©2012

Raymond Alton Zimmerman

ALL RIGHTS RESERVED

ASSOCIATION ANALYSIS OF CANDIDATE GENES FOR AUTISM AND A RELATED PHENOTYPE, NON-VERBAL MOTOR SPEECH DISORDER

by

RAYMOND A. ZIMMERMAN

A thesis 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

Master of Sciences

Graduate Program in Microbiology & Molecular Genetics

written under the direction of

Linda M. Brzustowicz, M.D.

and approved by

_____

_____

_____

New Brunswick, New Jersey

May 2012

ABSTRACT OF THE THESIS

Association Analysis of Candidate Genes for Autism and a Related Phenotype, Non-Verbal Motor Speech Disorder

By RAYMOND A. ZIMMERMAN

Thesis Director:
Linda M. Brzustowicz, M.D.

Using a bead-based oligonucleotide ligation SNP genotyping assay, 102 Single Nucleotide Polymorphisms (SNPs) were analyzed in a subset of Autism Genetic Resource Exchange (AGRE) families, looking at autism and a related phenotype, Non-Verbal Motor Speech Disorder (NVMSD). The NVMSD phenotype was defined using three variables from the Autism Diagnostic Interview – Revised (ADI-R), and subjects were chosen regardless of a diagnosis of autism. Genome-wide statistical analysis based on Affymetrix 5.0 SNP data revealed two large linkage regions on chromosomes 1 and 17 in the AGRE data set. Genes of interest were then identified using a software package (Ingenuity Pathway Analysis - IPA) that provides information on gene function and associated diseases. 102 SNPs in four of the genes of interest were genotyped in the AGRE families and statistically analyzed using the KELVIN software package.

ACKNOWLEDGEMENTS

I would like to thank Linda Brzustowicz for allowing me the opportunity to work in her laboratory, and for all of her patient guidance over the many years. James Millonig and Tara Matise for taking the time to serve as my committee members. Christopher Bartlett for teaching me how to think like a scientist, and instilling in me a commitment to both precision and accuracy in my laboratory work. Judy Flax for her inspirational work developing the NVMSD phenotype, and for having the grace to explain it to me several times. Abby Hare for all of her help throughout this entire project, I would never have finished without her. Marco Azaro for not only all of his help and direction during my project, but also for the hours of thought provoking scientific conversation as well as the not-so scientific conversations. A special thank you to Michael Moreau for being a good friend, and whose competitive nature has pushed me to be a better scientist.

I would also like to thank my fellow laboratory technicians Jared Hayter and Jaime Messenger for their assistance in the laboratory. Thanks to all of the Brzustowicz laboratory members who have made going to work a pleasure, Anthony Marcketta, Carmen Ramirez, Bill Manley, and Kathleen Tobin.

Thank you to my wonderful parents who have always been there for me, and encouraged me throughout the years. Thanks to all of my friends and family for their love and support. Finally, a most deserved thank you to Marley Nolan for her patience and understanding through my academic career, and for being supportive of my academic goals.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEGDMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES v

LIST OF ILLUSTRATIONS vi

INTRODUCTION 1

MATERIALS AND METHODS 16

RESULTS 24

DISCUSSION 27

REFERENCES 43

APPENDIX 48

LISTS OF TABLES

Table 1. KELVIN results for NVMSD:ALL as the affection status 24

Table 2. KELVIN results for ASD as affection status 25

Table 3. KELVIN results for NVMSD:ALL as affection status and ASD as a Liability Class 26

Table 4. Pairwise genotype comparison for SNPs in *NCOR1* 39

LIST OF ILLUSTRATIONS

Figure 1. Suggested model of overgrowth, followed by arrest and possible degeneration in ASD compared to normal brain growth 3

Figure 2. Regions of brain overgrowth shown in ASD 4

Figure 3. A flow chart of the NVMSD phenotype defining process 8

Figure 4. Genome-wide PPL linkage peaks for the NVMSD:ALL phenotype 10

Figure 5. PPL linkage peaks for chromosome 1 for the NVMSD:ALL phenotype 12

Figure 6. PPL linkage peaks for chromosome 17 for the NVMSD:ALL phenotype 13

Figure 7. Details of the bead-based oligonucleotide ligation SNP genotyping assay using the Luminex flow cytometer 17

Figure 8. NF1 involvement in different cellular pathways 30

Figure 9. OMG binding to Nogo Receptor and inhibiting neurite outgrowth in the NgR pathway 31

Figure 10. The role of p35 in the regulation of CDK5 and its downstream effects, and disease model implications 33

Figure 11. Interactions between CDK5, p35, and the Reelin signaling pathway through DAB 33

Figure 12. An example of the similar genotypes between four SNPs in *NCOR1* 38

Introduction

Autistic Disorder (AD) is a neurodevelopmental psychiatric disorder characterized by three criteria, (1) impairments in social interaction, (2) impairments in communication, and (3) restricted, repetitive, and stereotyped patterns of behavior. Impairments in social interaction include a variety of manifestations: (i) deficits in the use of multiple nonverbal behaviors, which would include a lack of eye contact, facial expressions, and the use of body posture in social interaction; (ii) failure to develop peer relationships; (iii) a lack of spontaneously seeking to share enjoyment or interests; and (iv) a lack of mutual social or emotional exchange with others. Communication impairments can range from a complete lack of spoken language to the use of adequate speech but with a notable deficit in the ability to start or continue a conversation. Communication impairments could also manifest as repetitive or idiosyncratic language, as well as lack of impulsive make-believe or social imitation play appropriate for the subject's developmental level. The stereotyped behaviors could present as abnormally intense or focused preoccupation with stereotyped and restricted patterns of interest, an inflexible devotion to routines or rituals, repetitive motor mannerisms, or a persistent preoccupation with parts of objects. Along with the main criteria listed above, symptoms must manifest prior to 3 years of age, and not be accounted for by Rett's Disorder or Childhood Disintegrative Disorder according to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision) (DSM-IV-TR).

Autism Spectrum Disorder (ASD) includes the definition of AD as well as other related disorders, such as Asperger Syndrome and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). ASD has a prevalence in the general population

of 1 in 88 according to the Center for Disease Control and Prevention (CDC), while the rates of AD are lower, with estimates between 10 and 22 in 10,000 (CDC, Fombonne 2003, Chakrabarti 2005). Males are diagnosed with ASD at a much higher rate than females; the average ratio of males-to-females is 4.3:1 (Newschaffer 2006).

Twin and sibling studies offer strong evidence for a genetic component to autism. In one study, monozygotic twins (MZ) were shown to have a 60% concordance rate for AD, while dizygotic (DZ) twins had 0% concordance. Using the ASD phenotype MZ twins had a concordance rate of 92%, while DZ twins had a concordance rate of 10% (Bailey 1995). More recent twin studies have shown a higher rate of 31% concordance in DZ twins using the ASD phenotype (Ronald 2011). Family studies reveal that siblings of AD children have a 2.9% chance of also having AD, while 12-20% develop the ASD phenotype (Bolton 1994). Compared to the prevalence of 1 in 88 (approximately 1.14%) in the general population, an increase to 12% - 20% prevalence in siblings of autistic children along with twin studies suggest a high level of genetic influence. There are also multiple genetic linkage and association studies that support there being the presence of a strong genetic component to autism.

Neuroanatomical characteristics

There are many characteristic differences in the brains of autistic patients. Understanding these physical abnormalities and the possible cellular mechanisms from which they arise is important when deciding which genes in a linkage region to study and what possible role they could be playing in the development of the brain. The most common cerebellar abnormality in the autistic brain is a reduced number of Purkinje

neurons (Allen 2004). Another hallmark of autism is an increase in head circumference associated with increased brain volume. In fact, several regions of the brain have been shown, through MRI studies, to have regional overgrowth. This initial period of early overgrowth is followed by an arrest of growth and possible degeneration (see Figure 1).

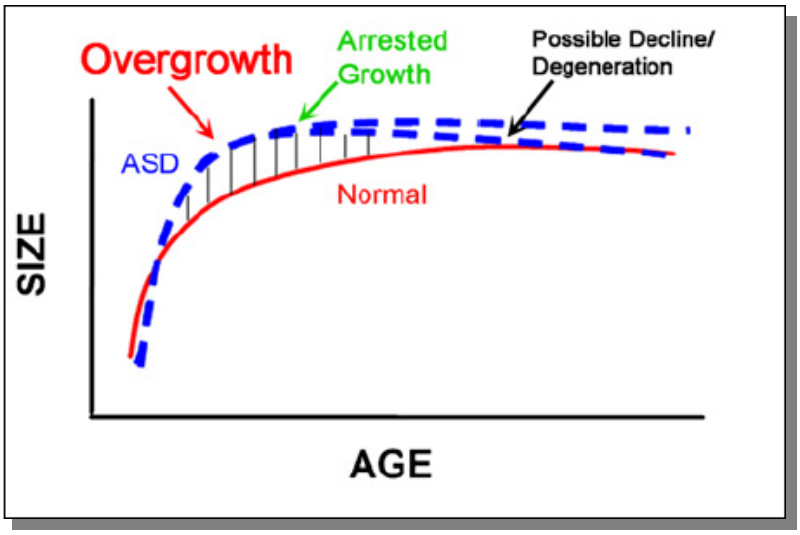

Figure 1. Suggested model of overgrowth, followed by arrest and possible degeneration in ASD compared to normal brain growth (Courchesne 2007).

This overgrowth could be caused by dysregulation of the cell cycle, failure of natural cell death, or a combination of both (Courchesne 2011). The timing of the overgrowth corresponds to developmentally important early years during which a child is acquiring language, social, and emotional skills, which are characteristically impaired in autistic children. The regions found to have the most significant early overgrowth in patients with autism are the frontal (and prefrontal) cortex, the temporal cortex, the cerebellum and the amygdala (Figure 2) (Courchesne 2007).

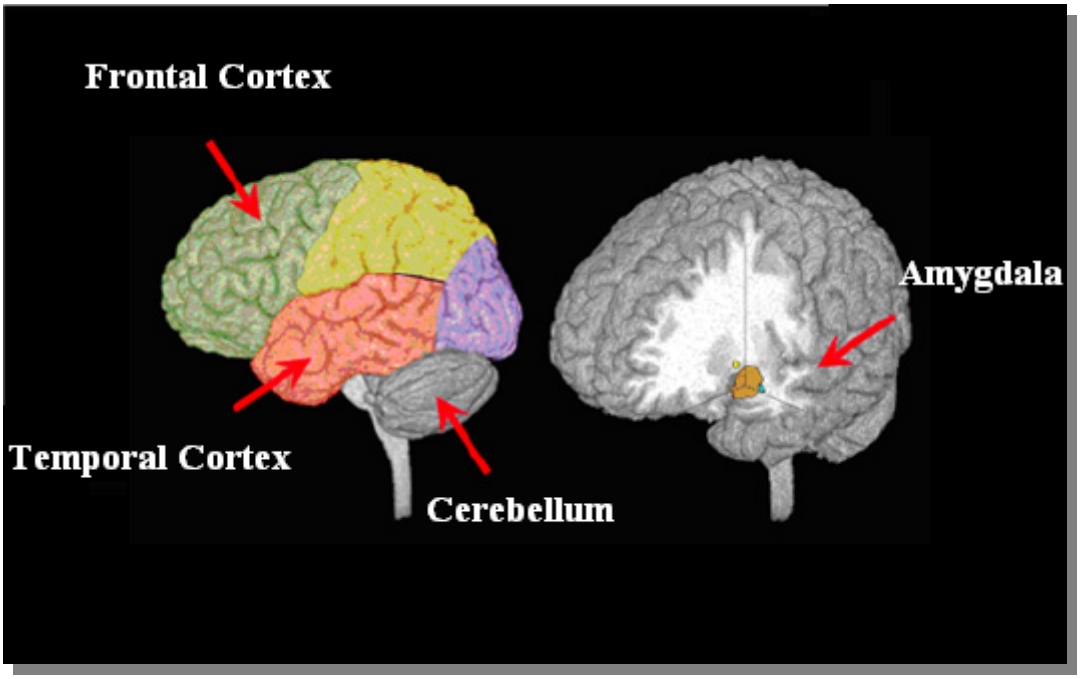

Figure 2. Regions of brain overgrowth shown in ASD (Courchesne 2007).

The frontal cortex (FC) has been implicated as one of the major regions of the brain that has shown a significant amount of overgrowth in young autistic patients (Courchesne 2007). The FC is involved in many higher order functions, such as problem solving, planning, judgment, parts of speech, and emotions, some of which are impaired in autistic patients. Located within the FC is the Broca's area which is critical for written and spoken language (Nolte 2002). A recent case-control study of the prefrontal cortex (PFC) of male children with autism showed a large increase in the total number of neurons. The children with autism had 67% more neurons in the PFC than control children; specifically the mean number of neurons for the autistic children was 1.94 billion, while the control children had a mean of 1.16 billion neurons. Two subdivisions of the PFC had the greatest increase, the dorsolateral PFC, with 79% more neurons, and the mesial PFC, with 29% more neurons (Courchesne 2011).

The temporal cortex (TC), which also shows significant overgrowth, is involved in a number of processes related to autism. The major functions of the TC are auditory perception, higher order processing of visual information, and complex aspects of learning and memory. The TC also contains the Wernicke's area, a region vital for the comprehension of language (Nolte 2002).

The amygdala is an almond shaped collection of nuclei located beneath the temporal lobe that goes through the characteristic early volume increase and subsequent volume decrease also seen in the frontal and temporal cortical regions of autistic children. Although it is formed early in gestation, the individual nuclei undergo differentiation postnatally (Baron-Cohen 2000). The amygdala has been shown through functional imaging studies to be important in the ability to recognize emotions in facial expressions (Adolphs 1999). It also plays a well-established role in recognizing and producing emotions, as well as modulating the same social behaviors that are impaired during the early stages of autism in toddlers, poor eye contact and a reduction in social interests (Schumann 2009).

