GENETIC AND PHENOTYPIC DISSECTION OF AUTISM SUSCEPTIBILITY

By

Jacob Lee McCauley

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Molecular Physiology and Biophysics

May, 2005

Nashville, Tennessee

Approved:

Professor James S. Sutcliffe

Professor Jonathan L. Haines

Professor Randy D. Blakely

Professor Jason H. Moore

Professor Doug P. Mortlock

Professor Scott M. Williams

To my parents for always believing in me
Aaron and Sharon McCauley

My children, the most beautiful part of my life
Anna and Alexis

And most importantly to my loving wife, without your support I would be lost
Glendia

ACKNOWLEDGEMENTS

The work presented here was supported by NIH grant MH61009 and a Vanderbilt Kennedy Center Hobbs Discovery Research Award to James S. Sutcliffe; NIH grant MH55135 to Susan E. Folstein; a grant from the Hobbs Society, Vanderbilt John F. Kennedy center to Randy D. Blakely, and both a Human Genetics Training Grant fellowship, Vanderbilt University and a predoctoral fellowship to Jacob L. McCauley from the National Alliance for Autism Research (NAAR).

I also acknowledge the vast contributions to this work from the staff in the Vanderbilt Center for Human Genetics Research, including the Data Analysis (Kim Gainer, Lan Jiang, and Lana Olson), DNA Resources (Cara Sutcliffe), Family Ascertainment (Genea Crockett and Kelly Taylor), and Bioinformatics (Janey Wang) Cores.

None of this work would have been possible without the families who have participated in this research. We recognize the contributions of families obtained by our collaborators at the New England Medical Center (Susan Folstein), those in the AGRE (The Autism Genetic Resource Exchange) consortium, those obtained through our ascertainment site at Vanderbilt University, and finally those of our collaborators in Canada, Ireland, and Portugal.

I'm greatly indebted to the members of my thesis committee (Dr. Randy Blakely, Dr. Jonathan Haines, Dr. Jason Moore, Dr. Doug Mortlock, and Dr. Scott Williams) for their invaluable time, guidance, friendship, and support.

I wish to thank the members of the Sutcliffe lab, past and present (Taneem Amin, Matt Conrad, Ryan Delahanty, Todd Edwards, Michelle Jacobs, Anita Lam, Erika Nurmi, and Ed Organ), for their contributions and dedication to all the ongoing projects in the lab. Taneem and Ed, I thank you both for the stories, the political debates, and the good times. Erika and Michelle, thanks for keeping me in line, giving advice, and always being entertaining. Ryan and Todd, you guys rock. To all, I'm blessed that our paths have crossed and wish you all the best.

Jim, a.k.a. Dr. James Sutcliffe, I want to thank you for your friendship and determination to help me succeed. Without your direction, none of my accomplishments would have been possible. Thank you for the freedom to express and pursue my ideas, no matter how crazy they might sometimes seem. I'm forever grateful of your constant unwavering support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5-HT	Serotonin/ 5-hydroxytryptamine
5-HTT/SERT/*SLC6A4*	Serotonin Transporter
5-HTTLPR	5-HT transporter gene-linked polymorphic region
ADI/ ADI-R	Autism Diagnostic Interview- Revised
ADOS	Autism Diagnostic Observation Schedule
AGRE	Autism Genetic Resource Exchange (http://www.agre.org/)
AS	Angelman Syndrome (OMIM #105830)
ASD	Autism Spectrum Disorder
ATP10A(ATP10C)	ATPase, class V, type 10A
BAP	Broader Autism Phenotype
bp	Base Pair
cDNA	Complementary Deoxyribonucleic Acid
Celera	Private Genome Assembly Database (http://www.celeradiscoverysystem.com/index.cfm)
CEPH	Centre d'Etude du Polymorphism Humain
CHGR	Center for Human Genetics Research
CLSA	Collaborative Linkage Study of Autism
cM	CentiMorgan
dbSNP	Public SNP database (http://www.ncbi.nlm.nih.gov/SNP/index.html)

deCode Genetics	Biopharmaceutical Company (http://www.decode.com/)
df	Degrees of Freedom
dNTPs	Deoxyribonucleotides
dHPLC	Denaturing High Performance Liquid Chromatography
DNA	Deoxyribonucleic Acid
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
DZ	Dizygotic Twins
Ensemble	Genome Browser (http://www.ensembl.org/)
FBAT	Family Based Association Tests (http://www.biostat.harvard.edu/~fbat/fbat.htm)
GABA	Gamma (γ)-aminobutyric acid
GABRA5	$GABA_A$ receptor, alpha 5
GABRB3	$GABA_A$ receptor, beta 3
GABRG3	$GABA_A$ receptor, gamma 3
GOLD	Graphical Overview of Linkage Disequilibrium
Haploview	Java-based Tool for Visualizing LD blocks (http://www.broad.mit.edu/mpg/haploview/index.php)
HLOD	Heterogeneity LOD Score
htSNP	Haplotype Tag Snp
HWE	Hardy-Weinberg Equilibrium
IC	Imprinting Center
IMGSAC	International Molecular Genetic Study of Autism

ITGB3	Integrin, beta3(platelet glycoprotein IIIa,antigen CD61)
kb	Kilobase
LD	Linkage Disequilibrium
LOD	Logarithm of the Odds
MAOA	Mono-amine Oxidase A
Mb	Megabase
MED	Maternal Expression Domain
ml	Milliliter
MLS	Multipoint LOD Score
MMLS	Maximum Multipoint LOD Score
MMR	Measles-Mumps-Rubella
mRNA	Messenger Ribonucleic Acid
MZ	Monozygotic twins
NEMC	New England Medical Center
ng	Nanogram
NIMH	National Institute of Mental Health (http://www.nimh.nih.gov/)
OCD	Obsessive-Compulsive Disorder
OMIM	Online Mendelian Inheritance in Man
OSA	Ordered-Subsets Analysis
PCR	Polymerase Chain Reaction
PDD	Pervasive Developmental Disorders
PDT	Pedigree Disequilibrium Test

PWS	Prader-Willi Syndrome (OMIM #176270)
QTL	Quantitative Trait Locus
SNP	Single Nucleotide Polymorphism
SSRI	Selective Serotonin Reuptake Inhibitor
T_A	Annealing Temperature
Taq	*Thermus Aquaticus* Polymerase
TD	Transmission Disequilibrium
TDT	Transmission Disequilibrium Test
TPH1	Tryptophan 5-hydroxylase 1
TPH2	Tryptophan 5-hydroxylase 2
UBE3A	Ubiquitin Protein Ligase E3A
UTR	Untranslated Region
VISTA	Visualization Tools for Alignment (http://www.gsd.lbl.gov/vista/)
VNTR	Variable Number Tandem Repeat
μg	Microgram

CHAPTER I

INTRODUCTION

Autism and the Broader Autism Spectrum

Autism (OMIM #209850) is a devastating neurological disorder in which affected individuals have life-long deficits related to three core phenotypic areas/domains: deficits in the development of language; inappropriate social understanding, interactions, and behavior; and patterns of repetitive, restrictive, and compulsive interests and behaviors. Classic autism, or autistic disorder, is distinguished from other Pervasive Developmental Disorders (PDDs), including Asperger's syndrome, Rett syndrome, and PDD not otherwise specified (PDD-NOS) through the use of diagnostic instruments that rate various aspects of the broader autism phenotype (BAP) [1]. These instruments include the DSM-IV, the Autism Diagnostic Interview (ADI) and its revision (ADI-R), and the Autism Diagnostic Observation Schedule (ADOS) [2-4]. The ADI and ADOS are used universally to obtain a research diagnosis of autism. A small percentage (~10%) of patients presenting with autistic features can be ascribed to single-gene disorders, such as fragile X syndrome, tuberous sclerosis complex [5], Angelman syndrome [6], and Rett syndrome [7]. Prominently identified features in patients with autism include mental retardation (~70%), macrocephaly (~20%), epilepsy

and/or seizures (~15-30%), gastrointestinal abnormalities, hypotonia, and motor stereotypies [8, 9].

History and Epidemiology

The term "autism" was coined by Eugen Bleuler, a Swiss psychiatrist and psychologist, around 1912 and originally referred to *"an escape from reality"*. In his 1943 article entitled "Autistic Disturbances of Affective Contact" in the journal *Nervous Child*, Leo Kanner adapted this term to describe children he believed to be afflicted with a syndrome not previously described. Kanner remarked that affected children *"have come into the world with innate inability to form the usual, biologically provided affective contact with people"*. His observations describe specific features of the clinical presentation: "*extreme autistic aloneness*"; delayed and abnormal speech with echolalia, pronominal reversal, literalness and inability to use language for communication; and monotonous, repetitive behaviors with an *"anxiously obsessive desire for the maintenance of sameness"*. Many children were believed to be deaf or hard of hearing due to their lack of response to questions or commands. Kanner indicates: *"Everything that is brought to the child from the outside, everything that changes his external or even internal environment, represents a dreaded intrusion"*. These intrusions may often be ignored. If persistent, however, they cause panic, stress, and despair, often being expressed as temper-tantrums [10, 11].

Physically, the eleven children Kanner originally reported were normal, however, he noted that five had enlarged heads. Three of the eleven were mute. He also noted that only three of the eleven were female. Other anecdotal observations included a curious commonality of backgrounds for these children. Kanner writes *"It is not easy to evaluate the fact that all of our patients have come of highly intelligent parents. This much is certain, that there is a great deal of obsessiveness in the family background. One other fact stands out prominently. In the whole group, there are very few really warmhearted fathers and mothers. For the most part, the parents, grandparents, and collaterals are persons strongly preoccupied with abstractions of a scientific, literary, or artistic nature, and limited in genuine interest in people"* [10]. These observations of non-autistic family members would later prove to be genetically important. At the time, the bias in psychiatry attempted to explain psychiatric disorders as a result of poor parenting, leading society to believe this was the cause of autism. Later studies disproved this by demonstrating no significant difference in the parents of autistic individuals to those of non-autistic children [8, 12].

Despite that fact that "classic autism" as we know it today was not formally described until 1943 by Leo Kanner, a few accounts of conditions in the eighteenth and nineteenth centuries are similar in nature. One case described by Uta Frith described 39 year-old Hugh Blair. When he appeared in court for a 1747 trial, witness depositions described him as lacking common sense and having a "silent madness". More specific descriptions led Frith to believe this to

be an early report of autism. Other early cases reports by John Haslam, Henry Maudsley, and Jean Itard are thought to represent case descriptions of individuals with pervasive developmental disorders (PDDs) (i.e. the broader autism spectrum), likely including Asperger syndrome. Kanner never believed these to meet his description of “autism”, while others argue these cases could have been autism. Regardless, over time many noteworthy investigators have refined Kanner’s description of autism. In doing so, they have shown that the clinical diagnosis and description for autism presents across a spectrum of severity involving the three core phenotypic domains [11].

Autism epidemiology is quite complex. Surveys of autism prevalence started in the mid-1960s in England and have since been conducted worldwide. Original surveys were based on a narrow definition of autistic disorder representing severe impairment of language, social interaction, and repetitive behaviors. Over time, diagnostic criteria have evolved to include cases on both less severely and more severely affected ends of the spectrum. This has complicated efforts to establish precise estimates of autism prevalence within and across populations over time [13]. Fombonne points out that rates for a narrow diagnosis of autism disorder in recent surveys has been consistently higher than 10/10,000 whereas previous prevalence estimates were on the range of ~4-5/10,000. For a broader diagnosis of autism spectrum disorder, the prevalence is ~60/10,000. The ratio of affected males to affected females is 4:1 for classic autism, and even higher in the broader spectrum. He concludes that

available evidence suggests that rates of autism spectrum disorders (ASDs) and autistic disorder appear 3 to 4 times higher now than rates estimated in the 1970s [9, 14-16]. Fombonne outlines other confounding factors that may help explain some or all of this apparent increase in prevalence [17].

Although classic autism is a distinct syndrome, current literature supports the description of autism as more appropriately presenting across a spectrum. DSM-IV-defined autistic disorder represents one slice of this spectrum of behavioral restriction and social and communication impairment [9].

Genetics

There has been long-standing support from twin, family, and segregation studies providing evidence of a substantial genetic component in the etiology of autism. Twin studies have shown a 60% to 90% concordance rate for classic autism among monozygotic (MZ) twins depending on use of narrow or broad diagnostic criteria, respectively. Dizygotic (DZ) twins, in contrast, show a 0-10% concordance under the same models [9, 18, 19]. Studies have estimated the recurrence risk for siblings of a child with autism to be anywhere from ~50-100 times greater than the population prevalence [19, 20]. Current estimates of the sibling recurrence risk are 6 to 8% for autistic disorder [8, 9]. The heritability, or the proportion of the total phenotypic variation due to genetic variation, of 90% indicates that autism is among the most genetic of neuropsychiatric conditions [8, 14]. Various studies indicate that between five and fifteen (and possibly more)

genes contribute to overall genetic risk for idiopathic autism. Models support oligogenic inheritance within individual families with locus heterogeneity resulting in different families possessing a different collection of susceptibility alleles [20, 21]. Traits or milder phenotypes presenting in relatives of individuals with autism may reflect inheritance of allelic subsets of those present in the affected individual [8]. Thus, autism is not merely a genetic disease, but a disorder with a complex genetic architecture. Despite such strong support for a genetic etiology, no autism susceptibility gene has been definitively identified.

Alternative hypotheses regarding the genetics of autism have been put forward. Some have proposed that epigenetic effects at susceptibility genes play an important role in autism etiology [22]. Epigenetics relates to alterations in gene expression without a necessarily accompanying sequence variation. Genomic imprinting is an example of an epigenetic phenomenon. There has also been substantial discussion involving environmental risk factors such as the Measles-Mumps-Rubella (MMR) vaccine or the earlier (higher mercury-containing) formulations of the vaccine preservative thimerosal [23]. The MMR hypothesis has been widely disproved [24]. While the case for thimerosal is less certain, studies are underway to evaluate the potential epidemiological effects for exposure to these older vaccine preservatives that contained relatively higher levels of mercury.

To identify autism susceptibility genes, several groups have undertaken unbiased, genome-wide linkage screens using varying numbers of families with

multiple affected individuals (e.g. affected sibling pairs) [21, 25-33]. These studies have identified numerous linkage peaks on many chromosomes. Comparison of these results indicates linkage on 7q and 2q as being most consistently replicated. However, no single gene has been convincingly demonstrated to underlie increased risk for autism on these or other chromosomes. Numerous candidate gene studies have been conducted in autism on the basis of available information concerning chromosomal abnormalities, linkage, altered neurobiology and neuropathology seen in autism [8].

Neurobiological Candidate Systems

While there are several neurobiological systems that are potentially involved with genetic risk for autism, the main focus of this dissertation is on the serotonin (5-hydroxytryptamine, 5-HT) and GABA (γ-aminobutyric acid) systems. Multiple lines of investigation point to abnormalities in both of these systems as being involved in autism. Several genes within these systems have been examined for involvement in disease risk, although no *consistent* results implicate common alleles at any particular gene.

Well-replicated observations of elevated platelet serotonin in ~20-25% of people with autism and first-degree relatives implicates serotonin as being relevant to autism etiology [34-36]. Selective serotonin reuptake inhibitors (SSRIs), preferentially targeting the serotonin transporter (SLC6A4), are often

effective in treating anxiety, rituals and aggression in autism and related disorders [37, 38]. Numerous association studies in autism have failed to yield consistent findings with serotonin-related genes in general and *SLC6A4* in particular. However, most of these studies have focused on 5-HTTLPR, an insertion/deletion polymorphism in the promoter of *SLC6A4*.

GABA is the major inhibitory neurotransmitter in the adult brain although it mediates excitatory transmission during development. Recent publications have shown: (1) the number of $GABA_A$ receptors are significantly decreased in brains of children with autism [39]; (2) plasma GABA, and its essential precursor glutamate, are elevated in children with autism [40-42]; (3) benzodiazepines, which are effective in treating seizures, anxiety, and social phobia that are seen in autism, bind to and act, on $GABA_A$ receptors [43]; (4) GABA-ergic transmission has important trophic actions during development; (5) genetic evidence points to the 15q11-q13 region in general and the *GABRB3* gene in particular [31, 44-47]. Based on these data, the $GABA_A$ receptor subunit genes, particularly those in 15q11-q13, represent excellent candidates, allelic variants of which could confer genetic susceptibility for development of autism.

Genomic Candidate Regions

Though there are multiple regions of interest based on linkage findings or chromosomal abnormalities (e.g. 7q, 2q), the first part of this dissertation will be focused on examination of candidate genes in two specific regions: 17q11.2 and

15q11-q13. The former region has been identified by ongoing linkage studies that are also part of this work. Initial linkage analysis of 158 multiplex autism families revealed a multipoint heterogeneity LOD (HLOD) score of 2.74 under a dominant model at ~53 cM on 17q11.2. This score increased to 3.65 when the families were analyzed separately based on relative affection for rigid-compulsive behaviors, a sub-phenotype of autism. The International Molecular Genetic Study of Autism Consortium (IMGSAC) has also observed suggestive linkage at this site [48]. The serotonin transporter locus (*SLC6A4*), long considered a functional candidate in autism, maps directly underneath the linkage peak. Given multiple lines of evidence, one aim of this thesis project is to test the hypothesis that *SLC6A4* is involved in autism susceptibility.

The second region, 15q11-q13, was hypothesized to harbor an autism susceptibility locus based on observations of duplications and triplications affecting this region in a small percentage of people with autism (reviewed in Veenstra-VanderWeele et al.) [9, 49, 50]. While not universally observed by all groups [30, 51, 52], both linkage and association have been observed in this region [31, 44-47]. Our group and Shao et al. have documented increased linkage in trait-based subsets of autism in a region containing a cluster of $GABA_A$ receptor subunit genes [53, 54]. Several groups have detected association at microsatellite and SNP markers in the *GABRB3* gene, and the data presented in this dissertation substantially extends these findings. A detailed analysis of allelic association throughout the GABA cluster, as well as the interval involved in

genomic imprinting control and that containing maternally-expressed genes (*UBE3A* and *ATP10A*) 5' of the $GABA_A$ receptor subunit gene cluster will be presented.

The final part of the dissertation describes a series of genome-wide analyses for autism and quantitative traits, representing subsets of the broader autism phenotype. These analyses point to additional potential candidate regions that may harbor autism-related risk factors. Regions showing significant or highly suggestive linkage were further investigated by analysis of additional markers to refine linkage data and test a limited number of promising candidate gene loci.

Genetic Analysis of Complex Disease

Three common approaches are used to facilitate identification of genes for complex genetic disorders. These are as follows: (1) Genome-wide linkage analysis, a biologically unbiased method for detecting regions of the genome more frequently inherited in common by affected individuals in a family than predicted by chance; (2) Test candidate systems likely to be involved in disease etiology identified based on altered physiology or pathology seen in patients with the disorder. Such information can then be used in speculation concerning particular biological systems or pathways, and ultimately specific proteins in those systems. Thus, by being well informed about altered biology in a disease condition, one may test specific genes, in a hypothesis-driven manner, directly for involvement in disease risk; (3) Examination of rare cases or families with the

disorder presenting with chromosomal abnormalities that may identify a relevant genomic interval or gene locus. Co-localization with linkage and/or functional candidate genes makes such findings even more compelling to investigate.

The purpose of a genomic linkage screen is to identify regions in the genome that are more frequently inherited by affected family members than would be expected by chance. The amount of allele-sharing within and across families is measured in LOD (logarithm of odds) score units. This is calculated based on the relationship between recombinant and non-recombinant alleles at a given marker and across nearby markers [55]. Traditional genome-wide analyses have exploited highly polymorphic microsatellite markers with an average genome-wide spacing of 10 cM or less to provide sufficient coverage and information content to identify regions for more detailed analysis. These intervals, which harbor potential functional (and by definition positional) candidate genes, may be selected for detailed study. Often, the first step in such efforts involves higher resolution genotyping to improve information content and refine genetic linkage data; this often has the effect of narrowing a candidate interval.

Identification of disease-related alleles within a targeted candidate region should be informed by at least two possible allele-disease risk paradigms. Common genetic markers (e.g. SNPs) may be employed to test for the presence of allelic association to detect the presence of common alleles. This may be done in an unbiased manner across the entire candidate interval, or within specific functional candidate genes, located within the candidate interval of linkage.

Alternatively, a variant screening-type approach may be used to test for possible heterogeneous disease-related variants. Such an approach is quite labor intensive and expensive, and in practice is generally reserved for a very small number of candidate loci. If more discreet genomic regions (in the case of a chromosomal abnormality) or specific genes (in the case of a functional candidate approach) are chosen, these methods permit a bypass of the genome-wide linkage process, by directly testing the hypothesis that a particular gene and/or functional candidate is involved in the disease.

Association studies test whether a common and specific allele is either present more often in cases than controls, or is more frequently transmitted to affected individuals within a family (i.e. exhibiting allelic transmission distortion). Thus, the two basic frameworks for association analysis are case-control and family-based studies. The case-control approach is generally regarded as more powerful and sensitive than family-based tests [56], although it can produce spurious associations due to population stratification [57]. Failure to match cases and controls for the same underlying population, ethnicity, and other factors, can lead to false positive or false negative results due to recent admixture and selection or drift between unlinked loci [58-60]. Historically, association tests have depended on the availability of numerous polymorphic markers in a given candidate region and the availability of abundant SNP markers makes this approach viable. It is estimated that SNPs occur on average every 500-1,000 base pairs and have a low mutation rate, both of which are advantageous in

association studies [61-64]. To avoid false negative results from an association study, it is important to perform a thorough analysis of linkage disequilibrium patterns and haplotype structures across a candidate locus or region to provide sufficient information content with which to detect and localize genetic effects. Without using multiple polymorphisms, any association study is likely to be vulnerable to type II error by failing to adequately cover the gene of interest [65, 66]. A current disadvantage of association studies is the need to identify a substantially narrowed genomic interval of interest or selected a candidate gene based on *a priori* knowledge. Recent technological advances will make the idea of genome-wide association studies a reality. However, as outlined in a recent review by Hirschhorn and Daly, there are multiple details that must be worked out before this can truly become practical [67]. These issues include but are not limited to; (1) the cost of such an experiment; (2) the criteria for choosing markers; (3) the limitation of false positive results.

Although, standard genetic analyses can detect genes of moderate to strong main effect, they have no ability to detect purely interactive effects. It is very likely that complex genetic disease in general and autism in particular will involve, at least in part, epistatic effects on disease risk. In fact, many researchers suggest that epistasis may be the rule in common complex disease and not the exception [68, 69]. This may be especially true given the lack of replication of single-locus results in association studies reviewed by Hirschhorn et al. [70]. In fact, Moore and Williams hypothesize that epistatic effects will prove

more important than independent main effects in complex disease [71]. Epistatic effects among multiple genes may play an important role in determining risk of autism, or broader phenotypes, and for some genes the interactive effects may be stronger than the independent single-gene effects. The Multifactor Dimensionality Reduction (MDR) method was developed by Moore and colleagues as a new approach to examine allelic interaction [72-74]. These kinds of new and innovative methods for examination of gene-gene interactions will play a key role in the future examination of complex disease.

A refinement of linkage and association strategies in genetic analysis of complex disease is to use trait-based phenotypic subsets and relevant statistical methodologies to identify susceptibility genes. These subsets can include endophenotypes, measurable components of the disease that become recognizable through detailed examination, and other intermediate quantitative traits [75]. Standard linkage and association analyses seek to identify genes of main effect and are typically employed using a global diagnosis for the disorder. Given the clinical and genetic heterogeneity in autism, I proposed to maximize use of available phenotypic data. Two recent reports describe trait subsets of autism derived from ADI information using a principal components analysis to identify clusters of ADI responses. This strategy has provided a means for performing genetic analyses using phenotypic subsets of autism. Our collaborator Dr. Susan Folstein and her group identified six distinct clusters of highly correlated variables from the Autism Diagnostic Interview (ADI): (1) spoken

language, (2) social intent, (3) compulsions and rigidity, (4) developmental milestones, (5) savant skills and (6) sensory aversions. Most of these are highly correlated between sibs and within affected sib-pair families, suggesting genetic relevance [76]. Calculation of scores for each of these sub-phenotypes would provide quantitative variables suitable for quantitative trait locus (QTL) analysis and quantitative trait association studies. Hauser and colleagues described an Ordered-Subsets Analysis (OSA) method that allows families to be ranked according to a family-specific (and genetically-relevant) covariate value. OSA starts with the (e.g.) highest or lowest ranked family, calculates an allele-sharing LOD score, subsequently adds families by rank and recalculates the LOD score with each addition until it defines the division in the dataset at which a maximum LOD score is reached [77]. These families, producing the maximum LOD score, are then defined as a subset for that covariate and can be used for subsequent analyses. I will describe application of these strategies for genome-wide and region/gene-specific studies.

Other endophenotypes in disease useful in homogenizing a sample include anatomical, biochemical, and neuropsychological measures. Several of these quantitative traits relevant to autism are reviewed by Veenstra-VanderWeele et al. and include: “level of intellectual functioning, degree of social or communication impairment, presence of seizure disorder, dysphmorphology, savant abilities, restrictive and repetitive behaviors”, and most notably “head circumference and whole blood serotonin” [9].

As mentioned by Weiss et al. another confounding effect of complex disease may be the presence of sex-specific genetic effects [78]. They suggest that failing to model for sex-specific architecture may limit the ability to detect "true" susceptibility loci in genome-wide analyses. Given that autism has roughly a 4:1 ratio of male to female-affected individuals, it may prove imperative to examine families with only male-affected individuals given that families containing female-affected individuals may have a different underlying landscape for genetic susceptibility. A recent study provides support that such a course may be warranted in autism [79].

Another innovation in the examination of candidate genes or regions potentially involved in complex disease involves applying the rapidly emerging availability of sequences from different genomes to identify regions of conservation that may harbor non-coding but regulatory functionality. A traditional bias in direct screening of candidate regions/genes, for rare mutations or to follow-up associated markers or haplotypes, is to focus on coding exons. Given a significant potential for complex disease-associated variants to affect expression, it is vitally important to represent potential regulatory sequences when designing genetic or variant screening strategies. If the alternative hypothesis of heterogeneous variation is involved in complex genetic disease, it is unlikely that coding variation will account for all disease alleles. It will be more likely that variation within conserved non-coding and/or promoter and enhancer regions of the genome will play a role in disease susceptibility. The web-based program

VISTA (Visualization Tools for Alignment) (www-gsd.lbl.gov/vista/) is a useful tool that allows identification of conserved, and therefore potentially functional, sequences. VISTA performs comparisons of sequences across species, and plots homology relationships as a function of nucleotide sequence identity. This is an important tool for identifying non-coding regions that should be targeted for genetic analysis and potentially by re-sequencing efforts to directly identify potential disease-associated variants.

Finally, many researchers seek to examine animal models to help explain the biology underlying complex diseases including autism [80]. In general, this approach can be fruitful in providing hints to gene and protein pathways involved. Unless a given gene is demonstrating extremely significant association with the human condition, and the nature of the underlying genetic alteration(s) that leads to disease risk is clearly understood, correlations between animal models and autism are necessarily weak. Even in such a scenario, there is always the caveat that the human disorder may be difficult or impossible to parallel in animal models.

With these available tools at our disposal, we must be mindful of the multiple complicating factors within complex genetic disease that make our endeavors of understanding complex genetic disease extremely difficult. A paper by Thornton-Wells et al. addresses numerous issues regarding the complexity of common disease and provides suggestion for methods and method development

to handle the expanding information required to elucidate a complex disease like autism [81].

CHAPTER II

HYPOTHESIS AND SPECIFIC AIMS

In this dissertation, I describe my genome-wide linkage studies of autism and traits comprising the aspects of the broader phenotype to identify autism susceptibility loci. I further document detailed molecular and genetic analyses of candidate genes in regions detected by linkage and specifically in 15q11-q13, identified on the basis of chromosomal duplications. Studies focus on genes acting within candidate neurobiological systems suspected of involvement in autism. Genetic analyses include construction of detailed linkage disequilibrium (LD) and corresponding haplotype maps across candidate loci and tests for transmission disequilibrium of single markers and haplotypes. Additionally, studies test for association to quantitative traits and potential allelic interactions. Molecular studies aim to identify functional variations on associated alleles or potential rare disease-related variants, and consider evolutionarily conserved sequence, in the absence of association. I hypothesize that there are allelic variants, which underlie genetic linkage and/or association to autism and related traits, and these contribute to autism susceptibility through both direct and interactive effects. Towards a genetic and phenotypic dissection of autism, I list the following aims for my dissertation:

Specific Aim I: To examine the serotonin transporter (*SLC6A4*) locus within the 17q11.2 autism candidate region

a. Identify and genotype single nucleotide polymorphisms (SNPs) across the *SLC6A4* locus in autism families

b. Characterize inter-marker linkage disequilibrium (LD) and corresponding haplotype structures

c. Test single markers and multi-marker haplotypes for evidence of allelic association

d. Screen functional/conserved sequences and/or associated allele(s) for evidence of disease-related variation

Specific Aim II: To examine the 15q11-q13 autism candidate region for involvement in inherited susceptibility to autism and related traits

a. Examine the $GABA_A$ subunit gene cluster for evidence of allelic association and disease-related sequence variation using the framework outlined in Specific Aim 1

b. Examine the Imprinting Center (IC) and Maternal Expression Domain (MED) for association to autism

Specific Aim III: Perform genome-wide linkage analyses of autism and individual traits representing autism sub-phenotypes

a. Perform a genomic linkage screen for autism using multiplex families

b. Perform a follow-up analysis of the most promising linked regions using SNPs to refine linkage data and perform limited candidate gene studies

CHAPTER III

GENETICS OF THE SEROTONIN TRANSPORTER (*SLC6A4*) LOCUS IN AUTISM[1]

Introduction

Evidence for Serotonin Involvement in Autism

Serotonin (5-hydroxytrptamine or 5-HT) has long been considered an attractive candidate system for involvement in autism etiology, given well replicated observations of elevated platelet serotonin in ~20-25% of probands and first-degree relatives [34-36, 82]. Given its role in 5-HT reuptake and presence in platelets, the serotonin transporter (*SLC6A4*) in particular has received significant attention as a logical candidate for this effect. Additionally, as the primary target for selective serotonin reuptake inhibitors (SSRIs) that are effective in treating behavioral symptoms including anxiety, rituals and aggression in autism and related disorders, the evidence for the serotonin transporter compounds [37, 38, 83-85]. Depleting tryptophan (an essential precursor for 5-HT biosynthesis) in adult subjects worsens autistic symptoms [86].

In addition to the biological and pharmacological findings providing links for serotonin involvement in autism, the emerging understanding of serotonin's

[1] Adapted from *Am J Med Genet B Neuropsychiatr Genet* 2004 May 15, 127B:104-12.

general role in behavior supports this premise. Serotonin is now known to affect sleep, mood, arousal, aggression, impulsivity, and affiliation, all of which are relevant to autism spectrum disorders. In fact, sleep disturbances have been observed in autistic individuals [87]. Genetic data from other neuropsychiatric conditions, some sharing behavioral traits with autism (mood, obsessive-compulsivity, anxiety, and social phobia), suggest involvement of multiple genes (e.g. MAOA [88], SERT [89, 90], 5-HT receptors [91-94], TPH1 [95], and TPH2 [96]) in the serotonergic pathway.

Previous Genetic Studies

Numerous association studies have been performed at this locus using two common variants: 5-HTTLPR, an insertion/deletion polymorphism in the promoter, and a variable number tandem repeat (VNTR) polymorphism located in intron 2. Both variants are proposed to exhibit allelic differences in transcription of *SLC6A4* message, although the data is more clear for 5HTTLPR [97-100]. The "long" (L) allele shows an approximately 2-fold level increased expression, compared to the "short" allele (S). Transmission disequilibrium test (TDT) studies in autism have failed to yield any consistent findings of allelic association with these markers in autism. This may in part be related to patterns of linkage disequilibrium, locus heterogeneity, allelic heterogeneity, or allelic interactions within or across loci [81]. Cook and colleagues reported preferential transmission of the short allele at 5-HTTLPR in autism [101], while others have reported

similar results for the long allele [102-104]. Other groups have failed to detect association [51, 105]. A single detailed study reported by the Cook lab describes high-density SNP genotyping across the *SLC6A4* locus and subsequent TDT analysis [106]. They replicate and extend upon previous findings, showing nominally significant association of a haplotype containing 5-HTTLPR and the VNTR, as well as the VNTR marker alone. Additionally, they identify significant linkage disequilibrium (LD) at other SNPs (near the 5′ end of the gene) in their dataset of autism families. One other group has recently found association with a number of 5′ single-nucleotide polymorphisms [107]. Recent work by both Stone et al. and Weiss et al. provide preliminary data suggesting that *SLC6A4* may be a male-specific susceptibility locus for autism and a sex-specific QTL for whole blood serotonin levels [78, 79]. The most recent evidence from Mulder et al. demonstrates association of the SERT intron two VNTR (12/12 genotype) with a rigid-compulsive domain of the autism phenotype [108].

Trait-Based Phenotypic Subsets

Growing evidence indicates that defining phenotypic subsets may improve power to detect genetic effects in complex disorders by identifying genetically more homogeneous sub-samples sharing common risk alleles[53, 54, 109-111]. To leverage phenotypic heterogeneity in autism for genetic studies, we need subsets defined by traits that (a) vary from one autistic person to the next; (b) are present, sometimes in milder form, in non-autistic family members much more

often than controls; and (c) aggregate in particular autism families. Towards this end, our collaborators completed a principal components analysis of items common to the Autism Diagnostic Interview (ADI) and its revision (ADI-R) and identified six phenotypic subsets, for which significant correlation of items within a given factor was observed [76]. These subsets are (1) spoken language, (2) social intent, (3) compulsions and rigidity, (4) developmental milestones, (5) savant skills and (6) sensory aversions. There was significant inter-sibling correlation in multiplex families for five of the six factor subsets, suggesting them to be genetically relevant. Social intent was not significantly correlated between sibs, but became so when age was covaried.

Study Design

We analyzed genotype data from an initial genomic screen of multiplex families for evidence of linkage using a categorical autism diagnosis. The analysis revealed suggestive linkage for autism to chromosome 17, and examination of the ADI-based phenotypic subsets revealed increased evidence supporting linkage in the families in which affected individuals are comparatively more affected for “rigid-compulsive” behaviors. Given that *SLC6A4* maps directly under the linkage peak, and the nature of the phenotypic subset demonstrating increased linkage, we examined this locus for evidence of allelic association. Our approach to this study included the use of markers previously shown by Kim et al. to associate with autism [106] in parent-child trio families, detailed

characterization of LD, and examination of haplotypes across the locus to test the hypothesis that common alleles at *SLC6A4* confer increased risk for development of autism and traits, such as rigid-compulsive behaviors, that are reflected in the broader phenotype.

Materials and Methods

Families

The sample for this initial linkage study consisted of 137 multiplex families. A total of 57 multiplex families were recruited from the Tufts/New England Medical Center (NEMC) and 80 affected sib-pair families were obtained from the AGRE consortium (www.agre.org). Association studies were conducted using a dataset of 123 multiplex families, all of which were included in the linkage study. All probands were at least four years of age and were clinically assessed with the ADI or ADI-R and most with the Autism Diagnostic Observation Schedule (ADOS). In multiplex families at least one sib had to meet ADI algorithm criteria for an autism diagnosis, while additional siblings may fall only one or two points short of meeting full criteria. Probands were excluded from the study if they had a known medical or neurological condition suspected to be associated with their autistic phenotype (e.g. fragile X syndrome). The procedures for clinically evaluating affected individuals for the AGRE families has been previously described [30].

Molecular Analyses

DNA was isolated from peripheral blood or lymphoblastoid cells using the PureGene kit according to the manufacturer's recommendations (Gentra Systems, Minneapolis, MN). 5-HTTLPR genotypes were determined using size discrimination of PCR products on 3% NuSeive (3:1) agarose (FMC Bioproducts; Rockland, ME) gels. The short allele corresponds to a product of 484 bp, while the longer allele is 528 bp; amplifying PCR primers have been described previously [101]. SNPs from the dbSNP and Celera SNP databases were selected based on their map position, minor allele frequency, and previous findings of allelic association to autism in trio families. Database reference numbers and other details for markers are cited in Table 3-1. PCR assays were developed and optimized to amplify an ~200 bp region flanking SNPs. Individual SNPs were genotyped by either fluorescent polarization template-directed dye terminator incorporation assay (FP-TDI) or TaqMan™. PCR primers and probes for assays are listed in Table 3-2. This information is unavailable for marker seven, which was obtained from Applied Biosystems (Foster City, CA) as an Assay-On-Demand™.

For FP-TDI genotyping, PCR reaction volumes were 8 μl, employing 10 ng genomic DNA template, 0.2 μM primers, 125 μM dNTPs and Applied Biosystems AmpliTaq Gold DNA polymerase and buffer (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 94 °C for 15 s, optimal annealing temperature (T_A°C) for 30 s, and 72

°C for 15 s, and a final extension at 72 °C for 10 min. FP-TDI analysis was performed using materials supplied in commercially-available Acycloprime™ kits according to the manufacturer's published protocols (Perkin-Elmer Lifesciences, Boston, MA) and as described elsewhere [112]. Samples were analyzed using a VICTOR$_2$™ multi-label plate reader instrument (Perkin-Elmer Life Sciences).

For TaqMan™ genotyping assays reactions were performed in a 5 μl volume according to manufacturer's recommendations (Applied Biosystems) [113]. Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 92 °C for 15 s, 60 °C for 1 min. Samples were analyzed using an ABI 7900HT™ Sequence Detection System.

Table 3-1. *SLC6A4* markers. Alleles are listed major/minor.

Marker No.	Marker Type	*SLC6A4* Region	dbSNP rs# /Celera hCV#	Alleles	Minor Allele Frequency	Intermarker Distance (kb)
1	ins/del	Promoter 1A	5-HTTLPR	528(l)/ 484(s)	0.45	–
2	SNP	Intron 1A	rs2066713/ hCV1841702	C/T	0.41	~12.5
3	SNP	Intron 1A	rs2020936/ hCV11414119	T/C	0.16	0.851
4	SNP	Intron 1A	rs2020937/ hCV1841703	T/A	0.41	0.057
5	SNP	Intron 2	rs2020942/ hCV1841709	G/A	0.4	3.843
6	SNP	Intron 5	rs140700/ hCV7473202	G/A	0.07	3.525
7	SNP	Intron 8	rs140701/ hCV7911143	T/C	0.43	4.857
8	SNP	Exon 14 3´ UTR	rs1042173	T/G	0.41	13.521

Table 3-2. *SLC6A4* PCR and genotyping primers. Labels are as follows: (F) Forward Primer, (R) Reverse Primer, (FP) FP-TDI extension primer, VIC and FAM are fluorescent labels for TaqMan™ allelic discrimination probes (Bold highlights SNP), AbD is an Assay-By-Design from ABI, AoD is an Assay-On-Demand from ABI. "Gel" means the product was run out on an agarose gel and genotyped through size discrimination. Sequence for primers and probes for AoD assays are proprietary information. MGB stands for minor groove binder, and NFQ stands for non-fluorescent quencher.

Marker No.	Primer sequences (5´-3´)		Product Size (bp)	T_A (°C)	Assay
1	F	GGCGTTGCCGCTCTGAATGC	484/ 528	63	Gel
	R	GAGGGACTGAGCTGGACAACCAC			
2	F	ACTGCTCACTGCTGCTGCTAAATG	127	58	FP
	R	GCATCACCCAAGCGTTCCC			
	FP-F	TTGCTTCTGAGATGGACCGCATTTCCCTTC			
3	F	GCCAGGCAGTAGCATAAATGGT	84	60	TaqMan (AbD)
	R	CAAACACCACTCAGAAGGATATGAA			
	VIC	VIC-AGAGCG**G**TCTCCATAA-MGB-NFQ			
	FAM	FAM-AAGAGCG**A**TCTCCA-MGB-NFQ			
4	F	CATATCCTTCTGAGTGGTGTTTGC	181	56	FP
	R	AATTTTAAAGGGATCGATTGTTGC			
	FP-F	TGTTTGCATTCTTGAGCCTGGGG			
5	F	AGGAAGGCCATCACGAGAACAC	187	58	FP
	R	CCTGCAGCCTGAGTTTTTAGCCTA			
	FP-F	AACACATGGTTTTATTCTCGAGCC			
6	F	TGCATAGTGGGCTCAGAGGTAGT	151	55	FP
	R	GGAGGTGGGTGAATGGATGTC			
	FP-F	TGATCTTTCTGCCACACCACCTC			
7	NA	Unavailable	NA	60	TaqMan (AoD)
8	F	GTAGGAGAGAACAGGGATGCTATC	153	52	FP
	R	CACACTATTTTTCATTTTAGCTTCTTACA			
	FP-R	AGGTTCTAGTAGATTCCAGCAATAAAATT			

Statistical Analyses

Genotype data from the following chromosome 17 microsatellite markers, located between 31cM and 90cM, was included in linkage analyses: D17S1852, D17S974, D17S1303, D17S969, D17S947, D17S799, D17S922, D17S900, D17S921, D17S839, D17S1857, D17S2196, D17S1850, D17S1871, D17S1824, D17S1294, D17S1800, D17S798, D17S1293, D17S1842, D17S933, D17S1867, D17S1299, D17S788, D17S957, and D17S916. The deCODE genetic map was

used to provide genetic distances for these markers [114]. For a few markers not present on the deCODE map, genetic map location was determined by relating the deCODE and Marshfield (http://research.marshfieldclinic.org/genetics/) maps. Microsatellite genotype data was analyzed following checks for quality control and Mendelian inconsistencies. HLOD values were calculated for the sample as a whole and for each subset under recessive and dominant models using GENEHUNTER-PLUS [115]. Disease allele frequencies were estimated to be q = 0.01 or 0.1 for dominant and recessive models, respectively. The phenocopy rate was 0.0005, and the penetrance value was set at 0.5 for the analysis. These simple models were chosen somewhat arbitrarily, simply making the disease allele frequency 10-fold greater under a recessive model than under the dominant model. The reduced penetrance value was set at 50% given the likelihood of oligogenic inheritance and the possibility of heterogeneity, such that not all individuals having the allele would present with disease. These parameters, while minimizing our power somewhat, have a smaller impact on the lod scores when there is either linkage or no linkage and were therefore selected to be robust [116].

Families were subdivided into two groups for each of the six factors, derived from items common to the ADI and ADI-R. For factors (including rigid-compulsive behaviors) whose scores had a unimodality of distribution across the overall dataset, the mean factor score was calculated for the two (sometimes three) probands, and families were split into two groups at the median of the

mean probands' scores for that factor. The two groups for each factor correspond to families either "positive" or relatively affected with higher scores on the phenotypic domain in question, or "negative" for less affected and lower scores for that domain. Linkage was analyzed in both positive and negative subsets for each factor as described above. A simulation study to calculate an empiric P-value for the rigid-compulsive linkage result was performed using SIMULATE (http://linkage.rockefeller.edu/ott/simulate.htm) to randomly draw, in each of 10,000 replicates, 58 (corresponding to the number of rigid-positive) families from the overall dataset. Linkage for all markers was calculated separately for this group and the remainder of families under both dominant and recessive models using GENEHUNTER-PLUS.

Genotype data from 5-HTTLPR and the seven SNPs were used in LD analysis. Initial analysis of SNP genotype data involved quality control checks consisting of verification of internal controls and assessment of Mendelian inconsistencies, followed by final haplotype consistency analyses using Simwalk2 [117]. Conformity with anticipated Hardy-Weinberg equilibrium expectations was established for all markers in the association study, and SNPs were examined for intermarker LD using both the r^2 and D´ measures calculated by the GOLD (Graphical Overview of Linkage Disequilibrium; [118]) software package. TD in autism families was determined using the PDT statistic, a variant of the transmission disequilibrium test (TDT), developed for use with general pedigrees [119]. Common haplotypes were determined using TRANSMIT [120],

and analysis of TD was performed using three SNPs (5-HTT-5, 5-HTT-6 and 5-HTT-7) that define all common (≥5%) haplotypes. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$.

Visualizations Tools for Alignment (VISTA) analysis was performed via web-based submission (http://www.gsd.lbl.gov/vista/) of human and mouse genomic sequence [121].

Results

To facilitate the identification of genes underlying genetic risk for autism, we performed a 10 cM genomic screen of multiplex autism families. While we present the initial details of that study as it relates to chromosome 17 here, the completed screen is presented in its entirety in Chapter VIII. We detected suggestive evidence for linkage on chromosome 17 (Figure 3-1), with a maximum multipoint HLOD of 2.74 at D17S1871 (~50cM) under a dominant model in our dataset of families (n = 137). In an exploratory analysis, we used the six ADI factors as a basis for subsetting our families for linkage analysis. We detected a substantial increase in linkage in the subset of families in which probands are relatively affected (positive) by rigid-compulsive behaviors (Figure 3-1). In the rigid-positive families (n = 58), the dominant HLOD increased to 3.62 at D17S1294 (~53cM), adjacent to D17S1871 in the genomic screen panel. A permutation test showed that linkage in the rigid-compulsive families was significant (P = 0.025). Thus, if the rigid-compulsive subsetting scheme had

nothing to do with the linkage signal on chromosome 17q, then the probability of our observing an HLOD ≥ 3.6 by chance would be ≤ 0.025.

