, 2001; Lewis et al., 2007; Rapin et al., 2009]. Additionally, a study conducted by Leyfer et al., 2008, demonstrated that a significant portion of children with SLI meet the criteria for social and communication impairments using ASD diagnostic tools. This overlap also exists in adolescents and young adults [Howlin et al., 2000; Mawhood et al., 2000; Conti-Ramsden et al., 2006]. Overall, the prevalence of ASD characteristics in individuals with SLI was found to be 3.9%, which is 10 times what is expected from the general population [Conti-Ramsden et al., 2006]. The increased prevalence in overlapping characteristics suggests that the two disorders are not completely independent.

In addition to clinical evidence for shared etiology, several linkage studies of autism exploring the relationship between language and autism have yielded findings on 2q24-32 [Buxbaum et al., 2001; Shao et al., 2002], 7q22-32, and 13q21-22 [Bradford et al., 2001]. These studies used the presence/absence of phrase speech delay (PSD) > 36 months to stratify ASD cohorts based on language delay in order to genetically differentiate between subgroups of autism. All of these studies were successfully in separating linked families from unlinked families *a priori*. The finding on 13q21-22 has not been replicated in other ASD studies, but was also linked to SLI [Bartlett et al., 2002]. In addition to these regions, molecular genetic studies have implicated *CNTNAP2* in autism susceptibility and language impairment [Alarcon et al., 2008; Arking et al., 2008; Bakkaloglu et al., 2008; Vernes et al., 2008; Whitehouse et al., 2011; Bradford et al., 2001]. *CNTNAP2*, Contactin Associated Protein-Like 2, is located on 7q35 and encodes a member of the neurexin family that functions in nervous system development and the differentiation of axons. CNTNAP2 is down-regulated by FOXP2, which is a transcription factor that has been implicated in speech and language disorders [Vernes et al., 2008]. In 2008, two studies identified linkage and association to *CNTNAP2* in two independent ASD samples

[Alarcon et al., 2008; Arking et al., 2008] . A third study identified an inversion on chromosome 7q that disrupts *CNTNAP2* in a child with cognitive and social delay [Bakkaloglu et al., 2008]. After subsequent resequencing of this gene in a sample individuals with ASD, 27 nonsynonomous changes were identified, 8 of which are predicted to have a deleterious effect. Association to *CNTNAP2* was also identified in a sample ascertained for SLI [Vernes et al., 2008] and to communication ability at age 2 in the general population [Whitehouse et al., 2011].

The genetic and clinical overlap of ASD and SLI led to the unique study design of the New Jersey Language and Autism Genetics Study (NJLAGS). Over the past decade, the NJLAGS project has collected families that contain at least one individual with autism and at least one other individual with SLI. This project was designed to investigate the role of shared genetics in ASD and SLI by testing the genetic overlap in pedigrees that were ascertained for both disorders within the same families. Family members were genotyped using the Affymetrix Axiom 1.0 array and were given a comprehensive testing battery that measures language and reading function, social ability, and obsessive-compulsive behaviors [Bartlett et al., 2013], Chapter 6). Two language and reading impairment phenotypes and three secondary social impairment and obsessive-compulsive behavior phenotypes were developed for linkage and association analysis from this data. Genome-wide linkage analysis identified evidence for linkage on 13q21.2 (obsessive-compulsive behaviors), 14q32.31 (social impairment), 15q25.1 (language impairment), 15q26.2 (social impairment) and 16p12.3 (reading impairment). In the analyses for language and reading impairments, all family members who met criteria for these impairments including individuals with ASD were coded as affected for these traits. For the case of language and reading impairments, if ASD and language and/or impairment are not related, the inclusion of the ASD proband will reduce the linkage signal. While these linkage regions do

not overlap with current findings for language impairment on Chromosome 7q, the strong evidence for linkage supports the hypothesis that ASD and SLI have some shared etiology.

Genome-wide association analysis was also conducted using these phenotypes. No compelling evidence for association was detected for any phenotype. The lack of evidence for association could be due to several reasons. First, the association analysis was conducted using all quality controlled genotype information from the Affymetrix Axiom 1.0 array which only effectively tags half of common haplotypes in our linkage regions. It is likely that the array density was not sufficient to capture subtle allelic association in these regions. A second possibility is that there are multiple causative variants within these regions that contribute to ASD/SLI susceptibility. Alternatively, there may be a contribution of rare variants that can only be detected through deep sequencing to ASD/SLI susceptibility.

This study aims to investigate the possibility that the genetic map was not of sufficient density to capture evidence for association. Seven candidate genes were selected from our linkage regions using Ingenuity Pathway Analysis. These genes were fine mapped and analyzed for association to our defined phenotypes. A lack of evidence for association in these genes would lend support to the possibility of multiple and/or rare causative variants.

Materials and Methods:

Participants

This study included the 78 families (375 individuals) from the NJLAGS cohort that were included in our preliminary linkage analysis [Bartlett et al., 2013], Chapter 6). All families in the NJLAGS study contained at least one individual with Autistic Disorder and at least one other individual who met criteria for SLI. All families were recruited from the New Jersey area and

were required to have at least 5 participating family members with English as their primary language. The 78 families were subdivided into 3 Tiers. Tier I (n = 53 families; 269 individuals) met the strictest NJLAGS criteria of containing one autism proband and at least one additional family member meeting the study criteria for SLI. Tier II (n = 7 families; 29 individuals) did not have a strict SLI proband, but had at least two individuals diagnosed with an ASD. Tier III (n = 11 families; 53 individuals) had an ASD proband and at least one individual with a social impairment or OCD. The remaining 7 families were trios with the following proband breakdown: one ASD and one SLI proband (n = 1 family), one ASD proband and one individual with a social impairment (n = 1 family), and ASD only (n = 5 families).

Phenotypes

Five phenotypes were derived from the NJLAGS testing battery (See Appendix 3; Supplemental Figure 3A.1) for use in linkage and association analyses ([Bartlett et al., 2013], Chapter 6). For the phenotypes defined below, a significant history of language/reading difficulties was defined as >2 years of intervention and/or childhood diagnosis of language and/or reading impairment.

1) Language Impairment (LI) was defined as an age appropriate CELF-4 core standard score ≤ 85, or at least 1 SD below the standard score on ≥ 60% of all language subtest scores with a significant history of language/reading difficulties. The language subtests included age appropriate CELF-4 subtests, CASL subtests (Ambiguous Sentences, Inferences, Nonliteral Language, Pragmatic Judgment, and Meaning from Context), and CTOPP subtests (Elision and Nonword Repetition).

2) Reading Impairment (RI) was defined as a score of at least one SD below the standard mean on at least 60% of the WRMT subtests (Word Identification and Word Attack), the GORT-4 comprehension standard score, and GORT-4 overall reading quotient.

3) Social Impairment – Dichotomous (SRS-DT) was defined using the raw scores on the Social Responsiveness Scale. Both the original SRS and a modified adult version of the SRS were analyzed. Raw scores were used to define affectedness since the adult version was not standardized at the time of assessment. The cut-off for a dichotomous social deficit was 54 for males and 45 for females. These raw scores are equivalent to the T-score >60 criteria used to identify mild-moderate impairment in children.

4) Social Impairment – Quantitative (SRS-QT) was a quantitative trait based on the raw scores from the SRS. Both the child and adult versions of the SRS were included.

5) Obsessive/Compulsive Behaviors (YBOCS) was defined based on the percent of maximum possible score for the compulsions and obsessions scales of the adult, child and PDD versions of the Yale-Brown Obsessive Compulsive Scale (YBOCS).

Each of these phenotypes were analyzed with (+ASD) and without (-ASD) the ASD proband included as affected. Association analysis was also conducted using autism diagnosis only as the affection status for all markers.

Selection of Candidate Genes for Fine Mapping Analysis

Genes within the linkage regions were identified using the UCSC Genome Browser (NCBI B37 assembly). A total of 587 genes reside within the linkage regions identified in Bartlett et al., 2013 (Chapter 6). Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, www.ingenuity.com) was used to categorize and prioritize candidate genes identified in our linkage regions by their

function. In order to model the domains of autism, the five linkage regions were combined into four groups (Table 7.1): Communication Impairment (LI+RI), Social Impairment (SRS-QT+SRS-DT), Restricted/Repetitive Behaviors (YBOCS), and Autism (LI+RI+SRS-QT+SRS-DT+YBOCS). Each

Table 7. 1: Division of Linkage Regions into Domains

Domain	Phenotype	Linkage Regions	Number of Genes
Communication Impairment	LI + RI	15q25.1, 16p12.3	354
Social Impairment	SRS	14q32.31, 15q26.2	213
Restricted/Repetitive Behaviors	YBOCS	13q21.2	20
Autism	LI + RI + SRS + YBOCS	13q21.2, 14q32.31, 15q25.1, 15q26.2 , 16p12.3	587

group was analyzed using the Core Analysis of IPA, which utilizes a right-tailed Fischer's exact test to identify the most common functions and pathways in a list of molecules. Genes that were implicated in functions and canonical pathways involved in 'Nervous System Development and Function', 'Neurological Disorders', and 'Psychological Disorders' were given a ranking score corresponding to number of relevant functions and pathways identified in IPA. Each analysis group was ranked separately and each domain group was combined with the Autism group score for a final ranked value. The Autism list was supplemented with functional information from the DAVID Functional Cluster Analysis [Huang et al., 2009a, 2009b] and pathway information from KEGG Pathway Analysis [Kanehisa and Goto, 2000; Kanehisa et al., 2012]. Genes with a ranking score above 5 were further prioritized by supporting literature findings for a total of 7 candidate genes (Table 7.2).

Table 7. 2: Prioritized List of Genes Selected for Fine Mapping.

GENE	CHROMOSOME	SIZE (bp)	tag SNPs	Phenotype	Position (cM)	PPL
AKT1	14	36,393	9	SRS	121-122	13.91-16.57
JAG2	14	37,085	8	SRS	121-122	13.91-16.57
PTPN9	15	122,165	7	LI	78	22.28
FES	15	21,297	5	LI	96	11.92
SEMA7A	15	33,953	16	LI	77-78	18.11-22.28
NTRK3	15	389,673	72	LI	92-93	6.68-8.47
SCNN1B	16	89,029	34	RI	47	4.45

All positions and tag SNPs were obtained from NCBI Build 37.

SNP Selection and Genotyping

Tag SNPs were selected using HapMap (HapMap Data Rel 24 Phase 1 & 2 – full dataset, NCBI B36 assembly, dbSNP b126) using the CEU population with a minor allele frequency (MAF) cuttoff ≥ 0.05 and $r^2 < 0.80$ for each gene region including 10kb upstream and downstream flanking regions. Tag SNPs were supplemented with markers from SNPbrowser (version 4.0, Applied Biosystems) and were validated using dbSNP [Sherry et al., 2001]. A total of 189 SNPs were selected for fine mapping analysis.

Multiplex PCR design templates were prepared for each SNP by extracting 500 bp of flanking upstream and downstream sequences from the current human genome build (hg19, Feb. 2009) and substituting all known neighboring SNPs obtained from dbSNP (Sherry et al., 2001; dbSNP Build ID: 131) with their respective IUB codes. Candidate primers were extracted from these flanks based on several criteria: (1) they did not overlap with neighboring SNPs; (2) they lacked simple tandem and inverted repeat elements; (3) their predicted melting temperatures fell within a 2°C range; and (4) they yielded a minimal number of secondary alignments when subjected to local BLAST analysis [The 1000 Genomes Project Consortium,

2010]. These filtered candidates were used to generate multiplex PCR primer sets for panels that ranged in size from 14 to 36 SNPs. For each multiplex panel, primer compatibility was enforced by requiring that no two PCR primers shared 3'-to-3'-end homology greater than 2 bp and that no given PCR primer had greater than 6 bp of homology between its 3' terminus and the internal region of any other primer. Amplicon sizes ranged from 98 bp to 434 bp.

SNPs were genotyped using the oligonucleotide ligation detection assay described in Bruse et al., 2008, that has been optimized to work in the 384 microtitre format. The 30 µL PCR reaction was scaled down to 20 µL to accommodate the smaller well volume of the 384 microtitre plate. The smaller reaction mixture contained 40 ng of genomic DNA, 1U AmpliTaq Gold® DNA polymerase, 10x reaction buffer (provided with AmpliTaq Gold®), 2.5 mM MgCl (Applied Biosystems), 200 µM dNTPs, and 3.33 pmol of each PCR primer, though thermocycling conditions remain unchanged. The OLA reaction volume was scaled up to 20 µL from 15 µL to allow for the use of a Perkin Elmer – Janus® Automated Workstation, using the original thermocycling conditions [Bruse et al., 2008]. The increased OLA reaction mixture contains 2 µL of the PCR product, 3 U of Taq DNA Ligase (New England Biolabs), 10 nM of each OLA probe, and a 10X reaction buffer (provided with the Taq DNA Ligase). The bead hybridization and fluorescent labeling steps remain as described in Bruse et al., 2008.

In preparation for association analysis, the genotype data were cleaned for missingness by marker and by individual. Using cut-offs of >5% missing marker data and >15% missing individual data, no SNPs or individuals were removed and the total assay failure rate was 3.8%. PedCheck [O'Connell and Weeks, 1998] was used to identify Mendelian errors and Merlin [Abecasis et al., 2002] was used to identify unlikely double recombination events yielding an overall error rate of 0.45%. While any SNP showing an error was removed for the entire family,

no families were eliminated due to excessive Mendelian or recombination errors. Following this cleaning procedure, there was a total missingness rate of 4.3%. It should be noted that the overall missingness rate for *JAG2* was 9.75%, which may skew association analysis (Table 7.3). Three SNPs were removed from analysis due to departures from Hardy Weinberg Equilibrium (p-value < 0.01).

Table 7. 3: Error Rate and Percentage of Missingness of Candidate Gene Fine Mapping

Gene	AKT1	JAG2	PTPN9	SEMA7a	FES	SCNN1b	NTRK3	Total
Total # tag SNPs	9	8	7	16	5	34	110	189
Total # SNPs Analyzed	5	5	6	14	5	30	102	167
Error Rate (%)	0.49	0.66	0.46	0.82	0.22	1.16	0.19	0.45
Failure Rate (%)	3.89	9.10	2.60	3.58	2.19	3.91	3.76	3.85
Missingness (%)	4.38	9.75	3.06	4.40	2.41	5.07	3.95	4.30

Statistical Analysis

All analyses were conducted using the software package KELVIN V2.4.0. Evidence for linkage was calculated using the PPL [Huang et al., 2006; Vieland, 2006, 1998]; the PPLD was used to calculate evidence for association. Genetic distances were obtained from the Rutgers genetic map (http://compgen.rutgers.edu/mapopmat; release 10/09/06; [Matise et al., 2007]). The PPL is a Bayesian statistical measure of linkage that is designed to accumulate evidence for or against linkage for complex traits. The PPL is parameterized in terms of a general approximating likelihood and utilizes a dichotomous trait model with an admixture parameter (α) representing the proportion of 'linked' pedigrees, the disease allele frequency (p), and the penetrance vector (f_i). The parameters of the trait model are integrated out of the algorithm, making this method of linkage analysis essentially model-free. The likelihood also contains the recombination fraction (θ) and the standardized LD parameter (D') as two location parameters. The PPL incorporates a prior probability of linkage of 2% [Vieland, 1998], therefore, PPL values

greater than 2% indicate evidence in favor of linkage, while PPL values less than 2% indicate evidence against linkage for that particular locus. As this measure can be interpreted directly as a probability, corrections for multiple testing are not required. Furthermore, the PPL is a measure of evidence without inherent significance levels. A PPL of 40%, therefore, can be interpreted as evidence that there is a 40% probability of a trait locus at the given marker [Elston and Lange, 1975].

As all markers selected reside within an already identified region of linkage, evidence for association was calculated using Posterior Probability of Linkage Disequilibrium Given Linkage (PPLD|L). This function uses a prior probability of linkage of 100%. As the highest PPL value identified in our linkage scan was 57%, the PPLD|L provides an inflated measurement of evidence for association. Therefore, any PPLD|L value >2% was scaled using the Combined Posterior Probability of Linkage Disequilibrium (cPPLD). This function uses the PPL values identified in the original linkage analysis as the prior probability of linkage [Huang, 2011].

AKT1 and *JAG2* were analyzed using the SRS-DT phenotype. *FES*, *NTRK3, PTPN9,* and *SEMA7A* were analyzed using the LI phenotype. *SCCN1B* was analyzed using the RI phenotype. All regions were analyzed using with (+ASD) and without (-ASD) the ASD diagnosis included in the affection status. All regions were also analyzed for ASD diagnosis only. Three analyses were conducted for each set: Tier I only, Tier I+II, Tier I+II+III to assess the effect of proband type.

Assessment of Sample Power

To assess the power of the sample to detect association, KELVIN was run using genotypes simulated by SLINK [Huang et al., 2007; Vieland and Huang, 2003]. Haplotypes that included one SNP disease marker and a SNP in LD with the disease marker (D' = 0.80) were generated for the NJLAGS dataset. The maximized models for each linkage region were used as

the genetic models for SLINK (Table 7.4). The percentage of unlinked families was varied from 0% to 100% in increments of 10% to model the heterogeneity. One hundred replicates were generated under each set of parameters and each was analyzed for association using the KELVIN framework described above. All replicates were conducted using estimations of the prevalence of each phenotype in the NJLAGS dataset. The LI+ASD and SRS-DT+ASD prevalence was 17% and the RI+ASD was 20%. As the LI and RI linkage analyses were conducted for Tier I only, the LI+ASD and RI+ASD phenotypes were analyzed for Tier I and Tier I+II+III. SRS-DT+ASD phenotype was analyzed for Tier I+II+III in the linkage analysis and all simulations.

Table 7. 4: Maximized linkage models for LI, RI, SRS-DT

Phenotype	Chromosome	Position (cM)	PPL (%)	alpha	DGF	P(DD)	P(Dd)	P(dd)
LI	15	83	55	1	0.001	0.999	0.7	0
LI	15	78	22	1	0.8	0.999	0.1	0
LI	15	88	12	1	0.1	0.999	0.7	0
RI	16	43	37	1	0.8	0.9	0	0
SRS-DT	14	117	37	1	0.1	0.8	0.2	0
SRS-DT	14	118	36	1	0.1	0.999	0.1	0
SRS-DT	14	120	23	1	0.3	0.4	0	0

DGF: Disease Gene Frequency

Results:

Selection of Candidate Genes

IPA was used to identify genes within our linkage regions that have functions that may be related to autism and language impairment. IPA identified 75 relevant functions and 13 canonical pathways for the genes in Communication Impairment group. For the Social Impairment group, 21 functions and 31 canonical pathways were identified. The Restricted/Repetitive Behavior group identified 6 functions. When all of these groups were combined into the overall

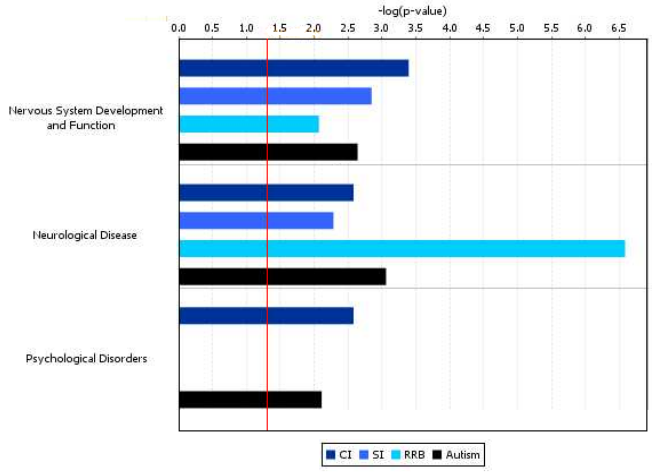


Figure 7. 1: Comparison of IPA Core Analyses of all three domains to the Autism model

Table 7. 5: Relevant Functions and Canonical Pathways Identified by IPA Core Analysis

Domain	Functions	Canonical Pathways
Communication Impairment	Specification of neurons, extension of axons, propagation of mitosis of neurons	Axonal Guidance Signaling, Calcium Signaling, mTOR Signaling
Social Impairment	Size of dendritic spines, abnormal morphology of nervous system, quantity of oligodendrocyte precursor cells	Neuregulin Signaling, Reelin Signaling in Neurons, mTOR signaling
Restricted/Repetitive Behaviors	Formation of dendrites, degeneration of Purkinje cells	-
Autism Model	Quantity of neurons, specification of neurons	Axonal Guidance Signaling, Neuregulin Signaling, nNOS Signaling in Neurons

Autism Model, 67 functions and 16 canonical pathways were identified. The most relevant functions and canonical pathways are listed in Table 7.5. Figure 7.1 compares the contribution of each domain to functions involved in 'Nervous System Development and Function', 'Neurological Disease', and 'Psychological Disorders'. All three groups and the Autism Model contribute to 'Nervous System Development and Function' while only the Communication Impairment group contributes to 'Psychological Disorders'. The Repetitive/Restricted Behavior group contributes strongly to 'Neurological Disease', however since this domain only encompassed 20 genes, the contribution is inflated. A total of 15 functions were implicated only in the overall Autism Model but not in the individual groups alone. Fifty functions were identified within the groups that were not identified in the overall Autism model.

Each gene was ranked by the number of relevant IPA functions and pathways identified by the IPA core analysis and 7 candidate genes were selected for fine-mapping based on gene function (Table 7.6 and Appendix Table 7.1). Two genes, *AKT1* and *JAG2*, were selected from the SRS-DT linkage region on Chromosome 14q32.31. *V-akt murine thymoma viral oncogene*

Table 7. 6: IPA functions and pathways of candidate genes selected for fine mapping analysis

Gene	Chromosome	Position (cM)	Domain	PPL (%)	Total Number of IPA Functions/ Pathways	IPA Functions/Pathways
AKT1	14	121-122	SRS	14-17	24	Neuregulin signaling, axonal guidance signaling
JAG2	14	121-122	SRS	14-17	9	Notch signaling
SEMA7a	15	77-78	LI	18-22	7	Axonal guidance signaling
PTPN9	15	78	LI	22	11	Neural tube defect, nervous system development
NTRK3	15	92-93	LI	7-8	24	Axonal guidance signaling, neurotrophin signaling, specification of neurons, development of afferent neurons
FES	15	96	LI	12	16	Axonal guidance signaling
SCNN1b	16	47	RI	4	6	Nervous system development and function, sodium channels

homolog 1 (*AKT1*) encodes a serine-threonine protein kinase that mediates growth factor-induced neuronal survival. *Jagged 2* (*JAG2*) encodes a ligand that functions in the Notch signaling pathway, which is important in embryonic development. A total of four genes, *NTRK3, PTPN9, FES,* and *SEMA7A,* were selected as candidate genes from the LI linkage region on Chromosome 15q23-26.2. *Protein tyrosine phosphatase, non-receptor type 9* (*PTPN9*) encodes a member of the protein tyrosine phosphatase family and is known to regulate cell growth and division and differentiation. *Semaphorin 7A* (*SEMA7A*) encodes a protein that functions in glycosylphosphatidylinositol (GPI) anchoring and has be implicated in axonal branching, neurite outgrowth, and neural development. *Feline sarcoma oncogene* (*FES*) encodes a tyrosine-specific protein kinase that is required for the maintenance of cellular transformation and axonal guidance signaling. *Neurotrophic tyrosine kinase, receptor, type 3* (*NTRK3*) encodes a membrane-bound receptor implicated in pathways including those for cell differentiation and development of neurons. One gene, *SCNN1b,* was selected from the RI linkage region on Chromosome 16p12.3. *Sodium channel, non-voltage-gated 1, beta subunit* (*SCNN1B*) encodes the beta subunit of a non-voltage gated sodium channel that controls water balance and distribution of electrolytes.

Association Results

A total of 167 SNPs were analyzed for association. For all analyses, cPPLD analysis reduced the PPLD|L values yielding no evidence for association (data not shown). When using the strictest criteria of one autism proband and one SLI proband (Tier I), 5 markers yielded PPLD|L > 2% (Table 7.7). For the SRS-DT+ASD phenotype, evidence for association strengthened for rs3784240 as Tiers II and III were included. Evidence for association to the RI+ASD phenotype increased for rs7204560 when Tier II was included, but decreased when the social impairment Tier (Tier III) was included. Association to the RI+ASD phenotype at rs8866113 and

rs8055868 remained constant at a PPLD|L = 3% for Tier I and II but decreased when Tier III was included. There was minimal evidence for association to the LI+ASD phenotype for rs6496461 (PPLD|L = 3%). All other markers showed evidence against association (See Appendix 6; Supplemental Tables 6A.2-10).

When ASD diagnosis was not included in the overall phenotype, the PPLD|L for rs3784240 decreased to 3% for all Tiers. For rs7204560, association to RI only decreased only when Tier III was included in the analysis. Evidence for association increased for three SNPs: rs6496461 (LI), rs886113 (RI), and rs8055868 (RI). The inclusion of Tier III decreased the evidence for association in all three SNPs. For rs3784432 and rs16941171, the exclusion of ASD diagnosis yielded evidence against association (Table 7.7).

When association to ASD diagnosis only was tested, rs6496461, rs886113, rs8055868, and rs7204560 showed evidence against association in all Tier groups. Evidence for association at rs3784240 was neutral. However, evidence for association to ASD diagnosis was positive at rs3784432 and rs16941171 (Table 7.7).

Table 7. 7: PPLD|L Results for Fine Mapping of Candidate Genes

Phenotype		SRS-DT	LI	LI	LI	RI	RI	RI
Gene		JAG2	NTRK3	NTRK3	NTRK3	SCNN1B	SCNN1B	SCNN1B
Chromosome		14	15	15	15	16	16	16
Position (cM)		124.12	95.90	95.90	95.96	49.66	49.69	49.69
Marker		rs3784240	rs3784432	rs16941171	rs6496461	rs886113	rs8055868	rs7204560
+ ASD	Tier I	4.0	2.4	2.0	3.0	3.0	3.0	4.0
	Tier I+II	4.0	2.4	2.0	3.0	3.0	3.0	5.0
	Tier I+II+III	5.0	2.1	2.2	1.7	2.5	2.4	4.0
- ASD	Tier I	3.0	2.0	1.7	6.0	4.0	3.0	4.0
	Tier I+II	3.0	1.6	1.6	7.0	4.0	4.0	5.0
	Tier I+II+III	3.0	1.3	1.6	2.2	3.0	2.2	2.4
ASD only	Tier I	3.0	4.0	4.0	1.7	1.5	1.8	1.9
	Tier I+II	2.4	3.0	4.0	1.7	1.4	1.9	1.9
	Tier I+II+III	2.3	3.0	3.0	1.6	1.4	1.9	2.0

Discussion:

Candidate Gene Selection

Genome-wide linkage analysis of the NJLAGS families yielded several findings on three chromosomes. A total of 587 genes reside within these linkage regions. As genome-wide association analysis did not provide strong evidence for association, candidate genes from these regions were selected for fine mapping analysis. The Core Analysis of IPA was used to prioritize the list of genes. Current genetic models of ASD propose that susceptibility to autism may be the result of a collection of variants contributing to each specific phenotypic domains of autism and that this polygenic collection of variants contributes to ASD susceptibility as a whole. To model this hypothesis, the linkage regions were analyzed in IPA in groups according to the domains of autism and combined into an overall autism model. IPA analysis of the overall autism model resulted in the identification of genes that function in axonal guidance, neuregulin signaling, and controlling the quantity of neurons. While the autism model represents these functions, each domain contributes to specific categories of functions within "Nervous System Development and Function". The Communication Impairment Domain specifically contributes to the "Axonal Guidance Signaling Pathway". This domain is also the sole contributor to the "Psychological Disorders" functional category in IPA. The Social Impairment Domain contains genes that function in the "Neuregulin Signaling Pathway" and the "Reelin Signaling Pathway." This domain also has a strong contribution to the control of the quantity of neurons. The Restricted/Repetitive Behaviors Domain only contained 20 genes causing a low membership within the IPA canonical pathways, but did contribute to the "Neurological Disorders" functional category in IPA. The contribution of the domain groups to specific aspects of the functions identified by the autism model supports the polygenetic model hypothesis.

Association Findings

While evidence for association to LI, RI, and SRS-DT phenotypes was minimal at best, subtle differences between each analysis are worth noting. The LI phenotype successfully identified linkage on Chromosome 15q25.1. Like our genome-wide association analysis using the Affymetrix Axiom array, fine mapping of 4 candidate genes in this region did not result in strong evidence for association. Overall, analyses in these regions produced evidence against association. The increase in evidence for association to the LI only phenotype for rs6496461 in Tiers I and I+II suggests that the LI diagnosis is a stronger contributor to association than the ASD diagnosis at this marker. The RI phenotype produced variable results for the different Tiers with and without the inclusion of ASD. The RI+ASD phenotype yielded the strongest findings for the entire dataset when including the ASD diagnosis. When ASD was not included in the affection status, there was a decrease in evidence for association to the RI phenotype when the Tier III families were included. The set of families in Tier I meets the strictest criteria for RI, while RI is not a requirement for Tiers II and III. Tiers II and III do have a strong ASD requirement, which may explain the consistent findings for RI+ASD in all three Tiers. The decreased evidence for Tier III also suggests that these families have fundamental genetic differences from Tiers I and II. The difference in Tiers also explains the variability in association findings for the SRS-DT phenotype at rs3784240. The strongest association findings for the SRS-DT+ASD phenotype included the Tier III families, which have social impairment as a requirement. As social impairment is a main characteristic of ASD, it is not surprisingly that there was a decrease evidence for association to SRS-DT without the inclusion of ASD diagnosis. The lower evidence for association to the ASD only phenotype in this SNP suggests that the SRS-DT and ASD diagnosis both contribute to the association signal seen in the SRS-DT+ASD analysis.

Despite the subtle differences in the results of each analysis, there was an overwhelming lack of evidence for association in this study. This lack of evidence does not fully discredit the hypothesis that the Axiom array utilized in our genome scan was not sufficiently dense enough to capture all LD information, as only 7 of the 587 genes located within the linkage regions were tested. Furthermore, these genes were selected primarily based on known function, not PPL value. Analysis of genes directly under the highest linkage peaks may yield more positive evidence for association. Another possibility is that the modest sample size limits power to detect common variants of small effect. Simulation analysis of recessive and dominant model parameters for each phenotype yielded variable power to detect association (Appendix 6; Supplemental Figure 6A.1 and 6A.2). Association is detectable for dominant models with a DGF > 5% and moderate to high disease risk for all three phenotypes. However, the NJLAGS sample does not have sufficient power to detect association under recessive models with low to moderate disease risk using the LI, RI, or SRS-DT phenotypes. Despite the variable power for generalized models, all of the maximized models for linkage regions identified for LI, RI, and SRS-DT had more than 80% power to detect association (Table 7.4 and Supplemental Figure 6A.3; Appendix 6). Five of the seven maximized linkage models were recessive models with medium to high disease risk. Both LI Model 2 and RI Model 1 are fully penetrant recessive models with a DGF = 80%; which is biologically unlikely for complex disorders like ASD and SLI. Simulation analysis of these two models produces strong power to detect association due to the extreme disease risk associated with these parameters. The remaining two maximized models (LI models 1 and 3) are dominant models with medium to high disease risk; these models have more tha 80% power to detect association in the NJLAGS families. Since the evidence for linkage in every region was under a model of homogeneity ($\alpha = 1$), it is unlikely that the lack of evidence for association is due to low sample power in these regions.

The lack of evidence for association in these genes lends support to the role of allelic heterogeneity and association analyses generally have not been replicated [Abrahams and Geschwind, 2008]. Individually, common variants that have been identified exhibit small effects on the risk for ASD [Anney et al., 2012] but may exhibit larger additive effects as a group [Klei et al., 2012]. Most recently, investigators have begun to evaluate the role of rare variants in ASD through the use of high throughput sequencing. Like association analyses, these studies have identified many single-instance variants with small effect on the risk for ASD [Sanders et al., 2012; Neale et al., 2012; O'Roak et al., 2011]. Association studies have been more successful in SLI studies with consistent findings with *CNTNAP2, FOXP2,* and *FOXP1* [Newbury et al., 2010]. However, there is also a high degree of allelic heterogeneity in SLI. The role of allelic heterogeneity across disorders is uncharacterized for ASD and SLI. As genetic studies have identified a link between ASD and SLI in *CNTNAP2*, there may also be a link between genetic variants that are relevant to both ASD and SLI. It is also possible that SLI and ASD have some of the same key genes in pathogenesis but not the same underlying variants or molecular mechanisms. This hypothesis cannot be directly addressed until functional alleles have been associated with each disorder.

Conclusions:

NJLAGS is the first study to investigate the shared genetics between ASD and SLI. The NJLAGS study design successfully developed phenotypes that measure language, reading, and social impairments, as well as obsessive-compulsive behaviors in families ascertained for ASD and SLI. Each of these phenotypes yielded strong evidence for linkage but did not show evidence for association. Functional analysis of the genes within each linkage region supports a polygenic model for ASD where a collection of variants contributes to each specific domain of ASD. The lack of evidence for association in these regions suggests that the shared genetics that

contribute to ASD and SLI susceptibility is not due to a collection of common variants, although this study did not comprehensively assess all genes under the linkage peaks, so the possibility of common variants in genes that were not tested cannot be excluded. Future studies will work to identify rare variants in the NJLAGS families that contribute to the genetics of ASD and SLI.

Chapter 8: Discussion

Review of Major Findings

The overall goal of these studies was to use communication impairments in the creation of endophenotypes for autism in order to reduce the heterogeneity in our genetic analyses. The communication domain of autism can manifest as either a speech impairment characterized by non-verbal, minimally verbal or unintelligible speech or language impairment (receptive and/or expressive) ranging in severity and type. Speech is the verbal means of communicating that includes articulation, voice, and fluency while language, refers to understanding what words mean, how to put words together, and the use of the appropriate word combinations in different situations. The studies presented in this thesis identified several speech and language phenotypes in families who were ascertained for autism. Chapters 2 and 3 focused on two motor speech phenotypes (with and without specifying for language comprehension) in nonverbal individuals from multiplex autism families. Chapters 4-7 place emphasis on language impairments in the New Jersey Language and Autism Genetics Study (NJLAGS) sample. NJLAGS specifically investigated oral language impairments (syntax, semantics, phonology, and pragmatics) and written language impairments (decoding, comprehension, reading fluency, and spelling), as both represent discreet forms of language impairments. Each of these communication phenotypes yielded evidence for linkage in non-overlapping regions, but did not produce strong evidence for association.

Reduction of Sample Heterogeneity

The work presented in these studies strives to reduce the genetic heterogeneity of autism samples by utilizing several methods of phenotype development. Our first study utilized a hybrid of two methods of heterogeneity reduction in the Autism Genetics Resource Exchange

(AGRE) dataset. We stratified the sample into clinically more similar groups by verbal ability. In order to do this, we developed two non-verbal motor speech phenotypes from variables in the Autism Diagnostic Interview-Revised (ADI-R). Our second study sought to investigate the shared genetics between Autism Spectrum Disorders (ASD) and Specific Language Impairment (SLI) by developing communication phenotypes that are specific to the NJLAGS sample. Both of these studies defined communication endophenotypes of autism that successfully met the 5 criteria set forth by Gottesman et al., 2003.

The first criterion of an endophenotype is the association of the trait with the disorder in question. Communication impairments (phonology, structure, content, or pragmatic) comprise the definition of the communication impairment domains of ASD and are therefore, associated with ASD. The Nonverbal Motor Speech Disorder (NVMSD) phenotypes were defined based on the ADI-R, which is an assessment of ASD diagnosis. Similarly, the Language Impairment (LI) and Reading Impairment (RI) phenotypes defined in the NJLAGS sample were based on assessments that are used to diagnose SLI, therefore supporting the association of these endophenotypes with SLI. The Social Responsiveness Scale (SRS) and Yale Brown Obsessive-Compulsive Scale (YBOCS) phenotypes were designed as secondary endophenotypes for measurement of the Broad Autism Phenotype (BAP) in the NJLAGS samples. As these families were ascertained for ASD and SLI, there is an increased loading for BAP traits among the family members in this dataset.

The NVMSD and NJLAGS phenotypes met the second endophenotype criterion of being heritable within the study sample. While heritability analysis was not conducted on the NVMSD phenotypes, this study required the presence of two probands, who were frequently siblings, in order to increase the genetic loading of the NVMSD traits within the families. Chapter 5

presents the heritability analysis of the testing variables used to define the communication phenotypes in the NJLAGS dataset [Bartlett et al., 2012].

Third, each endophenotype should be state-independent, meaning that the trait manifests in an individual despite current disease activity. The ASD status of an individual is a life-long diagnosis. However, this criterion is not directly applicable to the ASD status of an individual, but rather the communication impairment itself. By definition, the NVMSD phenotypes exist despite one's ASD diagnosis. However, NVMSD represents the one of the most severe forms of communication impairment, often persisting throughout one's lifetime and, in this study, coincides with ASD diagnosis. On the other hand, the language phenotypes defined in the NJLAGS sample can be variable on a longitudinal scale. Often, an individual with a language impairment may learn to compensate for their disability as an adult. In order to account for this compensation, higher order language tests, such as the Comprehension Assessment of Simple Language (CASL), were included in the definition of the language impairment phenotype and history of the impairment in terms of interventions and delayed language as a child were incorporated into the disease status. The SRS and YBOCS phenotypes are both measures of a snapshot in time of an individual's impairment. While the stability of these phenotypes was not measured over time in this sample, the snapshot assumes a state-independent trait.

The fourth requirement is that the endophenotype should co-segregate with the disease of interest. Even though the NVMSD phenotypes were defined using variables from the ADI-R, the NVMSD assessment was determined despite ASD diagnosis. However, of the 643 individuals who met the criteria for NVMSD:ALL 99% also met diagnostic criteria for ASD. As the purpose of the analyses of the NJLAGS dataset is to investigate the shared genetics of ASD and SLI, the

phenotypes must segregate with both disorders. Of the verbal individuals with ASD in the NJLAGS sample 68% met criteria for LI and 48% met the criteria for RI. When nonverbal individuals with ASD were included as impaired, 86% met criteria for LI and 83% met criteria for RI. While these phenotypes clearly co-segregate with ASD, the co-segregation of the SLI proband with the LI and RI phenotypes is inherent by definition. A total of 94% of the ASD probands met the cutoff for the SRS-DT phenotype. Approximately 20% of individuals without ASD who met criteria for LI or RI also met the criteria for SRS-DT. The low co-segregation of SRS-DT with SLI is not surprising, as pragmatic difficulties only persist in subsets of individuals with SLI.