Historically, the cerebellum has been a region of the brain thought to exclusively be involved in motor control and coordination. Not until a reappraisal of its function during the late 1980's into the 1990's did it become accepted that the cerebellum played a role in cognitive functions as well. It was eventually shown that the cerebellum is involved in learning and memory, visuospatial function, mood and affect, and language (O'Halloran 2011). In autistic individuals, along with the phenomenon of overgrowth, the cerebellum most frequently shows a reduction in Purkinje neurons or some other form of pathology in 95% of cases examined by autopsy. In conjunction with this pathology, it

has been demonstrated using functional MRI that in autistic children there is an abnormal patterning of cerebellar activation during simple motor tasks. The amount of activation in the cerebellum was greater in autistic children, and there was abnormal activity in cerebellar regions not normally associated with simple motor tasks (Allen 2004).

Non-Verbal Motor Speech Disorder

Non-Verbal Motor Speech Disorder (NVMSD) is a disorder related to ASD that is characterized by non-verbal or unintelligible behavior (NVMSD:ALL). A sub-phenotype of NVMSD:ALL that includes individuals without significant comprehension difficulties (NVMSD:C) was developed. These phenotypes were identified using variables from the Autism Diagnostic Interview-Revised (ADI-R) (Rutter 2003) in the Autism Genetics Resource Exchange (AGRE) data set. The ADI-R is a clinical diagnostic instrument that consists of responses from parents or caregivers of subjects suspected of being on the autism spectrum. The three variables used in creating the NVMSD phenotypes were Overall Level of Language, Articulation/Pronunciation, and Comprehension of Simple Language (Flax 2010).

Defining the Phenotype- Three ADI-R Variables

Overall Level of Language (**LEVELL**) is defined as "…Concerning whether subject uses at least three word phrases; this includes spontaneous speech or echoed or stereotyped speech if they are used functionally". **LEVELL** is scored as a 0, 1, or 2, with 2 being the score used in defining the NVMSD phenotype. In this case, a score of 2 means "fewer than five words total or speech not used on a daily basis", with 0 and 1

representing more functional use of language, for example, more frequent use of three-word phrases and use of speech on a daily basis (ADI-R).

Articulation/Pronunciation "refers to the enunciation of the sounds of language", and is scored on a scale of 0-3 with 0 denoting the highest level of articulation. A score of 2 is defined as "definite articulation difficulties such that some words are very difficult for strangers to understand until they get to know her/him", and a score of 3 is defined as "strangers find speech almost impossible to understand or parents have significant difficulties understanding because of articulation". Articulation/Pronunciation is broken into two types, **CARTIC** which represents current level of functioning and is scored if the subject scored is 4 years of age or older, and **ARTIC5**, meaning that if the subject is older than 5, what their retrospective score would have been at 5 years of age (ADI-R). Subjects with a **LEVELL** score of 2 or **CARTIC** (or **ARTIC5**) score of 2 or 3 are classified as NVMSD:ALL Affected. Subjects with a **LEVELL** score of 0 or 1, and a **CARTIC** (or **ARTIC5**) score of 0 or 1 are defined as NVMSD:ALL Unaffected, otherwise subjects were listed as NVMSD:ALL Unknown when other issues arose such as deafness or the relevant questions were simply not asked. (Figure 3)

The NVMSD:ALL Affected phenotype can further be sub-phenotyped based on the Comprehension of Simple Language (**CCOMPSL**) variable. **CCOMPSL** "deals with the subject's understanding of spoken language when there is an absence of other cues such as those that might be provided by gesture, pointing, or context", and is scored on a scale of 0-4. The major distinction for the NVMSD:C phenotype lies between a score of 2 and 3, where 2 means the subject "understands many words (more than 50)…", and a score of 3 indicates the subject "understands fewer than 50 words…" (ADI-R).

The distinction of understanding more or less than 50 words is the major component of the NVMSD:C phenotype, with a score of 0, 1, or 2 being defined as NVMSD:C Affected, and a score of 3 or 4 being defined as NVMSD:C Unknown since there is no way to determine whether the poor speech is caused by poor comprehension or an actual motor speech problem. The remainder of the subjects are classified as NVMSD:C Unaffected (Figure 3).

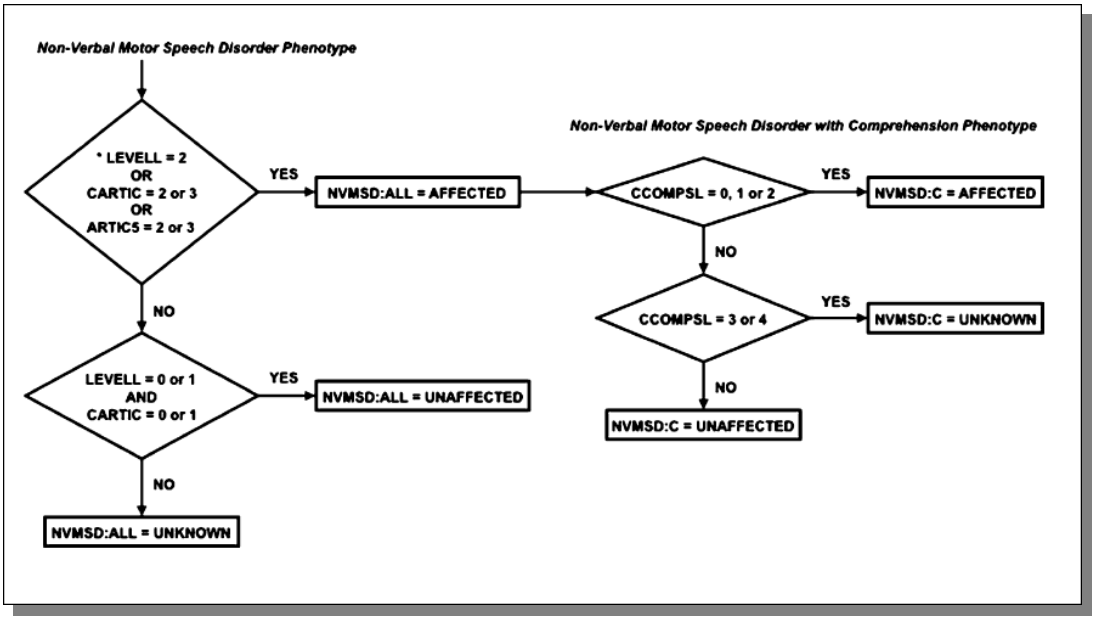

Figure 3. A flow chart of the NVMSD phenotype defining process (Flax 2010).

AGRE family selection

Families were chosen from the AGRE database based on the following criteria (Flax 2010). First, the families had to have ADI-R data available, which was used for defining the phenotypes. Second, the families needed to have Affymetrix 5.0 genotyping data available, which was used for the initial linkage and association analysis to locate potential susceptibility genes. The third criterion required that each family have "at least

two individuals who were either non-verbal, minimally verbal, or who had speech that was unintelligible to others" (Flax 2010). Though these families contain autistic probands, autism was not a factor in choosing the subjects for this study. However, these families were chosen from a set of families that include autistic individuals, thus a large portion of the NVMSD:ALL affected subjects used did meet criteria for autism. In total, 203 families (1,146 individuals) met these requirements.

Statistical Analysis

Genome wide linkage and association analysis was conducted on the chosen AGRE families using Affymetrix 5.0 SNP data downloaded from the AGRE site. Of the 443,106 SNPs interrogated by the Affymetrix data, 11,100 were used for the linkage analysis and were selected at 0.3 centimorgan (cM) intervals. For the association analysis, 263,334 of the SNPs were used. SNPs used in the linkage and association analysis were required to meet a minimum minor allele frequency (MAF) of 5% for the linkage analysis and 1% for the association analysis. SNPs chosen did not deviate from Hardy-Weinberg equilibrium, with a $p > 0.05$ for the linkage analysis and a $p \geq 10^{-10}$ for the LD analysis (Flax 2010).

The software package KELVIN (Vieland 2011) was used for both the linkage and association analyses. KELVIN uses the Posterior Probability of Linkage (PPL) class of models to measure the strength of the genetic evidence of linkage. KELVIN also reports the Posterior Probability of Linkage Disequilibrium (PPLD), as a measure of association, and the Posterior Probability of LD given linkage (PPLD|L), as a measure of association when analyzing a region that has known linkage. The probabilities are measured on a

zero to one scale, with a value of greater than 0.02 being evidence for linkage (or association) and a value less than 0.02 representing evidence against linkage (or association) (Vieland 1998; Wang 2000; Huang 2011). The results of the genome-wide linkage analysis (PPL) showed large peaks on chromosomes 1, 3, 4, 5, and several peaks on chromosome 17 (Figure 4). Note that for PPL the threshold value of 0.02 is used because there is a prior probability of linkage of 0.02 between any single locus and a marker (Elston and Lange 1975). Association analysis (PPLD) failed to produce any statistical evidence for LD, most likely due to insufficient sample size (Flax 2010).

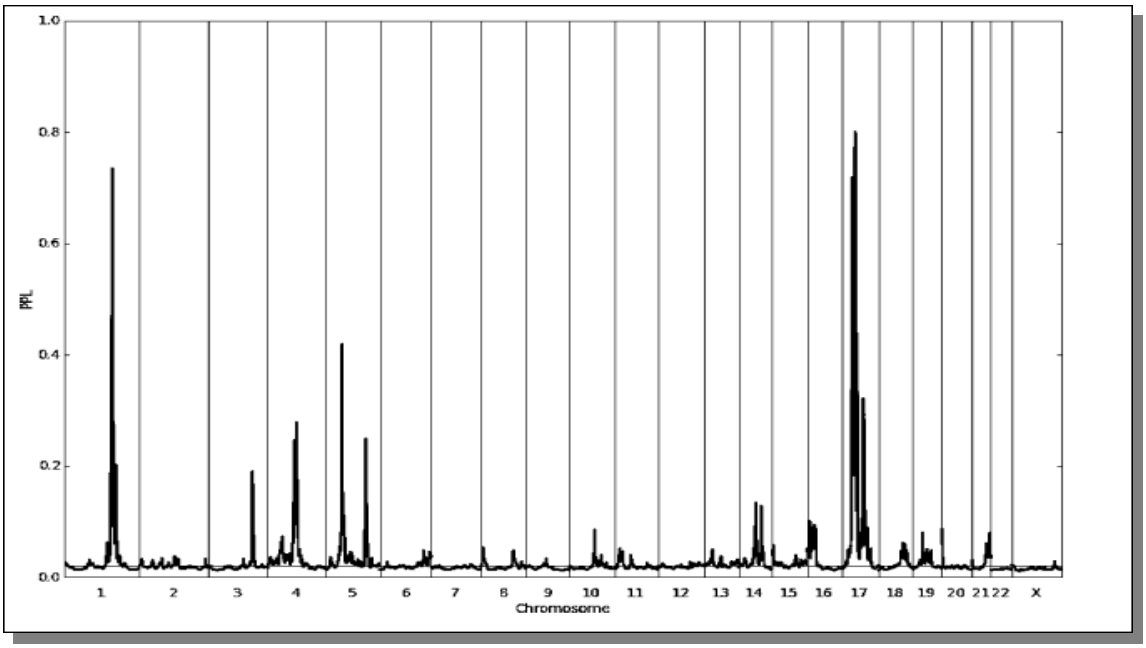

Figure 4. Genome-wide PPL linkage peaks for the NVMSD:ALL phenotype (Flax 2010).

Susceptibility Gene Identification

After the statistical analysis was completed and linkage regions were identified, Ingenuity Pathway Analysis (IPA) software was used to identify potential susceptibility genes within those regions (Ingenuity® Systems, www.ingenuity.com). The UCSC

Genome Browser (NCBI Build 36.0) was used to identify the genes under the linkage peaks, 645 in total. Of the 645 genes, 38 were excluded for being theoretical genes or because they were simply not in the IPA database. Analysis was done on the remaining 607 genes with the IPA Core Analysis which identifies biological functions and/or diseases associated with the genes. P-values were calculated to reflect the probability that each biological function and/or disease associated with the linkage analysis would occur by chance alone. Candidate genes were chosen from a list based on whether they had a significant p-value and a function related to Nervous System Development and Function, Neurological Disorders, Genetics Disorders, or Psychological Disorders. Fifty two genes met these criteria, and from those five candidate genes were selected based on the PPL results, and the relevancy of their known functions in relation to ASD and/or NVMSD.

Genes of Interest

LIM homeobox transcription factor 1 alpha (*LMX1A*), whose gene is located at 1q24.1 (Figure 5), is a transcription factor that plays a role in several processes during neuronal development. Transcription factors by nature either activate or repress the expression of other genes. *LMX1A* is involved in the differentiation and maturation of mesencephalic dopamine neurons, as well as their survival, and has previously been associated with schizophrenia (Bergman 2010). In human embryonic stem cells, expression of *LMX1A* has been shown to be a requirement for midbrain dopamine cell fate, and when *LMX1A* expression is knocked down using short interfering RNA (siRNA)

there is a corresponding reduction in the expression of many other genes related to the dopamine cell differentiation (Cai 2009).