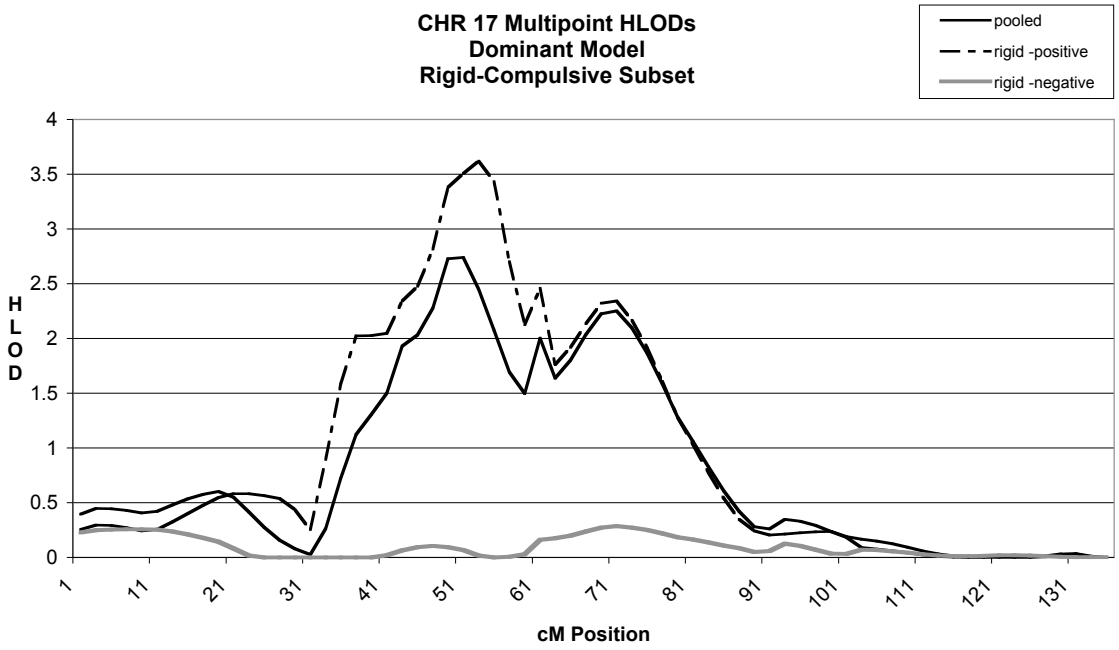


Figure 3-1. Linkage analysis of autism for chromosome 17. Dominant Multipoint HLOD scores are plotted for the overall (pooled) autism dataset (n=137), the rigid-positive (relatively affected; n=58), and the rigid-negative (relatively unaffected; n=79). The overall dataset has a peak multipoint HLOD of 2.74 at D17S1871, and this increase to 3.62 in the rigid-positive subset.

We noted that the serotonin transporter (*SLC6A4*) locus maps adjacent to D17S1294 (~140 kb). Given the long-standing hypothesis that 5-HTT is involved in serotonin-related autism etiology, we directly tested whether common alleles at this gene were associated with autism in our dataset. 5-HTTLPR and seven SNPs were initially selected as markers for this study providing an average marker density of ~5kb across the transcriptional unit (see Figure 3-2 and Table

3-1). Criteria for marker selection included previous evidence for association [106], high minor allele frequencies, and uniform spacing across the region. Markers were genotyped by PCR and size discrimination (5-HTTLPR), FP-TDI or TaqMan™ in a dataset of 123 multiplex families that represent a subset of the 137 families analyzed for linkage.

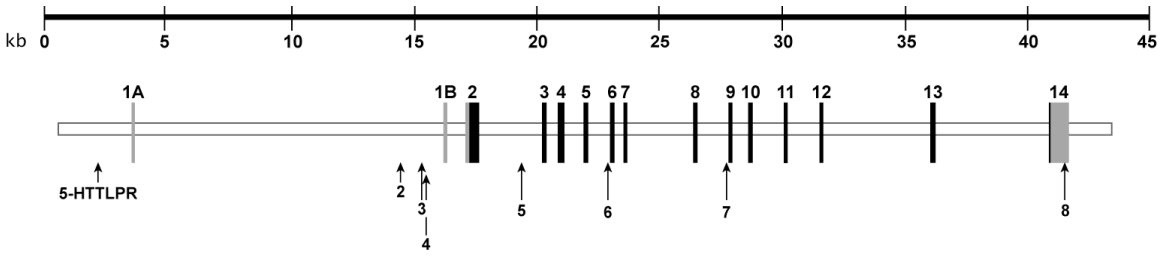


Figure 3-2. Schematic representation of *SLC6A4*. The 38kb transcriptional unit contains 15 total exons represented by the vertical boxes. Coding regions are indicated in black with the untranslated regions in gray. Markers are labeled and their positions indicated by arrows.

Genotypes at individual markers did not deviate from expectations of Hardy-Weinberg equilibrium (Data not shown). To discover all common alleles at this locus, we characterized intermarker LD to permit identification of haplotype blocks and corresponding haplotypes. Intermarker LD was assessed using GOLD, and markers 2-8 were found to be in strong LD with one another (Figure 3-3). By contrast 5-HTTLPR was in relatively weak LD ($D' < 0.5$) with these markers. These data suggest that at least two distinct haplotype blocks span the promoter and transcriptional unit of the *SLC6A4* locus.

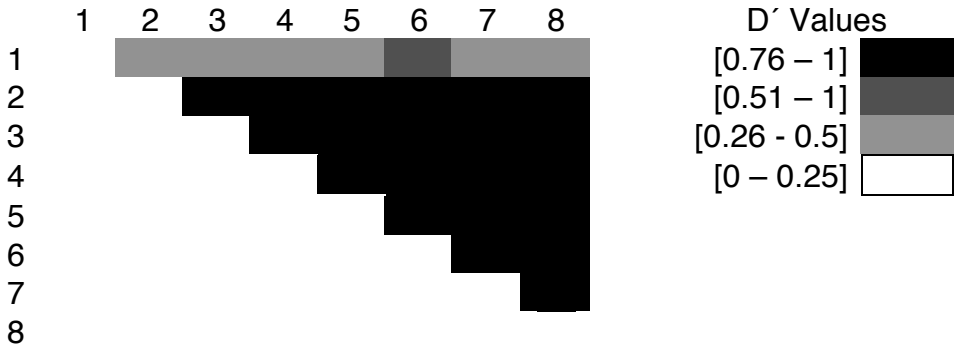


Figure 3-3. Intermarker linkage disequilibrium (LD) at *SLC6A4*. LD was measured for all markers using GOLD. D´ values are represented by shaded boxes to display intermarker LD relationships. The major haplotype block includes markers 2-8. 5-HTTLPR, for which nominal association is seen, is only in weak LD (D´< 0.5) with other markers and is not located in the main haplotype block.

Individual marker association analysis for both the overall and rigid-compulsive subset was tested using PDT. Nominal association ($P \leq 0.05$) was seen at 5-HTTLPR ($P = 0.01$) and 5-HTT-6 ($P = 0.02$) in the overall dataset ($n = 123$). We observed an over-transmission to affected individuals of the short allele at 5-HTTLPR (Table 3-3). None of these markers demonstrated evidence of association in the smaller rigid-compulsive subset.

Table 3-3. PDT analysis of the *SLC6A4* locus

Marker No.	Overall PDT (P)	Allele	# of Alleles Transmitted	# of Alleles Not Transmitted	Rigid Subset PDT (P)
1	**0.01**	484(s) 528(l)	221 237	180 278	0.07
2	0.46	T C	175 291	188 278	0.37
3	0.29	C T	77 387	89 375	0.82
4	0.5	A T	177 287	189 275	0.50
5	0.63	A G	175 291	184 282	0.88
6	**0.02**	A G	29 435	49 415	0.10
7	0.22	C T	205 253	184 274	0.38
8	0.29	G T	202 260	185 277	0.53

Haplotype frequencies and transmissions were assessed using TRANSMIT for the haplotype block defined by markers 2-8. We identified four haplotypes having estimated allele frequencies >5%. These haplotypes represent ~98% of the haplotypes in our sample. The common haplotypes may be differentiated by genotyping only three of the seven SNPs in the block. Using SNPs 5-HTT-5, 5-HTT-6, and 5-HTT-7 to represent these haplotypes, association analysis was performed on the overall dataset using TRANSMIT (Table 3-4). No significant evidence for association was detected with this combination of markers. Given the absence of association in the full autism dataset, we did not test the rigid-compulsive subset.

Table 3-4. TRANSMIT haplotype analysis for markers 5-6-7 in *SLC6A4*. Gray denotes htSNPs.

2-3-4-5-6-7-8 Haplotype	Frequency	Observed Transmissions	Expected Transmissions	χ^2	Global χ^2	P
C - T - T - G - G - C - G	0.43	215.5	203.0	2.23		
C - C - T - G - A - T - T	0.08	31.0	41.0	5.75		
T - T - A - A - G - T - T	0.37	175.0	178.2	0.12		
C - C - T - G - G - T - T	0.10	53.5	53.7	0.01		
Overall TRANSMIT analysis					8.00	0.09

VISTA analysis of human and mouse genomic sequence for *SLC6A4* revealed evolutionarily conserved coding and non-coding regions (Figure 3-4). Including 20kb of sequence both upstream of the transcriptional start site and downstream of the 3´ untranslated region (UTR), we used the web-based VISTA program (http://www.gsd.lbl.gov/vista/) to compare the two sequences. Figure 3-4 highlights the relative positions of SNP markers and the promoter variant on the VISTA plot, which shows the location of exons and evolutionarily conserved sequence. In addition to coding sequences, several regions of non-coding conservation are detected both inside and outside of the main haplotype block.

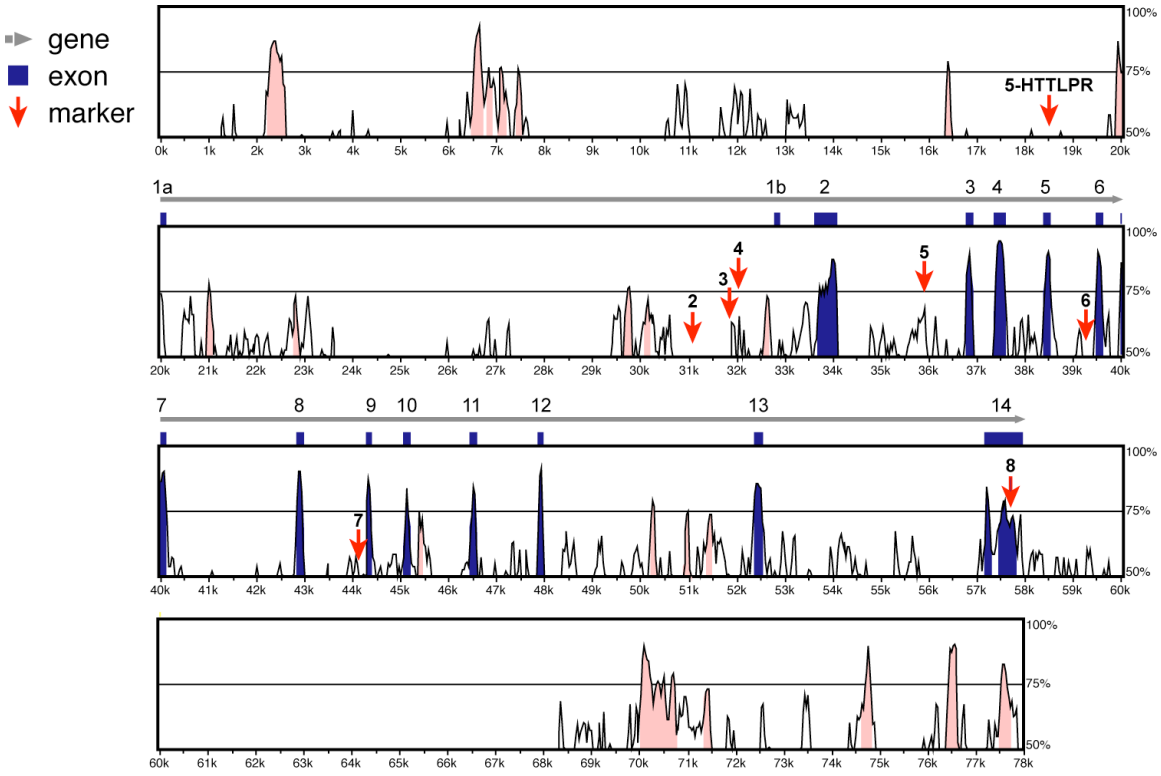


Figure 3-4. Evolutionary conservation at *SLC6A4*. Output from VISTA analysis of the *SLC6A4* transcriptional unit (indicated by gray arrow above the plot) is shown, with regions of non-coding sequence conservation (>75% identity) highlighted by pink shading and coding homology by blue shading. Red vertical arrows indicate the position of markers.

Discussion

It is well established that autism is one of the most genetic of neuropsychiatric disorders, and that multiple genes are likely involved in its etiology. Two complementary approaches for identifying genes underlying complex genetic disorders are (1) identification of genomic regions more frequently inherited in common by affected individuals in a family through genomic linkage screens and (2) analysis of genes in candidate systems suspected to be involved in the disease based on altered biology. Linkage-

oriented efforts to identify susceptibility genes tend to evolve into studies of positional candidates, once a region of linkage is identified. These two approaches are unified in this study, where an excellent functional candidate gene is found in a very strong region of linkage in our dataset.

Evaluation of a candidate gene must ultimately be able to detect the presence of either common or heterogeneous disease-associated variants. Allelic association studies permit identification of *common* variants at a disease susceptibility locus. However, the absence of association does not exclude the possibility of heterogeneous variants or rare mutations as underlying increased risk associated with a gene. Therefore, a thorough examination of a candidate gene must supplement allelic association analysis with direct screening to identify sequence variants. In this report, we identify *SLC6A4* as a positional and functional candidate in a region of linkage on chromosome 17. The potential for this gene as a candidate is strengthened by (1) the nature of the phenotypic subset in which we observe increased linkage and (2) the efficacy of SSRIs in treating aspects of the autism phenotype such as those present in this subset and related phenotypes (e.g. obsessive-compulsive disorder). We did not detect association in the main haplotype block corresponding to SNPs 2-8, although we saw nominal but not highly significant association at 5-HTTLPR, 5´ to this block. Thus, we conclude that a common disease-associated variant does not exist in this block. Future studies will examine the possibility of heterogeneous and/or rare variants in the main haplotype block in our dataset. Additionally, regions 5´

or 3´ to this block will be tested for the presence of either common or heterogeneous variants.

These results must be interpreted in the light of the significant association to autism in parent-child trio families reported by Cook and colleagues [106]. The comparison of theses studies is aided by a commonality of genetic markers but is complicated by the difference in family structure analyzed. With the exception of 5-HTTLPR, markers demonstrating association in the Kim et al. report do not show a similar effect in our dataset [106]. The possibility exists of biases towards different genetic mechanisms in multiplex families compared to singleton cases. These populations may harbor overlapping but significantly different constellations of autism susceptibility alleles. Considering the serotonin transporter in particular, it would be interesting to know whether affected individuals in multiplex and trio families differ in response to SSRIs. The number of transmissions analyzed in this study is slightly larger than that tested by Kim et al., so the current study is not limited by power. Ultimately, such detailed studies in independent datasets will be required to make a determination of how or if these data may be generalized across autism datasets.

Given the genetic heterogeneity in autism and other complex disorders, the need for approaches to identify genetically more homogeneous sub-samples is becoming increasingly apparent. For example, an alternative to the use of a categorical autism diagnosis to define affection status is the application of individual traits comprising the broader phenotype. There are a variety of ways in

which phenotype-based subsetting can be applied to genetic studies. In this report, we select families that are relatively affected for the previously defined ADI factors, and we use them for our analyses. We observed increased linkage in a rigid-compulsive subset, and a simulation study suggests this result may be significant (P = 0.025). However, this simulation does not account for testing multiple subsets, and Bonferroni correction would render this result non-significant. While we believe it is likely that these subsets are genetically relevant and will thus provide increased power to detect risk alleles, we recognize that splitting our dataset in this manner may reduce the overall power to detect effects due to decreasing the sample size.

Single marker association studies can be powerful in their ability to detect genetic effects. However, it is unlikely that a given polymorphic marker will be a disease-susceptibility variant. Rather, a positive finding of association probably indicates linkage disequilibrium to a nearby variant. We and others argue that characterization of all common haplotypes at a locus is important for meaningful association analysis [66]. The absence of strong positive findings may simply be the result of failing to test the relevant allele or the potential for allelic heterogeneity. Potential susceptibility alleles may exert their effect through perturbations in gene expression, and relevant sequences controlling gene expression may lay distant 5′ or 3′ to the transcriptional unit. This point highlights the utility of identifying conserved sequences around a gene locus under study,

so that potential functional sequences may be included in efforts to detect disease-associated variants.

CHAPTER IV

FOLLOW-UP ANALYSIS OF THE *SLC6A4* LOCUS

Introduction

Our initial analysis of the serotonin transporter lacked strong support for direct involvement of this gene in autism susceptibility. Although we saw nominally significant association with two of our eight markers, it was clear that these results were not explaining the strong evidence for linkage to this region seen in our dataset. Given increasing evidence for linkage by our and other groups and the significant biological rationale for involvement of this molecule, we pursued further examination of the transporter locus. While preliminary association studies well represented most of the locus in terms of LD patterns, the 5' end (e.g. around 5-HTTLPR and farther upstream in the promoter) was not covered. Thus a common allele at an upstream regulatory element, for example, would not have been detected. Additionally, some consideration to the alternative to the "*common disease-common variant*" hypothesis deserved consideration. To address these scenarios, additional studies of this locus were undertaken. They included (1) testing whether known, but otherwise extremely rare non-synonymous variants might be increased in frequency in autism families, (2) screening exons and regions showing high sequence identity between human and mouse, and (3) association analysis of markers previously demonstrating a

nominal level of significance in a larger dataset. By addressing these issues, we sought to further explore the involvement of this gene and the possibility of allelic heterogeneity at this locus.

The absence of highly significant allelic association at this locus led us to test the alternate hypothesis that otherwise rare coding variation may be enriched in the autism population. In testing a similar hypothesis, Fearnhead et al. present evidence suggesting that many rare variants collectively contribute to the inherited susceptibility to colorectal adenomas [122]. Glatt et al. reports the most detailed study in a non-selected (i.e. “normal”) Caucasian population and reported the presence of nine rare non-synonymous variants at the SERT locus, with most present only once in 900 chromosomes (minor allele frequency = 0.0011). The Gly56Ala variant was present on 4/900 chromosomes and corresponds to a 56Ala-encoding allele frequency of 0.0044 [123]. Another variant, Ile425Val, originally described by Glatt was recently rediscovered by Ozaki and colleagues in two unrelated families segregating a complex psychiatric phenotype containing Asperger’s syndrome, obsessive-compulsive disorder (OCD), and other co-morbid phenotypes [90]. Given the relationship of the phenotypes, particularly Asperger’s syndrome to autism, this provides yet further suggestion for the role for rare coding variation at *SLC6A4* in autism. In our initial rigid-compulsive subset we did not find increased evidence for association but several rare coding variants, including Ile425Val, might be present in this population of families. These families could in turn be those demonstrating

increased allele-sharing as detected by our linkage study. We hypothesize that one or more of these rare non-synonymous variants, multiple of which have been shown by the Blakely laboratory at Vanderbilt (manuscript in preparation) to exert functional effects on SERT activity, may be over-represented in the autism population.

Another hypothesis we aim to test is the influence of common alleles within evolutionarily conserved regions across the *SLC6A4* locus. Regions demonstrating conservation may harbor functional, possibly regulatory, sequences. Identifying sequence changes in these regions could uncover functionally significant variation.

Serotonin levels have shown to be sexually dimorphic in previous studies [124-127]. Recent evidence of sex-specific risk alleles in autism along with evidence of sex-specific whole blood serotonin levels lend more evidence for the role of serotonin and the serotonin transporter specifically in autism. Stone et al. completed a genome-wide linkage screen subdividing their dataset based on the sex of the affected child. They report a major male-specific linkage peak at 53cM on chromosome 17, according to the Marshfield map. The location of one of the markers (D17S1294) flanking this peak maps only 142 kb away from the *SLC6A4* locus [79]. Weiss et al. have performed a QTL analysis of whole blood serotonin levels demonstrating evidence for a male-specific QTL on 17q at the *ITGB3* locus (~16cM distal to the *SLC6A4* locus) in the Hutterite population [78]. Additionally, they have seen similar male-specific results near the serotonin transporter gene.

We hypothesize that association to autism at the *SLC6A4* locus is male-specific or at least provides for increased risk in autistic males. In addition, one can imagine that families containing autistic females have a different genetic architecture (therefore a potentially different set of susceptibility loci) than families containing only male-affected individuals. To test this hypothesis we plan on examining association within the subset of families containing only male-affected individuals.

Materials and Methods

Families

There are several subsets of samples used in these studies. One sample subset is the set of 23 affected but unrelated individuals determined to be the most "rigid" based on the "rigid-compulsive" factor score calculated by the algorithm obtained through use of a principle components analysis on the ADI-R (3 AGRE, 20 TUFTS) [76]. A second subset is the set of 24 affected but unrelated individuals (19 AGRE, 5 TUFTS) chosen from 24 strongly linked multiplex families (as determined from family-based multipoint LOD scores calculated in families demonstrating strongest linkage to the chromosome 17 peak using ALLEGRO). Both of these subsets were used for screening of exons and regions of conserved sequence across the *SLC6A4* locus. The larger dataset consisting of 384 families (283 AGRE, 98 TUFTS, 3 VANDY) was used for

screening of *known* rare non-synonymous variants, for a replication attempt of the nominal associated markers previously shown in Chapter III, and for examining SNPs detected through re-sequencing efforts. AGRE samples were obtained from the NIMH Center for Collaborative Genetic Studies on Mental Disorders (http://nimhgenetics.org/). All probands were at least four years of age and were clinically assessed with the ADI or ADI-R and most with the Autism Diagnostic Observation Schedule (ADOS). In multiplex families at least one sib had to meet ADI algorithm criteria for an autism diagnosis, while additional siblings need not meet full criteria. Probands were excluded from the study if they had a known medical or neurological condition suspected to be associated with their autistic phenotype (e.g. fragile X syndrome). The procedures for clinically evaluating affected individuals for the AGRE families has been previously described [30].

Molecular Analyses

DNA was isolated from peripheral blood or lymphoblastoid cells using the PureGene kit according to the manufacturer's recommendations (Gentra Systems, Minneapolis, MN). 5-HTTLPR and the intron 2 VNTR genotypes were determined using size discrimination of PCR products on 3% NuSeive (3:1) agarose (FMC Bioproducts; Rockland, ME) gels. Amplifying PCR primers for these two markers have been described previously described [101]. For the 5-HTTLPR genotyping, the short allele corresponds to a product of 484 bp, while

the longer allele is 528 bp. PCR reaction volumes were 10 μl, employing 12.5ng genomic DNA template, 0.22 μl at 10mM of each primer, 1 μl at 2.5mM dNTPs, 0.08 μl Applied Biosystems AmpliTaq Gold DNA polymerase and buffer (Applied Biosystems), and a GC-rich Kit purchased through Roche. For the VNTR genotyping, products amplified were 345bp (9 copy repeat), 360bp (10 copy repeat), and 390bp (12 copy repeat). PCR reaction volumes were 10 μl, employing 5ng genomic DNA template, 0.22 μl at 10mM of each primer, 1 μl at 2.5mM dNTPs, and Applied Biosystems AmpliTaq Gold DNA polymerase and buffer (Applied Biosystems). SNPs were selected based on previous findings of allelic association in our initial report, demonstration of association to autism through other previous studies, or by discovery through sequencing of exons or conserved regions in the lab [128]. In addition, 10 known rare non-synonymous variants were also selected for examination. Database reference numbers (both dbSNP and Celera IDs when available) and other details for all markers are cited in Table 4-1. Individual SNPs were genotyped through the TaqMan™ (Applied Biosystems) protocol using either the Assays-On-Demand or Assays-By-Design products. PCR primers and probes for all assays are listed in Table 4-2. This information is unavailable for any markers obtained from Applied Biosystems (Foster City, CA) as Assays-On-Demand™. When the assay permitted (compatible primers were able to be designed within exonic sequence) internal site-directed mutant and wild-type cDNA clones (obtained from the laboratory of Dr. Randy Blakely) for the rare allele of the non-synonymous variants were used

as homozygous (mutant cDNA) and heterozygous (made by using 50/50 mix of mutant cDNA with wildtype cDNA) controls to make unambiguous genotype when employing the TaqMan™ assays.

For TaqMan™ genotyping assays [113], reactions were performed in a 5 μl volume according to manufacturer's recommendations (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 92 °C for 15 s, 60 °C for 1 min. Samples were analyzed using an ABI 7900HT™ Sequence Detection System.

Variant screening was also performed on both exonic sequence and non-coding evolutionarily conserved regions across the *SLC6A4* locus. Our laboratory screened the conserved non-coding regions through direct sequencing efforts. Our collaborators in the Blakely laboratory employed the REVEAL™ System (Spectrumedix LLC, State College, PA) for variant screening, followed by sequencing confirmation of samples believed to contain real variation. Samples for re-sequencing were sent to the Vanderbilt Sequencing Core Facility after initial PCR reactions were performed in the lab. PCR reaction volumes were 8 μl, employing 10 ng genomic DNA template, 0.2 mM primers, 125 mM dNTPs and Applied Biosystems AmpliTaq Gold DNA polymerase and buffer (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 s, optimal annealing temperature (T_A°C) for 30 s, and 72 °C for 1 min, and a final extension at 72 °C for 10 min. PCR reaction cleanup was performed using Exo-Sap-It (USB) and then sent to

the Vanderbilt Sequencing Core for sequencing using the Big-Dye terminator system. Larger regions were amplified from both directions to get complete sequence information for the region of interest. Some sequencing reactions were carried out with universal primers (M13F (-21) 5′-TGTAAAACGACGGCCAGT-3′ and M13R 5′- CAGGAAACAGCTATGAC-3′). Sequencing primers for the regions of non-coding evolutionary "conservation" are shown in Table 4-3 along with primers used by the Blakely laboratory for screening of exonic sequence.

Table 4-1. *SLC6A4* follow-up markers. Alleles are listed major/minor.

Marker No.	Marker Type	SLC6A4 Region	dbSNP rs#/ Celera hCV#	Alleles	Minor Allele Frequency	Intermarker Distance (bp)
1	SNP	Promoter 1A	rs1050565/ hCV7473213	T/C	0.32	10,035
2	SNP	Promoter 1A	rs2020930/ hCV11424041	G/A	0.03	1,734
3	ins/del	Promoter 1A	5-HTTLPR	528(l)/ 484(s)	0.44	2,549
4	SNP	Intron 1A	rs2020933/ hCV11424045	A/T	0.07	295
5	SNP	Intron 1A	rs2020934/ hCV7911197	C/T	0.47	5
6	SNP	Intron 1A	rs2020935/ hCV11424046	T/A	0.07	11,477
7	SNP	Intron 1A	rs25528/ hCV1841705	A/C	0.16	80
8	SNP	Exon 1B	rs6354/ hCV1841706	A/C	0.17	931
9	SNP	Exon 2	Thr4Ala	A/G	_	157
10	SNP	Exon 2	Gly56Ala	G/C	0.01	310
11	VNTR	Intron 2	VNTR	9/ 10/ 12	0.02/ 0.39/ 0.59	3,269
12	SNP	Exon 4	Lys201Asn	G/C	_	40
13	SNP	Exon 4	Glu215Lys	G/A	_	1,802
14	SNP	Intron 5	rs140700/ hCV7473202	G/A	0.07	195
15	SNP	Exon 6	Ser293Phe	C/T	_	505
16	SNP	Exon 7	Pro339Leu	C/T	_	2,811
17	SNP	Exon 8	Leu362Met	C/A	_	1,504
18	SNP	Exon 9	Ile425Val	A/G	_	8,181
19	SNP	Exon 13	Lys605Asn	A/C	_	4,687
20	SNP	Exon 14	Pro621Ser	C/T	_	_

Table 4-2. Primers and probes for variants in follow-up study of *SLC6A4*. Labels are as follows: (F) Forward Primer, (R) Reverse Primer, VIC and FAM are fluorescent labels for TaqMan™ allelic discrimination probes (SNPs highlighted by bold text), AbD is an Assay-By-Design from ABI, AoD is an Assay-On-Demand from ABI. "Gel" means that the product was run out on an agarose gel and genotyped through size discrimination. Sequence for primers and probes for AoD assays is proprietary information. MGB stands for minor groove binder, and NFQ stands for non-fluorescent quencher.

Marker No.	Primer sequences (5´-3´)		Product Size (bp)	T_A (°C)	Assay
1	NA	Unavailable	NA	60	TaqMan (AoD)
2	F	GCTCAAGCAGGTGAACAAAGAAA	72	60	TaqMan (AbD)
	R	CTGGGCAGCTGGGAAGAG			
	VIC	VIC-AACTATTGCT**A**TGCGGTGAT-MGB-NFQ			
	FAM	FAM-TTGCT**G**TGCGGTGAT-MGB-NFQ			
3	F	GGCGTTGCCGCTCTGAATGC	484/ 528	61	Gel
	R	GAGGGACTGAGCTGGACAACCAC			
4	F	TGTATGTATTTTTACCATCAGTTTTGTCCAGAA	81	60	TaqMan (AbD)
	R	GAGAGTTAGCTAGCAGGCTCATAAAT			
	VIC	VIC-CATTGACC**A**GGTTCAC-MGB-NFQ			
	FAM	FAM-CATTGACC**T**GGTTCAC-MGB-NFQ			
5	F	TTTTCCTGCCACGCACTCT	80	60	TaqMan (AbD)
	R	GCACAAACCTCATAAGAACCTGCTT			
	VIC	VIC-ACCGTTCCAA**T**ATGG-MGB-NFQ			
	FAM	FAM-CCGTTCCAA**C**ATGG-MBG-NFQ			
6	F	TGGCAGTGACCGTTCCAA	68	60	TaqMan (AbD)
	R	TTGCTCAATTTGCACAAACCTCAT			
	VIC	VIC-CTGCTTCTCACTCA**T**CCA-MGB-NFQ			
	FAM	FAM-TGCTTCTCACTCA**A**CCA-MGB-NFQ			
7	F	CCCAGTGGAGGCACAGG	62	60	TaqMan (AbD)
	R	GAGTGTGCAGGTTACTGATGCT			
	VIC	VIC-TGGTTGGT**G**TCGCCG-MGB-NFQ			
	FAM	FAM-TGGTTGGT**T**TCGCCG-MGB-NFQ			
8	F	GGAGGCAAGGCGACCTT	58	60	TaqMan (AbD)
	R	CTGTGGCTAAGCCCCTTGTTATT			
	VIC	VIC-CTTGCCCTCT**A**TTGCAG-MGB-NFQ			
	FAM	FAM-TTGCCCTCT**C**TTGCAG-MGB-NFQ			
9	F	GTCATTTACTAACCAGCAGGATGGA	62	60	TaqMan (AbD)
	R	CGCTGATAGCTGCTTCTGAGA			
	VIC	VIC-ATTCAAGGGCG**T**CGTC-MGB-NFQ			
	FAM	FAM-CAAGGGCG**C**CGTC-MGB-NFQ			
10	F	GGGTACTCAGCAGTTCCAAGTC	56	60	TaqMan (AbD)
	R	GGGATAGAGTGCCGTGTGT			
	VIC	VIC-CTGGTGCGG**G**AGAT-MGB-NFQ			
	FAM	FAM-CTGGTGCGG**C**AGAT-MGB-NFQ			

11	F R	TGGATTTCCTTCTCTCAGTGATTGG TCATGTTCCTAGTCTTACGCCAGTG	345/ 360/ 390	57	Gel
12	F R VIC FAM	GCTATACTACCTCATCTCCTCCTTCAC TGGTGCAGTTGCCAGTGTT VIC-CCAGGAGTT**C**TTGCAGC-MGB-NFQ FAM-CAGGAGTT**G**TTGCAGC-MGB-NFQ	83	60	TaqMan (AbD)
13	F R VIC FAM	TCCTGGAACACTGGCAACTG GAATGGAGGGTCCAGGTGATG VIC-AATTACTTCTCC**G**AGGACAA-MGB-NFQ FAM-AATTACTTCTCC**A**AGGACAA-MGB-NFQ	65	60	TaqMan (AbD)
14	F R VIC FAM	ACTCCAAGGGTTGTGATCTTTCTG GGGTGAATGGATGTCAGTGTCTTTT VIC-ACCACCTC**A**CCCTCCT-MGB-NFQ FAM-CACCTC**G**CCCTCCT-MGB-NFQ	89	60	TaqMan (AbD)
15	F R VIC FAM	GTGACAGCCACCTTCCCTTATATC GTGGCACCCCTCACCAG VIC-CAGGACA**G**AAAGGAT-MGB-NFQ FAM-CAGGACA**A**AAAGGAT-MGB-NFQ	56	60	TaqMan (AbD)
16	F R VIC FAM	AGCCGCTCAGATCTTCTTCTCT TTGAACTTGTTGTAGCTAGCAAAAGC VIC-CAAAGCCC**G**GACCAA-MGB-NFQ FAM-CAAAGCCC**A**GACCAA-MGB-NFQ	75	60	TaqMan (AbD)
17	F R VIC FAM	CGGCCCCTTGGGTTTTC GAAGCTCGTCATGCAGTTCAC VIC-CCAGAGATGCC**C**TGGTGA-MGB-NFQ FAM-CAGAGATGCC**A**TGGTGA-MGB-NFQ	68	60	TaqMan (AbD)
18	F R VIC FAM	GCAGAAGCGATAGCCAACATG CAAGCCCAGCGTGATTAACATC VIC-TTTCTTTGCC**A**TCATCTT-MGB-NFQ VIC-TCTTTGCC**G**TCATCTT-MGB-NFQ	78	60	TaqMan (AbD)
19	F R VIC FAM	TCCCCACATATATAGCTTATCGGTTGA CAAAACAATTAGTAGTCTGAACACACACA VIC-CACGTACCTC**T**TTAAAT-MGB-NFQ FAM-ACGTACCTC**G**TTAAAT-MGB-NFQ	105	60	TaqMan (AbD)
20	F R VIC FAM	GCGTATTATTAAAAGTATTACCCCAGAAACAC CACAGCATTCAAGCGGATGTC VIC-CACAAG**G**AATTTCT-MGB-NFQ FAM-CACAAG**A**AATTTCT-MGB-NFQ	73	60	TaqMan (AbD)

Table 4-3. Variant screening primers for the *SLC6A4* locus. Labels are as follows: (F) Forward Primer, (R) Reverse Primer, T_A (annealing temperature). *Sequencing reactions were carried out with universal primers (M13F(-21) 5′-TGTAAAACGACGGCCAGT-3′ and M13R 5′- CAGGAAACAGCTATGAC-3′).

Gene	Conserved Non-coding Region	Primer sequences (5´-3´)		Product Size (bp)	T_A (°C)
SLC6A4	1	F	GTAGAGATGAGCCCAGGGTTCACAGT	540	56
		R	GCAATCACCGCACAGCAATAGTTT		
	2	F*	GGTCAAGAGAAAGCGGCACGAGCAGA	762	66
		R*	GCCCACTTCGAGCACTCCACGTTCCT		
	3	F*	ACTGTGCTACGTGGTTGAAGGATATA	654	56
		R*	AAGAAAAGAGGAAAAACCCATGC		
	4	F*	CATACGTGGGTGTGGAGCGAAAC	730	59
		R*	CATCACTGGGTTGCGTCTCTCCTT		
	5	F	TTGGCTCTATGACCTGTAACT	897	54
		R	TTTCTGCCTGCAAGCTC		
	6	F	GTCCAGTGTCATCTCAGCTAGGCA	382	59
		R	AGAGGGCAAGCAAGGTCGC		
	Exon	Primer sequences (5´-3´)		Product Size (bp)	T_A (°C)
	1a	F	GGCTATCTAGAGATCAGACCATGTG	280	66
		R	CTGGGGCGCATGCACCTCCT		
	1b	F	CCCAGCATCAGTAACCTGCAC	207	61
		R	TCTGCACATGGCCTTTCTGCGT		
	2	F	TTTCTCCTTCCTCTGTGTGTC	302	54
		R	CTGCTGAGTACCCATTGGATA		
	2	F	CGGAGTTCTACAGAAGGTTGTT	308	61
		R	GGGTCACAGCCACTACTCGCA		
	3	F	CTGTGATTCTCAGTCACAGATTGG	295	61
		R	GTCGCCAGCTGCCAGCCACTGATG		
	4	F	GGGCTGCGGGGCTCTCCAGTG	420	66
		R	CAGAAAGGGGTGAGGAGCCCTTGG		
	5	F	CCCGGGCACCCTCAAAGGAGC	315	67
		R	GTGCTGTCTGTCCAGGCTACT		
	6	F	CTCCCTGGAACAGCATGGTGA	212	59
		R	TGACAGACAGGTACACATATTTCCC		
	7	F	GCCTCTGCACTTAGCCACATGG	332	66
		R	GCAGTGGTATTAAGGCCTAAGCC		
	8	F	CCCTGATCTTGGAACTGTCTC	248	61
		R	GATCTTTACAAAGATTCAAAGCAAAGC		
	9	F	CTTTGTAGGACAGGTCTTGTCAAC	218	61
		R	CTCCTTTCCTCTTCATCCTCC		
	10	F	CCTGTTTACTGTCCTGAAGGCCACA	303	67
		R	GCCAGGGCACTGTGTGAGATGG		
	11	F	CTGGGGTACTCACGTTCGGTCCC	363	66
		R	CAGACCCATCATCGGGAGGTCAC		
	12	F	GTCCTTTCTTAGTCTCTGCCTC	192	66
		R	GGAAGTCTTTCGCCAGGGCAAG		
	13	F	TCGGAAACATCTCTATCTGAGTGG	293	61
		R	TTCTCCCAAAACAATTAGTAGTCTGAAC		
	14	F	TGTCAGTGAGACTATTCCAACTCG	309	61
		R	GGAAACTCATTCACTTGG		

Statistical Analyses

Initial analysis of SNP genotype data involved quality control checks consisting of verification of internal controls and assessment of Mendelian inconsistencies, followed by final haplotype consistency analyses using Simwalk2 [117]. Conformity with anticipated Hardy-Weinberg equilibrium expectations was established. SNPs (excluding the variants occurring less than 1%) were examined along with our previous marker set to determine intermarker LD using both the r^2 and D´ measures calculated by the publicly available Haploview program [129]. Transmission disequilibrium (only for the common, >0.01, follow-up markers) in autism families was determined using the PDT statistic, a variant of the transmission disequilibrium test (TDT), developed for use with general pedigrees [119]. Common haplotypes were determined using Haploview [129], and analyzed using the FBAT analysis software package [130]. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$.

In addition to TD being measured in all families, we performed analyses using only those families containing male-affected individuals. This approach was recently used by Stone et al. to examine the differences that sex of the autistic individuals would make in linkage analysis [79]. We excluded from these additional analyses any family containing a female-affected individual.

Visualization Tools for Alignment (VISTA) analysis was performed via web-based submission (http://www.gsd.lbl.gov/vista/) of human and mouse

genomic sequence [121]. Regions showing evolutionarily conserved non-coding sequence were identified for variant screening purposes.

Results

To examine in greater detail the serotonin transporter locus (*SLC6A4*), we chose additional markers and strategies to more closely examine a possible role for this gene in autism etiology. The details of our follow-up markers are highlighted in Table 4-1 and their relative positions across the transcriptional unit are illustrated in Figure 4-1.

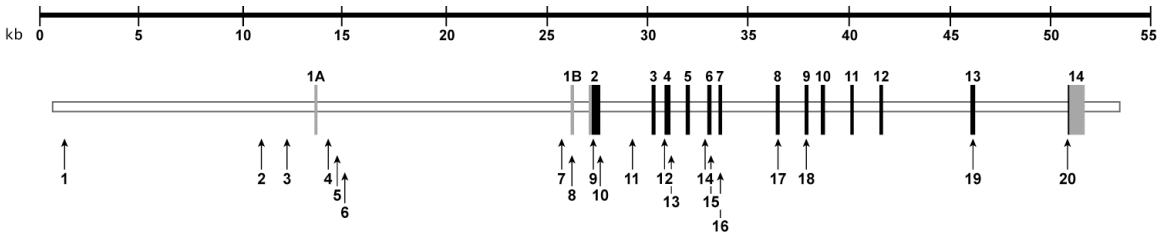


Figure 4-1. Schematic representation of *SLC6A4* detailing follow-up marker locations. The 38kb transcriptional unit contains 15 total exons represented by the vertical boxes. Coding regions are indicated in black with the untranslated regions in gray. Markers are labeled and their positions indicated by arrows.

In addition to testing our hypothesis with respect to common variation at this locus, it is quite plausible that allelic heterogeneity may play a role. To test this hypothesis we choose to examine the 10 known rare non-synonymous changes to determine if they are enriched in our autism sample ([123] and dbSNP http://www.ncbi.nlm.nih.gov/SNP/). In addition we, along with our collaborators in the Blakely laboratory, performed variant screening across

exonic sequence and non-coding regions identified through VISTA as being evolutionarily conserved with mouse. The non-coding regions screened through re-sequencing are illustrated in Figure 4-2.

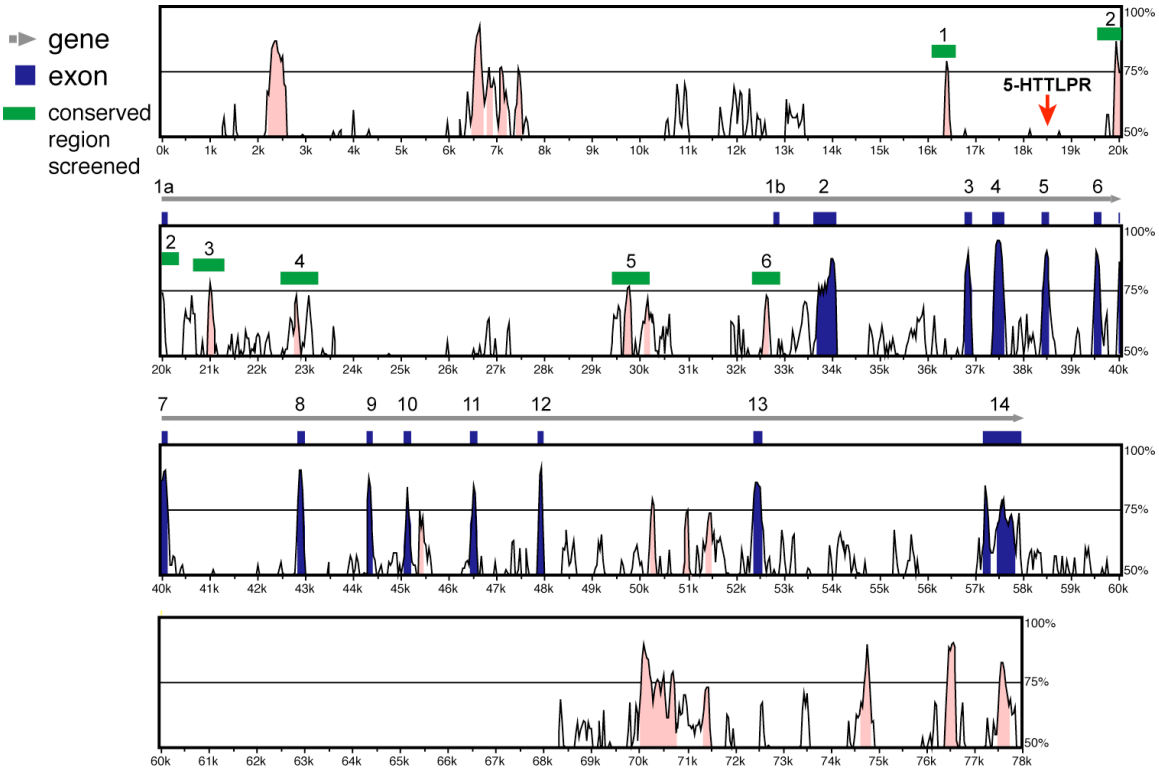


Figure 4-2. Variant screening at *SLC6A4* relative to sequence conservation. Output from VISTA analysis of the *SLC6A4* transcriptional unit (indicated by gray arrow above the plot) is shown, with regions of non-coding sequence conservation (>75% identity) highlighted by pink shading and coding homology by blue shading. Regions of conservation that were screened for variants are indicated with a green bar above them. The length of the bar covers approximately the amount of sequence examined.

Genotypes at individual markers did not deviate from expectations of Hardy-Weinberg equilibrium (Data not shown). We also employed the publicly available Haploview program to map linkage disequilibrium across the locus, in

hopes of identifying the common haplotypes seen in our sample. These common haplotypes in addition to our single markers were examined for association to disease. The single marker PDT results are presented in Table 4-4 for the common follow-up markers in both the overall dataset and the male only families dataset. Though the variant in intron 5 (hCV7473202) remains significant in the overall dataset of 384 families, the 5-HTTLPR variant is no longer significant. However, when we examine those families containing only male-affected individuals (235/384 or 61%), we observe that 5-HTTLPR is nominally significant and our intron 5 variant becomes even more significant.

Table 4-4. Follow-up PDT analysis of the *SLC6A4* locus

		Overall Sample (384 families)				Affected Male-Only Sample (235 families)			
Marker No.	Marker	PDT (P)	Allele	# of Alleles Transmitted	# of Alleles Not Transmitted	PDT (P)	Allele	# of Alleles Transmitted	# of Alleles Not Transmitted
1	hCV7473213	0.68				0.68			
2	hCV11424041	1.00				0.78			
3	5-HTTLPR	0.14				**0.03**	484(s) 584(l)	362 398	314 446
4	hCV11424045	0.31				0.55			
5	hCV7911197	0.28				0.25			
6	hCV11424046	0.21				0.64			
7	hCV1841705	0.93				0.91			
8	hCV1841706	0.31				0.52			
11	VNTR	0.65				0.72			
14	hCV7473202	**0.03**	A G	93 1215	124 1184	**0.01**	A G	47 711	76 682

Examination of linkage disequilibrium across all common markers studied, combining those in the previous report with our common (> 1%) follow-up markers, identifies three loosely defined haplotype blocks. The Haploview output is illustrated in Figure 4-3. Association analysis in both the overall dataset as well

as the male-affected only families, using FBAT, of all common haplotypes (≥ 0.05) yielded no significant association (Table 4-5).