The final criterion for an endophenotype is the increased prevalence of the trait in the dataset than in the general population. Both of these datasets were enriched for ASD and communication impairments. The rate of NVMSD:ALL in the entire AGRE database was 11%, which is remarkably higher than the estimated prevalence of 0.46% in the general population (Chapter 3). As stated in Chapter 4, the rate of LI and RI (42%) in the NJLAGS dataset after accounting for ascertainment bias was higher than in the general population (7-8%) [Tomblin et al., 1997]. Excluding ASD probands, a higher rate of obsessive compulsive behaviors was reported in NJLAGS families (5-8% in children and 8-14% in adults) as compared to the reported rates of OCD in the U.S. population which are reported as 0.4% in children and 1% in adults [Kessler et al., 2005; Wang et al., 2005; Weissman et al., 1994]. Similarly excluding ASD probands, there was a high rate of social impairments in the NJLAGS families (15-19%), which is higher than the rate in the general population (0.3%-1.4%) [Constantino and Todd, 2003].

The NVMSD, LI, and RI phenotypes successfully met all 5 criteria of an endophenotype. These phenotypes each characterized distinct linguistic levels of communication impairment

seen in ASD, including verbal ability and speech, structural language impairment, language comprehension and pragmatic use of language. The secondary endophenotypes of social impairment and obsessive-compulsive behaviors rounded out the exploration of the shared genetics between ASD and SLI, by investigating the role of the BAP within these families.

Linkage Findings

All linkage findings are listed in Table 8.1. Linkage analysis of the NVMSD endophenotypes produced several novel linkage signals. Analysis of NVMSD:ALL in the AGRE database yielded evidence for linkage on Chromosomes 1q24.2, 3q25.31, 4q22.3, 5p12, 5q33.1, 17p12, 17q11.2, and 17q22. With the exception of the regions on Chromosome 17, none of these linkage regions have been implicated in ASD or in language impairment prior to this study. Our linkage regions on Chromosome 17 overlap with linkage scans of the AGRE database [Stone et al., 2004; Cantor et al., 2005; Yonan et al., 2003; Bartlett et al., 2005; McCauley et al., 2005]. This is not surprising due to the large area encompassed by our linkage region on Chromosome 17 and also due to the overlap in AGRE samples among studies. Sequential updating of all linkage regions with additional AGRE families strengthened evidence for linkage in the following regions: 1q24.2, 3q25.31, 4q22.3, 5p12, and 17q22. The addition of genetic evidence from these families decreased evidence for linkage on 17p12 and 17q11.2. This finding suggests that there is a degree of genetic heterogeneity in these families for this region. This hypothesis is supported by an meta-analysis of several linkage scans of the AGRE database that revealed a significant amount of between-scan heterogeneity in this region [Trikalinos et al., 2006].

Despite the fact that the NVMSD:C sample is a subset of NVMSD:ALL sample, linkage analysis of both phenotypes did not yield overlapping regions. The NVMSD:C phenotype is a more narrowly defined nonverbal phenotype than the NVMSD:ALL, specifying that the individual

clearly understands language, but cannot physically reciprocate verbal communication. This definition is more is more closely related to the characteristics for Childhood Apraxia of Speech (CAS) which is a motor speech disorder that involves poor motor planning and results in speech output with compromised intelligibility. Genome-wide linkage analysis of NVMSD:C yielded evidence for linkage was on 4p15.2 and 21q22.2. The identification of linkage regions that are

Table 8. 1: Linkage regions identified using communication endophenotypes of autism.

Chromosome	Position (cM)	Phenotype	PPL (%)	Other Findings in Autism or Language Studies
1q24.2	180	NVMSD:ALL	82	Bartlett et al., 2005; Buxbaum et al., 2004; Davis et al., 2009
3q25.31	166	NVMSD:ALL	27	Alarcón et al., 2005; Auranen et al., 2002, 2003; Salyakina et al., 2010; Rehnström et al., 2010
4p15.2	45	NVMSD:C	84	-
4q22.3	109	NVMSD:ALL	28	Schellenberg et al., 2006
5p12	66	NVMSD:ALL	52	Yonan et al., 2003
5q33.1	156	NVMSD:ALL	21	-
13q21.2	58	YBOCS	36	Bradford et al., 2001; CLSA et al., 1999; Steele et al., 2011; Vorstman et al., 2006
14q32.31	117	SRS-QT	37	Qiao et al., 2013; Merritt et al., 2005
15q25.1	83	LI	57	-
15q26.2	120	SRS-DT	52	Chagnon, 2006
16p12.3	43	RI	36	-
17p12	36	NVMSD:ALL	45	Alarcón et al., 2005; Yonan et al., 2003; McCauley et al., 2004, 2005; Ylisaukko-oja et al., 2006; Sutcliffe et al., 2005; Vourc'h et al., 2003
	39	NVMSD:ALL	60	
	48	NVMSD:ALL	76	
17q11.2	55	NVMSD:ALL	44	Yonan et al., 2003; McCauley et al., 2005; Ylisaukko-oja et al., 2006
17q22	80	NVMSD:ALL	32	Cantor et al., 2005; Field et al., 2013
21q22.2	54	NVMSD:C	32	-

exclusive to NVMSD:C supports the hypothesis that a general nonverbal phenotype is distinct from a nonverbal phenotype with language comprehension.

The exclusivity of linkage signals for distinct phenotypes is echoed in the NJLAGS linkage analyses. The analysis of each endophenotype yielded evidence for linkage in non-overlapping regions: 13q21.2 (YBOCS), 14q32.31 (SRS-QT), 15q25.1 (LI), 15q26.2 (SRS-DT) and 16p12.3 (RI). By including ASD diagnosis in the phenotype as 'affected' for each linkage analysis, we were able to directly investigate shared loci between ASD and the phenotype of interest. If there were not a connection between the phenotype and ASD, then no evidence for linkage would be obtained. As each phenotype yielded evidence for linkage in distinct regions, there was also evidence for independent genetic connections between ASD and LI, RI, social impairments, and obsessive-compulsive behaviors, respectively. This evidence supports a genetic link between ASD and SLI within these families. Furthermore, the evidence supports a polygenic model for ASD where distinct collections of variants contribute to the impairments in each domain of ASD. Overall, the evidence for linkage in both the AGRE and NJLAGS samples supports the use of our defined endophenotypes to identify linkage by reducing the heterogeneity of our samples.

Effectiveness of Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA) is an effective tool for identifying known functions of biological molecules. IPA specializes in the identification of networks that connect a given set of molecules from several types of datasets including gene lists, expression data, microRNA targets, variant analysis, and drug interactions. In the studies presented in Chapters 2, 3, and 7, IPA's Core Analysis was utilized in a unique way to prioritize a list of candidate genes from several linkage analyses. The Core Analysis of IPA uses a Fischer's Exact Test to identify functions and canonical pathways that are unique to a given list of molecules. There are several

advantages to the use of IPA. First, the IPA Knowledge Base is a curated database of biological and chemical information on thousands of biological molecules and chemicals. The Knowledge Base provides the foundation of all of the computations performed in IPA and is updated frequently to provide the most up-to-date and accurate information. Second, the Knowledge Base includes canonical pathways that are well established in the scientific community. The Core Analysis identifies the pathways that are highly represented by the molecules in a given list. Third, the networks identified by IPA are not limited to direct connections and take into account cell type, species specificity, directionality, and mutations/modifications. Finally, IPA allows the user to input unpublished data and customized pathways, making analyses flexible to user needs. An important limitation of the use in IPA in these studies is the confinement of the analysis to coding regions located within the linkage regions. It is possible that the region responsible for a linkage signal is located within an intergenic, regulatory region that acts either in cis or trans within a biological pathway. It is important to note that the candidate gene selection procedure conducted with IPA is only a means of data reduction as a place to begin follow-up investigations.

In the analyses in Chapters 2 and 3, IPA was used to prioritize a list of 645 genes from the NVMSD:ALL linkage regions for candidate gene selection. The functional analysis of these genes produced 502 functions, 24 of which were directly related to “Nervous System Development and Function.” As many of the identified functions were general or unrelated to neurological disorders, candidate genes were selected only if the functions identified were directly related to “Nervous System Development and Function”, “Neurological Disorders”, or “Psychological Disorders”. One inherent limitation to this approach is the exclusion of potential candidate genes with immunological or metabolic functions. While the studies in Chapters 2 and 3 did not utilize the ranking system devised in Chapter 7, control analyses were conducted

to eliminate potential false positive results. False positives may arise from over-representation of a gene or function within the Knowledge Base. As only curated published findings are included in the Knowledge Base, over-representation could occur for functions or disorders that have been extensively investigated. The Fischer's Exact Test conducted by the Core Analysis does take this into account; however, the control analyses were designed to help to eliminate bias overrepresentation of a function in the IPA Knowledge Base. Another potential bias toward false positives is the analysis of a region that contains multiple variants of a large gene family. Each member of the gene family is analyzed as an independent entity, while it may be more appropriate to treat the family members as one unit that contributes to a biological function as a group. As a final control, the results from IPA were compared to the results obtained from the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [Huang et al., 2009b] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [Kanehisa and Goto, 2000; Kanehisa et al., 2012]. As each utility has distinct ontology nomenclature, direct comparison between IPA, DAVID, and KEGG is difficult. Therefore, these utilities were merely used as a means of comparison.

The analysis in Chapter 7 differs from that of Chapters 2 and 3 in that a ranking system was devised to select candidate genes. This helps to circumvent the arbitrary threshold that IPA imposes as a statistical cut-off for functions. As each NJLAGS phenotype represents each domain of ASD, the linkage regions were analyzed separately by domain. A combined autism model was also analyzed with all of the linkage regions. Overall, the ranking system identified functions that were specific to each domain of autism as well as functions that were specific to autism as a whole. While DAVID and KEGG were used as comparative tools in Chapters 2 and 3, the results from these utilities were included in the ranking system in Chapter 7.

Association Findings

Despite the strong evidence for linkage, evidence for genetic association was limited for all phenotypes. Several hypotheses may explain the lack of association within these populations. For the genome-wide association analyses, it is possible that the genotyping assays used were not comprehensive enough to sufficiently tag all haplotypes. The Affymetrix 5.0 SNP array effectively tags 65% of all common haplotypes genome-wide [Hodgkinson et al., 2008] and the Axiom 1.0 SNP array tags 56% of common haplotypes [Bartlett et al., 2013]. This hypothesis, however does not explain the lack of evidence for association in our fine mapping analyses. While it is possible that there is truly no association in these regions, it is also possible that our method of candidate gene selection was not inclusive enough. An alternative hypothesis is that while the AGRE and NJLAGS samples were suitable for linkage analysis, the samples may not have sufficient power to detect association. As presented in Chapter 3, the NVMSD:ALL sample had sufficient power to detect association as a whole. However, the sample did not have sufficient power to detect association in the candidate gene regions based on the parameters defined by the maximized models of these regions. The NJLAGS sample, on the other hand, had strong power to detect association based on the maximized models for each linkage region (Chapter 7).

A likely reason for the lack of association is allelic heterogeneity in ASD and, potentially, SLI genetics. Genetic association identifies the relationship between a phenotype and common variants in a population. The presence of multiple and/or rare variants that contribute to disease susceptibility would not be detected through association analysis. Due to the general lack of positive and replicative findings in genome-wide association studies, this hypothesis has become increasingly more popular. Furthermore, association analyses of candidate genes frequently produce conflicting results across different studies (reviewed in Freitag, 2007).

Through the studies conducted to date, it is estimated that 15-20% of ASD cases may be accounted for by rare *de novo* events that would not be detected through association analysis [Yu et al., 2013]. Neale et al., 2012 observed that genetic models that fit the data seen in these sequencing analyses support a polygenic model suggesting that the combination of variants, not the number of variants, contribute to ASD risk. Overall, the rate and distribution of known *de novo* mutations in ASD probands does not deviate from that of the general population and a single observation of a mutation is not sufficient evidence to implicate a gene as a risk factor for ASD due to its small effect. A study by Moreno-De-Luca et al., 2012, was able to add statistical support for several deleterious CNVs in ASD cohorts that may not have reached significance in smaller studies by combining CNV information from large clinical case-control cohorts. The combination of studies that identify rare variants that contribute to ASD susceptibility may also add statistical support to these findings and identify the role of these variants in a polygenic model of ASD.

Assessment of Power to Detect Association in the NVMSD and NJLAGS datasets

The NVMSD and NJLAGS cohorts are distinctive datasets with unique study designs. Both datasets are multiplex families that have been ascertained primarily for autism research. The multiplex study design allows for the study of inherited autism susceptibility factors with increased power, as opposed to *de novo* mutations that are commonly found in simplex populations. Both datasets successfully identified linkage to the endophenotypes discussed above; however did not yield evidence for association. As discussed in Chapter 3, the NVMSD dataset had sufficient power to detect association for dominant and recessive models with moderate to high disease risk. Power to detect association was limited for models with low disease risk or a high percentage of unlinked families. The maximized models for each NVMSD linkage region did not have more than 80% power to detect association due to the low disease

risk of each model. The two models that had moderate to high disease risks were not successful for different reasons. The first model, for the region containing *NKX6-1,* was only represented by three tag SNPs for fine mapping. The second model, for the region containing *LMX1A*, was a dominant model with high penetrance. This second model is not likely for the NVMSD dataset as, by definition, only offspring are affected for the NVMSD phenotypes.

While there was sufficient power for most models in the NVMSD families, the smaller sample size of NJLAGS families limits the power to detect association in this dataset. The NJLAGS dataset has sufficient power to detect association for dominant models with moderate to high disease risk. However, the NJLAGS dataset is more limited in its power to detect association for recessive models with low to moderate disease risk. Unlike the power to detect association for the maximized models identified in NVMSD analyses, the NJLAGS dataset has sufficient power to detect association for each of the maximized models identified by the NJLAGS linkage analyses. The majority of the maximized models are fully penetrant homogeneous recessive models ($\alpha = 1$) with high risk. Despite the limited power for recessive models, the combination of high disease risk and homogeneity contribute to the strong power to detect association for the maximized models in the NJLAGS families. Like the *LMX1A* model in the NVMSD dataset, several of the maximized models for the NJLAGS linkage regions are biologically unlikely. The fact that association was not detected with these models lends support for the role of multiple or rare variants in ASD and SLI.

The NVMSD and NJLAGS datasets have fundamentally different family structures that contribute to the difference in their power to detect association. The NVMSD dataset has the strongest power to detect recessive models since there is lack of founders who are affected for the NVMSD phenotype. Since the NJLAGS dataset was designed to increase the genetic loading

of language impairments in the family members of the ASD proband including parents, this dataset has the strongest power to detect dominant genetic models. The lack of strong association findings in these datasets supports the role allelic heterogeneity in communication impairment in ASD.

Future Directions

Increased Sample Size

Analyses of the NJLAGS dataset currently consist of 51 complete families who meet the strictest criteria of at least one autism proband and one SLI proband. These families represent a small percentage of all ASD families. Over the next five years, 100 additional families will be recruited using relaxed criteria of an ASD proband (as opposed to a strict definition of Autistic Disorder) and one SLI proband in order to increase eligibility rates. Alternatively, if no SLI proband is obtained, families will be eligible if they contain two ASD probands where at least one has a language learning impairment. These additional families will be used to increase the power of the NJLAGS dataset and to investigate the role of communication impairment in a broader group of ASD families. As shown in Chapter 3, increasing the sample size of the NVMSD:ALL phenotype added evidence for linkage for several linkage regions. Previous linkage scans of the NJLAGS dataset can be sequentially updated with the additional families in order to assess the role of our current phenotypes. As discussed in Chapters 3 and 7, the NVMSD and NJLAGS datasets had limited power to detect association. Increasing the sample size will also increase the power of the sample to detect association. In addition to increasing the sample size of the NJLAGS sample, a total of 7 families from the NJLAGS sample also meet the criteria for NVMSD:ALL. With the recruitment of more broad definition families, it is likely that this number will increase. Genetic evidence for linkage from these families can be added to the linkage analysis of the NVMSD families.

Assessment of Phenotype Stability

An important concept in endophenotype development is that ideal endophenotypes are a reflection of trait and not state; that is they should not fluctuate with course of illness or treatment but should reflect a steady state of underlying genetic liability. Of the 5 endophenotype criteria, this is the weakest point for the NVMSD and NJLAGS endophenotypes as each dataset provides a fixed representation of potentially dynamic traits. One inherent limitation of the AGRE database is the static availability of phenotypic information. Each assessment, while standardized, represents a snapshot in time of an individual's clinical traits, and cannot be directly evaluated for phenotypic stability. Unlike the AGRE database, the NJLAGS sample is collected and managed in-house, allowing for the possibility of sample reassessment. Reassessment of the NJLAGS families will allow us to identify stable phenotypic measures. This is particularly important for the language measures, which are standardized for specific age groups; a 10-year old subject does not receive the same set of subtests as a 14-year old subject. Stable measures from the test battery should be given higher priority for continued genetic analyses. In addition to age-specific standard scores, treatment and intervention programs for both autism and language impairment can affect an individual's phenotype. While these phenotypes may be variable, the identification of components of the autism or language impairment phenotypes that are amenable to therapy is valuable to the study of gene x intervention studies as candidate traits for future use in studies of the potential interaction between genetic factors and treatment response with the ultimate goal of personalized medicine. Currently, the NJLAGS team plans on reassessing 25 ASD and SLI probands to test for phenotypic stability.

Investigation of the Role of Rare Variants in ASD and SLI

The lack of association findings in the NVMSD and NJLAGS dataset could be due to allelic heterogeneity or insufficient density of SNPs in strong linkage disequilibrium to single risk alleles at each locus. Either possibility would limit power to detect association and can be addressed by direct sequencing of the genome. Direct sequencing can be conducted on several coverage levels: targeted candidate regions (i.e. linkage regions, candidate genes), whole exome level, and whole genome level. As discussed above, several studies have successfully identified rare variants in ASD probands through targeted and whole exome sequencing, which are limited to coding regions of the genome. Alternatively, whole genome sequencing would not be limited to coding regions of the genome, allowing for variant discovery and increasing the likelihood of identifying a regulatory variant. NJLAGS has recently obtained funding support for whole genome sequencing analysis of 25 unrelated ASD probands and has plans to expand this analysis to 25 additional unrelated SLI probands once funding is awarded. This would be the first study of its kind to investigate the role of allelic heterogeneity in families that contain both ASD and SLI probands. Following the identification of potential candidate variants, targeted sequencing and genotype verification will be conducted in family members of these selected probands.

Conclusions

The unique design of our communication endophenotypes revealed underlying genetic connections between each phenotype and ASD. Furthermore, the study designs effectively defined distinct endophenotypes that measure different aspects of communication impairment: speech, language, and reading. This is evident as analysis of each phenotype yielded non-overlapping linkage regions. Future studies will assess the stability of these endophenotypes through reassessment of selected NJLAGS probands. Even though we were unable to identify association to these phenotypes, it is an observation that is not uncommon in genetic studies of

ASD. The lack of evidence for association supports the hypothesis that a combination of multiple and/or rare variants may contribute to ASD susceptibility and communication impairment. This hypothesis will be investigated in the NJLAGS dataset via whole genome sequencing of ASD and SLI probands.

Appendix 1: Supplemental Materials for Chapter 2: Combined linkage and linkage disequilibrium analysis of a motor speech phenotype within families ascertained for autism risk loci

Supplemental Table A2. 1: ADI-R variables used to develop motor speech phenotypes: NVMSD:ALL and NVMSD:C.

Definition of ADI-R Variables	
Name	**Description**
LEVELL (Level of language)	Caregiver-based description of how much language child uses on a daily basis.
CARTIC (Current level of articulation)	Specifies the level of speech intelligibility at time of assessment (age 4+) on scale from 0-3.
ARTIC5 (Level of articulation at age 5)	Snapshot of intelligibility at age 5 - with CARTIC, used to determine temporal changes in articulation.
CCOMPSL (Comprehension of simple language)	Measures spoken language comprehension in absence of visual cues (e.g., pointing or gesturing).

Supplemental Table A2. 2: Sample demographics

	NVMSD:ALL	NVMSD:C
Individuals with ADI-R and 550K data	1146	778
Number of participating families	203	135
Number with autistic disorder/number of families	383/202	249/133
Number affected for phenotype	427	281
% Male/Female	79/21	80/20
Age at ADI-R assessment	98.83	107.02
(s.d.)	-63.18	-52.4

Supplemental Tables 2A.3a and 2A.3b. NVMSD:ALL and NVMSD:C genes of interest identified by IPA core analysis. Each gene of interest was found to be significant by at least one of our 3 core analyses in IPA. Each analysis (narrow, intermediate, or broad) is independent of the others and is dependent solely on the list of genes specific to that definition and their functions. As more genes are added by each definition, the number of functions increases and their frequency fluctuates between definitions. This is why some genes may appear significant in one analysis, but not all three.

Supplemental Table 2A.3a

NVMSD:ALL						
Chromosome	Gene	CM	PPL	Narrow	Intermediate	Broad
1	PBX1	173	5			B
1	LMX1A	173-174	15		I	B
1	RXRG	174	15			B
1	ATB1P1	179	73	N	I	B
1	SELE	180	63		I	B
1	MIR214	182	45	N		
1	FASLG	182-183	37	N	I	B
1	SERPINC1	183-184	18		I	B
1	TNR	184	18		I	B
1	ABL2	188	8			B
1	LHX4	188	8			B
1	LAMC1	191-192	14		I	B
1	NCF2	192	20		I	B
1	PLA2G4A	194	8			B
1	PTGS2	194	8			B
3	MME	165	15		I	B
3	KCNAB1	165	19		I	B
3	PPM1L	168	8			B
4	HPSE	96	11			B
4	NKX6-1	98	15		I	B
4	MAPK10	99	23		I	B
4	SPP1	101	24		I	
4	SNCA	103-104	13		I	B
4	GRID2	105	9		I	B
4	ATOH1	106	11		I	B
4	HPGDS	106	13		I	B
4	UNC5C	107	19		I	
4	NFKB1	112	9		I	B
4	CISD2	112-113	11		I	

NVMSD:ALL						
Chromosome	Gene	CM	PPL	Narrow	Intermediate	Broad
5	GDNF	61	5			B
5	SLC1A3	61	5			B
5	C9	64	30	N	I	
5	PRKAA1	65	36	N	I	
5	C7	65	36	N	I	
5	C6	65	36	N	I	B
5	PTGER4	65	36	N		B
5	GHR	66	41	N	I	B
5	HCN1	67	25			B
5	ISL1	67	16			B
5	ITGA2	68	10		I	B
5	GZMK	70	12		I	B
5	GZMA	70	12		I	B
5	MAP3K1	72	8			B
5	DPYSL3	153-154	21		I	
5	SPINK5	154	24			B
5	HTR4	154-155	24		I	B
5	MIR143	155-156	24	N		
5	MIR145	155-156	24	N		
5	PDGFRB	157	12		I	B
5	CDX1	157	12		I	B
5	CAMK2A	157	12		I	B
5	GLRA1	159	9			B
5	SYNPO	159	9			B
5	LIFR	62-63	10		I	B
5	ITGA1	67-68	12		I	B
5	IL6ST	70-71	13		I	B
6	POU4F3	152	6			B
16	NTN3	7	5			B
16	CREBBP	9	6			B
16	TRAP1	9	6			B
16	HMOX2	10	7			B
16	GRIN2A	25-26	9		I	B
17	MAP2K4	35	33			B
17	PMP22	46-47	77	N	I	B
17	UBB	47-48	79	N	I	B
17	TRPV2	48	79	N		
17	MPRIP	48-49	70		I	
17	LLGL1	50	33	N		
17	FLII	50	33	N		
17	ALDH3A2	50-51	25	N	I	B

NVMSD:ALL						
Chromosome	Gene	CM	PPL	Narrow	Intermediate	Broad
17	ULK2	50-51	25	N		
17	MAPK7	50-51	25			B
17	ALDH3A1	51	25			B
17	KSR1	52-53	18		I	B
17	NOS2	52-53	18		I	B
17	VTN	53	21		I	B
17	SLC6A4	54	25	N	I	B
17	NF1	54	34	N	I	B
17	GIT2	54	25	N		
17	SSH2	54	25	N		
17	OMG	54	34	N		
17	RAB11FIP4	54-55	42		I	
17	CDK5R1	55-56	39	N	I	B
17	ACCN1	57	21		I	B
17	ERBB2	67	5			B
17	ERBB2	67	5			B
17	NEUROD2	67	5			B
17	PPP1R1B	67-68	5			B
17	CCR7	68	5			B
17	RARA	68	5			B
17	THRA	68	5			B
17	CSF4	68	5			B
17	BECN1	71	7			B
17	HDAC1	71	7			B
17	PPY	71	7			B
17	PYY	71	7			B
17	NSF	72	7			B
17	CRHR1	72	7			B
17	WNT3	72	7			B
17	WNT9B	72	7			B
17	MAPT	72	7			B
17	GFAP	72	8			B
17	ITGB3	72-73	10		I	B
17	NPEPPS	73	8			B
17	HOXB1	73	8			B
17	HOXB2	73-74	8			B
17	HOXB8	73-74	8			B
17	NGFR	74-75	7			B
17	NME1	78	18		I	
17	NOG	84	7			B
17	DGKE	85	12		I	B
17	MPO	88-89	6			B
17	RPS6KB1	89	6			B
17	ACE	93	8			B
17	GH1	93	8			B
17	PECAM1	93-94	8			B
17	RGS9	94	6			B

Supplemental Table 2A.3b

NVMSD:C						
Chromosome	Gene	Position (cM)	PPL	Narrow	Intermediate	Broad
4	FGFR3	1	5			B
4	NAT8L	1-2	5			B
4	HTT	3	6			B
4	ADRA2C	4-5	7			B
4	OTOP	5	7			B
4	MSX1	6-7	5			B
4	EVC2	8-9	9		I	B
4	CRMP1	9	18	N	I	B
4	WFS1	11	9			B
4	PPP2R2C	11	8			B
4	SORCS2	15	6			B
4	DRD5	22	6			B
4	SLC2A9	22	6			B
4	RAB28	25-26	10			B
4	BST1	29	21	N	I	B
4	CD38	29-30	21	N		B
4	PROM1	30	23	N		B
4	FAM184B	33	52	N	I	B
4	SLIT2	35	35	N		B
4	PPARGC1A	38	19	N		B
4	SOD3	40	15			B
4	LGI1	40	13		I	B
4	ZCCHC4	41	15			B
4	RBPJ	41-42	35	N		B
4	CCKAR	44-45	83			B
4	STIM2	45	78	N		B
4	PGM2	55-56	19			B
4	TLR10	57	6			B
4	TLR6	58	6			B
4	FAM114A1	58	6			B
4	IL2	129	5			B
4	FGF2	129	5			B
21	DSCAM	51	10	N	I	B
21	BACE2	55	19	N	I	B
21	MX2	56	22	N		
21	MX1	56-57	22	N		
21	ABCG1	60-61	6			B
21	PKNOX1	62	8			B
21	CBS	62-63	7			B
21	TRPM2	65	9			B
21	ITGB2	65-66	12			B
21	COL18A1	67	12			B
21	PCBP3	68	7			B

Supplemental Tables 2A.4 a, b, and c: IPA Analysis Results. The number of genes analyzed for each analysis is dependent upon the number of genes that can be mapped to the IPA knowledge base. Each core analysis produced a list of biological functions/diseases that are unique to the dataset. From the list of significant functions and diseases, we identified functions that were related to nervous system development and function, psychological disorders, genetic disorders, and neurological disorders. The total number of relevant functions and candidate genes obtained for each analysis are listed in each table. The most prevalent relevant functions were tallied and the functions with more than 5 contributing genes are listed with the number of genes in parentheses. The candidate genes identified in each analysis and the number of their relevant significant functions was tallied and the genes with more than 5 relevant functions are listed with the number of their significant functions in parentheses.

Supplemental Table 2A.4a: NVMSD:ALL IPA Results. Results from the core analysis of 645 genes located within the Intermediate definition of linkage regions in NVMSD:ALL.

NVMSD:ALL IPA Results				
Number of Genes in IPA Knowledge Base	Number of Relevant Significant Functions	Number of Candidate Genes Identified	Most Prevalent Functions	Genes with Most Functions
607	83	52	Neurological Process of Cells (13) Cell Death of Neuroglia (9) Neurological Process of Axons (9) Cell Death of Oligodendrocytes (8) Cognition by Mice (8) Learning by Mice (7) Synaptic Plasticity (6) Apoptosis of Neuroglia (6) Apoptosis of Oligodendrocytes (5)	FASLG (5), MAPK10 (5), NKX6-1 (5), PMP22 (5)

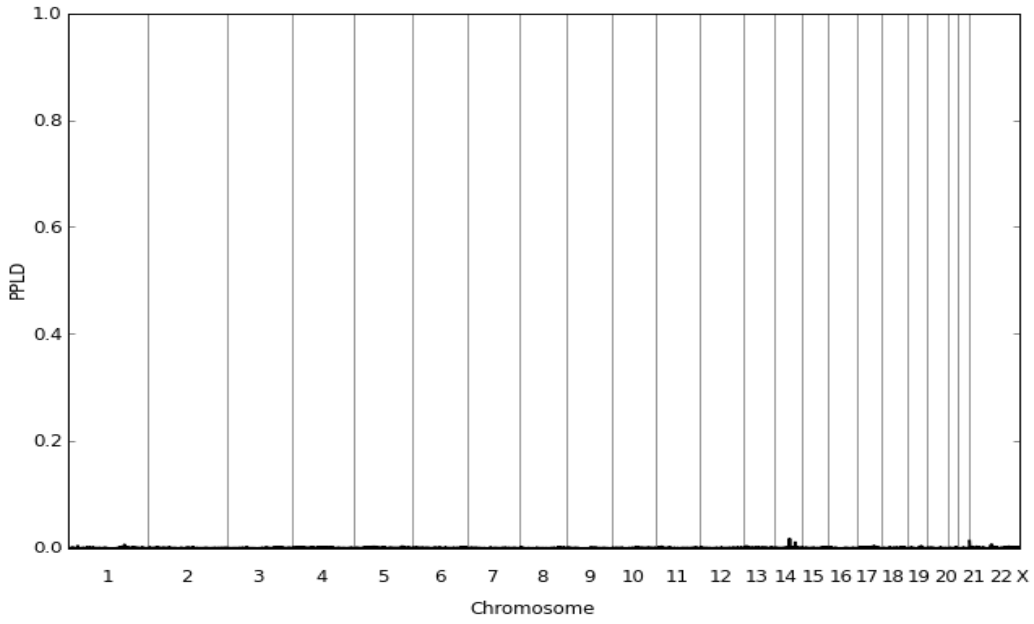
Supplemental Table 2A.4b. Control Analysis: Gene Number (C:GN) results. Each C:GN analysis mimicked the intermediate definition of the NVMSD:ALL linkage regions by containing 645 genes selected at random from the NVMSD:ALL control regions. From the 645 genes input into the core analysis of IPA, an average of 610 genes were mapped in the IPA knowledge base and were included in the core analysis.

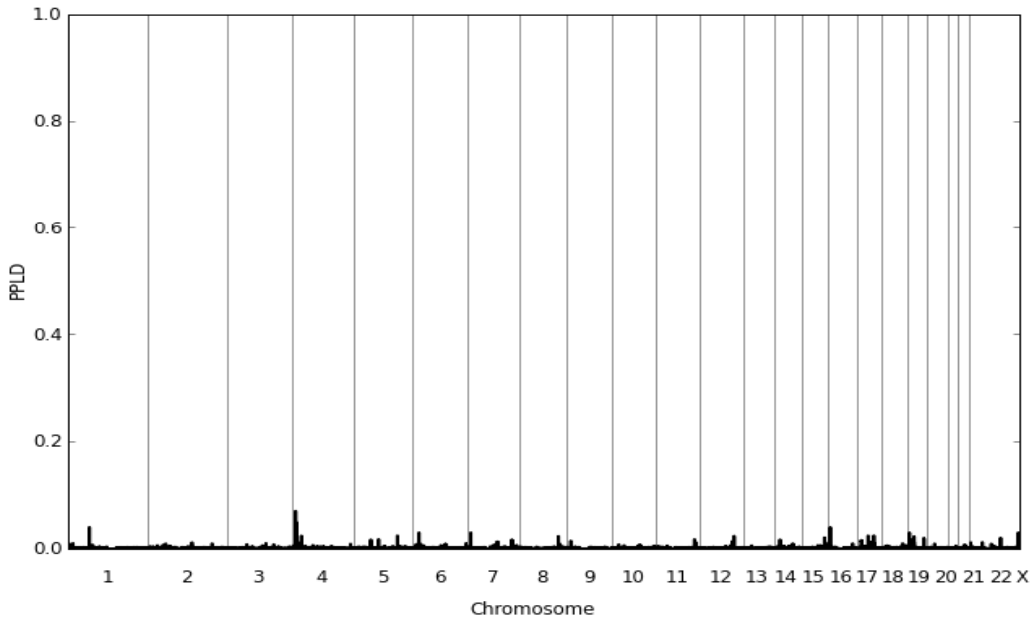
C:GN Results					
Analysis	**Number of Genes in IPA Knowledge Base**	**Number of Relevant Significant Functions**	**Number of Candidate Genes Identified**	**Most Prevalent Functions**	**Genes with Most Functions**
1	612	50	75	Progressive Motor Neuropathy (51) Branching of Neurites (6) Quantity of Brain Cells (6)	GAP43 (9), GLI2 (8), RTN4 (8), NFIA (7)
2	608	55	44	Neurogenesis (22) Outgrowth of Neurites (16) Growth of Neurites (14)	NTF3 (15), NTRK2 (12), GAP43 (8), BMP4 (7), CD9 (5)
3	614	28	37	Neurogenesis (21) Guidance of Axons (10)	NTF3 (7), GAP43 (6), GFRA3 (6), IL18 (5)
4	613	35	87	Progressive Motor Neuropathy (49) Parkinsons's (29) Neurogenesis (23)	GAP43 (6), CD9 (5), GFRA3 (5), IL18 (5)
5	606	41	69	Huntington's disease (28) Development of Nervous System (22) Development of CNS (21) Development of Brain (16) Synaptic Transmission (16)	DRD2 (11), ZIC1 (10), EN1 (8), DRD1 (6), CASP3 (6), FGFR3 (6), GFRA3 (5), IFT88 (5), SMAD9 (5), SMAD1 (5)
6	614	46	81	Huntington's Disease (31) Neurogenesis (22) Neurotransmission (17) Synaptic Transmission (16) Quantity of Neurons (10)	NTF3 (32), NTRK2 (19), HTT (13), BMP4 (9), ERBB3 (6), GJD2 (6), P2RX7 (6), CCL21 (5), GDF11 (5), IFNG (5)
7	614	60	52	Huntington's Disease (29) Synaptic Transmission (17)	NTF3 (9), BMP4 (7), ZIC1 (5)
8	608	28	32	Synaptic Transmission (14)	NCAM1 (5), SRC (5)
9	604	25	40	Development of Nervous System (17) Development of CNS (16) Development of Brain (12)	NTF3 (17), NTRK2 (12), OTX1 (7)
10	609	39	54	Schizophrenia (23) Developmental Process of Neurons (15) Neurotransmission (14) Synaptic Transmission (13)	DRD1 (5), GLI2 (5)

Supplemental Table 2A.4c: Control Analysis: Gene Number Genome (C:GNG) results. Each C:GNG analysis mimicked the intermediate definition of the NVMSD:ALL linkage regions by containing 645 genes selected at random from the genome. Similar to the C:GN analyses, an average of 610 genes were mapped in the IPA knowledge base and were used for each core analysis.

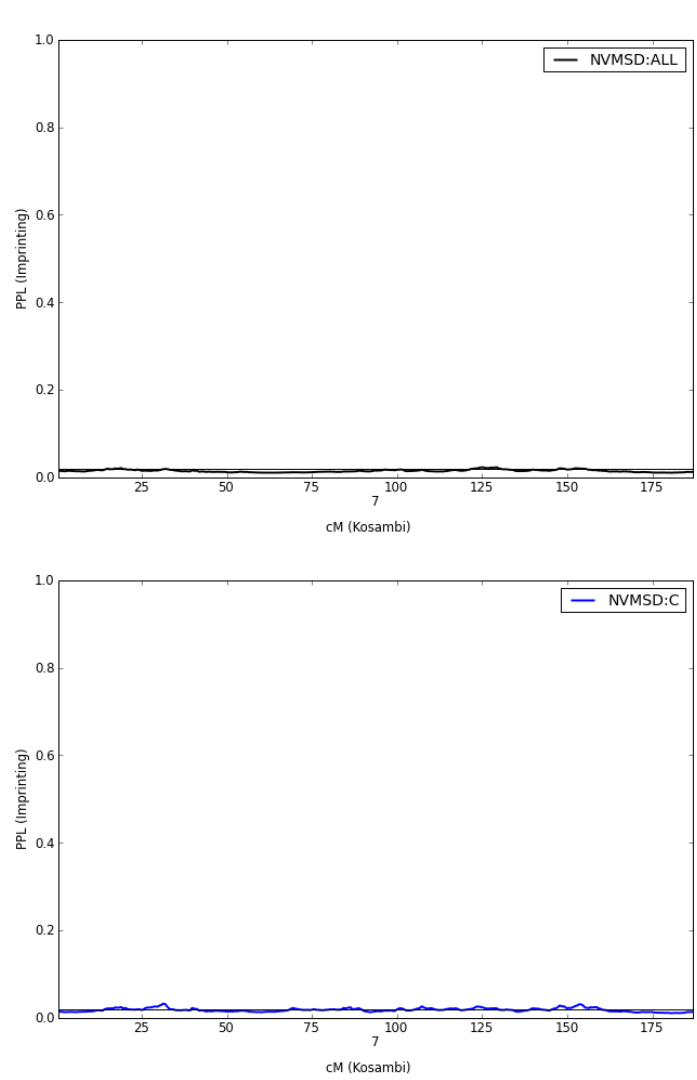
C:GNG Results					
Analysis	Number of Genes in IPA Knowledge Base	Number of Relevant Significant Functions	Number of Candidate Genes Identified	Most Prevalent Functions	Genes with Most Functions
1	610	95	77	Bipolar Affective Disorder (46) Survival of Neurons (8) Myelination (6) Development of Nervous Tissue (6) Survival of Motor Neurons (6) Obsessive-Compulsive Disorder (5) Learning by Mice (5) Neurological Process of Tissue (5) Cell Death of Neuroglia (5)	BDNF (59), NTF3 (37), LIF(13), NRP2 (18), SKI (8), OTX1 (6)
2	606	42	75	Neurodegenerative disorder (40) Alzheimer's disease (38) Cell Death of Neurons (17) Development of Central Nervous System (15) Apoptosis of Neurons (13) Cell Death of Sympathetic Neuron (5)	NTF3 (13), TP53 (13), BCL2 (10), EN1 (8), IL8RB (7), GRIN2B (7), ESR2 (6)
3	624	78	49	Development of Nervous System (17) Growth of Neurites (15) Outgrowth of Neurites (14) Development of Neurites (11) Morphogenesis of Neurites (9) Guidance of Axons (8) Development of Cerebral Cortex (6) Extension of Neurites (6) Growth of Axons (6) Morphogenesis of Dendrites (6) Branching of Neurites (5)	NRP1 (11), NEFL (10), POU4F3 (9), OTX2 (8), SEMA5A (7), ULK1 (6), ACTR3 (5)
4	608	42	30	Survival of Motor Neurons (5) Down's Syndrome (5)	ASCL1 (6), AGRN (5), HGF (5), KCNJ10 (5), NTF4 (5)
5	612	31	104	Psychological Disorder (72) Bipolar Affective Disorder (54) Neurogenesis (21) Developmental Process of Neurons (17) Developmental Process of Neuroglia (8)	IGF1 (17), NRP2 (9), SLIT2 (9), TP53 (7), CAMK2A (5), CTF1 (5)
6	616	21	178	Neurological Disorder (155) Mood Disorder (53) Huntington's disease (32) Neurogenesis (24) Survival of Neurons (9) Development of Neurons (8) Cell Death of Nervous Tissue Cell Lines (5)	BCL2 (8), NGFR (8), SLC1A3 (7), BAX (6), GABRA1 (6), GABRAG2 (6), TGFB1 (5), FOXG1 (5)
7	609	18	24	Neurological Process of Brain Cells (5)	-
8	613	48	27	-	BDNF (26), CHRM2 (7), GRIK2 (7)
9	612	15	20	-	-
10	618	39	54	Schizophrenia (23) Neurological Disorder of Humans (14) Psychological Disorder of Humans (14) Differentiation of Neurons (11) Schizophrenia in Humans (11) Seizures (8) Guidance of Axons (8)	NGFR (10), PSAP (6), NRP1 (6), GDF7 (5), PNPO (95), VAX1 (5)

Supplemental Figure 2A.1: Genome-wide PPLD results for (a) NVMSD:ALL and (b) NVMSD:C





Supplemental Figure 2A.2: PPL allowing for imprinting across chromosome 7.