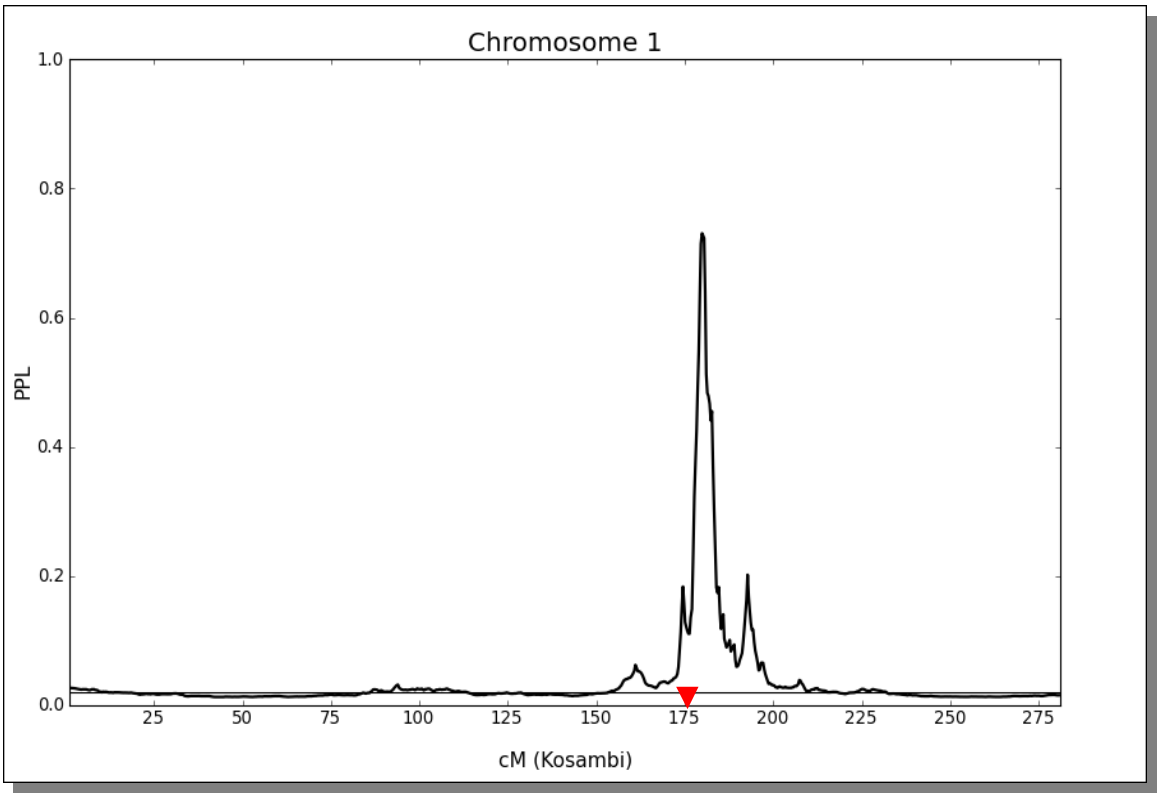

Figure 5. PPL linkage peaks for chromosome 1 for the NVMSD:ALL phenotype (Flax 2010). Red triangle indicates approximate gene position: *LMX1A* – 175.9 cM

Nuclear Receptor Co-Repressor 1 (*NCOR1*), whose gene is located at 17p11.2 (Figure 6), acts as a co-repressor for many transcription factors by recruiting histone deacetylases, and has multiple nuclear receptor interacting domains. *NCOR1* assists nuclear receptors in the down-regulation of DNA expression. In mice, the *NCOR1* homolog *Ncor* acts as a transcriptional co-regulatory protein that recruits histone deacetylase 3 (Hdac3), a chromatin modifying enzyme (Alenghat 2008). *Ncor* has been shown to repress the transcription factor CBF1 which acts downstream of notch, a factor that is implicated in the inhibition of neuronal differentiation (Hermanson 2002). *Ncor*

protein levels have been shown to be high in undifferentiated, proliferating neural progenitor cells in the embryonic brain. In one study, siRNA specifically targeting *Ncor* was introduced into the amygdala of developing rats, reducing the expression of *Ncor*. This reduction in expression showed an influence on the organization of juvenile social play behavior and an increase in anxiety-like behavior (Jessen 2011).

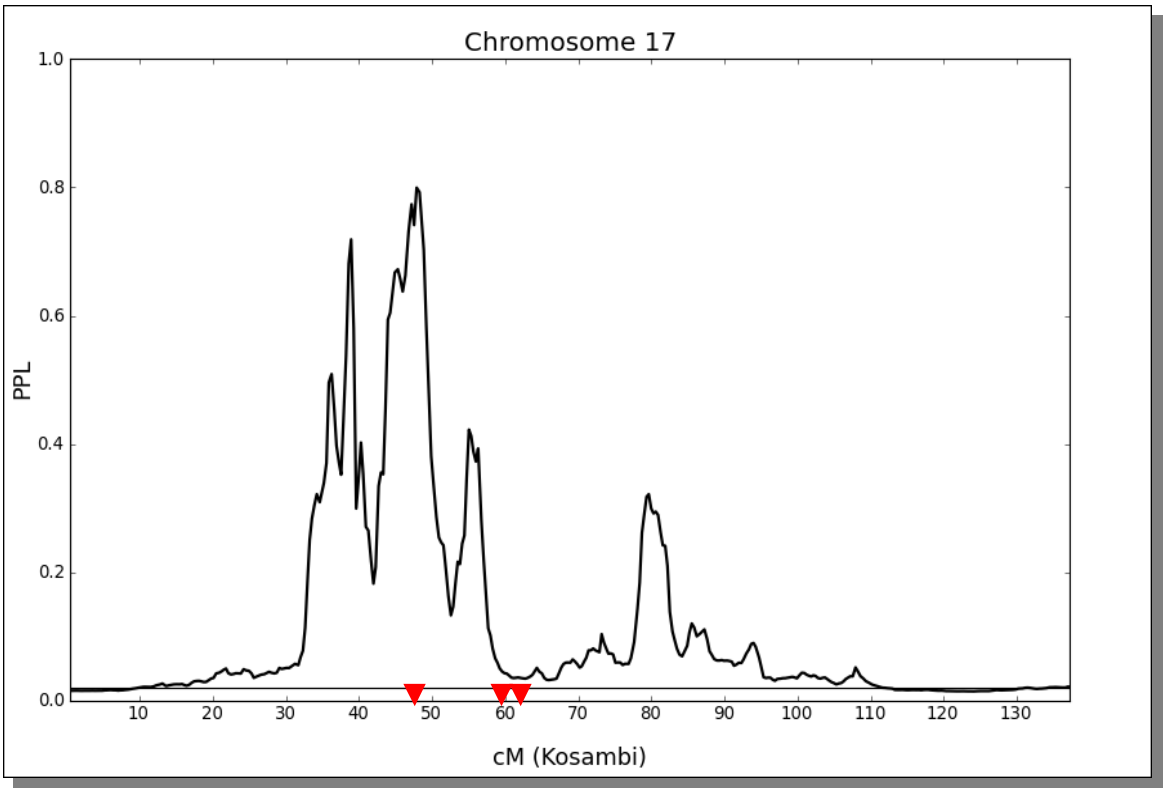

Figure 6. PPL linkage peaks for chromosome 17 for the NVMSD:ALL phenotype (Flax 2010). Red triangles indicate approximate gene positions: *NCOR1* – 48.1cM, *NF1* – 60.5cM, *OMG* – 60.8cM, *CDK5R1* – 63.1cM

Neurofibromin 1 (*NF1*) is a tumor suppressor gene located at 17q11.2 (Figure 6) that is over 350 Kb and consists of 61 exons. *NF1* is widely expressed in many cell types, including neurons, astrocytes, oligodendrocytes and Schwann cells (Trovó-Marqui 2006). Mutations in *NF1* are the cause of type I neurofibromatosis, a disorder that has an

increased prevalence in the autistic population over the general population ranging from a 48- to 190-fold increase depending on the study (Marui 2004; Martin 2007). The protein encoded by the *NF1* gene is neurofibromin (NF1), which has several functions within cell signaling pathways, regulates the expression of the RAS protein, and modulates adenylyl cyclase and cytoskeletal associations. By regulating RAS "neurofibromin plays a part in cellular transduction signal regulation involved in synaptic plasticity, memorizing and also in learning" (Hachon 2011). *NF1* has also previously been associated with learning deficits (Martin 2007). The possible role that NF1 may play in learning disabilities could be by indirectly causing an increase in gamma-aminobutyric acid (GABA) through increased RAS activity. The inhibitory neurotransmitter GABA acts in a region of the brain that is implicated in memory formation, the hippocampus (Trovó-Marqui 2006).

Oligodendrocyte-Myelin Glycoprotein (*OMG*) is a gene located within intron 27b of the *NF1* gene. The entire *OMG* gene has only two exons and one intron; it is the second exon that encodes for the entire protein (Martin 2007). OMG is expressed by neurons and oligodendrocytes and is involved in the myelination of the central nervous system (CNS). OMG also acts as a growth inhibitor that regulates neuronal morphology by binding to the Nogo-66 receptor which results in the inhibition of neurite outgrowth (Venturin 2006). *OMG* has also been previously implicated in Autistic Disorder in a study looking at the transmission disequilibrium of a G/A SNP that changes a glycine residue to an aspartic acid. This change of a small neutral amino acid to a large acidic amino acid located just four amino acids from a cleavage site, suggests that it could be functionally important (Martin 2007).

Cyclin-Dependent Kinase 5 Regulatory Subunit 1 (*CDK5R1*) also located at 17q11.2 (Figure 6), encodes the protein p35. p35 is a neuron-specific regulatory subunit of Cyclin-dependent kinase 5 (CDK5) which is required for proper development of the CNS. During central nervous system development CDK5 has an important role in the migration and differentiation of neurons (Moncini 2007). CDK5 and its regulatory subunit p35 form a complex that has several functions in the development and regulation of the CNS, including: (i) learning and memory processes, (ii) synaptic transmission, (iii) neuronal apoptosis, (iv) axonal regeneration, (v) cellular differentiation, and (vi) membrane trafficking during the outgrowth of neuronal processes (Venturin 2006). It has also been shown that CDK5 can modulate the Reelin signaling pathway which has been shown to be associated with autism in numerous studies (Fatemi 2005). CDK5 modulates the Reelin pathway in the developing mouse brain by phosphorylating DAB1, an adaptor protein that acts downstream of Reelin and regulates the Reelin signaling pathway (Ohshima 2007).

Materials and Methods

SNP Tagging

Once the genes of interest were identified, SNPs within these genes were chosen, or tagged. This was done with the use of three resources, HapMap, Applied Biosystems, and 1000 Genomes. First, HapMap was used to identify and tag SNPs; HapMap can be searched by chromosome and base position, 5 kb of sequence flanking both upstream and downstream of the genes were added to include possible regulatory regions. The parameters of the HapMap SNP tagging were: the restriction to the CEU (Caucasian of European descent) population, a minor allele frequency (MAF) of at least 5%, and an r^2 of 0.8. The r^2 is a measure of LD between SNPs, so an r^2 of 0.8 between two SNPs means that if the genotype of one SNP is known then the genotype of the other SNPs will be known 80% of the time. HapMap uses Tagger (de Bakker 2005) which ensures that no two SNPs in the final tagged panel have an r^2 score ≥ 0.8, and that there are no untagged SNPs that have low LD in all of the tagged SNPs. HapMap will then produce a list of SNPs in the region specified. The program SNPbrowser was then used to search for more SNPs in the Applied Biosystems database under the same parameters as HapMap, and another list of SNPs was generated. The lists from the two sources were then combined to eliminate any redundancies. Finally, to fill in any remaining gaps in the coverage of the genes, supplemental SNPs were chosen from the 1000 Genomes database. These SNPs were chosen to fill any large physical gaps that remained between the SNPs already chosen by HapMap and SNPBrowser.

SNP Genotyping

A total of 102 SNPs covering the five genes, *LMX1A*, *NCOR1*, *NF1*, *OMG*, and *CDK5R1*, were genotyped using a bead-based oligonucleotide ligation SNP genotyping assay run on a Luminex flow cytometer (Bruse 2008). The SNPs were divided into several multiplex panels and each panel of SNPs was assayed on eleven 96-well plates which contain 986 DNA samples of the 1,146 AGRE individuals used in the previous linkage and association analysis (Flax 2010). Some DNA samples however, were unavailable at the time they were ordered. This assay consists of several steps; (i) a multiplex polymerase chain reaction (PCR) is done to amplify the regions of DNA around the SNPs, followed by (ii) the oligonucleotide ligation assay (OLA), then (iii) a bead hybridization step, and the final step, (iv) the SAPE reaction, before assaying the panel on the Luminex flow cytometer (Figure 7).

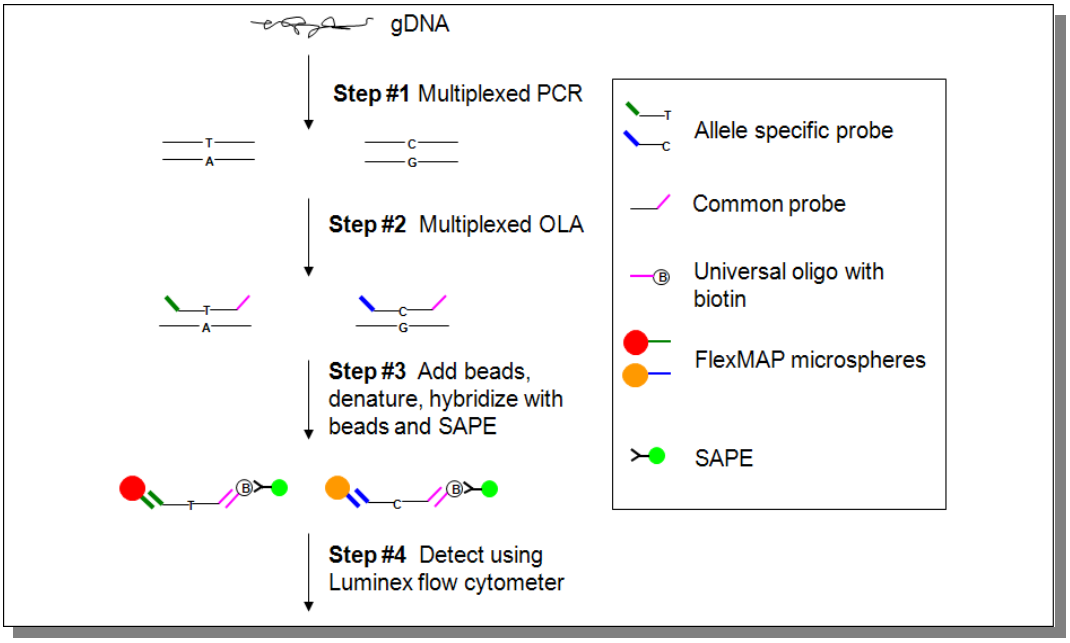

Figure 7. Details of the bead-based oligonucleotide ligation SNP genotyping assay using the Luminex flow cytometer (Bruse 2008).