Figure 4-3. LD analysis of *SLC6A4* using Haploview. The black outline surrounds each of our self-defined haplotype blocks with intermarker D´ values labeled within the individual boxes. Intermarker linkage disequilibrium with D´ values of 1.0 are never shown, but indicated by the dark red boxes for a LOD score ≥ 2 and indicated by the blue boxes for a LOD score < 2. White shading indicates a D´ < 1 and a LOD < 2, while shades of pink/red indicate D´ < 1 and a LOD ≥ 2.

Table 4-5. FBAT haplotype analysis for *SLC6A4*. Gray denotes htSNPs. Markers numbers are labeled relative to Figure 4-3.

Overall Sample (384 Families)

1-2-3 Haplotype	Frequency	# of Families	Observed Transmissions	Expected Transmissions	P
T - G - S	0.40	84	141	134	0.38
C - G - L	0.29	75	96	101	0.59
T - G - L	0.28	69	81	88	0.35

4-5-6 Haplotype	Frequency	# of Families	Observed Transmissions	Expected Transmissions	P
A - T - T	0.47	93	178	168	0.26
A - C - T	0.46	94	141	145	0.71
T - C - A	0.07	28	26	33	0.20

7-8-9-10-11-12-13-14-15-16-17 Haplotype	Frequency	# of Families	Observed Transmissions	Expected Transmissions	P
C - T - T - A - A - G - 12 - G - G - C - G	0.44	75	135	126	0.28
T - T - A - A - A - G - 10 - A - G - T - T	0.31	77	100	93	0.40
C - C - T - C - C - G - 12 - G - A - T - T	0.17	42	45	48	0.53

Affected Male-Only Sample (235 Families)

1-2-3 Haplotype	Frequency	# of Families	Observed Transmissions	Expected Transmissions	P
T - G - S	0.40	52	90	80	0.14
T - G - L	0.32	50	53	64	0.10
C - G - L	0.25	45	46	49	0.61

4-5-6 Haplotype	Frequency	# of Families	Observed Transmissions	Expected Transmissions	P
A - T - T	0.49	56	110	102	0.25
A - C - T	0.44	60	77	84	0.38
T - C - A	0.07	18	18	20	0.64

7-8-9-10-11-12-13-14-15-16-17 Haplotype	Frequency	# of Families	Observed Transmissions	Expected Transmissions	P
C - T - T - A - A - G - 12 - G - G - C - G	0.45	52	90	80	0.16
T - T - A - A - A - G - 10 - A - G - T - T	0.29	43	50	48	0.74
C - C - T - C - C - G - 12 - G - A - T - T	0.17	31	30	33	0.60

We examined all 384 families for the 10 rare non-synonymous variants. Only one of these rare variants was shown to have an allele frequency greater than 1% (GLY56ALA/rs6355/hCV11414113). Although this minor allele frequency did not differ *significantly* from a previous report by Glatt et al. [123], when tested

in a control panel of 100 unrelated Caucasian individuals obtained from the Coriell Cell Repository the minor allele frequency was 4% (8/200 chromosomes). This minor allele frequency is closer to that reported in dbSNP (0.02), but warrants further review and sequence confirmation. Rare alleles for three of the other non-synonymous variants were also detected: 1 heterozygous individual for the LEU362MET variant, 1 heterozygous individual for the LYS201ASN variant, and 3 heterozygous individuals for the LYS605ASN variant.

We re-sequenced six regions of conservation across the *SLC6A4* locus to identify variants that may or may not be autism specific. Variant screening across these regions was initially conducted with a panel of our "most rigid" unrelated individuals and subsequently with our "most linked" unrelated individuals as described in the Materials and Methods section of this chapter. Results for the re-sequencing efforts across the conserved regions are displayed in Table 4-6. Regions 1 and 2 did not show any variation in the samples we examined. With the exception of region 4, all other regions identified markers that have already been identified and listed in either dbSNP or the Celera database. In conserved region four we identified four heterozygous individuals for a single variant that to the best of our knowledge had not been previously identified. This A to G (common/rare) allele change is surrounded by the sequence GTGGATGTGAGGC**A**GTCTGACTGCCTT. Three of the markers (hCV7911197, hCV11424046, and hCV1841705) identified through screening were chosen to be genotyped on the entire sample and included in the association analysis

(Table 4-4 markers 5,6,and 7). Variants detected having relatively low frequency were excluded in the present study, since we hypothesized that a common variant in these conserved regions would be the source of our increased linkage.

Table 4-6. Results of variant screening across the *SLC6A4* locus. A panel of 24 affected but unrelated individuals was chosen for re-sequencing purposes. The individuals came from the 24 families having the highest family-based multipoint heterogeneity lod score at the 50cM location on chromosome 17.

Gene	Conserved Non-coding Region	# of Unique Sequence Variations Detected	dbSNP rs#/ Celera hCV#	Minor Allele Frequency
SLC6A4	1	0	_	_
	2	0	_	_
	3	3	rs2020934/ hCV7911197 rs2020935/ hCV11424046 rs2020933/ hCV11424045	0.08 0.40 0.08
	4	1	novel	0.08
	5	1	rs8071667/ hCV28964487	0.09
	6	2	rs2020940/ hCV11414117 rs25528/ hCV1841705	0.04 0.38

Discussion

The genes involved in autism have been elusive, likely due to the complex phenotype harboring a complex genetic architecture. Using multiple strategies to tease apart this complex disorder may be the only way to gain insight into its underlying epidemiology. This follow-up study of the serotonin transporter locus addresses several lines of investigation. Previous evidence and suggestion of serotonin involvement, in particular relative to the serotonin transporter, is strong in research surrounding autism and other neuropsychiatric disorders. However,

the specific details of its involvement are not only unproven, but to date are very unclear. As mentioned previously, numerous studies having shown whole blood serotonin levels to be sexually dimorphic. Combining this with the average sex-ratio of 4 affected males to 1 affected female in autism lends support for a biological mechanism involving the serotonergic pathway. However, without other strong evidence suggesting serotonin involvement in autism, such an argument would be necessarily weak given that any sexually dimorphic trait could be argued to play a role in autism susceptibility.

In this follow-up study we set out to test several hypotheses. We examined common alleles, rare coding alleles, and screened for novel variants across the serotonin transporter for involvement in autism. We did not find strong support for either a common allele, haplotype, or rare allele to increase autism susceptibility in our overall dataset of 384 autism families. However, we did find suggestive evidence for both the 5-HTTLPR variant and the intron 5 (rs140700/hCV7473202) to be associated in families containing only male-affected individuals. While other groups have detected association to the 5-HTTLPR polymorphism (detailed in Chapter III), this is the first data to suggest this association to be specific to autism families containing only male-affected individuals. This work extends upon findings by other groups suggesting linkage and more specifically linkage at 17q near the serotonin transporter is increased when examining the subset of families containing only male-affected individuals. We speculate that our findings suggest a yet unknown role for *SLC6A4*

involvement in this phenomenon. Future directions will include examination of the variant screening of exonic sequence and further examination for male-specific increased risk for autism.

CHAPTER V

GENETIC ANALYSIS OF THE 15q12 $GABA_A$ RECEPTOR SUBUNIT CLUSTER[1]

Introduction

Evidence for GABA and Glutamate Involvement in Autism

There are a number of findings that support a role for the GABA (γ-aminobutyric acid) neurotransmitter system in autism susceptibility. Two studies have shown decreased levels of $GABA_A$ receptors in autism. Blatt and colleagues showed reduced binding of radiolabelled $GABA_A$ receptor ligands in autopsy brain specimens from individuals with autism [39]. Chugani and colleagues used PET imaging in children with autism and also showed reduced $GABA_A$ receptors [131]. The $GABA_A$ receptor agonist benzodiazepine is effective in treating seizure and anxiety disorders common in autism. Finally, elevated levels of circulating GABA and its essential precursor glutamate have been observed in children with autism [40-42].

15q11-q13 Autism Candidate Region

A cluster of $GABA_A$ receptor subunit genes maps within the chromosome 15q11.2-q13 autism candidate region (Figure 5-1). Interstitial duplications of this

[1] Adapted from *Am J Med Genet B Neuropsychiatr Genet 2004* Nov 15, 131B:51-9.

region are associated with a significant risk of autism, and risk is greater for duplications of maternal compared to paternal origin [49, 132-136]. Maternal supernumerary pseudodicentric inverted duplicated marker chromosomes 15 (so-called idic(15) markers) carry two additional copies of a larger region and give rise to a more severe autistic phenotype [137-141]. Duplication-mediated autism arguably stems from a dosage effect of genes in the duplicated intervals. Maternal or paternal deletions of the same region affected by interstitial duplications give rise to Angelman (AS [MIM 105830]) or Prader-Willi (PWS [MIM 176270]) syndrome, respectively, because of the loss of expression of imprinted genes in the interval (reviewed in [142]; see Figure 5-1). This is noteworthy since symptoms of autism can be associated with both AS and PWS. In addition to regions of paternal-specific and maternal-specific gene expression, there is an apparently non-imprinted region containing a cluster of $GABA_A$ receptor subunit genes (*GABRB3*, *GABRA5* and *GABRG3*). While the GABA genes will be the focus of this chapter, greater detail of the proximal 15q11-q13 region will follow in Chapter VII, which focuses on examination of the imprinting center and maternal expression domain.

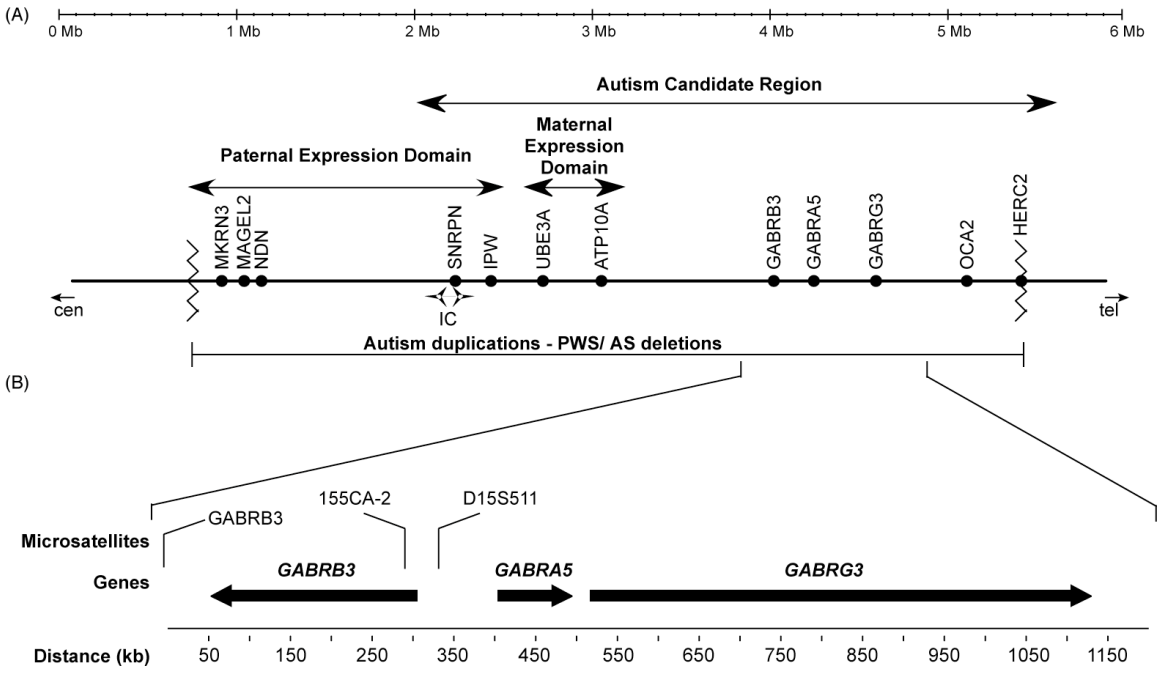


Figure 5-1. Schematic map of the 15q11-q13 autism candidate region. (A) Autism duplication and PWS/AS deletion interval. The region of chromosome 15q subject to interstitial duplication in cases of autism or deletion in PWS or AS is shown. Imprinted paternal expression (PWS) and maternal expression (AS) domains are indicated above the map relative to specific genes. The autism candidate region includes the 15q imprinting center (IC), maternally-expressed genes and the cluster of $GABA_A$ receptor subunit genes. A scale in Mb for the interval is provided above the map. (B) The $GABA_A$ receptor subunit gene cluster. The 1-Mb interval containing *GABRB3*, *GABRA5* and *GABRG3* is expanded and shows relative position and transcriptional orientation (arrows) of each gene. Reference microsatellite markers previously shown to be linked or associated are also indicated.

Accumulating genetic evidence also suggests the existence of genetic factor(s) in the GABA gene cluster region in idiopathic autism. Genomic linkage screens in autism have identified proximal 15q [27, 31, 109], but results are mixed as others have failed to detect significant linkage [21, 25, 28, 30]. Two recent reports, including one from our group, demonstrated that subsetting

autism families, based on variables from the Autism Diagnostic Interview-Revised (ADI-R), results in significantly increased evidence for linkage to the GABA region [53, 54]. Several reports have documented findings of allelic association at microsatellite [44-46, 143] and single nucleotide polymorphism (SNP) markers [144] in the GABA region. Two other groups, using microsatellite markers, did not identify association to autism in this region [51, 52]. Studies to date have involved analysis of only a small number of microsatellite or SNP markers and thus have not thoroughly surveyed this region for association. To test the hypothesis that common allele(s) in the GABA gene cluster confer risk for autism, we undertook a detailed analysis of linkage disequilibrium (LD) and allelic association across this 1-Mb region.

Materials and Methods

Families

The sample for this study consisted of 123 multiplex families. Forty-eight multiplex families were recruited at the Tufts/New England Medical Center and 75 affected sib-pair families were obtained from the Autism Genetics Resource Exchange (AGRE; http://agre.org). The vast majority of families (~98%) are of Caucasian ethnicity. All affected individuals were at least four years of age and were clinically assessed with the ADI or ADI-R and most with the Autism Diagnostic Observation Schedule (ADOS). At least one sib had to meet ADI

algorithm criteria for an autism diagnosis, while additional siblings may be on the broader autism spectrum. Families were excluded from the study if probands had a known medical or neurological condition suspected to be associated with their autistic phenotype (e.g. fragile X syndrome). The procedures for clinical evaluation of affected individuals for the AGRE families have been previously described [30].

Molecular Analysis

DNA was isolated from peripheral blood or lymphoblastoid cells using the PureGene kit according to the manufacturer's recommendations (Gentra Systems, Minneapolis, MN). SNPs from the dbSNP and Celera SNP databases were selected based on their map position, minor allele frequency, and, in one case, a finding of allelic association by Menold et al. [144]. Marker and exon locations and intermarker distances are based on the public (UCSC; July 2003 freeze; http://genome.cse.ucsc.edu/) and Celera (http://www.celeradiscoverysystem.com/index.cfm) assemblies and published gene structure information for *GABRB3* and *GABRA5* [145, 146]. Database reference numbers and other details for the markers studied are cited in Table 5-1. Genetic (cM) distances are based on the deCODE genetic map [114]. Individual SNPs were genotyped by either fluorescent polarization template-directed dye terminator incorporation assay (FP-TDI) or TaqMan™. FP-TDI PCR assays were developed and optimized to amplify an ~200 bp region flanking

SNPs. PCR primers and probes for assays are listed in Table 5-2. This information is proprietary for the majority of markers, for which Assays-On-Demand™ were obtained from Applied Biosystems (Foster City, CA).

For FP-TDI genotyping, PCR reaction volumes were 8 μl, employing 5 ng genomic DNA template, 0.2 □M primers, 125 □M dNTPs and Applied Biosystems AmpliTaq Gold DNA polymerase and buffer (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 94 °C for 15 s, optimal annealing temperature (T_A°C) for 30 s, and 72 °C for 15 s, and a final extension at 72 °C for 10 min. FP-TDI analysis was performed using materials supplied in commercially-available Acycloprime™ kits according to the manufacturer's published protocols (Perkin-Elmer Lifesciences, Boston, MA) and as described elsewhere [112]. Samples were analyzed using a VICTOR$_2$™ multi-label plate reader instrument (Perkin-Elmer Life Sciences).

For TaqMan™ genotyping assays, reactions were performed in a 5 μl volume, employing 2.5ng genomic DNA template, according to manufacturer's recommendations (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 92 °C for 15 s, 60 °C for 1 min. Samples were analyzed using an ABI 7900HT™ Sequence Detection System.

Table 5-1. SNP markers spanning the $GABA_A$ receptor subunit cluster. Bold highlights microsatellite markers previously studied.

Gene	SNP No.	Region	dbSNP rs# /Celera hCV#	Alleles	Minor Allele Frequency	Intermarker Distance (kb)
Intergenic		Intergenic	**GABRB3**			52.9
Intergenic	1	Intergenic	rs8025575/ hCV2911914	C/G	0.45	10.1
GABRB3	2	Intron 8	rs2081648/ hCV2911917	T/C	0.15	12.5
	3	Intron 7	rs1432007/ hCV8866669	A/G	0.47	10.4
	4	Intron 6	rs1426217/ hCV2901088	G/A	0.38	46.3
	5	Intron 3	rs2873027/ hCV2901140	T/C	0.45	0.5
	6	Intron 3	rs4542636/ hCV2901143	T/C	0.45	20.3
	7	Intron 3	rs754185/ hCV2901163	T/C	0.34	14.7
	8	Intron 3	rs12912421/ hCV2901177	G/A	0.50	8.5
	9	Intron 3	rs11161328/ hCV2901182	A/G	0.28	7.6
	10	Intron 3	rs1346149/ hCV2061398	A/G	0.47	10.2
	11	Intron 3	rs878960/ hCV8865198	C/T	0.38	9.6
	12	Intron 3	rs1863464/ hCV2901200	G/A	0.18	12.8
	13	Intron 3	rs11631421/ hCV245488	T/C	0.44	5.9
	14	Intron 3	rs981778/ hCV2901236	A/G	0.43	31.8
	15	Intron 3	rs970408/ hCV2901263	C/T	0.10	8.0
	16	Intron 3	rs2059574/ hCV2901280	T/A	0.49	14.0
		Intron 3	**155CA-2**			2.4
	17	Intron 3	rs3212337/ hCV218360	C/T	0.36	39.4
Intergenic		Intergenic	**D15S511**			1.9
Intergenic	18	Intergenic	rs4632100/ hCV30714164	A/G	0.18	15.7
Intergenic	19	Intergenic	rs4506865/ hCV30714186	A/G	0.10	47.7
GABRA5	20	Intron 3	rs2075716/ hCV1843341	C/T	0.38	5.4
	21	Intron 3	hCV474240	C/T	0.36	3.8
	22	Intron 3	hCV11298361	G/A	0.41	8.3

	23	Intron 5	hCV252720	T/C	0.43	9.8
	24	Intron 5	rs9745027/ hCV27725	C/A	0.50	18.2
	25	Intron 6	hCV42974	C/T	0.34	21.8
	26	Exon 7	rs140682/ hCV1028938	C/T	0.47	6.1
	27	Exon 9	rs140685/ hCV1028939	T/C	0.49	9.5
Intergenic	28	Intergenic	rs11263717/ hCV2078419	T/A	0.39	34.1
GABRG3	29	Intron 2	rs1432129/ hCV8866584	C/A	0.47	18.7
	30	Intron 2	hCV2078482	C/T	0.47	13.5
	31	Intron 2	rs6606855/ hCV2078497	A/G	0.45	8.8
	32	Intron 3	rs7172534/ hCV2078506	T/C	0.37	21.5
	33	Intron 3	rs4078843/ hCV2078548	G/A	0.31	45.3
	34	Intron 3	rs4555125/ hCV37817	A/G	0.49	30.8
	35	Intron 3	rs1029937/ hCV2665757	G/A	0.29	7.7
	36	Intron 3	rs208174/ hCV2665743	C/T	0.27	7.2
	37	Intron3	rs6606877/ hCV2665737	C/T	0.17	15.6
	38	Intron 3	rs2286946/ hCV2665715	A/G	0.48	5.2
	39	Intron 3	rs741121/ hCV2665706	G/T	0.46	17.1
	40	Intron 3	rs208129/ hCV2665692	A/T	0.42	1.6
	41	Intron 3	rs208126/ hCV2665687	G/T	0.43	29.4
	42	Intron 3	rs12907392/ hCV9408557	T/C	0.30	14.1
	43	Intron 3	rs897173/ hCV9408511	A/G	0.25	16.8
	44	Intron 3	rs897177/ hCV9408473	T/C	0.18	18.4
	45	Intron 3	rs6606891/ hCV9408434	C/T	0.44	3.7
	46	Intron 3	rs8043244/ hCV9408423	G/A	0.41	65.1
	47	Exon 5	rs140674/ hCV18418	T/C	0.04	5.1
	48	Intron 5	hCV435176	A/G	0.44	26.0
	49	Intron 5	rs9672931/ hCV376685	G/A	0.27	7.5
	50	Intron 5	hCV59714	A/G	0.27	26.0

	51	Intron 5	rs9635410/ hCV9399190	A/G	0.34	6.3
	52	Intron 5	rs12440080/ hCV458188	G/A	0.45	23.0
	53	Intron 5	rs11631444/ hCV473958	A/G	0.46	42.0
	54	Intron 5	rs4550406/ hCV374658	C/T	0.28	36.1
	55	Intron 6	rs11074283/ hCV34499	C/T	0.48	10.4
	56	Intron 6	rs1871019/ hCV11670850	A/G	0.40	8.4
	57	Intron 6	rs11631143/ hCV1846028	G/A	0.42	8.3
	58	Exon 8	rs140679/ hCV1845989	T/C	0.43	62.9
Intergenic	59	Intergenic	rs1382056/ hCV8926104	G/A	0.48	_

Table 5-2. PCR and genotyping primers for the GABA cluster SNPs. Labels are as follows: (F) Forward Primer, (R) Reverse Primer, (FP) FP-TDI extension primer, VIC and FAM are fluorescent labels for TaqMan™ allelic discrimination probes (SNPs indicated by bold text), AbD is an Assay-By-Design from ABI, AoD is an Assay-On-Demand from ABI. MGB stands for minor groove binder, and NFQ stands for non-fluorescent quencher. (*) SNPs 1, 4-8, 11-16, 21-46, 48-53, and 55-59 are AoDs. Sequence for primers and probes, for AoD assays, is proprietary information.

SNP No.	Primer sequences (5´-3´)		Product size (bp)	T_A (°C)	Assay
*	NA	Unavailable	< 200	60	TaqMan (AoD)
2	F	CATGAAAAGGGATTTGATAAATTGAGGTCAT	130	60	TaqMan (AbD)
	R	CAACATCAGATAGATTTTAAACATATAAGCTTACCATTT			
	VIC	VIC-ACTGTTT**G**AGATGCTG-MGB-NFQ			
	FAM	FAM-AACTGTTT**A**AGATGCTG-MGB-NFQ			
3	F	AGCCAGCCCATGCCTTTATC	86	60	TaqMan (AbD)
	R	GATTCCACTTTTTCTTCATGACAGCAT			
	VIC	VIC-AAAGCCAC**G**GAGGCA-MGB-NFQ			
	FAM	FAM-AAAGCCAC**A**GAGGCA-MGB-NFQ			
9	F	TGGTAAGCCACTGAGACAAGTAGG	116	51	FP
	R	GTTTGTGCGTGTGCTTGGTAA			
	FP-R	TAACAGTCTAAGGTCAGGGGGTTTCTACT			
10	F	GCCATGGAGCAAAGACCCTA	107	53	FP
	R	CTTGTGATGCACGTAGTTGTCTGA			
	FP-R	CTTTCACAGTCACTGCTTTCACAGC			
17	F	AAAGTCCTGTGAGCCCTTAATG	194	53	FP
	R	GCTGGTGAGGTAGGTAATAGAGGT			
	FP-F	GCTTTGTCCTGCAGTAAGAGTAAGAAAGT			
18	F	GGGTGGAACGTGCAGACA	56	60	TaqMan (AbD)
	R	CCCACACAGAGGCAGTCA			
	VIC	VIC-CCACTC**G**GCACCGC-MGB-NFQ			
	FAM	FAM-CCACTC**A**GCACCGC-MGB-NFQ			
19	F	GCCATGTTAATTTAATGTCATTTTGGAGATTCT	119	60	TaqMan (AbD)
	R	CATGTGATCTAACATAACAATAGATTTATCTGTTCCT			
	VIC	VIC-CACCTGA**C**AATTCT-MGB-NFQ			
	FAM	FAM-ACACCTGA**T**AATTCT-MGB-NFQ			
20	F	CAGTTGCCCTCCACGGTTC	168	55	FP
	R	ATCTTCAGGCACCTGTGGTTTATG			
	FP-F	AGTTGCCTTGAAAGCCAGGCC			
47	F	GAGCTAGACTGACACTTGGCTTTT	95	60	TaqMan (AbD)
	R	GGCAGGAGTGTTCGTCCAT			
	VIC	VIC-TGGCACTC**G**GCATT-MGB-NFQ			
	FAM	FAM-CTGGCACTC**A**GCATT-MGB-NFQ			
54	F	ATGCACTTATTACCCATTTGTATATCTTCCTT	105	60	TaqMan (AbD)
	R	CTTTCACACGTCAACAATACAAAGACA			
	VIC	VIC-AAATGG**G**CAAAGAAT-MGB-NFQ			
	FAM	FAM-TTAAAAATGG**A**CAAAGAAT-MGB-NFQ			

Statistical Analyses

Initial analysis of SNP genotype data involved quality control checks consisting of verification of internal controls and assessment of Mendelian inconsistencies, followed by final haplotype consistency analyses using Simwalk2 [117]. Conformity with anticipated Hardy-Weinberg equilibrium expectations was established, and SNPs were examined for intermarker LD using both the r^2 and D´ measures calculated by the GOLD (Graphical Overview of Linkage Disequilibrium) software package [118]. Similar to the definition described by Gabriel et al., SNP pairs were considered to be in strong LD if D´ values were >0.75 [147]. Neighboring SNPs were considered to be in a single LD block if all SNP pairs were in strong LD. Minimal block lengths were determined from intermarker spacing of SNPs defining the blocks. Transmission disequilibrium (TD) in autism families was determined using the pedigree disequilibrium test (PDT) statistic, developed for use with general pedigrees [119]. Common haplotypes (≥5%) were identified using TRANSMIT [120]; analysis of TD was performed using adjacent SNP pairs inclusive of loci significant in single marker analysis and haplotype tag SNPs for other multi-locus blocks. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$.

Results

We selected and genotyped 59 SNP markers spanning the 1-Mb interval containing the *GABRB3*, *GABRA5* and *GABRG3* (Table 5-1 and Figure 5-2) genes. Markers were primarily selected based on minor allele frequency and intermarker spacing to provide dense representation of regional LD. The average minor allele frequency was 0.37; six markers had a minor allele frequency less than 0.2 and one less than 0.1. This last marker (rs140674) was chosen based on a published report of nominal association in *GABRG3* [144]. The average intermarker spacing for the entire interval was 17.7 kb, while for individual genes it was 14.3, 10.3 and 18.7 kb, for *GABRB3*, *GABRA5* and *GABRG3*, respectively. Genotyping for all SNPs was performed on DNA samples from 123 multiplex (48 New England and 75 AGRE) families. Genotypes at individual markers did not deviate from expectations of HWE (data not shown).

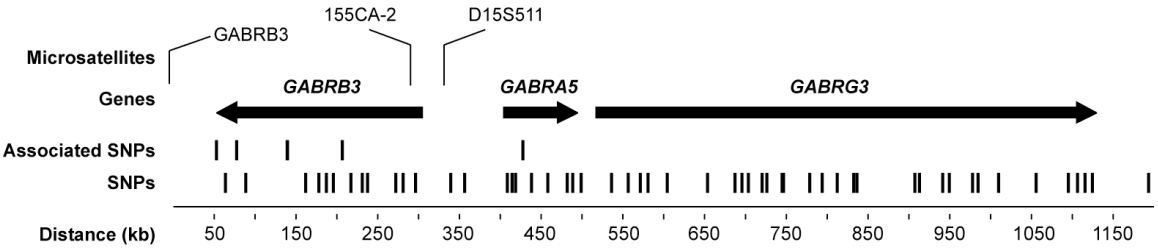


Figure 5-2. Schematic of markers located across the $GABA_A$ receptor cluster. SNP markers analyzed in this study are indicated by vertical hashes, above the scale in kb; proximity of SNPs 5/6, 21/22, 40/41 and 45/46 in relation to the scale is such that these markers are not distinguished by separate hashes. Associated markers are shifted up from other SNPs. Reference microsatellite markers previously shown to be linked and/or associated are also indicated.

With the goal of detecting and characterizing potential allelic effects on autism risk, we initially analyzed single marker genotypes for evidence of TD. Data were analyzed using the PDT, and resulting P-values are listed in Table 5-3. Five of 19 markers, representing three distinct locations across the *GABRB3* gene, demonstrated significant association at the nominal level ($P \leq 0.05$). The first location corresponds to SNP 1 (1df, χ^2=5.25; P=0.02) and SNP 3 (1df, χ^2=6.98; P=0.01), which are located towards the 3´ end of the gene. The second region has two sites showing significant allelic effects within intron 3. One involves adjacent SNPs 5 (1df, χ^2=4.18; P=0.04) and 6 (1df, χ^2=4.62; P=0.03). The final site corresponds to SNP 11 (1df, χ^2=4.27; P=0.04). One marker within intron 5 of *GABRA5* (SNP 23) also showed evidence for association (1df, χ^2=4.62; P=0.03). An initial examination of haplotypes, specifically at markers showing evidence for association, involved analysis of adjacent SNP-pairs using TRANSMIT. Table 5-4 details these results, and includes single marker allelic transmission data for significant SNPs. Consistent with results from individual markers, several two-SNP haplotypes demonstrated significant transmission distortion. None of the 30 SNPs located across the ~570-kb *GABRG3* gene showed evidence of association.

Table 5-3. Association analysis of single SNPs across the GABA cluster.

Gene	SNP No.	Overall PDT (P)		Gene	SNP No.	Overall PDT (P)
Intergenic	1	**0.02**				
GABRB3	2	0.37		*GABRG3*	29	0.18
	3	**0.01**			30	0.21
	4	0.21			31	0.25
	5	**0.04**			32	0.61
	6	**0.03**			33	0.89
	7	0.82			34	0.18
	8	0.94			35	0.94
	9	0.82			36	0.60
	10	0.07			37	0.60
	11	**0.04**			38	0.66
	12	0.80			39	0.85
	13	0.21			40	0.66
	14	0.28			41	0.16
	15	0.74			42	0.51
	16	0.61			43	0.06
	17	0.78			44	0.94
Intergenic	18	0.23			45	0.48
Intergenic	19	0.67			46	0.25
					47	0.71
					48	1.00
GABRA5	20	0.14			49	0.38
	21	0.78			50	0.28
	22	0.37			51	0.43
	23	**0.03**			52	0.55
	24	0.52			53	0.60
	25	0.08			54	0.36
	26	0.65			55	0.90
	27	0.84			56	0.33
Intergenic	28	0.89			57	0.79
					58	0.60
				Intergenic	59	0.75

Table 5-4. Pair-wise analysis of adjacent SNPs.

Gene	SNP No.	Overall PDT (P)	Allele	T	NT	χ^2	TRANSMIT (P)	
Intergenic	1	**0.02**	C G	260 188	226 222	5.25	**0.02**	
GABRB3	2	0.37						**0.02**
	3	**0.01**	A G	263 199	220 242	6.98	**0.04**	
	4	0.21						0.13
	5	**0.04**	C T	187 279	222 244	4.18	0.07	
	6	**0.03**	C T	182 272	218 236	4.62		0.13
	7	0.82						
	10	0.07					0.11	
	11	**0.04**	C T	265 191	296 160	4.27		**0.04**
	12	0.80						
GABRA5	22	0.37					0.07	
	23	**0.03**	C T	176 288	211 253	4.62		**0.01**
	24	0.52						

To characterize intermarker LD and haplotype structures across the entire 1-Mb region, D′ and r^2 measures were calculated from the genotype data using GOLD. D′ values for all intermarker combinations are represented in Figure 5-3, and 14 multi-SNP LD blocks (D′ > 0.75) are identified. Outside of these regions of relatively low haplotype diversity, LD between adjacent marker pairs is generally very low. While borders for individual blocks are not precisely defined, the multi-SNP blocks shown are represented by 38 SNPs and comprise a minimum of 263 kb of 1,040 kb, or 25%, of the entire interval. Average SNP coverage was higher

for *GABRB3* (14.3 kb^{-1}) and *GABRA5* (10.3 kb^{-1}), in comparison to *GABRG3* (18.7 kb^{-1}) and intergenic regions (see Table 5-1). Examination of minimal block lengths in proportion to total genomic DNA for genes individually reveals that multi-SNP blocks represent 40%, 45% and 24% of the DNA encoding *GABRB3*, *GABRA5* and *GABRG3*, respectively. Minimal block lengths ranged from 3.7 to 46.8 kb, and the average block size was estimated to be 18.8 kb for this Caucasian sample; this block size is consistent with previous reports [147]. Two sets of overlapping blocks were identified. Blocks 1-2-3 and 4-5-6 overlap based on strong LD between SNP 2 and 4-5-6, despite weak LD between other SNPs in the two blocks. SNP 22 was also found to bridge two overlapping blocks, and these findings are similar to those reported elsewhere [148, 149].

For all multi-SNP blocks, common haplotypes (>5%) were identified using TRANSMIT (Table 5-5). This permitted identification of haplotype tags, or the subset of SNPs (htSNPs) that detect all common haplotypes for a given block. Of the 38 SNPs present within blocks, 7 were eliminated as redundant. At the current resolution, 82% of the SNPs in multi-locus LD blocks, or 88% of the total, were required to represent common haplotypes across this interval. Genotype data for the reduced set of SNPs was then analyzed using TRANSMIT to test for TD with autism; results of this analysis are presented in Table 5-5. Transmissions at two multi-locus blocks, both containing SNPs showing association individually, were found to deviate significantly from that expected under the null hypothesis. The SNP 1-2-3 block (2df, χ^2=8.02; P=0.02), located at the 3′ end of *GABRB3*,

and the SNP 22-23-24 block (5df, χ^2=15.5; P=0.01), located in *GABRA5*, both showed significant results.

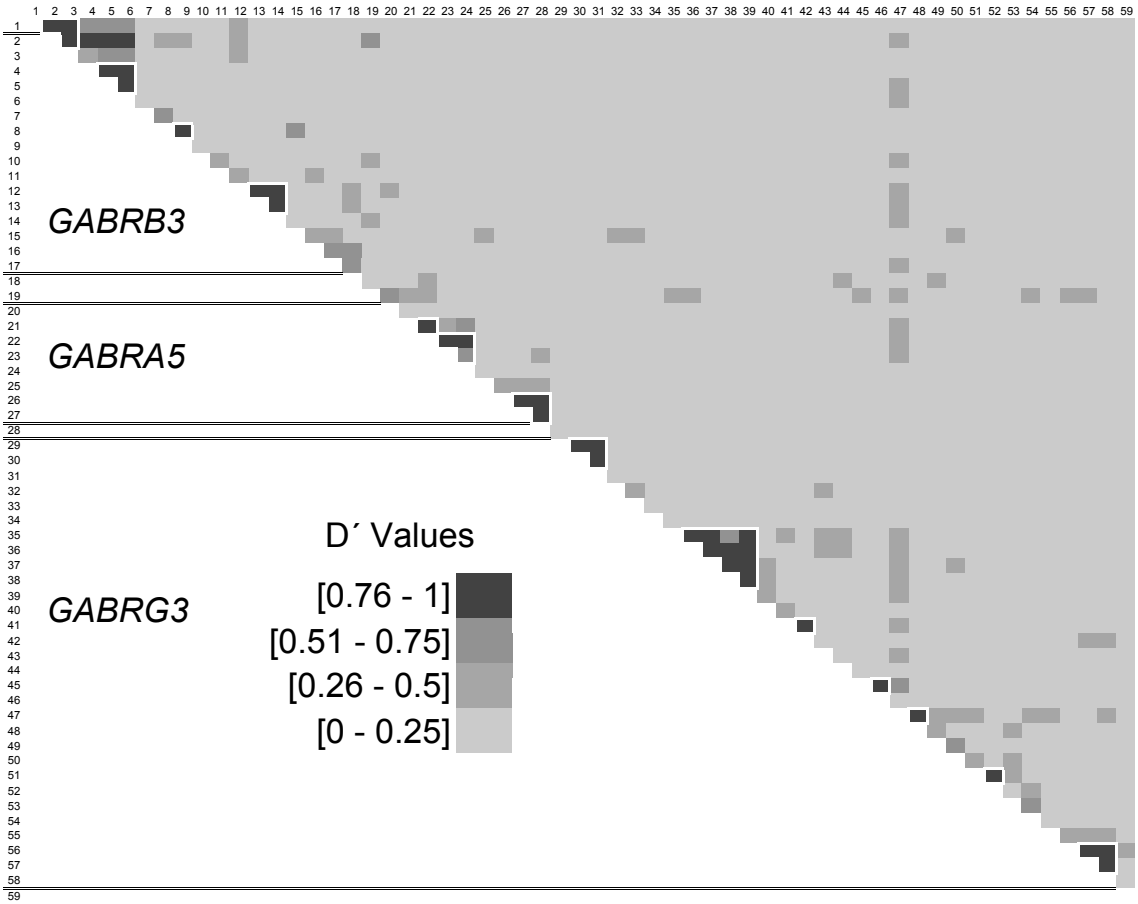


Figure 5-3. Intermarker LD and haplotype blocks in the 15q12 $GABA_A$ receptor subunit gene region. D´ values corresponding to each pair-wise SNP combination are plotted, and LD/haplotype blocks (14) are revealed as regions of high LD (black squares) along the diagonal and are outlined with white borders.

Table 5-5. TRANSMIT haplotype analysis for LD blocks across GABA cluster.

Haplotypes	Frequency	Transmissions Observed	Transmissions Expected	χ^2	TRANSMIT Global χ^2	TRANSMIT P
1-2-3						
C - C - A	0.17	93.8	87.8	1.24		
C - T - A	0.37	205.2	189.4	4.23		
G - T - T	0.45	201.0	222.8	8.01	8.02	**0.02**
4-5-6						
G - C - C	0.44	200.0	218.4	4.57		
A - T - T	0.36	195.0	182.9	2.30		
G - T - T	0.19	105.0	98.2	1.45	5.68	0.13
8-9						
A - A	0.21	104.1	106.4	0.10		
G - A	0.48	240.9	235.7	0.51		
A - G	0.30	151.9	152.3	0.01	3.39	0.34
12-13-14						
G - T - A	0.62	318.0	308.8	1.53		
A - C - G	0.16	77.0	79.5	0.18		
G - C - G	0.22	103.0	110.6	1.47	2.87	0.41
21-22						
C - A	0.41	209.9	200.8	1.45		
C - G	0.24	115.5	122.1	1.06		
T - G	0.35	171.6	173.0	0.04	2.56	0.46
22-23-24						
G - C - A	0.05	28.4	25.2	0.93		
A - T - A	0.39	199.9	194.1	0.52		
G - T - A	0.06	27.2	30.5	0.67		
G - C - C	0.37	163.6	184.9	6.93		
G - T - C	0.11	67.8	56.0	4.89	15.50	**0.01**
26-27-28						
C - C - A	0.36	171.2	179.8	1.36		
C - C - T	0.12	60.8	58.7	0.18		
C - T - T	0.06	30.4	28.2	0.39		
T - T - T	0.44	221.5	222.1	0.01	6.76	0.34
29-30-31						
A - C - G	0.44	209.8	219.8	1.77		
C - C - A	0.08	39.6	39.4	0.00		
C - T - A	0.47	244.5	234.7	1.82	1.95	0.58
35-36-37-38-39						
G - C - C - A - G	0.35	173.5	172.4	0.02		
G - C - T - A - G	0.17	78.0	81.2	0.42		
G - C - C - G - T	0.22	117.5	111.1	0.93		
A - T - C - G - T	0.25	124.0	128.7	0.50	2.38	0.79
41-42						
G - C	0.28	136.9	133.3	0.23		
G - T	0.29	156.7	151.0	0.64		
T - T	0.42	198.4	208.6	1.73	2.79	0.43
45-46						
T - A	0.33	171.9	164.2	1.00		
C - G	0.57	270.9	278.1	1.10		
T - G	0.07	36.2	37.5	0.12	1.43	0.70
47-48						
T - A	0.54	276.0	274.2	0.08		
T - G	0.46	224.0	225.8	0.08	0.80	0.78
51-52						
A - A	0.48	244.0	237.5	0.69		
A - G	0.21	104.0	103.8	0.01		
G - G	0.31	150.0	156.6	0.88	1.25	0.74
56-57-58						
G - A - C	0.38	180.0	188.3	1.17		
A - G - T	0.62	314.0	305.7	1.17	1.17	0.28

Discussion

Our results highlight the correlation between high local recombination rates, low LD and high haplotype diversity. The implication of this relationship is seen in the only modest reduction of genotyping using htSNPs and a requirement for dense SNP coverage for thorough representation of alleles in any association study. Application of this LD map to an autism dataset supports the existence of one or more risk alleles in the *GABRB3-GABRA5* region. Association was identified for a number of SNPs and haplotypes in *GABRB3*, as well as one SNP and corresponding haplotypes in *GABRA5*. These results correlate well with linkage previously reported in this region for autism [31] and phenotypic subsets of autism [53, 54].

We have described a first-generation LD map and corresponding haplotype structures for this 1-Mb autism candidate region. This is the first report to detail LD and provide dense analysis for allelic association to autism for the $GABA_A$ subunit gene cluster. The average haplotype block size (18.8 kb) is consistent with previous reports utilizing samples of European ancestry [147]. Therefore, our data generally agree with the haplotype block structure proposed for the human genome, and add to the literature of detailed LD analyses across large physical regions [147-152]. The International “HapMap” Project is working to develop such a picture of LD for the entire genome [153]. One important reason driving this effort is the argument that identification of htSNPs significantly reduces the number of markers necessary to perform genome-wide or regional

association studies. The combined minimum block lengths described here represent <50% of the DNA encoding these genes, and after identification of htSNPs, 88% of markers are necessary to represent alleles across the region.

This study reinforces the correlation between high local recombination, low LD and high haplotype diversity, noted previously by others [149, 154]. The sex-equal genetic map estimates that the rate for this interval is 4 cM Mb^{-1}, compared to a genome-wide average of 1.3 cM Mb^{-1}. Thus, our findings are not entirely unexpected. Dense SNP coverage and genotyping a large fraction of markers become prerequisites for conducting a thorough disease association study for regions of low LD. A higher resolution of SNPs, particularly in *GABRG3* and intergenic regions, will undoubtedly identify additional blocks, although a high proportion of SNPs will still be required to detect all common alleles across the region. While our study did not utilize a very dense SNP map (e.g. 5 kb^{-1}), the average minor allele frequency (0.37) was high, and markers were analyzed in a number of families sufficient to permit effective establishment of haplotype phase and structure. Some genome-wide estimates of the ability to eliminate redundant SNPs using haplotype tags [147] do not reflect the complexity of analyzing such regions of great haplotype diversity and low LD.

The findings of suggestive association in *GABRB3* occur in multiple locations within the gene. Those which cluster towards the 3′ end of the gene lie ~50 kb from the GABRB3 microsatellite marker shown to be linked [31, 54] and associated [45] by one group. It is unclear whether the GABRB3 microsatellite

marker, centromeric to our first SNP, lies within the same LD block containing SNPs 1-2-3. Two sites within the ~150-kb intron 3 showed association. SNPs 5 and 6 are in a single block located at the centromeric end of this intron. The more telomeric site, corresponding to SNP 11, lies ~80 kb centromeric to microsatellite marker 155CA-2 (see Figure 5-2), found associated to autism by two groups [44, 46]. Our own linkage studies in autism subsets have pointed to this region, and peak linkage occurs at the 5´ end of *GABRB3* at D15S511 [53], ~40 kb from 155CA-2.

Comparison of our data to that in a single published report of association analysis of SNPs in this region in autism shows only two markers in *GABRA5* and two in *GABRG3* common to both studies [144]. Neither of these markers demonstrated association in the current study, although the *GABRG3* exon 5 SNP showed nominal association in the Menold et al report. Unfortunately, a very low minor allele frequency (0.04 in our sample) for the *GABRG3* exon 5 marker substantially hampers power to detect association, and this could explain the difference. Additionally, the Menold et al report described analysis of both multiplex (91) and trio (135) families, whereas this study involves only multiplex families. In contrast to this report, the Menold et al study failed to detect association in *GABRB3* or *GABRA5*, although a much smaller number of markers (9) was examined for these genes.

The observation of association at alleles at three distinct locations within *GABRB3* as well as *GABRA5*, could be explained by the existence of multiple

autism risk alleles for these two genes. Such a scenario is consistent with the published data and a hypothesis that dup(15)-mediated autism is a contiguous gene duplication effect requiring the $GABA_A$ subunit genes in addition to the imprinted, maternally-expressed genes [155]. The relative strength of genetic effects, power to detect those effects, and potential phenotypic specificity of given genes or alleles bears consideration. Larger datasets will be required to provide power for detection of heterogeneous alleles and for analysis of phenotypic subsets. Increased evidence for linkage in families when data are co-varied for "insistence on sameness" [54] or in which affected individuals have savant skills [53] suggests a possible phenotypic specificity or bias of allelic effects in this region. While confounding clinical, locus and allelic heterogeneity can explain difficulties in detecting significant association, the absence of strong association does not allow us to exclude the possibility that one or more of the associated alleles represents a false-positive result.