Appendix 2: Supplemental Materials for Chapter 3: Follow-up Linkage and Association Analyses of a Nonverbal Motor Speech Phenotype Identified in the AGRE Data Set

Supplemental Methods:

Participants and NVMSD:ALL criteria

Subjects in this study were selected from families in the AGRE database. Each family was required to have two individuals who met the criteria for NVMSD:ALL, regardless of their ASD diagnosis. The NVMSD:ALL phenotype was derived from three variables in the Autism Diagnostic Interview-Revised (ADI-R): LEVELL, CARTIC, and ARTIC5 (Supplemental Figure 1). LEVELL is a measure of an individual's overall level of language abilities. Individuals who lack functional language received a '2' for this variable and were considered affected for the NVMSD:ALL phenotype. Additionally, individuals who scored a '2' or '3' for CARTIC or ARTIC5 have poor speech intelligibility and are also considered to be affected for NVMSD:ALL. Individuals had to be at least 2 years old to have a LEVELL score and had to be at least 4 years old to have a CARTIC score or 5 years old to have an ARTIC5 score. A total of 87 families met the inclusion criteria for NVMSD:ALL. Additionally, 34 individuals from the original 203 NVMSD:ALL families who did not have Affymetrix 5.0 genotype data available at the time of the study were included with the 87 families used in the sequential update of the PPL analysis.

Candidate Gene Selection for Fine Mapping

A total of 645 genes (UCSC Genome Browser, NCBI Build 36.1) were identified within the linkage regions from the original PPL analysis. Since no association was detected in these regions, candidate genes were selected from this list for fine mapping and further association analysis. In order to prioritize the list of 645 genes, Ingenuity Pathway Analysis – IPA (Ingenuity

Systems, www.ingenuity.com) was used to categorize the functions of these genes. The Core Analysis was run using the default settings and excluded uncategorized species. Genes categorized within 'Nervous System Development and Function', 'Neurological Disorders', and 'Psychological Disorders' were extracted and considered in candidate prioritization. A total of 53 genes were prioritized using these functions and were further sorted by PPL value. The final list of candidate genes for fine mapping included the following genes: *CDK5R1*, *EIF4E, GHR, LMX1A, NF1, NCOR1, NKX6-1, OMG, PMP22, TRPV2,* and *UBB*.

Determination of Sample Power

To assess the power of the sample to detect association, KELVIN was run using genotypes simulated by SLINK [Weeks et al., 1990]. A haplotype of two SNPs, the trait marker and another SNP in LD with the disease marker (D' = 0.80), was simulated for the NVMSD:ALL dataset. To assess all possible disease models, the penetrances of each genotype were derived by varying the disease allele risk and the minor allele frequency (MAF) of the disease marker. To model a semi-recessive model, the disease risk was set to 1 for the first allele and was varied from 1 to 10 in single increments for the second allele. As the second allele risk was increased, the penetrance of the homozygous disease trait was increased (Supplemental Table A2.1A). For the semi-dominant model, the disease risks for each allele were identical and were varied from 1 to 10 in single increments (Supplemental Table A2.1B). Disease risk of 1-2 signifies modest risk, 3-4 signifies moderate risk, 5-7 signifies medium risk, and 8-10 signifies high risk [Nussbaum et al., 2007; Coriell Personalized Medicine Collaborative, 2013]. The disease gene frequency (DGF) of the disease marker was simulated at percentages of 1, 2, 3, 4, 5, 10, 15, 20 and 30. The penetrances are also a function of the disease prevalence. All parameters were calculated using two different disease prevalences: the prevalence of NVMSD:ALL in AGRE (11.4%) and an estimation of the prevalence of NVMSD:ALL in the general population (0.46%). The prevalence

of NVMSD:ALL in the AGRE database was based on the percentage of individuals who met the criteria for the NVMSD:ALL phenotype based on available ADI-R data. The estimation of the prevalence of NVMSD:ALL in the general population was derived from the percentage of individuals with an autism diagnosis in AGRE who met the criteria for NVMSD:ALL (46%) and the CDC prevalence of ASD in the general population (1%). To model sample heterogeneity, the percentage of unlinked families was varied from complete linkage (0% unlinked families) to no linkage (100% unlinked families) in increments of 10%. Each set of parameters was replicated 100 times and was analyzed for association using the KELVIN software package. A previous simulation study of 10,000 replicates of unlinked data only 9 of the replicates yielded PPL values above 25% [Logue et al., 2003]. For comparability sake, the PPLD is set to the same scale as the PPL [Wratten et al., 2009], therefore we used 25% as our cut-off for positive evidence for association.

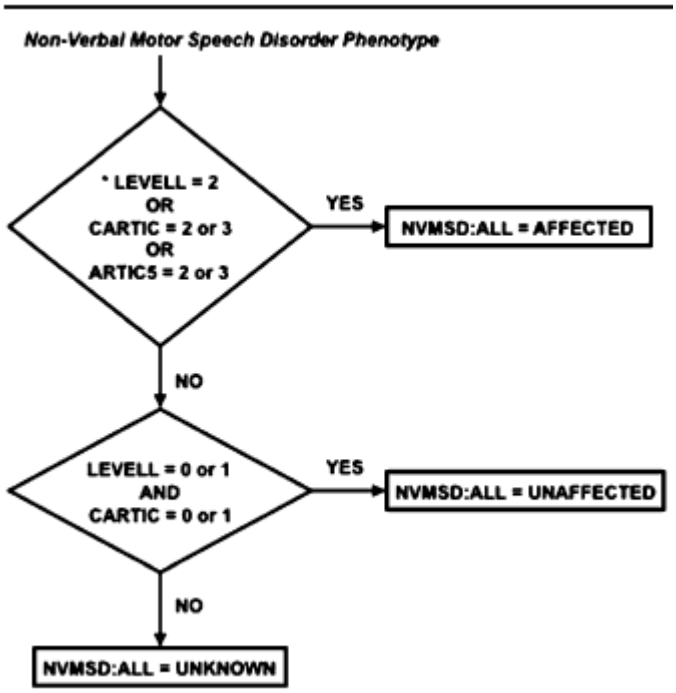
Supplemental Results:

Results of Determination of Sample Power

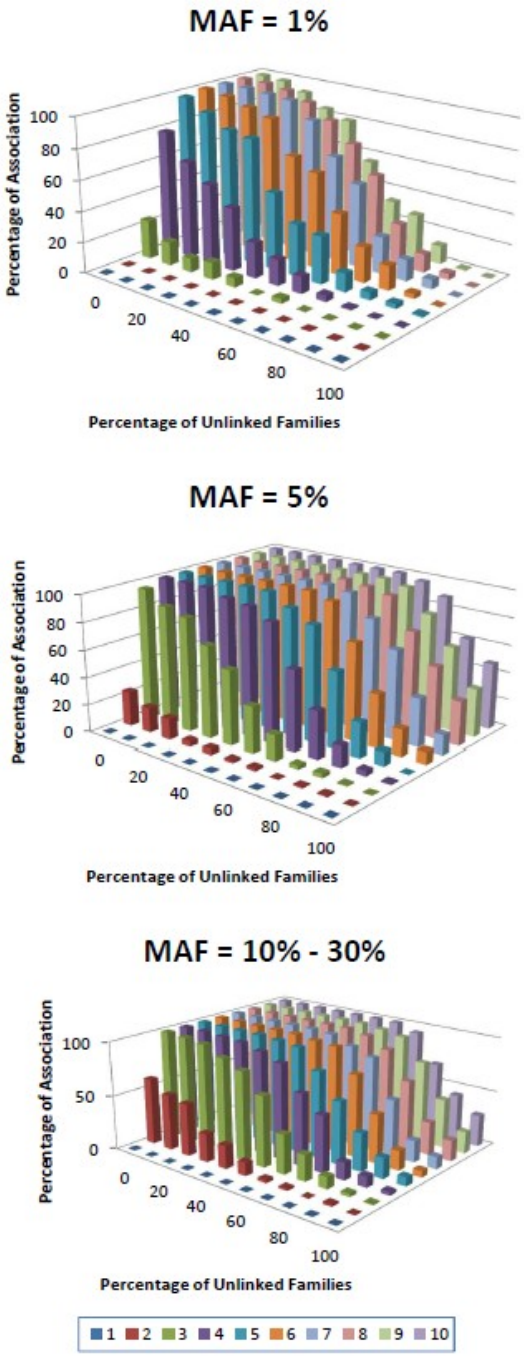
Using the prevalence of NVMSD:ALL in the AGRE population, simulations indicate that this sample has the power to detect association using both a semi-dominant and semi-recessive model. As expected, the NVMSD:ALL dataset does not have sufficient power to detect association for dominant and recessive models with low disease risks and models with completely unlinked families (Supplemental Figures A2.2 and A2.3). The difference in power to detect association between dominant and recessive models is dependent on DGF of the disease marker. There is sufficient power to detect association for DGFs above 5% using the dominant models (Supplemental Figure A2.2) and above 10% for the recessive models (Supplemental Figure A2.3). Overall, the power to detect association for disease markers with lower DGFs is more dramatically affected by a high percentage of unlinked families and low disease allele risk.

These results are replicated using the estimate of the NVMSD:ALL phenotype in the general population (data not shown).

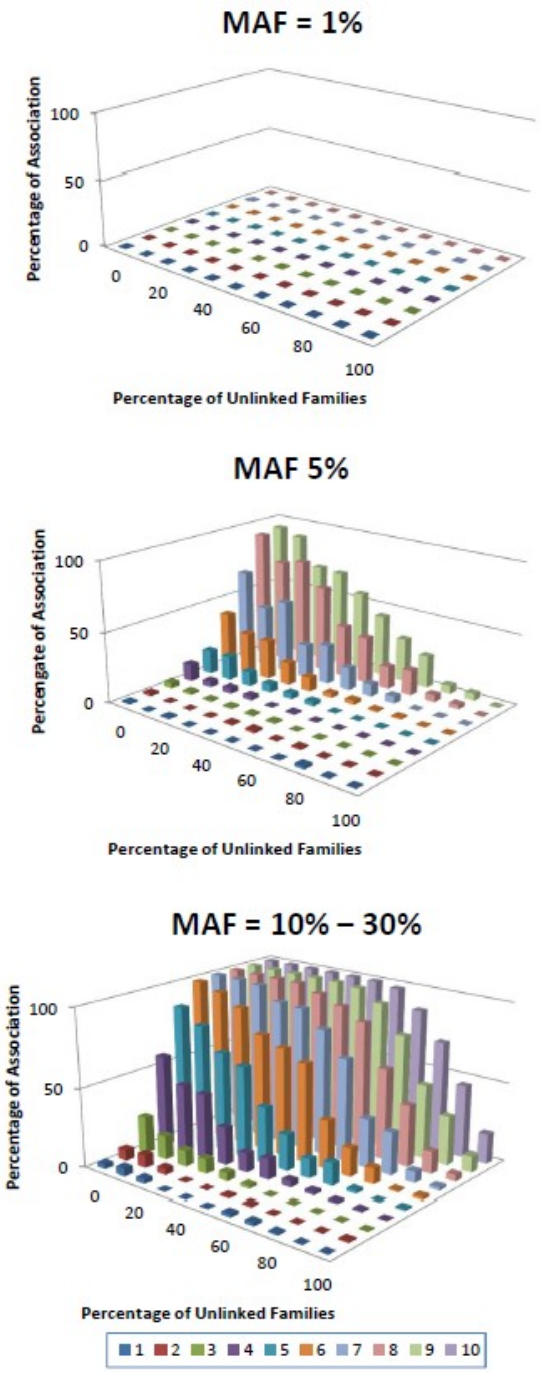
The power to detect association was tested for the maximized model of each linkage region (Table 4, Supplemental Figure 4). There is minimal power to detect association given the parameters for *EIF4E*, *GHR, PMP22, UBB, TRPV2, NCOR1, NF1/OMG,* and *CDK5R1* due to the low penetrance of the disease trait. The gene with the maximized model that is most likely to detect association in this dataset is *NKX6-1* (Supplemental Figure A2.4). The power to detect association in the NVMSD:ALL dataset is moderate for the maximized *LMX1A* model parameters. While the NVMSD:ALL sample has strong power to detect association for a dominant model at a disease marker of 1%, the power to detect association for a dominant model with 99% penetrance is reduced in this dataset due to the lack of parents affected for NVMSD:ALL (Supplemental Figures A2.2, A2.3, and A2.4).



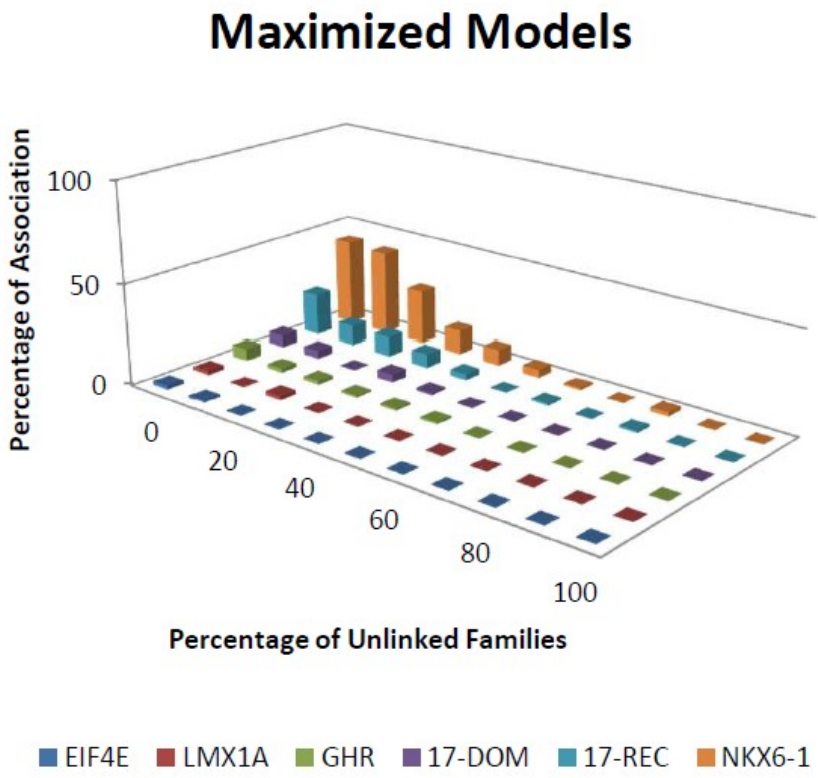
Supplemental Figure A2.1: Diagnostic Criteria for NVMSD:ALL phenotype using variables from the ADI-R. Individuals who lacked functional language (LEVELL = 2) or had poor speech intelligibility (CARIC/ARTIC5 = 2 or 3) were considered to be Affected for the NVMSD:ALL phenotype. Individuals who had functional language (LEVELL = 0 or 1) and intelligible language (CARIC/ARTIC5 = 0 or 1) were considered to be Unaffected for the NVMSD:ALL phenotype. Individuals with ADI-R data who lacked these variables were considered to be Unknown for the NVMSD:ALL phenotype.



Supplemental Figure A2.2: Determination of power to detect association in the NVMSD:ALL sample for semi-dominant models. For all figures, the Y-axis represents the percentage of replicates that produced PPLD values above 25%, the X-axis represents the percentage of UNLINKED families in each replication, and the Z axis represents increasing penetrance values for the disease trait, P(DD) and P(Dd). A) Simulated replicates using a MAF of 1% for the disease marker. B) Simulated replicates using a MAF of 5% for the disease marker. C) Simulated replicates using a MAF of 10, 15, 20, 25, and 30% for the disease marker.



Supplemental Figure A2.3: Determination of power to detect association in the NVMSD:ALL sample for semi-recessive models. For all figures, the Y-axis represents the percentage of replicates that produced PPLD values above 25%, the X-axis represents the percentage of UNLINKED families in each replication, and the Z axis represents increasing penetrance values for the disease trait, P(DD). A) Simulated replicates using a MAF of 1% for the disease marker. B) Simulated replicates using a MAF of 5% for the disease marker. C) Simulated replicates using a MAF of 10, 15, 20, 25, and 30% for the disease marker.



Supplemental Figure A2.4: Determination of power to detect association in the NVMSD:ALL sample for maximized models for each linkage region. Each set of replicates was simulated using the maximized model for its respective linkage region. *EIF4E, LMX1A, GHR,* and *NKX6-1* had unique maximized models and are represented by their gene name. *PMP22, UBB, TRPV2,* and *NCOR1* are located within the same linkage region and are represented by the maximized model called '17_DOM'. *NF1, OMG* and *CDK5R1* are represented by the '17_REC' maximized model. The Y-axis represents the percentage of replicates that produced PPLD values above 25%, the X-axis represents the percentage of UNLINKED families in each replication, and the Z axis represents increasing penetrance values for the disease trait, P(DD) and P(Dd).

Supplemental Table A2.1: Disease model parameters used in genotype data simulations using the AGRE prevalence.

The penetrances used for each model are listed for DGF equal to 1, 5, 10, 20, and 30%. The penetrance for a homozygous disease allele is P(DD), heterozygous is P(Dd), and homozygous non-disease allele is P(dd). Table 1A lists the recessive model parameters and Table 1B lists the dominant model parameters. Models in red have more than 80% power to detect association.

Supplemental Table 1A: Semi-Recessive Models

	Disease Gene Frequency														
	1%			5%			10%			20%			30%		
Disease Risk	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)
1	0.17	0.17	0.11	0.17	0.17	0.11	0.16	0.16	0.11	0.15	0.15	0.10	0.14	0.14	0.09
2	0.23	0.17	0.11	0.22	0.17	0.11	0.21	0.16	0.11	0.19	0.15	0.10	0.18	0.13	0.09
3	0.34	0.17	0.11	0.33	0.17	0.11	0.31	0.16	0.11	0.28	0.14	0.09	**0.25**	**0.13**	**0.08**
4	0.46	0.17	0.11	0.44	0.16	0.11	0.41	0.15	0.10	**0.36**	**0.14**	**0.09**	**0.31**	**0.12**	**0.08**
5	0.57	0.17	0.11	0.55	0.16	0.11	**0.51**	**0.15**	**0.10**	**0.44**	**0.13**	**0.09**	**0.37**	**0.11**	**0.07**
6	0.69	0.17	0.11	0.66	0.16	0.11	**0.61**	**0.15**	**0.10**	**0.51**	**0.13**	**0.09**	**0.42**	**0.10**	**0.07**
7	0.8	0.17	0.11	0.76	0.16	0.11	**0.71**	**0.15**	**0.10**	**0.58**	**0.12**	**0.08**	**0.46**	**0.10**	**0.07**
8	0.92	0.17	0.11	**0.87**	**0.16**	**0.11**	**0.80**	**0.15**	**0.10**	**0.64**	**0.12**	**0.08**	**0.50**	**0.09**	**0.06**
9	-	-	-	**0.98**	**0.16**	**0.11**	**0.89**	**0.15**	**0.10**	**0.71**	**0.12**	**0.08**	**0.54**	**0.09**	**0.06**
10	-	-	-	-	-	-	**0.98**	**0.15**	**0.10**	**0.76**	**0.11**	**0.08**	**0.57**	**0.09**	**0.06**

Supplemental Table 1B: Semi-Dominant Models

	Disease Gene Frequency														
	1%			5%			10%			20%			30%		
Disease Risk	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)	P(DD)	P(Dd)	P(dd)
1	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.12	0.12
2	0.23	0.23	0.11	0.21	0.21	0.11	0.19	0.19	0.10	0.17	0.17	0.09	0.15	0.15	0.08
3	0.33	0.33	0.11	**0.29**	**0.29**	**0.10**	**0.25**	**0.25**	**0.08**	**0.20**	**0.20**	**0.07**	**0.17**	**0.17**	**0.06**
4	0.44	0.44	0.11	**0.36**	**0.36**	**0.09**	**0.30**	**0.30**	**0.07**	**0.22**	**0.22**	**0.06**	**0.18**	**0.18**	**0.05**
5	**0.54**	**0.54**	**0.11**	**0.42**	**0.42**	**0.08**	**0.33**	**0.33**	**0.07**	**0.24**	**0.24**	**0.05**	**0.19**	**0.19**	**0.04**
6	**0.63**	**0.63**	**0.11**	**0.47**	**0.47**	**0.08**	**0.36**	**0.36**	**0.06**	**0.25**	**0.25**	**0.04**	**0.20**	**0.20**	**0.03**
7	**0.73**	**0.73**	**0.10**	**0.51**	**0.51**	**0.07**	**0.38**	**0.38**	**0.05**	**0.26**	**0.26**	**0.04**	**0.20**	**0.20**	**0.03**
8	**0.81**	**0.81**	**0.10**	**0.55**	**0.55**	**0.07**	**0.40**	**0.40**	**0.05**	**0.26**	**0.26**	**0.03**	**0.20**	**0.20**	**0.03**
9	**0.90**	**0.90**	**0.10**	**0.59**	**0.59**	**0.07**	**0.41**	**0.41**	**0.05**	**0.27**	**0.27**	**0.03**	**0.21**	**0.21**	**0.02**
10	0.98	0.98	0.10	**0.62**	**0.62**	**0.06**	**0.43**	**0.43**	**0.04**	**0.27**	**0.27**	**0.03**	**0.21**	**0.21**	**0.02**

Appendix 3: Supplemental Material for Chapter 4: Characterizing Language Profiles in Families Ascertained for Autism and Specific Language Impairment

ANOVA comparison of LI ONLY, RI ONLY, LI+RI, and ASD

The LI and RI groups were not mutually exclusive since there were impaired family members who met criteria for both oral and written language impairments. The RI group was then divided into those individuals who met criteria for just written language impairment (RI ONLY, n = 43) and those who met criteria for both oral and written language, (LI+RI, n = 34). A one-way ANOVA test with pre-specified contrasts was used to compare the mean scores of the the LI ONLY, RI ONLY, LI+RI, and ASD groups. A Bonferroni correction for multiple testing and post-hoc Tukey's test were conducted in SPSS.

When RI ONLY family members were compared to family member with ASD, the RI ONLY group performed significantly better on several segmental language measures and most of the higher order language measures. The RI ONLY group performed significantly better on Formulated Sentences (RI ONLY = 11.0, ASD = 5.7, $p < 0.001$); Recalling Sentences (RI ONLY = 8.7, ASD = 4.6, $p < 0.001$); Word Classes -Receptive (RI ONLY = 10.9, ASD = 7.3, $p < 0.001$); Word Classes- Expressive (RI ONLY = 9.9, ASD = 5.9, $p < 0.001$); Ambiguous Sentences (RI ONLY = 91.4, ASD = 77.9, $p < 0.001$); Inference (RI ONLY = 99.1, ASD = 70.1, $p < 0.001$); Meaning from Context (RI ONLY = 96.4, ASD = 71.1, $p < 0.001$); Non-literal Language (RI ONLY = 94.6, ASD = 70.8, $p < 0.001$, Pragmatic Judgment (RI ONLY = 96.9, ASD = 70.3, $p < 0.001$). However, when the ASD probands were compared to those language probands who were affected for LI+RI, results were strikingly different. The ASD group and the LI+RI groups only differed on the single word reading task with the ASD group performing significantly better than the LI+RI group (ASD = 99.1, LI+RI = 82.7, $p < 0.005$).

Analysis of the effect of familiality on score differences between ASD and SLI probands

For each phenotypic variable, we created a derived variable of the absolute pairwise differences among pairs of family members (N = 55) where one member was diagnosed as having SLI and the other with ASD. We used GWAS data (see Chapter 6; [Bartlett et al., 2013]) to estimate the proportion of SNPs identical by descent, denoted pi_hat. Primarily because of (possibly cryptic) consanguinity, these estimates vary from textbook values of 0.5 for siblings, 0.125 for cousins, and so on. We then used a mixed effects linear model to regress the absolute differences of the phenotypes in a pair against pi_hat, allowing for inherent correlation of differences among overlapping pairs within a family. Negative slopes would indicate that the more closely related a pair was the closer their phenotype scores independent of SLI or ASD categorization. Not all variables were assessed for each individual. We used the entire NJLAGS release 1.0 dataset to impute missing variables with the mice package in R [van Buuren and Groothuis-Oudshoorn, 2011]. Mice uses multivariate Gibbs sampling for imputation. For each variable, individuals with cross-imputation variability exceeding 1 standard deviation of the variable in the original data were discarded. Separate regressions of 10 imputations were combined; both within and across imputation variability was used to calculate standard errors [Rubin, 1987]. None of the phenotypic variables produced slopes that differed significantly from zero suggesting that familiality does not have a significant effect on score differences between SLI and ASD probands.

Supplemental Table 3A.1: NJLAGS Testing Battery including subtest descriptions and age ranges of tests.

Language & Cognitive Battery	Function	Age
Clinical Evaluation of Language Fundamentals- (CELF-4)- Core Score		
Concepts and Following Directions	Follow oral directions of increasing complexity	5+
Word Structure	Complete a sentence with the correct morphological inflection	5-8
Recalling Sentences	Imitation of sentences of increasing length and complexity	5+
Formulated Sentences	Create a meaningful sentence using a target word or phrase	5+
Word Classes	Identify two related words and explain	9+
Word Definitions	Define words with appropriate detail	13+
Clinical Evaluation of language Fundamentals-Preschool		
Sentence Structure	Interpret sentences of increasing length and complexity	3-5
Word Structure	Pronoun usage and complete sentences with the correct morphological inflection	3-5
Expressive Vocabulary	Label people, places, and actions	3-5
Comprehensive Assessment of Spoken Language (CASL)		
Non-literal language	Understand indirect and figurative language and sarcasm	7+
Meaning from Context	Infer meaning from linguistic context	11+
Inference	Use real world knowledge infer meaning	7-17
Ambiguous Sentences	Recognize ambiguity and verbalize it	11+
Pragmatic Judgment	Knowledge and use of pragmatic rules	3+
Comprehensive Test of Phonological Processing (CTOPP)		
Elision	Delete syllables and sounds from words to create new words	5+
Non-Word Repetition	Repeat nonsense words of increasing length	5+
Gray Oral Reading Tests (GORT-4)		
Rate	Timed oral reading	6+
Accuracy	Number or errors while reading	6+
Fluency	Combined score of rate and accuracy	6+
Comprehension	Answer questions about what was read	6+
Woodcock Reading Mastery Tests-Revised		
Word identification	Single word sight reading	6+
Word Attack	Nonsense word reading	6+
Wide Range Achievement Test 3 (WRAT3)		
Spelling	Spelling of dictated words	6+
Wechsler Abbreviated Test of Intelligence (WASI)		
Performance IQ	Composite non-verbal subtest scores: Block Design and Matrices	6+
Social Reciprocity Scale (SRS)		
Social Function	Assesses social functioning and deficits	All
Yale-Brown Obsessive Compulsive Scale (YBOCS)		
Adult Version	Rates severity of obsessive and compulsive symptoms	18+
CYBOCS	Child Version of YBOCS	< 18
PDD-CYBOCS	Rates obsessive symptoms in ASD probands	< 18

Supplemental Table 3A.2: Means and Standard Deviation of Language Test Scores for ASD, LI, RI and Unaffected individuals in the NJLAGS sample.

SEGMENTAL LANGUAGE								
	UNAFFECTED		ASD		LI		RI	
	N	MEAN (SD)	N	MEAN (SD)	N	MEAN (SD)	N	MEAN (SD)
CELF - Core Score	125	110.75 (9.41)	21	73.43 (21.90)	30	71.77 (12.61)	55	90.56(19.15)
CELF - Concepts and Following Directions	21	10.71 (2.37)	13	5.54 (4.45)	18	4.11 (2.65)	15	5.47 (3.54)
CELF - Formulated Sentences	126	11.85 (1.85)	23	5.70 (3.93)	32	5.61 (2.81)	58	9.17 (3.51)
CELF - Word Structure	7	10.57 (2.37)	8	5.12 (4.26)	14	4.07 (3.13)	8	4.50 (4.07)
CELF - Recalling Sentences	133	10.44 (2.35)	25	4.64 (3.39)	33	4.31 (2.23-5)	61	6.92 (3.43)
CELF - Word Classes Expressive Score	125	11.71 (2.32)	16	5.87 (3.18)	20	5.70 (2.36)	54	8.69 (2.96)
CELF - Word Classes Receptive Score	125	12.35 (1.42)	16	7.31 (3.22)	20	6.60 (2.91)	54	9.72 (2.83)
CELF - Word Classes Total Score	125	12.12 (1.76)	16	6.31 (2.89)	20	5.95 (2.34)	54	9.24 (2.91)
CELF - Word Definitions	112	13.29 (1.99)	11	7.94 (3.13)	17	7.82 (3.07)	46	10.51 (3.27)
CELF - Preschool Standard Score	7	100.00 (10.46)	3	70.67 (22.37)	2	77.00 (8.49)	-	-
CELF - Preschool Expressive Vocabulary	7	11.71 (2.50)	3	6.33 (4.73)	2	5.50 (2.12)	-	-
CELF - Preschool Sentence Structure	7	9.29 (1.89)	3	5.33 (3.79)	2	7.00 (0)	-	-
CELF - Preschool Word Structure	7	9.00 (2.77)	3	3.67 (3.06)	2	5.50 (2.12)	-	-
HIGHER ORDER LANGUAGE								
CASL - Ambiguous Sentences	117	100.62 (13.22)	11	77.91 (12.49)	19	74.00 (7.99)	46	86.21 (14.13)
CASL - Inference	26	100.92 (8.61)	10	70.10 (14.95)	10	74.50 (16.20	17	88.18 (20.12)
CASL - Meaning from Context	119	105.69 (10.75)	11	71.09 (16.31)	18	79.06 (8.73)	44	91.04 (12.30)
CASL - Nonliteral Language	123	101.24 (13.48)	20	70.80 (20.85)	28	79.07 (15.55)	57	88.76 (17.15)
CASL - Pragmatic Judgment	133	100.02 (9.72)	25	70.32 (20.35)	36	78.94 (14.90)	63	88.95 (17)
PHONOLOGICAL PROCESSING								
CTOPP - Elision	130	10.08 (2.46)	21	8.14 (3.89)	28	5.33 (3.10)	59	7.07 (3.50)
CTOPP - Nonword Repetition	125	9.18 (2.42)	23	8.30 (2.72)	28	6.81 (2.84)	57	8.00 (3.15)
WRITTEN LANGUAGE								
WRAT - Spelling	131	107.48 (11.63)	23	95.78 (17.74)	31	89.61 (16)	60	93.00 (17.15)
Woodcock - Word Attack	130	111.44 (10.78)	24	98.67 (18.83)	31	89.87 (13.81)	63	94.89 (15.67)
Woodcock - Word Identification	132	110.14 (10.61)	25	99.12 (20.99)	33	88.21 (18.44)	63	91.88 (15.87)
GORT4 - Accuracy	114	13.08 (13.08)	17	6.94 (3.77)	23	6.45 (3.54)	57	7.91 (3.95)
GORT4 - Comprehension	111	10.65 (2.05)	16	5.62 (3.61)	21	5.75 (3.11)	58	6.28 (2.62)
GORT4 - Fluency	114	13.92 (2.87)	17	6.59 (3.57)	23	5.50 (3.60)	57	7.35 (4.84)
GORT4 - Rate	114	11.88 (1.82)	17	7.35 (3.0)	23	6.41 (3.14)	57	7.79 (3.50)
GORT4 - Overall Reading Quotient	109	114.62 (11.65)	16	77.12 (19)	21	74.35 (15.56)	56	80.98 (15.31)

WASI - Performance IQ	129	109.42 (14.98)	28	93.89 (15.45)	33	89.58 (12.46)	62	99.19(14.84)

Supplementary Table 3A.3: T-test comparisons of oral and written language scores for individuals with ASD (who are verbal) and individuals with LI or RI. Comparisons that are significant after a Bonferroni correction are shown in red.

Segmental Language									
				LI vs ASD			RI vs ASD		
Subtest	Mean ASD (s.d.)	Mean LI (s.d.)	Mean RI (s.d.)	Difference	95% CI	Nom. p- value	Difference	95% CI	Nom. p- value	Comparison
CLFCFD	5.54 (4.45)	4.11 (2.65)	5.47 (3.70)	-1.67	(-3.82, 0.48)	0.316	0.04	(-2.48, 2.56)	0.963	
CLFFS	**5.70 (3.92)**	5.61 (2.81)	**9.17 (3.54)**	0.01	(-1.27, 1.29)	0.932	**3.57**	**(2.22, 4.91)**	**<.001**	**RI > ASD**
CLFWS	5.12 (4.26)	4.07 (3.12)	4.50 (4.07)	-1.11	(-3.98, 1.76)	0.522	0.05	(-3.89, 3.99)	0.769	
CLFRS	**4.64 (3.39)**	4.31 (2.25)	**7.02 (3.49)**	-0.33	(-1.71, 1.06)	0.679	**2.49**	**(1.15, 3.83)**	**0.005**	**RI > ASD**
CLFWCT	**6.31 (2.89)**	5.95 (2.34)	**9.24 (2.91)**	-0.37	(-1.75, 1.02)	0.683	**2.92**	**(1.6, 4.25)**	**0.001**	**RI > ASD**
CLFWD	**7.73 (3.17)**	7.94 (3.13)	**10.51 (3.27)**	0.25	(-1.6, 2.09)	0.866	**2.78**	**(1.11, 4.45)**	**0.013**	**RI > ASD**
Higher Order Language										
CASLAS	77.91 (12.49)	74.00 (7.99)	86.21 (14.13)	-2.96	(-12.63, 6.71)	0.368	9.46	(0.63, 18.29)	0.079	
CASLIN	**70.10 (14.95)**	74.50 (16.20)	**88.18 (20.12)**	3.7	(-6.81, 14.21)	0.536	**18.72**	**(7.69, 29.75)**	**0.021**	**RI > ASD**
CASLMC	**71.09 (16.31)**	79.06 (8.73)	**91.04 (12.30)**	7.97	(-0.83, 16.77)	0.104	**19.95**	**(11.99, 27.91)**	**<.001**	**RI > ASD**
CASLNL	**70.80 (20.99)**	79.07 (15.55)	**88.76 (17.15)**	8.27	(-0.7, 17.24)	0.127	**17.96**	**(9.92, 26)**	**<.001**	**RI > ASD**
CASLPJ	**70.32 (20.25)**	78.94 (14.90)	**88.95 (16.86)**	8.62	(2.16, 15.09)	0.078	**18.95**	**(12.64, 25.26)**	**<.001**	**RI > ASD**
Phonological Processes										
C-TOPPEL	**8.14 (3.89)**	**5.33 (3.10)**	7.07 (3.50)	**-2.74**	**(-4.4, -1.09)**	**0.008**	-0.98	(-2.44, 0.48)	0.242	**ASD > LI**
C-TOPPNR	8.30 (2.72)	6.81 (2.84)	8.02 (3.13)	-1.25	(-2.65, 0.14)	0.066	0.06	(-1.17, 1.29)	0.701	
Written Language										
GRTACC	6.94 (3.77)	6.45 (3.54)	7.91 (3.95)	-0.49	(-2.67, 1.7)	0.681	0.97	(-0.73, 2.67)	0.372	
GRTCOMP	5.62 (3.61)	5.75 (3.11)	6.28 (2.62)	0.12	(-1.76, 2.01)	0.912	0.65	(-0.67, 1.97)	0.509	
GRTFLU	6.59 (3.57)	5.50 (3.60)	7.35 (4.84)	-1	(-3.59, 1.58)	0.353	0.76	(-1.22, 2.74)	0.484	
GRTRQ	77.12 (19.00)	74.35 (15.56)	80.98 (15.31)	-2.77	(-14.33, 8.78)	0.349	3.86	(-3.75, 11.46)	0.403	
GRTRATE	7.35 (3.00)	6.41 (3.14)	7.79 (3.50)	-0.95	(-2.69, 0.8)	0.633	0.44	(-0.95, 1.83)	0.643	
WRMWA	98.67 (18.83)	89.93 (14.04)	95.05 (15.60)	-8.73	(-16.11, -1.36)	0.056	-3.62	(-9.93, 2.7)	0.363	
WRMWID	**99.12 (20.98)**	**88.12 (18.73)**	91.88 (15.87)	**-9.49**	**(-16.81, -2.16)**	**0.042**	-6.16	(-12.37, 0.05)	0.082	**ASD > LI**
WRAT	95.78 (17.74)	88.97 (15.86)	93.00 (17.01)	-6.02	(-13.72, 1.68)	0.147	-2.78	(-9.5, 3.94)	0.511	
WASI PIQ	93.89 (15.45)	89.81 (12.59)	99.19 (14.84)	-3.79	(-10.9, 3.31)	0.295	5.81	(-0.65, 12.28)	0.078	

Supplemental Table 3A.4. Heritability of Language and Reading Measures with/without ASD as a Covariate.