Multiplex Polymerase Chain Reaction

20 μL PCR reactions were set up in 384-well microtiter plates, using a Perkin Elmer – Janus® Automated Workstation. Starting with 40 ng of genomic DNA in each well, a multiplex PCR was performed using a solution containing: AmpliTaq Gold DNA polymerase, a proprietary 1x reaction buffer, 2.5 mM MgCl (Applied Biosystems), dNTPs, and 3.33 pmol of each PCR primer. The PCR products were all designed to amplify a region of DNA between 100 and 400 base pairs in length that contained a given target SNP. The plate containing the genomic DNA and the PCR master mix solution was then put on a thermocycler and the following program was run: an initial denaturing step of 95°C for 10 minutes, then 40 cycles of 95°C for 40 seconds, 60°C for 30 seconds, ramping from 60°C to 65°C using fifty 0.1°C 2 second cycles, and 72°C for 2 minutes, lastly a final extension step of 72°C for 10 minutes. The temperature ramping during the annealing step from 60°C to 65°C was to accommodate a range of primer annealing temperatures in the multiplexed mixture. Regardless of the number of SNPs in each multiplex panel the same PCR protocol was followed.

Oligonucleotide Ligation Assay - OLA

The OLA is an assay that ligates two oligonucleotide probes which have complementarily hybridized to the PCR product containing the SNP. The OLA requires the design of two allele specific probes and a common probe for each SNP (Figure 7). The allele specific probes consist of a 5´-tag sequence that is 24-base pairs and has complementarity to a sequence unique to a specific bead that is introduced in the next step, and a 3´ terminal base specific to one of the two possible alleles of the target SNP.

The common probe consists of a 5´ locus-specific region that is designed to start one base downstream of the SNP, and is 5´-phosphorylated to allow for ligation. The 3´ region of the common probe is complementary to the doubly-biotinylated universal oligonucleotide, also introduced in the next step.

An OLA product is only made if the allele specific probe can properly anneal to the PCR product containing the SNP and subsequently be ligated to the common probe. If, for example a G/T SNP were being assayed, one allele specific probe would end with a C and the other allele specific probes would end with an A. If the genomic DNA sample being assayed had the homozygous genotype G/G, then only one of the allele specific probes would properly hybridize to the PCR product and be ligated, specifically the one ending in the C nucleotide. If the sample were heterozygous (G/T), then both allele specific probes would hybridize to their respective PCR products, and become ligated to the common probe.

The 20 µL OLA reaction is set up in a separate 384-well plate also using the Perkin Elmer – Janus® Automated Workstation. 2 µL of the multiplex PCR product is added to a solution containing a 1 pmol mixture of the appropriate OLA probes, a proprietary 1x reaction buffer and Taq DNA ligase (New England Biolabs). OLA reactions are also run on a thermocycler using the following program: 95°C for 1 minute, 32 cycles of 95°C for 15 seconds (to denature the newly formed ligation product from the PCR product) and 58°C for 2 minutes (to allow for annealing and ligation). The protocol is the same independent of the number of SNPs being genotyped per multiplex panel. The OLA product is then used for the next step.

Bead Hybridization

A set of 100 FlexMAP beads (Luminex Corporation) are used when designing assays, each bead is covered in a 24-base nucleotide sequence referred to as a 'tag'. Because of volume limitations of the 384-well plates (30 µL total volume per well), the subsequent reactions are done in a 96-well microtiter plate (200 µL total volume per well) to accommodate the 65 µL total volume of the bead hybridization reaction. The next step combines 15 µL of the OLA reaction product with a 50 µL solution containing a hybridization buffer, the beads corresponding to the allele specific probes used in the specific multiplex panel, and the doubly-biotinylated universal oligonucleotide (which is the anti-tag to bead 100 which is not used when designing assays). The amount of universal oligonucleotide must be empirically determined prior to each assay by performing an optimization reaction. The bead hybridization reaction once again takes place on a thermocycler, using a two step program. During the first step the mixture is heated to 95°C for 2 minutes to denature the OLA products from their corresponding PCR products, while the second step is a 20 minute incubation at 37°C that allows for hybridization of the OLA product to the beads as well as the universal oligonucleotide. The purpose of this reaction is to create a physical bridge from the FlexMAP bead, hybridized to the 5′ region of the OLA product, to the universal oligonucleotide which is hybridized to the 3′ region of the OLA product.

SAPE Reaction

A 6 μL mixture of hybridization buffer (also used in the bead hybridization step) and Streptavidin-R-phycoerythrin (SAPE) (Invitrogen Corporation) is added directly to the bead hybridization product for a total final volume of 71 μL. SAPE binds non-covalently to the biotin on the universal oligonucleotide through the streptavidin portion of the molecule, and the phycoerythrin is the portion of the molecule that fluoresces later in the assay. The mixture of bead hybridization product and SAPE is then placed on a thermocycler for one hour at 37°C.

Flow Cytometry Analysis

After the SAPE reaction has completed, the reaction plate is loaded into the Luminex flow cytometer. The Luminex sorts and categorizes the beads and their associated amount of SAPE, which fluoresces under laser excitation. If the OLA product is successfully made, then the corresponding bead is associated with a commensurate degree of SAPE fluorescence. When no OLA product is made during the OLA reaction, then no SAPE fluorescence is associated with the corresponding bead. The genotype data for each well (DNA sample) is then saved and imported into Excel for initial analysis and file formatting.

Genotype clustering

The raw data from the Luminex is imported into Microsoft Excel (2007) and a simple macro is run that calculates the percent of the first allele's fluorescent signal with respect to the total signal for both possible alleles in each well. Note that the first allele is

simply being assigned alphabetically. This "percent of allele one" is then plotted on a graph for the entire set of samples with percent of allele one on the y-axis and the individuals along the x-axis. Once graphed (see Figure 12), three distinct clusters representing each of the three possible genotypes should be visible and any obvious errors in the data are removed.

Error Checking

The raw genotypes, acquired in the previous step, are run though two error checking programs, PedCheck and Merlin. PedCheck is a program that looks for mendelian inconsistencies in the genotyping data. It employs the *Nuclear-Family Algorithm* that "uses the individual's genotypes as given in the pedigree to check for inconsistencies between parents and offspring" (O'Connell 1998). PedCheck requires two input files, one containing the genotypes and pedigree information (.pre file), and another containing a list of the markers, in this case SNPs, and their allele frequencies (.dat file). Any PedCheck errors, reported by marker and family ID, were manually reviewed, and in most cases, the genotypes of that family for the specific marker were zeroed out. Merlin is a program that checks genotyping data for errors based on unlikely recombinations. Merlin requires three input files, a .pre and .dat file similar to the ones used in PedCheck, and a third file that contains the genetics distances in centimorgans between the markers (.map file). Again, any errors were manually reviewed and erroneous genotypes were zeroed out.

In total 429 genotypes were removed due to errors discovered by PedCheck and Merlin, out of approximately 100,000 genotypes. The average amount of missing

genotypes per SNP was 1.05% before error checking, and rose to 1.5% after error checking. The data was also cleaned for missingness by marker; any SNP missing more than 5% of its genotype data was abrogated from this analysis. SNPs were then checked to make sure they were in Hardy-Weinberg equilibrium (HWE) with a p-value threshold of 0.01; only one SNP was out of HWE and was removed from the analysis.

KELVIN Statistical Analysis

Once all errors had been removed the data was analyzed using KELVIN, the same software package used to conduct the linkage analysis of the initial Affymetrix 5.0 SNP data (Flax 2010). Before the data is ready for KELVIN though, two more programs, Mega2 and Mendel, are necessary to prepare the files. Mega2 is used to rearrange the .ped, .dat, and .map files into the proper format for Mendel and in turn, Mendel is the used to calculate the actual allele frequencies of the SNPs in our data set and produces a .mrk file. Two additional configuration files are necessary for the analysis. The qrun file, which is not inherently necessary for running the KELVIN software, is required for running the software on our computing grid. Lastly, the run file contains input- and output-file definitions, as well as runtime parameters.

Results

When analyzing SNPs that have been identified under linkage peaks the Posterior Probability of Linkage Disequilibrium given Linkage (PPLD|L) is the most statistically relevant measure of association. This analysis using NVMSD:ALL as the affection status showed PPLD|L values above 0.02, which is considered evidence for LD, at several SNPs across the genes assayed (See Table 1). The highest PPLD|L values were seen in SNPs located inside *LMX1A* (rs1506944 at 0.07, rs1506946 at 0.06, and rs4657430 at 0.05) and *CDK5R1* (rs9889771 at 0.05).

Gene	Chr	Marker	Position (cM)	PPLD|L
LMX1A	1	rs1506946	175.9271	0.0600
LMX1A	1	rs1506944	175.9275	0.0700
LMX1A	1	rs4657430	175.9701	0.0500
LMX1A	1	rs10753669	176.0220	0.0300
LMX1A	1	rs6662349	176.0244	0.0400
NCOR1	17	rs3744328	48.0996	0.0205
NCOR1	17	rs11650682	48.1000	0.0300
NCOR1	17	rs7224403	48.1012	0.0205
CDK5R1	17	rs756787	63.1244	0.0232
CDK5R1	17	rs9889771	63.1262	0.0500

Table 1 – PPLD|L values above 0.02 for all SNPs analyzed using NVMSD:ALL as the affection status.

The analysis was then run using ASD as the affection status instead of NVMSD:ALL. However, in our sample set there were twenty three individuals that were classified as Not Quite Autism (NQA). For this analysis, these NQA individuals were coded as affected. The PPLD|L values over 0.02 for this analysis are given in Table 2.

Gene	Chr	Marker	Position (cM)	PPLD|L
LMX1A	1	rs1506946	175.9271	0.0400
LMX1A	1	rs1506944	175.9275	0.0400
LMX1A	1	rs4657428	175.9586	0.0215
LMX1A	1	rs4657430	175.9701	0.0241
LMX1A	1	rs10753669	176.0220	0.0300
LMX1A	1	rs6662349	176.0244	0.0500
LMX1A	1	rs7549447	176.0713	0.0208
NCOR1	17	rs11650682	48.1000	0.0400
NF1	17	rs2953014	60.4942	0.0231
CDK5R1	17	rs9889771	63.1262	0.0400

Table 2 –PPLD|L values above 0.02 for all SNPs analyzed using autism as the affection status and NQA individuals classified as 'affected'.

The diagnosis of autism was then used as a liability class rather than as the affection status, which was NVMSD:ALL. A liability class allows for the penetrance of the locus being evaluated to depend on a covariate - in this case the covariate is a diagnosis of ASD (Vieland 2011). The analysis was again performed and the PPLD|L results are shown in Table 3.

Gene	Chr	Marker	Position (cM)	PPLD|L
LMX1A	1	rs1506946	175.9271	0.0600
LMX1A	1	rs1506944	175.9275	0.0800
LMX1A	1	rs4657430	175.9701	0.0500
LMX1A	1	rs10753669	176.0220	0.0300
LMX1A	1	rs6662349	176.0244	0.0400
NCOR1	17	rs11650682	48.1000	0.0300
CDK5R1	17	rs756787	63.1244	0.0244
CDK5R1	17	rs9889771	63.1262	0.0400

Table 3 –PPLD|L values above 0.02 for all SNPs analyzed using NVMSD:ALL as the affection status and autism as the liability class; NQA-V individuals were classified as 'affected'.

Discussion

In an effort to identify genes associated with autism, and its related phenotype NVMSD, in regions where linkage has previously been identified (Flax 2010), IPA software was utilized (Ingenuity® Systems, www.ingenuity.com). Candidate genes were chosen based on their known functions and how they could plausibly relate to the developing autistic brain. Once a reasonable set of promising genes were picked, SNPs within those genes were tagged using HapMap, SNPBrowser, and 1000 Genomes as resources. The SNPs were genotyped using a bead-based oligonucleotide ligation assay on a Luminex flow cytometer (Bruse 2008). The genotype data was error checked using the programs Pedcheck and Merlin, and statistical analysis was performed using the software package KELVIN.

Possible candidate gene involvement in autistic brain abnormalities

Each of the genes considered in this study could play a possible role in the neuroanatomical abnormalities shown to occur in the autistic brain. Excessive neuronal overgrowth resulting in the significantly increased number of neurons seen in the autistic PFC could be explained by unchecked rampant neural growth or a failure to naturally prune back the numbers of neurons during development through apoptosis, or a combination of the two phenomena (Courchesne 2011). The roles that each of these genes play in normal neuronal development gives insight into what problems they may cause when functioning improperly.

LMX1A has been associated with the neurological disorders Parkinson's disease and schizophrenia (Bergman 2010). Schizophrenia shares several phenotypic characteristics with autism, such as increased risk of epilepsy, intellectual disability, and social and communication deficits. In mouse studies the *LMX1A* homolog *Lmx1a* was shown, through positional cloning, to be the gene responsible for the *dreher* mutant mouse, which is characterized by, among other brain abnormalities, hypoplasia of the cerebellum. In the *dreher* mouse, *Lmx1a* is expressed in the roof plate of the developing spinal cord, the neural crest, and the rhombic lip of the hindbrain, and is responsible for roof plate formation as well as the specification for dorsalization of cells in the CNS (Millonig 2000). In another murine study, *dreher* mutant mice were shown to have perturbations in neuronal migration; neuroblasts destined to move into layers II and III of the cerebral cortex 'over-migrate' into layer I (Costa 2001). These disruptions in the function of LMX1A, migration errors in the cortex and perturbations of cerebellar development, could be related to the abnormalities seen in autistic brains.