While suggestive association was detected in *GABRB3* and *GABRA5*, the current study involves multiple analyses on genotype data for a large number of SNP markers, and none of these data are corrected for multiple comparisons. Therefore, given a concern over potential type I error, these association results must be interpreted cautiously. We analyzed 59 SNPs, although not all tests were independent, since a number of these markers are in LD. If we assumed 35 independent tests (one test per block and per SNP between blocks), then an adjusted significance threshold could be estimated to be 0.05/35=0.0014. Using

this criterion, none of these results are statistically significant. On the other hand, using this model, we would expect only 35x0.05=1.75 (i.e. ~2) significant ($P \leq 0.05$) results, and there is only a 10% chance of observing 4 or more independent nominally significant results in this dataset. We see six individually significant SNPs representing four sites not in LD. There is also a clustering of positive results in *GABRB3* in proximity to markers previously shown to exhibit linkage and association. These considerations argue against type I error in this case. Ultimately, replication of these findings in independent samples will be key to determination of significance for these data.

CHAPTER VI

A REPLICATION STUDY OF ALLELIC ASSOCIATION IN THE *GABRB3* AND *GABRA5* LOCI IN 943 AUTISM FAMILIES

Introduction

Evidence from Chapter V demonstrated a clustering of positive association results across the $GABA_A$ receptor subunit gene *GABRB3* in a dataset of 123 multiplex families [128]. Our current replication study examines 10 of these single-nucleotide polymorphism markers, which had previously shown nominally significant ($P \leq 0.05$) association, in a total dataset of 943 families. Autistic families included in this study were identified and ascertained in the USA, Canada, Ireland, and Portugal.

The background and logic for examining this region has been previously stated in great detail. It is our hypothesis that if the association we were detecting in our original study is central to autism, then we will be able to replicate our findings in a much larger study. Recall the previous association results fall in line with other positive linkage and association findings within this region of 15q11-q13. The relative location of the follow-up markers across the $GABA_A$ receptor subunit cluster of genes is indicated in Figure 6-1.

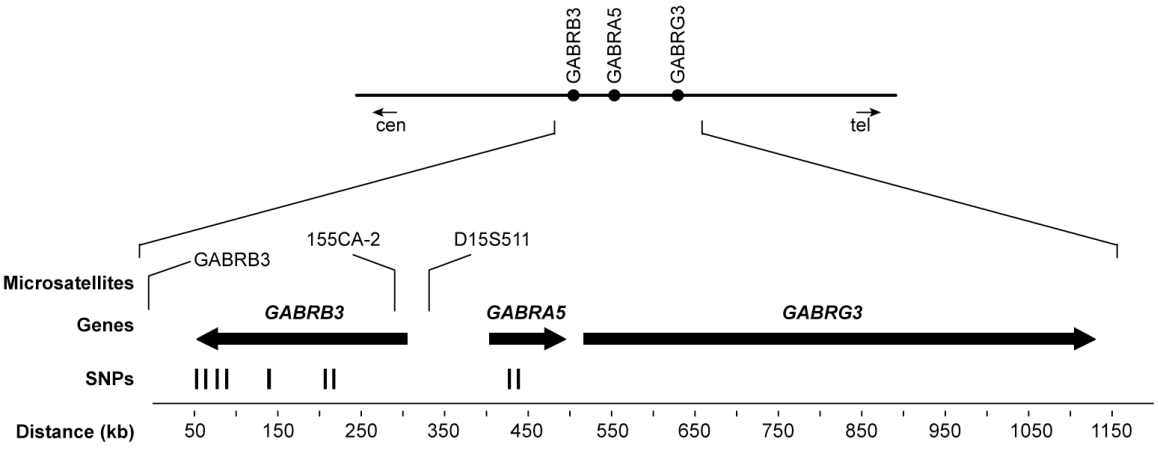


Figure 6-1. Schematic map showing markers included in the GABA replication study. Vertical hashes above the scale in kb indicate SNP markers analyzed in this study; proximity of SNPs 5/6 in relation to the scale is such that separate hashes do not distinguish these markers.

Materials and Methods

Families

The sample for this study consisted of 943 autism families. Two-hundred eighty-three (255 multiplex and 28 trio) families are part of the Autism Genetics Resource Exchange (AGRE; http://agre.org) (A number of which were obtained from the NIMH Center for Collaborative Genetic Studies on Mental Disorders (http://nimhgenetics.org/), 261 families (107 multiplex and 154 trio) were obtained through the Hospital for Sick Children, Toronto, Ontario, Canada and McMaster University, Hamilton, Canada; 85 trio families from the Trinity Centre for Health Sciences, St. James' Hospital, Dublin, Ireland; 213 families (6 multiplex and 207 trio) from the Hospital Pediatrico de Colmbra, Colmbra, Portugal; 98 families (70

multiplex and 28 trio) from Tufts/ New England Medical Center; 3 multiplex families from Vanderbilt University. The demographics of the entire sample are outlined in Table 6-1. The majority of families are of Caucasian ethnicity. IQ estimates were assessed using different measurement scales and have been omitted. Affected individuals were clinically assessed with the ADI or ADI-R and most with the Autism Diagnostic Observation Schedule (ADOS). At least one sib had to meet ADI algorithm criteria for an autism diagnosis, while additional siblings may be on the broader autism spectrum. Families were excluded from the study if probands had a known medical or neurological condition suspected to be associated with their autistic phenotype (e.g. fragile X syndrome). The procedures for clinical evaluation of affected individuals for the AGRE families have been previously described [30].

Table 6-1. Breakdown of families included in the GABA replication study.

Statistics	**Overall**	AGRE	Tufts	Vanderbilt	Canadian	Irish	Portuguese
Total Families	943	283	98	3	261	85	213
Multiplex	440	255	70	2	107	0	6
Trio	503	28	28	1	154	85	207
Affected individuals	1445	584	173	5	379	85	219
Male	1139	450	135	4	302	65	183
Female	306	134	38	1	77	20	36
Unaffected Relatives	2402	807	228	6	739	170	434
Ethnicity (%)							
Caucasian	67%	57%	69%	100%	44%	100%	95%
Hispanic-Latino	2%	4%	_	_	3%	_	_
African or African-American	2%	1%	3%	_	2%	_	3%
Asian	2%	1%	1%	_	6%	_	_
Multi-ethnic	<1%	_	_	_	1%	_	_
Other	2%	8%	_	_	_	_	_
Unknown	24%	27%	27%	_	44%	_	2%
Age at ADI (range)	2-47yrs	2-39yrs	2-47yrs	6-9yrs	2-40yrs	4-34yrs	2-15yrs

Molecular Analyses

SNPs from the dbSNP and Celera SNP databases were previously selected based on their map position, minor allele frequency, and, in one case, a finding of allelic association [144]. This replication study focuses on ten of those markers demonstrating association in our original analysis. Database reference numbers and other details for the markers studied are cited in Table 6-2. Marker and exon locations and intermarker distances are based on the public (UCSC; July 2003 freeze; http://genome.cse.ucsc.edu/) and Celera (http://www.celeradiscoverysystem.com/index.cfm) assemblies and published gene structure information for *GABRB3* and *GABRA5* [145, 146]. Individual SNPs were genotyped using the TaqMan™ protocol (Applied Biosystems). Assays-By-Design™ primers and probes are listed in Table 6-3. Information on markers for which Assays-On-Demand™ were obtained from Applied Biosystems (Foster City, CA) is proprietary.

For TaqMan™ genotyping assays, reactions were performed in a 5 μl volume, employing 2.5ng genomic DNA template, according to manufacturer's recommendations (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 92 °C for 15 s, 60 °C for 1 min. Samples were analyzed using an ABI 7900HT™ Sequence Detection System. The genotyping of the Canadian samples was completed, by our collaborators at the Hospital for Sick Children, using the TaqMan™ protocol.

Table 6-2. SNP markers for the GABA follow-up study. Alleles are listed major/minor. Allele frequencies are based on the entire dataset of 943 families. Flanking microsatellite markers are highlighted in bold. Note: allele frequencies across collection sites were similar for all markers.

Gene	SNP No.	Region	dbSNP rs# /Celera hCV#	Alleles	Minor Allele Frequency	Intermarker Distance (kb)
Intergenic		Intergenic	**GABRB3**			52.9
Intergenic	1	Intergenic	rs8025575/ hCV2911914	G/C	0.47	10.1
GABRB3	2	Intron 8	rs2081648/ hCV2911917	T/C	0.14	12.5
	3	Intron 7	rs1432007/ hCV8866669	A/G	0.48	10.4
	4	Intron 6	rs1426217/ hCV2901088	A/G	0.46	46.3
	5	Intron 3	rs2873027/ hCV2901140	C/T	0.46	0.5
	6	Intron 3	rs4542636/ hCV2901143	C/T	0.46	61.3
	7	Intron 3	rs878960/ hCV8865198	T/C	0.46	9.6
	8	Intron 3	rs1863464/ hCV2901200	G/A	0.20	70.3
		Intron 3	**155CA-2**			39.1
Intergenic		Intergenic	**D15S511**			84.9
GABRA5	9	Intron 6	hCV252720	T/C	0.42	9.8
	10	Intron 6	rs9745027/ hCV27725	C/A	0.48	–

Table 6-3. PCR and genotyping primers for the GABA follow-up SNPs.
Labels are as follows: (F) Forward Primer, (R) Reverse Primer, VIC and FAM are fluorescent labels for TaqMan allelic discrimination probes (the SNP is indicated by bold text), AbD is an Assay-By-Design from Applied Biosystems. MGB stands for minor groove binder, and NFQ stands for non-fluorescent quencher.

SNP No.		Primer sequences (5´-3´)	Product size (bp)	Assay
1	F	AGAAATGTAAAAAACTTGTCCAAACAATGGAA	112	TaqMan (AbD)
	R	ACCTCCAAATTTCCTAACTGTTAAGCAA		
	VIC	VIC-CCTGTAAA**G**AATTCC-MGB-NFQ		
	FAM	FAM-CCTGTAAA**C**AATTCC-MGB-NFQ		
2	F	CATGAAAAGGGATTTGATAAATTGAGGTCAT	133	TaqMan (AbD)
	R	ACTCAACATCAGATAGATTTTAAACATATAAGCTTACC		
	VIC	VIC-ACTGTTT**G**AGATGCTG-MGB-NFQ		
	FAM	FAM-AACTGTTT**A**AGATGCTG-MGB-NFQ		
3	F	AGCCAGCCCATGCCTTTATC	86	TaqMan (AbD)
	R	GATTCCACTTTTTCTTCATGACAGCAT		
	VIC	VIC-AAAGCCAC**A**GAGGCA-MGB-NFQ		
	FAM	FAM-AAAGCCAC**G**GAGGCA-MGB-NFQ		
4	F	ATCTGTTCTCCACATATCTTGCGAAA	101	TaqMan (AbD)
	R	TGGGATATACAGTTTTTGCCAAATCTGA		
	VIC	VIC-AAAGACAACTCT**G**AAGTGA-MGB-NFQ		
	FAM	FAM-CAAAGACAACTCT**A**AAGTGA-MGB-NFQ		
5	F	GGCTTGCCCACAAAATTACCA	68	TaqMan (AbD)
	R	AACTAACTGCTTTGGCAAAACGAAA		
	VIC	VIC-CCATCACGT**A**GATATT-MGB-NFQ		
	FAM	FAM-CATCACGT**G**GATATT-MGB-NFQ		
6	F	TTGGAATATAGAGGCTTAGGGACTGA	82	TaqMan (AbD)
	R	CTGGTTTTGTTTCCCCTTCTTGTG		
	VIC	VIC-ATCTGAAGACAAATC**A**ATGT-MGB-NFQ		
	FAM	FAM-ATCTGAAGACAAATC**G**ATGT-MGB-NFQ		
7	F	ACAAACAACCTGTTAGCCACTCTAA	98	TaqMan (AbD)
	R	ATGCCATTCTTTCCTTTTCCACAAG		
	VIC	VIC-CTCATGAGTGTAT**A**AGAGTGT-MGB-NFQ		
	FAM	FAM-ATGAGTGTAT**G**AGAGTGT-MGB-NFQ		
8	F	GCTGGCTCAGCGTTCCTA	73	TaqMan (AbD)
	R	CCTCTCTTATGTTTCCTCTTGCTGTT		
	VIC	VIC-CAGTCTCTAAAC**A**TTTC-MGB-NFQ		
	FAM	FAM-CAGTCTCTAAAC**G**TTTC-MGB-NFQ		
9	F	TGGCTATGCAAACTACTGGTGAAAT	80	TaqMan (AbD)
	R	GGACCCTGGCACTGAAGTG		
	VIC	VIC-ATTCGTGCTTT**G**GTGATT-MGB-NFQ		
	FAM	FAM-ATTCGTGCTTT**A**GTGATT-MGB-NFQ		
10	F	TGGAAAGAGAGGTCCCTTCACT	70	TaqMan (AbD)
	R	GCTGCTTTGGCTGGTTAAAATTCA		
	VIC	VIC-CGCTCATGTATTC**T**CTATAAT-MGB-NFQ		
	FAM	FAM-CTCATGTATTC**G**CTATAAT-MGB-NFQ		

Statistical Analyses

Initial analysis of SNP genotype data involved quality control checks consisting of verification of internal controls and assessment of Mendelian inconsistencies, followed by final haplotype consistency analyses using Simwalk2 [117]. Conformity with anticipated Hardy-Weinberg equilibrium expectations was established. Transmission disequilibrium (TD) in autism families was determined using the pedigree disequilibrium test (PDT) statistic, developed for use with general pedigrees [119]. Maternal and paternal specific transmissions were also assessed using the T_{SP} test [156]. Common haplotypes (≥5%) were identified and analyzed using FBAT [130]; analysis of TD was performed using a sliding window approach of adjacent SNP pairs and haplotype tag SNPs for other multi-locus blocks. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$.

Results

We find little evidence of significant association in our total replication dataset of 943 families. We examined for both single marker associations and multi-marker haplotype associations in the total dataset as well as a number of subsets. We individually tested for association within the Canadian, Irish, Portuguese, and our sample (AGRE, Tufts, Vanderbilt) subsets. Additional subsets included multiplex families only, trio families only, families with male-

affected individuals only, and families containing female-affected individuals only (Table 6-1).

Given the relevance of genomic imprinting across the 15q11-q13 region, we examined the overall dataset for evidence of parental-specific transmission to affected individuals using the T_{SP} test. One marker was nominally significant in the overall dataset when transmitted from the father (hCV252720). This marker demonstrated greater significance, while another marker (hCV27725) became nominally significant when we examined the families containing only male-affected individuals (Table 6-4). These two markers located within intron 6 of the *GABRA5* gene are in strong linkage disequilibrium with one another (data previously shown). Recall from previous discussion the examination of the affected male only family subset may be a valid approach to investigate the sexually dimorphic findings in autism, and potentially the male to female ratio of affected individuals. The affected-female-containing families did not show any single marker associations. However, this is not surprising given the greatly decreased sample size and resulting loss in power. We also examined each sample based on the location of ascertainment and found no association (data not shown). The allelic transmission counts for the two significant makers (hCV252720 and hCV27725) in the overall dataset are found in Table 6-5.

Table 6-4. PDT and T_{sp} analysis of GABA follow-up markers.

			Overall Sample						Affected Male-Only Families					
			Total			Multiplex	Trio	Original Sample	Total			Multiplex	Trio	Original Sample
			943 Families			440 Families	503 Families	123 Multiplex Families	677 Families			265 Families	406 Families	75 Multiplex Families
Gene	SNP No.	Celera hCV#/ dbSNP rs#	PDT (P)	Tsp Father (P)	Tsp Mother (P)	PDT (P)	PDT (P)	Tsp Father (P)	PDT (P)	Tsp Father (P)	Tsp Mother (P)	PDT (P)	PDT (P)	Tsp Father (P)
	1	hCV2911914/ rs8025575	0.42	0.18	0.86	0.27	0.88	**0.01**	0.91	0.67	0.89	0.67	0.73	0.11
GABRB3	2	hCV2911917/ rs2081648	0.52	0.57	0.67	0.78	0.46	0.75	0.19	0.61	0.34	0.49	0.19	0.32
	3	hCV8866669/ rs1432007	0.34	0.70	0.96	0.26	0.96	**0.03**	0.69	0.90	0.84	0.75	0.81	0.12
	4	hCV2901088/ rs1426217	0.73	0.82	0.95	0.50	0.73	0.82	0.55	0.40	0.49	0.54	0.84	0.77
	5	hCV2901140/ rs2873027	0.34	0.45	0.27	0.74	0.22	0.44	0.20	0.60	0.27	0.45	0.26	0.69
	6	hCV2901143/ rs4542636	0.46	0.39	0.44	0.69	0.47	0.35	0.52	0.95	0.47	0.68	0.61	0.79
	7	hCV8865198/ rs878960	0.26	0.31	0.62	0.13	0.83	0.41	0.21	0.14	0.90	0.18	0.77	0.08
	8	hCV2901200/ rs1863464	0.86	0.63	0.73	0.74	0.88	0.59	0.34	0.14	0.93	0.54	0.46	0.09
GABRA5	9	hCV252720	0.50	**0.05**	0.46	0.47	0.91	0.13	0.43	**0.02**	0.22	0.29	0.89	0.19
	10	hCV27725/ rs9745027	0.33	0.16	0.87	0.45	0.54	0.83	0.23	**0.04**	0.90	0.22	0.71	0.57

Table 6-5. T_{sp} allelic transmission data for GABA follow-up significant markers. The number of fathers represents the number of heterozygous fathers used for each test. The number of transmissions (T) refers to the number of times a father, heterozygous for that marker, transmitted that particular allele to an affected child. The number of non-transmissions (NT) refers to the number of times a father, heterozygous for that marker, did not transmit that particular allele to an affected child.

	Marker	Paternal Transmissions									
		Tsp Father (P)	Father χ^2	Triads				Sib-Pairs			
				No. of Fathers	Allele	T	NT	No. of Fathers	Allele	T	NT
Overall Sample (943 Families)	hCV252720	**0.05**	3.84	127	C	60	67	93	C	79	107
					T	67	60		T	107	79
	hCV27725/ rs9745027	0.16	1.97	124	C	58	66	109	C	100	118
					A	66	58		A	118	100
Affected Male Only (677 Families)	hCV252720	**0.02**	5.35	92	C	42	50	53	C	40	66
					T	50	42		T	66	40
	hCV27725/ rs9745027	**0.04**	4.27	96	C	42	54	67	C	57	77
					A	54	42		A	77	57
Affected Female Containing (266 Families)	hCV252720	0.92	0.01	35	C	18	17	40	C	39	41
					T	17	18		T	41	39
	hCV27725/ rs9745027	0.56	0.35	28	C	16	12	42	C	43	41
					A	12	16		A	41	43

Discussion

We were unable to replicate our previous findings of association in this larger dataset of families. There are multiple reasons for this lack of replication. First, we tested a very specific hypothesis; namely that a common variant at this locus is involved in autism risk. However, allelic heterogeneity may play an important role at this locus. The addition of more families, which may or may not share the involvement of this locus in their autism susceptibility, may cause the association to be non-significant. Secondly, our initial results could reflect false positive findings. Our findings do not rule out allelic heterogeneity within or across our sample populations.

In an exploratory analysis, we do find evidence of association within the *GABRA5* gene in the overall subset when we examine preferential transmission of alleles from the father. We detected a nominally significant over-transmission of the T allele for hCV252720 from fathers to affected individuals in our overall dataset of 943 families. This over-transmission is more significant when we examine the families only containing male-affected individuals. When we examine our original dataset of 123 families for this effect within the entire dataset and the male-only dataset we find no evidence of association with these two markers, however be do detect nominal significance with two other markers, but this significance is not found in the male-affected-only families. If the preferential transmission from fathers to male-affected individuals is indeed a real effect, we may have lacked sufficient power to detect this in our original study.

The subset of male only families in the original dataset was relatively small in comparison to the completed replication dataset (77 families and 677 families respectively).

We have not corrected any of our results for multiple testing, so we can only speculate on their true significance. Any correction would cause these results to be non-significant. Future investigations into the apparent transmission biased within intron 6 of *GABRA5* may warrant variant screening of this region for other putative variation, which may prove functional.

Other approaches would include the examination of relevant phenotypic subsets. Recent reports, including one from our group, demonstrated that subsetting autism families based on variables from the Autism Diagnostic Interview-Revised (ADI-R) results in significantly increased evidence for linkage to the GABA region of chromosome 15 [53, 54]. By using more sophisticated methods for examining quantitative traits, we could test the hypothesis that the $GABA_A$ receptor subunit genes on chromosome 15 exert their effects within specific phenotypic traits across the broader autism phenotype. One such tool for association studies of quantitative traits is the QTDT program [157].

Our lack of replication may suggest an interaction effect across these loci. Examination of gene-gene interaction by way of allelic interaction studies between these genes, whose proteins are known to biological interact, may prove useful. Using methods such as the multifactor dimensionality reduction (MDR)

method, developed by Moore and colleagues, to examine allelic interaction will be essential to future studies within this region [72, 73].

CHAPTER VII

ASSOCIATION ANALYSIS ACROSS THE IMPRINTING CENTER AND MATERNAL EXPRESSION DOMAIN IN 15q11.2

Introduction

An interval in proximal chromosome 15q, corresponding to 15q11-q13, has been targeted as a candidate region based on observations of duplications in a small percentage of autism cases (illustrated in Figure 7-1). These duplications are the most frequent chromosomal abnormality in autism, occurring in an estimated 1-3 percent of cases (reviewed in [155]) [49, 50]. Duplications of 15q11-q13 are seen in two forms: (1) interstitial, tandem duplication of an approximately 5-Mb segment or (2) supernumerary isodicentric, inverted, duplicated chromosomes 15 (so-called idic(15) marker chromosomes). Idic(15) marker chromosomes are almost always derived from maternal chromosomes, possibly due to maternal meiotic nondisjunction events. In contrast, interstitial duplication of 15q11-q13 may arise on a maternal or paternally-derived chromosome.

Proximal 15q duplications often present clinically with an autistic phenotype. Compared to interstitial dup(15), idic(15) duplications contain *two* additional copies of a larger interval, extending farther in a telomeric direction and including an additional 1-2 Mb of DNA. These cases typically present with a more severe phenotype, related to gene dosage and/or degree of chromosomal

imbalance. Interstitial dup(15) can occur on either parentally-derived chromosome, although studies suggest a greater correlation with dup(15) and autism when the duplication arises on a maternally-derived chromosome. One detailed study showed that seven of ten such maternal interstitial duplications were associated with an autism phenotype, while only a few cases of paternal dup(15) are seen in the context of an autism diagnosis. Thus there appears to be significant bias in parent-of-origin of a dup(15) with the risk of autism, or more broadly, the autism spectrum. This phenomenon should be understood in the context of two other genomic disorders involving chromosomal abnormalities of this region.

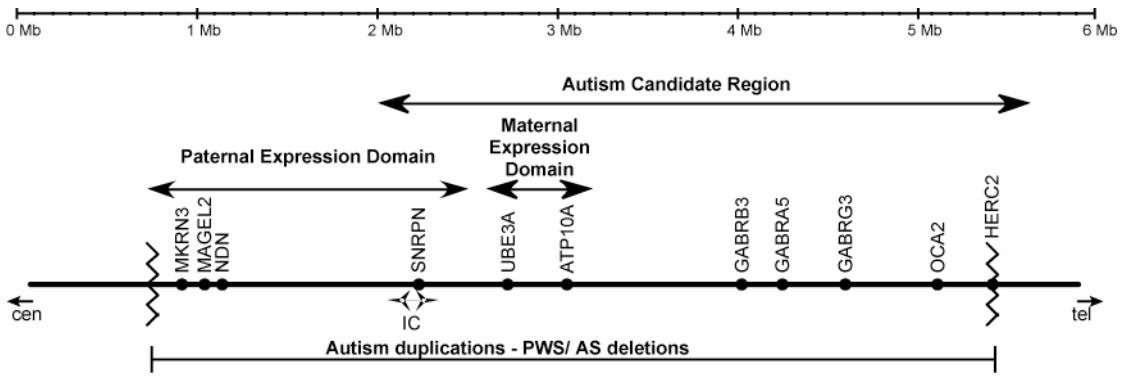


Figure 7-1. Schematic of the 15q11-q13 region.

Interstitial *deletions* of 15q11-q13 affecting the same 5-Mb interval involved with interstitial dup(15) are the major cause of the two other disorders: Prader-Willi Syndrome (PWS; OMIM# 176270) or Angelman Syndrome (AS; OMIM# 105830), depending on the parental origin of the sporadic deletion event.

Paternal deletions give rise to PWS, while maternal deletions give rise to AS; this reflects the corresponding loss of paternal-specific gene expression in PWS and maternal-specific gene expression in AS (reviewed in [142, 158]). Interstitial deletion in PWS and AS and duplication in cases of autism represents reciprocal meiotic products of unequal meiotic cross-over events mediated by large sequence duplicates, or so-called duplicons, located at either end of the deletion interval [159, 160]. In addition to deletion, occurring in 65-70% of PWS and AS cases, maternal uniparental disomy (UPD) accounts for ~20-25% of PWS cases and paternal UPD for ~2-3% of AS cases.

Underscoring the epigenetic and specifically genomic imprinting phenomena occurring in this region, imprinting mutations occur in ~5-7% of both disorders, typically involving microdeletions (e.g. 100 kb or less), and point to a region extending from the 5' end of the paternally-expressed gene *SNRPN* (small nuclear ribonucleoprotein polypeptide N) as harboring the 15q11-q13 imprinting center (IC; [161, 162]); this empirically-determined functional locus has an apparent bipartite structure and is required for normal control of imprinted gene expression in the region [163]. Specifically, the IC is required to facilitate the switch (if necessary) during gametogenesis in defining the appropriate maternal or paternal identity of the inherited chromosome (reviewed in [155]). Small deletions or other unidentified mutations in this region result in a failure of the imprinting switch process, subsequently leading to inappropriate expression patterns of imprinted loci. A recent study in 58 autism families by Jiang and

colleagues described excess paternal compared with maternal allele sharing, which appeared maximized closer to the IC [22].

A number of paternally-expressed transcripts are located within the proximal third of the PWS/AS deletion interval [142, 158]. While evidence points towards involvement of a gene termed Necdin (gene symbol: *NDN*; [164]) as being a significant contributor to the PWS phenotype [165, 166], there remains a lack of certainty as to which gene(s) contribute to the human condition.

Identification of maternal, *de novo*, loss-of-function mutations in the E6-AP ubiquitin-protein ligase gene (*UBE3A*) [167, 168], identified this locus as being responsible for the AS phenotype. *UBE3A* is located telomeric from the IC and *SNRPN*, is adjacent to another maternally-expressed gene termed *ATP10A* (previously *ATP10C*) that encodes a putative aminophospholipid translocase possibly involved in CNS signaling [169], and is thus located within a domain of maternally-expressed genes; we have termed this the maternal expression domain (MED; [47]).

Clinically, PWS and AS are quite distinct, but of relevance is the presence of features in common with autism in both disorders. PWS is characterized by infantile hypotonia, failure to thrive, and feeding difficulties. By two years of age, PWS patients develop hyperphagia and subsequent obesity, the most obvious aspect of the phenotype. People with PWS display mild to moderate mental retardation and physical findings that include decreased stature, small hands and feet, almond-shaped eyes, and hypogonadism [170]. Behavioral abnormalities are common, and include aggression, self-abuse, preoccupation with ordering

and arranging, resistance to change in daily routines, and food foraging. With the exception of food-related behaviors, these other behavioral features are things frequently seen in autism. AS is more severe, with profound mental retardation, absent speech, epilepsy, ataxic gait, hand-flapping, and inappropriate laughter. Motor stereotypes (such as hand-flapping), seizures or actual epilepsy, severely compromised language skills, and poor coordination are also features found within the autism population [6]. Therefore, with both of these very different disorders, there is phenotypic overlap or commonality with autism. An outstanding question, therefore, is whether these overlapping commonalities reflect genetic variants or deficiencies present in the idiopathic autism population.

The bias of maternal dup(15) in association with autism, has been interpreted by some as a significant indicator of *maternal UBE3A* (and *ATP10A*) over-expression (two copies in the case of maternal interstitial duplications) as being the sole or primary factor causing (or conferring substantial risk for) autism. If this were the case one would predict that maternal UPD PWS cases would be "more autistic". While one published report attempts to make this argument [171], the data in support of this premise are mostly anecdotal and not compelling. Thus in the absence of information clearly documenting that maternal UPD, and documenting that two expressed copies of *UBE3A* and *ATP10A* are necessary and sufficient to result in autism or a higher rate of autistic traits, we propose that dup(15)-autism is the result of a contiguous gene duplication defect within 15q11-q13 rather than a simple gene-dosage effect involving the maternally expressed genes.

The argument proposing a contiguous gene duplication effect in genetic risk in dup(15) autism cases, points to the cluster of $GABA_A$ receptor subunit genes (examination of these genes is described in Chapters V and VI), for which better genetic evidence exists suggesting a role in idiopathic autism. However, it in no way excludes the hypothesis that epigenetic dysfunction, which would be very difficult to detect by traditional analytical methods, plays a role in conferring risk for autism. In fact, a recent report documents reduced *UBE3A* and *GABRB3* expression in *MECP2*-deficient brain tissue [172]. This supports hypothetical involvement of MeCP2, the methyl-DNA-binding protein deficient in Rett Syndrome (another Pervasive Developmental Disorder), in facilitating paternal silencing of *UBE3A* by acting *in trans* at the differentially-methylated 15q IC or elsewhere in the region. It also raises the specter that epigenetic gene silencing in 15q11-q12 extends beyond *UBE3A* and *ATP10A*. Paternal silencing of *UBE3A* in the brain is speculated to involve a paternal antisense transcript (ATS), promoting from the *SNRPN* promoter and transcribed through the MED [173]. The antisense orientation relative to *UBE3A* and *ATP10A* (and possibly *GABRB3* as well) is thought to promote silencing via RNA interference mechanisms.

The purpose of this study is to focus on the MED and corresponding imprinting center to test the hypotheses that a common allele or set of alleles in this region might (1) be involved in causing abnormal epigenetic effects or imprinting regulation or (2) affect one of the maternally-expressed genes directly. To test these hypotheses, we examined SNPs and haplotypes across the region

for evidence of transmission disequilibrium generally, or in a parental-specific manner in autism families.

Materials and Methods

Families

The sample for this study consisted of 384 autism families (327 multiplex and 57 trio families). Two-hundred eighty-three families were obtained from the Autism Genetics Resource Exchange (AGRE; http://agre.org), 98 multiplex and trios from Tufts/ New England Medical Center, and 3 multiplex families from Vanderbilt University. The majority of families are Caucasian. All affected individuals were at least four years of age and were clinically assessed with the ADI or ADI-R and most with the Autism Diagnostic Observation Schedule (ADOS). At least one sib had to meet ADI algorithm criteria for an autism diagnosis, while additional siblings may be on the broader autism spectrum. Families were excluded from the study if probands had a known medical or neurological condition suspected to be associated with their autistic phenotype (e.g. fragile X syndrome). The procedures for clinical evaluation of affected individuals for the AGRE families have been previously described [30].

Molecular Analyses

DNA from Tufts and Vanderbilt samples was isolated from peripheral blood or lymphoblastoid cells using the PureGene kit according to the manufacturer's recommendations (Gentra Systems, Minneapolis, MN). A number of our samples were or obtained from the NIMH Center for Collaborative Genetic Studies on Mental Disorders (http://nimhgenetics.org/), involving a contract repository at Rutgers University. SNPs from the dbSNP and Celera SNP databases were selected based on their map position, and minor allele frequency, and potential for having a functional effect. Visualization Tools for Alignment (VISTA) analysis was performed via web-based submission (http://www.gsd.lbl.gov/vista/) of human and mouse genomic sequence [121]. Regions showing evolutionarily conserved non-coding sequence were identified as an additional means of identifying regions wherein variants were selected for examination. The reasoning is that conserved regions are more likely to harbor functionally significant (e.g. regulatory) sequences. Therefore, such regions are specifically targeted to provide the ability to represent potential unknown functional (but non-coding) variation. Marker and exon locations and intermarker distances are based on the public (Ensembl; Dec 2004 release 27.35a.1; http://www.ensembl.org) and Celera (http://www.celeradiscoverysystem.com/index.cfm) assemblies. Database reference numbers and other details for the markers studied are cited in Table 7-1. All SNPs were genotyped using the TaqMan™ protocol (Applied Biosystems,

Foster City, CA). PCR primers and probes for assays are listed in Table 7-2. This information is proprietary for the majority of markers, for which Assays-On-Demand™ were obtained from Applied Biosystems (Foster City, CA).

For TaqMan™ genotyping assays, reactions were performed in a 5 μl volume, employing 2.5ng genomic DNA template, according to manufacturer's recommendations (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 92 °C for 15 s, 60 °C for 1 min. Samples were analyzed using an ABI 7900HT Sequence Detection System.

Table 7-1. SNP markers spanning the Imprinting Center (IC) and Maternal Expression Domain (MED) region on 15q11-q13. Alleles are listed major/minor.

Gene	SNP No.	Region	dbSNP rs# /Celera hCV#	Alleles	Minor Allele Frequency	Intermarker Distance (kb)
Intergenic	1	Intergenic	rs12902137/ hCV1796103	T/C	0.27	5.7
Intergenic	2	Intergenic	rs12905620/ hCV9710866	T/C	0.24	49.9
Intergenic	3	Intergenic	rs975907/ hCV8378225	A/G	0.41	20.6
Intergenic	4	Intergenic	rs1463292/ hCV8378415	A/G	0.42	32.9
Intergenic	5	Intergenic	rs3913224/ hCV7516185	T/C	0.18	30.6
Intergenic	6	Intergenic	rs11161139/ hCV1244494	C/T	0.38	7.8
SNRPN	7	Promoter	hCV1244509	T/G	0.19	21
	8	Intron 2	rs736008/ hCV2979338	T/C	0.44	16.6
	9	Intron 2	rs5001649/ hCV2979365	T/C	0.29	17.7

	10	Intron 2	rs2047433/ hCV2979396	G/T	0.39	10.9
	11	Intron 3	rs11634496/ hCV2979419	C/A	0.30	8.1
	12	Intron 3	rs8037745/ hCV1223678	A/G	0.39	7.6
	13	Intron 3	rs4906695/ hCV1223674	G/C	0.17	10.4
	14	Intron 4	rs2736708/ hCV16287658	A/G	0.21	34.4
	15	Intron 4	rs220030/ hCV3214754	T/C	0.36	0.5
	16	Intron 4	rs220029/ hCV3214752	G/A	0.08	9.9
	17	Intron 5	rs220034/ hCV1025117	G/A	0.09	5.9
	18	Intron 6	rs2554428/ hCV2066559	C/T	0.34	3.4
	19	Exon 7 (5' UTR)	rs705/ hCV2066555	T/C	0.46	101.5
Intergenic	20	Intergenic	rs4906699/ hCV2066488	T/C	0.43	3.6
Intergenic	21	Intergenic	rs11161166/ hCV2066485	T/A	0.45	7.5
Intergenic	22	Intergenic	rs1549478/ hCV2066478	C/T	0.44	20.1
Intergenic	23	Intergenic	rs7162559/ hCV152343	G/A	0.34	127
Intergenic	24	Intergenic	rs2714758/ hCV132223	A/G	0.06	44.4
Intergenic	25	Intergenic	rs1977036/ hCV422420	C/T	0.13	4.8
Intergenic	26	Intergenic	hCV422410	T/C	0.10	47.2
Intergenic	27	Intergenic	rs4906951/ hCV2625778	T/C	0.28	4.8
UBE3A	28	Intron 11	hCV11487073	G/A	0.12	19.9
	29	Intron 9	rs12907375/ hCV2558359	A/G	0.26	5.8
	30	Intron 6	rs10162823/ hCV2558365	G/A	0.14	11.8
	31	Intron 5	rs4906708/ hCV11487149	C/T	0.26	21.2
	32	Intron 4	rs2340625/ hCV2558398	C/G	0.22	32.1
	33	Intron 1	rs7496951/ hCV2558422	G/C	0.27	8.6
		Intron 1	**D15S122**			2.5

	34	Intron 1	hCV2558441	G/A	0.13	2.8
	35	Promoter	rs2526025/ hCV2558443	C/A	0.18	25.7
Intergenic	36	Intergenic	rs969860/ hCV8766820	C/G	0.13	36.4
Intergenic	37	Intergenic	rs4906719/ hCV28014321	G/T	0.47	0.5
Intergenic		Intergenic	**D15S210**			42.3
Intergenic	38	Intergenic	rs2189713/ hCV2564743	T/C	0.41	54.6
Intergenic	39	Intergenic	rs2925280/ hCV15943093	T/C	0.42	10.5
Intergenic	40	Intergenic	rs2930599/ hCV26647614	G/A	0.32	26.1
Intergenic	41	Intergenic	rs1385388/ hCV431962	T/C	0.30	39.6
Intergenic	42	Intergenic	rs7181116/ hCV454609	T/C	0.33	3.9
ATP10A	43	Exon 22	rs1047700/ hCV1157519	T/C	0.18	1.6
	44	Exon 19	rs2076745/ hCV15863547	T/C	0.01	3.1
	45	Intron 17	rs8041681/ hCV1157511	A/G	0.31	7.8
	46	Intron 15	rs2066705/ hCV12080790	T/C	0.36	16.5
	47	Exon 12	rs2066704/ hCV12080806	C/T	0.08	8.6
	48	Exon 10	rs2066703/ hCV12080811	G/A	0.06	2
	49	Intron 8	rs2014053/ hCV320540	T/C	0.47	19.4
	50	Intron 3	rs4906629/ hCV394756	G/A	0.19	4.8
	51	Intron 3	rs12901627/ hCV394762	C/T	0.38	21.6
	52	Intron 3	rs11161217/ hCV125931	A/C	0.45	7.4
	53	Intron 3	rs7165728/ hCV24876	T/C	0.22	7
	54	Intron 3	rs2044311/ hCV12064096	A/T	0.25	0.9
	55	Intron 3	rs8038726/ hCV232863	G/A	0.20	2.6
	56	Intron 2	rs11632263/ hCV11300397	C/T	0.49	2.9

	57	Intron 2	rs11636393/ hCV399638	C/T	0.39	6.3
	58	Intron 2	rs12439329/ hCV11300400	A/T	0.49	10.4
	59	Intron 2	rs872537/ hCV8864618	A/T	0.42	0.4
	60	Intron 2	rs1867511/ hCV9400334	A/G	0.22	9.1
	61	Intron 2	rs1345098/ hCV8864621	T/G	0.45	0.4
	62	Intron 2	rs1345099/ hCV9400402	A/G	0.49	10.6
	63	Intron 2	rs11161232/ hCV423475	C/T	0.24	7.6
	64	Intron 2	rs11630555/ hCV504244	T/C	0.44	11.3
	65	Intron 2	rs11633552/ hCV60097	C/G	0.17	11.4
	66	Intron 2	hCV402689	T/C	0.18	10.6
	67	Intron 1	rs2076749/ hCV15863526	A/T	0.26	1.3
	68	Promoter	rs1444621/ hCV8864860	C/T	0.07	29.6
Intergenic	69	Intergenic	rs4906642/ hCV1591365	A/T	0.37	42
Intergenic		Intergenic	**D15S1513**			34
Intergenic	70	Intergenic	rs1397855/ hCV8866740	C/T	0.10	2.3
Intergenic	71	Intergenic	rs1511493/ hCV8866723	G/A	0.45	2.1
Intergenic		Intergenic	**D15S540**			77.4
Intergenic	72	Intergenic	rs1403590/ hCV8865982	T/C	0.43	80.1
Intergenic	73	Intergenic	rs693798/ hCV578309	G/A	0.08	23.4
Intergenic	74	Intergenic	rs2030601/ hCV11669761	C/T	0.45	79.6
Intergenic	75	Intergenic	rs1435831/ hCV8864717	C/T	0.39	57.6
Intergenic	76	Intergenic	rs4906872/ hCV240706	A/C	0.37	52.5
Intergenic	77	Intergenic	rs3922665/ hCV26111547	T/C	0.47	63.4
Intergenic	78	Intergenic	rs4906673/ hCV2911753	A/G	0.38	81
Intergenic		Intergenic	**GABRB3**			_

Table 7-2. PCR and genotyping primers for the IC-MED region SNPs. Labels are as follows: (F) Forward Primer, (R) Reverse Primer, VIC and FAM are fluorescent labels for TaqMan allelic discrimination probes (SNPs indicated by bold text), AbD is an Assay-By-Design from ABI, AoD is an Assay-On-Demand from ABI. Sequence for primers and probes for AoD assays are proprietary information. MGB stands for minor groove binder, and NFQ stands for non-fluorescent quencher.