Assessment	ASD^{+}	ASD^{COVAR}	Δh^2
CELF			
Core Score	0.43 (0.19)	0.53 (0.18)	***-0.10****
Formulating Sentences	0.24 (0.18)	0.24 (0.19)	0
Repeating Sentences	0.62 (0.17)	0.65 (0.16)	-0.03
Word Classes Expressive	0.34 (0.17)	0.37 (0.17)	-0.03
Word Classes Receptive	0.47 (0.20)	0.48 (0.19)	-0.01
Word Classes Total	0.36 (0.20)	0.40 (0.19)	-0.04
Word Definitions	0.64 (0.16)	0.65 (0.15)	-0.01
CASL			
Ambiguous Sentences	0.38 (0.21)	0.46 (0.19)	***-0.08****
Meaning from Context	0.59 (0.17)	0.70 (0.14)	***-0.11****
Non-literal Language	0.39 (0.16)	0.45 (0.16)	***-0.06****
Pragmatic Judgment	0.38 (0.18)	0.44 (0.20)	***-0.06****
CTOPP			
Elision	0.46 (0.23)	0.48 (0.25)	-0.02
Nonword Repetition	0.86 (0.13)	0.86 (0.13)	0
GORT			
Accuracy	0.57 (0.14)	0.56 (0.14)	0
Comprehension	0.36 (0.18)	0.35 (0.18)	0.01
Fluency	0.46 (0.10)	0.46 (0.10)	0
Oral Reading Quotient	0.45 (0.12)	0.45 (0.12)	0
Rate	0.64 (0.15)	0.64 (0.15)	0
WRMT			
Word Attack	0.69 (0.16)	0.69 (0.17)	0
Word Identification	0.81 (0.12)	0.81 (0.13)	0
WASI			
PIQ	0.43 (0.20)	0.43 (0.20)	0
WRAT			
Spelling	0.51 (0.16)	0.53 (0.16)	-0.02

Supplemental Table 3A.5: Effect of Familiality on Differences in Scores.

Test	Subtest	number of pairs	number of imputed scores	slope	p-value
CELF-4	Core Score	11	42	1.59	0.94
	Concepts and Following Directions	5	44	1.25	0.81
	Formulated Sentences	11	32	-0.07	0.99
	Recalling Sentences	14	29	-0.81	0.84
	Word Classes - Expressive	8	34	-2.60	0.44
	Word Classes - Receptive	8	33	-2.40	0.35
	Word Classes - Total	8	40	-1.93	0.50
	Word Definitions	6	35	-0.44	0.90
CASL	Ambiguous Sentences	7	31	2.91	0.86
	Meaning from Context	6	32	-10.91	0.52
	Non-literal Language	9	30	2.73	0.87
	Pragmatic Judgement	13	20	-9.11	0.67
CTOPP	Elision	12	21	1.7	0.59
	Non-word Repetition	13	17	-0.34	0.92
GORT-4	Accuracy	8	38	-8.20	0.12
	Comprehension	7	48	0.78	0.86
	Fluency	8	35	-5.51	0.25
	Reading Quotient	7	43	-10.32	0.64
	Rate	8	34	-4.25	0.30
WRMT	Word Attack	12	26	-15.49	0.43
	Word Identification	14	25	-12.29	0.58
WRAT	Spelling	12	25	-31.36	0.13
WASI	PIQ	17	23	-3.24	0.83

Appendix 4: Supplemental Material for Chapter 5: Gene x Gene Interaction in Shared Etiology of Autism and Specific Language Impairment

The Variance Component Model

A vector of quantitative trait values for a pedigree of n number of subjects is defined as Y=(y1….yn), has an assumed multivariate normal mean μ and variance-covariance matrix **Ω.** To allow for covariates, μ is replaced with μ+β $\mathbf{X}$, where $\mathbf{X}$ is the vector of covariates and β is the vector of coefficients for the covariates. The phenotypic covariance between relatives is the sum of the individual covariances between relatives:

$$\Omega=2\mathbf{\Phi}\sigma g2+\mathbf{I}\sigma e2$$

Where $\mathbf{\Phi}$ is the genetic kinship matrix, $\sigma g2$ is the additive genetic variance, $\mathbf{I}$ is the identity matrix, 2 is the environmental variance. Additional genetics effects may be modeled by adding terms to this equation with the appropriate matrix to define the genetic effects (such as dominance, gene-gene interactions) and the likewise variance component.

Analysis Procedures

Since variance components analysis has a documented sensitivity to violations of multivariate normality, we have applied robust estimation by assuming a multivariate t-distribution as suggested for scores with kurtosis > 1.5 [Allison et al., 1999; Blangero et al., 2000]. The multivariate t-distribution down-weights extreme values relative to the multivariate normal distribution. Note that the robust estimator is not required for parameter estimation, but is necessary for statistically valid hypothesis testing [Rao et al., 1987].

Significance of parameters estimates are determined by comparing the likelihood of the model with heritability estimates constrained between zero and one, versus the likelihood of the model with the heritability constrained to zero. This is a likelihood ratio test with a test statistic distributed as a mixture of a chi-square and a point mass of zero [Self and Liang, 1987]. Standard errors were derived by inversion of Fisher's information matrix.

Appendix 5: Supplemental Material for Chapter 6: A Genome-scan for Loci Shared by Autism Spectrum Disorder and Language Impairment

Families

Criteria for family recruitment for the study included: 1) at least one individual with a diagnosis of autistic disorder with no known genetic cause (i.e. Fragile X or Rett's), 2) at least one other individual who met criteria for specific language impairment (SLI), 3) at least three other family members willing to participate, and 4) English as the primary language of all individuals participating. Prior to behavioral testing, all subjects gave informed consent conforming to the guidelines for treatment of human subjects at Rutgers University. All subjects were tested with a comprehensive neuropsychology battery administered by an experienced psychometrician, speech language pathologist, or psychologist.

Autism proband criteria: To be identified as the Autism Proband, the etiology must be unknown (for example, no Fragile X or Rett's) and the individual was required to meet the cut-off for Autism or Autistic Disorder on at least two of the three following measures (all were administered to all autism probands): 1) Autism Diagnostic Interview-Revised, 2) Autism Diagnostic Observation Scale, 3) Diagnostic and Statistical Manual-IV.

SLI proband criteria: In order to be identified as an SLI proband, a person had to meet the following inclusionary/exclusionary criteria:

1. A core standard score of <= 85 on the age appropriate version of the Comprehensive Test of Language Fundamentals [Wiig et al., 2004; Semel et al., 2003]; or subtest scores of at least one standard deviation below peers on 60% of all language measures plus a significant history of language and reading difficulties as measured by at least 2+ years

of intervention and/or previous childhood diagnosis of language and/or reading impairment.

2. A non-verbal IQ >= 80 on the Wechsler Abbreviated Scale for Intelligence [Weschler, 1999] and always greater than the Core Language Score of the Comprehensive Test of Language Fundamentals.
3. Hearing within normal limits [positive identification of 500 Hz at 30 dB (SPL), and 1000, 2000, and 4000 Hz at 20 dB (SPL)].
4. No motor impairments or oral structural deviations affecting speech or non-speech movement of the articulators as assessed by a speech-language pathologist.
5. No history of autism or frank neurological disorders such as mental retardation, seizure disorder, or brain injury as determined from parental report. When autism spectrum behaviors were suspected upon parental interview or if observed by the Speech Language Pathologist during the language assessment, the ADI-R and ADOS were administered to formally rule out ASD.
6. Native English speaker with English as the primary language spoken at home.

Measures

All family members as well as higher functioning family members with ASD received age appropriate measures of language and reading. See Supplementary Table 5A.2 for a summary of test and subtests. Briefly, the standardized language and reading measures in the battery included:

a. The Clinical Evaluation of Language Fundamentals (CELF-4 and CELF Preschool) [Wiig et al., 2004; Semel et al., 2003], A Core Language Score is derived from 3-4 subtests scaled

scores (age depending) that address areas of language comprehension, expression, and structure.

b. The Comprehensive Assessment of Spoken Language (CASL) [Carrow-Woolfolk, 1999] contains subtests addressing metalinguistic language skills that tap into complex language include abstraction, inference, and also include a subtest that addresses the pragmatic aspects of language. *These areas are of great relevance to older children and adults as well as higher functioning individuals with autism who may be challenged by meaning that cannot be accessed directly through lexical and grammatical information; these areas of language are not assessed by most other standardized language measures.*

c. The Comprehensive Test of Phonological Processing (CTOPP) [Wagner et al., 1999], Elison and Non-word Repetition subtests only. *The Elision subtest was used to measure deletion and phonological manipulation of sounds in words while the Non-word Repetition task measured phonological short-term memory; both have a strong documented relationship with oral language abilities and reading.*

d. Gray Oral Reading Tests (GORT-4) [Wiederholt and Bryant, 2001] assesses oral reading rate, accuracy, and comprehension.

e. The Woodcock Reading Mastery Tests-Revised [Woodcock, 1987], Word Attack and Word Identification subtests only, *Subjects age 6 and older received the Word Attack (non-word reading) subtest consisting of mono- and polysyllabic pseudowords to assess decoding abilities, and the Word Identification subtests, single word reading of real words arranged in order of increasing difficulty.*

f. The Wide Range Achievement Test 3 (WRAT) [Wilkinson, 1984], Spelling subtest only.

Genotyping

DNA samples were obtained in most cases from cell lines established from peripheral blood by the Rutgers University Cell and DNA Repository (RUCDR) as part enrolling subjects into the National Institute of Mental Health Autism Collection. For a few subjects who did not consent to drawing blood or for whom cell lines were not successfully established (N=5), DNA was extracted from saliva using Oragene DNA sample collection kits using the recommended protocol in our lab.

Samples were genotyped using Affymetrix Axiom™ 1.0 arrays by the RUCDR. Genotype calling was conducted on 567,893 SNP genotypes on 440 individuals with the Affymetrix Power Tools software package using the Axiom™ GT1 algorithm, which incorporates a novel modification of the BLRMM-P algorithm. Quality control on SNP genotypes was conducted as described previously [Simmons et al., 2010]. Briefly, SNPs and individuals with high missing rates were first excluded from further analysis (missingness ≥ 0.1). Call rates for saliva DNA samples were slightly lower than compared to cell line derived DNA (98.2% versus 99.4%) but no systematic differences were observed in copy number, sex determination, Mendelian inconsistency rate or other indicators of quality. The linkage markers (also used for ancestry clustering) were chosen from the subset of markers, post quality control, that had minor allele frequency > 0.3, were on average 0.3 cM apart, and had negligible linkage disequilibrium as measured by $r^2 < 0.2$. Marker allele frequencies were estimated using the maximum likelihood option in MERLIN v 1.11 [Abecasis et al., 2002]. Marker-to-markers linkage disequilibrium was estimated using PLINK v1.07 [Purcell et al., 2007] using unrelated persons (founders) in the dataset. These SNPs were used in a relationship checking analysis by RELCHECK [Broman and

Weber, 1998]. After correcting inconsistencies in the sample IDs or analysis files, all SNPs were tested for Hardy-Weinberg equilibrium (and dropped if $p < 0.001$) and Mendelian inconsistencies (dropped if rate > 0.05). Samples with missingness > 0.03 were dropped from further analysis. SNPs with missingness > 0.05 were dropped from further analysis. Two sets of duplicate samples yielded genotype concordance of 99.82% on SNPs from the final quality controlled set. Additionally, 36 samples were also genotyped with an Affymetrix 250k array with 99.76% concordance for 35,094 SNP in common to both arrays. For association analysis, SNPs with minor allele frequency < 0.05 were not included in the analysis.

Statistical Analysis

For all language measures, standardized scores provided with the test were used in downstream analysis. The Social Responsiveness Scale has two forms, a 65-item child form that is standardized using T-Scores and an adult version that has been minimally modified from the child version to make the wording more appropriate but has not been standardized to a T-Score. We therefore used total raw scores assuming both the adult and child data reflect the same underlying quantitative trait. The cut-off for a dichotomous trait social deficit (SRS-DT) was 54 for males and 45 for females which is equivalent to the T-score > 60 criteria used to identify mild impairment in children. Analysis using a graded response model indicated this procedure did not lose appreciable sensitivity compared to item response theory based scoring (Spearman's rho = 0.99). The Y-BOCS had three versions of the test by age and diagnostic status (adult, child and PDD). The PDD scale has only half the number of items as the other scales. We used percent of the maximum possible score for each scale as a single quantitative trait.

Derivation of phenotypic factor scores for quantitative trait analysis. Factor analysis of 21 phenotypic measures stratified across age bands (N=19 for ages 5-8; N=20 for ages 9-13;

N=19 for ages 13+) as some language measures are only given at certain ages, though N=17 are administered to all subjects. We used the matrix of genetic correlations from SOLAR v4.3 [Almasy and Blangero, 1998] as the basis for a genetic factor analysis. Since the matrix was not itself positive definite, we first determined the closest symmetric positive definite matrix, utilizing Higham's method [Higham, 2002], and worked with that matrix. We used parallel analysis to settle on 3 factors [Horn, 1965]. With no *a priori* reason to think that the factors would be orthogonal, we used the oblimin oblique factor rotation. We then determined factor scores for each individual; for individuals with missing data, we created 32 imputations, calculated factor scores based on those imputations, and used the mean score. For individuals reliant on imputations we set all factor scores to missing if any factor had a standard deviation greater than 0.25 standard units across imputations for that individual.

Ancestry checking. Population ancestry was examined through principal components analysis as implemented in EIGENSTRAT, which computes principal components analysis scores over the input SNP data, 8068 SNPs from the linkage map as defined above. All samples that passed quality control (N=440) and all available HapMap samples with at least 85% of the 8068 linkage analysis SNPs were analyzed to provide clear references for the 3 major continental groupings. The first 4 principal components were visualized graphically with HapMap samples color-coded. Outliers were defined as prescribed in the EIGENSTRAT documentation [Price et al., 2006].

Linkage/association analysis methods. We chose the PPL framework to conduct all analyses. Historically the PPL was developed as a linkage analysis method (indeed PPL stood for posterior probability of linkage), which was an improvement upon traditional categorical trait LOD score analysis using the pedigree likelihood but could account for multiple trait models

(additive, dominant, recessive and any single locus variation between) without inflationary effects of multiple testing or parameter maximization. This was accomplished by using Bayesian statistical tools whereby trait parameters were removed from the likelihood by integration (not maximization as is done in commonly used statistical frameworks) as nuisance parameters. However, the same underlying likelihood need only be slightly modified for association analysis, even on non-pedigrees such as case-control datasets. Joint linkage-association is also possible as well as many extension that allow for quantitative traits, imprinting, sex specific recombination rates and epistasis. Rather than use a new acronym for each variation of the initial PPL likelihoods, it is preferable to use PPL not as an acronym for posterior probability of linkage but as an identifier for a statistical framework that allows for flexible model of complex disease datasets.

Briefly, the PPL is on the probability scale and regardless of the number of parameters in the model the same scale is always in effect. Thus, analysis from any of the PPL variations can all be directly compared to one another and directly interpreted to quantify the relative strength of the evidence for loci from categorical and quantitative analysis [Vieland et al., 2011]. Categorical traits were analyzed as described previously [9]. For quantitative phenotypes we used a PPL threshold model originally designed for the present dataset and empirically evaluated [Hou et al., 2012; Bartlett and Vieland, 2005]. As some persons with autism are non-verbal, have behavioral issues that interfere with quantitative language/reading assessment or perform too poorly for those assessments to be valid, analysis of quantitative data in a traditional framework would require either ignoring those subjects in the analysis (missing data) or setting their quantitative value to a low, but statistically plausible value (arbitrary constant). Instead of ad hoc data imputation, the PPL includes a threshold parameter that assumes the untestable subjects performed below a threshold that is left unspecified in the PPL analysis. The

advantage of this method is that we retain power by including the individuals with ASD who were without quantitative data in the analysis. It is always possible to simply remove those individuals with ASD from the analysis as a contrast condition for elucidating the role of ASD at a locus that may be ostensible linked to a language phenotype.

Primary linkage analysis was conducted on each of the three tiers separately and the linkage evidence was sequentially updated across the three tiers to provide a single metric for linkage evidence. While the primary linkage outcome is for Tier I, which contains ASD and SLI (or in five cases an "autism language impaired" subject), we also examined results from the other two tiers and the results sequentially updated linkage results over all three tiers (similar to a meta-analysis). Association analysis was conducted similarly; however, a family that contained persons that were not of European ancestry (N=1 in Tier III) was dropped from the association analysis, and several trios (one ASD and two parents) with data from our phenotypic battery were included (N=9, added to Tier III). Due to the sample sizes, the primary analysis is the sequentially updated evidence across all three tiers though Tier I is of interest alone.

To determine the extent of SNP haplotype tag coverage under regions of interest define by linkage peaks, all HapMap [International HapMap Consortium, 2003] SNPs segregating in the CEU HapMap population with minor allele frequency >=0.03 within each linkage regions were acquired via bulk download from HapMap Data Release 28 Phase II & III (www.hapmap.org) on NCBI assembly B36 and the dbSNP 126 dataset. Tag SNP panels were generated using the Tagger algorithm as implemented in the client version of Haploview 4.2 [Barrett et al., 2005]. Comparisons were made between the number of additional SNPs that were required to cover each region when the original Axiom™ SNPs were specified as "force includes" (i.e., mandatory tags) and the number of SNPs that were required to cover the same region without

specifying any force includes. Supplementary Table 5A.5 summarizes the results. To determine the fraction of each chromosomal region that was successfully tagged when only the original Axiom™ SNPs are considered, we divided the number of Axiom™ SNPs by the total number of SNPs needed to tag the region (Axiom™ SNPs plus additional HapMap SNPs).

Assessing the relative contribution of the three proband types to the final PPL. The maximized LOD (MOD) was calculated for each cM position by subsets for families with the presence of at least one autism nonverbal subject forming the nonverbal family group, from the remaining families the presence of at least one autism language impaired forming the language impaired family group, with other families classified as the language normal family group. For both linked and unlinked regions, each subset contributes to the MOD score as an increasing function of sample size (the exact function depends on the pedigree structure), but in linked regions the linked subsets make far larger contributions to the MOD score than the unlinked subsets. Therefore, for each genomic position we define:

$$\omega = \sum_{i} \alpha_{N,i} - \alpha_{MOD,i}$$

where *i* is the subset index, $\alpha_{N,i}$ is the proportion of families in subset *i*, and $\alpha_{MOD,i}$ is the proportion of the total MOD score attributable to subset *i*. The average of ω in unlinked regions, $\omega_{non-peak}$, can be compared to ω_{peak}, calculated for linked regions, where the ratio would be close to 1 if the linked region were homogeneously linked across the subsets and deviate from 1 as any subset contributed more to the final MOD. We examined $\omega_{peak}/\omega_{non-peak}$ using our defined linkage regions versus the remaining chromosomal positions for the language-related linkage peaks on chromosome 15 and 16. The ratios were 1.02 and 1.06

for the two peaks, both consistent with all three subgroups providing the same proportional linkage signal.

Assessing the within-family relative contributions of language impairment and ASD to the final PPL. For language related linkage peaks, we sought to understand the relative contribution of language impairment versus ASD to the observed score by either removing all SLI probands from the analysis or, removing all autism probands from the analysis, by setting those phenotypes to be missing data in the analysis. However, lack of power under these missing data scenarios may confound interpretation. To assess the probability of observing reductions in the PPL as great or greater than those reported, we performed a permutation study in the linked regions, randomly removing one phenotyped person (unaffected, LI/RI or ASD) from each pedigree and repeating the analysis. The resulting null distribution allows us to test the hypothesis that SLI and/or autism proband status in particular is important for the observed linkage finding. On chromosome 15, only 1% of the permutations resulted in a lower PPL than removing either SLI probands only or autism probands only (which gave roughly equal drops), indicating a significant relationship between SLI, autism and the maximum PPL without power loss as a confounding factor in interpretation. On chromosome 16, the opposite trend was observed, whereby 75% of the permutations resulted in a lower PPL than that obtained by dropping one RI (non-ASD) subject or the autism proband. Therefore, on chromosome 16 there is no evidence to distinguish the drop in the PPL when RI or autism subjects are removed from the drop in the PPL, (presumably) caused by reduced sample size, when subjects are removed randomly without regard to LI or ASD status. Another way to show that SLI and ASD do not provide unique contributions to a linkage peak is to repeat the permutation procedure, but restrict it to affected persons only (i.e., randomly remove one person with either LI (chr 15)/RI (chr 16) or ASD in each pedigree). A non-significant result from the permutation procedure that

is conditional on “affected” for removal indicated any affected phenotype is exchangeable in terms of induced power loss. On chromosome 15, the permutation p-values were 0.48 and 0.55 for ASD and SLI respectively, indicating full exchangeability of the two phenotypes for inducing power loss at that locus. On chromosome 16, the results are not as clear. For SLI p = 0.42 while for ASD p = 0.005, indicating a greater effect of ASD on the linkage signal than SLI at that locus. However, the unconditional permutation failed to indicate an effect of diagnostic status. Additional work to find the variants that underlie the linkage peak will be necessary to understand the relationship between those variants to SLI and ASD.

Supplementary Tables

Supplementary Table 5A.1: Demographic Table for All Tiers - Nuclear and Extended Family Profiles

	N	Mean age (SD)	Range
Family members meeting criteria for Autism	77	13.7 (9.6)	4.0-40.0
Family members meeting criteria for ASD	15	8.0 (4.5)	3.1-18.0
Family members meeting criteria for LI only	43	19.8 (20.3)	4.1-79.1
Family members meeting criteria for RI only	8	34.2 (23.6)	6.1-74.0
Family members meeting criteria for LI & RI only	9	30.4 (17.1)	7.1-51.1
All other family members	202	34.6 (16.9)	3.0-80.1
Total	354		

Supplementary Table 5A.2: Quantitative Cognitive Phenotypic Battery

Language & Cognitive Battery	Function	Age
Clinical Evaluation of Language Fundamentals- (CELF-4)- Core Score		
Concepts and Following Directions	Follow oral directions of increasing complexity	5+
Word Structure	Complete a sentence with the correct morphological inflection	5-8
Recalling Sentences	Imitation of sentences of increasing length and complexity	5+
Formulated Sentences	Create a meaningful sentence using a target word or phrase	5+
Word Classes	Identify two related words and explain	9+
Word Definitions	Define words with appropriate detail	13+
Clinical Evaluation of language Fundamentals-Preschool		
Sentence Structure	Interpret sentences of increasing length and complexity	3-5
Word Structure	Pronoun usage and complete sentences with the correct morphological inflection	3-5
Expressive Vocabulary	Label people, places, and actions	3-5
Comprehensive Assessment of Spoken Language (CASL)		
Non-literal language	Understand indirect and figurative language and sarcasm	7+
Meaning from Context	Infer meaning from linguistic context	11+
Inference	Use real world knowledge infer meaning	7-17
Ambiguous Sentences	Recognize ambiguity and verbalize it	11+
Pragmatic Judgment	Knowledge and use of pragmatic rules	3+
Comprehensive Test of Phonological Processing (CTOPP)		
Elision	Delete syllables and sounds from words to create new words	5+
Non-Word Repetition	Repeat nonsense words of increasing length	5+
Gray Oral Reading Tests (GORT-4)		
Rate	Timed oral reading	6+
Accuracy	Number or errors while reading	6+
Fluency	Combined score of rate and accuracy	6+
Comprehension	Answer questions about what was read	6+
Woodcock Reading Mastery Tests-Revised		
Word identification	Single word sight reading	6+
Word Attack	Nonsense word reading	6+
Wide Range Achievement Test 3 (WRAT3)		
Spelling	Spelling of dictated words	6+
Wechsler Abbreviated Test of Intelligence (WASI)		
Performance IQ	Composite non-verbal subtest scores: Block Design and Matrices	6+

Supplementary Table 5A.3: Phenotypic correlation matrix

	CELF				CTOPP		GORT					CSL				WRAT	WRM	
	FS	RS	WCE	WCR	EL	NR	AS	CS	FS	RS	ORQ	AS	MC	NL	PJ	Spell	WA	WID
CLFFS	1	0.7	0.58	0.66	0.4	0.2	0.6	0.5	0.6	0.5	0.59	0.6	0.7	0.6	0.7	0.48	0.4	0.5
CLFRS	0.7	1	0.61	0.68	0.5	0.3	0.6	0.6	0.6	0.6	0.68	0.7	0.7	0.6	0.7	0.55	0.5	0.6
CLFWCE	0.6	0.6	1	0.7	0.4	0.2	0.5	0.5	0.5	0.4	0.57	0.6	0.6	0.6	0.6	0.37	0.4	0.5
CLFWCR	0.7	0.7	0.7	1	0.4	0.2	0.6	0.6	0.6	0.6	0.69	0.6	0.7	0.6	0.6	0.53	0.5	0.6
CTOPEL	0.4	0.5	0.38	0.42	1	0.3	0.5	0.5	0.4	0.4	0.5	0.5	0.4	0.4	0.4	0.47	0.5	0.6
CTOPNR	0.2	0.3	0.2	0.19	0.3	1	0.1	0.2	0.2	0.2	0.21	0.3	0.3	0.2	0.2	0.29	0.2	0.3
GRT4AS	0.6	0.6	0.49	0.63	0.5	0.1	1	0.6	1	0.8	0.88	0.5	0.5	0.4	0.5	0.66	0.6	0.7
GRT4CS	0.5	0.6	0.53	0.6	0.5	0.2	0.6	1	0.6	0.6	0.82	0.5	0.6	0.5	0.5	0.47	0.5	0.6
GRT4FS	0.6	0.6	0.49	0.63	0.4	0.2	1	0.6	1	0.9	0.92	0.6	0.6	0.4	0.5	0.71	0.6	0.7
GRT4RS	0.5	0.6	0.44	0.62	0.4	0.2	0.8	0.6	0.9	1	0.86	0.5	0.6	0.4	0.5	0.73	0.6	0.7
GRT4ORQ	0.6	0.7	0.57	0.69	0.5	0.2	0.9	0.8	0.9	0.9	1	0.6	0.7	0.5	0.6	0.68	0.6	0.7
CSLAS	0.6	0.7	0.57	0.58	0.5	0.3	0.5	0.5	0.6	0.5	0.61	1	0.7	0.7	0.6	0.5	0.5	0.6
CSLMC	0.7	0.7	0.62	0.71	0.4	0.3	0.5	0.6	0.6	0.6	0.65	0.7	1	0.8	0.7	0.49	0.5	0.6
CSLNL	0.6	0.6	0.61	0.58	0.4	0.2	0.4	0.5	0.4	0.4	0.5	0.7	0.8	1	0.8	0.42	0.4	0.5
CSLPJ	0.7	0.7	0.6	0.62	0.4	0.2	0.5	0.5	0.5	0.5	0.55	0.6	0.7	0.8	1	0.38	0.3	0.5
WRAT	0.5	0.7	0.66	0.6	0.6	0.4	0.2	0.5	0.5	0.5	0.47	0.6	0.6	0.7	0.8	1	0.4	0.3
WRMWA	0.4	0.5	0.55	0.37	0.5	0.5	0.3	0.7	0.5	0.7	0.73	0.7	0.5	0.5	0.4	0.38	1	0.6
WRMWID	0.4	0.4	0.48	0.44	0.5	0.5	0.2	0.6	0.5	0.6	0.62	0.6	0.5	0.5	0.4	0.33	0.6	1

Supplementary Table 5A.4: Factor loading for the derived quantitative traits

Test	**Subscale**	**Factor 1**	**Factor 2**	**Factor 3**
CELF	Formulating Sentences	0.995	0.132	-0.219
	Repeating Sentences	0.647	0.379	0.315
	Word Classes - Expressive	0.73	0.451	0.011
	Word Classes - Receptive	0.944	0.014	0.106
CTOPP	Elision	0.082	-0.056	0.974
	Nonword Repetition	0.005	0.137	0.9
GORT	Accuracy	0.903	-0.094	-0.015
	Comprehension	0.372	0.653	0.251
	Fluency	0.892	0.039	-0.257
	Reading Score	0.855	0.171	0.198
CASL	Ambiguous Sentences	0.569	0.382	0.405
	Meaning from Context	0.608	0.404	0.112
	Nonliteral Language	0.284	0.769	-0.053
	Pragmatic Judgment	-0.204	0.775	0.506
WRAT	Spelling	0.95	-0.295	0.103
Woodcock Reading	Word Attack	0.848	-0.544	0.234
	Word ID	0.857	-0.111	0.288

Supplementary Table 5A.5: Fraction of linkage regions tagged for association analysis

Trait	Chromosome	bp range (kb)	Axiom-HapMap overlap	Additional HapMap SNPs needed	Fraction of region tagged
LI*	15	68361-91690	3596	2499	0.52
RI*	16	16022-26460	1468	962	0.55
SRS-QT	14	98547-106332	1444	1037	0.28
SRS-DT	15	93955-100181	2213	1039	0.53
YBOCS	13	52365-69872	2485	1110	0.63

Supplementary Table 5A.6: Follow-up Genotyping from GWAS

SNP	Chromosome	bp	cPPLD	Trait
rs3835700*	1	35422400	0.0014	f2
rs2279749	3	4772781	0.0114	f2
rs12490375	3	4773489	0.0026	f2
rs3792495*	3	4774443	0.2	f2
rs3792498	3	4776967	0.0024	f3
rs9831960	3	4777573	0.0028	f1
rs11732255	4	184678911	0.0003	f1
rs12646153*	4	184681169	0.0005	f1
rs7658620	4	184684075	0.0004	f1
rs10011451	4	185495233	0.0015	f1
rs12505290	4	185495461	0.0018	f1
rs6854407	4	185496329	0.0015	f1
rs12514616	5	120374182	0.0011	f2
rs983571	5	120375136	0.09	f2
rs2973177*	5	120375499	0.0072	f2
rs10973702	9	3818714	0.0012	f1
rs16919730*	9	3819188	0.13	f1
rs7029652	9	3819212	0.0003	f1
rs2987740	9	102914868	0.005	f1
rs1360098	9	102915925	0.0027	f1
rs10760791	9	102918028	0.0142	f1
rs7958047	12	39662871	0.06	f3
rs11179592*	12	39720021	0.0025	f3
rs1797990*	12	39803135	0.1	f3
rs1797988	12	39805150	0.18	f3
rs1626744	12	39805182	0.1	f3
rs764350*	14	49731682	0.0003	f2
rs9930784*	16	20602872	0.0038	f2
rs9930741	16	20602987	0.024	f2
rs12928136	16	20604843	0.0039	f2
rs11643793	16	20607517	0.0115	f2
rs764138	16	20650973	0.04	f2
rs11646042	16	20651537	0.004	f2
rs8044864	16	20653822	0.04	f2
rs9923588	16	25849211	0.18	RI*
rs4511540	16	25850199	0.1	RI*
rs8089400	18	43417118	0.18	LI*
rs8088661*	18	43417377	0.0012	LI*

Appendix 6: Supplemental Material for Chapter 7: Fine Mapping and Association Analysis of Candidate Genes for Autism Spectrum Disorder and Language Impairment in the NJLAGS Sample

Supplemental Table 6A.1: IPA Gene Ranking

GENE	SCORE
AKT1	27
NTRK3	25
IGF1R	23
BBS4	18
CHRNA3	18
CHRNB4	17
ST8SIA2	17
FES	16
CHRNA5	15
mir-9	14
PPP2R5C	11
PTPN9	11
NDE1	10
PRKCB	10
CYP1A2	9
JAG2	9
KLHL1	8
SEMA7A	8
CYP1A1	7
POLG	7
CYP11A1	6
FURIN	6
HSP90AA1	6
ISL2	6
NR2E3	6

GENE	SCORE
SCNN1B	6
SCNN1G	6
ALDH1A3	5
ARNT2	5
HEXA	5
KIF7	5
KLC1	5
LINGO1	5
PRC1	5
RLBP1	5
ATXN8OS	4
BCL11B	4
NRG4	4
HOMER2	3
IQGAP1	3
ISLR2	3
MESP2	3
MYH11	3
PKM2	3
RASGRF1	3
RGMA	3
SEMA4B	3
CSK	2
DIAPH3	2
DIO3	2

GENE	SCORE
IDH2	2
MEF2A	2
MESP1	2
mir-134	2
mir-7	2
NMB	2
NR2F2	2
OTOA	2
PCSK6	2
PLD4	2
RHCG	2
RPS15A	2
RPS17/RPS17L	2
UMOD	2
ARID3B	1
BAG5	1
CSPG4	1
IREB2	1
KIF26A	1
MFGE8	1
NPTN	1
PLIN1	1
SELS	1
SIVA1	1
TRAF3	1

Supplemental Table 6A.2: +ASD All Tiers

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0136
14	rs3803304	123.5126	0.0138
14	rs2494735	123.5191	0.0132
14	rs2494738	123.5253	0.0167
14	rs10138227	123.5472	0.0186
14	rs1022431	124.0897	0.0196
14	rs2816671	124.0903	0.015
14	rs909236	124.094	0.0169
14	rs2293806	124.1126	0.0186
14	rs3784240	124.1211	0.05
15	rs11854442	78.9473	0.0173
15	rs741761	78.95	0.0118
15	rs5015024	78.9503	0.0158
15	rs12594462	78.9509	0.0166
15	rs2734633	78.9518	0.019
15	rs2075589	78.953	0.0174
15	rs2072649	78.9532	0.0162
15	rs11857558	78.9546	0.0186
15	rs11856835	78.956	0.0238
15	rs8023274	78.9569	0.0174
15	rs1992145	78.9596	0.0159
15	rs7174305	78.9622	0.0188
15	rs1866113	78.9624	0.03
15	rs11072484	78.963	0.0166
15	rs8034317	79.5798	0.0157
15	rs12592505	79.5972	0.0196
15	rs8023268	79.6318	0.0172
15	rs10851882	79.6561	0.0137
15	rs7163907	79.6661	0.0172
15	rs11635996	79.6774	0.0161
15	rs3743164	95.6781	0.0129
15	rs16940941	95.6895	0.0174
15	rs1435393	95.6903	0.0144
15	rs11073750	95.6918	0.0139
15	rs7176429	95.6968	0.0126
15	rs2117655	95.7083	0.014
15	rs1369430	95.7108	0.0138
15	rs11631866	95.7153	0.0125
15	rs16940994	95.724	0.0186
15	rs11634284	95.7263	0.0145
15	rs6496453	95.728	0.0155
15	rs3784441	95.7361	0.0139
15	rs7161806	95.7371	0.0129
15	rs3903308	95.7414	0.0128
15	rs12595693	95.7435	0.016
15	rs1369423	95.7478	0.0136

Chr	Marker	Position	PPLD|L
15	rs11855377	95.7507	0.0125
15	rs1946698	95.7592	0.014
15	rs879131	95.7608	0.0159
15	rs2043516	95.7639	0.0147
15	rs2277580	95.7657	0.0195
15	rs920069	95.7733	0.0148
15	rs1991981	95.7766	0.0155
15	rs7178071	95.7864	0.0161
15	rs7170215	95.7914	0.0138
15	rs8038245	95.7966	0.014
15	rs7170976	95.8023	0.015
15	rs1461214	95.8066	0.0163
15	rs7172184	95.813	0.0151
15	rs922231	95.8154	0.0141
15	rs4887342	95.8209	0.0151
15	rs16941103	95.824	0.014
15	rs4887344	95.8293	0.0132
15	rs2009966	95.8395	0.0139
15	rs12324332	95.8484	0.0156
15	rs4989257	95.8544	0.0163
15	rs898707	95.8544	0.0154
15	rs1350799	95.857	0.0168
15	rs10468138	95.8611	0.0149
15	rs999905	95.8639	0.0176
15	rs1381112	95.866	0.0165
15	rs4887348	95.8719	0.0165
15	rs7167737	95.8969	0.0136
15	rs3784432	95.8977	0.0206
15	rs1948066	95.8986	0.0138
15	rs16941171	95.9013	0.0218
15	rs12438418	95.9018	0.0169
15	rs4887351	95.9032	0.0141
15	rs8035178	95.9044	0.0143
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0185
15	rs12595249	95.9128	0.0133
15	rs3825885	95.9169	0.013
15	rs16941206	95.9246	0.0138
15	rs4887353	95.9342	0.0127
15	rs17755717	95.9499	0.0156
15	rs8035265	95.9619	0.0198
15	rs6496461	95.9622	0.0174
15	rs3784421	95.9632	0.0188
15	rs16941252	95.9698	0.0168
15	rs3784419	95.9846	0.0207
15	rs3784415	95.9937	0.016

Chr	Marker	Position	PPLD|L
15	rs12148590	95.9942	0.0135
15	rs11073763	95.9949	0.0132
15	rs7168277	95.998	0.0144
15	rs4887363	96.0015	0.0157
15	rs2114251	96.0112	0.0164
15	rs994068	96.0141	0.0148
15	rs3784407	96.0169	0.0146
15	rs12911150	96.0172	0.0151
15	rs1426301	96.0179	0.0158
15	rs3825882	96.0183	0.0125
15	rs10163123	96.0278	0.0128
15	rs4887368	96.0296	0.0148
15	rs1991287	96.0392	0.0135
15	rs16941328	96.0473	0.0149
15	rs7168478	96.0512	0.0197
15	rs16941334	96.057	0.0137
15	rs11073767	96.0662	0.0115
15	rs17830422	96.0687	0.0172
15	rs10520676	96.0694	0.0151
15	rs8025158	96.0743	0.0133
15	rs12148100	96.0809	0.013
15	rs1346164	96.0934	0.016
15	rs6496469	96.1015	0.0138
15	rs1105962	96.1029	0.0135
15	rs4887381	96.1043	0.0143
15	rs8025146	96.105	0.019
15	rs17831280	96.1115	0.0173
15	rs16941424	96.1121	0.0167
15	rs1107292	96.1339	0.0162
15	rs7169789	96.1361	0.0147
15	rs7176520	96.1616	0.0118
15	rs7182602	96.1766	0.0141
15	rs11852905	96.1831	0.0167
15	rs16941563	96.1838	0.0129
15	rs1105693	96.1885	0.012
15	rs878646	96.2033	0.0166
15	rs878647	96.2036	0.0162
15	rs7164531	96.2148	0.02
15	rs11635754	96.2218	0.0165
15	rs16941593	96.2239	0.0158
15	rs4702	104.0594	0.0144
15	rs1894401	104.0679	0.0144
15	rs2521501	104.0968	0.0186
15	rs1029420	104.1095	0.0178
15	rs2677737	104.1099	0.0177
16	rs7205273	49.6515	0.03

Chr	Marker	Position	PPLD|L
16	rs886113	49.661	0.025
16	rs17841809	49.6698	0.03
16	rs17841810	49.6713	0.0157
16	rs463664	49.6847	0.0143
16	rs7190829	49.6881	0.0183
16	rs8055868	49.6896	0.0245
16	rs7204560	49.692	0.04
16	rs239349	49.6971	0.0144
16	rs238547	49.7091	0.0115
16	rs8044970	49.7118	0.0154
16	rs152733	49.7132	0.0153
16	rs152745	49.7193	0.0129
16	rs2301601	49.7194	0.0229
16	rs152742	49.7222	0.0135
16	rs238551	49.7304	0.0097
16	rs17199599	49.7329	0.0171
16	rs239350	49.7341	0.0239
16	rs2303157	49.738	0.0135
16	rs12596831	49.7401	0.0211
16	rs12446463	49.7423	0.0156
16	rs4967999	49.7425	0.0191
16	rs168748	49.7432	0.0143
16	rs250570	49.7434	0.0126
16	rs2303153	49.7452	0.0113
16	rs4968000	49.7461	0.0152
16	rs3743966	49.7467	0.0148
16	rs250567	49.7496	0.0179
16	rs181835	49.7524	0.0152
16	rs36064924	49.7547	0.0188