Alterations in *Ncor* expression has been shown to cause increased anxiety-like behavior and alter juvenile social play in rats when its expression is downregulated through targeted siRNA treatment in the developing rat amygdala (Jessen 2010). Abnormal amygdala morphology has been shown in autistic patients (Courchesne 2007), and increased anxiety and poor social behavior is commonly associated with autism. On the cellular level NCOR1 functions to aid in the repression of transcription factors, a role that if disrupted could cause any number of problems in a developing brain, not just the amygdala. One such role could be that NCOR1 co-represses CBF1, also called RBPJ, a

downstream transcription factor in the notch signaling pathway that is involved in normal neural development (Hermanson 2002).

Approximately 50% of individuals with Neurofibromatosis type 1, or von Recklinghausen disease, suffer from macrocephaly (abnormally large head) that is related to megalencephaly (abnormally large brain). MRI studies revealed that this megalencephaly occurs predominantly in the white matter of the frontal and parietal regions of the brain (Hachon 2011). This increase in frontal brain volume in about half of Neurofibromatosis type 1 patients corresponds to similar regions shown to have increased volume and increased total neuron count in autistic patients (Courchesne 2011). The cause of the overgrowth could be related to the function of NF1 in its control of Ras signaling, which is involved in cell proliferation (Hachon 2011). NF1 is mainly expressed in neurons, Schwann cells, and oligodendrocytes and plays a part in regulating cell proliferation through Ras signaling and adenylyl cyclase, as well as association with microtubules and cytoskeleton (see Figure 8).

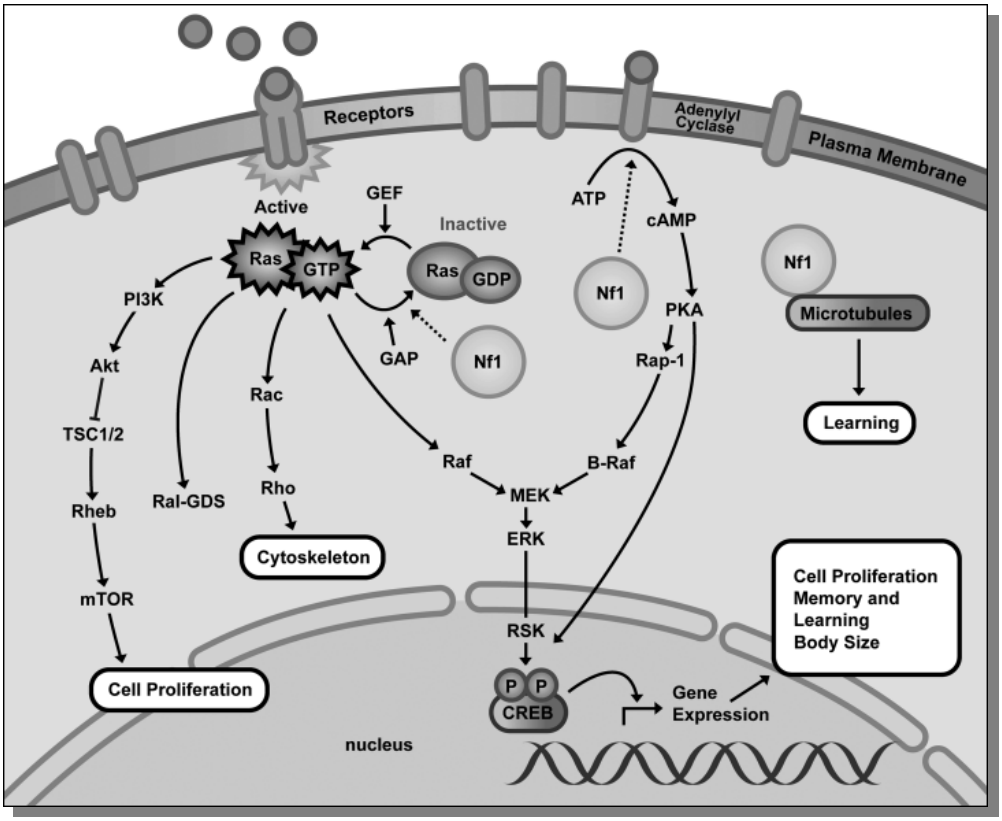

Figure 8. NF1 involvement in different cellular pathways. (Trovó-Marqui 2006).

Located within the *NF1* gene, the *OMG* gene encodes the oligodendrocyte-myelin glycoprotein (OMG). OMG is expressed on the surface of oligodendrocytes where it can bind to the Nogo Receptor on nearby neurons and act to inhibit their axonal outgrowth (Figure 9) (Wang 2002; Lee 2003; Venturin 2006). Any mutations in this gene could drastically alter neuronal growth and connectivity; too much OMG would result in too much inhibition leading to a lack of growth, whereas too little OMG would result in too little inhibition and, consequently, an overgrowth of neurites. A known mutation caused by a SNP (rs11080149) in the coding region of this relatively small gene has already been associated with autism in one study (Martin 2007).

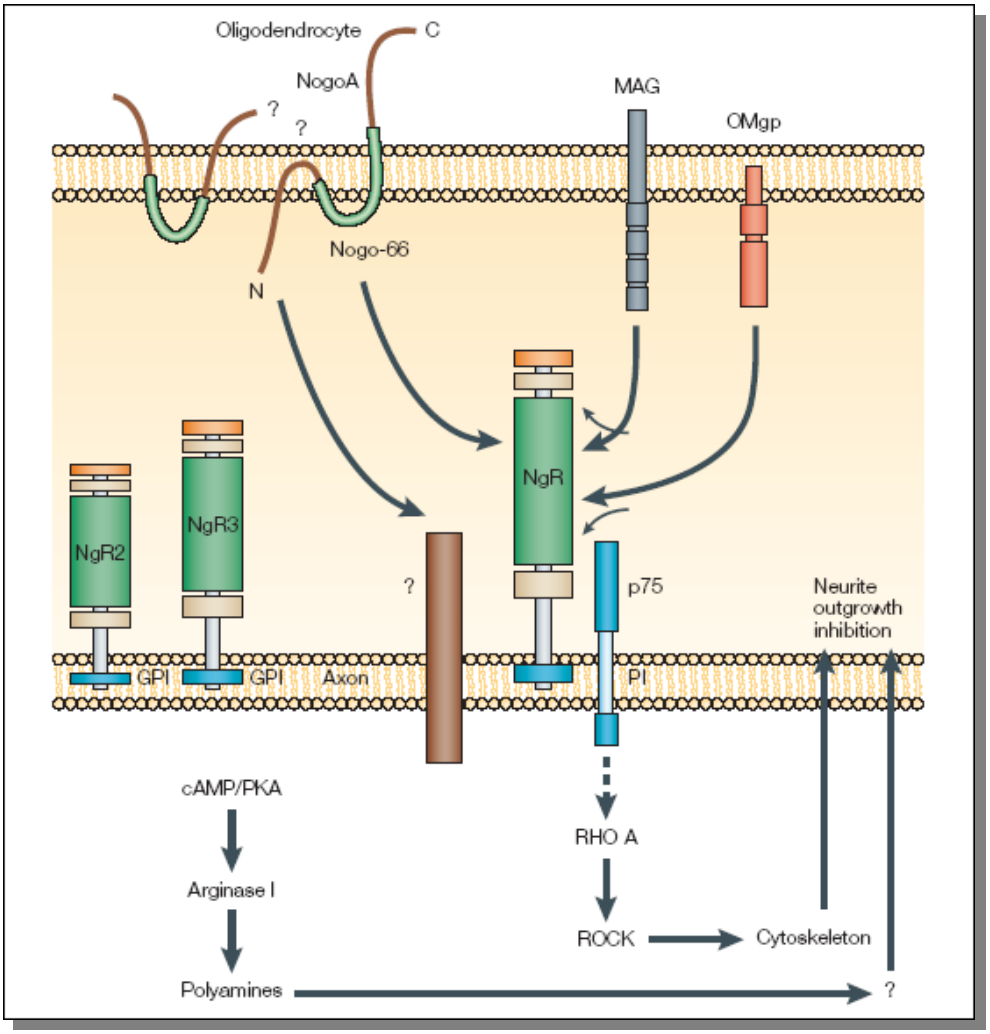

Figure 9. OMG (here OMgp) binding to Nogo Receptor and inhibiting neurite outgrowth in the NgR pathway (Lee 2003).

The p35 protein encoded for by *CDK5R1* functions as an activator of CDK5, a kinase that plays multiple roles during normal neuronal development, neuronal survival, modulation of cytoskeleton, and synaptic plasticity. p35 and CDK5 have also been shown to be factors in multiple neurodegenerative disorders: Parkinson's Disease (PD), Amyotrophic Lateral Sclerosis (ALS) or Lou Gehrig's disease, Alzheimer's Disease, and others. Normally, when a cell is under certain kinds of stress that result in increased intracellular Ca^{2+}, such as excitotoxicity, mitochondrial dysfunction, oxidative stress, and

inflammation to name a few, a proteolytic enzyme called calpain cleaves p35 into its truncated form p25. p25 forms a much more stable complex with CDK5, and this complex becomes hyperactive. Hyperactivity of the CDK5/p25 complex leads to hyperphosphorylation of many cytoskeletal proteins, and eventually to apoptosis of the cell, as illustrated in Figure 10 (Shukla 2011). In mice, transient levels of p25 increase synaptogenesis and facilitate learning, while prolonged, high levels of p25 lead to neuronal loss, a decrease in long-term potentiation, and learning impairments (Fischer 2005). Disturbances in the regulation of this system may play a part in the lack of neuronal pruning seen in the autistic brain, which could be a factor in neuronal overgrowth. CDK5 also acts to phosphorylate DAB1 (Figure 11) such that it inhibits reelin signaling. Reelin signaling is critical in cortical layer formation in the developing brain, and disruption in the pathway drastically perturbs the organization of the cortical layers (Ohshima 2007).

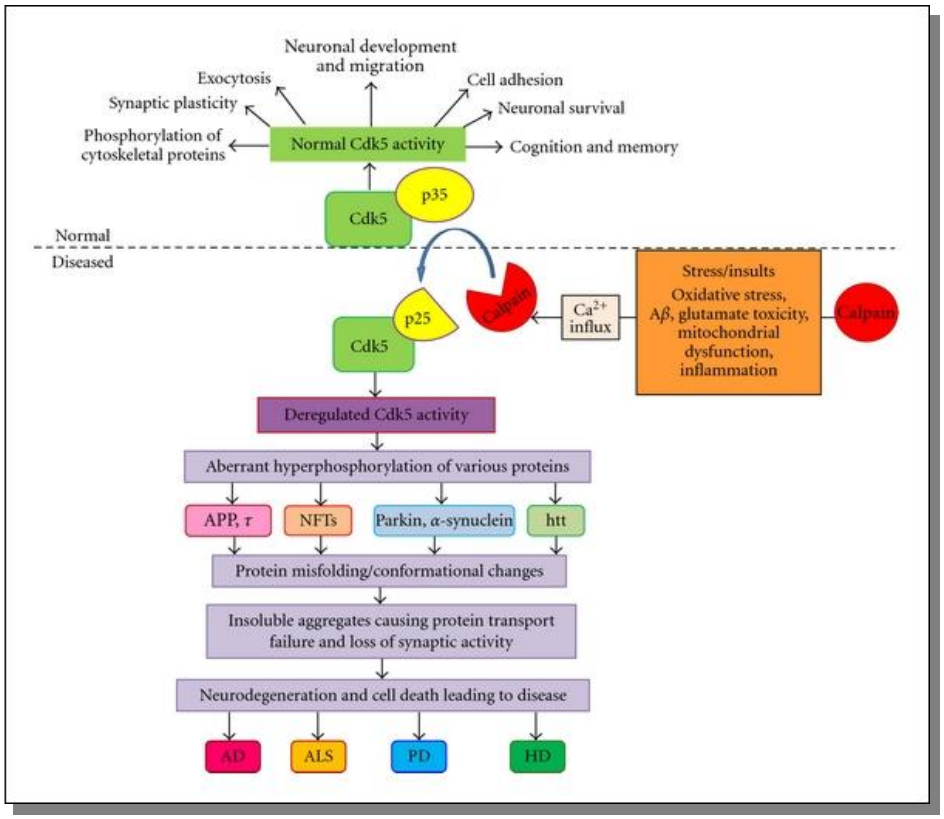

Figure 10. The role of p35 in the regulation of CDK5 and its downstream effects, and disease model implications. (Shukla 2011)

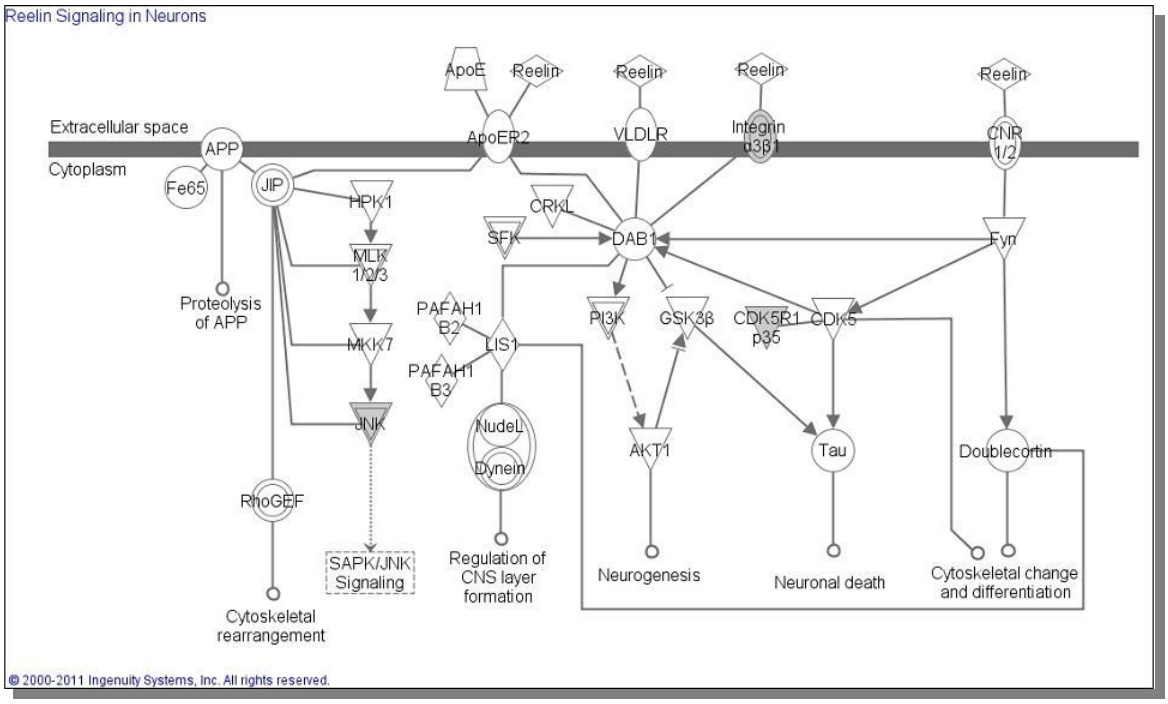

Figure 11. Interactions between CDK5, p35, and the Reelin signaling pathway through DAB.