SNP No.	Primer sequences (5´-3´)		Product size (bp)	Assay
*	NA	Unavailable	< 200	TaqMan (AoD)
1	F	CCTCTCCAGAACTCAGGTTGTG	109	TaqMan (AbD)
	R	GATGCACCACCTGTGTTTTGTC		
	VIC	VIC-ATGTGTACTGAC**G**GAGAA-MGB-NFQ		
	FAM	FAM-ATGTGTACTGAC**A**GAGAA-MGB-NFQ		
2	F	CATCCTGGCACCATTAACCATCA	82	TaqMan (AbD)
	R	CAAATGCTAGTCTATGTAGTTTGCATGAA		
	VIC	VIC-TTACTGCATA**C**AGAGAGC-MGB-NFQ		
	FAM	FAM-TTTTACTGCATA**T**AGAGAGC-MGB-NFQ		
3	F	GTGTATGCATAGCACCTGTAGGA	85	TaqMan (AbD)
	R	CCTTTTAGATCCATCTTAAATCTTTGATTCACTG		
	VIC	VIC-CCAGTGGA**C**GACCTT-MGB-NFQ		
	FAM	FAM-CCAGTGGA**T**GACCTT-MGB-NFQ		
4	F	GCTCGCAATACAAAGGAAAAATGTGT	106	TaqMan (AbD)
	R	ACTGTGTATAGGTCAATTATGATCTTCCCT		
	VIC	VIC-TCGATAATTAAACTCC**A**TTTGTA-MGB-NFQ		
	FAM	FAM-CGATAATTAAACTCC**G**TTTGTA-MGB-NFQ		
5	F	CAACTTCAAGCTTTCCTGGTAATATTTAGTC	114	TaqMan (AbD)
	R	TGAGATTAAAAAAAGAAAACTGCAAACAAATGTT		
	VIC	VIC-CCAAAACAA**A**CAAACAA-MGB-NFQ		
	FAM	FAM-CAAAACAA**G**CAAACAA-MGB-NFQ		
6	F	CGTGATGTCACAGGCACAGA	72	TaqMan (AbD)
	R	TCCATTCCATGAGAGAGTTATGTTTGG		
	VIC	VIC-CTCAGCAC**G**GTGTCC-MGB-NFQ		
	FAM	FAM-CTCAGCAC**A**GTGTCC-MGB-NFQ		
7	F	GGCCAACTTCATTCAAATGTCTTCTC	90	TaqMan (AbD)
	R	CCAGCCATTTTCTACAGAGACATG		
	VIC	VIC-CTTCCC**G**TTTTCTC-MGB-NFQ		
	FAM	FAM-CCTTCCC**T**TTTTCTC-MGB-NFQ		
8	F	TCCTGCGAAAGATACAAAGTCAGAAA	103	TaqMan (AbD)
	R	GCTGGTAAGGTGAAATTAAGGCATGA		
	VIC	VIC-AAAACGACATTT**C**GGCACC-MGB-NFQ		
	FAM	FAM-AAAACGACATTT**T**GGCACC-MGB-NFQ		
9	F	CTTAATAACTAAATGAATTGCCCTGCTTTGT	81	TaqMan (AbD)
	R	CCCCTGAAATGAGGAGACATGAATC		
	VIC	VIC-TTCAGGAATGATTA**G**GCAAA-MGB-NFQ		
	FAM	FAM-TTCAGGAATGATTA**A**GCAAA-MGB-NFQ		
10	F	TGTGAAATTCCAACTTCTGGACAGT	80	TaqMan (AbD)
	R	CCTGCTTACTTGGCTGCAAAG		
	VIC	VIC-TCTGATTTT**C**CTGGCTTC-MGB-NFQ		
	FAM	FAM-TTCTGATTTT**A**CTGGCTTC-MGB-NFQ		
11	NA	Unavailable	< 200	TaqMan (AoD)

12	F	CATTTCCGTTGCATTTTTGTGAACA	107	TaqMan (AbD)
	R	CCTCAAGTACAGGAAGAAAGCTGAA		
	VIC	VIC-TGTTGCATGT**A**GTAGATG-MGB-NFQ		
	FAM	FAM-TTGCATGT**G**GTAGATG-MGB-NFQ		
13	F	GTATTTATTTTTATCCTGTATCACCAAGCGTTT	82	TaqMan (AbD)
	R	GAAATTTAGAAGACGATGATGAGAGGGTAA		
	VIC	VIC-CCGAGTATC**G**TAATCAG-MGB-NFQ		
	FAM	FAM-CCGAGTATC**C**TAATCAG-MGB-NFQ		
14	F	CCAGTGGCTGAATCTACTTTCTCT	79	TaqMan (AbD)
	R	CAAGTGGAAGGTAAGGAGGAAGAG		
	VIC	VIC-CTGGAGATA**T**ATAAAATT-MGB-NFQ		
	FAM	FAM-TGGAGATA**C**ATAAAATT-MGB-NFQ		
15	F	ATTGATTGTGGTTATGGCGCATTT	77	TaqMan (AbD)
	R	CTCACCCTCAGGTCTTCCTATGT		
	VIC	VIC-CCAGCTTTTT**T**GTACCGC-MGB-NFQ		
	FAM	FAM-CAGCTTTTT**C**GTACCGC-MGB-NFQ		
16	F	AGCGGCCACTTTTATTCATCAGATA	86	TaqMan (AbD)
	R	ACAAAGGACTTTAGGGCCCAAATT		
	VIC	VIC-TTGGAGTACT**G**AATAAA-MGB-NFQ		
	FAM	FAM-CTTGGAGTACT**A**AATAAA-MGB-NFQ		
17	F	TGTGCGTTTTCATTTTGAGAGTAATTGG	94	TaqMan (AbD)
	R	CCTGAGAATCCTAAACACATGGACAA		
	VIC	VIC-CAGGATTG**G**TTAACACT-MGB-NFQ		
	FAM	FAM-CAGGATTG**A**TTAACACT-MGB-NFQ		
18	F	CTCTATGGAGACCCTCTATAATTTTAAATTTTAACAGA	96	TaqMan (AbD)
	R	CTAGGGTTGGCCAAGGCA		
	VIC	VIC-CATAGCAAATGA**G**ACACAC-MGB-NFQ		
	FAM	FAM-ATAGCAAATGA**A**ACACAC-MGB-NFQ		
19	F	GCCCAGCTTGCATTGTTTCTAG	93	TaqMan (AbD)
	R	TCCACAGTTAAAACTTGATGCTTCTGA		
	VIC	VIC-TGACGC**G**GGTTCT-MGB-NFQ		
	FAM	FAM-TGACGC**A**GGTTCT-MGB-NFQ		
20	F	TGGTGGATACCTCAGGATGGT	71	TaqMan (AbD)
	R	CTCATACCTGCAAACATGGAACATT		
	VIC	VIC-CCTTTG**G**AAAACCA-MGB-NFQ		
	FAM	FAM-CCTTTG**A**AAAACCA-MGB-NFQ		
21	F	AGTTTGAGAACAGTGGAGTTCCAATC	66	TaqMan (AbD)
	R	CTCACAATGCATGGAACACAACAT		
	VIC	VIC-ACCTGGC**A**TAGCTTT-MGB-NFQ		
	FAM	FAM-ACCTGGC**T**TAGCTTT-MGB-NFQ		
22	F	CATAGCCTCTTTGCCGCATTTC	71	TaqMan (AbD)
	R	GCCTGATCATGCAATACACAAATCC		
	VIC	VIC-AACATATGGGAT**G**TGACATG-MGB-NFQ		
	FAM	FAM-TAACATATGGGAT**A**TGACATG-MGB-NFQ		
23	F	GCATTAGATGATAGATTAAATTAGATGCTTCATTAAACATT	147	TaqMan (AbD)
	R	AGAATCTCTCTTAAAATATACAAAGACTCAATGAAGG		
	VIC	VIC-CCAACAATATCCA**T**CTAAAA-MGB-NFQ		
	FAM	FAM-CAACAATATCCA**C**CTAAAA-MGB-NFQ		
24	F	AGCTCTGACCCTTCCTGACT	74	TaqMan (AbD)
	R	TGCATCTAGCAGGACCTGGATAT		
	VIC	VIC-CTGAGCTGTG**A**TGACC-MGB-NFQ		
	FAM	FAM-TGAGCTGTG**G**TGACC-MGB-NFQ		
25	F	CAGTGGCATCCTGTCTTATGAGT	71	TaqMan (AbD)
	R	ACAATTAAAGCAAGTCTTGGGTCACT		
	VIC	VIC-CACACTAAAGT**G**CACACTG-MGB-NFQ		
	FAM	FAM-CACACTAAAGT**A**CACACTG-MGB-NFQ		

26	F	CACAGACTTCACCTTAGGGTTTTACA	90	TaqMan (AbD)
	R	AAGCACTGTACACAATCAACAACAAAT		
	VIC	VIC-CACCAG**G**CCCCTCC-MGB-NFQ		
	FAM	FAM-CACCAG**A**CCCCTCC-MBG-NFQ		
27	F	CAGGACAGTGGAGTAAAGAAAACAGA	78	TaqMan (AbD)
	R	ATGACTGCCTAGTTCCCTCCT		
	VIC	VIC-AAATTTTAAGCCTC**G**TGAGAGT-MGB-NFQ		
	FAM	FAM-AAATTTTAAGCCTC**A**TGAGAGT-MGB-NFQ		
28	F	AAGTGAAGATGCTGGGTAGGAATG	64	TaqMan (AbD)
	R	CCTTTCCAATTTGGCCTTGTGTTAC		
	VIC	VIC-CAGCGAG**A**AAAGTGT-MGB-NFQ		
	FAM	FAM-AGCGAG**G**AAAGTGT-MGB-NFQ		
29	F	AACAGACACATTTAAACCATGTATCCATCT	97	TaqMan (AbD)
	R	TTGGGACTTCCTGGTTTGCTTAA		
	VIC	VIC-TTTAAGAGAGTACAAT**A**TATTTG-MGB-NFQ		
	FAM	FAM-AGAGAGTACAAT**G**TATTTG-MGB-NFQ		
30	F	ATTTCAATTAAGAAAAATACACTGAATACTAAAGCAGTTT	97	TaqMan (AbD)
	R	GGGTTCTAGTCCATTAAAAGGTGGTAA		
	VIC	VIC-TCTCTCACA**A**TAAGACTTA-MGB-NFQ		
	FAM	FAM-CTCTCACA**G**TAAGACTTA-MGB-NFQ		
31	F	AGGTTAATTTTAGAGGCACACAGACAA	100	TaqMan (AbD)
	R	GTGGGAGTCATTAAAAATTCTACATGTTGT		
	VIC	VIC-TTGGC**G**CAATGGA-MGB-NFQ		
	FAM	FAM-ATATTTGGC**A**CAATGGA-MGB-NFQ		
32	F	CTGTCTTCTCTGATTCCAGGTGTTA	74	TaqMan (AbD)
	R	GAGGATGCTTGGGAATAGCTAAGAA		
	VIC	VIC-CTGCCTAGAACA**C**TATGA-MGB-NFQ		
	FAM	FAM-CTGCCTAGAACA**G**TATGA-MGB-NFQ		
33	NA	Unavailable	< 200	TaqMan (AoD)
34	F	GTATTTCTTAAACTTTAGGTCGTCTCCCA	85	TaqMan (AbD)
	R	GGATCATTTTACTGCCTGCAATATTGAG		
	VIC	VIC-CAGCAAATTATTA**T**AGATTTT-MGB-NFQ		
	FAM	FAM-CAAATTATTA**C**AGATTTT-MGB-NFQ		
35	F	GCCTCACAACTATAACCAACCATTG	95	TaqMan (AbD)
	R	GGCTTAATACTATCATTATCATCAGTTGTCCT		
	VIC	VIC-AAGCTAGAAGA**A**TTTAAG-MGB-NFQ		
	FAM	FAM-AAGCTAGAAGA**C**TTTAAG-MGB-NFQ		
36	F	TTGTGCCCCAGGTTCAGTAC	73	TaqMan (AbD)
	R	CTGAAGTGGGTTAGAAGATGGGAAT		
	VIC	VIC-TCACCA**C**AACATGGC-MGB-NFQ		
	FAM	FAM-CACCA**G**AACATGGC-MGB-NFQ		
37	F	AGCTCTAAATTGTGGTGCATACTGA	73	TaqMan (AbD)
	R	GCCTTACAGCTGACGAAATGC		
	VIC	VIC-CTCAGAGAAATAAAAT**A**AGTG-MGB-NFQ		
	FAM	FAM-CTCAGAGAAATAAAAT**C**AGTG-MGB-NFQ		
38	F	GTCAGGTACCTACACTGGCTTTAC	88	TaqMan (AbD)
	R	GCTCTGTGTTGATTTTGCAAGGAAT		
	VIC	VIC-TGTACTATATC**A**CATACCCC-MGB-NFQ		
	FAM	FAM-CTATATC**G**CATACCCC-MGB-NFQ		
39	F	GCCTAGAGTACTTATTGCAACACCAA	95	TaqMan (AbD)
	R	CTACGTTGATGCCTTCTAAGGTTACT		
	VIC	VIC-AAAGGGAATCAA**A**GTCCAT-MGB-NFQ		
	FAM	FAM-AAGGGAATCAA**G**GTCCAT-MGB-NFQ		

40	F	TTCCAAAACATAGTTATCTGTTGGCTGTA	98	TaqMan (AbD)
	R	ACATTAGGAAAATCTGTGAACAGTTGAAGA		
	VIC	VIC-CATTCCTTC**A**TTTGGAAAT-MGB-NFQ		
	FAM	FAM-TTCCTTC**G**TTTGGAAAT-MGB-NFQ		
41	F	CTCTCTGCCCCTCTTTGTTCTG	86	TaqMan (AbD)
	R	AGCCCAAAGGAGCAAAGCA		
	VIC	VIC-CCTGAAGG**G**CCAGCCT-MGB-NFQ		
	FAM	FAM-CCTGAAGG**A**CCAGCCT-MGB-NFQ		
42	F	CTCAGGGCCCGTGTAACAG	65	TaqMan (AbD)
	R	CACTCAGAAGGGAAGCCATCTC		
	VIC	VIC-CCACACCA**C**TGAGAGG-MGB-NFQ		
	FAM	FAM-CCACACCA**T**TGAGAGG-MGB-NFQ		
43	F	CGCCTTGAAGATGCTCCTATAAGTA	64	TaqMan (AbD)
	R	ACTCAGGCCGATCAGGACTT		
	VIC	VIC-TGGTCTGGCCC**C**TG-MGB-NFQ		
	FAM	FAM-TGGTCTGGCCC**T**TG-MGB-NFQ		
44	F	CCGGATCCAGCTGGTAAGAAAA	91	TaqMan (AbD)
	R	CGACGCCGCCTTCCA		
	VIC	VIC-TTTGCTTTTCC**A**TTCCTTA-MGB-NFQ		
	FAM	FAM-TTGCTTTTCC**G**TTCCTTA-MGB-NFQ		
45	F	GCAAAGCCAATGACCTCATCAAC	56	TaqMan (AbD)
	R	GGTCCCCGTGGACAATCA		
	VIC	VIC-ATCGCTGC**A**GTGACC-MGB-NFQ		
	FAM	FAM-CGCTGC**G**GTGACC-MGB-NFQ		
46	F	GCACAGGCAAGAAGAATTGTTTGAT	102	TaqMan (AbD)
	R	CGCTTTCACAGGACTGAGTTCA		
	VIC	VIC-CAAGAAAATCTA**G**CACATTG-MGB-NFQ		
	FAM	FAM-CAAGAAAATCTA**A**CACATTG-MGB-NFQ		
47	F	GAGGTAATTCTGAGTTTTGCTCCGA	95	TaqMan (AbD)
	R	CCAGTTCTTACCCATGTTTCTGCTT		
	VIC	VIC-CAACCTAGTTGAC**A**CCAGAG-MGB-NFQ		
	FAM	FAM-ACCTAGTTGAC**G**CCAGAG-MGB-NFQ		
48	F	ACCTTCTCCAGCAGCTTTGG	64	TaqMan (AbD)
	R	CTGCCAGCCTCCCTTTCTC		
	VIC	VIC-AGAAGGATATCA**T**GCCCGAC-MGB-NFQ		
	FAM	FAM-AAGGATATCA**C**GCCCGAC-MGB-NFQ		
49	F	TCCTGTCTGGATGAAATGCTATGG	88	TaqMan (AbD)
	R	CTGAGTCCAAACCCTAGCTGTAG		
	VIC	VIC-CACTCCCAGA**C**TGCACA-MGB-NFQ		
	FAM	FAM-ACTCCCAGA**T**TGCACA-MGB-NFQ		
50	F	CCATCAGTCAGACTTGTCAACCT	77	TaqMan (AbD)
	R	CCAGAGCAGAAGGGTGGAT		
	VIC	VIC-CATGGACAAA**A**GTAG-MGB-NFQ		
	FAM	FAM-CATGGACAAA**G**GTAG-MGB-NFQ		
51	F	CTGGAGAAAAGGTGCTCTCTCT	85	TaqMan (AbD)
	R	TGCCTTTAGGGTCCCTTTTATTACAAA		
	VIC	VIC-CCCTCTTAC**G**CTGTGCA-MGB-NFQ		
	FAM	FAM-CCCTCTTAC**A**CTGTGCA-MGB-NFQ		
52	F	CTCACCTTAATTCTTCCAGCTCTGT	85	TaqMan (AbD)
	R	AAAGGGTTTGCGTCTCTGTGA		
	VIC	VIC-ATCATGAATACAATACC**T**CTGTCT-MGB-NFQ		
	FAM	FAM-ATGAATACAATACC**G**CTGTCT-MGB-NFQ		
53	F	TCCACATCCCAGTAGGCAATAGT	67	TaqMan (AbD)
	R	AGGAATCGACCTTTAAGGAAAGCAA		
	VIC	VIC-CTCGACTTCAG**G**AAAC-MGB-NFQ		
	FAM	FAM-CGACTTCAG**A**AAAC-MGB-NFQ		

54	F	CGCCCCACTCCCTACAG	84	TaqMan (AbD)
	R	TCGGCACCAGCCCATC		
	VIC	VIC-CTGTCTG**T**GGACCTC-MGB-NFQ		
	FAM	FAM-CTGTCTG**A**GGACCTC-MGB-NFQ		
55	NA	Unavailable	< 200	TaqMan (AoD)
56	F	CCTAGATGCTGAAGAGAGGAGTGT	83	TaqMan (AbD)
	R	GGTTTGGTGGATCTTGAATTAGAGGAT		
	VIC	VIC-CCCTG**C**CATTTTA-MGB-NFQ		
	FAM	FAM-CCCCTG**T**CATTTTA-MGB-NFQ		
57	F	CACATCCATGAAGGCAGACAGT	65	TaqMan (AbD)
	R	GCTACAGCTGGGCATGAAATG		
	VIC	VIC-CTTGAGTGATGC**A**CAAGA-MGB-NFQ		
	FAM	FAM-TTGAGTGATGC**G**CAAGA-MGB-NFQ		
58	F	AATTCAGATCTTAGAGCACGAAGCA	73	TaqMan (AbD)
	R	AGTCTCCTGGAGGTGAATTAACACT		
	VIC	VIC-CCAGGC**A**CCTCCAG-MGB-NFQ		
	FAM	FAM-ACCAGGC**T**CCTCCAG-MGB-NFQ		
59	F	GTTTCCTCCAGGTGCTCTTATGG	70	TaqMan (AbD)
	R	CCGGATTTGCCAGGAATTATCTCA		
	VIC	VIC-TCACAAAGAC**T**CAGCCCT-MGB-NFQ		
	FAM	FAM-CACAAAGAC**A**CAGCCCT-MGB-NFQ		
60	F	TGGGTGCTGGGACATTTTCTC	64	TaqMan (AbD)
	R	CGAGTTGTCCTAGCCATTCAGT		
	VIC	VIC-TCCTGCTTCTC**A**GGTACA-MGB-NFQ		
	FAM	FAM-CTGCTTCTC**G**GGTACA-MGB-NFQ		
61	F	TTTCTTTTTTCATTCAAAGCGAGGGAA	83	TaqMan (AbD)
	R	AAATGCAAGTGTTCTTGGAAAAGGTT		
	VIC	VIC-AATAACAAGGAC**T**CCCTAAAT-MGB-NFQ		
	FAM	FAM-AAGGAC**G**CCCTAAAT-MGB-NFQ		
62	F	TTTGAAAGTGATTACAAACCTCTACTTCTCA	96	TaqMan (AbD)
	R	CATGCATTTCTTACTTATTTGATGCAGAAGA		
	VIC	VIC-CAGTGCA**A**GATTTA-MGB-NFQ		
	FAM	FAM-CAGTGCA**G**GATTTA-MGB-NFQ		
63	F	ACTTTCAGAAATTAACTATTCAAAGAAGCAACATT	119	TaqMan (AbD)
	R	CACTGTGTGCTTTCCAAGCC		
	VIC	VIC-CAAACCCCA**G**TTTCTGT-MGB-NFQ		
	FAM	FAM-AAACCCCA**A**TTTCTGT-MGB-NFQ		
64	F	CTGCATTGTCCCATGAGGTCAT	86	TaqMan (AbD)
	R	GCACGTTTACTAAAGCCACTTGT		
	VIC	VIC-CTTTTGAAAC**G**GGTTTAA-MGB-NFQ		
	FAM	FAM-TTCTTTTGAAAC**A**GGTTTAA-MGB-NFQ		
65	F	TGGGCCACCTGTTGCT	62	TaqMan (AbD)
	R	CCAGGAGACCCAGACCCA		
	VIC	VIC-CTTTT**G**CCCATCCACC-MGB-NFQ		
	FAM	FAM-CTTTT**C**CCCATCCACC-MGB-NFQ		
66	F	ACCAGCAGGAGTCTAACAGTCA	61	TaqMan (AbD)
	R	GGCTTCCCCACGTTTTGG		
	VIC	VIC-CCCCTT**C**TCAACCTT-MGB-NFQ		
	FAM	FAM-CCCCTT**T**TCAACCTT-MGB-NFQ		
67	F	TTTCTGTTCTTTCTTCATGGAAAGGAGAA	89	TaqMan (AbD)
	R	GAGGGAAACAGACACTCAAATACCT		
	VIC	VIC-CAAATTCTGTCTT**T**GCTATAG-MGB-NFQ		
	FAM	FAM-AAATTCTGTCTT**A**GCTATAG-MGB-NFQ		

68	F	GATTTCCAGGAAGGCCTAGCT	67	TaqMan (AbD)
	R	GATTCCCCAGATCACACAGCTT		
	VIC	VIC-AGACCTTTCC**A**GAGGTGA-MGB-NFQ		
	FAM	FAM-ACCTTTCC**G**GAGGTGA-MGB-NFQ		
69	F	CAGCTGAAGAAAACAGAAAGATGAAACT	95	TaqMan (AbD)
	R	GATCATTTCAATTTGAATTCCCAGTGCTT		
	VIC	VIC-CAAAGAAAAGTTGC**A**TGTTAA-MGB-NFQ		
	FAM	FAM-CAAAGAAAAGTTGC**T**TGTTAA-MGB-NFQ		
70	F	TCCCCTAAGTACTTCCAGTAGATGTAAC	103	TaqMan (AbD)
	R	AGCTCAATAAATCTATCCCTGTGCATT		
	VIC	VIC-ACTGCTGG**A**CTGTATG-MGB-NFQ		
	FAM	FAM-CTGCTGG**G**CTGTATG-MGB-NFQ		
71	F	AGGAGCCATGCATTAAGTTGCT	137	TaqMan (AbD)
	R	GAACCTGAAAGACCCTCTGTGAAA		
	VIC	VIC-CAAGCACTC**A**GCTGTG-MGB-NFQ		
	FAM	FAM-AAGCACTC**G**GCTGTG-MGB-NFQ		
72	F	TGGGCTGCACTTGCCAAT	56	TaqMan (AbD)
	R	GCCGCTCTGCTTCTTAGCTAT		
	VIC	VIC-TTGCCAG**G**TGTTGCT-MGB-NFQ		
	FAM	FAM-TTTTGCCAG**A**TGTTGCT-MGB-NFQ		
73	F	GGTGGGAGGGTGATAGACAGAA	69	TaqMan (AbD)
	R	CCAATGCATTCTAATCCCCTACCA		
	VIC	VIC-ACACCATGCT**A**TCCCGT-MGB-NFQ		
	FAM	FAM-CACCATGCT**G**TCCCGT-MGB-NFQ		
74	F	CCAAAATGAGTCCCAGGAAGCT	75	TaqMan (AbD)
	R	CTGCCCCACCACAGAATAGG		
	VIC	VIC-CTGACCCT**C**ACGCCCA-MGB-NFQ		
	FAM	FAM-CTCTGACCCT**T**ACGCCCA-MGB-NFQ		
75	F	TCAGCCCCATTTCTCTACCAGA	95	TaqMan (AbD)
	R	AGGACGAAAGCAGCTGGTATTT		
	VIC	VIC-CTAATTCAAGTAATGAC**A**TCCTAT-MGB-NFQ		
	FAM	FAM-TTCAAGTAATGAC**G**TCCTAT-MGB-NFQ		
76	F	CTGCCTGTGTTCTAAGCAACTCT	85	TaqMan (AbD)
	R	AGAGACAACAGAATTAGCAGAGCATAAA		
	VIC	VIC-CTGCCAAT**A**GATGCC-MGB-NFQ		
	FAM	FAM-CTGCCAAT**C**GATGCC-MGB-NFQ		
77	F	GATGGGTTCAGTGGCTAAAGGAA	81	TaqMan (AbD)
	R	CCTGAGGCATTGAGTTTTCAGTACA		
	VIC	VIC-AGCTCATTTC**A**GCCCCAT-MGB-NFQ		
	FAM	FAM-CTCATTTC**G**GCCCCAT-MGB-NFQ		
78	F	GCAGCACCATGGGTAATATGAAATG	89	TaqMan (AbD)
	R	TGCATCCCAACCCAAGGAAAA		
	VIC	VIC-TTTATTCTAGGGTCA**T**CTGTCAC-MGB-NFQ		
	FAM	FAM-TCTAGGGTCA**C**CTGTCAC-MGB-NFQ		

Statistical Analyses

Initial analysis of SNP genotype data involved quality control checks consisting of verification of internal controls and assessment of Mendelian inconsistencies, followed by final haplotype consistency analyses using Simwalk2 [117]. Conformity with anticipated Hardy-Weinberg equilibrium expectations was established. Transmission disequilibrium (TD) in autism families was determined using the pedigree disequilibrium test (PDT) statistic, developed for use with general pedigrees [119]. Maternal and paternal specific transmissions were also assessed using the T_{SP} test [156]. Common haplotypes (≥5%) were determined using Haploview [129], and analyzed using the FBAT analysis software package [130]. TD for haplotypes was performed using haplotype tag SNPs for multi-marker blocks. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$.

Results

We examined 78 single nucleotide polymorphisms across the IC and MED domains of 15q11.2-q12 (Figure 7-2). No nominally significant ($P \leq 0.05$) single marker association results were found in the total dataset of 384 families. However, we would have expected to find 78x0.05=3.9 (i.e. ~4) significant ($P \leq 0.05$) results. Extending from work previously done in the lab [47, 143], we examined the entire dataset for maternal and paternal transmissions using the T_{SP} test; this statistic also allows for simultaneous analysis of both multiplex and

singleton families, while allowing for separate analyses of paternal- and maternal-specific transmissions. Three markers (4,12, and 16) located in and around *SNRPN* demonstrated nominal ($P \leq 0.05$) evidence for TD when maternally-inherited. Two additional markers (45 and 76) showed nominal TD when paternally-transmitted. Though these results are well within expectations given the testing of all 78 markers for parental transmissions, they warrant follow-up study.

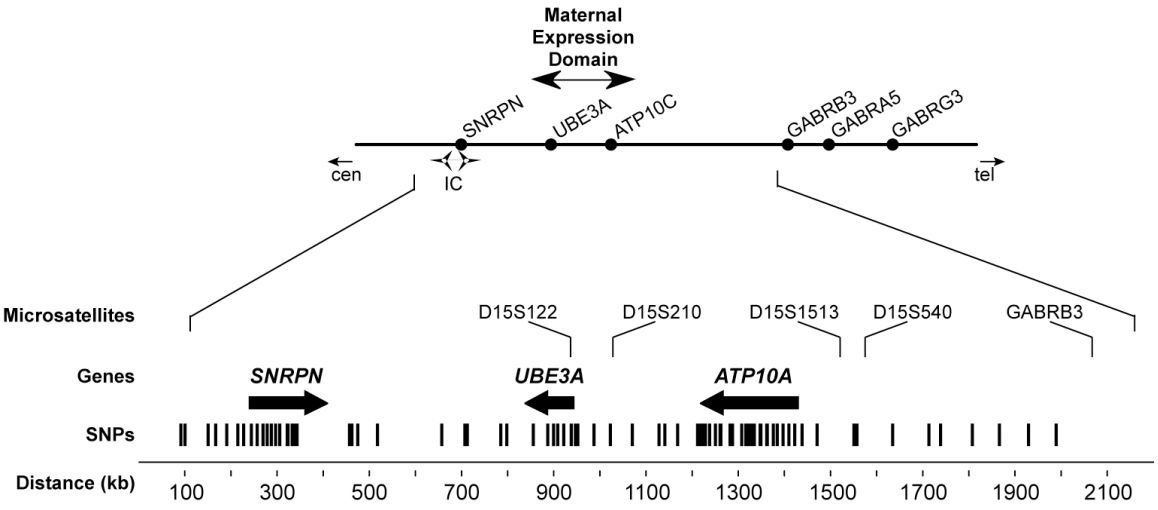


Figure 7-2. Schematic map detailing SNPs genotyped across the Imprinting Center (IC) and Maternal Expression Domain (MED). The autism candidate region includes the 15q imprinting center (IC), maternally-expressed genes (*UBE3A* and *ATP10A*) and the cluster of $GABA_A$ receptor subunit genes. The 2-Mb interval containing *SNRPN*, *UBE3A*, and *ATP10A* is expanded and shows relative position and transcriptional orientation (arrows) of each gene. Microsatellite markers used in our genome-wide linkage screen are indicated as references. Vertical hashes above the scale in kb indicate SNP markers analyzed in this study. The proximity of a large number of SNPs in relation to the scale is such that separate hashes do not distinguish these markers.

Intermarker LD was examined using the Haploview program [129]. Examination of LD across all markers included in the present study identifies 12 loosely defined haplotype blocks, compared with 58 blocks detected using data (markers with minor allele frequency > 0.05) from the latest HapMap release (#14, 2004-12-10) and the default algorithm of Gabriel et al. within the Haploview program [147]. Haploview output for such a large region would be impractical to show, but markers defining the 12 blocks described are indicated in Table 7-3. Association analysis in the overall dataset, using FBAT, of all common haplotypes yielded no significant association (data not shown).

In addition to examination of the entire dataset (384 families), we also examined those families only containing male-affected individuals (235 families) separately from those families containing female-affected individuals (149 families). These analyses were performed in the context that other groups have recently presented findings with regard to sex-specific differences on other chromosomes as described in Chapters III and IV [78, 79]. All single marker results for the described datasets are presented in Table 7-3.

Table 7-3. Single marker analysis across the IC-MED region. Blocks of markers in high LD are highlighted in green.

Gene	#	Marker	All Families (384)			Affected Male Only Families (235)			Affected Female Containing Families (149)		
			PDT	Tsp Father	Tsp Mother	PDT	Tsp Father	Tsp Mother	PDT	Tsp Father	Tsp Mother
	1	hCV1796103	0.892	0.280	0.388	0.754	0.726	0.178	0.599	0.239	0.758
	2	hCV9710866	0.312	0.941	0.481	0.889	0.666	0.745	0.143	0.612	0.502
	3	hCV8378225	0.868	0.586	0.331	0.906	0.527	0.677	0.727	0.908	0.335
	4	hCV8378415	0.468	0.689	**0.047**	0.502	0.853	0.493	0.733	0.441	**0.017**
	5	hCV7516185	0.241	0.131	0.458	0.687	0.311	0.579	0.183	0.248	0.623
	6	hCV1244494	0.485	0.730	0.499	0.824	0.491	0.469	0.159	0.189	0.691
	7	hCV1244509	0.753	0.587	1.000	0.235	0.104	0.274	0.066	0.248	0.159
	8	hCV2979338	0.609	0.564	0.663	0.881	0.518	0.490	0.308	0.898	0.895
	9	hCV2979365	0.679	0.358	0.463	0.908	0.351	0.373	0.592	0.777	0.898
	10	hCV2979396	0.215	0.160	0.203	0.550	0.322	0.596	0.227	0.317	0.152
	11	hCV2979419	0.182	0.055	0.932	0.068	0.122	0.838	0.823	0.257	0.876
	12	hCV1223678	0.140	0.551	**0.007**	0.126	0.683	**0.020**	0.636	0.663	0.228
SNRPN	13	hCV1223674	0.137	0.325	0.835	0.092	0.392	0.903	0.732	0.599	0.366
SNRPN	14	hCV16287658	0.368	0.586	1.000	0.408	0.375	0.855	0.692	0.800	0.647
SNRPN	15	hCV3214754	0.537	0.217	0.944	0.102	0.356	0.112	0.346	0.413	0.250
SNRPN	16	hCV3214752	0.111	0.917	**0.030**	0.119	0.803	**0.016**	0.553	0.847	0.083
SNRPN	17	hCV1025117	0.205	0.751	0.152	0.112	0.522	**0.046**	1.000	0.706	0.096
SNRPN	18	hCV2066559	0.517	0.165	0.885	0.096	0.404	0.265	0.360	0.251	0.083
SNRPN	19	hCV2066555	0.965	0.592	0.655	0.349	0.919	0.428	0.272	0.352	0.092
SNRPN	20	hCV2066488	0.659	0.357	0.588	0.668	0.919	0.726	0.856	0.198	0.201
SNRPN	21	hCV2066485	0.849	0.107	0.241	0.679	0.371	0.541	0.461	0.159	**0.015**
SNRPN	22	hCV2066478	0.789	0.057	0.195	0.758	0.278	0.932	0.455	0.100	**0.035**
SNRPN	23	hCV152343	0.608	0.641	0.837	0.953	0.827	0.413	0.419	0.377	0.211
	24	hCV132223	0.143	0.273	0.406	0.325	0.450	0.732	0.262	0.157	0.411
	25	hCV422420	0.302	0.535	0.384	0.380	0.182	0.198	0.574	0.578	1.000
	26	hCV422410	0.500	0.558	0.113	0.932	0.138	0.192	0.222	0.157	0.297
	27	hCV2625778	0.799	0.601	0.665	0.636	0.588	0.336	0.394	0.837	0.081
UBE3A	28	hCV11487073	0.590	0.522	0.149	0.694	0.249	0.133	0.713	0.564	0.535
UBE3A	29	hCV2558359	0.763	0.704	0.394	0.817	0.503	0.157	0.846	0.803	0.921
UBE3A	30	hCV2558365	1.000	0.361	0.337	0.941	0.204	0.286	0.932	0.866	0.718
UBE3A	31	hCV11487149	0.571	0.877	0.477	0.560	0.626	0.157	0.844	0.701	0.765
UBE3A	32	hCV2558398	0.435	0.805	1.000	0.679	0.811	0.486	0.490	0.909	0.345
UBE3A	33	hCV2558422	0.421	0.502	0.549	0.179	0.376	0.208	0.783	1.000	0.777
UBE3A	34	hCV2558441	0.817	0.404	0.324	0.941	0.128	0.286	0.782	0.465	0.701
UBE3A	35	hCV2558443	0.464	0.193	0.802	0.242	0.196	0.726	0.810	0.662	0.406
UBE3A	36	hCV8766820	0.907	0.600	0.435	0.820	0.172	0.286	0.647	0.239	1.000
	37	hCV28014321	0.811	0.840	0.823	1.000	0.799	0.629	0.712	0.508	0.814
	38	hCV2564743	0.664	0.331	0.099	0.571	0.353	0.071	1.000	0.706	0.739
	39	hCV15943093	0.136	0.210	0.127	0.088	0.085	0.138	0.880	0.895	0.639
	40	hCV26647614	0.965	0.538	0.732	0.391	0.404	0.566	0.260	0.869	0.134
	41	hCV431962	0.061	**0.054**	0.627	0.063	**0.047**	0.347	0.527	0.622	0.633
	42	hCV454609	0.725	0.592	0.086	0.628	0.922	0.465	0.939	0.332	**0.051**
ATP10A	43	hCV1157519	0.645	0.401	0.514	0.947	0.751	0.467	0.525	0.336	0.884
ATP10A	44	hCV15863547	0.317	1.000	0.317	0.317	NA	0.317	1.000	1.000	1.000
ATP10A	45	hCV1157511	0.122	**0.018**	0.651	0.275	**0.022**	0.673	0.272	0.387	0.335
ATP10A	46	hCV12080790	0.200	0.086	1.000	0.490	0.072	0.331	0.251	0.617	0.340
ATP10A	47	hCV12080806	0.256	0.225	0.706	0.419	0.456	0.873	0.414	0.297	0.683
ATP10A	48	hCV12080811	0.435	0.592	0.362	0.482	0.691	0.480	0.710	0.715	0.564
ATP10A	49	hCV320540	0.760	0.883	0.596	0.924	0.441	0.697	0.701	0.497	0.706
ATP10A	50	hCV394756	0.524	0.454	0.398	0.182	0.193	0.914	0.455	0.456	0.181
ATP10A	51	hCV394762	0.769	1.000	0.586	0.404	0.289	0.448	0.610	0.225	0.096
ATP10A	52	hCV125931	1.000	0.360	0.941	0.306	0.063	0.615	0.206	0.285	0.583
ATP10A	53	hCV24876	0.748	0.688	0.883	0.703	0.608	0.753	0.386	0.197	0.917
ATP10A	54	hCV12064096	0.227	0.190	0.313	0.081	0.180	0.099	0.801	0.674	0.662
ATP10A	55	hCV232863	0.705	0.793	0.605	0.949	0.371	0.666	0.524	0.484	0.761
ATP10A	56	hCV11300397	0.704	0.812	0.758	0.215	0.490	0.574	0.270	0.181	0.138
ATP10A	57	hCV399638	0.835	0.728	0.603	0.367	0.819	0.782	0.494	0.423	0.211
ATP10A	58	hCV11300400	0.808	0.203	0.524	0.824	0.292	0.635	0.907	0.475	0.112
ATP10A	59	hCV8864618	0.571	0.157	0.492	0.174	0.117	0.531	0.442	0.793	0.751
ATP10A	60	hCV9400334	0.811	0.505	1.000	0.332	0.221	0.588	0.389	0.564	0.486
ATP10A	61	hCV8864621	0.328	0.663	0.741	1.000	0.928	0.798	0.124	0.541	0.835
ATP10A	62	hCV9400402	0.254	0.766	0.776	0.726	0.719	0.535	**0.022**	0.285	0.722
ATP10A	63	hCV423475	0.111	0.814	0.153	0.426	0.583	0.437	0.133	0.371	0.239
ATP10A	64	hCV504244	0.155	0.560	0.937	0.834	1.000	0.553	**0.047**	0.317	0.362
ATP10A	65	hCV60097	0.348	0.730	0.300	0.635	0.399	0.502	**0.042**	0.710	**0.014**
ATP10A	66	hCV402689	0.869	0.728	0.662	0.845	0.916	0.536	0.540	0.647	0.869
ATP10A	67	hCV15863526	0.758	0.780	0.861	0.700	0.157	0.233	1.000	0.063	0.086
ATP10A	68	hCV8864860	0.411	0.272	0.806	0.345	0.078	0.884	0.895	0.706	0.819
	69	hCV1591365	0.635	0.056	0.429	0.351	0.239	1.000	0.748	0.103	0.251
	70	hCV8866740	0.656	0.270	0.917	0.869	0.267	0.541	0.366	0.715	0.433
	71	hCV8866723	0.496	0.470	0.611	0.237	0.178	0.584	0.701	0.612	0.904
	72	hCV8865982	0.232	0.354	0.820	0.726	0.915	0.340	**0.042**	0.174	0.378
	73	hCV578309	0.934	1.000	0.889	0.831	0.739	0.532	0.696	0.695	0.450
	74	hCV11669761	0.061	0.484	0.265	0.069	0.531	**0.028**	0.465	0.752	0.346
	75	hCV8864717	0.265	0.377	0.421	0.910	0.259	0.833	0.074	1.000	0.237
	76	hCV240706	0.560	**0.021**	0.161	0.307	**0.003**	0.608	0.715	0.748	0.063
	77	hCV26111547	0.695	0.440	0.723	0.346	0.200	0.931	0.565	0.710	0.623
	78	hCV2911753	0.509	0.411	0.566	0.435	0.353	0.825	**0.035**	**0.007**	0.336

Table 7-4. T_{sp} allelic transmission data for significant IC-MED markers. The number of fathers/mothers represents the number of heterozygous fathers/mothers used for each test. The number of transmissions (T) refers to the number of times a parent, heterozygous for that marker, transmitted that particular allele to an affected child. The number of non-transmissions (NT) refers to the number of times a parent, heterozygous for that marker, did not transmit that particular allele to an affected child.

	Marker	Tsp Father (P)	Father χ^2	Allele	Triads: No. of Fathers	Triads: T	Triads: NT	Sib-pair: No. of Fathers	Sib-pair: T	Sib-pair: NT	Tsp Mother (P)	Mother χ^2	Allele	Triads: No. of Mothers	Triads: T	Triads: NT	Sib-pair: No. of Mothers	Sib-pair: T	Sib-pair: NT
Overall Sample (384 Families)	hCV8378415										**0.047**	3.95	A	29	12	17	75	63	87
													G		17	12		87	63
	hCV1223678										**0.007**	7.32	A	18	9	9	72	89	55
													G		9	9		55	89
	hCV3214752										**0.030**	4.74	G	1	0	1	27	36	18
													A		1	0		18	36
	hCV431962	**0.054**	3.71	T	23	12	11	64	74	54									
				C		11	12		54	74									
	hCV1157511	**0.018**	5.64	A	25	15	10	73	56	90									
				G		10	15		90	56									
	hCV240706	**0.021**	5.29	A	22	7	15	79	68	90									
				C		15	7		90	68									
Affected Male Only (235 Families)	hCV1223678										**0.020**	5.44	A	11	6	5	48	60	36
													G		5	6		36	60
	hCV3214752										**0.016**	5.83	G	1	0	1	15	22	8
													A		1	0		8	22
	hCV1025117										**0.046**	4.00	G	4	2	2	16	22	10
													A		2	2		10	22
	hCV431962	**0.047**	3.95	T	14	7	7	41	50	32									
				C		7	7		32	50									
	hCV1157511	**0.022**	5.24	A	17	8	9	46	35	57									
				G		9	8		57	35									
	hCV11669761										**0.028**	4.86	T	25	10	15	37	28	46
													C		15	10		46	28
	hCV240706	**0.003**	8.78	A	15	5	10	43	32	54									
				C		10	5		54	32									
Affected Female Containing (149 Families)	hCV8378415										**0.017**	5.73	A	13	4	9	29	21	37
													G		9	4		37	21
	hCV2066485										**0.015**	5.95	A	9	7	2	38	48	28
													T		2	7		28	48
	hCV2066478										**0.035**	4.46	T	7	5	2	37	46	28
													C		2	5		28	46
	hCV454609										**0.051**	3.81	T	11	7	4	22	28	16
													C		4	7		16	28
	hCV60097										**0.014**	6.00	C	6	4	2	24	32	16
													G		2	4		16	32
	hCV2911753	**0.007**	7.41	A	10	4	6	29	40	18									
				G		6	4		18	40									

Discussion

There is compelling evidence for the involvement of 15q11-q13 in risk for autism spectrum disorders. The evidence is direct and quite clear for the small percentage of cases exhibiting interstitial or idic(15) duplications. When taken as a whole, genetic studies within families affected by idiopathic autism seem to point towards the cluster of $GABA_A$ receptor subunit genes, and in particular *GABRB3* and *GABRA5*. The picture is quite complicated, however. Maternal bias of genetic risk for duplication to be associated with autism suggests involvement of maternally-expressed genes. We acknowledge the maternal bias, but given the risk for autism in the case of paternal duplication and the absence of compelling data suggesting the paternal UPD PWS represents a class of PWS patients with more autistic behaviors, we conclude that dup(15) autism is (at least to some degree) a contiguous gene duplication effect.

We must now consider this background in the context of a recent report implicating epigenetic effects involving not just *UBE3A*, but also *GABRB3*, in Rett syndrome, AS and autism [172]. The fact that *GABRB3* has a common transcriptional orientation with *UBE3A* and *ATP10A*, makes the hypothesis of possible RNA interference-based epigenetic gene silencing effect for *GABRB3* at least plausible. Equally plausible, however, is the presence of, probably heterogeneous, variants in *GABRB3* and/or *GABRA5* that result in partial or complete loss-of-function alleles of one of these genes.

This analysis of the IC/MED region does not allow for a definitive conclusion on these points, however. While a number of markers do show nominal evidence for an effect, interpretation is hindered by the absence of a very *strong* main allelic effect at any one of these markers. While many sectors within this very large stretch of DNA were sufficiently covered relative to LD patterns, this is not universally true for all regions (data not shown) given the high recombination rate across this stretch of DNA. Therefore, we cannot conclude that a "negative" result in the absence of thorough coverage of all LD intervals, excludes the possibility that a significant effect was not missed.

To some degree, this survey was exploratory. We can hypothesize effects on either maternal or paternal alleles or both, there could be gender effects in terms of potential risk, and possibly even both parent-of-origin and sex-specific genetic effects. Any attempt to query for all of these scenarios necessarily represents a hypothesis-generating exercise. The number of tests performed reduces any given result to non-significance, if corrected for multiple testing; this is certainly true for the conservative application of a Bonferroni correction, but also true if one were to consider all markers in one block of LD to represent one independent test. We also have no *a priori* understanding of the size of any given genetic effect. These results would be consistent with potential allelic heterogeneity or any given putative TD affecting a relatively small number of families individually.

Future directions will include insuring complete coverage based on LD, for regions not completely represented; this is particularly true for the IC and regions of sequence homology. Given studies in PWS and AS cases with imprinting mutations that have led to well defined intervals within the imprinting center that are critical for the maternal-to-paternal or paternal-to-maternal imprint switch, it may be desirable to re-sequence or screen these short regions for evidence of sequence alteration more common to autism cases than in a control group. Similarly, regions of significant sequence homology in non-coding regions within the MED may be screened. Both of these situations would address the possibility that multiple different sequence variants (arising independently) can exert a functional effect on imprinting control or on *UBE3A*, *ATP10A*, or even *GABRB3*. Using larger datasets, such as with the Autism Genome Project, or smaller samples available through the NIMH repository, we may want to test for excess paternal (or maternal) allele-sharing across this region, to follow-up on a previous report [22].

CHAPTER VIII

A SECOND GENERATION GENOME-WIDE LINKAGE ANALYSIS TO IDENTIFY LOCI WITH AUTISM SUSCEPTIBILITY GENES[1]

Introduction

Several groups, including ours, have undertaken genome-wide screens of multiplex autism families for susceptibility loci [21, 25, 27-32, 109, 174, 175]. Comparing the results from all studies identifies a few genomic regions in common across multiple studies. Chromosomes 7q and 2q have received the greatest attention [28, 29, 48, 110, 111, 174, 176-178], and observations of chromosomal abnormalities in isolated autistic probands reinforce the plausibility of these regions for involvement in idiopathic autism. Genetic studies of autism are substantially complicated by both clinical and locus heterogeneity, and it is possible that epistatic or epigenetic mechanisms may play important roles in genetic etiology [8, 22]. Analytical strategies that address the latter concerns are limited, and most studies to date have focused on analysis of main effects using a global autism diagnosis to define affection status. Moving forward, more sophisticated approaches are being proposed in which trait-based subsets of the broader autism phenotype are used in genetic analyses. Similarly, given the interdependence of genes and their protein products within biological systems,

[1] Adapted from *BMC Med Genet* 2005 Jan 12, 6:1

analytical approaches that address potential interaction between susceptibility loci will also be critical to characterizing gene-phenotype relationships in autism.

We report a second generation 10-cM microsatellite-based genomic screen of multiplex autism families. The dataset for this screen includes 71 families recruited by the Tufts/New England Medical Center, a well-characterized set of 85 families from the Autism Genetics Resource Exchange (AGRE) and 2 families from Vanderbilt University. Several sites of suggestive linkage are identified, although none meet criteria for genome-wide significance. The loci with greatest support for linkage were 17q11.2 and 19p13; the latter site demonstrated significantly increased allele-sharing when the Ordered-Subset Analysis (OSA) algorithm was employed using a quantitative trait-based autism phenotypic subset related to specific developmental milestones as a covariate to rank families.

Materials and Methods

Families

The demographics for the 158 family dataset comprising the studies in this report are shown in Table 8-1. Families were recruited through three sites: (a) 71 families were recruited through the Tufts/NEMC site, (b) 2 families were recruited from the Vanderbilt University site, and (c) the remainder of families (85) were chosen from the AGRE repository based on criteria identical to those used for our own recruitment. Multiplex families (mostly affected sibling-pairs) had one

affected individual who met full criteria for autistic disorder based on Autism Diagnostic Interview-Revised (ADI-R; [3, 4, 179]) algorithm scores, while the second individual either met criteria or in some cases was under the cut-off by only one or two points. Exclusion criteria to enrich for idiopathic autism include dysmomrphology, abnormal chromosomal karyotype, diagnosis of fragile X syndrome, and other genetic disorders of known etiology. Individuals were assessed by the respective groups using the ADI-R at a developmental age >18 months; Tufts/NEMC and Vanderbilt groups included individuals between the ages of 4 and 22; in cases in which ADI-R interviews were performed initially at <4 years, they were repeated when the probands reached 4 years of age. All individuals were additionally assessed using the Autism Diagnostic Observation Schedule [3, 180] and the Vineland Adaptive Behavior Scales – Interview Edition [181, 182].

Table 8-1. Genomic screen sample demographics. *ADI-Rs performed <4 yrs were repeated at 4 yrs for Tufts/NEMC families. **IQ estimates are based on the Vineland Daily Living standard scores. ***IQ estimates are based on the overall Vineland Adaptive Behavior standard scores.