Supplemental Table 6A.3: +ASD Tier I+II

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0139
14	rs3803304	123.5126	0.0148
14	rs2494735	123.5191	0.0146
14	rs2494738	123.5253	0.0168
14	rs10138227	123.5472	0.0204
14	rs1022431	124.0897	0.0194
14	rs2816671	124.0903	0.0163
14	rs909236	124.094	0.0181
14	rs2293806	124.1126	0.0188
14	rs3784240	124.1211	0.04
15	rs11854442	78.9473	0.0194
15	rs741761	78.95	0.0137
15	rs5015024	78.9503	0.0167
15	rs12594462	78.9509	0.0174
15	rs2734633	78.9518	0.0197
15	rs2075589	78.953	0.0163
15	rs2072649	78.9532	0.0169
15	rs11857558	78.9546	0.021
15	rs11856835	78.956	0.0249
15	rs8023274	78.9569	0.0185
15	rs1992145	78.9596	0.0168
15	rs7174305	78.9622	0.0201
15	rs1866113	78.9624	0.03
15	rs11072484	78.963	0.0171
15	rs8034317	79.5798	0.0152
15	rs12592505	79.5972	0.0195
15	rs8023268	79.6318	0.0156
15	rs10851882	79.6561	0.0145
15	rs7163907	79.6661	0.0155
15	rs11635996	79.6774	0.0173
15	rs3743164	95.6781	0.013
15	rs16940941	95.6895	0.0175
15	rs1435393	95.6903	0.014
15	rs11073750	95.6918	0.0152
15	rs7176429	95.6968	0.0135
15	rs2117655	95.7083	0.0147
15	rs1369430	95.7108	0.0145
15	rs11631866	95.7153	0.0132
15	rs16940994	95.724	0.0184
15	rs11634284	95.7263	0.0144
15	rs6496453	95.728	0.0157
15	rs3784441	95.7361	0.0143
15	rs7161806	95.7371	0.0145
15	rs3903308	95.7414	0.0134
15	rs12595693	95.7435	0.0163
15	rs1369423	95.7478	0.0139
15	rs11855377	95.7507	0.0133
15	rs1946698	95.7592	0.0163
15	rs879131	95.7608	0.0167
15	rs2043516	95.7639	0.0159
15	rs2277580	95.7657	0.0194
15	rs920069	95.7733	0.0155
15	rs1991981	95.7766	0.0173
15	rs7178071	95.7864	0.0163
15	rs7170215	95.7914	0.0143
15	rs8038245	95.7966	0.0153
15	rs7170976	95.8023	0.0163
15	rs1461214	95.8066	0.016
15	rs7172184	95.813	0.0172
15	rs922231	95.8154	0.0144
15	rs4887342	95.8209	0.0162
15	rs16941103	95.824	0.0144
15	rs4887344	95.8293	0.0136
15	rs2009966	95.8395	0.0147
15	rs12324332	95.8484	0.0167
15	rs4989257	95.8544	0.018
15	rs898707	95.8544	0.0163
15	rs1350799	95.857	0.0171
15	rs10468138	95.8611	0.0158
15	rs999905	95.8639	0.0211
15	rs1381112	95.866	0.0171
15	rs4887348	95.8719	0.0198
15	rs7167737	95.8969	0.014
15	rs3784432	95.8977	0.0237
15	rs1948066	95.8986	0.0141
15	rs16941171	95.9013	0.0229
15	rs12438418	95.9018	0.0167
15	rs4887351	95.9032	0.0149
15	rs8035178	95.9044	0.0155
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0185
15	rs12595249	95.9128	0.0144
15	rs3825885	95.9169	0.0132
15	rs16941206	95.9246	0.0141
15	rs4887353	95.9342	0.0138
15	rs17755717	95.9499	0.0164
15	rs8035265	95.9619	0.0238
15	rs6496461	95.9622	0.03
15	rs3784421	95.9632	0.0189
15	rs16941252	95.9698	0.0172
15	rs3784419	95.9846	0.0206
15	rs3784415	95.9937	0.0187
15	rs12148590	95.9942	0.0148
15	rs11073763	95.9949	0.0143
15	rs7168277	95.998	0.0144
15	rs4887363	96.0015	0.0162
15	rs2114251	96.0112	0.0171
15	rs994068	96.0141	0.017
15	rs3784407	96.0169	0.0166
15	rs12911150	96.0172	0.0159
15	rs1426301	96.0179	0.0185
15	rs3825882	96.0183	0.0133
15	rs10163123	96.0278	0.0146
15	rs4887368	96.0296	0.016
15	rs1991287	96.0392	0.0141
15	rs16941328	96.0473	0.0153
15	rs7168478	96.0512	0.0197
15	rs16941334	96.057	0.0144
15	rs11073767	96.0662	0.0118
15	rs17830422	96.0687	0.0177
15	rs10520676	96.0694	0.015
15	rs8025158	96.0743	0.0142
15	rs12148100	96.0809	0.0154
15	rs1346164	96.0934	0.0158
15	rs6496469	96.1015	0.0142
15	rs1105962	96.1029	0.0139
15	rs4887381	96.1043	0.0144
15	rs8025146	96.105	0.021
15	rs17831280	96.1115	0.0177
15	rs16941424	96.1121	0.0172
15	rs1107292	96.1339	0.0165
15	rs7169789	96.1361	0.0151
15	rs7176520	96.1616	0.0116
15	rs7182602	96.1766	0.0148
15	rs11852905	96.1831	0.017
15	rs16941563	96.1838	0.0154
15	rs1105693	96.1885	0.0138
15	rs878646	96.2033	0.0161
15	rs878647	96.2036	0.0175
15	rs7164531	96.2148	0.0196
15	rs11635754	96.2218	0.0146
15	rs16941593	96.2239	0.0175
15	rs4702	104.0594	0.0146
15	rs1894401	104.0679	0.0143
15	rs2521501	104.0968	0.0203
15	rs1029420	104.1095	0.0161
15	rs2677737	104.1099	0.0164
16	rs7205273	49.6515	0.0225
16	rs886113	49.661	0.0236
16	rs17841809	49.6698	0.0222
16	rs17841810	49.6713	0.0175
16	rs463664	49.6847	0.0147
16	rs7190829	49.6881	0.0229
16	rs8055868	49.6896	0.03
16	rs7204560	49.692	0.05
16	rs239349	49.6971	0.016
16	rs238547	49.7091	0.0128
16	rs8044970	49.7118	0.0155
16	rs152733	49.7132	0.0189
16	rs152745	49.7193	0.0135
16	rs2301601	49.7194	0.0234
16	rs152742	49.7222	0.0151
16	rs238551	49.7304	0.014
16	rs17199599	49.7329	0.0201
16	rs239350	49.7341	0.0216
16	rs2303157	49.738	0.0146
16	rs12596831	49.7401	0.0218
16	rs12446463	49.7423	0.0157
16	rs4967999	49.7425	0.0185
16	rs168748	49.7432	0.0143
16	rs250570	49.7434	0.0148
16	rs2303153	49.7452	0.0128
16	rs4968000	49.7461	0.0165
16	rs3743966	49.7467	0.0151
16	rs250567	49.7496	0.0178
16	rs181835	49.7524	0.0158
16	rs36064924	49.7547	0.0171

Supplemental Table 6A.4: +ASD Tier 1

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0144
14	rs3803304	123.5126	0.0149
14	rs2494735	123.5191	0.0146
14	rs2494738	123.5253	0.017
14	rs10138227	123.5472	0.021
14	rs1022431	124.0897	0.0194
14	rs2816671	124.0903	0.0168
14	rs909236	124.094	0.0178
14	rs2293806	124.1126	0.0188
14	rs3784240	124.1211	0.04
15	rs11854442	78.9473	0.0194
15	rs741761	78.95	0.0136
15	rs5015024	78.9503	0.0178
15	rs12594462	78.9509	0.0177
15	rs2734633	78.9518	0.0201
15	rs2075589	78.953	0.0172
15	rs2072649	78.9532	0.0185
15	rs11857558	78.9546	0.0195
15	rs11856835	78.956	0.0183
15	rs8023274	78.9569	0.0182
15	rs1992145	78.9596	0.0153
15	rs7174305	78.9622	0.0205
15	rs1866113	78.9624	0.03
15	rs11072484	78.963	0.0166
15	rs8034317	79.5798	0.0156
15	rs12592505	79.5972	0.0196
15	rs8023268	79.6318	0.0156
15	rs10851882	79.6561	0.015
15	rs7163907	79.6661	0.0155
15	rs11635996	79.6774	0.021
15	rs3743164	95.6781	0.0135
15	rs16940941	95.6895	0.0179
15	rs1435393	95.6903	0.0143
15	rs11073750	95.6918	0.0168
15	rs7176429	95.6968	0.015
15	rs2117655	95.7083	0.0161
15	rs1369430	95.7108	0.0146
15	rs11631866	95.7153	0.0144
15	rs16940994	95.724	0.0187
15	rs11634284	95.7263	0.0148
15	rs6496453	95.728	0.0158
15	rs3784441	95.7361	0.0149
15	rs7161806	95.7371	0.0168
15	rs3903308	95.7414	0.0145
15	rs12595693	95.7435	0.0166
15	rs1369423	95.7478	0.0151
15	rs11855377	95.7507	0.0147
15	rs1946698	95.7592	0.03
15	rs879131	95.7608	0.0168
15	rs2043516	95.7639	0.0177
15	rs2277580	95.7657	0.0194
15	rs920069	95.7733	0.0156
15	rs1991981	95.7766	0.0191
15	rs7178071	95.7864	0.0168
15	rs7170215	95.7914	0.0144
15	rs8038245	95.7966	0.0162
15	rs7170976	95.8023	0.0177
15	rs1461214	95.8066	0.0159
15	rs7172184	95.813	0.0161
15	rs922231	95.8154	0.0147
15	rs4887342	95.8209	0.0147
15	rs16941103	95.824	0.0145
15	rs4887344	95.8293	0.014
15	rs2009966	95.8395	0.0156
15	rs12324332	95.8484	0.0166
15	rs4989257	95.8544	0.0177
15	rs898707	95.8544	0.0159
15	rs1350799	95.857	0.0177
15	rs10468138	95.8611	0.0159
15	rs999905	95.8639	0.03
15	rs1381112	95.866	0.0182
15	rs4887348	95.8719	0.0225
15	rs7167737	95.8969	0.0141
15	rs3784432	95.8977	0.0244
15	rs1948066	95.8986	0.0145
15	rs16941171	95.9013	0.02
15	rs12438418	95.9018	0.017
15	rs4887351	95.9032	0.0152
15	rs8035178	95.9044	0.016
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0186
15	rs12595249	95.9128	0.0148
15	rs3825885	95.9169	0.0136
15	rs16941206	95.9246	0.0152
15	rs4887353	95.9342	0.0145
15	rs17755717	95.9499	0.0174
15	rs8035265	95.9619	0.0246
15	rs6496461	95.9622	0.03
15	rs3784421	95.9632	0.0189
15	rs16941252	95.9698	0.0177
15	rs3784419	95.9846	0.0208
15	rs3784415	95.9937	0.0193
15	rs12148590	95.9942	0.0157
15	rs11073763	95.9949	0.0142
15	rs7168277	95.998	0.0147
15	rs4887363	96.0015	0.0161
15	rs2114251	96.0112	0.0178
15	rs994068	96.0141	0.0168
15	rs3784407	96.0169	0.0156
15	rs12911150	96.0172	0.0157
15	rs1426301	96.0179	0.0176
15	rs3825882	96.0183	0.0139
15	rs10163123	96.0278	0.0153
15	rs4887368	96.0296	0.0157
15	rs1991287	96.0392	0.0143
15	rs16941328	96.0473	0.0151
15	rs7168478	96.0512	0.0198
15	rs16941334	96.057	0.0151
15	rs11073767	96.0662	0.0121
15	rs17830422	96.0687	0.0186
15	rs10520676	96.0694	0.0151
15	rs8025158	96.0743	0.0154
15	rs12148100	96.0809	0.0173
15	rs1346164	96.0934	0.0152
15	rs6496469	96.1015	0.0139
15	rs1105962	96.1029	0.0138
15	rs4887381	96.1043	0.0144
15	rs8025146	96.105	0.0179
15	rs17831280	96.1115	0.0178
15	rs16941424	96.1121	0.0169
15	rs1107292	96.1339	0.0169
15	rs7169789	96.1361	0.0155
15	rs7176520	96.1616	0.0122
15	rs7182602	96.1766	0.0152
15	rs11852905	96.1831	0.0164
15	rs16941563	96.1838	0.0154
15	rs1105693	96.1885	0.0139
15	rs878646	96.2033	0.0185
15	rs878647	96.2036	0.0215
15	rs7164531	96.2148	0.0199
15	rs11635754	96.2218	0.015
15	rs16941593	96.2239	0.0188
15	rs4702	104.0594	0.0152
15	rs1894401	104.0679	0.015
15	rs2521501	104.0968	0.0212
15	rs1029420	104.1095	0.0163
15	rs2677737	104.1099	0.0161
16	rs7205273	49.6515	0.0212
16	rs886113	49.661	0.03
16	rs17841809	49.6698	0.0205
16	rs17841810	49.6713	0.0189
16	rs463664	49.6847	0.0152
16	rs7190829	49.6881	0.0217
16	rs8055868	49.6896	0.03
16	rs7204560	49.692	0.04
16	rs239349	49.6971	0.016
16	rs238547	49.7091	0.0136
16	rs8044970	49.7118	0.0158
16	rs152733	49.7132	0.0183
16	rs152745	49.7193	0.0127
16	rs2301601	49.7194	0.025
16	rs152742	49.7222	0.017
16	rs238551	49.7304	0.0157
16	rs17199599	49.7329	0.0236
16	rs239350	49.7341	0.0224
16	rs2303157	49.738	0.0169
16	rs12596831	49.7401	0.0209
16	rs12446463	49.7423	0.015
16	rs4967999	49.7425	0.0165
16	rs168748	49.7432	0.0134
16	rs250570	49.7434	0.0148
16	rs2303153	49.7452	0.0126
16	rs4968000	49.7461	0.0215
16	rs3743966	49.7467	0.0152
16	rs250567	49.7496	0.0178
16	rs181835	49.7524	0.0162
16	rs36064924	49.7547	0.0178

Supplemental Table 6A.5: -ASD All Tiers

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0131
14	rs3803304	123.5126	0.0142
14	rs2494735	123.5191	0.0129
14	rs2494738	123.5253	0.0161
14	rs10138227	123.5472	0.0175
14	rs1022431	124.0897	0.0201
14	rs2816671	124.0903	0.016
14	rs909236	124.094	0.0196
14	rs2293806	124.1126	0.0189
14	rs3784240	124.1211	0.03
15	rs11854442	78.9473	0.0178
15	rs741761	78.95	0.0142
15	rs5015024	78.9503	0.0144
15	rs12594462	78.9509	0.0171
15	rs2734633	78.9518	0.0177
15	rs2075589	78.953	0.0177
15	rs2072649	78.9532	0.0158
15	rs11857558	78.9546	0.0238
15	rs11856835	78.956	0.0204
15	rs8023274	78.9569	0.0172
15	rs1992145	78.9596	0.0132
15	rs7174305	78.9622	0.019
15	rs1866113	78.9624	0.03
15	rs11072484	78.963	0.0191
15	rs8034317	79.5798	0.0149
15	rs12592505	79.5972	0.0198
15	rs8023268	79.6318	0.0151
15	rs10851882	79.6561	0.0143
15	rs7163907	79.6661	0.0151
15	rs11635996	79.6774	0.0174
15	rs3743164	95.6781	0.0119
15	rs16940941	95.6895	0.0183
15	rs1435393	95.6903	0.0129
15	rs11073750	95.6918	0.0129
15	rs7176429	95.6968	0.0115
15	rs2117655	95.7083	0.0132
15	rs1369430	95.7108	0.0132
15	rs11631866	95.7153	0.0125
15	rs16940994	95.724	0.0179
15	rs11634284	95.7263	0.014
15	rs6496453	95.728	0.0158
15	rs3784441	95.7361	0.0137
15	rs7161806	95.7371	0.0123
15	rs3903308	95.7414	0.0126
15	rs12595693	95.7435	0.0141
15	rs1369423	95.7478	0.0129
15	rs11855377	95.7507	0.0124
15	rs1946698	95.7592	0.0135
15	rs879131	95.7608	0.0154
15	rs2043516	95.7639	0.0125
15	rs2277580	95.7657	0.0194
15	rs920069	95.7733	0.0141
15	rs1991981	95.7766	0.0131
15	rs7178071	95.7864	0.0166
15	rs7170215	95.7914	0.0137
15	rs8038245	95.7966	0.0124
15	rs7170976	95.8023	0.0136
15	rs1461214	95.8066	0.0177
15	rs7172184	95.813	0.0172
15	rs922231	95.8154	0.016
15	rs4887342	95.8209	0.0149
15	rs16941103	95.824	0.0127
15	rs4887344	95.8293	0.0136
15	rs2009966	95.8395	0.0128
15	rs12324332	95.8484	0.0146
15	rs4989257	95.8544	0.0145
15	rs898707	95.8544	0.0131
15	rs1350799	95.857	0.0169
15	rs10468138	95.8611	0.0129
15	rs999905	95.8639	0.0175
15	rs1381112	95.866	0.017
15	rs4887348	95.8719	0.0153
15	rs7167737	95.8969	0.0128
15	rs3784432	95.8977	0.013
15	rs1948066	95.8986	0.013
15	rs16941171	95.9013	0.0159
15	rs12438418	95.9018	0.018
15	rs4887351	95.9032	0.0146
15	rs8035178	95.9044	0.0139
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0193
15	rs12595249	95.9128	0.0127
15	rs3825885	95.9169	0.0128
15	rs16941206	95.9246	0.0129
15	rs4887353	95.9342	0.0134
15	rs17755717	95.9499	0.0156
15	rs8035265	95.9619	0.0212
15	rs6496461	95.9622	0.0228
15	rs3784421	95.9632	0.0189
15	rs16941252	95.9698	0.0173
15	rs3784419	95.9846	0.0209
15	rs3784415	95.9937	0.0179
15	rs12148590	95.9942	0.0136
15	rs11073763	95.9949	0.0139
15	rs7168277	95.998	0.0143
15	rs4887363	96.0015	0.0148
15	rs2114251	96.0112	0.0163
15	rs994068	96.0141	0.0153
15	rs3784407	96.0169	0.0137
15	rs12911150	96.0172	0.0149
15	rs1426301	96.0179	0.0164
15	rs3825882	96.0183	0.0124
15	rs10163123	96.0278	0.0141
15	rs4887368	96.0296	0.0159
15	rs1991287	96.0392	0.0132
15	rs16941328	96.0473	0.0154
15	rs7168478	96.0512	0.0199
15	rs16941334	96.057	0.0153
15	rs11073767	96.0662	0.0117
15	rs17830422	96.0687	0.0171
15	rs10520676	96.0694	0.0148
15	rs8025158	96.0743	0.0154
15	rs12148100	96.0809	0.0152
15	rs1346164	96.0934	0.013
15	rs6496469	96.1015	0.015
15	rs1105962	96.1029	0.0133
15	rs4887381	96.1043	0.0147
15	rs8025146	96.105	0.0213
15	rs17831280	96.1115	0.018
15	rs16941424	96.1121	0.02
15	rs1107292	96.1339	0.0199
15	rs7169789	96.1361	0.03
15	rs7176520	96.1616	0.0117
15	rs7182602	96.1766	0.021
15	rs11852905	96.1831	0.03
15	rs16941563	96.1838	0.0183
15	rs1105693	96.1885	0.0159
15	rs878646	96.2033	0.0136
15	rs878647	96.2036	0.014
15	rs7164531	96.2148	0.0174
15	rs11635754	96.2218	0.0144
15	rs16941593	96.2239	0.0168
15	rs4702	104.0594	0.0123
15	rs1894401	104.0679	0.0119
15	rs2521501	104.0968	0.0109
15	rs1029420	104.1095	0.0121
15	rs2677737	104.1099	0.0121
16	rs7205273	49.6515	0.03
16	rs886113	49.661	0.03
16	rs17841809	49.6698	0.0236
16	rs17841810	49.6713	0.015
16	rs463664	49.6847	0.0141
16	rs7190829	49.6881	0.0182
16	rs8055868	49.6896	0.0223
16	rs7204560	49.692	0.0244
16	rs239349	49.6971	0.0116
16	rs238547	49.7091	0.0143
16	rs8044970	49.7118	0.0147
16	rs152733	49.7132	0.013
16	rs152745	49.7193	0.0137
16	rs2301601	49.7194	0.0242
16	rs152742	49.7222	0.0144
16	rs238551	49.7304	0.0099
16	rs17199599	49.7329	0.0171
16	rs239350	49.7341	0.0172
16	rs2303157	49.738	0.0144
16	rs12596831	49.7401	0.0175
16	rs12446463	49.7423	0.0162
16	rs4967999	49.7425	0.0155
16	rs168748	49.7432	0.0153
16	rs250570	49.7434	0.0121
16	rs2303153	49.7452	0.0132
16	rs4968000	49.7461	0.0132
16	rs3743966	49.7467	0.0136
16	rs250567	49.7496	0.0174
16	rs181835	49.7524	0.0148
16	rs36064924	49.7547	0.0196

Supplemental Table 6A.6: -ASD Tier I+II

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.014
14	rs3803304	123.5126	0.0151
14	rs2494735	123.5191	0.0145
14	rs2494738	123.5253	0.0175
14	rs10138227	123.5472	0.0185
14	rs1022431	124.0897	0.0187
14	rs2816671	124.0903	0.0182
14	rs909236	124.094	0.0195
14	rs2293806	124.1126	0.0188
14	rs3784240	124.1211	0.03
15	rs11854442	78.9473	0.0233
15	rs741761	78.95	0.0159
15	rs5015024	78.9503	0.0151
15	rs12594462	78.9509	0.0178
15	rs2734633	78.9518	0.0186
15	rs2075589	78.953	0.0163
15	rs2072649	78.9532	0.0161
15	rs11857558	78.9546	0.03
15	rs11856835	78.956	0.03
15	rs8023274	78.9569	0.0186
15	rs1992145	78.9596	0.0163
15	rs7174305	78.9622	0.0214
15	rs1866113	78.9624	0.03
15	rs11072484	78.963	0.0193
15	rs8034317	79.5798	0.0152
15	rs12592505	79.5972	0.0196
15	rs8023268	79.6318	0.015
15	rs10851882	79.6561	0.0152
15	rs7163907	79.6661	0.0151
15	rs11635996	79.6774	0.0186
15	rs3743164	95.6781	0.0121
15	rs16940941	95.6895	0.0182
15	rs1435393	95.6903	0.0131
15	rs11073750	95.6918	0.0134
15	rs7176429	95.6968	0.0123
15	rs2117655	95.7083	0.0136
15	rs1369430	95.7108	0.0136
15	rs11631866	95.7153	0.0132
15	rs16940994	95.724	0.0181
15	rs11634284	95.7263	0.0138
15	rs6496453	95.728	0.0164
15	rs3784441	95.7361	0.0141
15	rs7161806	95.7371	0.0132
15	rs3903308	95.7414	0.0129
15	rs12595693	95.7435	0.0144
15	rs1369423	95.7478	0.0136
15	rs11855377	95.7507	0.0128
15	rs1946698	95.7592	0.0157
15	rs879131	95.7608	0.0159
15	rs2043516	95.7639	0.0137
15	rs2277580	95.7657	0.0197
15	rs920069	95.7733	0.0145
15	rs1991981	95.7766	0.0146
15	rs7178071	95.7864	0.0167
15	rs7170215	95.7914	0.0135
15	rs8038245	95.7966	0.0135
15	rs7170976	95.8023	0.0138
15	rs1461214	95.8066	0.016
15	rs7172184	95.813	0.0222
15	rs922231	95.8154	0.0148
15	rs4887342	95.8209	0.016
15	rs16941103	95.824	0.0133
15	rs4887344	95.8293	0.0144
15	rs2009966	95.8395	0.0137
15	rs12324332	95.8484	0.0157
15	rs4989257	95.8544	0.0155
15	rs898707	95.8544	0.0138
15	rs1350799	95.857	0.0168
15	rs10468138	95.8611	0.0137
15	rs999905	95.8639	0.0217
15	rs1381112	95.866	0.0169
15	rs4887348	95.8719	0.0177
15	rs7167737	95.8969	0.0143
15	rs3784432	95.8977	0.0163
15	rs1948066	95.8986	0.0141
15	rs16941171	95.9013	0.0161
15	rs12438418	95.9018	0.0185
15	rs4887351	95.9032	0.0158
15	rs8035178	95.9044	0.0153
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0193
15	rs12595249	95.9128	0.0136
15	rs3825885	95.9169	0.0134
15	rs16941206	95.9246	0.0142
15	rs4887353	95.9342	0.0148
15	rs17755717	95.9499	0.0164
15	rs8035265	95.9619	0.03
15	rs6496461	95.9622	0.07
15	rs3784421	95.9632	0.0191
15	rs16941252	95.9698	0.0178
15	rs3784419	95.9846	0.0204
15	rs3784415	95.9937	0.03
15	rs12148590	95.9942	0.0146
15	rs11073763	95.9949	0.0161
15	rs7168277	95.998	0.0147
15	rs4887363	96.0015	0.017
15	rs2114251	96.0112	0.017
15	rs994068	96.0141	0.02
15	rs3784407	96.0169	0.0168
15	rs12911150	96.0172	0.0156
15	rs1426301	96.0179	0.0213
15	rs3825882	96.0183	0.0137
15	rs10163123	96.0278	0.0166
15	rs4887368	96.0296	0.0178
15	rs1991287	96.0392	0.0141
15	rs16941328	96.0473	0.0165
15	rs7168478	96.0512	0.0199
15	rs16941334	96.057	0.0171
15	rs11073767	96.0662	0.0125
15	rs17830422	96.0687	0.0184
15	rs10520676	96.0694	0.0161
15	rs8025158	96.0743	0.0164
15	rs12148100	96.0809	0.0178
15	rs1346164	96.0934	0.0135
15	rs6496469	96.1015	0.016
15	rs1105962	96.1029	0.0137
15	rs4887381	96.1043	0.0148
15	rs8025146	96.105	0.03
15	rs17831280	96.1115	0.0176
15	rs16941424	96.1121	0.0231
15	rs1107292	96.1339	0.0185
15	rs7169789	96.1361	0.0198
15	rs7176520	96.1616	0.0135
15	rs7182602	96.1766	0.0195
15	rs11852905	96.1831	0.03
15	rs16941563	96.1838	0.03
15	rs1105693	96.1885	0.0231
15	rs878646	96.2033	0.0145
15	rs878647	96.2036	0.0169
15	rs7164531	96.2148	0.0169
15	rs11635754	96.2218	0.0139
15	rs16941593	96.2239	0.0183
15	rs4702	104.0594	0.0138
15	rs1894401	104.0679	0.0164
15	rs2521501	104.0968	0.0119
15	rs1029420	104.1095	0.0129
15	rs2677737	104.1099	0.0133
16	rs7205273	49.6515	0.0196
16	rs886113	49.661	0.04
16	rs17841809	49.6698	0.0228
16	rs17841810	49.6713	0.0216
16	rs463664	49.6847	0.0163
16	rs7190829	49.6881	0.024
16	rs8055868	49.6896	0.04
16	rs7204560	49.692	0.05
16	rs239349	49.6971	0.0131
16	rs238547	49.7091	0.0143
16	rs8044970	49.7118	0.016
16	rs152733	49.7132	0.0143
16	rs152745	49.7193	0.014
16	rs2301601	49.7194	0.03
16	rs152742	49.7222	0.0161
16	rs238551	49.7304	0.0146
16	rs17199599	49.7329	0.0193
16	rs239350	49.7341	0.0171
16	rs2303157	49.738	0.0157
16	rs12596831	49.7401	0.0187
16	rs12446463	49.7423	0.016
16	rs4967999	49.7425	0.0176
16	rs168748	49.7432	0.0155
16	rs250570	49.7434	0.0143
16	rs2303153	49.7452	0.0159
16	rs4968000	49.7461	0.0172
16	rs3743966	49.7467	0.0144
16	rs250567	49.7496	0.0184
16	rs181835	49.7524	0.0171
16	rs36064924	49.7547	0.0165

Supplemental Table 6A.7: -ASD Tier I

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0141
14	rs3803304	123.5126	0.015
14	rs2494735	123.5191	0.0144
14	rs2494738	123.5253	0.0177
14	rs10138227	123.5472	0.0184
14	rs1022431	124.0897	0.0189
14	rs2816671	124.0903	0.0182
14	rs909236	124.094	0.0192
14	rs2293806	124.1126	0.0188
14	rs3784240	124.1211	0.03
15	rs11854442	78.9473	0.0227
15	rs741761	78.95	0.0153
15	rs5015024	78.9503	0.0171
15	rs12594462	78.9509	0.0178
15	rs2734633	78.9518	0.0178
15	rs2075589	78.953	0.0166
15	rs2072649	78.9532	0.0172
15	rs11857558	78.9546	0.03
15	rs11856835	78.956	0.0216
15	rs8023274	78.9569	0.018
15	rs1992145	78.9596	0.0156
15	rs7174305	78.9622	0.0237
15	rs1866113	78.9624	0.03
15	rs11072484	78.963	0.0182
15	rs8034317	79.5798	0.0153
15	rs12592505	79.5972	0.0196
15	rs8023268	79.6318	0.0155
15	rs10851882	79.6561	0.0149
15	rs7163907	79.6661	0.0156
15	rs11635996	79.6774	0.0207
15	rs3743164	95.6781	0.0124
15	rs16940941	95.6895	0.0186
15	rs1435393	95.6903	0.0136
15	rs11073750	95.6918	0.014
15	rs7176429	95.6968	0.0129
15	rs2117655	95.7083	0.0141
15	rs1369430	95.7108	0.0138
15	rs11631866	95.7153	0.0136
15	rs16940994	95.724	0.0182
15	rs11634284	95.7263	0.0141
15	rs6496453	95.728	0.0166
15	rs3784441	95.7361	0.0143
15	rs7161806	95.7371	0.0144
15	rs3903308	95.7414	0.0129
15	rs12595693	95.7435	0.0148
15	rs1369423	95.7478	0.0146
15	rs11855377	95.7507	0.0134
15	rs1946698	95.7592	0.0191
15	rs879131	95.7608	0.0162
15	rs2043516	95.7639	0.015
15	rs2277580	95.7657	0.0197
15	rs920069	95.7733	0.015
15	rs1991981	95.7766	0.0151
15	rs7178071	95.7864	0.0174
15	rs7170215	95.7914	0.0133
15	rs8038245	95.7966	0.0144
15	rs7170976	95.8023	0.0142
15	rs1461214	95.8066	0.0165
15	rs7172184	95.813	0.0168
15	rs922231	95.8154	0.0142
15	rs4887342	95.8209	0.0148
15	rs16941103	95.824	0.0136
15	rs4887344	95.8293	0.0144
15	rs2009966	95.8395	0.0145
15	rs12324332	95.8484	0.0155
15	rs4989257	95.8544	0.0159
15	rs898707	95.8544	0.0142
15	rs1350799	95.857	0.0172
15	rs10468138	95.8611	0.0141
15	rs999905	95.8639	0.0226
15	rs1381112	95.866	0.0175
15	rs4887348	95.8719	0.0191
15	rs7167737	95.8969	0.0149
15	rs3784432	95.8977	0.0204
15	rs1948066	95.8986	0.0143
15	rs16941171	95.9013	0.0166
15	rs12438418	95.9018	0.0182
15	rs4887351	95.9032	0.0158
15	rs8035178	95.9044	0.0158
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0189
15	rs12595249	95.9128	0.0142
15	rs3825885	95.9169	0.0138
15	rs16941206	95.9246	0.0154
15	rs4887353	95.9342	0.0158
15	rs17755717	95.9499	0.0175
15	rs8035265	95.9619	0.0236
15	rs6496461	95.9622	0.06
15	rs3784421	95.9632	0.0191
15	rs16941252	95.9698	0.0181
15	rs3784419	95.9846	0.0204
15	rs3784415	95.9937	0.03
15	rs12148590	95.9942	0.0153
15	rs11073763	95.9949	0.0158
15	rs7168277	95.998	0.0152
15	rs4887363	96.0015	0.0162
15	rs2114251	96.0112	0.0176
15	rs994068	96.0141	0.0181
15	rs3784407	96.0169	0.0158
15	rs12911150	96.0172	0.0157
15	rs1426301	96.0179	0.0191
15	rs3825882	96.0183	0.0141
15	rs10163123	96.0278	0.017
15	rs4887368	96.0296	0.0166
15	rs1991287	96.0392	0.0142
15	rs16941328	96.0473	0.0158
15	rs7168478	96.0512	0.0199
15	rs16941334	96.057	0.0173
15	rs11073767	96.0662	0.013
15	rs17830422	96.0687	0.0201
15	rs10520676	96.0694	0.0163
15	rs8025158	96.0743	0.0164
15	rs12148100	96.0809	0.0173
15	rs1346164	96.0934	0.0138
15	rs6496469	96.1015	0.0148
15	rs1105962	96.1029	0.0142
15	rs4887381	96.1043	0.0152
15	rs8025146	96.105	0.0201
15	rs17831280	96.1115	0.0176
15	rs16941424	96.1121	0.0211
15	rs1107292	96.1339	0.0184
15	rs7169789	96.1361	0.0202
15	rs7176520	96.1616	0.0134
15	rs7182602	96.1766	0.0207
15	rs11852905	96.1831	0.0223
15	rs16941563	96.1838	0.03
15	rs1105693	96.1885	0.0208
15	rs878646	96.2033	0.0156
15	rs878647	96.2036	0.0192
15	rs7164531	96.2148	0.0176
15	rs11635754	96.2218	0.0139
15	rs16941593	96.2239	0.0185
15	rs4702	104.0594	0.0141
15	rs1894401	104.0679	0.0163
15	rs2521501	104.0968	0.0125
15	rs1029420	104.1095	0.0133
15	rs2677737	104.1099	0.0136
16	rs7205273	49.6515	0.0189
16	rs886113	49.661	0.04
16	rs17841809	49.6698	0.0215
16	rs17841810	49.6713	0.0234
16	rs463664	49.6847	0.0173
16	rs7190829	49.6881	0.0231
16	rs8055868	49.6896	0.03
16	rs7204560	49.692	0.04
16	rs239349	49.6971	0.0135
16	rs238547	49.7091	0.0137
16	rs8044970	49.7118	0.0155
16	rs152733	49.7132	0.0145
16	rs152745	49.7193	0.0143
16	rs2301601	49.7194	0.03
16	rs152742	49.7222	0.0152
16	rs238551	49.7304	0.0139
16	rs17199599	49.7329	0.0214
16	rs239350	49.7341	0.0191
16	rs2303157	49.738	0.0168
16	rs12596831	49.7401	0.0189
16	rs12446463	49.7423	0.015
16	rs4967999	49.7425	0.016
16	rs168748	49.7432	0.0144
16	rs250570	49.7434	0.0139
16	rs2303153	49.7452	0.0137
16	rs4968000	49.7461	0.0205
16	rs3743966	49.7467	0.0143
16	rs250567	49.7496	0.0179
16	rs181835	49.7524	0.0165
16	rs36064924	49.7547	0.0162