Discussion of KELVIN

The PPL framework that is the heart of the KELVIN software package is a novel method for analyzing genetic data. It differs from traditional statistical genetics methods in that it is model free; this is possible because trait parameters are treated as nuisance parameters by integrating them out, which makes it ideal for complex, polygenic disorders that may not follow simple dominant or recessive models. Another important difference is that not only can it show evidence in favor of linkage/association, but it also has the ability to report evidence against linkage/association (Huang 2011).

KELVIN reports several different statistical measures, and knowing which statistic applies to the data being analyzed is paramount. The PPL is the posterior probability of linkage, which directly measures the probability that linkage has been found, and uses a prior probability of linkage (π_L) equal to 0.02. This prior probability of linkage is based on the fact that any two randomly selected loci in the human genome have a 2% chance of being linked (Elston and Lange 1975). Developed from this PPL framework, the posterior probability of linkage disequilibrium measures association and represents the joint posterior probability of both linkage and linkage disequilibrium (LD). PPLD is utilized when looking for association in a region where there is no evidence for linkage, and employs the joint prior probability of linkage and LD (π_{LD}), which is obtained from the product of π_L (0.02) and π_{LD} (0.02), yielding 0.0004. When analyzing loci in a region where linkage is known to exist, the posterior probability of linkage disequilibrium given linkage (PPLD|L) is the relevant statistic. PPLD|L also uses the joint prior probability of linkage and LD, however in this case π_L is set to 1 while π_{LD}

remains 0.02. Therefore, when calculating PPLD|L, the prior probability of both linkage and LD is 0.02; this is calculated in the same way as for the PPLD.

Interpreting KELVIN Results

The analysis procedure described above was utilized to analyze the final genotype data in a number of ways. The initial analysis used NVMSD:ALL as the affection status, and in this case the PPLD|L is the relevant statistic because these genes were chosen based on prior evidence of PPL (Flax 2010). This yielded the nominally suggestive results seen in Table 1. The highest values for this analysis were observed for rs1506944, rs1506946, and rs4657430 in the *LMX1A* gene with PPLD|L scores of 0.07, 0.06, and 0.05 respectively. Also notable are rs9889771 in the flanking region of *CDK5R1* with a PPLD|L score of 0.05, and rs11650682 in *NCOR1* with a PPLD|L score of 0.03. To look a little deeper into the data, a few other analysis strategies were employed.

Since these samples were all obtained from the AGRE data set, a significant proportion of the individuals are also diagnosed with autism, so KELVIN analysis was run on the same genotype data but using autism (specifically ASD) as the affection status instead of NVMSD:ALL. Twenty three of the probands in this study were classified as Not Quite Autism (NQA); twenty were NQA and non-verbal (NQA-NV) and three individuals were NQA and verbal (NQA-V). Initially all NQA individuals were assumed to be ASD. This analysis was run again classifying the three NQA-V individuals as 'unknown', but only yielded small variations in the results (Appendix).

The PPLD|L scores of some of the SNPs in *LMX1A* were diminished, and others slightly increased, while new SNPs rose above threshold. For *CDK5R1*, rs756787

dropped below threshold, and rs9889771 decreased slightly. Two of the three SNPs fell below threshold for *NCOR1*, while the PPLD|L score for rs11650682 increased to 0.04. Interestingly, using autism as an affection status yielded PPLD|L values above threshold for SNPs not seen with the previous analysis using NVMSD:ALL as the affection status, including several new SNPs in *LMX1A*. rs2953014 (located in *NF1*) scored above 0.02 in this analysis suggesting that *NF1* could potentially be involved in the ASD phenotype, but not in NVMSD phenotype.

In the last analysis, the diagnosis of NVMSD:ALL was again used as the affection status, while autism was treated as a liability class, with the NQA individuals again classified as affected. The results of using ASD as a liability class saw little variation in PPLD|L scores across the genotyped SNPs, though some fell below threshold. All SNPs located in *NF1* and *OMG* failed to rise above threshold in this analysis. However, rs1506944 in *LMX1A* gave the highest PPLD|L value seen across all the analyses, yielding a value of 0.08.

Problems with SNP selection

When clustering the raw genotype data for the *NCOR1* SNPs, an unexpected pattern in the genotypes became visually apparent. Several of the SNPs appeared to have extremely similar genotype patterns and looked like they were in very high linkage disequilibrium (LD) with each other (Figure 12). When picking SNPs to genotype a given region, any documented LD between SNPs is taken into consideration, and those SNPs that have highly correlated genotypes are avoided. If the chosen SNPs are in high LD with each other, they yield redundant information, in that knowing the genotype of

any one of the SNPs would allow knowledge of the genotype of any other SNPs with high confidence. This introduces an impractical redundancy that would waste time and reagents, which is why representative tag SNPs are selected so that no two tag SNPs share an r^2 score ≥ 0.8. When these SNPs were originally selected, HapMap and Applied Biosystems (SNPbrowser) were the main sources, but when there was not enough SNP coverage of a gene, meaning large physical gaps in coverage, 1000 Genomes was used. An inherent problem exists that since these three databases were used, and they all use different populations for their genotype data, LD can only be calculated within the populations, but not between them. Thus, it was expected that there would be some degree of redundancy; however the main source of uninformative SNPs was 1000 Genomes. Another issue was at the time of this phase of the study, 1000 Genomes lacked sufficient genotype data to accurately detect LD between SNPs using a standard software package like PLINK. Another observation regarding SNPs chosen from the 1000 Genomes database was that several of the SNPs lacked sufficient validation checks, and turned out to be monomorphic, meaning they may not have been true polymorphisms.

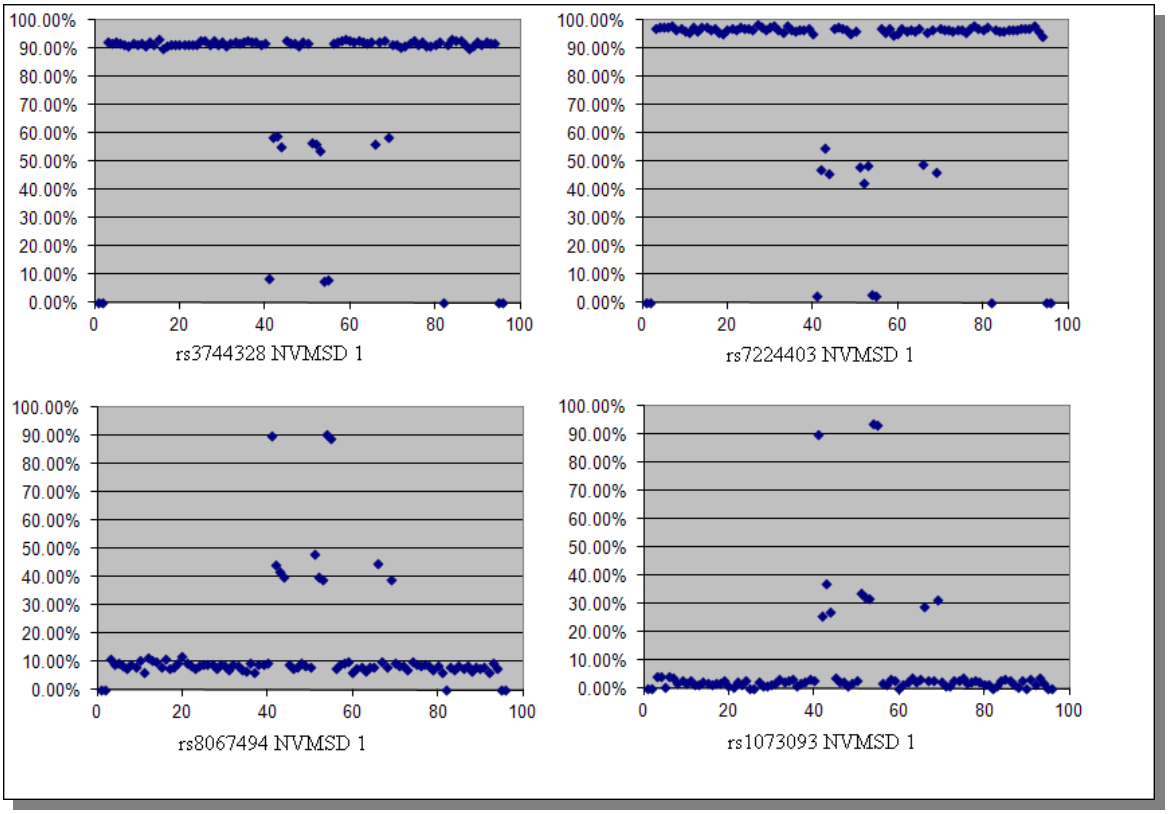

Figure 12. An example of the similar genotypes between four SNPs in *NCOR1*, corresponding to the same set of individuals. Y-axis represents the Percent of allele one; X-axis represents the individuals.

After genotyping the 29 SNPs in the *NCOR1* gene, it was apparent that several SNPs shared a high degree of LD, thus a Python script was used to calculate the percent of pairwise genotype similarity between all of the SNPs observed for all individuals assayed for this gene. This data comparison was put into Excel and color coded to better visualize which SNPs are in the highest LD with each other. Anything that has been highlighted in Table 7 is above the r^2 of 0.8 cutoff used in the SNP tagging process.

0.33	0.33	0.33	0.99	0.33	0.97	0.33	0.33	0.98	0.91	0.33	0.91	0.33	0.87	0.33	0.75	0.98	0.88	0.33	0.33	0.98	0.33	0.33	0.33	0.33	0.33	0.83	0.99	0.33
	0.84	1	0.33	0.99	0.34	0.68	0.99	0.33	0.34	0.99	0.34	0.68	0.32	0.96	0.31	0.33	0.32	0.96	0.87	0.33	0.87	0.99	0.98	0.98	0.99	0.29	0.33	0.99
		0.84	0.33	0.86	0.34	0.69	0.86	0.32	0.29	0.85	0.3	0.69	0.32	0.83	0.32	0.32	0.31	0.83	0.87	0.32	0.87	0.86	0.86	0.86	0.86	0.34	0.32	0.86
			0.33	0.99	0.34	0.68	0.99	0.33	0.34	0.99	0.33	0.68	0.32	0.96	0.31	0.33	0.32	0.96	0.87	0.33	0.87	0.99	0.98	0.98	0.99	0.29	0.33	0.99
				0.33	0.98	0.33	0.33	0.99	0.91	0.33	0.91	0.33	0.87	0.33	0.76	0.99	0.89	0.33	0.34	0.99	0.34	0.33	0.33	0.33	0.33	0.84	0.99	0.33
					0.34	0.7	1	0.33	0.33	0.99	0.33	0.7	0.31	0.96	0.31	0.32	0.31	0.96	0.89	0.32	0.89	1	1	1	1	0.29	0.33	1
						0.33	0.34	0.97	0.9	0.33	0.9	0.33	0.87	0.33	0.75	0.97	0.88	0.33	0.34	0.97	0.34	0.34	0.34	0.34	0.34	0.83	0.98	0.34
							0.7	0.34	0.38	0.69	0.39	1	0.41	0.67	0.34	0.34	0.4	0.67	0.79	0.34	0.79	0.69	0.7	0.7	0.69	0.44	0.34	0.69
								0.33	0.33	0.99	0.33	0.7	0.31	0.96	0.31	0.32	0.31	0.96	0.89	0.32	0.89	1	1	1	1	0.29	0.33	1
									0.92	0.33	0.92	0.34	0.88	0.33	0.74	0.99	0.89	0.32	0.33	0.99	0.33	0.33	0.33	0.32	0.33	0.84	0.99	0.32
										0.34	1	0.38	0.96	0.33	0.67	0.92	0.97	0.33	0.29	0.92	0.29	0.33	0.33	0.33	0.33	0.91	0.92	0.33
											0.33	0.69	0.32	0.97	0.31	0.32	0.32	0.97	0.88	0.33	0.88	0.99	0.99	0.99	0.99	0.29	0.33	0.99
												0.39	0.96	0.33	0.68	0.92	0.97	0.33	0.29	0.92	0.29	0.33	0.33	0.33	0.33	0.91	0.92	0.33
													0.41	0.67	0.34	0.34	0.4	0.67	0.79	0.34	0.79	0.69	0.7	0.7	0.69	0.44	0.33	0.69
														0.33	0.65	0.89	0.99	0.33	0.29	0.89	0.29	0.31	0.31	0.31	0.31	0.96	0.88	0.31
															0.3	0.32	0.33	1	0.85	0.33	0.85	0.96	0.96	0.96	0.96	0.31	0.33	0.96
																0.74	0.66	0.3	0.31	0.75	0.31	0.31	0.31	0.31	0.31	0.62	0.76	0.31
																	0.9	0.32	0.33	1	0.33	0.32	0.32	0.32	0.32	0.85	0.99	0.32
																		0.33	0.28	0.9	0.28	0.31	0.31	0.31	0.31	0.95	0.89	0.31
																			0.85	0.32	0.85	0.96	0.96	0.96	0.96	0.31	0.32	0.96
																				0.33	1	0.89	0.89	0.89	0.89	0.32	0.33	0.89
																					0.33	0.32	0.32	0.32	0.32	0.85	0.99	0.32
																						0.89	0.89	0.89	0.89	0.32	0.33	0.89
																							1	1	1	0.29	0.33	1
																								1	1	0.29	0.33	1
																									1	0.29	0.33	1
																										0.29	0.33	1
																											0.85	0.29
																												0.32

Table 4. Yellow represents pairwise genotype correlations between 80% and 90%; Orange ranges between 90% and 99%; and Red indicates anything at 100%.