Families in Linkage Screen	**158**	**Ethnicity**		**I.Q. estimate distributions**	
Tufts/NEMC	71	**Caucasian (242)**	**73.0%**	Tufts/NEMC** (148)	
AGRE	85	Tufts/NEMC (130)	87.8%	<30	13
Vanderbilt	2	AGRE (108)	59.7%	30-49	21
Affected Individuals	**333**	Vanderbilt (4)	100.0%	50-69	37
Males	257	**Hispanic-Latino (14)**	**4.2%**	70+	17
Females	76	AGRE (14)	7.7%	Unknown	60
Tufts/NEMC Total	**148**	**African-American (8)**	**2.4%**	**AGRE*** (181)**	
Males	117	Tufts/NEMC (6)	4.1%	<30	17
Females	31	AGRE (2)	1.1%	30-49	27
AGRE Total	181	**Asian (8)**	**2.4%**	50-69	15
Males	137	Tufts/NEMC (2)	1.4%	70+	13
Females	44	AGRE (6)	3.3%	Unknown	109
Vanderbilt Total	**4**	**Multi-ethnic (14)**	**4.2%**	**Vanderbilt*** (4)**	
Males	3	AGRE (14)	7.7%	<30	1
Females	1	**Other (2)**	**0.6%**	30-49	1
Age at ADI (range)	**2-46.7**	Tufts/NEMC (2)	1.4%	50-69	0
Tufts/NEMC*	2-46.7	**Unknown (45)**	**13.5%**	70+	2
AGRE	2-38.0	Tufts/NEMC (8)	5.4%	Unknown	0
Vanderbilt	6.2-9.2	AGRE (37)	20.4%		

Genotype Data and Statistical Analyses

DNA from Tufts and Vanderbilt samples was obtained from peripheral blood or immortalized lymphoblastoid cell lines using the PureGene Kit (Gentra Systems). While a minority of families from the Tufts/NEMC cohort had been genotyped previously [109], both new and previously genotyped families were genotyped by deCODE (Reykjavik, Iceland) using their 500 marker (~8 cM intermarker spacing) panel and corresponding genetic map [114]. Genotype data were obtained from the AGRE website (http://agre.org) for families whose samples were included in this study. Clinical procedures and genotyping for the

AGRE sample has been described previously [30, 183]. Because we utilized existing genotype data, AGRE samples had a distinct but overlapping panel of markers compared to the Tufts or Vanderbilt families. All AGRE genetic markers were carefully placed on the deCODE map, with order and spacing (both genetic and physical) properly insured through exhaustive comparisons between genotyped markers, available genetic maps, and physical DNA sequence assemblies in both public and Celera databases.

Genotype data for each chromosome underwent thorough error detection and genotype confirmation. Initially, data were tested for Mendelian inconsistencies using PEDCHECK [184] and RELPAIR (http://www.sph.umich.edu/statgen/boehnke/relpair.html), followed by application of SIMWALK2 (v2.89) [185] to construct haplotypes in all genotyped family members to detect genotyping errors reflected by unlikely double recombinants. In the event that error checking indicated that a genotype was highly improbable, the genotype data for that marker were excluded for the entire family.

Allele frequencies were estimated using genotype data from all unrelated individuals in the combined dataset, consisting of more than 300 chromosomes. These allele frequencies were compared with available data from other Caucasian populations, and no significant differences were observed (data not shown). The LAPIS program of the PEDIGENE system [186] was used to output appropriate analysis files for the different programs.

Linkage was analyzed using both model-dependent and model-independent methods. For autosomes, two-point and multipoint heterogeneity LOD (HLOD) scores were calculated under both dominant and recessive models using Allegro [187]. Disease allele frequency was estimated to be 0.01 and 0.1 for dominant and recessive models, respectively. The phenocopy rate was 0.0005, and the penetrance value was set at 0.5 for the analysis. The reduced penetrance value was set at 50% given the likelihood of oligogenic inheritance and the possibility of heterogeneity, such that not all individuals having the allele would present with disease. These parameters, while minimizing our power somewhat, have a smaller impact on the lod scores when there is either linkage or no linkage and were therefore selected to be robust [116]. Phenotypic status was only considered for affected individuals, and other family members were designated as having an unknown phenotypic status. Nonparametric analysis involved calculating allele-sharing LOD* values using affected relative pair data based on an exponential model that uses the S_{pairs} scoring function as recommended by McPeek [188]. NPL scores and corresponding P-values were also calculated by Allegro. Data from the X chromosome were analyzed using ASPEX v2.5 (http://aspex.sourceforge.net/) and FASTLINK [189] v4 to calculate two-point and multipoint MLOD scores. Peak parametric (HLOD) or nonparametric LOD* scores ≥1.5 were considered as “suggestive” evidence for linkage and listed in Table 8-2, along with corresponding peak marker, deCODE

cM location, and chromosomal band position. The LOD ≥ 1.5 cutoff was chosen to provide a small and focused group of regions for initial follow-up.

The nonparametric genome-wide significance threshold [190, 191] for linkage at the P=0.05 level was determined by conducting simulations using Merlin [192] with the current dataset. The Simulate option in Merlin was used to produce 1000 random datasets that preserve the properties of the original data for marker informativeness, spacing and missing data patterns. An empirical significance threshold was determined by using the 95th percentile of the resulting distribution.

OSA [193] identifies genetically more homogeneous subsets of the overall data by ordering families according to covariate trait values in ascending or descending order. OSA takes the first family and calculates an allele-sharing LOD* score. In an iterative process, OSA successively adds families, re-calculating LOD* scores with each addition, and it identifies the division in the dataset at which maximum linkage is obtained on the chromosome being analyzed. Permutation testing using randomized data is used to determine the empirical significance of the observed results. This approach has been applied with success to identify or increase evidence in support of linkage to complex disease susceptibility loci [54, 194, 195].

To explore potential genetic interaction or other genetic correlations between sites of main effect (i.e. suggestive linkage), OSA was applied using family-specific LOD scores as the covariate trait. Families were ranked in

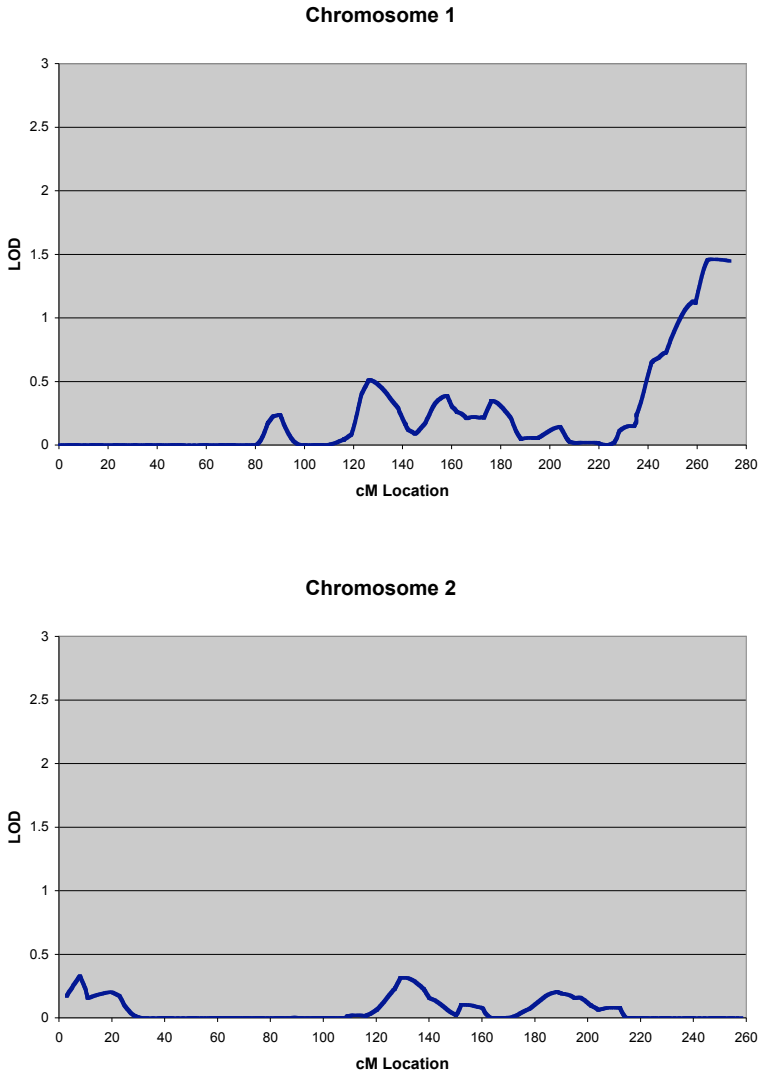
descending or ascending order, according to the family-specific LOD score at peak linkage for sites demonstrating LOD scores ≥1.5. Nonparametric allele-sharing analysis was then performed for the other six chromosomes (see Table 8-2) showing LOD scores ≥1.5. For instances of empirically significant increases in evidence for linkage, we explored the nature of the genetic correlation to ask whether it reflected clinical correlations in the respective subsets. We employed previously published clusters of correlated variables, identified by principal components analyses of ADI/ADI-R items, to represent putative phenotypic subsets in autism [76, 196]. The ADI-based factor subsets correspond to "(1) language, (2) social intent, (3) developmental milestones, (4) rigid-compulsive behaviors, (5) savant skills, and (6) sensory aversion", as determined by Folstein and colleagues [76]; and (7) "insistence on sameness" as described by Cuccaro and colleagues [196]. The description of these factors reveals significant inter-sibling correlation in the case of affected sib-pair families for all of these factors with the exception of "social intent" [76]. We thus compared the seven ADI-based factor score means (both the mean of family means and the mean of affected individuals) using a t-test for the families above and below the OSA-determined split in the dataset resulting in maximal linkage. Subsequent analysis involved specific examination of the "developmental milestones" cluster. The milestones factor indexes on the following ADI items: "*(1) To walk unaided; (2) to sit unaided on flat surface; (3) age of first single words; (4) age of first phrase; (5-6) acquisition of bladder control: daytime, night; (7) acquisition of bowel control.*"

Analysis of the “developmental milestones” factor as a potential phenotypic subset related to the autism linkage correlations was performed by applying the OSA algorithm. We used “developmental milestones” family means, normalized via SAS v9.1.2 and Box-Cox transformation procedures, as an ascending ranking covariate. LOD* scores were calculated according to the OSA algorithm, and the resulting increase in linkage achieved with the OSA-determined family subset was analyzed through permutation testing.

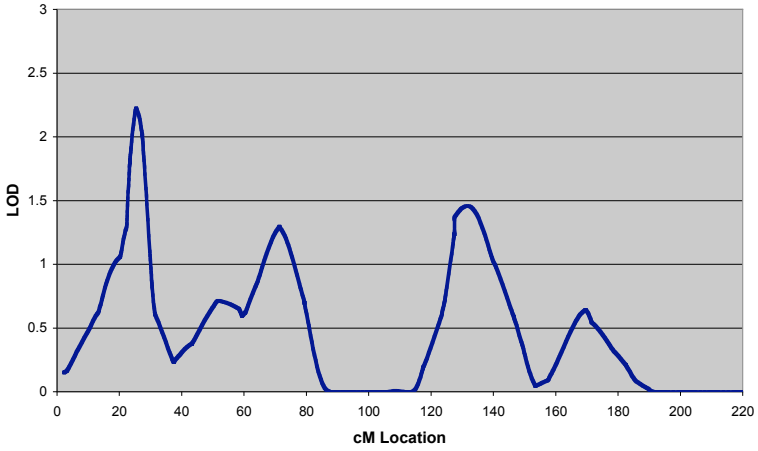
Results

Seven chromosomes revealed one or more regions of linkage with a model-dependent or model-independent LOD score ≥1.5 (Figure 8-1). No locus reached the empirically derived genome-wide significance level of 2.92. These suggestive loci include 3p25, 6q23, 12p12, 16p12-p13, 17q11, 17q21 and 19p13 (Table 8-2). Data provide the most compelling support for 17q11.2 and 19p13 as harboring autism susceptibility loci.

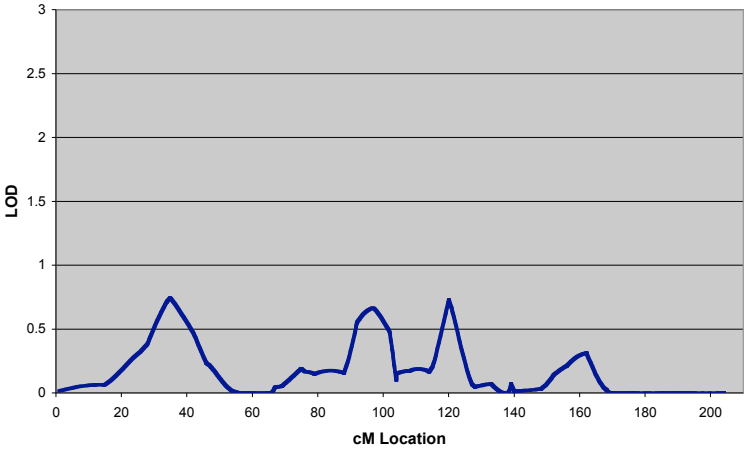
Figure 8-1. Genome-wide nonparametric linkage analysis in 158 multiplex families for autism loci. Individual plots show allele-sharing LOD* scores calculated for autosomes using Allegro and MLOD scores for the X chromosome calculated using ASPEX.



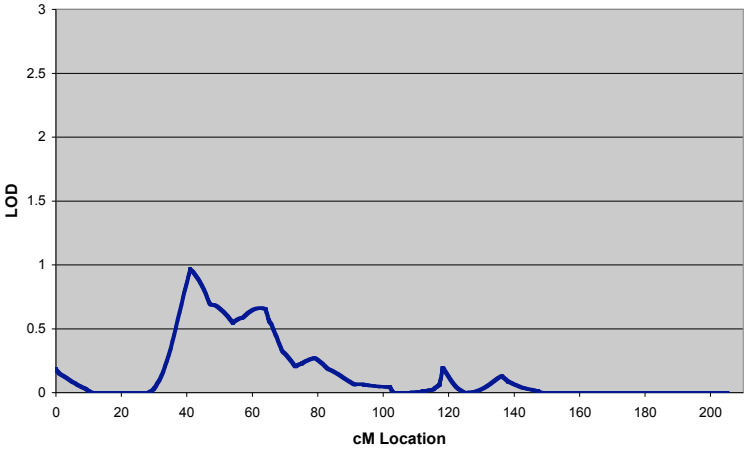
Chromosome 3



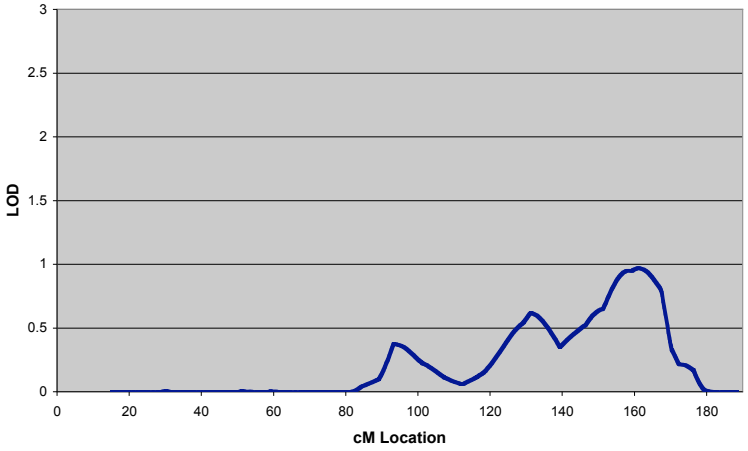
Chromosome 4



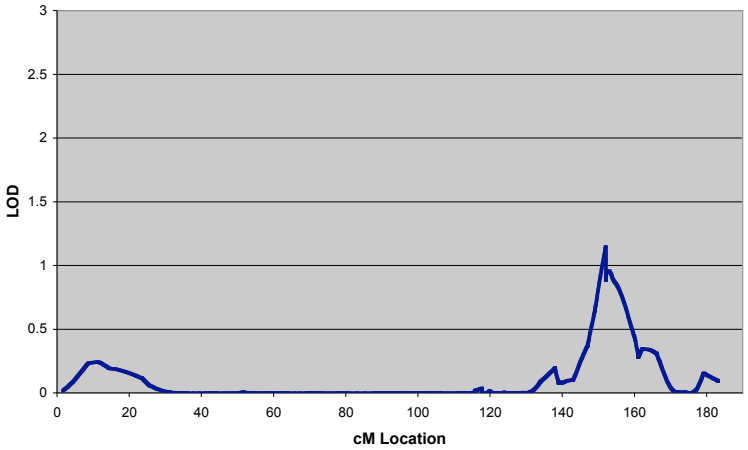
Chromosome 5



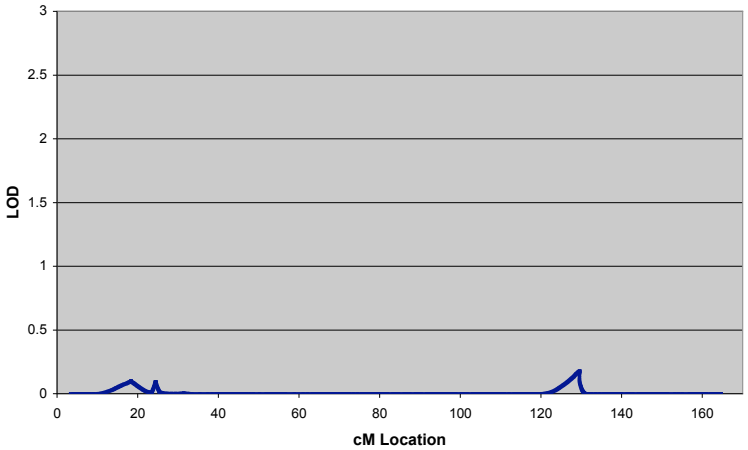
Chromosome 6



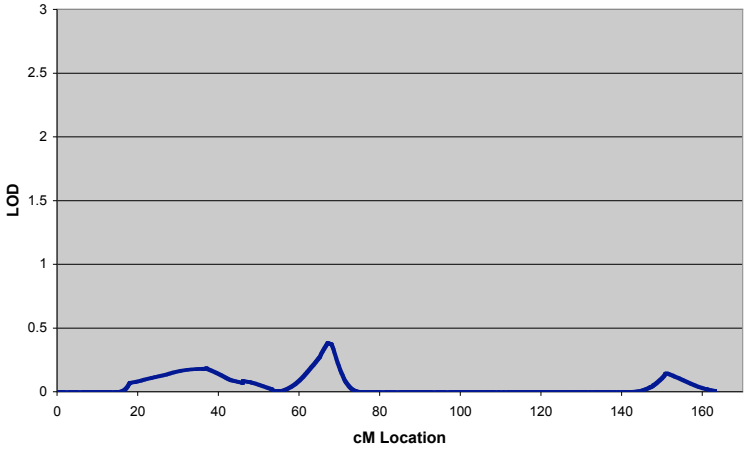
Chromosome 7



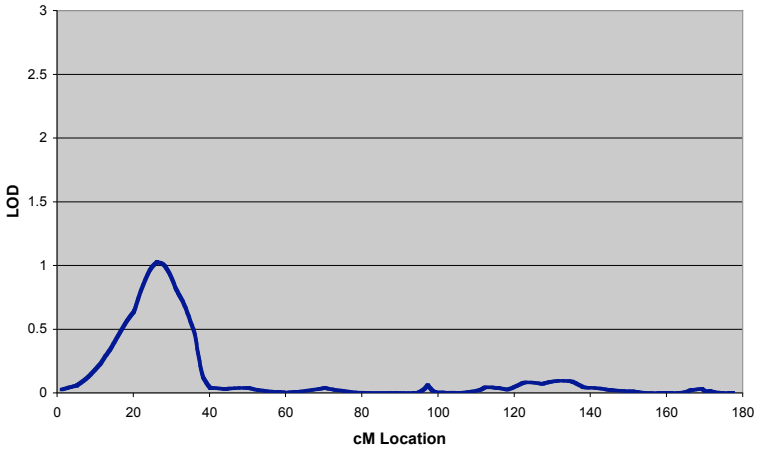
Chromosome 8



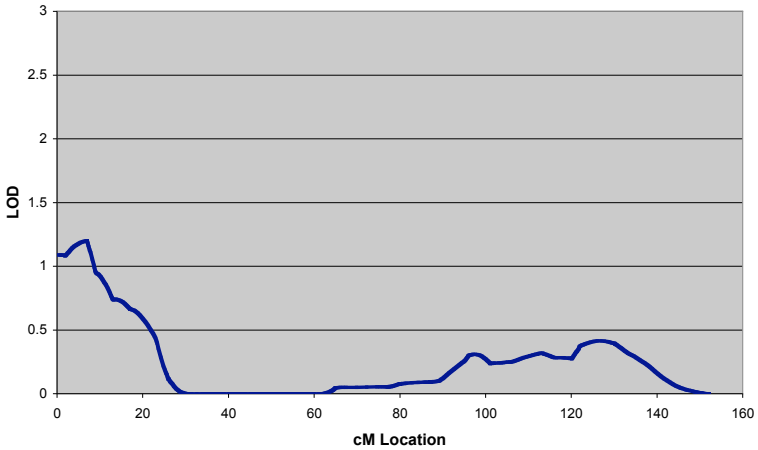
Chromosome 9



Chromosome 10

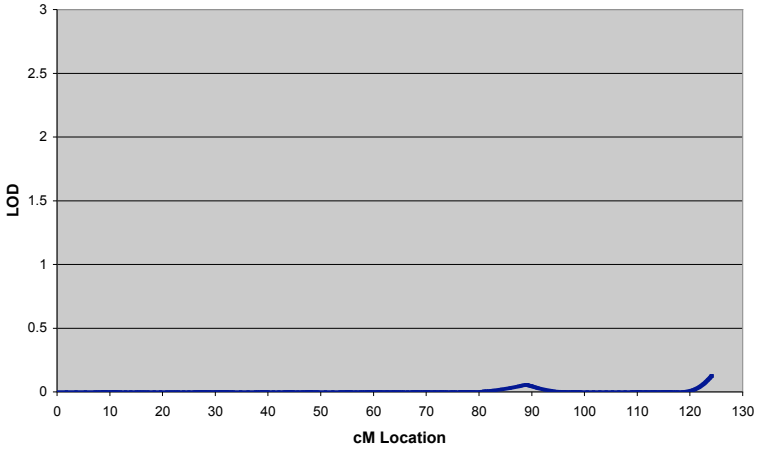


Chromosome 11

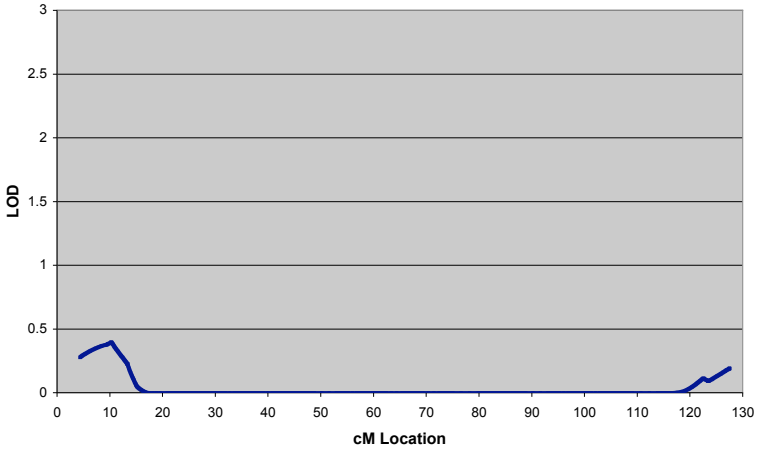


Chromosome 12

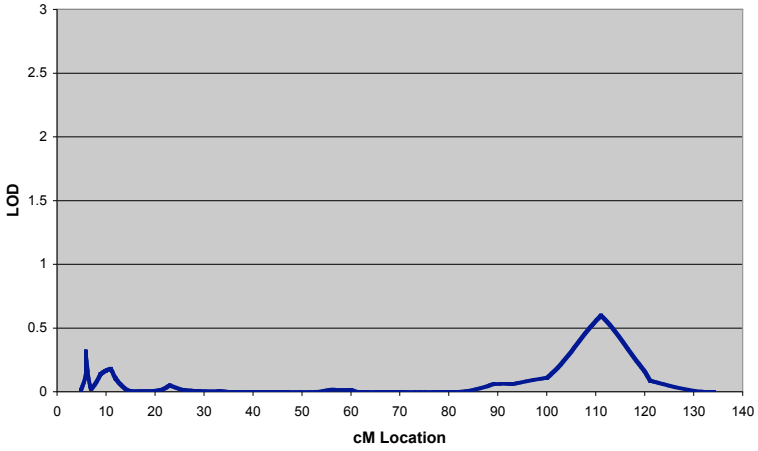
Chromosome 13



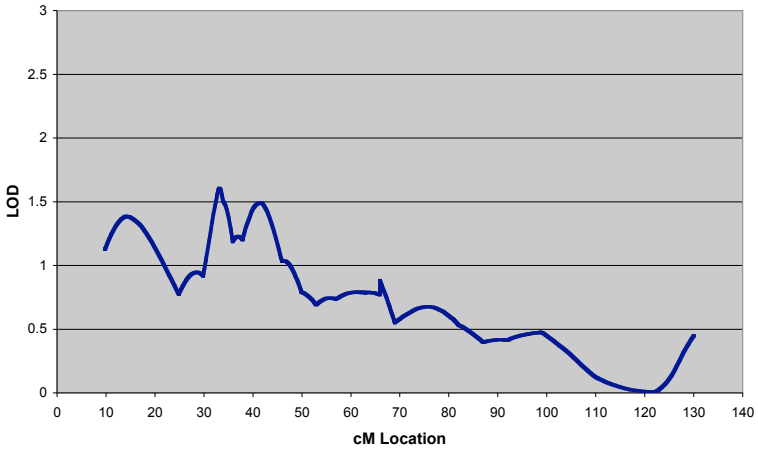
Chromosome 14



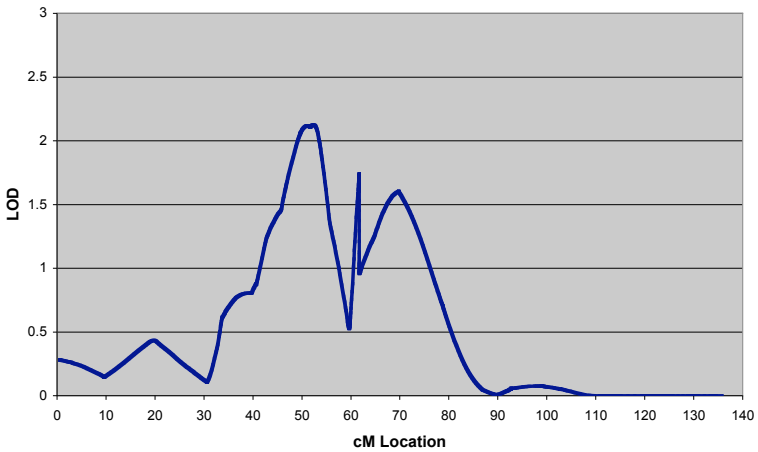
Chromosome 15



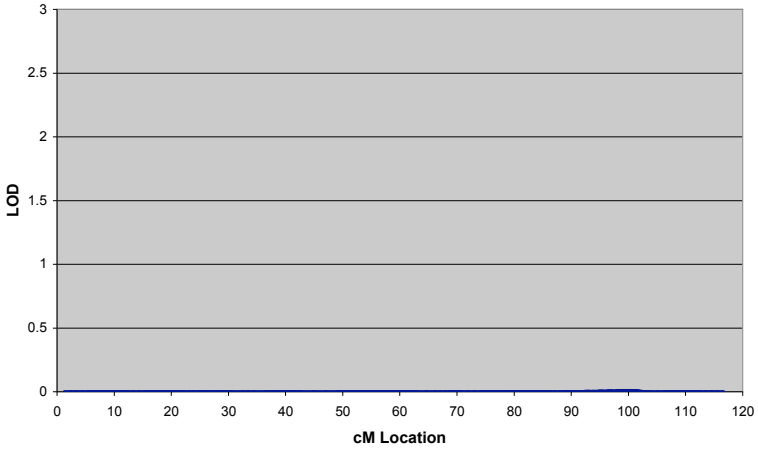
Chromosome 16



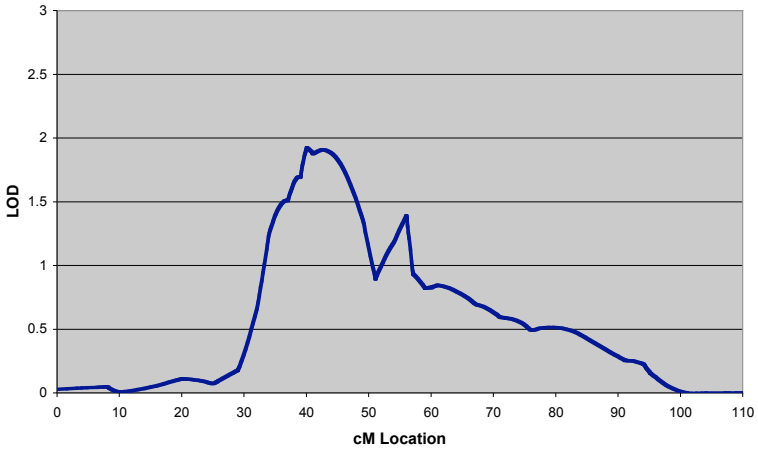
Chromosome 17



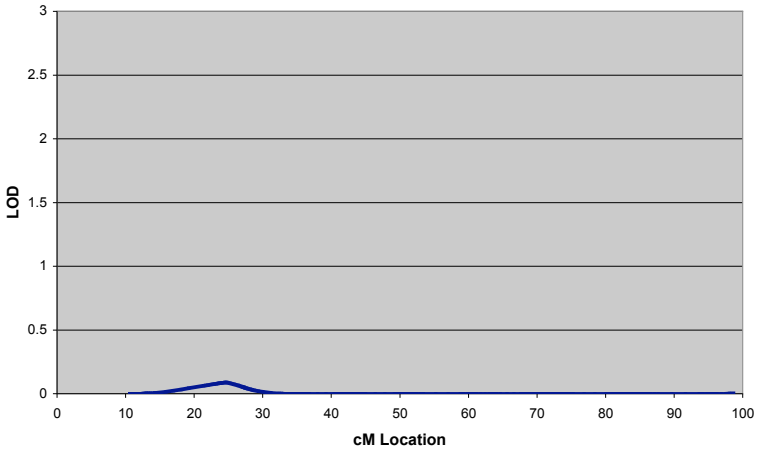
Chromosome 18



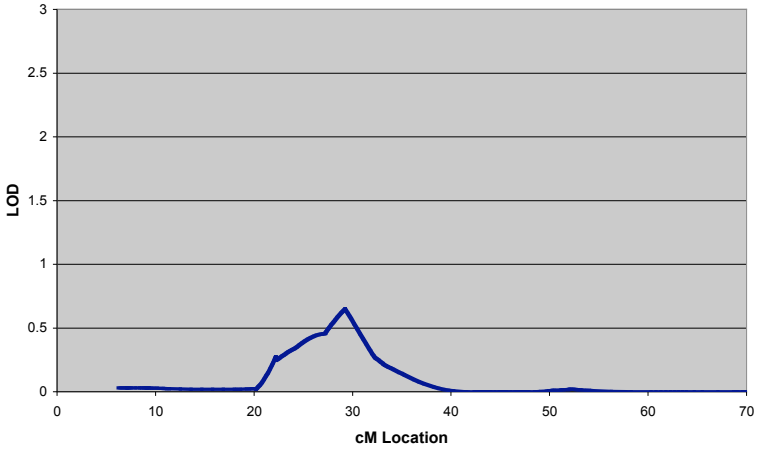
Chromosome 19



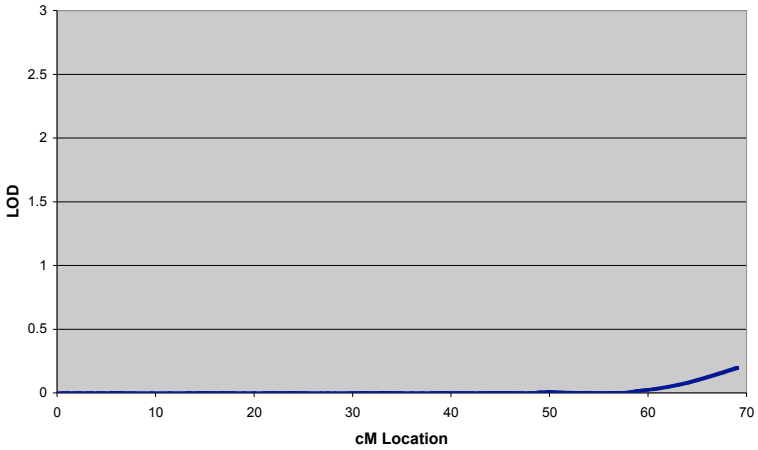
Chromosome 20



Chromosome 21



Chromosome 22



Chromosome X

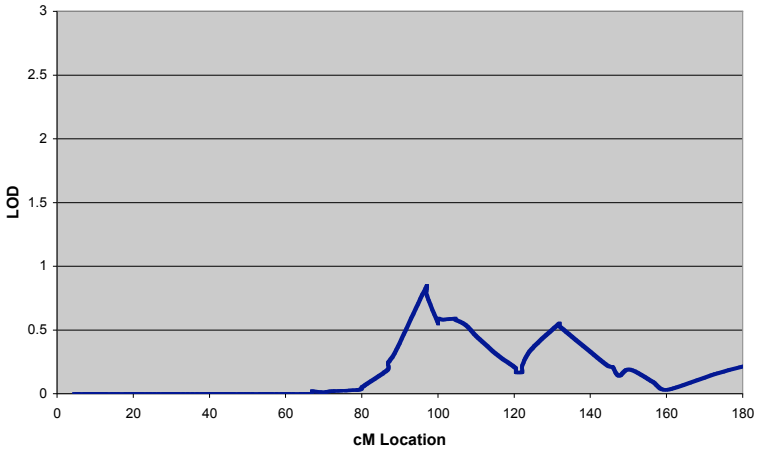


Table 8-2. Linkage data for loci with LOD scores > 1.5. Superscript (D) represents a score calculated assuming a dominant model of inheritance and superscript (R) represents a score calculated assuming a recessive model of inheritance.

Chromosome	deCODE cM	Marker	HLOD	LOD*
3p25.3	25	D3S3691	1.76^{R}	**2.22**
6q23.2	131	D6S1656	**1.61^{R}**	0.62
7q35	152	D7S2195	**1.65^{R}**	1.14
12p12.1	45	D12S1591	**1.50^{R}**	1.43
16p13.2	15	ATA41E04	**1.64^{D}**	1.38
16p13.12	33	D16S3062	**1.87^{D}**	1.60
16p12.3	43	D16S490	**1.80^{D}**	1.49
17q11.2	53	D17S1294	**2.85^{D}**	2.13
17q21.2	69	D17S1299	**1.90^{D}**	1.60
19p13.11	40	D19S930	**2.55^{R}**	1.92
19p13.11	56	D19S113	**2.20^{R}**	1.39

For 17q11.2, peak linkage was observed at 53 cM on the deCODE genetic map, corresponding to marker D17S1294 (Table 8-2), at which we see a multipoint HLOD of 2.85. Nonparametric multipoint analysis revealed an allele-sharing LOD* score of 2.13 and an NPL score of 2.84 (P=0.0024). A second telomeric linkage peak can be distinguished on 17 at ~69 cM, corresponding to 17q21.2. Marker D17S1299 at this site yielded a HLOD of 1.9 and nonparametric results were a LOD* of 1.66 and an NPL score of 2.26 (P=0.012). The more proximal peak at ~53 cM lies in close proximity (~150 kb) to the serotonin transporter (*SLC6A4*) locus, long considered to be an attractive functional candidate gene for autism and other neuropsychiatric conditions. Figure 8-2 shows multipoint LOD score plots for both dominant and recessive parametric (HLOD) and nonparametric allele-sharing LOD* values for chromosomes 17 and 19.

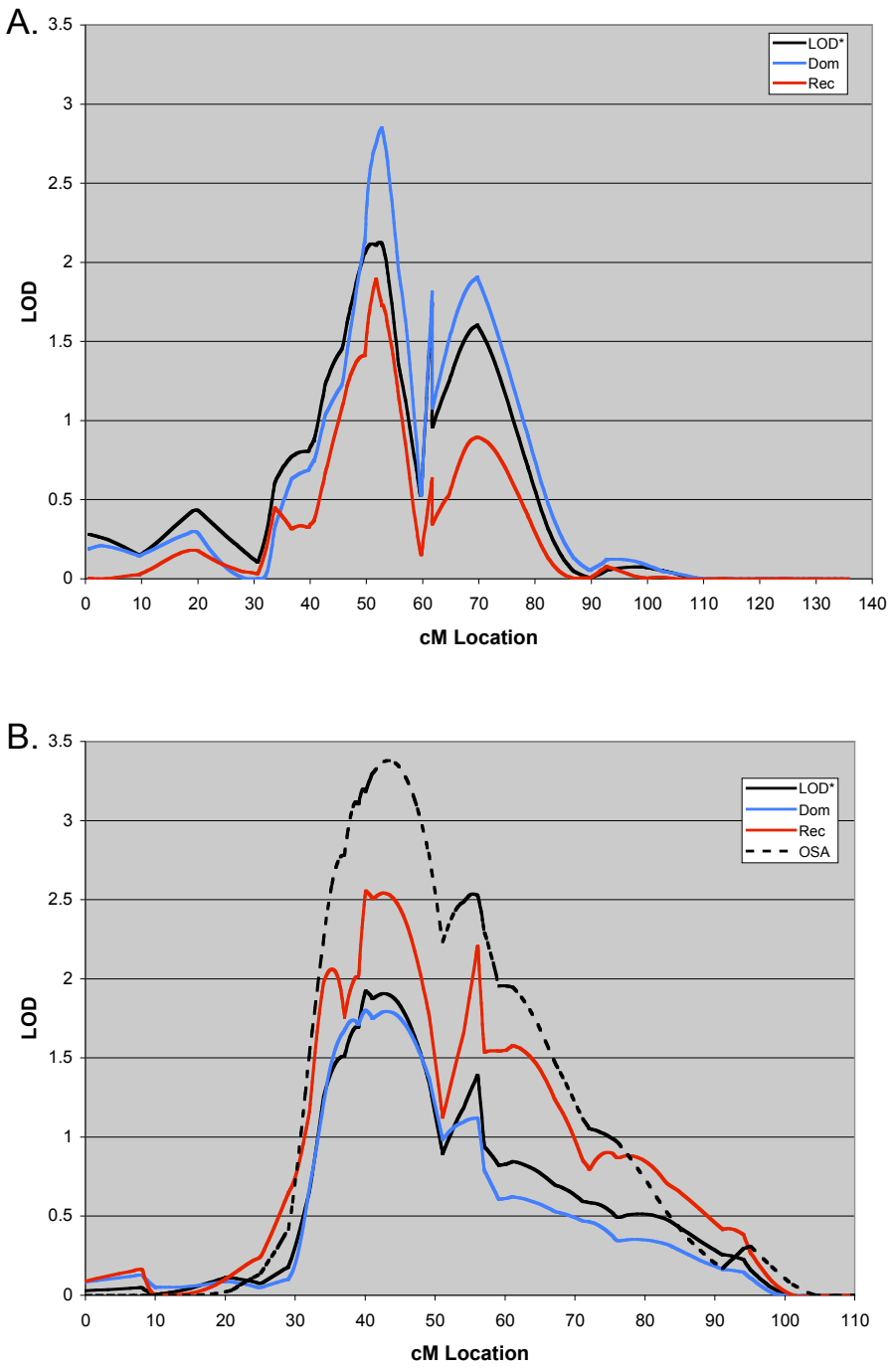


Figure 8-2. Multipoint linkage analysis under all models for chromosomes 17 (A) and 19 (B). Multipoint parametric HLOD plots for both dominant (blue) and recessive (red) models, and nonparametric allele-sharing LOD* values (black) are displayed across the respective chromosomes. OSA analysis using ascending "developmental milestones" factor scores to order families is shown for chromosome 19, for which a 92-family optimal subset was identified and used to calculate allele-sharing LOD* scores (dashed black line).

The second most significant result was observed on 19p13, where peak linkage was detected at marker D19S930, mapping to ~40 cM and yielded a multipoint HLOD of 2.55 (Table 8-2 and Figure 8-2). Nonparametric analyses at this locus showed a LOD* of 1.92 and a corresponding NPL of 2.77 (P=0.003). As with chromosome 17, the multipoint analyses show a second more telomeric peak, corresponding to marker D19S113. The recessive HLOD at this site was 2.20, with model-independent LOD* and NPL values of 1.39 and 2.10 (P=0.018), respectively.

To address the possibility of gene-gene interaction, we applied the OSA approach with family-specific LOD scores as the ranking trait. Families, almost all of which have affected sib-pairs, were ranked in both ascending and descending order using family-specific LOD scores. The three most significant correlations are presented in Figure 8-3. Using chromosome 19 lod scores (High to Low) as the covariate, the results on chromosome 17q, while non-significant (P=0.1), showed an increase in linkage at the more distal peak on 17q21.1 from a LOD* of 1.7 to 3.6 and identified an optimal subset of 52 families. Applying the same covariate, a significant increase was seen on chromosome 6q, with a smaller, completely overlapping, 30-family optimal subset. This subset resulted in an increase in LOD* values from 1.0 to 3.6 at ~164 cM (P=0.004). Another significant finding involves the 7q region, possibly representing the most replicated site of linkage in autism [28, 48, 111, 174, 176-178]. Given a substantial focus on this region over several years, we lessened our criteria to

examine any other chromosome demonstrating a LOD score >1. Application of OSA using chromosome 7q linkage data, again ranking families based on LOD scores in a descending manner, lead to a significant increase in linkage on 5p at ~41 cM from a LOD* of 1.1 to 3.3 in a 41-family subset. Thus, in these three cases, notwithstanding the non-significance of the 19p13/17q21 result, there is a positive correlation of linkage in varying but overlapping subsets of the data between these respective pair-wise locus combinations within the same set of families. There were no significant correlations found using ranked LOD scores from low to high as a covariate in any of our OSA analyses.

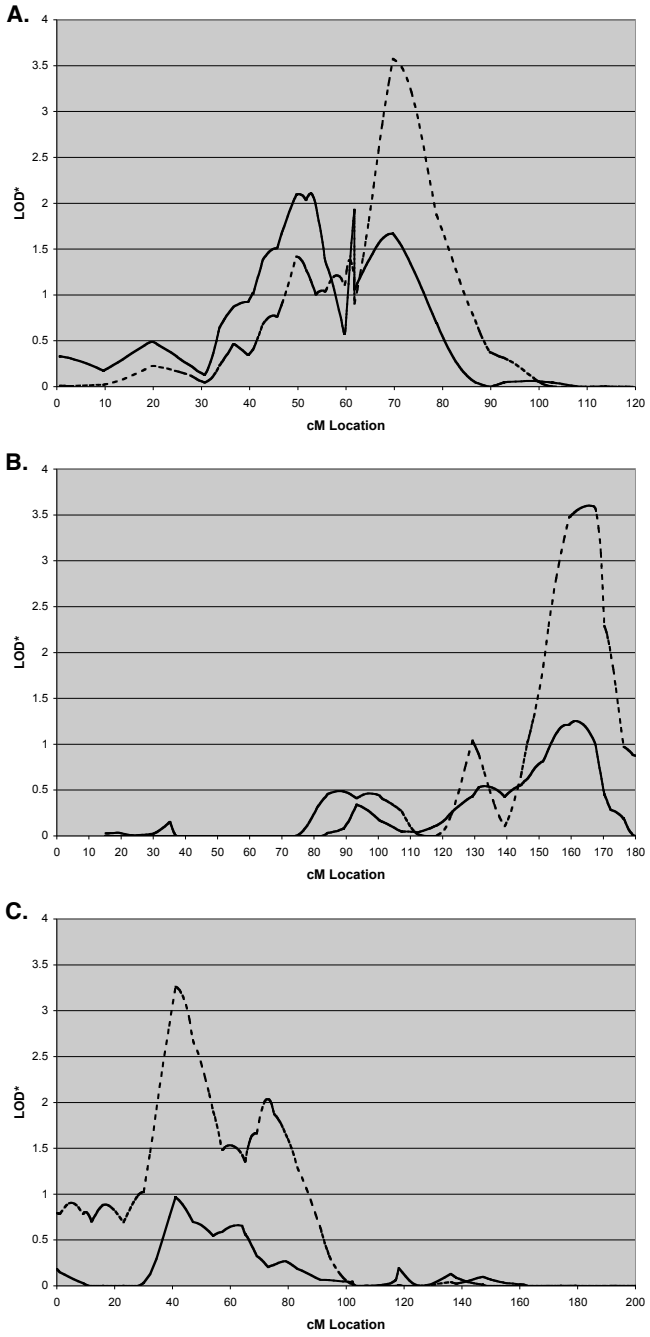


Figure 8-3. OSA using family-specific LOD scores as the ranking covariate. Families were ordered based on descending LOD scores at peak linkage for 19p13 and allele-sharing LOD scores calculated in the optimal subset for (A) chromosome 17 or (B) chromosome 6. Families were also ranked based on descending LOD scores at peak linkage on chromosome 7q (C), and LOD scores calculated for chromosome 5. Solid lines reflect multipoint LOD scores corresponding to the entire dataset for the chromosome being analyzed, while dashed lines represent analysis of the optimal subset (above the dataset division in all cases) identified from OSA; these were 52 families for chromosome 17, 30 for chromosome 6 and 41 families for chromosome 7.

To further explore the basis of the observed results, we tested the hypothesis that underlying phenotypic correlates might explain genetic correlations. We tested for differences in the mean values for the seven factor traits in the optimal subsets compared to the means of the remaining families using a t-test. This comparison for all seven available factors revealed a nominally significant difference in the chromosome 19 subsets identified through OSA analysis, of both chromosomes 17 and 6, for the "developmental milestones" factor [76]. The families in the optimal OSA subset have lower scores and therefore are more rapidly achieving developmental milestones. A similar procedure for the chromosome 7-based subsets revealed no obvious differences in any of the factors (data not shown).

To directly test the hypothesis that chromosome 19 linkage was related to reduced affection for the "developmental milestones" factor, we performed an OSA analysis in which families were ranked in ascending order based on mean values for the milestones factor score. Figure 8-2 shows the results from this analysis, which generated increased evidence for linkage to 19p13 with peak LOD* scores increasing from 1.9 to 3.4. Permutation testing revealed this increase to be empirically significant (P=0.04), thus further supporting this region as harboring a genetic risk factor.