Supplemental Table 6A.8: ASD only All Tiers

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0166
14	rs3803304	123.5126	0.0165
14	rs2494735	123.5191	0.0171
14	rs2494738	123.5253	0.0173
14	rs10138227	123.5472	0.0161
14	rs1022431	124.0897	0.0164
14	rs2816671	124.0903	0.0152
14	rs909236	124.094	0.0145
14	rs2293806	124.1126	0.0171
14	rs3784240	124.1211	0.0234
15	rs11854442	78.9473	0.0188
15	rs741761	78.95	0.0133
15	rs5015024	78.9503	0.0145
15	rs12594462	78.9509	0.0175
15	rs2734633	78.9518	0.017
15	rs2075589	78.953	0.03
15	rs2072649	78.9532	0.0158
15	rs11857558	78.9546	0.0151
15	rs11856835	78.956	0.0144
15	rs8023274	78.9569	0.0187
15	rs1992145	78.9596	0.0162
15	rs7174305	78.9622	0.0163
15	rs1866113	78.9624	0.0155
15	rs11072484	78.963	0.016
15	rs8034317	79.5798	0.0147
15	rs12592505	79.5972	0.0197
15	rs8023268	79.6318	0.0149
15	rs10851882	79.6561	0.0132
15	rs7163907	79.6661	0.0154
15	rs11635996	79.6774	0.0172
15	rs3743164	95.6781	0.0148
15	rs16940941	95.6895	0.018
15	rs1435393	95.6903	0.0152
15	rs11073750	95.6918	0.0153
15	rs7176429	95.6968	0.015
15	rs2117655	95.7083	0.0144
15	rs1369430	95.7108	0.0151
15	rs11631866	95.7153	0.015
15	rs16940994	95.724	0.0187
15	rs11634284	95.7263	0.0156
15	rs6496453	95.728	0.0163
15	rs3784441	95.7361	0.0155
15	rs7161806	95.7371	0.0142
15	rs3903308	95.7414	0.0141
15	rs12595693	95.7435	0.0164
15	rs1369423	95.7478	0.0143
15	rs11855377	95.7507	0.0146
15	rs1946698	95.7592	0.0144
15	rs879131	95.7608	0.0161
15	rs2043516	95.7639	0.0168
15	rs2277580	95.7657	0.0193
15	rs920069	95.7733	0.0159
15	rs1991981	95.7766	0.0164
15	rs7178071	95.7864	0.0172
15	rs7170215	95.7914	0.0148
15	rs8038245	95.7966	0.0147
15	rs7170976	95.8023	0.017
15	rs1461214	95.8066	0.0157
15	rs7172184	95.813	0.0145
15	rs922231	95.8154	0.014
15	rs4887342	95.8209	0.0191
15	rs16941103	95.824	0.015
15	rs4887344	95.8293	0.0152
15	rs2009966	95.8395	0.016
15	rs12324332	95.8484	0.03
15	rs4989257	95.8544	0.03
15	rs898707	95.8544	0.0185
15	rs1350799	95.857	0.0172
15	rs10468138	95.8611	0.0167
15	rs999905	95.8639	0.0163
15	rs1381112	95.866	0.0169
15	rs4887348	95.8719	0.0159
15	rs7167737	95.8969	0.0158
15	rs3784432	95.8977	0.03
15	rs1948066	95.8986	0.0146
15	rs16941171	95.9013	0.03
15	rs12438418	95.9018	0.0179
15	rs4887351	95.9032	0.0149
15	rs8035178	95.9044	0.0143
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0184
15	rs12595249	95.9128	0.0138
15	rs3825885	95.9169	0.0149
15	rs16941206	95.9246	0.0135
15	rs4887353	95.9342	0.0133
15	rs17755717	95.9499	0.0148
15	rs8035265	95.9619	0.0155
15	rs6496461	95.9622	0.0162
15	rs3784421	95.9632	0.0188
15	rs16941252	95.9698	0.0182
15	rs3784419	95.9846	0.0197
15	rs3784415	95.9937	0.0158
15	rs12148590	95.9942	0.0147
15	rs11073763	95.9949	0.0141
15	rs7168277	95.998	0.0149
15	rs4887363	96.0015	0.0154
15	rs2114251	96.0112	0.0161
15	rs994068	96.0141	0.0151
15	rs3784407	96.0169	0.0143
15	rs12911150	96.0172	0.0159
15	rs1426301	96.0179	0.0156
15	rs3825882	96.0183	0.0136
15	rs10163123	96.0278	0.0144
15	rs4887368	96.0296	0.0151
15	rs1991287	96.0392	0.0142
15	rs16941328	96.0473	0.0151
15	rs7168478	96.0512	0.0195
15	rs16941334	96.057	0.0153
15	rs11073767	96.0662	0.0135
15	rs17830422	96.0687	0.0175
15	rs10520676	96.0694	0.0159
15	rs8025158	96.0743	0.0152
15	rs12148100	96.0809	0.0156
15	rs1346164	96.0934	0.0144
15	rs6496469	96.1015	0.0151
15	rs1105962	96.1029	0.0147
15	rs4887381	96.1043	0.0157
15	rs8025146	96.105	0.0154
15	rs17831280	96.1115	0.0172
15	rs16941424	96.1121	0.0162
15	rs1107292	96.1339	0.0163
15	rs7169789	96.1361	0.0164
15	rs7176520	96.1616	0.0134
15	rs7182602	96.1766	0.0156
15	rs11852905	96.1831	0.0157
15	rs16941563	96.1838	0.0141
15	rs1105693	96.1885	0.0146
15	rs878646	96.2033	0.0147
15	rs878647	96.2036	0.0147
15	rs7164531	96.2148	0.0176
15	rs11635754	96.2218	0.0169
15	rs16941593	96.2239	0.0167
15	rs4702	104.0594	0.0122
15	rs1894401	104.0679	0.011
15	rs2521501	104.0968	0.013
15	rs1029420	104.1095	0.0118
15	rs2677737	104.1099	0.0116
16	rs7205273	49.6515	0.017
16	rs886113	49.661	0.0141
16	rs17841809	49.6698	0.018
16	rs17841810	49.6713	0.014
16	rs463664	49.6847	0.0152
16	rs7190829	49.6881	0.0176
16	rs8055868	49.6896	0.0185
16	rs7204560	49.692	0.0196
16	rs239349	49.6971	0.015
16	rs238547	49.7091	0.0136
16	rs8044970	49.7118	0.0165
16	rs152733	49.7132	0.0153
16	rs152745	49.7193	0.0159
16	rs2301601	49.7194	0.0181
16	rs152742	49.7222	0.015
16	rs238551	49.7304	0.0137
16	rs17199599	49.7329	0.0172
16	rs239350	49.7341	0.0168
16	rs2303157	49.738	0.0156
16	rs12596831	49.7401	0.0192
16	rs12446463	49.7423	0.0151
16	rs4967999	49.7425	0.0132
16	rs168748	49.7432	0.0139
16	rs250570	49.7434	0.0136
16	rs2303153	49.7452	0.014
16	rs4968000	49.7461	0.0165
16	rs3743966	49.7467	0.0141
16	rs250567	49.7496	0.0173
16	rs181835	49.7524	0.016
16	rs36064924	49.7547	0.0172

Supplemental Table 6A.9: ASD only Tier I+II

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0153
14	rs3803304	123.5126	0.0165
14	rs2494735	123.5191	0.0144
14	rs2494738	123.5253	0.0174
14	rs10138227	123.5472	0.0163
14	rs1022431	124.0897	0.0169
14	rs2816671	124.0903	0.0189
14	rs909236	124.094	0.0174
14	rs2293806	124.1126	0.0169
14	rs3784240	124.1211	0.0239
15	rs11854442	78.9473	0.0206
15	rs741761	78.95	0.0157
15	rs5015024	78.9503	0.016
15	rs12594462	78.9509	0.018
15	rs2734633	78.9518	0.0172
15	rs2075589	78.953	0.0219
15	rs2072649	78.9532	0.0161
15	rs11857558	78.9546	0.0163
15	rs11856835	78.956	0.0145
15	rs8023274	78.9569	0.0185
15	rs1992145	78.9596	0.0158
15	rs7174305	78.9622	0.0165
15	rs1866113	78.9624	0.0164
15	rs11072484	78.963	0.0169
15	rs8034317	79.5798	0.015
15	rs12592505	79.5972	0.0196
15	rs8023268	79.6318	0.0145
15	rs10851882	79.6561	0.0141
15	rs7163907	79.6661	0.0149
15	rs11635996	79.6774	0.0172
15	rs3743164	95.6781	0.0152
15	rs16940941	95.6895	0.018
15	rs1435393	95.6903	0.0159
15	rs11073750	95.6918	0.0175
15	rs7176429	95.6968	0.0158
15	rs2117655	95.7083	0.015
15	rs1369430	95.7108	0.0159
15	rs11631866	95.7153	0.0163
15	rs16940994	95.724	0.0184
15	rs11634284	95.7263	0.0158
15	rs6496453	95.728	0.0167
15	rs3784441	95.7361	0.0163
15	rs7161806	95.7371	0.0149
15	rs3903308	95.7414	0.0148
15	rs12595693	95.7435	0.0166
15	rs1369423	95.7478	0.0147
15	rs11855377	95.7507	0.0152
15	rs1946698	95.7592	0.0148
15	rs879131	95.7608	0.0172
15	rs2043516	95.7639	0.0174
15	rs2277580	95.7657	0.0193
15	rs920069	95.7733	0.0157
15	rs1991981	95.7766	0.0165
15	rs7178071	95.7864	0.0171
15	rs7170215	95.7914	0.0153
15	rs8038245	95.7966	0.0154
15	rs7170976	95.8023	0.0175
15	rs1461214	95.8066	0.0159
15	rs7172184	95.813	0.0147
15	rs922231	95.8154	0.015
15	rs4887342	95.8209	0.0177
15	rs16941103	95.824	0.0153
15	rs4887344	95.8293	0.0147
15	rs2009966	95.8395	0.0162
15	rs12324332	95.8484	0.03
15	rs4989257	95.8544	0.03
15	rs898707	95.8544	0.0182
15	rs1350799	95.857	0.0174
15	rs10468138	95.8611	0.0167
15	rs999905	95.8639	0.0178
15	rs1381112	95.866	0.0193
15	rs4887348	95.8719	0.0169
15	rs7167737	95.8969	0.0158
15	rs3784432	95.8977	0.03
15	rs1948066	95.8986	0.0149
15	rs16941171	95.9013	0.04
15	rs12438418	95.9018	0.0185
15	rs4887351	95.9032	0.0156
15	rs8035178	95.9044	0.0153
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0188
15	rs12595249	95.9128	0.0149
15	rs3825885	95.9169	0.0152
15	rs16941206	95.9246	0.0142
15	rs4887353	95.9342	0.0146
15	rs17755717	95.9499	0.0171
15	rs8035265	95.9619	0.0161
15	rs6496461	95.9622	0.0174
15	rs3784421	95.9632	0.0189
15	rs16941252	95.9698	0.0191
15	rs3784419	95.9846	0.0195
15	rs3784415	95.9937	0.0165
15	rs12148590	95.9942	0.0163
15	rs11073763	95.9949	0.0154
15	rs7168277	95.998	0.0164
15	rs4887363	96.0015	0.0164
15	rs2114251	96.0112	0.0167
15	rs994068	96.0141	0.0162
15	rs3784407	96.0169	0.0146
15	rs12911150	96.0172	0.0166
15	rs1426301	96.0179	0.0168
15	rs3825882	96.0183	0.0147
15	rs10163123	96.0278	0.0149
15	rs4887368	96.0296	0.0161
15	rs1991287	96.0392	0.0147
15	rs16941328	96.0473	0.0154
15	rs7168478	96.0512	0.0195
15	rs16941334	96.057	0.0158
15	rs11073767	96.0662	0.0139
15	rs17830422	96.0687	0.0184
15	rs10520676	96.0694	0.016
15	rs8025158	96.0743	0.0155
15	rs12148100	96.0809	0.0162
15	rs1346164	96.0934	0.0147
15	rs6496469	96.1015	0.0149
15	rs1105962	96.1029	0.0153
15	rs4887381	96.1043	0.016
15	rs8025146	96.105	0.0154
15	rs17831280	96.1115	0.0176
15	rs16941424	96.1121	0.0171
15	rs1107292	96.1339	0.0167
15	rs7169789	96.1361	0.0168
15	rs7176520	96.1616	0.0131
15	rs7182602	96.1766	0.0157
15	rs11852905	96.1831	0.0159
15	rs16941563	96.1838	0.015
15	rs1105693	96.1885	0.0152
15	rs878646	96.2033	0.0152
15	rs878647	96.2036	0.0157
15	rs7164531	96.2148	0.0189
15	rs11635754	96.2218	0.0163
15	rs16941593	96.2239	0.0172
15	rs4702	104.0594	0.013
15	rs1894401	104.0679	0.012
15	rs2521501	104.0968	0.0139
15	rs1029420	104.1095	0.0131
15	rs2677737	104.1099	0.0133
16	rs7205273	49.6515	0.0178
16	rs886113	49.661	0.0143
16	rs17841809	49.6698	0.0183
16	rs17841810	49.6713	0.0144
16	rs463664	49.6847	0.0158
16	rs7190829	49.6881	0.0181
16	rs8055868	49.6896	0.0187
16	rs7204560	49.692	0.0189
16	rs239349	49.6971	0.0154
16	rs238547	49.7091	0.0148
16	rs8044970	49.7118	0.0165
16	rs152733	49.7132	0.016
16	rs152745	49.7193	0.0158
16	rs2301601	49.7194	0.0187
16	rs152742	49.7222	0.016
16	rs238551	49.7304	0.0145
16	rs17199599	49.7329	0.0193
16	rs239350	49.7341	0.0156
16	rs2303157	49.738	0.0165
16	rs12596831	49.7401	0.0191
16	rs12446463	49.7423	0.0164
16	rs4967999	49.7425	0.013
16	rs168748	49.7432	0.014
16	rs250570	49.7434	0.015
16	rs2303153	49.7452	0.0148
16	rs4968000	49.7461	0.0199
16	rs3743966	49.7467	0.0166
16	rs250567	49.7496	0.0209
16	rs181835	49.7524	0.0162
16	rs36064924	49.7547	0.0171

Supplemental Table 6A.10: ASD only Tier I

Chr	Marker	Position	PPLD|L
14	rs2494730	123.503	0.0149
14	rs3803304	123.5126	0.0157
14	rs2494735	123.5191	0.0151
14	rs2494738	123.5253	0.0175
14	rs10138227	123.5472	0.0165
14	rs1022431	124.0897	0.0171
14	rs2816671	124.0903	0.0185
14	rs909236	124.094	0.0179
14	rs2293806	124.1126	0.0177
14	rs3784240	124.1211	0.03
15	rs11854442	78.9473	0.0201
15	rs741761	78.95	0.0151
15	rs5015024	78.9503	0.0167
15	rs12594462	78.9509	0.0178
15	rs2734633	78.9518	0.0172
15	rs2075589	78.953	0.022
15	rs2072649	78.9532	0.0165
15	rs11857558	78.9546	0.0165
15	rs11856835	78.956	0.014
15	rs8023274	78.9569	0.018
15	rs1992145	78.9596	0.0152
15	rs7174305	78.9622	0.0166
15	rs1866113	78.9624	0.0164
15	rs11072484	78.963	0.017
15	rs8034317	79.5798	0.0158
15	rs12592505	79.5972	0.0196
15	rs8023268	79.6318	0.0151
15	rs10851882	79.6561	0.0158
15	rs7163907	79.6661	0.0154
15	rs11635996	79.6774	0.0176
15	rs3743164	95.6781	0.0163
15	rs16940941	95.6895	0.0182
15	rs1435393	95.6903	0.0174
15	rs11073750	95.6918	0.0206
15	rs7176429	95.6968	0.017
15	rs2117655	95.7083	0.016
15	rs1369430	95.7108	0.016
15	rs11631866	95.7153	0.0167
15	rs16940994	95.724	0.0186
15	rs11634284	95.7263	0.0164
15	rs6496453	95.728	0.0164
15	rs3784441	95.7361	0.0169
15	rs7161806	95.7371	0.0148
15	rs3903308	95.7414	0.0155
15	rs12595693	95.7435	0.0176
15	rs1369423	95.7478	0.0155
15	rs11855377	95.7507	0.0169
15	rs1946698	95.7592	0.0166
15	rs879131	95.7608	0.0173
15	rs2043516	95.7639	0.0205
15	rs2277580	95.7657	0.0194
15	rs920069	95.7733	0.0164
15	rs1991981	95.7766	0.0186
15	rs7178071	95.7864	0.0178
15	rs7170215	95.7914	0.0155
15	rs8038245	95.7966	0.0159
15	rs7170976	95.8023	0.0202
15	rs1461214	95.8066	0.0166
15	rs7172184	95.813	0.0149
15	rs922231	95.8154	0.0157
15	rs4887342	95.8209	0.0174
15	rs16941103	95.824	0.0158
15	rs4887344	95.8293	0.0149
15	rs2009966	95.8395	0.0186
15	rs12324332	95.8484	0.03
15	rs4989257	95.8544	0.03
15	rs898707	95.8544	0.0192
15	rs1350799	95.857	0.0175
15	rs10468138	95.8611	0.018
15	rs999905	95.8639	0.0187
15	rs1381112	95.866	0.0201
15	rs4887348	95.8719	0.0174
15	rs7167737	95.8969	0.0162
15	rs3784432	95.8977	0.04
15	rs1948066	95.8986	0.0155
15	rs16941171	95.9013	0.04
15	rs12438418	95.9018	0.0187
15	rs4887351	95.9032	0.0159
15	rs8035178	95.9044	0.0169
15	rs449832	95.9093	0.02
15	rs12592807	95.9105	0.0188
15	rs12595249	95.9128	0.0162
15	rs3825885	95.9169	0.0153
15	rs16941206	95.9246	0.0159
15	rs4887353	95.9342	0.0167
15	rs17755717	95.9499	0.0182
15	rs8035265	95.9619	0.017
15	rs6496461	95.9622	0.0172
15	rs3784421	95.9632	0.019
15	rs16941252	95.9698	0.0185
15	rs3784419	95.9846	0.0196
15	rs3784415	95.9937	0.0166
15	rs12148590	95.9942	0.0188
15	rs11073763	95.9949	0.0179
15	rs7168277	95.998	0.0182
15	rs4887363	96.0015	0.0163
15	rs2114251	96.0112	0.0169
15	rs994068	96.0141	0.0162
15	rs3784407	96.0169	0.0154
15	rs12911150	96.0172	0.0168
15	rs1426301	96.0179	0.0165
15	rs3825882	96.0183	0.0162
15	rs10163123	96.0278	0.0165
15	rs4887368	96.0296	0.0166
15	rs1991287	96.0392	0.0161
15	rs16941328	96.0473	0.0154
15	rs7168478	96.0512	0.0196
15	rs16941334	96.057	0.0165
15	rs11073767	96.0662	0.0148
15	rs17830422	96.0687	0.0199
15	rs10520676	96.0694	0.017
15	rs8025158	96.0743	0.0169
15	rs12148100	96.0809	0.0173
15	rs1346164	96.0934	0.015
15	rs6496469	96.1015	0.0147
15	rs1105962	96.1029	0.0155
15	rs4887381	96.1043	0.0164
15	rs8025146	96.105	0.0154
15	rs17831280	96.1115	0.0179
15	rs16941424	96.1121	0.0174
15	rs1107292	96.1339	0.0166
15	rs7169789	96.1361	0.0179
15	rs7176520	96.1616	0.013
15	rs7182602	96.1766	0.0167
15	rs11852905	96.1831	0.0163
15	rs16941563	96.1838	0.0152
15	rs1105693	96.1885	0.0154
15	rs878646	96.2033	0.0154
15	rs878647	96.2036	0.0156
15	rs7164531	96.2148	0.019
15	rs11635754	96.2218	0.0168
15	rs16941593	96.2239	0.0172
15	rs4702	104.0594	0.0133
15	rs1894401	104.0679	0.0123
15	rs2521501	104.0968	0.0139
15	rs1029420	104.1095	0.0133
15	rs2677737	104.1099	0.0134
16	rs7205273	49.6515	0.0177
16	rs886113	49.661	0.0147
16	rs17841809	49.6698	0.0179
16	rs17841810	49.6713	0.0153
16	rs463664	49.6847	0.0165
16	rs7190829	49.6881	0.0182
16	rs8055868	49.6896	0.0183
16	rs7204560	49.692	0.0185
16	rs239349	49.6971	0.016
16	rs238547	49.7091	0.0151
16	rs8044970	49.7118	0.0168
16	rs152733	49.7132	0.0163
16	rs152745	49.7193	0.016
16	rs2301601	49.7194	0.0189
16	rs152742	49.7222	0.0166
16	rs238551	49.7304	0.0153
16	rs17199599	49.7329	0.0205
16	rs239350	49.7341	0.0159
16	rs2303157	49.738	0.0177
16	rs12596831	49.7401	0.0177
16	rs12446463	49.7423	0.0176
16	rs4967999	49.7425	0.0139
16	rs168748	49.7432	0.015
16	rs250570	49.7434	0.0149
16	rs2303153	49.7452	0.0154
16	rs4968000	49.7461	0.0232
16	rs3743966	49.7467	0.0171
16	rs250567	49.7496	0.02
16	rs181835	49.7524	0.0161
16	rs36064924	49.7547	0.0174

Supplemental Table 6A.11: Disease model parameters used in genotype data simulations using the.

The penetrances used for each model are listed for DGF equal to 1, 5, 10, 20, and 30%. The penetrance for a homozygous disease allele is P(DD), heterozygous is P(Dd), and homozygous non-disease allele is P(dd). Table A lists the recessive model parameters and Table B lists the dominant model parameters. Models in red have more than 80% power to detect association for the LI (L), RI (R), and SRS-DT (S) phenotypes.

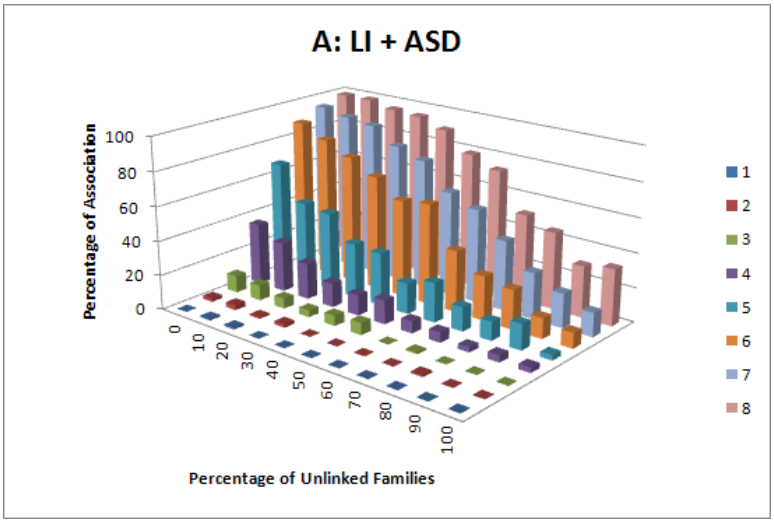
A: Recessive Models

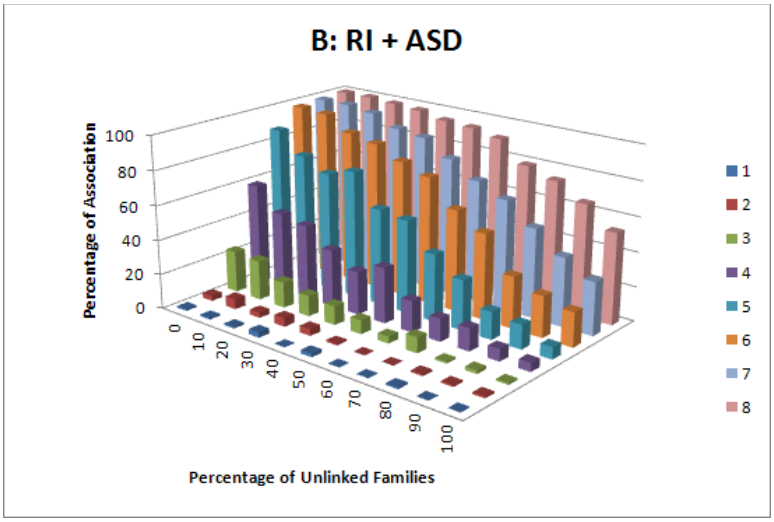
	Disease Gene Frequency																			
	1%				5%				10%				20%				30%			
Disease Risk	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%
1	0.17	0.17	0.17		0.17	0.17	0.17		0.17	0.17	0.17		0.17	0.17	0.17		0.17	0.17	0.17	
2	0.34	0.25	0.17		0.32	0.24	0.16		0.31	0.23	0.15		0.28	0.21	0.14		0.26	0.20	0.13	
3	0.50	0.25	0.17		0.48	0.24	0.16		0.46	0.23	0.15		0.41	0.21	0.14		0.37	0.18	0.12	
4	0.67	0.25	0.17		0.64	0.24	0.16		0.61	0.23	0.15		0.53	0.20	0.13		0.46	0.17	0.11	
5	0.84	0.25	0.17		0.80	0.24	0.16		0.75	0.23	0.15		0.64	0.19	0.13	R	0.54	0.16	0.11	L,R,S
6	-	-	-		0.96	0.24	0.16		0.89	0.22	0.15		0.75	0.19	0.13	L,R,S	0.61	0.15	0.10	L,R,S
7	-	-	-		-	-	-		-	-	-		0.85	0.18	0.12	L,R,S	0.68	0.15	0.10	L,R,S
8	-	-	-		-	-	-		-	-	-		0.94	0.18	0.12	L,R,S	0.74	0.14	0.09	L,R,S
9	-	-	-		-	-	-		-	-	-		-	-	-		0.79	0.13	0.09	L,R,S
10	-	-	-		-	-	-		-	-	-		-	-	-		0.84	0.13	0.08	L,R,S

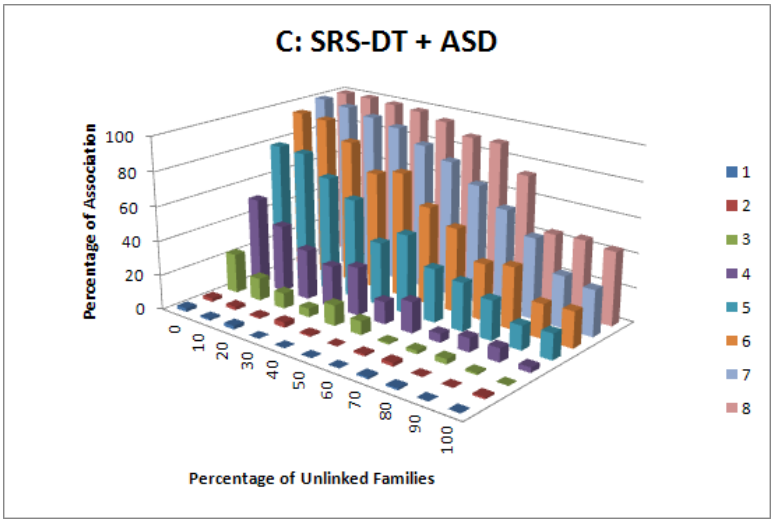
B: Dominant Models

	Disease Gene Frequency																			
	1%				5%				10%				20%				30%			
Disease Risk	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%	P(DD)	P(Dd)	P(dd)	80%
1	0.17	0.17	0.17		0.17	0.17	0.17		0.17	0.17	0.17		0.17	0.17	0.17		0.17	0.17	0.17	
2	0.33	0.33	0.17		0.31	0.31	0.15		0.29	0.29	0.14		0.25	0.25	0.13		0.23	0.23	0.11	
3	0.49	0.49	0.16		0.43	0.43	0.14		0.37	0.37	0.12		0.30	0.30	0.10		0.25	0.25	0.08	R,S
4	0.64	0.64	0.16		0.53	0.53	0.13	L,R,S	0.43	0.43	0.11	L,R,S	0.33	0.33	0.08	L,R,S	0.27	0.27	0.07	R,S
5	0.79	0.79	0.16	R	0.61	0.61	0.12	L,R,S	0.48	0.48	0.10	L,R,S	0.35	0.35	0.07	L,R,S	0.28	0.28	0.06	L,R,S
6	0.93	0.93	0.15	R,S	0.69	0.69	0.11	L,R,S	0.52	0.52	0.09	L,R,S	0.36	0.36	0.06	L,R,S	0.29	0.29	0.05	L,R,S
7	-	-	-		0.75	0.75	0.11	L,R,S	0.56	0.56	0.08	L,R,S	0.38	0.38	0.05	L,R,S	0.29	0.29	0.04	L,R,S
8	-	-	-		0.81	0.81	0.10	L,R,S	0.58	0.58	0.07	L,R,S	0.39	0.39	0.05	L,R,S	0.30	0.30	0.04	L,R,S
9	-	-	-		0.86	0.86	0.10	L,R,S	0.61	0.61	0.07	L,R,S	0.39	0.39	0.04	L,R,S	0.30	0.30	0.03	L,R,S
10	-	-	-		0.91	0.91	0.09	L,R,S	0.63	0.63	0.06	L,R,S	0.40	0.40	0.04	L,R,S	0.30	0.30	0.03	L,R,S

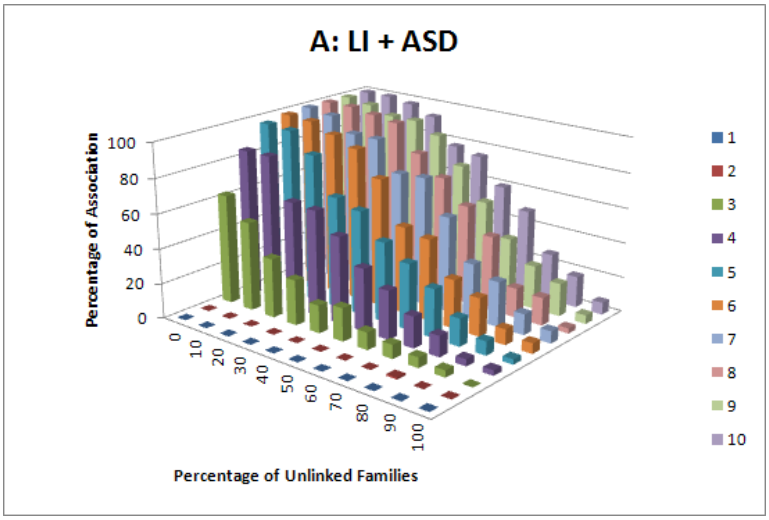
Supplemental Figure 6A.1: NJLAGS Power Determination for Recessive Models. DGF = 20%

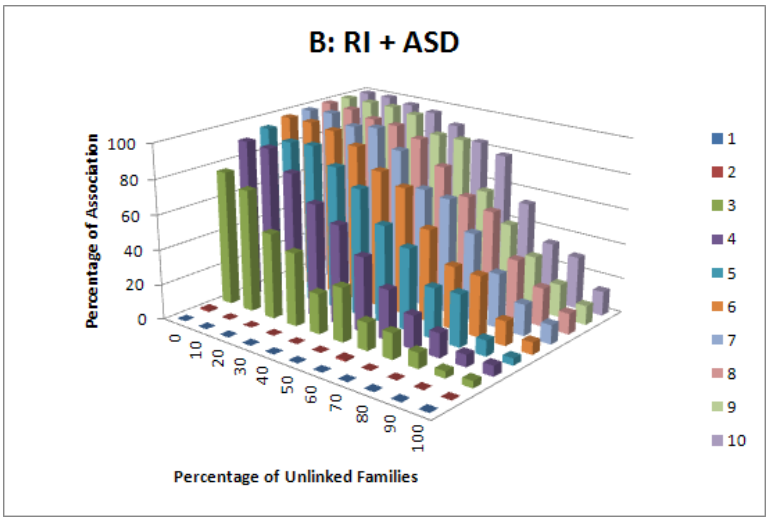


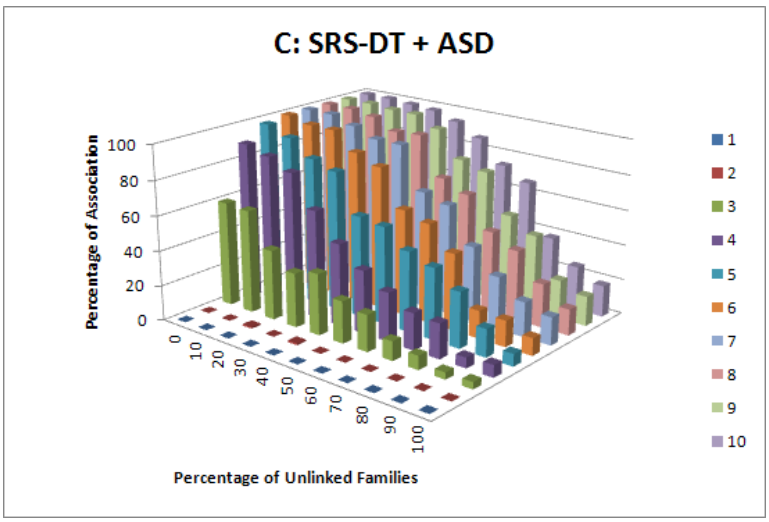




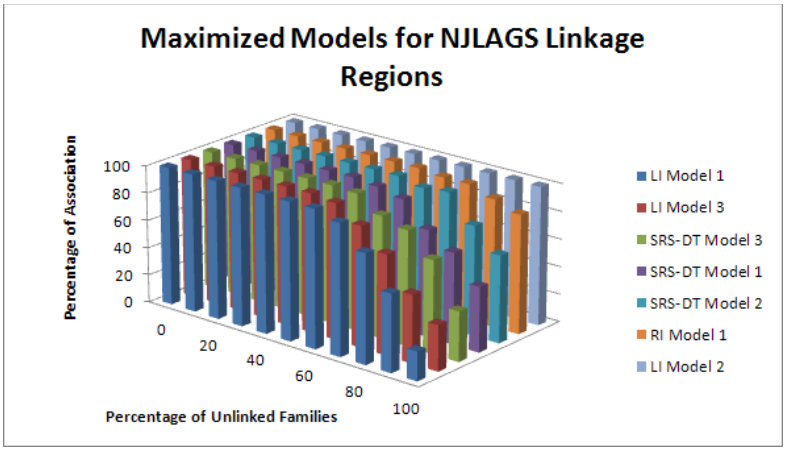
Supplemental Figure 6A.2: NJLAGS Power Determination for Dominant Models. DGF = 20%







Supplemental Figure 6A.3: Power Determination of Maximized Models for NJLAGS Linkage Regions. Simulation results for the maximized models for LI, RI, and SRS-DT (PPLD ≥ 50%). The Models represented are: LI Model 1: α = 1, DGF = 0.001, P(DD) = 0.999, P(Dd) = 0.7, P(dd) = 0; LI Model 2: α = 1, DGF = 0.8, P(DD) = 0.999, P(Dd) = 0.1, P(dd) = 0; LI Model 3: α = 1, DGF = 0.1, P(DD) = 0.999, P(Dd) = 0.7, P(dd) = 0; RI Model 1: α = 1, DGF = 0.8, P(DD) = 0.9, P(Dd) = 0, P(dd) = 0; SRS-DT Model 1: α = 1, DGF = 0.1, P(DD) = 0.8, P(Dd) = 0.2, P(dd) = 0; SRS-DT Model 2: α = 1, DGF = 0.1, P(DD) = 0.999, P(Dd) = 0.1, P(dd) = 0; SRS-DT Model 3: α = 1, DGF = 0.3, P(DD) = 0.4, P(Dd) = 0, P(dd) = 0.



References

Abecasis, GR, Cherny, SS, Cookson, WO, Cardon, LR. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. 30: 97-101.

Abrahams, B, Geschwind, D. 2012. Genetics of Autism. In: Antonarakis, S, Speicher, M, Motulsky, A, editors. Vogel and Motulsky's human genetics: problems and approaches. Berlin: Springer-Verlag, p 699-714.

Abrahams, BS, Geschwind, DH. 2008. Advances in autism genetics: on the threshold of a new neurobiology. Nature reviews. Genetics 9: 341-55.

Abramson, R, Wright, H, Carpenter, R, Brennan, W, Lumpuy, O, Cole, E, Young, S. 1989. Elevated blood serotonin in autistic probands and their first-degree relatives. J Autism Dev Disord 19: 397-407.

Alarcon, M, Abrahams, BS, Stone, JL, Duvall, JA, Perederiy, JV, Bomar, JM, Sebat, J, Wigler, M, Martin, CL, Ledbetter, DH, Nelson, SF, Cantor, RM, Geschwind, DH. 2008. Linkage , Association , and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. Journal of Human Genetics: 150-159.

Alarcon, M, Cantor, RM, Liu, J, Gilliam, TC, Autism Genetic Resource Exchange Consortium, the, Geschwind, DH. 2002. Evidence for a Language Quantitative Trait Locus on Chromosome 7q. Am J Hum Genet.: 60-71.

Alarcón, M, Cantor, RM, Liu, J, Gilliam, TC, Geschwind, DH. 2002. Evidence for a Language Quantitative Trait Locus on Chromosome 7q in Multiplex Autism Families. The American Journal of Human Genetics 70: 60-71.

Alarcón, M, Yonan, a L, Gilliam, TC, Cantor, RM, Geschwind, DH. 2005. Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs. Molecular psychiatry 10: 747-57.

Allen, HL, Estrada, K, Lettre, G, Berndt, SI, Michael, N, Rivadeneira, F, Willer, CJ, Jackson, AU, Raychaudhuri, S, Ferreira, T, Wood, AR, Weyant, RJ. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832-838.

Allen-Brady, K, Cannon, D, Robison, R, McMahon, WM, Coon, H. 2010. A unified theory of autism revisited: linkage evidence points to chromosome X using a high-risk subset of AGRE families. Autism research : official journal of the International Society for Autism Research 3: 47-52.

Allison, D, Neale, M, Zannolli, R, Schork, N, Amos, C, Blangero, J. 1999. Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. Am J Hum Genet 65: 531-544.

Almasy, L, Blangero, J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62: 1198-211.

Altschul, S, Madden, T, Schaffer, A, Zhang, J, Zhang, Z, Miller, W, Lipman, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Anderson, G, Freedman, D, Cohen, D, Volkmar, F, Hoder, E, McPhedran, P, Minderaa, R, Hansen, C, Young, J. 1987. Whole blood serotonin in autistic and normal subjects. J Child Psychol Psychiatry 28: 885-900.

Anney, R, Klei, L, Pinto, D, Almeida, J, Bacchelli, E, Baird, G, Bolshakova, N, Bölte, S, Bolton, PF, Bourgeron, T, Brennan, S, Brian, J, Casey, J, Conroy, J, Correia, C, Corsello, C, Crawford, EL, de Jonge, M, Delorme, R, Duketis, E, Duque, F, Estes, A, Farrar, P, Fernandez, B a, Folstein, SE, Fombonne, E, Gilbert, J, Gillberg, C, Glessner, JT, Green, A, Green, J, Guter, SJ, Heron, E a, Holt, R, Howe, JL, Hughes, G, Hus, V, Igliozzi, R, Jacob, S, Kenny, GP, Kim, C, Kolevzon, A, Kustanovich, V, Lajonchere, CM, Lamb, J a, Law-Smith, M, Leboyer, M, Le Couteur, A, Leventhal, BL, Liu, X-Q, Lombard, F, Lord, C, Lotspeich, L, Lund, SC, Magalhaes, TR, Mantoulan, C, McDougle, CJ, Melhem, NM, Merikangas, A, Minshew, NJ, Mirza, GK, Munson, J, Noakes, C, Nygren, G, Papanikolaou, K, Pagnamenta, AT, Parrini, B, Paton, T, Pickles, A, Posey, DJ, Poustka, F, Ragoussis, J, Regan, R, Roberts, W, Roeder, K, Roge, B, Rutter, ML, Schlitt, S, Shah, N, Sheffield, VC, Soorya, L, Sousa, I, Stoppioni, V, Sykes, N, Tancredi, R, Thompson, AP, Thomson, S, Tryfon, A, Tsiantis, J, Van Engeland, H, Vincent, JB, Volkmar, F, Vorstman, J a S, Wallace, S, Wing, K, Wittemeyer, K, Wood, S, Zurawiecki, D, et al. 2012. Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Human molecular genetics 21: 4781-92.

Anney, R, Klei, L, Pinto, D, Regan, R, Conroy, J, Magalhaes, TR, Correia, C, Abrahams, BS, Sykes, N, Pagnamenta, AT, Almeida, J, Bacchelli, E, Bailey, AJ, Baird, G, Battaglia, A, Berney, T, Bolshakova, N, Bölte, S, Bolton, PF, Bourgeron, T, Brennan, S, Brian, J, Carson, AR, Casallo, G, Casey, J, Chu, SH, Cochrane, L, Corsello, C, Crawford, EL, Crossett, A, Dawson, G, de Jonge, M, Delorme, R, Drmic, I, Duketis, E, Duque, F, Estes, A, Farrar, P, Fernandez, B a, Folstein, SE, Fombonne, E, Freitag, CM, Gilbert, J, Gillberg, C, Glessner, JT, Goldberg, J, Green, J, Guter, SJ, Hakonarson, H, Heron, E a, Hill, M, Holt, R, Howe, JL, Hughes, G, Hus, V, Igliozzi, R, Kim, C, Klauck, SM, Kolevzon, A, Korvatska, O, Kustanovich, V, Lajonchere, CM, Lamb, J a, Laskawiec, M, Leboyer, M, Le Couteur, A, Leventhal, BL, Lionel, AC, Liu, X-Q, Lord, C, Lotspeich, L, Lund, SC, Maestrini, E, Mahoney, W, Mantoulan, C, Marshall, CR, McConachie, H, McDougle, CJ, McGrath, J, McMahon, WM, Melhem, NM, Merikangas, A, Migita, O, Minshew, NJ, Mirza, GK, Munson, J, Nelson, SF, Noakes, C, Noor, A, Nygren, G, Oliveira, G, Papanikolaou, K, Parr, JR, Parrini, B, Paton, T, Pickles, A, Piven, J, Posey, DJ, et al. 2010. A genome-wide scan for common alleles affecting risk for autism. Human molecular genetics 19: 4072-82.