Further Study - Increasing sample size

Increasing the sample size has the potential to increase the significance of the results, as more samples can provide increased statistical power. Increasing the sample size in conjunction with altering our statistical analysis parameters could have quite an impact on the results in the future. Fortunately, there is the option of looking at these SNPs in more families, which are available from the AGRE database. In this study, only 986 individuals were genotyped, though DNA from thousands more subjects are available from AGRE. The initial 1,146 individuals were used because there was both ADI-R and Affymetrix 5.0 SNP data available for them; excluding many others that failed to meet criteria. The requirement for the Affymetrix data is no longer a restriction as it was only initially used to find the linkage and association peaks, but presently, as long as additional AGRE families have ADI-R data and at least two non-verbal individuals, they would be eligible.

Further Study – Supplemental SNPs and Novel genes

Though none of the SNPs analyzed in this study showed a large amount of evidence for association does not mean the corresponding genes are not still candidates for further investigation. It would be worthwhile to assay more SNPs in some of these genes simply because of the poor coverage due to design issues and the high degree of LD between SNPs assayed. *NCOR1* in particular, warrants additional SNP genotyping simply because of the 29 SNPs assayed, only a handful gave unique genotypes, and of those, rs11650682 gave consistent results across all analyses.

Interestingly, NCOR1 protein levels have been shown to be significantly downregulated by estradiol in breast cancer cells. Estradiol does this by up-regulating Siah2, a ubiquitin ligase that targets NCOR1 for degredation in the proteasome (Frasor 2005). Reduction in estradiol levels would consequently lead to an increase in NCOR1 protein levels and therefore increased co-repression of the targets of NCOR1. Estradiol proves to be an interesting hormone with regards to autism in that it has been linked to dendritic density in the rat hippocampus (Woolley 1994), and reduced dendritic density in the hippocampus of autistic individuals has been demonstrated in a postmortem study (Raymond 1996). Estradiol has been shown to have many effects on neuronal development; depending on the region, estradiol can prevent or initiate apoptosis, promote or retard synaptogenesis, and be either neuroprotective or neurodamaging (McCarthy 2009).

In the final step of the estradiol biosynthesis pathway the enzyme aromatase is directly responsible for the conversion of testosterone into estradiol (Ruta 2011). Males must rely on this pathway, specifically the enzyme aromatase, as a source of estradiol in the developing brain. Disturbances in the estradiol biosynthesis pathway could account for the unexplained sex ratio differences in autism based on its activity during sex determination of the developing brain and the fact that it plays a prolific role in neuronal morphology and survival. Females may not be as susceptible to disruptions in this pathway since the developing female brain does not rely on the testosterone to estradiol conversion pathway but receives estradiol from the ovaries; alternately, estradiol could possibly be acting as a protective agent in females.

The fairly consistent signal seen across the different analyses for rs9889771 in the *CDK5R1* gene is enough to warrant its further investigation. rs9889771 is actually located in the flanking region of *CDK5R1* and happens to be within the *MYO1D* gene which has been previously implicated in autism (Stone 2007) and itself would be an interesting candidate for study. The *MYO1D* gene encodes myosin-1d, a class I myosin, which is a monomeric molecular motor expressed in the mammalian brain, and is developmentally regulated in the mouse brain, being expressed in the Purkinje and granule cell layers of the cerebellum (Benesh 2012).

Complex, heterogeneous disorders like autism are often very difficult to study, and require new and unique approaches to truly understand their etiologies. This study has demonstrated that the use of alternative related phenotypes as a way to identify risk loci can be helpful for studying these complex genetic disorders, in this case NVMSD. Despite the issues that arose in the SNP tagging process and the power limitations of using a small sample set, the results produced are interesting enough to warrant further study of the genes included, as well as genes in the surrounding areas that fall under the previously found PPL peaks.

References

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, a J., Phelps, E. A., Anderson,A., et al. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, 37(10), 1111-7.

Alenghat, T., Meyers, K., Mullican, S. E., Leitner, K., Adeniji-Adele, A., Avila, J., Bucan, M., et al. (2008). Nuclear receptor corepressor and histone deacetylase 3 govern circadian metabolic physiology. *Nature*, 456(7224), 997-1000.

Allen, G., Müller, R.-A., & Courchesne, E. (2004). Cerebellar Function in Autism: Functional Magnetic Resonance Image Activation During a Simple Motor Task. *Biological psychiatry*, 56(4), 269-78.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th ed text revision).* Washington, DC: Author.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine*, 25(1), 63-77.

Baron-Cohen, S, Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., & Williams, S. C. (2000). The amygdala theory of autism. *Neuroscience and Biobehavioral Reviews*, 24(3), 355-64.

Benesh, A. E., Fleming, J. T., Chiang, C., Carter, B. D., & Tyska, M. J. (2012). Expression and localization of myosin-1d in the developing nervous system. *Brain Research*, 1440, 9-22.

Bergman, O., Westberg, L., Nilsson, L.-göran, Adolfsson, R., & Eriksson, E. (2010). Preliminary evidence that polymorphisms in dopamine-related transcription factors LMX1A , LMX1B and PITX3 are associated with schizophrenia. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 34(6), 1094-1097.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., Bailey, A., et al. (1994). A case-control family history study of autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 35(5), 877-900.

Bruse, S., Moreau, M., Azaro, M., Zimmerman, R., & Brzustowicz, L. (2008). Improvements to bead-based oligonucleotide ligation SNP genotyping assays. *BioTechniques*, 45(5), 559-71.

Cai, J., Donaldson, A., Yang, M., German, M. S., Enikolopov, G., & Iacovitti, L. (2009). The role of Lmx1a in the differentiation of human embryonic stem cells into midbrain dopamine neurons in culture and after transplantation into a Parkinson's disease model. *Stem cells*, 27(1), 220-9.

Center for Disease Control and Prevention. Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveill Summ.* 2012; 61(SS03);1-19

Chakrabarti, S., & Fombonne, E. (2005). Pervasive developmental disorders in preschool children: Confirmation of High Prevalence. *The American Journal of Psychiatry*, 162(6), 1133-41.

Costa, C., Harding, B., & Copp, A J. (2001). Neuronal migration defects in the Dreher (Lmx1a) mutant mouse: role of disorders of the glial limiting membrane. *Cerebral cortex*, 11(6), 498-505.

Courchesne, E., Campbell, K., & Solso, S. (2011). Brain growth across the life span in autism: age-specific changes in anatomical pathology. *Brain Research*, 1380, 138-45.

Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., Barnes, C. C., et al. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA: The Journal of the American Medical Association*, 306(18), 2001-10.

Courchesne, E., Pierce, K., Schumann, C. M., Redcay, E., Buckwalter, J. A, Kennedy, D. P., & Morgan, J. (2007). Mapping early brain development in autism. *Neuron*, 56(2), 399-413.

de Bakker, P. I. W., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature Genetics*, 37(11), 1217-23.

Elston, R. C., Lange, K. (1975). The prior probability of autosomal linkage. *Ann. Hum. Genet., Lond.* 38:341-350

Fatemi, S. H., Snow, A. V., Stary, J. M., Araghi-Niknam, M., Reutiman, T. J., Lee, S., Brooks, A. I., et al. (2005). Reelin signaling is impaired in autism. *Biological Psychiatry*, 57(7), 777-87.

Flax, J. F., Hare, A., Azaro, M. A., Vieland, V. J., & Brzustowicz, L. M. (2010). Combined linkage and linkage disequilibrium analysis of a motor speech phenotype within families ascertained for autism risk loci. *Journal of Neurodevelopmental Disorders*, 2(4), 210-223.

Fombonne, E. (2003). The Prevalence of Autism. *JAMA: The Journal of the American Medical Association*, 289(1), 87-9.

Frasor, J., Danes, J. M., Funk, C. C., & Katzenellenbogen, B. S. (2005). Estrogen down-regulation of the corepressor N-CoR: mechanism and implications for estrogen derepression of N-CoR-regulated genes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37), 13153-7.

Hachon, C., Iannuzzi, S., & Chaix, Y. (2011). Behavioural and cognitive phenotypes in children with neurofibromatosis type 1 (NF1): the link with the neurobiological level. *Brain & development*, 33(1), 52-61.

Hermanson, O., Jepsen, K., & Rosenfeld, M. G. (2002). N-CoR controls differentiation of neural stem cells into astrocytes. *Nature*, 419(6910), 934–939.

Huang, Yungui. "Association statistics under the PPL framework." doctoral dissertation, University of Iowa, 2011.

Jessen, H. M., Kolodkin, M. H., Bychowski, M. E., Auger, C. J., & Auger, A. P. (2010). The nuclear receptor corepressor has organizational effects within the developing amygdala on juvenile social play and anxiety-like behavior. *Endocrinology*, 151(3), 1212-20.

Lee, D. H. S., Strittmatter, S. M., & Sah, D. W. Y. (2003). Targeting the Nogo receptor to treat central nervous system injuries. *Nature reviews: Drug Discovery*, 2(11), 872-8.

Martin, I., Gauthier, J., D'Amelio, M., Védrine, S., Vourc'h, P., Rouleau, G. A., Persico, A. M., et al. (2007). Transmission disequilibrium study of an oligodendrocyte and myelin glycoprotein gene allele in 431 families with an autistic proband. *Neuroscience Research*, 59(4), 426-30.

Marui, T., Hashimoto, O., Nanba, E., Kato, C., Tochigi, M., Umekage, T., Ishijima, M., et al. (2004). Association between the neurofibromatosis-1 (NF1) locus and autism in the Japanese population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 131B(1), 43-7.

Mccarthy, M. M. (2009). The Two Faces of Estradiol: Effects on the Developing Brain, *Neuroscientist*, 15(6), 599-610.

Millonig, J. H., Millen, K. J., & Hatten, M. E. (2000). The mouse *Dreher* gene *Lmx1a* controls formation of the roof plate in the vertebrate CNS. *Nature*, 403(6771), 764-9.

Moncini, S., Bevilacqua, A., Venturin, M., Fallini, C., Ratti, A., Nicolin, A., & Riva, P. (2007). The 3' untranslated region of human Cyclin-Dependent Kinase 5 Regulatory subunit 1 contains regulatory elements affecting transcript stability. *BMC Molecular Biology*, 8: 111.

Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., Mandell, D. S., et al. (2006). The Epidemiology of Autism Spectrum. *Annual Review of Public Health*, 28:21.1–21.24.

Nolte, J. (2002), The Human Brain: An Introduction to Its Functional Anatomy. (5th ed.). Mosby Inc.

Ohshima, T., Suzuki, H., Morimura, T., Ogawa, M., & Mikoshiba, K. (2007). Modulation of Reelin signaling by Cyclin-dependent kinase 5. *Brain Research*, 1140, 84-95.

O'Connell, J. R., & Weeks, D. E. (1998). PedCheck: A Program for Identification of Genotype Incompatibilities in Linkage Analysis. *American Journal of Human Genetics*, 63(1), 259-66.

O'Halloran, C. J., Kinsella, G. J., & Storey, E. (2011). The cerebellum and neuropsychological functioning: A critical review. *Journal of Clinical and Experimental Neuropsychology*, 34:1 35-56.

Raymond, G. V., Bauman, M. L., & Kemper, T. L. (1996). Hippocampus in autism: a Golgi analysis. *Acta Neuropathologica*, 91(1), 117-9.

Ronald, A., & Hoekstra, R.A. (2011). Autism Spectrum Disorders and Autistic Traits: A Decade of New Twin Studies. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 156: 255-74.

Ruta, L., Ingudomnukul, E., Taylor, K., Chakrabarti, B., & Baron-Cohen, S. (2011). Increased serum androstenedione in adults with autism spectrum conditions. *Psychoneuroendocrinology*. 36:8, 1154 –1163.

Rutter, M., Le Couteur, A., Lord, C. (2003). The autism diagnostic interview, revised (ADI-R). Western Psychological Services.

Schumann, Cynthia Mills, Barnes, C. C., Lord, C., & Courchesne, E. (2009). Amygdala enlargement in toddlers with autism related to severity of social and communication impairments. *Biological Psychiatry*, 66(10), 942-9.

Shukla, V., Mishra, S. K., & Pant, H. C. (2011). Oxidative Stress in Neurodegeneration. Advances in Pharmacological Sciences, 572634.