Discussion

We have presented evidence in support of autism susceptibility loci on chromosomes 17q and 19p. Our results suggest that the 19p locus is related to a phenotypic profile involving a more rapid achievement of particular "developmental milestones". Features indexed in this ADI-based factor are: (1) ability to walk unaided; (2) ability to sit unaided on a flat surface; (3) age of first single words; (4) age of first phrase; (5-6) acquisition of bladder control: daytime and night; and (7) acquisition of bowel control. Analyses leading to this conclusion also showed positive genetic correlations between optimal OSA-defined subsets contributing to linkage at 19p13 and increases in linkage at loci on 17q21 and 6q23. A similar positive genetic correlation was shown for chromosomes 7q and 5p. However this observation lacks evidence of an underlying phenotypic relationship based on the clusters of clinical variables, or autism subsets, tested in the current study. While the increase in linkage at 17q21 was not empirically significant, the differences in "milestone" score means between the optimal chromosome 19 subsets seen for both chromosomes 17(52 families) and 6q (30 families) were significant. These exploratory data led to the significant finding of increased linkage in the single direct test of our hypothesis concerning the phenotypic correlation related to chromosome 19 linkage.

Despite the significance of the final results on 19, we remain cautious in the interpretation of the overall results. As with a number of other genomic screens in autism, no single main effect locus achieved genome-wide

significance. Support for a number of these loci, particularly at 17q11.2 and 19p13 comes from similar suggestive linkage in other genomic screens for autism. Although not all screens detect these loci (not an uncommon finding in linkage studies for complex genetic disorders), the evidence is strong regarding an effect at 19p, within 10 cM of our peak: (1) The Duke/South Carolina Collaborative Autism Team reported a maximum multipoint lod score (MMLS) of 1.21 and a MLOD=1.38 [31]; (2) PARISS (Paris Autism Research International Research Study) reported an MMLS=1.37 [27]; (3) The International Molecular Genetic Study of Autism Consortium (IMGSAC) reported an MLS of 1.16 [25]; (4) The Mt. Sinai group reported an NPL of 1.56 which increased to 2.31 when only families with obsessive-compulsive behaviors were considered for this region [175]. While a portion of our sample overlaps with those of other studies, we

Similarly, several groups have reported evidence for linkage at 17q11. The recently published AGRE follow-up genomic screen identified an MLS of 2.83 near *SLC6A4* [32]. A genome scan for attention deficit/hyperactivity disorder (ADHD) identified an MLS of 2.98 near this locus [197]. An IMGSAC follow-up screen for autism [48] reported a maximum multipoint LOD score of 2.34, with sex-specific analysis suggesting an excess sharing of paternal alleles at HTTINT2 in the *SLC6A4* gene on chromosome 17q11.2. Our own more preliminary analysis of linkage in this region, using a highly overlapping dataset to that in the current study, revealed very similar results [198]. Given our inclusion of some AGRE families, it is not completely unexpected that 17q11.2

linkage is similar to that seen the larger AGRE 2nd-stage screen [32], however AGRE families only represented about half of the overall dataset. Families recruited from the Tufts/NEMC site clearly contribute to this linkage based on the LOD score-based family subset compositions. In fact, if you examine the multipoint HLOD scores under a dominant model of inheritance for the AGRE only families versus the Tufts/NEMC families, the HLOD score at marker D17S1294 is 1.25 for the AGRE families versus 1.84 for the Tufts/NEMC.

The 17q21 locus is worth further consideration. Our data support the premise that the adjacent linkage peaks represent distinct loci and are not an artifact of primary linkage at 17q11.2. The evidence for linkage at 17q21, while weaker than that at 17q11.2 only 16 cM centromeric, specifically showed an, albeit non-significant, interlocus correlation with 19p13 linkage. Linkage at 17q11.2 in this subset of families actually decreases slightly. Of particular interest is the fact that the distal region harbors the integrin β3 (*ITGB3*) locus, which was identified recently from a genome-wide quantitative trait locus (QTL) association screen for platelet serotonin levels [199]. We see nominal evidence of linkage to autism at this site, and ~20-25% of individuals with autism have elevated levels of circulating serotonin.

The other "suggestive" (LOD ≥ 1.5) loci reported here have also been detected in other genome-wide scans for autism loci. A broad region of 7q has been detected in most screens [28, 48, 111, 174, 176, 178]. The 16p region has been identified by IMGSAC, and others [25, 30, 48, 175]. Chromosomal

abnormalities have also been reported for this region in cases of autism (reviewed in [9]). Linkage at 3p was reported by at least two groups [28, 31]. Linkage has also been reported at our 6q locus by at least one other group [27]. Thus, while not significant, the replication of these linkage observations provides support for the likelihood that many of these loci represent true sites of main effect in autism.

The application of OSA to detect putative interlocus correlations between the 19p13 and 17q21, 19p13 and 6q23, and between 7q35 and 5p are limited to some degree in significance by their highly exploratory and hypothesis-generating nature. Given the number of comparisons between loci, and the number of comparisons between optimal subset pairs (on 19p or 7q) for the traits means, the potential for type I error is increased. Therefore our interpretation must be made with appropriate caveats. Nevertheless, the multiple exploratory comparisons generated a hypothesis: *that linkage to 19p13 was related to a more rapid achievement for specific milestones.* We were able to test this hypothesis with a single analysis revealing an empirically significant increase for linkage at this site. A combination of our results related to autism linkage and Ordered-Subset Analysis for ascending milestone scores, taken in the context of replicated observations of suggestive linkage by other groups, strengthens support for the presence of an autism gene at this site. In the end, ultimate interpretation will rely upon replication of these phenomena with independent samples to confirm these observations.

Finally, our results highlight the utility of using trait-based subsets of autism to identify putative susceptibility loci for this complex disorder. We and others have hypothesized a likely increased specificity of individual risk genes and corresponding alleles for traits or subphenotypes comprising the broader autism spectrum. Therefore methods such as (1) the Ordered-Subset Analysis that has power to identify more homogeneous samples from the overall population of families, or (2) QTL linkage and association analyses have the ability to provide greater sensitivity in the discovery of disease genes in the context of locus and clinical heterogeneity. Additionally, OSA or other forms of conditional linkage analyses have the ability to uncover potential interactions between loci. This is an important concept since the inherent interdependence of proteins in common pathways or networks acting during development and normal neuronal function could be easily imagined to act genetically in concert with one another.

CHAPTER IX

REFINED LINKAGE ANALYSIS AND ASSOCIATION STUDIES IN 17q AND 19p

Introduction

The genome-wide linkage screen using the broader autism phenotype provided several potential regions for follow-up study. Our two most prominent regions, 17q and 19p, have been identified by several groups as described in Chapter VIII. To narrow our regions of interest at these two loci, we chose to incorporate the use of SNPs across a LOD-based confidence interval of approximately 1.0 for chromosome 17 and 1.5 for chromosome 19. Although linkage at a given locus is more traditionally performed using microsatellite markers, rapid evolution of SNP genotyping technologies and availability of many SNP markers with a minor allele frequency (MAF) near 0.5 makes SNP-based follow-up a reasonable and cost-effective alternative. In fact, a recent report documents that a relatively dense SNP follow-up panel can provide greater information content than that afforded by microsatellites [200]. We hypothesize that follow-up genotyping will narrow our regions of interest, and thereby limit the size of physical intervals and corresponding list of potential candidate genes. This biologically unbiased strategy will permit us to establish a potential list of positional candidates and/or provide a narrowed region for LD-based mapping of common allele-based disease risk in autism.

Materials and Methods

Families

The largest sample for this study consisted of 384 autism families (327 multiplex and 57 trio families). Two-hundred eighty-three families were obtained from the Autism Genetics Resource Exchange (AGRE; http://agre.org), 98 multiplex and trios from Tufts/ New England Medical Center, and 3 multiplex families from Vanderbilt University. A number of the AGRE samples were obtained from the NIMH Center for Collaborative Genetic Studies on Mental Disorders (http://nimhgenetics.org/). Initial samples for the linkage follow-up analysis were smaller and completed prior to the acquisition of additional families. A significant majority of "known" families are of Caucasian ethnicity; however, a number of AGRE samples do not have ethnicity information ("unknown"). All affected individuals were at least four years of age and were clinically assessed with the ADI or ADI-R and most with the Autism Diagnostic Observation Schedule (ADOS). At least one sib had to meet ADI algorithm criteria for an autism diagnosis, while additional siblings may be on the broader autism spectrum. Families were excluded from the study if probands had a known medical or neurological condition suspected to be associated with their autistic phenotype (e.g. fragile X syndrome). The procedures for clinical evaluation of affected individuals for the AGRE families have been previously described [30].

Molecular Analyses

SNPs from the dbSNP and Celera SNP databases were selected for linkage follow-up based on their map position, minor allele frequency, and potential for functional effects within constraints of their position and allele frequency. SNPs across the *CRSP7* gene region were chosen based on the above criteria with the exception of constraining selection on intermarker distance and allele frequency in the general population. Marker and exon locations and intermarker distances are based on the Celera (http://www.celeradiscoverysystem.com/index.cfm) assembly. Database reference numbers and other details for the markers studied are cited in Tables 9-1, 9-2, and 9-3. Genetic (cM) distances indicated in Table 9-1 and Table 9-2 are based on the deCODE genetic map [114]. Individual SNPs were genotyped using the TaqMan™ system developed by Applied Biosystems (Foster City, CA). PCR primers and probes for assays performed across the chromosome 19 genes are listed in Table 9-4. This information is proprietary for all the markers chosen for linkage follow-up, for which Assays-On-Demand™ were obtained from Applied Biosystems (Foster City, CA).

For TaqMan™ genotyping assays, reactions were performed in a 5 μl volume, employing 2.5 ng genomic DNA template, according to manufacturer's recommendations (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 92 °C for 15 s, 60 °C

for 1 min. Samples were analyzed using an ABI 7900HT Sequence Detection System.

Table 9-1. SNP markers for chromosome 17 linkage follow-up. Alleles are listed major/minor. Microsatellite markers spanning the follow-up region are highlighted in bold.

Gene	SNP No.	Region	Marker	Alleles	Minor Allele Frequency	Approx. cM Location	Intermarker Distance (kb)
			D17S900			41	35
HS3ST3B1	1	Intron 1	hCV2587060	C/T	0.43	41	31
			D17S921			41	225
			D17S839			43	649
PMP22	2	Exon 5 3' UTR	rs13422/ hCV774376	G/T	0.47	44	1074
TRPV2	3	Intron 6	rs7222754/ hCV12125529	C/T	0.38	45	1146
			D17S1857			46	1060
Intergenic	4	Intergenic	rs9915758/ hCV2036889	G/A	0.44	47	86
			D17S2196			47	521
NP_055949	5	Exon 4	rs3744137/ hCV1385623	C/A	0.49	47	997
MYO15A	6	Exon 16	rs2280777/ hCV2601396	T/C	0.46	48	1703
PRPSAP2	7	Intron 4	rs2305064/ hCV1129714	A/G	0.46	49	752
C17orf35	8	Intron 1	rs4450456/ hCV7444583	C/T	0.44	50	342
			D17S1871			50	1158
Intergenic	9	Intergenic	rs7215373/ hCV25636564	C/T	0.50	51	585
			D17S1824			52	776
MYO18A	10	Exon 16	rs8076604/ hCV1555487	G/A	0.47	52	485
NP_689558	11	Intron 1	rs3110492/ hCV2923617	G/A	0.43	53	460
			D17S1294			53	380
CPD	12	Intron 8	rs9913237/ hCV2980234	C/G	0.47	53	1104
RAB11FIP4	13	Intron 3	rs2017548/ hCV2181821	G/A	0.46	54	130

			D17S1800			54	883
MYO1D	14	Intron 21	rs1018866/ hCV2536985	T/G	0.39	55	461
			D17S798			56	120
ACCN1	15	Intron 3	rs7217619/ hCV2033368	C/T	0.45	56	727
			D17S1850			56	64
	16	Intron 1	rs915484/ hCV1820633	A/C	0.38	57	358
			D17S1293			60	728
CCT6B	17	Intron 1	rs1482103/ hCV7447761	G/A	0.42	63	1475
TRIP3	18	Intron 4	rs7222903/ hCV1163237	G/T	0.46	64	529
			D17S1867			65	528
AP1GBP1	19	Intron 14	rs3110621/ hCV2554333	T/C	0.39	66	960
LASP1	20	Intron 1	rs1873050/ hCV11935244	C/T	0.46	67	1111
PSMD3	21	Intron 2	rs2305482/ hCV15977478	C/A	0.43	68	652
SMARCE1	22	Intron 7	rs1474454/ hCV7571606	C/T	0.46	69	203
			D17S1299			70	1149
DNAJC7	23	Intron 2	rs12948970/ hCV11613990	T/A	0.49	71	763
Q8N4A7	24	Intron 1	rs1078523/ hCV2160077	A/G	0.48	72	1287
G6PC3	25	Exon 1 5' UTR	rs228758/ hCV557390	C/T	0.50	73	873
Intergenic	26	Intergenic	rs4793165/ hCV7915790	A/G	0.41	74	847
Intergenic	27	Intergenic	hCV2544808	A/G	0.47	74	891
GOSR2	28	Intron 6	rs758391/ hCV2275279	G/A	0.40	75	980
SP2	29	Exon 3	rs2228251/ hCV95252	T/C	0.46	76	1002
NP_075567	30	Intron 4	rs12453374/ hCV1242900	C/G	0.50	76	581
NGFR	31	Intron 1	rs575791/ hCV2305277	G/A	0.47	77	218
NP_110429	32	Intron 7	rs2017835/ hCV7476484	A/G	0.48	77	1339
SPAG9	33	Intron 3	rs2041319/ hCV11936873	T/A	0.45	78	714
CA10	34	Intron 4	rs1989800/ hCV2536374	T/A	0.38	79	451
			D17S788			79	1232

Intergenic	35	Intergenic	rs1502501/ hCV8732620	G/T	0.47	80	378
NP_115948	36	Exon 1	rs3803824/ hCV7957804	C/T	0.38	81	1295
Intergenic	37	Intergenic	rs244356/ hCV2576424	T/G	0.42	83	1204
NP_694960	38	Intron 2	rs966793/ hCV2665614	T/C	0.39	84	786
AKAP1	39	Intron 2	rs2241073/ hCV349718	G/T	0.46	85	289
			D17S957			87	_

Table 9-2. SNP markers for chromosome 19 linkage follow-up. Alleles are listed major/minor. Microsatellite markers spanning the follow-up region are highlighted in bold.

Gene	SNP No.	Region	Marker	Alleles	Minor Allele Frequency	Approx. cM Location	Intermarker Distance (kb)
			D19S221			32	59
MAN2B1	1	Exon 8	rs1054487/ hCV3004808	G/A	0.4	32	601
CACNA1A	2	Intron 27	rs2074880/ hCV346304	A/C	0.38	33	231
	3	Intron 1	rs7250452/ hCV2883344	A/G	0.42	34	240
			D19S840			34	454
LPHN1	4	Intron 2	rs40282/ hCV8728798	T/C	0.42	35	407
			D19S179			37	383
SLC1A6	5	Intron 2	rs2285980/ hCV2569724	C/G	0.46	37	647
			D19S714			37	24
Intergenic	6	Intergenic	rs1290617/ hCV7496758	T/G	0.38	37	442
TPM4	7	Intron 1	rs2278006/ hCV16006818	A/T	0.46	38	173
			D19S917			39	142
EPS15L1	8	Intron 17	rs1870071/ hCV11459169	T/C	0.28	39	179
SLC35E1	9	Promoter	rs8110418/ hCV3057319	A/G	0.11	39	60
CRSP7	10	Promoter	rs12461484/ hCV11699366	G/C	0.43	39	48
MGC3169	11	Intron 4	rs901792/ hCV1975557	T/C	0.21	39	64
Intergenic	12	Intergenic	rs812847/ hCV7494847	A/G	0.5	40	113
			D19S930			40	132
CPAMD8	13	Intron 13	rs4447554/ hCV11699091	T/A	0.49	40	208
			D19S593			41	46
NR2F6	14	Intron 1	rs891205/ hCV7493310	A/C	0.36	41	816
IL12RB1	15	Exon 18 3¢ UTR	rs404733/ hCV795434	A/T	0.46	42	844
COPE	16	Exon 8	rs2074797/ hCV2585304	T/C	0.46	43	40
HOMER3	17	Intron 1	rs1122821/ hCV11462980	A/G	0.41	44	241
MEF2B	18	Intron 4	rs2239369/ hCV2537047	T/A	0.47	44	819
Intergenic	19	Intergenic	rs386976/ hCV2305565	T/C	0.47	45	1360
Intergenic	20	Intergenic	rs2012013/ hCV7497240	C/T	0.46	46	1149
Q96IR2	21	Intron 3	rs2915929/ hCV8163344	A/C	0.41	48	2467
			D19S419			49	533
Intergenic	22	Intergenic	rs347787/ hCV3057412	G/A	0.46	50	1478
CCNE1	23	Intron 4	rs997669/ hCV2885582	T/C	0.37	51	108
			D19S433			51	_

Table 9-3. SNP markers across the *CRSP7* and *SLC35E1* region of chromosome 19p13. Alleles are listed major/minor.

Gene	SNP No.	Region	dbSNP rs# /Celera hCV#	Alleles	Minor Allele Frequency	Intermarker Distance (bp)
EPS15L1	1	Intron 2	rs2305781/ hCV16191652	G/A	0.13	75,910
CHERP	2	Exon 17	rs17029/ hCV2885433	C/T	0.27	14,193
	3	Intron 5	rs8101084/ hCV29287905	G/A	0.12	19,156
SLC35E1	4	Intron 6	rs12461181/ hCV3057302	T/C	0.07	3,658
	5	Intron 5	rs731617/ hCV3057305	C/T	0.34	17,776
	6	Promoter	rs8110418/ hCV3057319	A/G	0.10	33,210
CRSP7	7	Intron 1	rs751788/ hCV795308	G/C	0.01	13,351
	8	Intron 1	rs3786602/ hCV3057145	A/G	0.33	1,524
	9	Intron 1	rs11668934/ hCV3057143	G/A	0.05	1,665
	10	Intron 1	rs10408776/ hCV3057144	C/T	0.14	2,707
	11	Intron 1	rs3786604/ hCV27482120	A/G	0.16	4,023
	12	Promoter	rs3760673/ hCV31765862	T/C	0.33	1,138
	13	Promoter	pcr3SNP1	G/A	0.11	296
	14	Promoter	rs7351094/ hCV29287914	G/T	0.11	197
	15	Promoter	rs10425272/ hCV30483070	A/G	0.10	481
	16	Promoter	pcr2SNP1	T/C	0.05	983
	17	Promoter	rs12461484/ hCV11699366	G/C	0.43	12,973
Intergenic	18	Intergenic	rs4808051/ hCV27896147	C/T	0.07	1,526
MGC2747	19	Exon 4	rs10402/ hCV11459109	T/C	0.05	201
	20	Exon 3	rs706762/ hCV8931897	G/A	0.08	32,950
MGC3169	21	Exon 3	rs730120/ hCV970369	G/A	0.05	584
	22	Intron 4	rs901792/ hCV1975557	T/C	0.20	_

Table 9-4. PCR and genotyping primers for the SNPs spanning the 19p13 region. Labels are as follows: **(**F) Forward Primer, (R) Reverse Primer, (FP) FP-TDI extension primer, VIC and FAM are fluorescent labels for TaqMan allelic discrimination probes (Bold highlights SNP), AbD is an Assay-By-Design from ABI, AoD is an Assay-On-Demand from ABI. MGB stands for minor groove binder, and NFQ stands for non-fluorescent quencher.

SNP No.	Primer sequences (5´-3´)		Product size (bp)	Assay
1	F	GCTGACAGTGAACAGAAAAGGACAA	78	TaqMan (AbD)
	R	GCATGCCTGAGCTGAAAGC		
	VIC	VIC-CAGAATAGAAGCC**T**GGCTG-MGB-NFQ		
	FAM	FAM-CAGAATAGAAGCC**C**GGCTG-MGB-NFQ		
2	NA	Unavailable	< 200	TaqMan (AoD)
3	F	GTGTGCGTGCAAATCAGTGA	63	TaqMan (AbD)
	R	CAGCCACGGCCTCCTT		
	VIC	VIC-CTGTGTCTGT**T**TCCTG-MGB-NFQ		
	FAM	FAM-CTGTGTCTGT**C**TCCTG-MGB-NFQ		
4	NA	Unavailable	< 200	TaqMan (AoD)
5	F	CGCCAAGATCATAAACCAGTCAGT	66	TaqMan (AbD)
	R	TCCCGGGCGCTCAAG		
	VIC	VIC-ATGTCA**G**AGCTTTGC-MGB-NFQ		
	FAM	FAM-TATGTCA**A**AGCTTTGC-MGB-NFQ		
6	NA	Unavailable	< 200	TaqMan (AoD)
7	F	GCTGTCTGGTGTGGTAGTGATC	71	TaqMan (AbD)
	R	CCCTTCTTCCTTCACTCAAAACTGA		
	VIC	VIC-CAAGGGATA**G**AGCATC-MGB-NFQ		
	FAM	FAM-CAAGGGATA**C**AGCATC-MGB-NFQ		
8	F	GGTGGCCACATCCTCGTT	65	TaqMan (AbD)
	R	AGGAGAGCCGTTAGCTACCAT		
	VIC	VIC-CCAGGGTGATGT**A**GTAGT-MGB-NFQ		
	FAM	FAM-CAGGGTGATGT**G**GTAGT-MGB-NFQ		
9	F	CATCAGAAACCTTCTGGAAGCTTCA	87	TaqMan (AbD)
	R	GGAGTAGGTTCTGTCCTCGAATTTT		
	VIC	VIC-CATAACAACAG**G**TGTGCTT-MGB-NFQ		
	FAM	FAM-CATAACAACAG**A**TGTGCTT-MGB-NFQ		
10	F	CCCATGCTTGGAGCACCTT	58	TaqMan (AbD)
	R	TCCACACCCTGCTCTAAGAGT		
	VIC	VIC-AAAGC**G**GCTGTTTTG-MGB-NFQ		
	FAM	FAM-AAAAGC**A**GCTGTTTTG-MGB-NFQ		
11	F	CCTGCTGCTATAGGAATTCAAATGC	90	TaqMan (AbD)
	R	CCTTGACAAAGGAATTCTTAGGAGGAA		
	VIC	VIC-CTCTCACTTT**T**AGCCCC-MGB-NFQ		
	FAM	FAM-CTCACTTT**C**AGCCCC-MGB-NFQ		
12	F	GAGGCCGCCGAGCTT	56	TaqMan (AbD)
	R	AGCTCTGTTGGAAATGCTCTTGT		
	VIC	VIC-CCTTTCCA**T**CCCGTCTC-MGB-NFQ		
	FAM	FAM-CTTTCCA**C**CCCGTCTC-MGB-NFQ		

13	F	CCAAGTCATGTCAGTGCTTCTTAGT	100	TaqMan (AbD)
	R	CTGGGTGACAGAGCAAGACT		
	VIC	VIC-TCCAGATGTTT**G**TTTATTT-MGB-NFQ		
	FAM	FAM-TCCAGATGTTT**A**TTTATTT-MGB-NFQ		
14	F	ACACCCTTTTACGTACATTAAATGTGGAA	118	TaqMan (AbD)
	R	GAGCTATTTCTGGAATGAATGAAATACTTCATAAAAT		
	VIC	VIC-CCGTGTGTGT**G**TTGTGTA-MGB-NFQ		
	FAM	FAM-CCGTGTGTGT**T**TTGTGTA-MGB-NFQ		
15	F	CGAACCATCACAGCAAGAGGAATTA	90	TaqMan (AbD)
	R	CAACTTTCTGCATTGAGACATCCTT		
	VIC	VIC-CATCTTCGCATA**A**CTGACA-MGB-NFQ		
	FAM	FAM-TCTTCGCATA**G**CTGACA-MGB-NFQ		
16	F	CCCATGGCAAGAGCAAGTCAT	97	TaqMan (AbD)
	R	CAGAGCGAGACTCCATCACAAATAT		
	VIC	VIC-TCCTCTCCCTTT**T**TTCTAT-MGB-NFQ		
	FAM	FAM-CCTCTCCCTTT**C**TTCTAT-MGB-NFQ		
17	F	AGACCTCAATTTTTTAATCCCATCTTCAAGT	94	TaqMan (AbD)
	R	CACCCCCCAACCAGGTT		
	VIC	VIC-ACTGGAACCTTT**G**TGTATC-MGB-NFQ		
	FAM	FAM-ACTGGAACCTTT**C**TGTATC-MGB-NFQ		
18	F	GTCTGAATAAGGCAGGAAGGTTTCT	81	TaqMan (AbD)
	R	GAACAGGCACGCCAAGATC		
	VIC	VIC-CAGGCCA**A**ATGAAT-MGB-NFQ		
	FAM	FAM-CAGGCCA**G**ATGAAT-MGB-NFQ		
19	NA	Unavailable	< 200	TaqMan (AoD)
20	F	CGTGGTGGGTGGAAAAGTGT	52	TaqMan (AbD)
	R	ACGTGGCTGGGAAACCAT		
	VIC	VIC-CGCC**G**AGGTGTGC-MGB-NFQ		
	FAM	FAM-TCGCC**A**AGGTGTGC-MGB-NFQ		
21	F	GAGCAGATCAGCCAGGATGTAG	52	TaqMan (AbD)
	R	GCTGTGCGCCATGCT		
	VIC	VIC-CTCCC**G**AAGCAATG-MGB-NFQ		
	FAM	FAM-CTCCC**A**AAGCAATG-MGB-NFQ		
22	F	ACAATGAAAAATCATCCCAGTGGAGAA	79	TaqMan (AbD)
	R	GGAGCAATGGCTTTCCTAGCT		
	VIC	VIC-CCACTG**G**ACCAAAAG-MGB-NFQ		
	FAM	FAM-CCACTG**A**ACCAAAAG-MGB-NFQ		

Variant screening was performed on exonic sequence and non-coding putative promoter regions across the *CRSP7* and *SLC35E1* loci. Our laboratory screened these regions through direct sequencing. PCR products for re-sequenced samples were submitted to the Vanderbilt Shared Sequencing Resource Facility after initial PCR reactions were performed in the lab. PCR reaction volumes were 8 μl, employing 10 ng genomic DNA template, 0.2 μM

primers, 125 μM dNTPs and AmpliTaq Gold DNA polymerase and buffer (Applied Biosystems). Cycling conditions included an initial denaturation at 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 s, optimal annealing temperature (T_A°C) for 30 s, and 72 °C for 1 min, and a final extension at 72 °C for 10 min. Post-PCR reaction cleanup was performed using Exo-Sap-It (USB) and then sent to the core for sequencing using the Big-Dye terminator system. Larger regions were amplified from both directions to get complete sequence information for the region of interest. Some sequencing reactions were carried out with universal primers (M13F (-21) 5′-TGTAAAACGACGGCCAGT-3′ and M13R 5′-CAGGAAACAGCTATGAC-3′). Sequencing primers for all sequenced intervals are shown in Table 9-5.

Table 9-5. Sequencing primers for variant screening across the *CRSP7* and *SLC35E1* loci. Labels are as follows: (F) Forward Primer, (R) Reverse Primer, (T_A) Annealing temperature. *Note that some sequencing reactions were carried out with universal primers (M13F(-21) 5′-TGTAAAACGACGGCCAGT-3′ and M13R 5′- CAGGAAACAGCTATGAC-3′) added to the 5′ end of the listed primer.

Gene	Region		Primer sequences (5′-3′)	Product Size (bp)	T_A (°C)
CRSP7	Snp1	F*	CGCTGGCAGAAATTTTTAAAACCCTT	708	64
		R*	CCACAGGGTCAGAGGGTATGAGCTT		
	PCR1	F*	GGAAAAGGCAAGGAAGTTAGTGTC	942	59
		R*	CCTTTCTGTATCCCCATCCCTAC		
	PCR2	F*	TCATCAAATACACGGTAGCG	913	59
		R*	ACTTGACCTCCTGACACTAACTTC		
	PCR3	F*	TTTTGCATGGGTCAGGTTATT	930	59
		R*	TGTGTGTTGTGTATCCTGGTGTAT		
	PCR4	F*	AGACCCTTCTACCCGCCTCT	772	59
		R*	CTTAAAATAACCTGACCCATGCAA		
Gene	Exon		Primer sequences (5′-3′)	Product Size (bp)	T_A (°C)
CRSP7	1	F	CGGCGGGTACTTACGTTGCTC	423	59.3
		R	CTCCCAACCTCCTCAGCGTG		
	2	F	CATCAAGGCCACTGTAAGGA	802	52.0
		R	CATCAGATGGCAGAGTAGAGG		
	3	F*	ATTTTGTTGAGGTGGGGGTAGT	895	57.5
		R*	GTCATGTGCCAGAAGCTTCTC		
	3	F	AGCTTCTGGCACATGACA	655	60.2
		R	GTCAGGACACACAGGGTA		
	3	F	CACTGGGCTTTCCGCACTG	962	57.5
		R	TTCACTGTCAGGGCCACTTCC		
SLC35E1	1	F	GCTTCCTCCTGCCACGGCTGTT	764	62.6
		R	CGACCAATGGGAAACGGCGTAG		
	2	F	AACGGGGAGGAAAATTGGATCTGA	575	57.3
		R	TTTGGTAGGGGAGGGACACGC		
	3	F	CCGGCCAGAAAAACAAACTATT	445	62.8
		R	TTTCAAAGTGTTAGGTGGTATCCC		
	4	F	TGAGGCATTACAGAGGACTGACTG	283	53.3
		R	TCCAGGTCTTGCGAGATTCAC		
	5	F	CAGTGGCAAAGCTCTGACATACAA	395	56.3
		R	AGTGCTGTCTTTACCTTCCGGTCT		
	6	F	GCGACACCATGCCCGGCTAAT	1512	64.0
		R	CGCAGCCAGCCAGTTGATTGACTT		
	7	F	ACTTCCCATGTTATTCCGTTCAAT	1112	56.8
		R	GCGCAGTCTCAGCTCACTACAA		

Statistical Analyses

Initial analysis of SNP genotype data involved quality control checks consisting of verification of internal controls and assessment of Mendelian inconsistencies, followed by final haplotype consistency analyses using Simwalk2 [117]. Conformity with anticipated Hardy-Weinberg equilibrium expectations was established.

Linkage was analyzed using both model-dependent and model-independent methods. Multipoint heterogeneity LOD (HLOD) scores were calculated under both dominant and recessive models using Allegro [187]. Disease allele frequency was estimated to be 0.01 and 0.1 for dominant and recessive models, respectively. The phenocopy rate was 0.0005, and the penetrance value was set at 0.5 for the analysis. These simple models were chosen somewhat arbitrarily, simply making the disease allele frequency 10-fold greater under a recessive model than under the dominant model. The reduced penetrance value was set at 50% given the likelihood of oligogenic inheritance and the possibility of heterogeneity, such that not all individuals having the allele would present with disease. These parameters, while minimizing our power somewhat, have a smaller impact on the lod scores when there is either linkage or no linkage and were therefore selected to be robust [116]. Phenotypic status was only considered for affected individuals, and other family members were designated as having an unknown phenotypic status. Nonparametric analysis involved calculating allele-sharing LOD* values using affected relative pair data

based on an exponential model that uses the S_{pairs} scoring function as recommended by McPeek [188].

Transmission disequilibrium (TD) in autism families was determined using the pedigree disequilibrium test (PDT) statistic, developed for use with general pedigrees [119]. Maternal and paternal specific transmissions were also assessed using the T_{SP} test [156]. Common haplotypes (≥5%) were determined using Haploview [129], and analyzed using the FBAT analysis software package [130]. TD for haplotypes was performed using haplotype tag SNPs for multi-marker blocks. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$. Common haplotypes were determined using Haploview, and analyzed using the FBAT analysis software package [130]. Results were considered significant at the nominal level for markers or haplotypes with $P \leq 0.05$.

In addition to TD being measured in all families, based on the well-known male bias in people affected by autism and observations of sex-specific biases in genetic risk by Stone et al., we performed analyses using only those families containing male-affected individuals [79]. Thereby excluding from these additional analyses any family containing a female-affected individual.

Visualization Tools for Alignment (VISTA) analysis was performed via web-based submission (http://www.gsd.lbl.gov/vista/) of human and mouse genomic sequence [121]. Regions showing evolutionarily conserved non-coding sequence were identified for future variant screening purposes.

Results

To move forward with a more detailed examination of apparent genetic risk being detected in 19p and 17q, we attempted to define a narrowed interval of linkage for both regions, by increasing information content through genotyping SNP markers in an ~1-cM grid across linked intervals. Markers were chosen to cover LOD-based confidence intervals. The chromosome 17 follow-up involved 39 markers, spanning approximately 46 cM. For chromosome 19p, 23 markers were genotyped across approximately 19 cM. Markers were spaced at approximately 1-cM intervals and chosen specifically to lie within genes in the two intervals. An identical 158-family dataset, examined in the initial genome-wide linkage scan and detailed in Chapter VIII, was used for our linkage follow-up analysis. Figure 9-1 shows the multipoint parametric HLOD plots for both chromosome 17 (under a dominant model) and chromosome 19 (under a recessive model). Plots reflecting data from only the SNP follow-up panel are shown, as well as a plot for combined data in which the SNP genotype information was integrated with the existing microsatellite data. Although the information content across the follow-up regions did not significantly increase with the addition of our follow-up markers, our results did help to prioritize the region of greatest interest below our original broader peaks. The plots for chromosome 17 demonstrate a narrowing of the interval, for the centromeric (more prominent) peak, to roughly 6 cM (50-56 cM on the deCODE map) from the original ~20 cM region (40-60 cM). The peak corresponding to the narrowed

interval remains centered directly under *SLC6A4*. The plot containing combined data for chromosome 19 not only substantially narrows the region of interest from the 18 cM (32-50 cM) to a 2 cM region (~39-41 cM), but also increases our peak LOD score from 2.6 to 3.1.

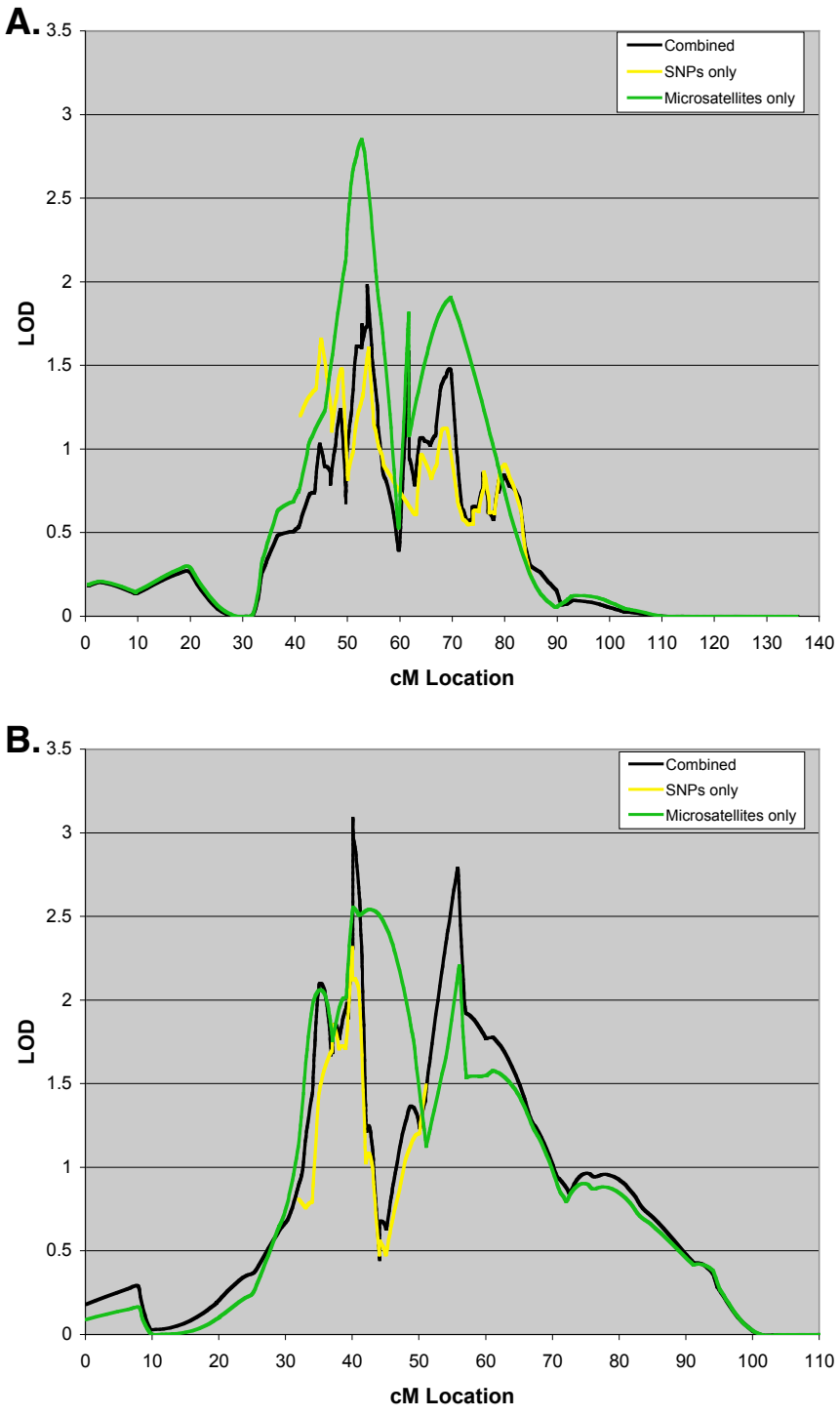


Figure 9-1. Follow-up multipoint linkage analysis for chromosomes 17 (A) and 19 (B). Multipoint parametric HLOD plots for chromosome 17 (Dominant model) and chromosome 19 (Recessive model) are shown. Linkage analysis showing our original "Microsatellites only" (green), our follow-up "SNPs only" (yellow), and the "Combined" set of microsatellite and SNP markers (black).

While recognizing a substantial improbability of detecting meaningful association using markers spaced at 1 cM intervals, we nevertheless took advantage of availability to this data by testing available families for evidence of allelic association. Given that follow-up markers were chosen not based on map position and high minor allele frequency alone, but also based on their location within genes; thus we hoped to bias towards the ability to detect, by virtue of strong LD, the potential for functional effects. Family-based TD analysis was performed using PDT, and we tested in a more exploratory fashion for maternal- or paternal-specific TD effects using the T_{sp} statistic. Results of these analyses are shown in Table 9-6. While the data for chromosome 17 are generally negative, we identified several markers on 19p that demonstrated nominal evidence for association to autism. Although there is a cluster of significant values, linkage follow-up marker 10 (hCV11699366) yielded the most significant finding.

Table 9-6. PDT analysis of follow-up SNPs across 17q and 19p.

Chromosome	Marker No.	Overall PDT (P)	Tsp Father (P)	Tsp Mother (P)
17	1	0.21	0.29	0.16
	2	0.13	0.37	0.65
	3	0.33	0.31	0.52
	4	0.95	0.29	0.81
	5	0.68	0.26	0.40
	6	0.58	0.38	0.77
	7	0.24	0.29	0.09
	8	0.62	0.41	0.31
	9	0.66	0.86	0.91
	10	0.46	0.50	0.53
	11	0.30	0.34	0.23
	12	0.69	0.54	1.00
	13	0.08	0.11	0.57
	14	0.81	0.73	0.20
	15	0.13	**0.03**	0.48
	16	0.86	0.75	0.81
	17	0.95	0.60	0.67
	18	0.32	0.61	0.37
	19	0.25	0.43	**0.04**
	20	0.65	1.00	0.83
	21	0.13	0.20	0.44
	22	0.86	0.62	0.91
	23	0.60	1.00	0.77
	24	0.09	0.31	0.46
	25	0.76	0.68	0.59
	26	0.82	0.61	0.92
	27	0.96	0.83	0.48
	28	0.13	0.17	0.09
	29	0.77	0.92	0.92
	30	0.80	1.00	0.80
	31	0.50	0.78	0.30
	32	0.25	0.85	0.25
	33	0.72	0.21	0.74
	34	0.31	0.56	0.80
	35	0.15	0.90	0.35
	36	0.95	0.92	0.63
	37	0.07	0.18	0.82
	38	0.23	0.61	0.82
	39	0.39	0.23	0.84

Chromosome	Marker No.	dbSNP rs#/ Celera hCV#	Overall PDT (P)	Tsp Father (P)	Tsp Mother (P)
19	1		0.16	0.92	0.33
	2		0.32	0.77	**0.02**
	3		0.60	1.00	0.50
	4		0.67	0.23	0.12
	5		0.72	0.27	0.83
	6		0.36	0.64	0.23
	7		0.95	0.64	0.53
	8		**0.04**	0.23	**0.03**
	9	rs8110418/ hCV3057319	**0.01**	0.25	**0.04**
	10	rs12461484/ hCV11699366	**0.01**	0.10	0.09
	11		0.38	0.12	0.51
	12		0.15	0.55	0.85
	13		0.20	0.49	0.36
	14		0.74	0.92	0.19
	15		0.49	0.30	0.25
	16		0.58	0.71	0.36
	17		0.23	0.84	0.54
	18		**0.05**	0.30	0.55
	19		0.39	0.59	0.70
	20		0.49	0.09	0.06
	21		0.86	1.00	0.29
	22		0.08	0.92	**0.05**
	23		0.65	0.74	0.76

The nominally significant TD values were clustering around a group of genes at approximately 40 cM. Continuing with the unbiased approach to detect common disease-associated risk alleles, additional markers were selected and genotyped across the ~300-kb region wherein association was detected. Particular focus was placed around marker hCV11699366, demonstrating

relatively significant TD (P = 0.0012, rounded to 2 decimal places in Table 9-6). Analysis of additional markers in the 158-family dataset surrounding hCV11699366, located in the promoter of *CRSP7*, revealed five SNPs (including hCV11699366) that showed nominally significant evidence for TD. Moreover, examination of haplotype blocks using both HapMap data, as well as our own genotype data, revealed that all five of these markers were located within a single haplotype block containing three genes: *CRSP7*, *SLC35E1*, and *CHERP*. Specific markers, significant associations, corresponding genes, intermarker LD relationships based on D´ values are all shown in Figure 9-2. HapMap LD data, based on 30 CEPH trios, are compared to that generated from our much larger dataset; the resulting increased information is reflected in the difference between the Haploview LD plot using HapMap data (Figure 9-2, lower half) and Haploview output generated from our own data (Figure 9-2, upper half).

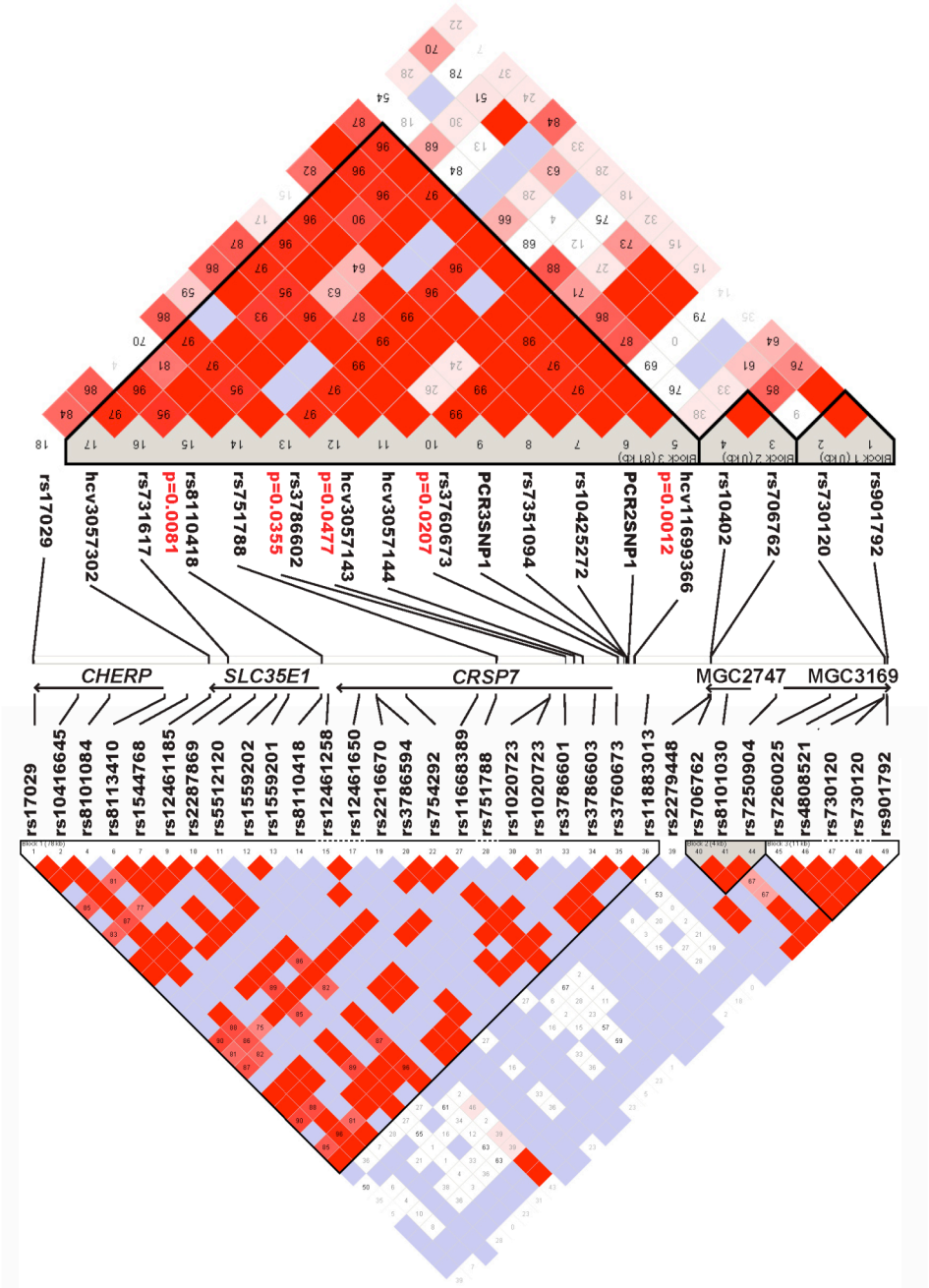


Figure 9-2. Schematic and analysis of the 19p13 region. Markers surrounding hCV11699366 are illustrated. Markers demonstrating significant association have accompanying p values. Relative intermarker spacing is also shown. LD blocks are numbered *right to left*. The black outline surrounds each of our self-defined haplotype blocks with intermarker D´ values labeled within the individual boxes. Intermarker linkage disequilibrium with D´ values of 1.0 are never shown, but indicated by the dark red boxes for a LOD score ≥ 2 and indicated by the blue boxes for a LOD score < 2. White shading indicates a D´ < 1 and a LOD < 2, while shades of pink/red indicate D´ < 1 and a LOD ≥ 2.