Arking, DE, Cutler, DJ, Brune, CW, Teslovich, TM, West, K, Ikeda, M, Rea, A, Guy, M, Lin, S, Jr, EHC, Chakravarti, A. 2008. A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. Journal of Human Genetics: 160-164.

Auranen, M, Vanhala, R, Varilo, T, Ayers, K, Kempas, E, Ylisaukko-Oja, T, Sinsheimer, JS, Peltonen, L, Järvelä, I. 2002. A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25-27. American journal of human genetics 71: 777-90.

Auranen, M, Varilo, T, Alen, R, Vanhala, R, Ayers, K, Kempas, E, Ylisaukko-Oja, T, Peltonen, L, Järvelä, I. 2003. Evidence for allelic association on chromosome 3q25-27 in families with autism spectrum disorders originating from a subisolate of Finland. Molecular psychiatry 8: 879-84.

Badner, J a, Gershon, ES. 2002. Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. Molecular psychiatry 7: 56-66.

Bailey, a, Phillips, W, Rutter, M. 1996. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. Journal of child psychology and psychiatry, and allied disciplines 37: 89-126.

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

Bakkaloglu, B, O'Roak, B, Louvi, A, Gupta, A, Abelson, J, Morgan, T. 2008. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet 82: 165-173.

Barrett, J, Fry, B, Maller, J, Daly, M. 2005. Haploview: Analysis and visualization of LD and haplotype markers. Bioinformatics 21: 263-265.

Bartlett, C, Hou, L, Flax, J, Hare, A, Cheong, S, Fermano, Z, Zimmerman-Bier, B, Cartwright, C, Azaro, M, Buyske, S, Brzustowicz, L. 2013. A Genome-scan for Loci Shared by Autism Spectrum Disorder and Language Impairment. Am J Psychiatry in press.

Bartlett, CW, Flax, J, Logue, M, Vieland, VJ, Bassett, AS, Tallal, P, Brzustowicz, LM. 2002. A Major Susceptibility Locus for Specific Language Impairment is Located on 13q21. American Journal of Human Genetics 71: 45-55.

Bartlett, CW, Flax, JF, Fermano, Z, Hare, A, Hou, L, Petrill, S a, Buyske, S, Brzustowicz, LM. 2012. Gene × gene interaction in shared etiology of autism and specific language impairment. Biological psychiatry 72: 692-9.

Bartlett, CW, Flax, JF, Logue, MW, Smith, BJ, Vieland, VJ, Tallal, P, Brzustowicz, LM. 2004. Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. Hum Hered 57: 10-20.

Bartlett, CW, Goedken, R, Vieland, VJ. 2005. Effects of updating linkage evidence across subsets of data: reanalysis of the autism genetic resource exchange data set. American journal of human genetics 76: 688-95.

Bartlett, CW, Vieland, VJ. 2005. Two novel quantitative trait linkage analysis statistics based on the posterior probability of linkage: application to the COGA families. BMC genetics 6 Suppl 1: S121.

Baudet, M-L, Rattray, D, Harvey, S. 2007. Growth hormone and its receptor in projection neurons of the chick visual system: retinofugal and tectobulbar tracts. Neuroscience 148: 151-63.

Becker, K. 2004. The common variants/multiple disease hypothesis of common complex genetic disorders. Medical Hypotheses 62: 309-317.

Bell, PA, Chaturvedi, S, Gelfand, CA, Huang, CY, Kochersperger, M, Kopla, R, Modica, F, Pohl, M, Varde, S, Zhao, R, Zhao, X, Boyce-jacino, MT. 2002. SNPstream ® UHT : Ultra-High Throughput Drug Discovery. Discovery.

Ben-David, E, Shifman, S. 2012. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. PLoS genetics 8: e1002556.

Benayed, R, Choi, J, Matteson, PG, Gharani, N, Kamdar, S, Brzustowicz, LM, Millonig, JH. 2009. Autism-associated haplotype affects the regulation of the homeobox gene, ENGRAILED 2. Biological psychiatry 66: 911-7.

Benayed, R, Gharani, N, Rossman, I, Mancuso, V, Lazar, G, Kamdar, S, Bruse, SE, Tischfield, S, Smith, BJ, Zimmerman, R a, Dicicco-Bloom, E, Brzustowicz, LM, Millonig, JH. 2005. Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. American journal of human genetics 77: 851-68.

Berna-Erro, A, Braun, A, Kraft, R, Kleinschnitz, C, Schuhmann, MK, Stegner, D, Wultsch, T, Eilers, J, Meuth, SG, Stoll, G, Nieswandt, B. 2009. STIM2 Regulates Capacitive Ca2+ Entry in Neurons and Plays a Key Role in Hypoxic Neuronal Cell Death. Science Signaling 2: ra67.

Betancur, C, Corbex, M, Spielewoy, C, Philippe, a, Laplanche, JL, Launay, JM, Gillberg, C, Mouren-Siméoni, MC, Hamon, M, Giros, B, Nosten-Bertrand, M, Leboyer, M. 2002. Serotonin transporter gene polymorphisms and hyperserotonemia in autistic disorder. Molecular psychiatry 7: 67-71.

Bishop, D. 2010. Overlaps between autism and language impairment: Phenomimicry or shared etiology? Behav Genet 40: 618-629.

Bishop, D, Adams, C, Norbury, C. 2004a. Using nonword repetition to distinguish genetic and environmental influences on early literacy development: a study of 6-year-old twins. Am J Med Genet B Neuropsychiatr Genet 129B: 94-6.

Bishop, D, Bishop, S, Bright, P, James, C, Delaney, T, Tallal, P. 1999. Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. J Speech Lang Hear Res 42: 155-68.

Bishop, D, Mayberry, M, Wong, D, Maley, A, Hill, W, Hallmayer, J. 2004b. Are phonological processing deficits part of the broad autism phenotype? Am J Med Genet B Neuropsychiatr Genet 128B: 54-60.

Bishop, D, North, T, Donlan, C. 1996. Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. J Child Psychol Psychiatry 37: 391-403.

Bishop, DV, Maybery, M, Wong, D, Maley, A, Hallmayer, J. 2006. Characteristics of the broader phenotype in autism: a study of siblings using the children's communication checklist-2. Am J Med Genet B Neuropsychiatr Genet 141B: 117-122.

Bishop, DVM, Maybery, M, Maley, A, Wong, D, Hill, W, Hallmayer, J. 2004c. Using self-report to identify the broad phenotype in parents of children with autistic spectrum disorders: a study using the Autism-Spectrum Quotient. Journal of Child Psychology and Psychiatry 45: 1431-1436.

Blangero, J, Williams, J, Almasy, L. 2000. Robust LOD scores for variance component-based linkage analysis. Genet Epidemiol 19 Suppl 1: S8-14.

Bohl, D, Liu, S, Blanchard, S, Hocquemiller, M, Haase, G, Heard, J-M. 2008. Directed evolution of motor neurons from genetically engineered neural precursors. Stem cells (Dayton, Ohio) 26: 2564-75.

Bolte, S, Poustka, F, JN, C. 2008. Assessing autistic traits: cross-cultural validation of the social responsivenss scale (SRS). Autism research : official journal of the International Society for Autism Research 1: 354-363.

Bonora, E, Beyer, KS, Lamb, J a, Parr, JR, Klauck, SM, Benner, a, Paolucci, M, Abbott, a, Ragoussis, I, Poustka, a, Bailey, a J, Monaco, a P. 2003. Analysis of reelin as a candidate gene for autism. Molecular psychiatry 8: 885-92.

Boucard, A a, Chubykin, A a, Comoletti, D, Taylor, P, Südhof, TC. 2005. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron 48: 229-36.

Bradford, Y, Haines, J, Hutcheson, H, Gardiner, M, Braun, T, Sheffield, V, Cassavant, T, Huang, W, Wang, K, Vieland, V, Folstein, S, Santangelo, S, Piven, J. 2001. Incorporating language phenotypes strengthens evidence of linkage to autism. Am J Med Genet 105: 539-47.

Brielmaier, J, Matteson, PG, Silverman, JL, Senerth, JM, Kelly, S, Genestine, M, Millonig, JH, DiCicco-Bloom, E, Crawley, JN. 2012. Autism-relevant social abnormalities and cognitive deficits in engrailed-2 knockout mice. PloS one 7: e40914.

Broman, K, Weber, J. 1998. Estimation of pairwise relationships in the presence of genotyping errors. Am J Hum Genet 63: 1563-4.

Bruse, S, Moreau, M, Azaro, M, Zimmerman, R, Brzustowicz, L. 2008. Improvements to Bead Based Oligonucleotide Ligation SNP Genotyping Assays. Biotechniques 45: 559-571.

van Buuren, S, Groothuis-Oudshoorn, K. 2011. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 45: 1-67.

Buxbaum, JD, Silverman, J, Keddache, M, Smith, CJ, Hollander, E, Ramoz, N, Reichert, JG. 2004. Linkage analysis for autism in a subset families with obsessive-compulsive behaviors: evidence for an autism susceptibility gene on chromosome 1 and further support for susceptibility genes on chromosome 6 and 19. Mol Psychiatry 9: 144-150.

Buxbaum, JD, Silverman, JM, Smith, CJ, Kilifarski, M, Reichert, J, Hollander, E, Lawlor, BA, Fitzgerald, M, Greenberg, DA, Davis, KL. 2001. Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. Am J Hum Genet 68: 1514-20.

CDC. 2012. Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002) 61: 1-19.

CLSA, Barrett, S, Beck, JC, Bernier, R, Bisson, E, Braun, TA, Casavant, TL, Childress, D, Folstein, SE, Garcia, M, Gardiner, MB, Gilman, S, Haines, JL, Hopkins, K, Landa, R, Meyer, NH, Mullane, JA, Nishimura, DY, Palmer, P, Piven, J, Purdy, J, Santangelo, SL, Searby, C, Sheffield, V, Singleton, J, Slager, S, et al. 1999. An autosomal genomic screen for autism. Collaborative linkage study of autism. Am J Med Genet 88: 609-15.

Cantor, RM, Kono, N, Duvall, J a, Alvarez-Retuerto, A, Stone, JL, Alarcón, M, Nelson, SF, Geschwind, DH. 2005. Replication of autism linkage: fine-mapping peak at 17q21. American journal of human genetics 76: 1050-6.

Carbon, S, Ireland, A, Mungall, CJ, Shu, S, Marshall, B, Lewis, S. 2009. AmiGO: online access to ontology and annotation data. Bioinformatics (Oxford, England) 25: 288-9.

Carrow-Woolfolk, E. 1999. Comprehensive Assessment of Spoken Language. Circle Pines, MN: AGS.

Carter, M, Scherer, S. 2013. Autsim spectrum disorder in the genetics clinic: a review. Clinical Genetics: 1-9.

Chadman, KK, Gong, S, Scattoni, ML, Boltuck, SE, Gandhy, SU, Heintz, N, Crawley, JN. 2008. Minimal aberrant behavioral phenotypes of neuroligin-3 R451C knockin mice. Autism research : official journal of the International Society for Autism Research 1: 147-58.

Chagnon, Y. 2006. Shared susceptibility region on chromosome 15 between autisma and catatonia. Int Rev Neurobiol 72: 165-178.

Chapman, N, Estes, A, Munson, J, Bernier, R, Webb, S, Rothstein, J, Minshew, N, Dawson, G, Schellenberg, G, Wijsman, E. 2011. Genome-scan for IQ discrepancy in autism: evidence for loci on chromosomes 10 and 16. Human genetics 129: 59-70.

Cheh, M a, Millonig, JH, Roselli, LM, Ming, X, Jacobsen, E, Kamdar, S, Wagner, GC. 2006. En2 knockout mice display neurobehavioral and neurochemical alterations relevant to autism spectrum disorder. Brain research 1116: 166-76.

Chen, GK, Kono, N, Geschwind, DH, Cantor, RM. 2006. Quantitative trait locus analysis of nonverbal communication in autism spectrum disorder. Molecular psychiatry 11: 214-20.

Chen, K, Northington, FJ, Martin, LJ. 2010. Inducible nitric oxide synthase is present in motor neuron mitochondria and Schwann cells and contributes to disease mechanisms in ALS mice. Brain structure & function 214: 219-34.

Colton, CA, Wilcock, DM, Wink, DA, Davis, J, Van, WE, Vitek, MP. 2009. The Effects of NOS2 Gene Deletion on Mice Expressing Mutated Human AβPP. J Alzheimers Dis 15: 571-587.

Connolly, JJ, Glessner, JT, Hakonarson, H. 2013. A genome-wide association study of autism incorporating autism diagnostic interview-revised, autism diagnostic observation schedule, and social responsiveness scale. Child development 84: 17-33.

Conroy, J, Meally, E, Kearney, G, Fitzgerald, M, Gill, M, Gallagher, L. 2004. Serotonin transporter gene and autism: a haplotype analysis in an Irish autistic population. Molecular psychiatry 9: 587-93.

Constantino, J, Davis, S, Todd, R, Schindler, M, Gross, M, Brophy, S, Metzger, L, Shoushtari, C, Splinter, R, Reich, W. 2003. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. J Autism Dev Disord 33: 427-33.

Constantino, J, Gruber, C, Davis, S, Hayes, S, Passanante, N, Przybeck, T. 2004. The factor structure of autistic traits. Journal of child psychology and psychiatry, and allied disciplines 45: 719-726.

Constantino, J, Todd, R. 2003. Autistic Traits in the General Population: A Twin Study. Arch Gen Psychiatry 60: 524-530.

Constantino, JN, Todd, RD. 2005. Intergenerational transmission of subthreshold autistic traits in the general population. Biol Psychiatry 57: 655-660.

Constantino, JN, Zhang, Y, Frazier, T, Abbacchi, AM, Law, P. 2010. Sibling recurrence and the genetic epidemiology of autism. Psychiatry: Interpersonal and Biological Processes 167: 1349-1356.

Conti-Ramsden, G, Simkin, Z, Botting, N. 2006. The prevalence of autistic spectrum disorders in adolescents with a history of specific language impairment (SLI). J Child Psychol Psychiatry 47: 621-628.

Coon, H, Villalobos, M, Robison, R, Camp, N, Cannon, D, Allen-Brady, K, Miller, J, McMahon, W. 2010. Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. Molecular autism 1: 8.

Coriell Personalized Medicine Collaborative. 2013. Types of Risk: Absolute Risk and Relative Risk. Genetic Education: Understanding Risk.

Coutinho, a M, Oliveira, G, Morgadinho, T, Fesel, C, Macedo, TR, Bento, C, Marques, C, Ataíde, a, Miguel, T, Borges, L, Vicente, a M. 2004. Variants of the serotonin transporter gene (SLC6A4) significantly contribute to hyperserotonemia in autism. Molecular psychiatry 9: 264-71.

Cross, S, Kim, S-J, Weiss, L a, Delahanty, RJ, Sutcliffe, JS, Leventhal, BL, Cook, EH, Veenstra-Vanderweele, J. 2008. Molecular genetics of the platelet serotonin system in first-degree relatives of patients with autism. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 33: 353-60.

Davis, LK, Meyer, KJ, Rudd, DS, Librant, a L, Epping, E a, Sheffield, VC, Wassink, TH. 2009. Novel copy number variants in children with autism and additional developmental anomalies. Journal of neurodevelopmental disorders 1: 292-301.

Devlin, B, Bennett, P, Dawson, G, Figlewicz, D a, Grigorenko, EL, McMahon, W, Minshew, N, Pauls, D, Smith, M, Spence, MA, Rodier, PM, Stodgell, C, Schellenberg, GD. 2004. Alleles of a reelin CGG repeat do not convey liability to autism in a sample from the CPEA network. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 126B: 46-50.

Devlin, B, Cook, EH, Coon, H, Dawson, G, Grigorenko, EL, McMahon, W, Minshew, N, Pauls, D, Smith, M, Spence, M a, Rodier, PM, Stodgell, C, Schellenberg, GD. 2005. Autism and the serotonin transporter: the long and short of it. Molecular psychiatry 10: 1110-6.

Devlin, B, Scherer, SW. 2012. Genetic architecture in autism spectrum disorder. Current opinion in genetics & development 22: 229-37.

Dutta, S, Sinha, S, Ghosh, S, Chattergjee, A, Ahmed, S, Usha, R. 2008. Genetic analysis of reelin gene (RELN) SNPs: no association with autism spectrum disorder in the Indian population. Neurosci Lett 441: 56-60.

Duvall, JA, Lu, A, Cantor, RM, Todd, RD, Constantino, JN, Geschwind, DH. 2007. A quantitative trait locus analysis of social responsiveness in multiplex autism families. Am J Psychiatry 164: 656-662.

Ek, U, Norrelgen, F, Westerlund, J, Dahlman, A, Hultby, E, Fernell, E. 2012. Teenage outcomes after speech and language impairment at preschool age. Neuropsychiatric Disease and Treatment 8: 221-227.

Elia, J, Capasso, M, Zaheer, Z, Lantieri, F, Ambrosini, P, Berrettini, W, Devoto, M, Hakonarson, H. 2009. Candidate gene analysis in an on-going genome-wide association study of attention-deficit hyperactivity disorder: suggestive association signals in ADRA1A. Psychiatric Genetics 19: 134-141.

Elia, J, Gai, X, Xie, H, Perin, J, Geiger, E, Glessner, J. 2010. Rare structural variants found in attention-deficit hyper activity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry 15: 637-646.

Elston, R, Stewart, J. 1971. A general model for the genetic analysis of pedigree data. Hum Hered 21: 523-42.

Elston, RC, Lange, K. 1975. The prior probability of autosomal linkage. Ann Hum Genet 38: 341-350.

Etherton, M, Földy, C, Sharma, M, Tabuchi, K, Liu, X, Shamloo, M. 2011. Autism-linked neuroligin-3 R451C mutation differentially alters hippocampal and cortical synaptic function. PNAS 108: 13764-9.

Etherton, MR, Blaiss, CA, Powell, CM, Su, TC. 2009. Mouse neurexin-1 alpha deletion causes correlated electrophysiological and behavioral changes. PNAS 106: 17998-18003.

Ey, E, Leblond, CS, Bourgeron, T. 2011. Behavioral profiles of mouse models for autism spectrum disorders. Autism research : official journal of the International Society for Autism Research 4: 5-16.

Fanarraga, ML, Charité, J, Hage, WJ, De Graaff, W, Deschamps, J. 1997. Hoxb-8 gain-of-function transgenic mice exhibit alterations in the peripheral nervous system. Journal of neuroscience methods 71: 11-8.

Fatemi, SH, Snow, AV, Stary, JM, Araghi-Niknam, M, Reutiman, TJ, Lee, S, Brooks, AI, Pearce, D a. 2005. Reelin signaling is impaired in autism. Biological psychiatry 57: 777-87.

Fatemi, SH, Stary, JM, Egan, EA. 2002. Reduced blood levels of reelin as a vulnerability factor in pathophysiology of autistic disorder. Cellular and molecular neurobiology 22: 139-52.

Feuk, L, Kalervo, A, Lipsanen-Nyman, M, Skaug, J, Nakabayashi, K, Finucane, B, Hartung, D, Innes, M, Kerem, B, Nowaczyk, MJ, Rivlin, J, Roberts, W, Senman, L, Summers, A, Szatmari, P, Wong, V, Vincent, JB, Zeesman, S, Osborne, LR, Cardy, JO, Kere, J, Scherer, SW, Hannula-Jouppi, K. 2006. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. American journal of human genetics 79: 965-72.

Field, LL, Shumansky, K, Ryan, J, Truong, D, Swiergala, E, Kaplan, BJ. 2013. Dense-map genome scan for dyslexia supports loci at 4q13, 16p12, 17q22; suggests novel locus at 7q36. Genes, brain, and behavior 12: 56-69.

Flax, J, Hare, A, Fermano, Z, Zimmerman-Bier, B, Cartwright, C, Bartlett, C, Buyske, S, Brzustowicz, L. 2013. Characterizing Language Profiles in Families Ascertained for Autism and Specific Language Impairment. in preparation.

Flax, J, Realpe-Bonilla, T, Brzustowicz, L, Bartlett, C, Tallal, P. 2003. Specific language impairment in families: Evidence for co-occurrence with reading impairments. Journal of Speech, Language, and Hearing Research: JSLHR 46: 530-543.

Flax, JF, Hare, A, Azaro, M a, Vieland, VJ, Brzustowicz, LM. 2010. Combined linkage and linkage disequilibrium analysis of a motor speech phenotype within families ascertained for autism risk loci. Journal of neurodevelopmental disorders 2: 210-223.

De Fosse, L, Hodge, S, Makris, N, Kennedy, D, Caviness, Jr, V, McGrath, L, Steele, S, Ziegler, D, Herbert, M, Frazier, J, Tager-Flusberg, H, Harris, G. 2004. Language-association cortex asymmetry in autism and specific language impairment. Annals of Neurology 56: 757-66.

Freitag, CM. 2007. The genetics of autistic disorders and its clinical relevance: a review of the literature. Molecular psychiatry 12: 2-22.

Friedman, J, Vrijenhoek, T, Markx, S, Janssen, I, van der Vliet, W, Faas, B. 2008. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry 13: 261-266.

Gathercole, S, Willis, C, Baddeley, A, Emslie, E. 1994. The Children's Test of Nonword Repetition: A test of phonological working memory. Memory 2: 103-127.

Gernsbacher, MA, Sauer, E a, Geye, HM, Schweigert, EK, Hill Goldsmith, H. 2008. Infant and toddler oral- and manual-motor skills predict later speech fluency in autism. Journal of child psychology and psychiatry, and allied disciplines 49: 43-50.

Geschwind, DH. 2011. Genetics of autism spectrum disorders. Trends in cognitive sciences 15: 409-16.

Gharani, N, Benayed, R, Mancuso, V, Brzustowicz, LM, Millonig, JH. 2004. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. 9: 474-484.

Gibbs, V, Aldridge, F, Chandler, F, Witzlsperger, E, Smith, K. 2012. Brief report: an exploratory study comparing diagnostic outcomes for autism spectrum disorders under DSM-IV-TR with the proposed DSM-5 revision. Journal of autism and developmental disorders 42: 1750-6.

Gish, W. 2003. WUBLAST alignment engine.

Goodman, W, Price, L, Rasmussen, S, Mazure, C, Delgado, P, Heninger, G, Charney, D. 1989a. The Yale-Brown Obsessive Compulsive Scale. II. Validity. Archives of General Psychiatry 46: 1012-1016.

Goodman, W, Price, L, Rasmussen, S, Mazure, C, Fleischmann, R, Hill, C, Heninger, G, Charney, D. 1989b. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Archives of General Psychiatry 46: 1006-1011.

Gotham, K, Pickles, A, Lord, C. 2009. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. J Autism Dev Disord 39: 693-705.

Gotham, K, Risi, S, Pickles, A, Lord, C. 2007. The Autism Diagnostic Obersvation Schedule: Revised algorithms for improved diagnostic validity. J Autism Dev Disord 37: 613-627.

Gottesman, II, Ph, D, Gould, TD. 2003. Reviews and Overviews The Endophenotype Concept in Psychiatry : Etymology and Strategic Intentions. Am J Psychiatry 160: 636-645.

Grados, M a, Walkup, J, Walford, S. 2003. Genetics of obsessive-compulsive disorders: new findings and challenges. Brain & development 25 Suppl 1: S55-61.

Gregor, A, Albrecht, B, Bader, I, Bijlsma, E, Ekici, A, Engels, H. 2011. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC Med Genet 12: 106.

Grzadzinski, R, Huerta, M, Lord, C. 2013. DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. Molecular autism 4: 12.

Hallmayer, J. 1999. Chromosomes 1, 2, and 7 workshops. Am J Med Genet 88: 219-223.

Hallmayer, J, Cleveland, S, Torres, A, Phillips, J, Cohen, B, Torigoe, T, Miller, J, Fedele, A, Collins, J, Smith, K, Lotspeich, L, Croen, L a, Ozonoff, S, Lajonchere, C, Grether, JK, Risch, N. 2011. Genetic heritability and shared environmental factors among twin pairs with autism. Archives of general psychiatry 68: 1095-102.

Hammond, R, Vivancos, V, Naeem, A, Chilton, J, Mambetisaeva, E, Mambitisaeva, E, Andrews, W, Sundaresan, V, Guthrie, S. 2005. Slit-mediated repulsion is a key regulator of motor axon pathfinding in the hindbrain. Development (Cambridge, England) 132: 4483-95.

Harvey, S. 2001. Growth hormone in neural tissues of the chick embryo. Journal of Endocrinology 169: 487-498.

Harvey, S, Lavelin, I, Pines, M. 2002. Growth Hormone (GH) Action in the Brain. Journal of Molecular Neuroscience 18: 89-95.

Hayden, EP, Nurnberger, JI. 2006. Molecular genetics of bipolar disorder. Genes, brain, and behavior 5: 85-95.

He, Y, Xun, G, Xia, K, Hu, Z, Lv, L, Deng, Z, Zhao, J. 2001. No significant association between RELN polymorphism and autism in case-control and family-based association study in Chinese Han population. Psychiatry research 187: 462-464.

Herold, E, Connors, E. 2012. DSM-5 Proposed Criteria for Autism Spectrum Disorder Designed to Provide More Accurate Diagnosis and Treatment. Arlington, VA.

Higham, N. 2002. Computing the nearest correlation matrix - a problem from finance. IMA Journal of Numerical Analysis 22: 329-343.

Hirata, Y, Souza, RP, Lieberman, J a, Meltzer, HY, Kennedy, JL. 2010. Lack of association between HTR4 gene polymorphisms and schizophrenia in case-control and family-based samples. Psychiatry research 175: 176-8.

Hodgkinson, C a, Yuan, Q, Xu, K, Shen, P-H, Heinz, E, Lobos, E a, Binder, EB, Cubells, J, Ehlers, CL, Gelernter, J, Mann, J, Riley, B, Roy, A, Tabakoff, B, Todd, RD, Zhou, Z, Goldman, D. 2008. Addictions biology: haplotype-based analysis for 130 candidate genes on a single array. Alcohol and alcoholism (Oxford, Oxfordshire) 43: 505-15.

Horn, J. 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30: 179-185.

Hou, L, Phillips, C, Azaro, M, Brzustowicz, L, Bartlett, C. 2011. Validation of a cost-efficient multi-purpose SNP panel for disease based research. PloS one 6: e19699.

Hou, L, Wang, K, Bartlett, C. 2012. Evaluation of a bayesian model integration-based method for censored data. Human Heredity 74: 1-11.

Howlin, P, Mawhood, L, Rutter, M. 2000. Autism and developmental language disorder -- a follow-up comparison in early adult life. II: Social, behavioural, and psychiatric outcomes. Journal Of Child Psychology And Psychiatry 41: 561-578.

Huang, D, Sherman, B, Lempicki, R. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13.

Huang, D, Sherman, B, Lempicki, R. 2009b. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc 4: 44-57.

Huang, Y. 2011. Association statistics under the PPL framework. 143.

Huang, Y, Bartlett, CW, Segre, AM, O'Connell, JR, Mangin, L, Vieland, VJ. 2007. Exploiting gene x gene interaction in linkage analysis. BMC proceedings 1 Suppl 1: S64.

Huang, Y, Segre, A, O'Connell, J, Wang, H, Vieland, VJ. 2006. KELVIN: a 2nd generation distributed multiprocessor linkage and linkage disequilibrium analysis program.

Huerta, M, Bishop, SL, Duncan, A, Hus, V, Lord, C. 2012. Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. The American journal of psychiatry 169: 1056-64.

IBM Corp. 2012. IBM SPSS Statistics for Windows,21e. Armonk, NY: IBM Corp.

IMGSAC. 1998. A full genome screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. Hum Mol Genet 7: 571-8.

IMGSAC. 2001. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. American journal of human genetics 69: 570-81.

Ichtchenko, K, Hata, Y, Nguyen, T, Ullrich, B, Missler, M, Moomaw, C, Sudhof, T. 1995. Neuroligin 1: A Splice Site-Specific Ligand for Beta-Neurexins. Cell 81: 435-443.

International HapMap Consortium. 2003. The International HapMap Project. Nature 426: 789-796.

Iossifov, I, Ronemus, M, Levy, D, Wang, Z, Hakker, I, Rosenbaum, J, Yamrom, B, Lee, Y, Narzisi, G, Leotta, A, Kendall, J, Grabowska, E, Ma, B, Marks, S, Rodgers, L, Stepansky, A, Troge, J, Andrews, P, Bekritsky, M, Pradhan, K, Ghiban, E, Kramer, M, Parla, J, Demeter, R, Fulton, L, Fulton, R, Magrini, V, Ye, K, Darnell, J, Darnell, R, Mardis, E, Wilson, R, Schatz, M, McCombie, W, Wigler, M. 2012. De novo gene disruptions in children on the autistic spectrum. Neuron 74: 285-299.

Jamain, S, Quach, H, Betancur, C, Rastam, M, Colineaux, C, Gillberg, I, Soderstrom, H, Giros, B, Leboyer, M, Gillberg, C, Bourgeron, T, Study, PARIS (PARIS). 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34: 27-29.

James, SJ, Shpyleva, S, Melnyk, S, Pavliv, O, Pogribny, IP. 2013. Complex epigenetic regulation of Engrailed-2 (EN-2) homeobox gene in the autism cerebellum. Translational psychiatry 3: e232.

Joseph, R, Tager-Flusberg, H, Lord, C. 2002. Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. J Child Psychol Psychiatry 43: 807-821.

Kamio, Y, Inada, N, Moriwaki, A, Kuroda, M, Koyama, T, Tsujii, H, Kawakub, Y, Kuwabara, H, Tsuchiya, K, Uno, Y, Constantino, J. 2012. Quantitative autistic traits ascertained in a national survey of 22,529 Japanese schoolchildren. Acta psychiatrica Scandinavica.

Kanehisa, M, Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28: 27-30.

Kanehisa, M, Goto, S, Sato, Y, Furumichi, M, Tanabe, M. 2012. KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res 40: D109-D114.

Ke, X. 2012. Presence of multiple independent effects in risk loci of common complex human diseases. American journal of human genetics 91: 185-92.

Kent, W, Sugnet, C, Furey, T, Roskin, K, Pringle, T, Zahler, A. 2002. The human genome browser at UCSC. Genome Res 12: 996-1006.

Kessler, R, Chiu, W, Demler, O, Walters, E. 2005. Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). Archives of General Psychiatry 62: 617-27.

Kim, S-j, Cox, N, Courchesne, R, Lord, C, Corsello, C, Akshoomoff, N, Guter, S. 2002. Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder. Molecular Psychiatry 1: 278-288.

Kjelgaard, MM, Tager-Flusberg, H. 2001. An investigation of language impairment in autism: Implications for genetic subgroups. Language & Cognitive Processes 16: 287-308.

Klauck, SM, Poustka, F, Benner, a, Lesch, KP, Poustka, a. 1997. Serotonin transporter (5-HTT) gene variants associated with autism? Human molecular genetics 6: 2233-8.

Klei, L, Sanders, SJ, Murtha, MT, Hus, V, Lowe, JK, Willsey, a J, Moreno-De-Luca, D, Yu, TW, Fombonne, E, Geschwind, D, Grice, DE, Ledbetter, DH, Lord, C, Mane, SM, Martin, CL, Martin, DM, Morrow, EM, Walsh, C a, Melhem, NM, Chaste, P, Sutcliffe, JS, State, MW, Cook, EH, Roeder, K, Devlin, B. 2012. Common genetic variants, acting additively, are a major source of risk for autism. Molecular autism 3: 9.

Koishi, S, Yamamoto, K, Matsumoto, H, Koishi, S, Enseki, Y, Oya, A, Asakura, A, Aoki, Y, Atsumi, M, Iga, T, Inomata, J, Inoko, H, Sasaki, T, Nanba, E, Kato, N, Ishii, T, Yamazaki, K. 2006. Serotonin transporter gene promoter polymorphism and autism: a family-based genetic association study in Japanese population. Brain & development 28: 257-60.

Krakowiak, P, Walker, CK, Bremer, A a, Baker, AS, Ozonoff, S, Hansen, RL, Hertz-Picciotto, I. 2012. Maternal metabolic conditions and risk for autism and other neurodevelopmental disorders. Pediatrics 129: e1121-8.

Krebs, MO, Betancur, C, Leroy, S, Bourdel, MC, Gillberg, C, Leboyer, M. 2002. Absence of association between a polymorphic GGC repeat in the 5’ untranslated region of the reelin gene and autism. Molecular psychiatry 7: 801-4.

Lai, CS, Fisher, SE, Hurst, J a, Vargha-Khadem, F, Monaco, a P. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413: 519-23.

Lander, E, Green, P. 1987. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci U1SA 84: 2363-7.

Lauritsen, MB, Pedersen, CB, Mortensen, PB. 2005. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. Journal of child psychology and psychiatry, and allied disciplines 46: 963-71.

Leboyer, M, Philippe, a, Bouvard, M, Guilloud-Bataille, M, Bondoux, D, Tabuteau, F, Feingold, J, Mouren-Simeoni, MC, Launay, JM. 1999. Whole blood serotonin and plasma beta-endorphin in autistic probands and their first-degree relatives. Biological psychiatry 45: 158-63.

Lee, SK, Pfaff, SL. 2001. Transcriptional networks regulating neuronal identity in the developing spinal cord. Nature neuroscience 4 Suppl: 1183-91.

Lewis, B a, Freebairn, L a, Hansen, A, Gerry Taylor, H, Iyengar, S, Shriberg, LD. 2004. Family pedigrees of children with suspected childhood apraxia of speech. Journal of communication disorders 37: 157-75.

Lewis, B, Freebaim, L, Hansen, A, Miscimarra, L, Iyengar, S, Taylor, H. 2007. Speech and language skills of parents of children with speech sound disorders. Am J Speech Lang Pathol 16: 108-18.

Leyfer, OT, Tager-Flusberg, H, Dowd, M, Tomblin, JB, Folstein, SE. 2008. Overlap between autism and specific language impairment: comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule scores. Autism research : official journal of the International Society for Autism Research 1: 284-96.

Li, H, Li, Y, Shao, J, Li, R, Qin, Y, Xie, C, Zhao, Z. 2008. The association analysis of RELN and GRM8 genes with autistic spectrum disorder in Chinese Han population. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 147B: 194-200.

Li, J, Liu, J, Zhao, L, Ma, Y, Jia, M, Lu, T, Ruan, Y, Li, Q, Yue, W, Zhang, D, Wang, L. 2013. Association study between genes in Reelin signaling pathway and autism identifies DAB1 as a susceptibility gene in a Chinese Han population. Progress in neuro-psychopharmacology & biological psychiatry 44C: 226-232.

Li, J, Nguyen, L, Gleason, C, Lotspeich, L, Spiker, D, Risch, N, Myers, RM. 2004. Lack of evidence for an association between WNT2 and RELN polymorphisms and autism. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 126B: 51-7.

Li, X, Zou, H, Brown, WT. 2012. Genes associated with autism spectrum disorder. Brain Research Bulletin 88: 543-552.

Lichtenstein, P, Carlstrom, E, Rastam, M, Gillberg, C, Anckarsater, H. 2010. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry 167: 1357-1363.

Lindgren, K a, Folstein, SE, Tomblin, JB, Tager-Flusberg, H. 2009. Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives. Autism research : official journal of the International Society for Autism Research 2: 22-38.

Liu, X-Q, Paterson, AD, Szatmari, P. 2008. Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. Biological psychiatry 64: 561-70.

Logan, J, Petrill, S, Flax, J, Justice, L, Hou, L, Bassett, A. 2011. Genetic covariation underlying reading, language and related measures in a sample selected for specific language impairment. Behav Genet 41: 651-659.

Logue, MW, Vieland, VJ, Goedken, RJ, Crowe, RR. 2003. Bayesian analysis of a previously published genome screen for panic disorder reveals new and compelling evidence for linkage to chromosome 7. Am J Med Genet B Neuropsychiatr Genet 121B: 95-99.

Lord, C, Risi, S, Pickles, A. 2004. Developmental language disorders: From phenotypes to etiologies. In: Rice, M, Warren, S, editors. Trajectory of language development in autism spectrum disorders. Mawah, NJ: Lawrence Earlbaum, p 7-29.

Lord, C, Rutter, M, Le Couteur, A. 1994. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord Oct 24: 659-85.

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

Losh, M, Adolphs, R, Poe, MD, Couture, S, Penn, D, Baranek, GT, Piven, J. 2009. Neuropsychological Profile of Autism and the Broad Autism Phenotype. Arch Gen Psychiatry 66: 518-526.

Losh, M, Childress, D, Lam, K, Piven, J. 2008. Defining key features of the broad autism phenotype: a comparison across parents of multiple- and single-incidence autism families. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 147B: 424-33.

Luedi, PP, Dietrich, FS, Weidman, JR, Bosko, JM, Jirtle, RL, Hartemink, AJ. 2007. Computational and experimental identification of novel human imprinted genes. Genome research 17: 1723-30.

MacNeil, LK, Mostofsky, SH. 2012. Specificity of dyspraxia in children with autism. Neuropsychology 26: 165-71.

Maher, B, Reimers, M, Riley, B, Kendler, K. 2010. Allelic Heterogeneity in Genetic Association Meta-Analysis: An Application to DTNBP1 and Schizophrenia. Human Heredity 69: 71-79.

De Marco Garcia, NV, Jessel, TM. 2008. Early motor neuron pool identity and muscle nerve trajectory defined by post-mitotic restrictions in Nkx6.1 activity. New York 57: 217-231.

Maski, KP, Jeste, SS, Spence, SJ. 2011. Common neurological co-morbidities in autism spectrum disorders. Current opinion in pediatrics 23: 609-15.

Matis, C, Oury, F, Remacle, S, Lampe, X, Gofflot, F, Picard, JJ, Rijli, FM, Rezsohazy, R. 2007. Identification of Lmo1 as part of a Hox-dependent regulatory network for hindbrain patterning. Developmental dynamics : an official publication of the American Association of Anatomists 236: 2675-84.

Matise, T, Chen, F, Chen, W, De La Vega, F, Hansen, M, He, C. 2007. A second-generation combined linkage physical map of the human genome. Genome Res 17: 1783-6.

Mattila, M-L, Kielinen, M, Linna, S-L, Jussila, K, Ebeling, H, Bloigu, R, Joseph, RM, Moilanen, I. 2011. Autism spectrum disorders according to DSM-IV-TR and comparison with DSM-5 draft criteria: an epidemiological study. Journal of the American Academy of Child and Adolescent Psychiatry 50: 583-592.e11.

Mawhood, L, Howlin, P, Rutter, M. 2000. Autism and developmental language disorder -- a comparative follow-up in early adult life. I: Cognitive and language outcomes. Journal Of Child Psychology And Psychiatry 41: 547-559.

McCanlies, EC, Fekedulegn, D, Mnatsakanova, A, Burchfiel, CM, Sanderson, WT, Charles, LE, Hertz-Picciotto, I. 2012. Parental occupational exposures and autism spectrum disorder. Journal of autism and developmental disorders 42: 2323-34.