Stone, J. L., Merriman, B., Cantor, R. M., Geschwind, D. H., & Nelson, S. F. (2007). High density SNP association study of a major autism linkage region on chromosome 17. *Human Molecular Genetics*, 16(6), 704-15.

Trovó-Marqui, a B., & Tajara, E. H. (2006). Neurofibromin: a general outlook. *Clinical Genetics*, 70(1), 1-13.

Venturin, M., Moncini, S., Villa, V., Russo, S., Bonati, M. T., Larizza, L., & Riva, P. (2006). Mutations and novel polymorphisms in coding regions and UTRs of CDK5R1 and OMG genes in patients with non-syndromic mental retardation. *Neurogenetics*, 7(1), 59-66.

Vieland, Veronica J. (1998). Bayesian Linkage Analysis, or: How I Learned to Stop Worrying and Love the Posterior Probability of Linkage. *American Journal of Human Genetics, The*, 63: 947-954.

Vieland, Veronica J, Huang, Y., Seok, S.-C., Burian, J., Catalyurek, U., O'Connell, J., Segre, A., et al. (2011). KELVIN: A Software Package for Rigorous Measurement of Statistical Evidence in Human Genetics. *Human Heredity*, 72(4), 276-88.

Wang, K. C., Koprivica, V., Kim, J. A., Sivasankaran, R., Guo, Y., Neve, R. L., & He, Z. (2002). Oligodendrocyte-myelin glycoprotein is a Nogo receptor ligand that inhibits neurite outgrowth. *Nature*, 47:941-944.

Wang, K., Huang, J., & Vieland, V. J. (2000). The consistency of the posterior probability of linkage. *Annals of Human Genetics*, 64(Pt 6), 533-53.

Woolley, S., & McEwen, B. S. (1994). Estradiol Regulates Hippocampal Dendritic Spine Density via an N-Methyl-D-Aspartate Receptor-Dependent Mechanism. *The Journal of Neuroscience*, 14(12): 7680-7687.

Appendix

Gene	Chr	Marker	Position (cM)	PPLD|L NVMSD:ALL	PPLD|L ASD NQA-V unknown	PPLD|L ASD NQA-V affected	PPLD|L ASD Liability Class NQA-V zeroed	PPLD|L ASD Liability Class NQA-V affected
LMX1A	1	rs1819768	175.8149	0.0107	0.0113	0.0118	0.0104	0.0102
LMX1A	1	rs12023709	175.8185	0.0175	0.0135	0.0134	0.0166	0.0169
LMX1A	1	rs4657411	175.8251	0.0146	0.0137	0.0134	0.0138	0.0139
LMX1A	1	rs16841013	175.8284	0.0127	0.0127	0.0127	0.0122	0.0123
LMX1A	1	rs4656435	175.8293	0.0138	0.0131	0.0131	0.0132	0.0131
LMX1A	1	rs12742667	175.8321	0.0123	0.0102	0.0109	0.0119	0.0118
LMX1A	1	rs4657412	175.8332	0.0132	0.0141	0.0140	0.0127	0.0127
LMX1A	1	rs3738193	175.8456	0.0147	0.0136	0.0138	0.0141	0.0140
LMX1A	1	rs6668493	175.8493	0.0159	0.0154	0.0154	0.0158	0.0159
LMX1A	1	rs6668715	175.8500	0.0141	0.0130	0.0133	0.0140	0.0136
LMX1A	1	rs10737508	175.8552	0.0133	0.0143	0.0143	0.0130	0.0128
LMX1A	1	rs6686960	175.8654	0.0091	0.0094	0.0103	0.0096	0.0094
LMX1A	1	rs10800076	175.8703	0.0183	0.0202	0.0182	0.0166	0.0170
LMX1A	1	rs969099	175.8766	0.0153	0.0152	0.0152	0.0145	0.0144
LMX1A	1	rs1394436	175.8783	0.0136	0.0120	0.0123	0.0134	0.0132
LMX1A	1	rs10494437	175.8816	0.0149	0.0140	0.0148	0.0147	0.0145
LMX1A	1	rs1807837	175.8828	0.0157	0.0151	0.0150	0.0153	0.0153
LMX1A	1	rs10494438	175.8844	0.0144	0.0136	0.0144	0.0143	0.0141
LMX1A	1	rs1394440	175.9186	0.0133	0.0129	0.0131	0.0125	0.0125
LMX1A	1	rs1394436	175.9201	0.0185	0.0148	0.0145	0.0190	0.0194
LMX1A	1	rs1506946	175.9271	0.0600	0.0500	0.0400	0.0500	0.0600
LMX1A	1	rs1506944	175.9275	0.0700	0.0500	0.0400	0.0700	0.0800
LMX1A	1	rs7535109	175.9395	0.0169	0.0194	0.0183	0.0162	0.0163
LMX1A	1	rs10918144	175.9441	0.0156	0.0155	0.0150	0.0149	0.0151
LMX1A	1	rs4657428	175.9586	0.0186	0.0230	0.0215	0.0166	0.0169
LMX1A	1	rs4657430	175.9701	0.0500	0.0300	0.0241	0.0500	0.0500
LMX1A	1	rs1504694	175.9837	0.0157	0.0147	0.0147	0.0153	0.0152
LMX1A	1	rs1343554	175.9864	0.0158	0.0189	0.0181	0.0147	0.0149
LMX1A	1	rs1833030	175.9906	0.0225	0.0300	0.0300	0.0203	0.0211
LMX1A	1	rs12752351	175.9919	0.0170	0.0194	0.0183	0.0158	0.0161
LMX1A	1	rs12405561	176.0014	0.0162	0.0172	0.0166	0.0151	0.0153
LMX1A	1	rs10753669	176.0220	0.0300	0.0300	0.0300	0.0400	0.0300
LMX1A	1	rs6662349	176.0244	0.0400	0.0600	0.0500	0.0400	0.0400
LMX1A	1	rs11799765	176.0529	0.0167	0.0166	0.0164	0.0159	0.0158
LMX1A	1	rs12138441	176.0579	0.0160	0.0151	0.0151	0.0152	0.0152
LMX1A	1	rs1811714	176.0594	0.0151	0.0165	0.0160	0.0146	0.0144

Gene	Chr	Marker	Position (cM)	PPLD|L NVMSD:ALL	PPLD|L ASD NQA-V unknown	PPLD|L ASD NQA-V affected	PPLD|L ASD Liability Class NQA-V zeroed	PPLD|L ASD Liability Class NQA-V affected
LMX1A	1	rs2174696	176.0621	0.0160	0.0205	0.0195	0.0154	0.0151
LMX1A	1	rs7549447	176.0713	0.0148	0.0191	0.0208	0.0139	0.0137
LMX1A	1	rs10800086	176.0766	0.0150	0.0300	0.0248	0.0154	0.0155
LMX1A	1	rs12564138	176.0775	0.0140	0.0169	0.0184	0.0134	0.0132
LMX1A	1	rs10800087	176.0831	0.0140	0.0131	0.0132	0.0133	0.0131
LMX1A	1	rs1948857	176.0875	0.0139	0.0144	0.0143	0.0137	0.0135
LMX1A	1	rs7549587	176.0878	0.0155	0.0148	0.0148	0.0148	0.0148
LMX1A	1	rs10082152	176.1016	0.0145	0.0131	0.0132	0.0134	0.0133
LMX1A	1	rs6665018	176.1065	0.0160	0.0150	0.0150	0.0152	0.0152
LMX1A	1	rs6659644	176.1122	0.0140	0.0139	0.0139	0.0138	0.0136
LMX1A	1	rs1330390	176.1223	0.0147	0.0130	0.0133	0.0141	0.0139
LMX1A	1	rs10918147	176.1232	0.0187	0.0138	0.0143	0.0184	0.0177
LMX1A	1	rs6686310	176.1300	0.0159	0.0165	0.0162	0.0155	0.0152
LMX1A	1	rs34533308	176.1351	0.0149	0.0149	0.0149	0.0146	0.0145
LMX1A	1	rs1506954	176.1400	0.0131	0.0098	0.0117	0.0128	0.0124
NCOR1	17	rs35168566	48.0994	0.0114	0.0093	0.0085	0.0108	0.0110
NCOR1	17	rs3744328	48.0996	0.0205	0.0157	0.0160	0.0203	0.0198
NCOR1	17	rs11650682	48.1000	0.0300	0.0500	0.0400	0.0300	0.0300
NCOR1	17	rs7224403	48.1012	0.0205	0.0156	0.0159	0.0203	0.0197
NCOR1	17	rs1075901	48.1018	0.0111	0.0083	0.0076	0.0105	0.0107
NCOR1	17	rs8067494	48.1029	0.0193	0.0166	0.0168	0.0191	0.0186
NCOR1	17	rs4792718	48.1036	0.0127	0.0131	0.0118	0.0120	0.0122
NCOR1	17	rs11657857	48.1044	0.0168	0.0125	0.0127	0.0169	0.0160
NCOR1	17	rs1073093	48.1050	0.0191	0.0159	0.0161	0.0188	0.0184
NCOR1	17	rs11652025	48.1055	0.0121	0.0101	0.0088	0.0114	0.0116
NCOR1	17	rs9898313	48.1082	0.0108	0.0075	0.0069	0.0101	0.0104
NCOR1	17	rs3760295	48.1086	0.0161	0.0145	0.0145	0.0161	0.0159
NCOR1	17	rs12947205	48.1128	0.0112	0.0078	0.0072	0.0104	0.0107
NCOR1	17	rs62073621	48.1159	0.0166	0.0123	0.0126	0.0165	0.0156
NCOR1	17	rs4791661	48.1171	0.0120	0.0091	0.0085	0.0113	0.0115
NCOR1	17	rs4500784	48.1178	0.0154	0.0143	0.0142	0.0156	0.0153
NCOR1	17	rs11658564	48.1213	0.0118	0.0097	0.0085	0.0113	0.0114
NCOR1	17	rs11655773	48.1234	0.0109	0.0065	0.0058	0.0103	0.0105
NCOR1	17	rs4792723	48.1240	0.0152	0.0141	0.0140	0.0155	0.0152
NCOR1	17	rs62072469	48.1283	0.0174	0.0155	0.0156	0.0175	0.0174
NCOR1	17	rs8079969	48.1301	0.0121	0.0102	0.0089	0.0115	0.0117
NCOR1	17	rs62072471	48.1318	0.0174	0.0155	0.0156	0.0175	0.0174

Gene	Chr	Marker	Position (cM)	PPLD|L NVMSD:ALL	PPLD|L ASD NQA-V unknown	PPLD|L ASD NQA-V affected	PPLD|L ASD Liability Class NQA-V zeroed	PPLD|L ASD Liability Class NQA-V affected
NCOR1	17	rs62072471	48.1318	0.0174	0.0155	0.0156	0.0175	0.0174
NCOR1	17	rs16959544	48.1322	0.0181	0.0162	0.0163	0.0180	0.0176
NCOR1	17	rs16959547	48.1325	0.0186	0.0164	0.0166	0.0185	0.0181
NCOR1	17	rs7225813	48.1374	0.0185	0.0163	0.0165	0.0183	0.0179
NCOR1	17	rs7220699	48.1402	0.0186	0.0164	0.0166	0.0185	0.0180
NCOR1	17	rs21180	48.1440	0.0109	0.0094	0.0090	0.0103	0.0105
NCOR1	17	rs178821	48.1441	0.0118	0.0104	0.0091	0.0112	0.0114
NCOR1	17	rs7212874	48.1569	0.0180	0.0162	0.0163	0.0179	0.0175
NF1	17	rs2905791	60.4336	0.0126	0.0083	0.0078	0.0120	0.0124
NF1	17	rs2952979	60.4373	0.0125	0.0079	0.0073	0.0119	0.0122
NF1	17	rs2953016	60.4910	0.0130	0.0120	0.0121	0.0129	0.0126
NF1	17	rs2953014	60.4942	0.0178	0.0224	0.0231	0.0165	0.0160
NF1	17	rs2905869	60.5552	0.0137	0.0098	0.0094	0.0131	0.0134
NF1	17	rs2905873	60.5904	0.0146	0.0146	0.0151	0.0133	0.0128
NF1	17	rs2905796	60.6042	0.0119	0.0077	0.0072	0.0114	0.0116
NF1	17	rs8067021	60.6504	0.0126	0.0083	0.0078	0.0120	0.0124
NF1	17	rs12600974	60.7414	0.0124	0.0085	0.0080	0.0117	0.0120
OMG	17	rs11080149	60.8333	0.0159	0.0193	0.0161	0.0155	0.0167
NF1	17	rs9902893	60.8398	0.0126	0.0085	0.0079	0.0121	0.0124
NF1	17	rs2040792	60.8478	0.0124	0.0086	0.0078	0.0115	0.0118
CDK5R1	17	rs4523973	63.1116	0.0173	0.0122	0.0122	0.0181	0.0187
CDK5R1	17	rs9899310	63.1138	0.0179	0.0125	0.0120	0.0169	0.0179
CDK5R1	17	rs756785	63.1153	0.0173	0.0122	0.0122	0.0182	0.0190
CDK5R1	17	rs8192474	63.1183	0.0120	0.0093	0.0095	0.0114	0.0114
CDK5R1	17	rs735555	63.1202	0.0116	0.0100	0.0105	0.0113	0.0113
CDK5R1	17	rs731880	63.1215	0.0189	0.0131	0.0124	0.0168	0.0177
CDK5R1	17	rs1047336	63.1226	0.0199	0.0199	0.0199	0.0199	0.0199
CDK5R1	17	rs1047208	63.1234	0.0119	0.0087	0.0083	0.0112	0.0114
CDK5R1	17	rs756787	63.1244	0.0232	0.0134	0.0130	0.0237	0.0244
CDK5R1	17	rs9889771	63.1262	0.0500	0.0500	0.0400	0.0400	0.0400