McCauley, JL, Li, C, Jiang, L, Olson, LM, Crockett, G, Gainer, K, Folstein, SE, Haines, JL, Sutcliffe, JS. 2005. Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. BMC medical genetics 6: 1.

McCauley, JL, Olson, LM, Dowd, M, Amin, T, Steele, a, Blakely, RD, Folstein, SE, Haines, JL, Sutcliffe, JS. 2004. Linkage and association analysis at the serotonin transporter (SLC6A4) locus in a rigid-compulsive subset of autism. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 127B: 104-12.

McGregor, K, Berns, A, Owen, A, Michels, S, Duff, D, Bahnsen, A, Lloyd, M. 2012. Associations between syntax and the lexicon among children with or without ASD and language impairment. Journal of Autism and Developmental Disorders 42: 35-47.

McPartland, JC, Reichow, B, Volkmar, FR. 2012. Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. Journal of the American Academy of Child and Adolescent Psychiatry 51: 368-83.

Mefford, H, Muhle, H, Ostertag, P, von Spiczak, S, Buysse, K, Baker, C. 2010. Genome-wide copy number variation in epilepsy: Novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet 6: e1000962.

Merritt 2nd, J, Jalal, S, Barbaresi, W, Babovic-Vuksanovic, D. 2005. 14q32.3 deletion syndrome with autism. Am J Hum Genet A 133A: 99-100.

Micali, S, Fombonne, E, Chakrabarti, N. 2004. The broad autism phenotype: Findings from an epidemiological survey. Autism 8: 16.

Michaelson, JJ, Shi, Y, Gujral, M, Zheng, H, Malhotra, D, Jin, X, Jian, M, Liu, G, Greer, D, Bhandari, A, Wu, W, Corominas, R, Peoples, A, Koren, A, Gore, A, Kang, S, Lin, GN, Estabillo, J, Gadomski, T, Singh, B, Zhang, K, Akshoomoff, N, Corsello, C, McCarroll, S, Iakoucheva, LM, Li, Y, Wang, J, Sebat, J. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 151: 1431-42.

Missler, M, Zhang, W, Rohlmann, A, Kattenstroth, G, Hammer, RE, Gottmann, K, Südhof, TC. 2003. Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis. Nature 423: 939-48.

Molloy, CA, Keddache, M, Martin, LJ. 2005. Evidence for linkage on 21q and 7q in a subset of autism characterized by developmental regression. Mol Psychiatry 10: 741-746.

Moreno-De-Luca, D, Sanders, SJ, Willsey, a J, Mulle, JG, Lowe, JK, Geschwind, DH, State, MW, Martin, CL, Ledbetter, DH. 2012. Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Molecular psychiatry: 1-6.

Mukhopadhyay, S. 2013. HALO: Helping Autism through Learning and Outreach. Soma & Tito.

Mulder, EJ, Anderson, GM, Kema, IP, Brugman, AM, Ketelaars, CEJ, de Bildt, A, van Lang, NDJ, den Boer, J a, Minderaa, RB. 2005. Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 133B: 93-6.

Murphy, M, Bolton, P, Pickles, A, Fombonne, E, Piven, J, Rutter, M. 2000. Personality Traits of the Relatives of Autistic Probands. Psychological Medicine 30: 1411-1424.

National Fragile X Foundation. 2013. Autism and Fragile X Syndrome.

Neale, BM, Kou, Y, Liu, L, Ma'ayan, A, Samocha, KE, Sabo, A, Lin, C-F, Stevens, C, Wang, L-S, Makarov, V, Polak, P, Yoon, S, Maguire, J, Crawford, EL, Campbell, NG, Geller, ET, Valladares, O, Schafer, C, Liu, H, Zhao, T, Cai, G, Lihm, J, Dannenfelser, R, Jabado, O, Peralta, Z, Nagaswamy, U, Muzny, D, Reid, JG, Newsham, I, Wu, Y, Lewis, L, Han, Y, Voight, BF, Lim, E, Rossin, E, Kirby, A, Flannick, J, Fromer, M, Shakir, K, Fennell, T, Garimella, K, Banks, E, Poplin, R, Gabriel, S, DePristo, M, Wimbish, JR, Boone, BE, Levy, SE, Betancur, C, Sunyaev, S, Boerwinkle, E, Buxbaum, JD, Cook, EH, Devlin, B, Gibbs, R a, Roeder, K,

Schellenberg, GD, Sutcliffe, JS, Daly, MJ. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485: 242-5.

Newbury, DF, Bonora, E, Lamb, J a, Fisher, SE, Lai, CSL, Baird, G, Jannoun, L, Slonims, V, Stott, CM, Merricks, MJ, Bolton, PF, Bailey, a J, Monaco, a P. 2002. FOXP2 is not a major susceptibility gene for autism or specific language impairment. American journal of human genetics 70: 1318-27.

Newbury, DF, Fisher, SE, Monaco, AP. 2010. Recent advances in the genetics of language impairment. Genome medicine 2: 6.

Newbury, DF, Paracchini, S, Scerri, TS, Winchester, L, Addis, L, Richardson, AJ, Walter, J, Stein, JF, Talcott, JB, Monaco, a P. 2011. Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. Behavior genetics 41: 90-104.

Nsengimana, J, Bishop, D. 2012. Design Considerations for Genetic Linkage and Association Studies. Methods in Molecular Biology 850: 237-262.

Nussbaum, R, McInnes, R, Williard, H. 2007. Genetics of Common Disorders with Complex Inheritance. In: Thompson & Thompson Genetics in Medicine. Philadelphia, PA: Saunders Elsevier, p 600.

OMIM. 2013. Tuberous Sclerosis. Online Mendelian Inheritance in Man, OMIM®: MIM Number: #613254.

Ozonoff, S, Young, GS, Carter, A, Messinger, D, Yirmiya, N, Zwaigenbaum, L, Bryson, S, Carver, LJ, Constantino, JN, Dobkins, K, Hutman, T, Iverson, JM, Landa, R, Rogers, SJ, Sigman, M, Stone, WL. 2011. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics 128: e488-95.

O'Connell, JR, Weeks, DE. 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. 63: 259-266.

O'Roak, BJ, Deriziotis, P, Lee, C, Vives, L, Schwartz, JJ, Girirajan, S, Karakoc, E, Mackenzie, AP, Ng, SB, Baker, C, Rieder, MJ, Nickerson, D a, Bernier, R, Fisher, SE, Shendure, J, Eichler, EE. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature genetics 43: 585-9.

O'Roak, BJ, Vives, L, Girirajan, S, Karakoc, E, Krumm, N, Coe, BP, Levy, R, Ko, A, Lee, C, Smith, JD, Turner, EH, Stanaway, IB, Vernot, B, Malig, M, Baker, C, Reilly, B, Akey, JM, Borenstein, E, Rieder, MJ, Nickerson, D a, Bernier, R, Shendure, J, Eichler, EE. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485: 246-50.

Pampanos, A, Volaki, K, Kanavakis, E, Papandreou, O, Youroukos, S, Thomaidis, L, Karkelis, S, Tzetis, M, Kitsiou-Tzeli, S. 2009. A substitution involving the NLGN4 gene associated with autistic behavior in the Greek population. Genet Test Mol Biomarkers 13: 611-5.

Persico, a M, D'Agruma, L, Maiorano, N, Totaro, a, Militerni, R, Bravaccio, C, Wassink, TH, Schneider, C, Melmed, R, Trillo, S, Montecchi, F, Palermo, M, Pascucci, T, Puglisi-Allegra, S, Reichelt, KL, Conciatori, M, Marino, R, Quattrocchi, CC, Baldi, a, Zelante, L, Gasparini, P, Keller, F. 2001. Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. Molecular psychiatry 6: 150-9.

Persico, a M, Militerni, R, Bravaccio, C, Schneider, C, Melmed, R, Conciatori, M, Damiani, V, Baldi, a, Keller, F. 2000. Lack of association between serotonin transporter gene promoter variants and autistic disorder in two ethnically distinct samples. American journal of medical genetics 96: 123-7.

Peter, B, Raskind, WH, Matsushita, M, Lisowski, M, Vu, T, Berninger, VW, Wijsman, EM, Brkanac, Z. 2011. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. Journal of neurodevelopmental disorders 3: 39-49.

Petit, E, Hérault, J, Martineau, J, Perrot, a, Barthélémy, C, Hameury, L, Sauvage, D, Lelord, G, Müh, JP. 1995. Association study with two markers of a human homeogene in infantile autism. Journal of medical genetics 32: 269-74.

Petrin, A, Giacheti, C, Maximino, L, Abramides, D, Zanchetta, S, Rossi, N. 2010. Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. Am J Med Genet A 152A: 3164-3172.

Peñagarikano, O, Geschwind, DH. 2012. What does CNTNAP2 reveal about autism spectrum disorder? Trends in molecular medicine 18.

Philippe, A, Martinez, M, Guilloud-Bataille, M, Gillberg, C, Rastam, M, Sponheim, E, Coleman, M, Zappella, M, Aschauer, H, Van Maldergem, L, Penet, C, Feingold, J, Brice, A, Leboyer, M, van Malldergerme, L. 1999. Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. Hum Mol Genet 8: 805-12.

Pickles, A, St Clair, MC, Conti-Ramsden, G. 2012. Communication and Social Deficits in Relatives of Individuals with SLI and Relatives of Individuals with ASD. Journal of autism and developmental disorders.

Pilowsky, T, Yirmiya, N, Shalev, R, Gross-Tsur, V. 2003. Language abilities of siblings of children with autism. J Child Psychol Psychiatry 44: 914-925.

Pinto, D, Pagnamenta, AT, Klei, L, Anney, R, Merico, D, Regan, R, Conroy, J, Magalhaes, TR, Correia, C, Abrahams, BS, Almeida, J, Bacchelli, E, Bader, GD, Bailey, AJ, Baird, G, Battaglia, A, Berney, T, Bolshakova, N, Bölte, S, Bolton, PF, Bourgeron, T, Brennan, S, Brian, J, Bryson, SE, Carson, AR, Casallo, G, Casey, J, Chung, BHY, Cochrane, L, Corsello, C, Crawford, EL, Crossett, A, Cytrynbaum, C, Dawson, G, de Jonge, M, Delorme, R, Drmic, I, Duketis, E, Duque, F, Estes, A, Farrar, P, Fernandez, B a, Folstein, SE, Fombonne, E, Freitag, CM, Gilbert, J, Gillberg, C, Glessner, JT, Goldberg, J, Green, A, Green, J, Guter, SJ, Hakonarson, H, Heron, E a, Hill, M, Holt, R, Howe, JL, Hughes, G, Hus, V, Igliozzi, R, Kim, C, Klauck, SM,

Kolevzon, A, Korvatska, O, Kustanovich, V, Lajonchere, CM, Lamb, J a, Laskawiec, M, Leboyer, M, Le Couteur, A, Leventhal, BL, Lionel, AC, Liu, X-Q, Lord, C, Lotspeich, L, Lund, SC, Maestrini, E, Mahoney, W, Mantoulan, C, Marshall, CR, McConachie, H, McDougle, CJ, McGrath, J, McMahon, WM, Merikangas, A, Migita, O, Minshew, NJ, Mirza, GK, Munson, J, Nelson, SF, Noakes, C, Noor, A, Nygren, G, Oliveira, G, Papanikolaou, K, Parr, JR, Parrini, B, Paton, T, et al. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466: 368-72.

Price, A, Patterson, N, Plenge, R, Weinblatt, M, Shadick, N, Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-909.

PubMed Health. 2012. Rett Syndrome. A.D.A.M. Medical Encyclopedia.

Purcell, S, Neale, B, Todd-Brown, K, Thomas, L, Ferreira, M, Bender, D, Maller, J, Sklar, P, de Bakker, P, Daly, M, Sham, P. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

Qiao, Y, Tyson, C, Hrynchak, M, Lopez-Rangel, E, Hildebrand, J, Martell, S, Fawcett, C, Kasmara, L, Calli, K, Harvard, C, Liu, X, Holden, J, Lewis, S, Rajcan-Separovic, E. 2013. Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. Clinical Genetics 83: 145-154.

Ransome, MI, Goldshmit, Y, F, P, Waters, MJ, Turnley, AM. 2004. Comparative analysis of CNS populations in knockout mice with altered growth hormone responsiveness. Neuroscience 19.

Rao, D, Vogler, G, Borecki, I, Province, M, Russell, J. 1987. Robustness of path analysis of family resemblance against deviations from multivariate normality. Hum Hered 37: 107-112.

Rapin, I. 1996. Practitioner review: Developmental language disorders: A clinical update. Journal of Child Psychology and Psychiatry, and Allied Disciplines 37: 643-655.

Rapin, I, Dunn, M. 2003. Update on the language disorders of individuals on the autistic spectrum. Brain and Development 25: 166-172.

Rapin, I, Dunn, M, Allen, D, Stevens, M, Fein, D. 2009. Subtypes of language disorders in school-age children with autism. Developmental Neuropsychology 34: 66-84.

Raskind, W, Hsu, L, Berninger, V, Thomson, J, Wijsman, E. 2000. Familial aggregation of dyslexia phenotypes. Behav Genet 30: 385-396.

Rehnström, K, Ylisaukko-oja, T, Nummela, I, Ellonen, P, Vanhala, R, Wendt, LV, Järvelä, I, Peltonen, L. 2010. Europe PMC Funders Group Allelic Variants in HTR3C Show Association With Autism. American Journal Of Medical Genetics Part B Neuropsychiatric Genetics.

Reichelt, a C, Rodgers, RJ, Clapcote, SJ. 2012. The role of neurexins in schizophrenia and autistic spectrum disorder. Neuropharmacology 62: 1519-26.

Reichow, B, Salamack, S, Paul, R, Volkmar, F, Klin, A. 2008. Pragmatic assessment in autism spectrum disorders: A comparison of a standard measure with parent report. Commun Disord Q 29: 169-176.

Riches, N, Loucas, T, Baird, G, Charman, T, Simonoff, E. 2011. Nonword repetition in adolescents with specific language impairment and autism plus language impairments: A qualitative analysis. J Commun Disord 44: 23-26.

Ricketts, J. 2011. Research review: Reading comprehension in developmental disorders of language and communication. Journal Of Child Psychology And Psychiatry, and Allied Disciplines 52: 1111-1123.

Risch, N, Spiker, D, Lotspeich, L, Nouri, N, Hinds, D, Hallmayer, J, Kalaydjieva, L, McCague, P, Dimiceli, S, Pitts, T, Nguyen, L, Yang, J, Harper, C, Thorpe, D, Vermeer, S, Young, H, Hebert, J, Lin, A, Ferguson, J, Chiotti, C, Wiese-Slater, S, Rogers, T, Salmon, B, Nicholas, P, Myers, RM, et al. 1999. A genomic screen of autism: evidence for a multilocus etiology. Am J Hum Genet 65: 493-507.

Roberts, J, Rice, M, Tager-Flusberg, H. 2004. Tense marking in children with autism. Applied Psycholinguistics 25: 429-448.

Robinson, E, Koenen, K, McCormick, M, Munir, K, Hallett, V, Happe, F. 2012. A multivariate twin study of autistic traits in 12-year-olds: Testing the fractionable autism triad hypothesis. Behav Genet 42: 245-255.

Robinson, E, Koenen, K, McCormick, M, Munir, K, Hallett, V, Happe, F. 2011. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). Arch Gen Psychiatry 68: 1113-1121.

Rosenberg, R, Daniels, A, Law, P, Kaufmann, W. 2009. Trends in autism spectrum disorder diagnoses: 1994-2007. J Autism Dev Disord 39: 1099-1111.

Royall, R. 1997. Statistical evidence: a likelihood paradigm. London.

Rubin, D. 1987. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.

Ruser, T, Arin, D, Dowd, M, Putnam, S, Winklosky, B, Rosen-Sheidley, B. 2007. Communicative competence in parents of children with autism and parents of children with specific language impairment. J Autism Dev Disord 37: 1323-1336.

Rutter, M, Le Couteur, A, Lord, C. 2003. The autism diagnostic interview, revised (ADI-R). Los Angeles: Western Psychological Services.

Salyakina, D, Ma, DQ, Jaworski, JM, Konidari, I, Whitehead, PL, Henson, R, Martinez, D, Robinson, JL, Sacharow, S, Wright, HH, Abramson, RK, Gilbert, JR, Cuccaro, ML, Pericak-Vance, M a. 2010. Variants in several genomic regions associated with asperger disorder. Autism research : official journal of the International Society for Autism Research 3: 303-10.

Sanders, SJ, Murtha, MT, Gupta, AR, Murdoch, JD, Raubeson, MJ, Willsey, a J, Ercan-Sencicek, a G, DiLullo, NM, Parikshak, NN, Stein, JL, Walker, MF, Ober, GT, Teran, N a, Song, Y, El-Fishawy, P, Murtha, RC, Choi, M, Overton, JD, Bjornson, RD, Carriero, NJ, Meyer, K a, Bilguvar, K, Mane, SM, Sestan, N, Lifton, RP, Günel, M, Roeder, K, Geschwind, DH, Devlin, B, State, MW. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485: 237-41.

Scahill, L, McDougle, C, Williams, S, Dimitropoulos, A, Aman, M, McCracken, J, Tierney, E, Arnold, L, Cronin, P, Grados, M, Ghuman, J, Koenig, K, Lam, K, McGough, J, Posey, D, Ritz, L, Swiezy, N, Vitiello, B. 2006. Children's Yale-Brown Obsessive Compulsive Scale modified for pervasive developmental disorders. Journal of the American Academy of Child and Adolescent Psychiatry 45: 1114-1123.

Scahill, L, Riddle, M, McSwiggin-Hardin, M, Ort, S, King, R, Goodman, W, Cicchetti, D, Leckman, J. 1997. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. Journal of the American Academy of Child and Adolescent Psychiatry 69: 844-852.

Schellenberg, GD, Dawson, G, Sung, YJ, Estes, A, Munson, J, Rosenthal, E, Rothstein, J, Flodman, P, Smith, M, Coon, H, Leong, L, Yu, CE, Stodgell, C, Rodier, PM, Spence, MA, Minshew, N, McMahon, WM, Wijsman, EM. 2006. Evidence for multiple loci from a genome scan of autism kindreds. Mol Psychiatry 11: 979,1049-1060.

Schmidt, R, Hansen, R, Hartiala, J, Allayee, H, Schmidt, L, Tancredi, D, Tassone, F, Hertz-Picciotto, I. 2011. Prenatal vitamins, one-carbon metabolism gene variants, and risk for autism. Epidemiology 22: 476-485.

Self, S, Liang, K. 1987. Asymptotic properties of maxiumum likelihood estimators and likelihood ratio tests under nonstandard conditions. j Am Stat Assoc 82: 605-610.

Semel, E, Wiig, E, Secord, W. 2003. Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4). Toronto: The Psychological Corporation/A Harcourt Assessment Company.

Serajee, FJ, Zhong, H, Mahbubul Huq, a HM. 2006. Association of Reelin gene polymorphisms with autism. Genomics 87: 75-83.

Shao, Y, Cuccaro, ML, Hauser, ER, Raiford, KL, Menold, MM, Wolpert, CM, Ravan, S a, Elston, L, Decena, K, Donnelly, SL, Abramson, RK, Wright, HH, DeLong, GR, Gilbert, JR, Pericak-Vance, M a. 2003. Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. American journal of human genetics 72: 539-48.

Shao, Y, Raiford, KL, Wolpert, CM, Cope, HA, Ravan, SA, Ashley-Koch, AA, Abramson, RK, Wright, HH, DeLong, RG, Gilbert, JR, Cuccaro, ML, Pericak-Vance, MA. 2002. Phenotypic Homogeneity Provides Increased Support for Linkage on Chromosome 2 in Autistic Disorder. American Journal of Human Genetics 70: 1058-1061.

Sharma, J, Arieff, Z, Gameeldien, H, Davids, M, Kaur, M, van der Merwe, L. 2013. Association analysis of two single-nucleotide polymorphisms of the RELN gene with autism in the South African population. Genet Test Mol Biomarkers 17: 93-8.

Shelton, JF, Hertz-picciotto, I, Pessah, IN. 2012. Review Tipping the Balance of Autism Risk : Potential Mechanisms Linking Pesticides. Environmental Health Perspectives 120: 944-951.

Sherry, S, Ward, M, Kholodov, M, Baker, J, Phan, L, Smigielski, E, Sirotkin, K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29: 308-11.

Shinoda, Y, Sadakata, T, Furuichi, T. 2013. Animal Models of Autism Spectrum Disorder (ASD): A Synaptic-Level Approach to Autistic-Like Behavior in Mice. Experimental animals / Japanese Association for Laboratory Animal Science 62: 71-8.

Shriberg, L, Aram, D, Kwiatkowski, J. 1997. Developmental apraxia of speech: I. Descriptive and theoretical perspectives. J Speech Lang Hear Res 40: 273-85.

Shriberg, LD, Paul, R, Black, LM, van Santen, JP. 2011. The hypothesis of apraxia of speech in children with autism spectrum disorder. Journal of autism and developmental disorders 41: 405-26.

Sikora, DM, Pettit-kekel, K, Penfield, J, Merkens, LS, Steiner, RD. 2006. The Near Universal Presence of Autism Spectrum Disorders in Children With Smith – Lemli – Opitz Syndrome. American Journal of Medical Genetics Part A 140: 1511-1518.

Simmons, TR, Flax, JF, Azaro, MA, Hayter, JE, Justice, LM, Petrill, SA, Bassett, AS, Tallal, P, Brzustowicz, LM, Bartlett, CW. 2010. Increasing genotype-phenotype model determinism: application to bivariate reading/language traits and epistatic interactions in language-impaired families. Human Heredity. 70: 232-244.

Skaar, D a, Shao, Y, Haines, JL, Stenger, JE, Jaworski, J, Martin, ER, DeLong, GR, Moore, JH, McCauley, JL, Sutcliffe, JS, Ashley-Koch, a E, Cuccaro, ML, Folstein, SE, Gilbert, JR, Pericak-Vance, M a. 2005. Analysis of the RELN gene as a genetic risk factor for autism. Molecular psychiatry 10: 563-71.

Smit, A, Hubley, R, Green, P. 2010. RepeatMasker Open 3.0.

Smith, C. 1963. Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet 27: 175-82.

Sparrow, S, Balla, D, Cicchetti, D. 1984. Vineland Adaptive Behavior Scales, Expanded Edition. Circle Pines, MN: American Guidance Service.

Spence, S, Cantor, R, Chung, L, Kim, S, Geschwind, D, Alarcon, M. 2006. Stratification based on language-related endophenotypes in Autism: attempt to replicate reported linkage. Am J Hum Genet Par B Neuropsychiatr Genet 141B: 591-8.

Steele, M, Al-Adeimi, M, Siu, V, Fan, Y. 2011. Brief report: A case of autism with interstitial deltion of chromosome 13. J Autism Dev Disord 31: 231-234.

Steer, C, Golding, J, Bolton, P. 2010. Traits contributing to the autistic spectrum. PLoS ONE 5.

Stone, JL, Merriman, B, Cantor, RM, Yonan, AL, Gilliam, TC, Geschwind, DH, Nelson, SF. 2004. Evidence for sex-specific risk alleles in autism spectrum disorder. American journal of human genetics 75: 1117-23.

Stromswold, K. 2008. The genetics of speech and language impairments. The New England Journal of Medicine 359: 2381-2383.

Stromswold, K. 2001. The heritability of language: A review and metaanalysis of twin adoption and linkage studies. Language 77: 647-723.

Sullivan, K, Sharda, M, Greenson, J, Dawson, G, Singh, NC. 2013. A novel method for assessing the development of speech motor function in toddlers with autism spectrum disorders. Frontiers in integrative neuroscience 7: 17.

Sutcliffe, JS, Delahanty, RJ, Prasad, HC, McCauley, JL, Han, Q, Jiang, L, Li, C, Folstein, SE, Blakely, RD. 2005. Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. American journal of human genetics 77: 265-79.

Suzuki, T, Iwata, N, Kitamura, Y, Kitajima, T, Yamanouchi, Y, Ikeda, M, Nishiyama, T, Kamatani, N, Ozaki, N. 2003. Association of a haplotype in the serotonin 5-HT4 receptor gene (HTR4) with Japanese schizophrenia. American Journal of Medical Genetics B Neuropsychiatric Genetics 121B: 7-13.

Tabuchi, K, Blundell, J, Etherton, MR, Hammer, RE, Liu, X, Powell, CM, Südhof, TC. 2007. A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. Science (New York, N.Y.) 318: 71-6.

Tager-Flusberg, H, Paul, R, Lord, C. 2005. Language and communication in autism. In: Volkmar, F, Paul, R, Klin, A, Cohen, D, editors. Handbook of autism and pervasive developmental disorders. New York: Wiley, p 335-64.

Takeuchi, K, Gertner, M, Zhou, J, Parada, L, Bennett, M, Zukin, R. 2013. Dysregulation of synaptic plasticity precedes appearance of morphological defects in a Pten conditional knockout mouse model of autism. PNAS 110: 4738-4743.

Tallal, P, Hirsch, LS, Realpe-Bonilla, T, Miller, S, Brzustowicz, LM, Bartlett, C, Flax, JF. 2001. Familial aggregation in specific language impairment. J Speech Lang Hear Res 44: 1172-1182.

Taniai, H, Nishiyama, T, Miyachi, T, Imaeda, M, Sumi, S. 2008. Genetic influences on the broad spectrum of autism: Study of proband-ascertained twins. Am J Med Genet B Neuropsychiatr Genet 147B: 844-849.

Taper, M, Lele, S. 2004. The nature of scientific evidence. Chicago: Univ Chicago Press.

The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073.

Tomblin, B. 2011. Co-morbidity of autism and SLI: Kinds, kin and complexity. Int J Lang Commun Disord 46: 127-137.

Tomblin, JB. 1989. Familial Concentration of Developmental Language Impairment. J Speech Hear Disord 54: 287-295.

Tomblin, JB, Records, NL, Buckwalter, P, Zhang, X, Smith, E, O'Brien, M. 1997. Prevalence of specific language impairment in kindergarten children. JSHR 40: 1245-1260.

Tordjman, S, Gutknecht, L, Carlier, M, Spitz, E, Antoine, C, Slama, F, Carsalade, V, Cohen, DJ, Ferrari, P, Roubertoux, PL, Anderson, GM. 2001. Role of the serotonin transporter gene in the behavioral expression of autism. Molecular psychiatry 6: 434-9.

Trikalinos, T a, Karvouni, a, Zintzaras, E, Ylisaukko-oja, T, Peltonen, L, Järvelä, I, Ioannidis, JP a. 2006. A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. Molecular psychiatry 11: 29-36.

Tuchman, R, Rapin, I, Shinnar, S. 1991. Autistic cysphasic children. II: Epilepsy. Pediatrics 88: 1219-1225.

Vernes, SC, Newbury, DF, Abrahams, BS, Winchester, L, Nicod, J, Groszer, M, Alarcon, M, Oliver, PL, Davies, KE, Geschwind, DH, Monaco, AP, Fisher, SE. 2008. A functional genetic link between distinct developmental language disorders. N Engl J Med 359: 2337-2345.

Vieland, VJ. 1998. Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage. Am J Hum Genet 63: 947-954.

Vieland, VJ. 2006. Thermometers: something for statistical geneticists to think about. 61: 144-156.

Vieland, VJ, Hallmayer, J, Huang, Y, Pagnamenta, AT, Pinto, D, Khan, H, Monaco, AP, Paterson, AD, Scherer, SW, Sutcliffe, JS, Szatmari, P. 2011a. Novel method for combined linkage and genome-wide association analysis finds evidence of distinct genetic architecture for two subtypes of autism. Journal of neurodevelopmental disorders 3: 113-23.

Vieland, VJ, Huang, J. 2003. Two-locus heterogeneity cannot be distinguished from two-locus epistasis on the basis of affected-sib-pair data. American journal of human genetics 73: 223-32.

Vieland, VJ, Huang, Y, Seok, S-C, Burian, J, Catalyurek, U, O'Connell, J, Segre, A, Valentine-Cooper, W. 2011b. KELVIN: A Software Package for Rigorous Measurement of Statistical Evidence in Human Genetics. Human heredity 72: 276-88.

Volk, HE, Hertz-Picciotto, I, Delwiche, L, Lurmann, F, McConnell, R. 2011. Residential proximity to freeways and autism in the CHARGE study. Environmental health perspectives 119: 873-7.

Vorstman, J a S, Morcus, MEJ, Duijff, SN, Klaassen, PWJ, Heineman-de Boer, J a, Beemer, F a, Swaab, H, Kahn, RS, van Engeland, H. 2006. The 22q11.2 deletion in children: high rate of autistic disorders and early onset of psychotic symptoms. Journal of the American Academy of Child and Adolescent Psychiatry 45: 1104-13.

Vourc'h, P, Martin, I, Marouillat, S, Adrien, JL, Barthélémy, C, Moraine, C, Müh, JP, Andres, C. 2003. Molecular analysis of the oligodendrocyte myelin glycoprotein gene in autistic disorder. Neuroscience letters 338: 115-8.

Wagner, R, Torgesen, J, Rashotte, C. 1999. Comprehensive Test of Phonological Processing. Austin, TX: PRO-ED.

Wang, K, Zhang, H, Ma, D, Bucan, M, Glessner, JT, Brett, S, Salyakina, D, Imielinski, M, Bradfield, JP, Patrick, MA, Kim, CE, Hou, C, Frackelton, E, Chiavacci, R, Sakurai, T, Rappaport, E, Lajonchere, CM, Munson, J, Estes, A, Korvatska, O, Piven, J, Sonnenblick, LI, Alvarez, AI, Herman, EI, Dong, H, Hutman, T, Sigman, M, Klin, A, Owley, T, Sweeney, JA, Brune, CW, Rita, M, Bernier, R, Gilbert, JR, Cuccaro, ML, Mcmahon, WM, Miller, J, State, MW, Wassink, TH, Coon, H, Levy, SE, Schultz, RT, Jr, JIN, Haines, JL, Sutcliffe, JS, Cook, EH, Minshew, NJ, Buxbaum, JD, Dawson, G, Grant, FA, Geschwind, DH, Pericak-vance, MA, Gerard, D. 2010. Common genetic variants on 5p14.1 associate with autism spectrum disorders. 459: 528-533.

Wang, P, Lane, M, Olfson, M, Pincus, H, Wells, K, Kessler, R. 2005. Twelve month use of mental health services in the United States. Archives of general psychiatry 6: 629-640.

Warburton, P, Baird, G, Chen, W, Morris, K, Jacobs, B, Hodgson, S, Docherty, Z. 2000. Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. American Journal of Human Genetics 96: 228-234.

Wechsler, D. 1999. Wechsler Abbreviated Scale of Intelligence (WASI). San Antonio, Texas: Harcourt Assessment.

Weeks, D, Ott, J, Latrhop, G. 1990. SLINK: a general simulation program for linkage analysis. Am J Hum Genet 47.

Weiss, L, Arking, D, Consortium, TGDP of JHA. 2010. A GENOME-WIDE LINKAGE AND ASSOCIATION SCAN REVEALS NOVEL LOCI FOR AUTISM. October 461: 802-808.

Weissman, M, Bland, R, Canino, G, Greenwald, S, Hwu, H, Kyoon, C, Newman, S, Oakley-Browne, M, Rubio-Stipec, M, Wickramaratne, P, Wittchen, H, Eng-Kung, Y, Group, for the CNC. 1994. The cross national epidemiology of obsessive-compulsive disorder. J Clin Psychiatry 55(suppl): S5-S10.

Weitzdoerfer, R, Fountoulakis, M, Lubec, G. 2001. Aberrant expression of dihydropyrimidinase related proteins-2,-3 and -4 in fetal Down syndrome brain. J Neural Transm Suppl.: 95-107.

Weschler, D. 1999. Wecshler Abbreviated Scale of Intelligence. San Antonio, Texas: The Psychological Corporation.

Whitehouse, A, Barry, J, Bishop, D. 2008. Further defining the language impairment of autism: Is there a specific language impairment subtype? J Commun Disord 41: 319-336.

Whitehouse, A, Barry, J, Bishop, D. 2007. The broader language phenotype of autism: A comparison with specific language impairment. J Child Psychol Psychiatry 48: 822-830.

Whitehouse, A, Bishop, D, Ang, Q, Pennel, C, Fisher, S. 2011. CNTNAP2 variants affect early language development in the general population. Genes Brain Behav 10: 451-456.

Whitehouse, A, Coon, H, Miller, J, Salisbury, B, Bishop, D. 2010. Narrowing the broader autism phenotype: A study using the Communication Checklist-Adult Version (CC-A). Autism 14: 559-574.

Whitehouse, A, Line, E, Watt, H, Bishop, D. 2009. Qualitative aspects of developmental language impairment related to language and literacy outcomein adulthood. International Journal of Language & Communication Disorders/Royal College of Speech & Language Therapists 44: 489-510.

Wiederholt, J, Bryant, B. 2001. Gray oral reading test-4, Pro-Ed. Austin, TX.

Wigham, S, McConachie, H, Tandos, J, Le Couteur, A. 2012. The reliability and validity of the Social Responsivenss Scale in a UK general child population. Research in developmental disabilities 33: 944-950.

Wiig, E, Secord, W, Semel, E. 2004. Clinical Evaluation of Language Fundamentals -- Preschool, Second Edition (CELF Preschool-2). Toronto: The Psychological Corporation/A Harcourt Assessment Company.

Wilkinson, G. 1984. Wide range achievement test-revised 3. Wilmington, DE: Jastak Associates.

Williams, D, Botting, N, Boucher, J. 2008. Language in autism and specific language impairment: Where are the links? Psychological Bulletin 134: 944-963.

Woodcock, R. 1987. Woodcock Reading Mastery Tests Revised. Minnesota: American Guidance Service, Inc.

Worley, J, Matson, J. 2012. Comparing symptoms of autism spectrum disorders using the current DSM-IV-TR diagnostic criteria and the proposed DSM-5 diagnostic criteria. Research in Autism Spectrum Disorders 6: 965-972.

Wratten, N, Memoli, H, Huang, Y, Dulencin, A, Matteson, P, Cornacchia, M, Azaro, M, Messenger, J, Hayter, J, Bassett, A, Buyske, S, Millonig, J, Vieland, V, Brzustowicz, L. 2009a. PubMed Central CANADA. Am J Psychiatry 166: 434-441.

Wratten, NS, Memoli, H, Huang, Y, Dulencin, AM, Matteson, PG, Cornacchia, MA, Azaro, MA, Messenger, J, Hayter, JE, Bassett, AS, Buyske, S, Millonig, JH, Vieland, VJ, Brzustowicz, LM. 2009b. Identification of a schizophrenia-associated functional noncoding variant in NOS1AP. 166: 434-441.

Yamashita, N, Uchida, Y, Ohshima, T, Hirai, S-ichi, Nakamura, F, Taniguchi, M, Mikoshiba, K, Honnorat, J, Kolattukudy, P, Thomasset, N, Takei, K, Takahashi, T, Goshima, Y. 2006. Collapsin response mediator protein 1 mediates reelin signaling in cortical neuronal migration. The Journal of neuroscience : the official journal of the Society for Neuroscience 26: 13357-62.

Yan, J, Feng, J, Schroer, R, Li, W, Skinner, C, Schwartz, C, Cook, E, Sommer, S. 2008. Analysis of the neuroligin 4Y gene in patients with autism. Psychiatric Genetics2 18: 204-207.

Yang, X, Huang, J, Logue, MW, Vieland, VJ. 2005. The posterior probability of linkage allowing for linkage disequilibrium and a new estimate of disequilibrium between a trait and a marker. Hum Hered 59: 210-9. Epub 2005 Jul 7.

Ylisaukko-oja, T, Alarcón, M, Cantor, RM, Auranen, M, Vanhala, R, Kempas, E, von Wendt, L, Järvelä, I, Geschwind, DH, Peltonen, L. 2006. Search for autism loci by combined analysis of Autism Genetic Resource Exchange and Finnish families. Annals of neurology 59: 145-55.

Yonan, AL, Alarcón, M, Cheng, R, Magnusson, PKE, Spence, SJ, Palmer, A a, Grunn, A, Juo, S-HH, Terwilliger, JD, Liu, J, Cantor, RM, Geschwind, DH, Gilliam, TC. 2003. A genomewide screen of 345 families for autism-susceptibility loci. American journal of human genetics 73: 886-97.

Yu, TW, Chahrour, MH, Coulter, ME, Jiralerspong, S, Okamura-Ikeda, K, Ataman, B, Schmitz-Abe, K, Harmin, D a, Adli, M, Malik, AN, D'Gama, AM, Lim, ET, Sanders, SJ, Mochida, GH, Partlow, JN, Sunu, CM, Felie, JM, Rodriguez, J, Nasir, RH, Ware, J, Joseph, RM, Hill, RS, Kwan, BY, Al-Saffar, M, Mukaddes, NM, Hashmi, A, Balkhy, S, Gascon, GG, Hisama, FM, Leclair, E, Poduri, A, Oner, O, Al-Saad, S, Al-Awadi, S a, Bastaki, L, Ben-Omran, T, Teebi, AS, Al-Gazali, L, Eapen, V, Stevens, CR, Rappaport, L, Gabriel, SB, Markianos, K, State, MW, Greenberg, ME, Taniguchi, H, Braverman, NE, Morrow, EM, Walsh, C a. 2013. Using Whole-Exome Sequencing to Identify Inherited Causes of Autism. Neuron 77: 259-273.

Zeeland, AAS-van, Abrahams, BS, Alvarez-retuerto, AI, Lisa, I, Rudie, JD, Ghahremani, D, Mumford, JA, Russell, A, Dapretto, M, Geschwind, DH, Susan, Y. 2011. NIH Public Access. Russell The Journal Of The Bertrand Russell Archives 2.

Zerbo, O, Iosif, A-M, Walker, C, Ozonoff, S, Hansen, RL, Hertz-Picciotto, I. 2013. Is maternal influenza or fever during pregnancy associated with autism or developmental delays? Results from the CHARGE (CHildhood Autism Risks from Genetics and Environment) study. Journal of autism and developmental disorders 43: 25-33.

Zhang, C, Milunsky, JM, Newton, S, Ko, J, Zhao, G, Maher, A, Tager-flusberg, H, Bolliger, MF, Carter, AS. 2009. HHMI Author Manuscript A Neuroligin-4 Missense Mutation Associated With Autism Impairs Neuroligin-4 Folding and ER Export Clinical evaluation. Journal of Neuroscience 29: 10843-10854.

Zhang, H, Liu, X, Zhang, C, Mundo, E, Macciardi, F, Grayson, DR, Guidotti, a R, Holden, JJ a. 2002. Reelin gene alleles and susceptibility to autism spectrum disorders. Molecular psychiatry 7: 1012-7.

Zweier, C, de Jong, E, Zweier, M, Orrico, A, Ousager, L, Collins, A. 2009. CNAPT2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am J Hum Genet 85: 655-666.