Given that hCV1169936 was located in the promoter of *CRSP7*, we evaluated the possibility of this marker correlating with one or more transcription factor binding sites. Sequence corresponding to either allele was tested using the web-based CONSITE program [201], and haplotypes for the significant SNP-containing block (block #3) were detailed using Haploview (Figure 9-3). While significant caveats accompany any such *in silico* prediction, we found that the over-transmitted allele [C] at hCV1169936, corresponds to a non-consensus residue in the consensus binding site sequence for a basic helix-loop-helix factor Spz1.

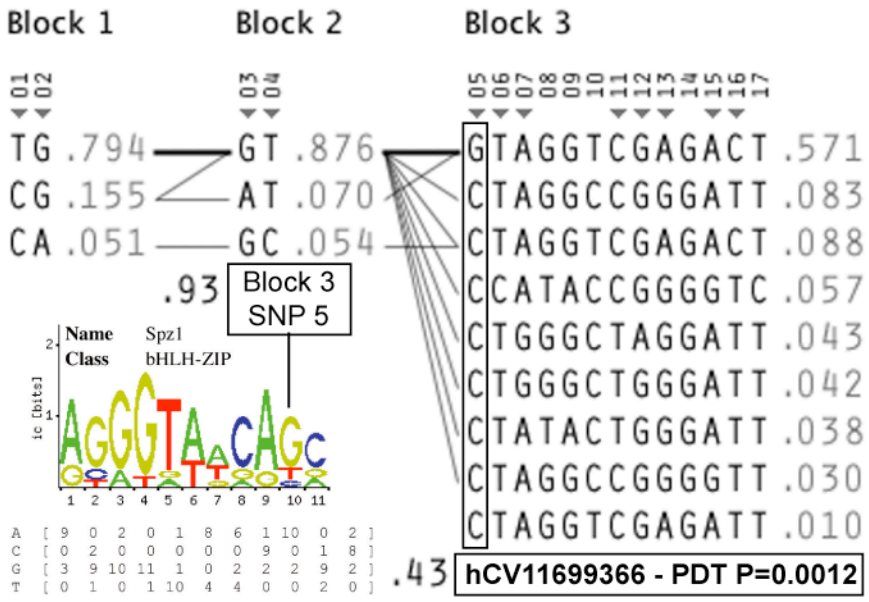


Figure 9-3. CONSITE output of the 19p13 region LD block #3.

Given the likelihood of identifying this number of markers *by chance* demonstrating significant association within a single block seemed highly

unlikely, we further tested the significance of these data. These and a few additional SNP markers were genotyped in an additional 226 (predominantly multiplex AGRE) families from the NIMH repository. Single marker PDT analysis of resulting genotype data revealed only two of the previously significant markers replicated a nominal significance, with hCV11699366 in the *CRSP7* promoter yielding a P = 0.05, and rs8110418 in *SLC35E1* also showing nominal TD (P = 0.02) in the total dataset (Table 9-7). Single marker association results and transmission counts for significant markers can be found in Table 9-7. Association analysis in the overall dataset, using FBAT, of all common haplotypes yielded no significant association (data not shown).

.

Table 9-7. PDT analysis of markers across the 19p13 region.

Marker No.	dbSNP rs#/ Celera hCV#	Overall PDT (P)	Allele	# of Alleles Transmitted	# of Alleles Not Transmitted
1		0.91			
2		0.69			
3		0.80			
4		0.21			
5		0.18			
6	rs8110418/ hCV3057319	**0.02**	A	540	566
			G	78	52
7		0.23			
8		0.20			
9		0.40			
10		1.00			
11		0.83			
12		0.22			
13		0.28			
14		0.16			
15		0.86			
16		0.24			
17	rs12461684/ hCV11699366	**0.05**	G	712	762
			C	592	542
18		0.51			
19		0.91			
20		0.78			
21		0.30			
22		0.61			

In a concurrent study, we addressed the alternative to the "*common-disease-common variant*" hypothesis by screening for heterogeneous disease-related alleles. We selected a panel of 24 affected and unrelated individuals from families demonstrating evidence for linkage to 19p13, as determined from ranking families for the highest multipoint HLOD score at the 40 cM location on chromosome 19p13. We sequenced all known *CRSP7* and *SLC35E1* exons for these 24 individuals. In addition we sequenced 5 overlapping regions 5´ of *CRSP7* exon one surrounding the putative promoter region and including marker hCV11699366 (our most significant marker in the chromosome 19 follow-up SNPs). While several variants were rediscovered, one (hCV3057306/rs2287869, a synonymous variant in exon 5 of *SLC35E1*) was observed to be heterozygous in 9 individuals and homozygous (for the rarer allele) in 6 individuals within our screening panel. This may at first glance seem significant; however, it is most likely non-significant given this marker has a dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) reported minor allele frequency of ~0.40 in the general population. Two other markers, 13 (pcr3SNP1) and 16 (pcr2SNP1) in Table 9-3 are novel. These novel markers were subsequently genotyped in the entire dataset and included in both the single marker and haplotype analysis, but yielded no significant association.

Discussion

We identified 19p13 as harboring a putative autism susceptibility locus, based on very suggestive linkage to autism detected in a genome-wide survey for autism risk loci. A follow-up analysis of these two loci to refine genetic linkage was very successful in narrowing linked intervals. In the case of chromosome 17q, there was a decrease in the observed LOD scores. However, intervals for both the primary peak underlying the *SLC6A4* locus and a more distal peak at 17q21 (~70 cM) were substantially narrowed when using a LOD score value of 1.0 as a cutoff. While the reduction in linkage is somewhat disturbing, other data from more recent analyses of linkage to this region suggest that the overall evidence for linkage of autism to this chromosome is very strong. The more distal peak could correspond to a distinct genetic effect at the integrin β3 (*ITGB3*) locus. Weiss and colleagues recently reported identification of this locus in a genome-wide QTL association screen for circulating 5-HT levels in a population isolate [199]. The same group subsequently reported a "sex-specific architecture" for allelic association at *ITGB3* and *SLC6A4* for 5-HT levels. That the narrowed peak in the current analysis and subsequent analyses remains coincident with *SLC6A4* also underscores this gene as an attractive candidate.

Follow-up SNP-based linkage analyses on 19p not only identified a narrowed, more significant, peak at 19p13 (~40 cM), but it also suggests a second locus at ~55 cM. While this pattern of linkage was present to some degree in the initial microsatellite-based genomic screen, addition of the SNP

genetic information makes this pattern more clear. The conclusion we can draw from these studies is that maximizing genetic information content in a region of interest is an important step in, not just *refining*, but in some cases actually *defining* focused intervals within which gene discovery should be pursued.

By continuing a biologically unbiased approach in search of autism susceptibility genes through SNP-based follow-up, we exploited the availability of data spanning the initially-linked interval to perform exploratory association analyses. With little *a priori* likelihood of identifying biologically-relevant association, given a sparse (relative to general expectations of linkage disequilibrium) ~1-cM grid of markers, we identified an intriguing pattern of association in adjacent markers, leading us to further explore the region. This strategy led to identification of a haplotype block containing three genes *CRSP7, SLC35E1*, and *CHERP*. While these genes co-localize with peak linkage (at ~40 cM), and thus represent positional candidates for an autism risk locus, these genes would not make *a priori* compelling candidate loci for autism. However, the principle underlying this biologically unbiased approach is to allow excess allele sharing within a disease population, to lead one to the relevant risk factors.

Though these genes may not initially make compelling biological candidates, it is important to point out their function, if known. *CRSP7* encodes a subunit of the CRSP co-factor that binds to the C-terminal domain of RNA polymerase II to mediate Sp1 transcriptional activation [202]. *SLC35E1* is a solute carrier of unknown function. *CHERP* encodes a protein important in maintaining Ca^{++} homeostasis in the endoplasmic reticulum, and is thought to be

important generally in intracellular Ca^{++} mobilization, and cellular growth and proliferation [203]. Without evidence of specific function for *SLC35E1*, *CHERP*, which is expressed in the brain based on UniGene expression profiling, may represent the most promising of the three genes for predicting functional relevance to autism.

Identification of a total of five SNPs within this block demonstrating evidence for association is very compelling but difficult to interpret. Concerns regarding multiple testing and Type I error may be reduced somewhat, because we were following up an initial observation and more carefully characterizing the association in that region. Regardless, given the number of tests performed within that block (13) and region (18), hCV11699366 (P = 0.0012) would remain significant based on Bonferroni correction. Considering the initial follow-up SNP association test, which involved 23 markers, association at hCV11699366 remains significant following Bonferroni correction for 23 tests. However, analysis of association within this block in a larger dataset containing 384 combined multiplex and trio families yielded reduced significance; only hCV11699366 (P = 0.05) and another previously associated marker (rs8110418; P = 0.02) from the original 158 families retain nominal evidence for association. It may be important to consider that similar follow-up linkage analysis in the 327 multiplex families (from the total 384-family dataset), was virtually identical in magnitude to that observed in the 158-family dataset. Therefore, the observation of reduced significance in tests of allelic association in the *CRSP7-SLC35E1-CHERP* haplotype block may not be at all surprising using this entire dataset.

Given the additional families examined for association may not be linked to chromosome 19, future studies will likely include the examination of the subset of linked families for association at these and other markers across the interval.

In an exploratory manner, the examination of marker hCV11699366 using CONSITE revealed plausible evidence this marker may alter a putative transcription factor binding site for Spz1. While these findings are initial and highly speculative, they demonstrate the usefulness and potential of such tools for examination of alleles at markers with unknown functionality.

Though we did not provide compelling evidence of a common allele at this locus to be involved in autism susceptibility, these studies are still in their infancy and provide the groundwork for future experiments. Studies to date also lack evidence for heterogeneous disease-risk alleles.

CHAPTER X

CONCLUSIONS AND FUTURE DIRECTIONS

Science is in a state of constant evolution, and research into complex disease is no exception. Advances have been made not only in the technology being used, but also in the basic way we define disease. Complex diseases are beginning to be examined as specific traits or phenotypes that characterize a particular disease or class of diseases. Scientists are testing whether these traits, which may or may not be quantitative, may hold the key to unraveling some of the complexity. Even more prominent is the idea of a complex disease having a genetic and/or environmental susceptibility landscape. This idea implies that once a given threshold, whether it be genetic or environmental or a combination of both, is exceeded a person presents with disease. Understanding this susceptibility landscape as well as potential environmental triggers will set the stage for preventative and therapeutic research.

The goals set forth in this body of work were to dissect both genetic and phenotypic susceptibility for a severe neurodevelopmental disorder, which is characterized by life-long deficits. Autism may be justifiably viewed as one of the more complex of the complex diseases. From the complex phenotypic spectrum, which includes other known genetic disorders (i.e. Fragile X syndrome, Rett syndrome, and tuberous sclerosis), to the male bias seen within the autism

spectrum of disorders, to numerous positive yet not replicated molecular and genetic findings, and finally to the high potential for gene-gene interactions, gene-environment interactions, and/or epigenetic and imprinting effects there is substantial evidence for complexity.

To examine the genetic susceptibility landscape we began our genetic studies by examining two of the more prominent regions of the genome and then expanded to a third region based on new lines of evidence. All of these regions arguably harbor potential candidate genes, which have previously been discussed. We characterized linkage disequilibrium patterns, haplotype block structures, and identified haplotype tag markers where relevant. In addition we tested for both single marker associations and multi-marker haplotypes associations in our overall dataset as well as different phenotypic subsets. Finally, in smaller samples we examined potential functional sequences at selected loci through use of molecular screening methods.

An overall conclusion is that there does not appear to be a *common* variant across any of the loci that is responsible for autism susceptibility in all cases or even a majority of cases. This conclusion is not a great surprise given the likelihood of both locus and allelic heterogeneity, but further details the complexity surrounding this disease. The lack of strong association to a *common* marker at any of the loci does not rule out their involvement in disease, but may merely suggest allelic heterogeneity at the locus or allelic interactions of common alleles with other loci. We do however detect instances of single marker and

haplotype associations in phenotypic subsets. These results provide preliminary findings for further study and illustrate the potential usefulness and desire to homogenize this diverse dataset. In the very near future, larger sample sizes will offset the power issues present with current subsets of the data. However, replication issues may still persist if larger samples sizes are used in the initial studies and cannot be found for replication studies.

Technology is also improving at a great rate. As genotyping throughput increases and costs decrease, genome-wide association studies of large autism samples will soon offer the potential to detect genes of small to moderate effect without requiring hypotheses about pathophysiology or chromosomal location. Currently such hypotheses, which can sometimes be quite elaborate, are necessary to narrow a large region of suggestive linkage to a manageable number of candidate loci.

Additional technological advances must be made to handle the large amount of data now being generated. In addition to the sheer volume of data to handle, resources and methods must be created to test for inter-locus interactions. With little to no replicable evidence with regard to genes or alleles of major effect within autism, there may be numerous allelic interactions that affect disease susceptibility. It is a very plausible hypothesis that genes in a common biological pathway may interact to yield disease; however, the complexity of such a pathway is likely greater than our current understanding. Moreover, strategies

to tease apart both locus and allelic heterogeneity must be developed and will likely play a key role in examination of complex genetic disease.

Another potential issue beyond the scope of this project involves epigenetic variations. Unknown and unexamined epigenetic differences may have a great affect on particular genotypes as they relate to disease risk. It may be clearer to see the effects of epigenetics when it comes to diseases that present more gradually in the adult. Epigenetic variation may be a necessary and specific target for environmental influence, upon which alteration can demonstrate adverse effects over time. In the case of developmental disorders such variation in developmentally differentiated genes may have more immediate irreversible outcomes. Though this is purely speculative, the reality of autism is most likely a combination of genetic, interactive, environmental, and epigenetic effects that we are only at the tip of understanding.

REFERENCES

1. Piven, J., The broad autism phenotype: a complementary strategy for molecular genetic studies of autism. *Am. J. Med. Genet.*, 2001. **105**: p. 34-5.

2. Association, A. P., *Diagnostic and statistical manual of mental disorders*. 4th ed. 1994, Washington, D.C.

3. Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., and Schopler, E., Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *Journal of Autism & Developmental Disorders.*, 1989. **19**: p. 185-212.

4. Lord, C., Rutter, M., and Le Couteur, A., Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.*, 1994. **24**: p. 659-85.

5. Gutierrez, G. C., Smalley, S. L., and Tanguay, P. E., Autism in tuberous sclerosis complex. *J. Autism Dev. Disord.*, 1998. **28**: p. 97-103.

6. Williams, C. A., Angelman, H., Clayton-Smith, J., Driscoll, D. J., Hendrickson, J. E., Knoll, J. H., Magenis, R. E., Schinzel, A., Wagstaff, J., Whidden, E. M., and et, a., Angelman syndrome: consensus for diagnostic criteria.Angelman Syndrome Foundation. *Am. J. Med. Genet.*, 1995. **56**: p. 237-8.

7. Zappella, M., Meloni, I., Longo, I., Canitano, R., Hayek, G., Rosaia, L., Mari, F., and Renieri, A., Study of MECP2 gene in Rett syndrome variants and autistic girls. *Am J Med Genet*, 2003. **119B**: p. 102-7.

8. Folstein, S. E. and Rosen-Sheidley, B., Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Reviews Genetics*, 2001. **2**: p. 943-55.

9. Veenstra-VanderWeele, J., Christian, S. L., and Cook, J. E. H., Autism as a Paradigmatic Complex Genetic Disorder. *Annual Review of Genomics and Human Genetics 2003 Jan 01;5(1):379*, 2004. **5**: p. 407-441.

10. Kanner, L., Autistic disturbance of affective contact. *Nervous Child*, 1943. **2**: p. 217-250.

11. Wolff, S., The history of autism. *Eur Child Adolesc Psychiatry*, 2004. **13**: p. 201-8.

12. Rutter, M., Concepts of autism: a review of research. *J Child Psychol Psychiatry*, 1968. **9**: p. 1-25.

13. Tidmarsh, L. and Volkmar, F. R., Diagnosis and epidemiology of autism spectrum disorders. *Can J Psychiatry*, 2003. **48**: p. 517-25.

14. Chakrabarti, S. and Fombonne, E., Pervasive developmental disorders in preschool children. *JAMA*, 2001. **285**: p. 3093-9.

15. Fombonne, E., The prevalence of autism. *Jama*, 2003. **289**: p. 87-9.

16. Veenstra-VanderWeele, J. and Cook, E. H., Molecular genetics of autism spectrum disorder. *Mol Psychiatry*, 2004.

17. Fombonne, E., The epidemiology of autism: a review. *Psychol. Med.*, 1999. **29**: p. 769-786.

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

19. Rutter, M., Macdonald, H., Le Couteur, A., Harrington, R., Bolton, P., and Bailey, A., Genetic factors in child psychiatric disorders--II. Empirical findings. *J. Child. Psychol. Psychiatry*, 1990. **31**: p. 39-83.

20. Pickles, A., Bolton, P., Macdonald, H., Bailey, A., Le Couteur, A., Sim, C. H., and Rutter, M., Latent-class analysis of recurrence risks for complex

phenotypes with selection and measurement error: a twin and family history study of autism. *Am. J. Hum. Genet.*, 1995. **57**: p. 717-26.

21. 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., Petersen, P. B., Pingree, C., McMahon, W., Wong, D. L., Cavalli-Sforza, L. L., Kraemer, H. C., and Myers, R. M., A genomic screen of autism: Evidence for a multilocus etiology. *Am. J. Hum. Genet.*, 1999. **65**: p. 493-507.

22. Jiang, Y.-h., Sahoo, T., Michaelis, R. C., Bercovich, D., Bressler, J., Kashork, C. D., Liu, Q., Shaffer, L. G., Schroer, R. J., Stockton, D. W., Spielman, R. S., Stevenson, R. E., and Beaudet, A. L., A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. *Am. J. Med. Genet.*, 2004. **In Press**.

23. Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E., and Walker-Smith, J. A., Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children [see comments]. *Lancet*, 1998. **351**: p. 637-41.

24. Madsen, K. M., Hviid, A., Vestergaard, M., Schendel, D., Wohlfahrt, J., Thorsen, P., Olsen, J., and Melbye, M., A population-based study of measles, mumps, and rubella vaccination and autism. *N Engl J Med*, 2002. **347**: p. 1477-82.

25. IMGSAC, 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.*, 1998. **7**: p. 571-8.

26. Barrett, S., Beck, J. C., Bernier, R., Bisson, E., Braun, T. A., Casavant, T. L., Childress, D., Folstein, S. E., Garcia, M., Gardiner, M. B., Gilman, S., Haines, J. L., Hopkins, K., Landa, R., Meyer, N. H., Mullane, J. A., Nishimura, D. Y., Palmer, P., Piven, J., Purdy, J., Santangelo, S. L., Searby, C., Sheffield, V., Singleton, J., Slager, S., and et al., An autosomal genomic screen for autism. Collaborative linkage study of autism. *Am. J. Med. Genet.*, 1999. **88**: p. 609-15.

27. Philippe, A., Martinez, M., Guilloud-Bataille, M., Gillberg, C., Rastam, M., Sponheim, E., Coleman, M., Zappella, M., Aschauer, H., van Malldergerme, L., Penet, C., Feingold, J., Brice, A., and Leboyer, M., Genome-wide scan for autism susceptibility genes. *Hum. Mol. Genet.*, 1999. **8**: p. 805-812.

28. Auranen, M., Nieminen, T., Majuri, S., Vanhala, R., Peltonen, L., and Jarvela, I., Analysis of autism susceptibility gene loci on chromosomes 1p, 4p, 6q, 7q, 13q, 15q, 16p, 17q, 19q and 22q in Finnish multiplex families. *Mol. Psychiatry*, 2000. **5**: p. 320-2.

29. Buxbaum, J. D., Silverman, J. M., Smith, C. J., Kilifarski, M., Reichert, J., Hollander, E., Lawlor, B. A., Fitzgerald, M., Greenberg, D. A., and Davis, K. L., Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am. J. Hum. Genet.*, 2001. **68**: p. 1514-20.

30. Liu, J., Nyholt, D. R., Magnussen, P., Parano, E., Pavone, P., Geschwind, D., Lord, C., Iversen, P., Hoh, J., Ott, J., Gilliam, T. C., and The Autism Genetic Resource Exchange, C., A genomewide screen for autism susceptibility loci. *Am. J. Hum. Genet.*, 2001. **69**: p. 327-40.

31. Shao, Y., Wolpert, C. M., Raiford, K. L., Menold, M. M., Donnelly, S. L., Ravan, S. A., Bass, M. P., McClain, C., von Wendt, L., Vance, J. M., Abramson, R. H., Wright, H. H., Ashley-Koch, A., Gilbert, J. R., DeLong, R. G., Cuccaro, M. L., Pericak-Vance, M. A., and McCoy, P. A., Genomic screen and follow-up analysis for autistic disorder. *Am. J. Med. Genet.*, 2002. **114**: p. 99-105.

32. Yonan, A. L., Alarcon, M., Cheng, R., Magnusson, P. K., Spence, S. J., Palmer, A. A., Grunn, A., Hank Juo, S. H., Terwilliger, J. D., Liu, J., Cantor, R. M., Geschwind, D. H., and Gilliam, T. C., A genomewide screen of 345 families for autism-susceptibility Loci. *Am J Hum Genet*, 2003. **73**: p. 886-97.

33. McCauley, J. L., Li, C., Jiang, L., Olson, L. M., Crockett, G., Gainer, K., Folstein, S. E., Haines, J. L., and Sutcliffe, J. S., 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 Med Genet*, 2005. **6**: p. 1.

34. Schain, R. J. and Freedman, D., Studies on 5-hydroxyindole metabolism in autistic disorder and other mentally retarded children. *Journal of Pediatrics*, 1961. **58**: p. 315-320.

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

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

37. Gordon, C. T., State, R. C., Nelson, J. E., Hamburger, S. D., and Rapoport, J. L., A double-blind comparison of clomipramine, desipramine, and placebo in the treatment of autistic disorder. *Archives of General Psychiatry*, 1993. **50**: p. 441-7.

38. McDougle, C. J., Naylor, S. T., Cohen, D. J., Volkmar, F. R., Heninger, G. R., and Price, L. H., A double-blind, placebo-controlled study of fluvoxamine in adults with autistic disorder [see comments]. *Archives of General Psychiatry*, 1996. **53**: p. 1001-8.

39. Blatt, G. J., Fitzgerald, C. M., Guptill, J. T., Booker, A. B., Kemper, T. L., and Bauman, M. L., Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study. *J. Autism Dev. Disord.*, 2001. **31**: p. 537-43.

40. Moreno-Fuenmayor, H., Borjas, L., Arrieta, A., Valera, V., and Socorro-Candanoza, L., Plasma excitatory amino acids in autism. *Investigacion Clinica*, 1996. **37**: p. 113-28.

41. Dhossche, D., Applegate, H., Abraham, A., Maertens, P., Bland, L., Bencsath, A., and Martinez, J., Elevated plasma gamma-aminobutyric acid (GABA) levels in autistic youngsters: stimulus for a GABA hypothesis of autism. *Med Sci Monit*, 2002. **8**: p. PR1-6.

42. Aldred, S., Moore, K. M., Fitzgerald, M., and Waring, R. H., Plasma amino acid levels in children with autism and their families. *J Autism Dev Disord*, 2003. **33**: p. 93-7.

43. Olsen, R. W. and Macdonald, R. L., *GABAA receptor complex: Structure and function*, in *Glutamate and GABA receptors and Transporters*, J. Egebjerg, A. Schousboe, and P. Krogsgaard-Larsen, Editors. 2002, Taylor & Francis: London. p. 202-235.

44. Cook, E. H., Courchesne, R. Y., Cox, N. J., Lord, C., Gonen, D., Guter, S. J., Lincoln, A., Nix, K., Haas, R., Leventhal, B. L., and Courchesne, E., Linkage-Disequilibrium Mapping of Autistic Disorder, With 15q11-13 Markers. *Am. J. Hum. Genet.*, 1998. **62**: p. 1077-1083.

45. Martin, E. R., Menold, M. M., Wolpert, C. M., Bass, M. P., Donnelly, S. L., Ravan, S. A., Zimmerman, A., Gilbert, J. R., Vance, J. M., Maddox, L. O., Wright, H. H., Abramson, R. K., DeLong, G. R., Cuccaro, M. L., and Pericak-Vance, M. A., Analysis of linkage disequilibrium in gamma-aminobutyric acid receptor subunit genes in autistic disorder. *Am. J. Med. Genet.*, 2000. **96**: p. 43-8.

46. Buxbaum, J. D., Silverman, J. M., Smith, C. J., Greenberg, D. A., Kilifarski, M., Reichert, J., Cook, E. H., Fang, Y., Song, C. Y., and Vitale, R., Association between a GABRB3 polymorphism and autism. *Mol. Psychiatry*, 2002. **7**: p. 311-6.

47. Nurmi, E. L., Amin, T., Olson, L. M., Jacobs, M. M., McCauley, J. L., Lam, A. Y., Organ, E. L., Folstein, S. E., Haines, J. L., and Sutcliffe, J. S., Dense linkage disequilibirum mapping in the 15q11-q13 maternal expression domain yields evidence for association in autism. *Mol. Psychiatry*, 2003. **in press**.

48. International Molecular Genetic Study of Autism, C., A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am. J. Hum. Genet.*, 2001. **69**: p. 570-81.

49. Cook, E. H., Jr., Lindgren, V., Leventhal, B. L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C., and Courchesne, E., Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.*, 1997. **60**: p. 928-34.

50. Schroer, R. J., Phelan, M. C., Michaelis, R. C., Crawford, E. C., Skinner, S. A., Cuccaro, M., Simensen, R. J., Bishop, J., Skinner, C., Fender, D., and Stevenson, R. E., Autism and Maternally Derived Aberrations of Chromosome 15q. *Am. J. Med. Genet.*, 1998. **76**: p. 327-336.

51. Maestrini, E., Lai, C., Marlow, A., Matthews, N., Wallace, S., Bailey, A., Cook, E. H., Weeks, D. E., and Monaco, A. P., Serotonin transporter (5-HTT) and gamma-aminobutyric acid receptor subunit beta3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSA families. The International Molecular Genetic Study of Autism Consortium. *Am. J. Med. Genet.*, 1999. **88**: p. 492-6.

52. Salmon, B., Hallmayer, J., Rogers, T., Kalaydjieva, L., Petersen, P. B., Nicholas, P., Pingree, C., McMahon, W., Spiker, D., Lotspeich, L., Kraemer, H., McCague, P., Dimiceli, S., Nouri, N., Pitts, T., Yang, J., Hinds, D., Myers, R. M., and Risch, N., Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism. *Am. J. Med. Genet.*, 1999. **88**: p. 551-6.

53. Nurmi, E. L., Dowd, M., Tadevosyan-Leyfer, O., Haines, J. L., Folstein, S. E., and Sutcliffe, J. S., Exploratory Subsetting of Autism Families Based on Savant Skills Improves Evidence of Genetic Linkage to 15q11-q13. *J Am Acad Child Adolesc Psychiatry*, 2003. **42**: p. 856-863.

54. Shao, Y., Cuccaro, M. L., Hauser, E. R., Raiford, K. L., Menold, M. M., Wolpert, C. M., Ravan, S. A., Elston, L., Decena, K., Donnelly, S. L., Abramson, R. K., Wright, H. H., DeLong, G. R., Gilbert, J. R., and Pericak-Vance, M. A., Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. *Am J Hum Genet*, 2003. **72**: p. 539-48.

55. Morton, N. E., LODs past and present. *Genetics*, 1995. **140**: p. 7-12.

56. Morton, N. E. and Collins, A., Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A*, 1998. **95**: p. 11389-93.

57. Ewens, W. J. and Spielman, R. S., The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet*, 1995. **57**: p. 455-64.

58. Khoury, M. J. and Beaty, T. H., Applications of the case-control method in genetic epidemiology. *Epidemiol Rev*, 1994. **16**: p. 134-50.

59. Lander, E. S. and Schork, N. J., Genetic dissection of complex traits. *Science*, 1994. **265**: p. 2037-48.

60. Weeks, D. E. and Lathrop, G. M., Polygenic disease: methods for mapping complex disease traits. *Trends Genet*, 1995. **11**: p. 513-9.

61. Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S., and Schmidtke, J., An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet*, 1985. **69**: p. 201-5.

62. Li, W. H. and Sadler, L. A., Low nucleotide diversity in man. *Genetics*, 1991. **129**: p. 513-23.

63. Kruglyak, L., The use of a genetic map of biallelic markers in linkage studies. *Nat Genet*, 1997. **17**: p. 21-4.

64. Landegren, U., Nilsson, M., and Kwok, P. Y., Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res*, 1998. **8**: p. 769-76.

65. Wall, J. D. and Pritchard, J. K., Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet*, 2003. **4**: p. 587-97.

66. Fullerton, S. M., Buchanan, A. V., Sonpar, V. A., Taylor, S. L., Smith, J. D., Carlson, C. S., Salomaa, V., Stengard, J. H., Boerwinkle, E., Clark, A. G., Nickerson, D. A., and Weiss, K. M., The effects of scale: variation in the APOA1/C3/A4/A5 gene cluster. *Hum Genet*, 2004. **115**: p. 36-56.

67. Hirschhorn, J. N. and Daly, M. J., Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 2005. **6**: p. 95-108.

68. Templeton, A. R., *Epistasis and complex traits*, in *Epistasis and the evolutionary Process*, J. Wolf, E.D. Brodie III, and M.J. Wade, Editors. 2000, Oxford University Press: New York.

69. Moore, J. H., The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered*, 2003. **56**: p. 73-82.

70. Hirschhorn, J. N., Lohmueller, K., Byrne, E., and Hirschhorn, K., A comprehensive review of genetic association studies. *Genet Med*, 2002. **4**: p. 45-61.

71. Moore, J. H. and Williams, S. M., New strategies for identifying gene-gene interactions in hypertension. *Ann Med*, 2002. **34**: p. 88-95.

72. Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., and Moore, J. H., Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*, 2001. **69**: p. 138-47.

73. Ritchie, M. D., Hahn, L. W., and Moore, J. H., Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol*, 2003. **24**: p. 150-7.

74. Hahn, L. W., Ritchie, M. D., and Moore, J. H., Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*, 2003. **19**: p. 376-82.

75. Gottesman, II and Gould, T. D., The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*, 2003. **160**: p. 636-45.

76. Tadevosyan-Leyfer, O., Dowd, M., Mankoski, R., Winklosky, B., Putnam, S., McGrath, L., Tager-Flusberg, H., and Folstein, S. E., A Principal Components Analysis of the Autism Diagnostic Interview-Revised. *J Am Acad Child Adolesc Psychiatry*, 2003. **42**: p. 864-872.

77. Hauser, E. R., Bass, M. P., Martin, E. R., Watanabe, W. L., Duren, W., and Boehnke, M., Power of the ordered subset method for detection and localization of genes in linkage analysis of complex traits. *Am J Hum Genet Suppl*, 2001. **69**: p. A 2038.

78. Weiss, L. A., Abney, M., Cook, E. H., and Ober, C., Sex-specific genetic architecture of whole blood serotonin levels. *Am J Hum Genet*, 2005. **76**: p. 33-41.

79. Stone, J. L., Merriman, B., Cantor, R. M., Yonan, A. L., Gilliam, T. C., Geschwind, D. H., and Nelson, S. F., Evidence for sex-specific risk alleles in autism spectrum disorder. *Am J Hum Genet*, 2004. **75**: p. 1117-23.

80. Insel, T. R., Mouse models for autism: report from a meeting. *Mamm Genome*, 2001. **12**: p. 755-7.

81. Thornton-Wells, T. A., Moore, J. H., and Haines, J. L., Genetics, statistics and human disease: analytical retooling for complexity. *Trends Genet*, 2004. **20**: p. 640-7.

82. Hanley, H. G., Stahl, S. M., and Freedman, D. X., Hyperserotonemia and amine metabolites in autistic and retarded children. *Arch Gen Psychiatry*, 1977. **34**: p. 521-31.

83. Fatemi, S. H., Realmuto, G. M., Khan, L., and Thuras, P., Fluoxetine in treatment of adolescent patients with autism: a longitudinal open trial. *J Autism Dev Disord*, 1998. **28**: p. 303-7.

84. Hollander, E., Phillips, A., King, B. H., Guthrie, D., Aman, M. G., Law, P., Owley, T., and Robinson, R., Impact of recent findings on study design of future autism clinical trials. *CNS Spectr*, 2004. **9**: p. 49-56.

85. Hollander, E., Pallanti, S., Allen, A., Sood, E., and Rossi, N. B., Does sustained-release lithium reduce impulsive gambling and affective instability versus placebo in pathological gamblers with bipolar spectrum disorders? *Am J Psychiatry*, 2005. **162**: p. 137-45.

86. McDougle, C. J., Naylor, S. T., Cohen, D. J., Aghajanian, G. K., Heninger, G. R., and Price, L. H., Effects of tryptophan depletion in drug-free adults with autistic disorder [see comments]. *Archives of General Psychiatry*, 1996. **53**: p. 993-1000.

87. Hoshino, Y., Watanabe, H., Yashima, Y., Kaneko, M., and Kumashiro, H., An investigation on sleep disturbance of autistic children. *Folia Psychiatr Neurol Jpn*, 1984. **38**: p. 45-51.

88. Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., and Poulton, R., Role of genotype in the cycle of violence in maltreated children. *Science*, 2002. **297**: p. 851-4.

89. Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., and Poulton, R., Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, 2003. **301**: p. 386-9.

90. Ozaki, N., Goldman, D., Kaye, W. H., Plotnicov, K., Greenberg, B. D., Lappalainen, J., Rudnick, G., and Murphy, D. L., Serotonin transporter missense mutation associated with a complex neuropsychiatric phenotype. *Mol Psychiatry*, 2003. **8**: p. 895, 933-6.

91. Rothe, C., Gutknecht, L., Freitag, C., Tauber, R., Mossner, R., Franke, P., Fritze, J., Wagner, G., Peikert, G., Wenda, B., Sand, P., Jacob, C., Rietschel, M., Nothen, M. M., Garritsen, H., Fimmers, R., Deckert, J., and Lesch, K. P., Association of a functional 1019C>G 5-HT1A receptor gene polymorphism with panic disorder with agoraphobia. *Int J Neuropsychopharmacol*, 2004. **7**: p. 189-92.

92. Rothe, C., Koszycki, D., Bradwejn, J., King, N., De Luca, V., Shaikh, S., Franke, P., Garritsen, H., Fritze, J., Deckert, J., and Kennedy, J. L., Association study of serotonin-2A receptor gene polymorphism and panic disorder in patients from Canada and Germany. *Neurosci Lett*, 2004. **363**: p. 276-9.

93. Serretti, A., Artioli, P., and De Ronchi, D., The 5-HT2C receptor as a target for mood disorders. *Expert Opin Ther Targets*, 2004. **8**: p. 15-23.

94. Serretti, A., Artioli, P., Lorenzi, C., Pirovano, A., Tubazio, V., and Zanardi, R., The C(-1019)G polymorphism of the 5-HT1A gene promoter and antidepressant response in mood disorders: preliminary findings. *Int J Neuropsychopharmacol*, 2004: p. 1-8.

95. Chotai, J., Serretti, A., and Lorenzi, C., Interaction between the Tryptophan Hydroxylase Gene and the Serotonin Transporter Gene in Schizophrenia but Not in Bipolar or Unipolar Affective Disorders. *Neuropsychobiology*, 2004. **51**: p. 3-9.

96. Zhang, X., Gainetdinov, R. R., Beaulieu, J. M., Sotnikova, T. D., Burch, L. H., Williams, R. B., Schwartz, D. A., Krishnan, K. R., and Caron, M. G., Loss-of-Function Mutation in Tryptophan Hydroxylase-2 Identified in Unipolar Major Depression. *Neuron*, 2005. **45**: p. 11-6.

97. Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., and Lesch, K. P., Allelic variation of human serotonin transporter gene expression. *J Neurochem*, 1996. **66**: p. 2621-4.

98. Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., Benjamin, J., Muller, C. R., Hamer, D. H., and Murphy, D. L., Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, 1996. **274**: p. 1527-31.

99. Fiskerstrand, C. E., Lovejoy, E. A., and Quinn, J. P., An intronic polymorphic domain often associated with susceptibility to affective disorders has allele dependent differential enhancer activity in embryonic stem cells. *FEBS Lett*, 1999. **458**: p. 171-4.

100. MacKenzie, A. and Quinn, J., A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. *Proc Natl Acad Sci U S A*, 1999. **96**: p. 15251-5.

101. Cook, E. H., Jr., Courchesne, R., Lord, C., Cox, N. J., Yan, S., Lincoln, A., Haas, R., Courchesne, E., and Leventhal, B. L., Evidence of linkage between the serotonin transporter and autistic disorder. *Mol. Psychiatry*, 1997. **2**: p. 247-50.

102. Klauck, S. M., Poustka, F., Benner, A., Lesch, K. P., and Poustka, A., Serotonin Transporter (5-Htt) Gene Variants Associated With Autism. *Hum. Mol. Genet.*, 1997. **6**: p. 2233-2238.

103. Tordjman, S., Gutknecht, L., Carlier, M., Spitz, E., Antoine, C., Slama, F., Carsalade, V., Cohen, D. J., Ferrari, P., Roubertoux, P. L., and Anderson, G. M., Role of the serotonin transporter gene in the behavioral expression of autism. *Mol. Psychiatry*, 2001. **6**: p. 434-9.

104. Yirmiya, N., Pilowsky, T., Nemanov, L., Arbelle, S., Feinsilver, T., Fried, I., and Ebstein, R. P., Evidence for an association with the serotonin transporter promoter region polymorphism and autism. *Am. J. Med. Genet.*, 2001. **105**: p. 381-6.

105. Persico, A. M., Militerni, R., Bravaccio, C., Schneider, C., Melmed, R., Conciatori, M., Damiani, V., Baldi, A., and Keller, F., Lack of association between serotonin transporter gene promoter variants and autistic disorder in two ethnically distinct samples. *Am. J. Med. Genet.*, 2000. **96**: p. 123-7.

106. Kim, S. J., Cox, N., Courchesne, R., Lord, C., Corsello, C., Akshoomoff, N., Guter, S., Leventhal, B. L., Courchesne, E., and Cook Jr, E. H., Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder. *Mol. Psychiatry*, 2002. **7**: p. 278-88.

107. Conroy, J., Meally, E., Kearney, G., Fitzgerald, M., Gill, M., and Gallagher, L., Serotonin transporter gene and autism: a haplotype analysis in an Irish autistic population. *Mol Psychiatry*, 2004.

108. Mulder, E. J., Anderson, G. M., Kema, I. P., Brugman, A. M., Ketelaars, C. E., de Bildt, A., van Lang, N. D., den Boer, J. A., and Minderaa, R. B., Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder. *Am J Med Genet B Neuropsychiatr Genet*, 2005. **133B**: p. 93-6.

109. CLSA, Barrett, S., Beck, J. C., Bernier, R., Bisson, E., Braun, T. A., Casavant, T. L., Childress, D., Folstein, S. E., Garcia, M., Gardiner, M. B., Gilman, S., Haines, J. L., Hopkins, K., Landa, R., Meyer, N. H., Mullane, J. A., Nishimura, D. Y., Palmer, P., Piven, J., Purdy, J., Santangelo, S. L., Searby, C., Sheffield, V., Singleton, J., Slager, S., and et al., An autosomal genomic screen for autism. Collaborative linkage study of autism. *Am. J. Med. Genet.*, 1999. **88**: p. 609-15.

110. Shao, Y., Raiford, K. L., Wolpert, C. M., Cope, H. A., Ravan, S. A., Ashley-Koch, A. A., Abramson, R. K., Wright, H. H., DeLong, R. G., Gilbert, J. R., Cuccaro, M. L., and Pericak-Vance, M. A., Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am. J. Hum. Genet.*, 2002. **70**: p. 1058-61.

111. Alarcon, M., Cantor, R. M., Liu, J., Gilliam, T. C., Geschwind, D. H., and Consortium, A. G. R. E., Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am. J. Hum. Genet.*, 2002. **70**: p. 60-71.

112. Hsu, T. M., Chen, X., Duan, S., Miller, R. D., and Kwok, P. Y., Universal SNP genotyping assay with fluorescence polarization detection. *Biotechniques*, 2001. **31**: p. 560, 562, 564-8, passim.

113. Holloway, J. W., Beghe, B., Turner, S., Hinks, L. J., Day, I. N., and Howell, W. M., Comparison of three methods for single nucleotide polymorphism typing for DNA bank studies: sequence-specific oligonucleotide probe hybridisation, TaqMan liquid phase hybridisation, and microplate array diagonal gel electrophoresis (MADGE). *Hum Mutat*, 1999. **14**: p. 340-7.

114. Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., Shlien, A., Palsson, S. T., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., and Stefansson, K., A high-resolution recombination map of the human genome. *Nat Genet*, 2002. **31**: p. 241-7.

115. Kong, A. and Cox, N. J., Allele-sharing models: LOD scores and accurate linkage tests. *Am. J. Hum. Genet.*, 1997. **61**: p. 1179-88.

116. Clerget-Darpoux, F., Bonaiti-Pellie, C., and Hochez, J., Effects of misspecifying genetic parameters in lod score analysis. *Biometrics*, 1986. **42**: p. 393-9.

117. Sobel, E. and Lange, K., Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet*, 1996. **58**: p. 1323-37.

118. Abecasis, G. R. and Cookson, W. O., GOLD--graphical overview of linkage disequilibrium. *Bioinformatics*, 2000. **16**: p. 182-3.

119. Martin, E. R., Monks, S. A., Warren, L. L., and Kaplan, N. L., A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.*, 2000. **67**: p. 146-54.

120. Clayton, D., A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am. J. Hum. Genet.*, 1999. **65**: p. 1170-7.

121. Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., Pachter, L. S., and Dubchak, I., VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, 2000. **16**: p. 1046-7.

122. Fearnhead, N. S., Wilding, J. L., Winney, B., Tonks, S., Bartlett, S., Bicknell, D. C., Tomlinson, I. P., Mortensen, N. J., and Bodmer, W. F., Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci U S A*, 2004. **101**: p. 15992-7.

123. Glatt, C. E., DeYoung, J. A., Delgado, S., Service, S. K., Giacomini, K. M., Edwards, R. H., Risch, N., and Freimer, N. B., Screening a large reference sample to identify very low frequency sequence variants: comparisons between two genes. *Nat Genet*, 2001. **27**: p. 435-8.

124. Ashcroft, G. W., Crawford, T. B., Binns, J. K., and Macdougall, E. J., Estimation Of 5-Hydroxytryptamine In Human Blood. *Clin Chim Acta*, 1964. **45**: p. 364-9.

125. Wirz-Justice, A., Lichtsteiner, M., and Feer, H., Diurnal and seasonal variations in human platelet serotonin in man. *J Neural Transm*, 1977. **41**: p. 7-15.

126. Gonzales, G. F., Blood levels of 5-Hydroxytryptamine in human beings under several physiological situations. *Life Sci*, 1980. **27**: p. 647-50.

127. Ortiz, J., Artigas, F., and Gelpi, E., Serotonergic status in human blood. *Life Sci*, 1988. **43**: p. 983-90.

128. McCauley, J. L., Olson, l. M., Delahanty, R., Amin, T., Nurmi, E. L., Organ, E. L., Jacobs, M. M., Folstein, S. E., Haines, J. L., and Sutcliffe, J. S., A linkage disequilibrium map of the 1-Mb 15q12 GABAA receptor subunnit cluster and association to autism. *Am J Med Genet*, 2004. **In Press.**

129. Barrett, J. C., Fry, B., Maller, J., and Daly, M. J., Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 2005. **21**: p. 263-5.

130. Horvath, S., Xu, X., and Laird, N. M., The family based association test method: strategies for studying general genotype--phenotype associations. *Eur J Hum Genet*, 2001. **9**: p. 301-6.

131. Chugani, D. C., Muzik, O., Juhasz, C., Janisse, J. J., Ager, J., and Chugani, H. T., Postnatal maturation of human GABAA receptors measured with positron emission tomography. *Annals of Neurology*, 2001. **49**: p. 618-26.

132. Clayton-Smith, J., Webb, T., Cheng, X. J., Pembrey, M. E., and Malcolm, S., Duplication of chromosome 15 in the region 15q11-13 in a patient with developmental delay and ataxia with similarities to Angelman syndrome. *J. Med. Genet.*, 1993. **30**: p. 529-31.

133. Browne, C. E., Dennis, N. R., Maher, E., Long, F. L., Nicholson, J. C., Sillibourne, J., and Barber, J. C. K., Inherited Interstitial Duplications Of Proximal 15q - Genotype-Phenotype Correlations. *Am. J. Hum. Genet.*, 1997. **61**: p. 1342-1352.

134. Repetto, G. M., White, L. M., Bader, P. J., Johnson, D., and Knoll, J. H., Interstitial duplications of chromosome region 15q11q13: clinical and molecular characterization. *Am. J. Med. Genet.*, 1998. **79**: p. 82-9.

135. Mohandas, T. K., Park, J. P., Spellman, R. A., Filiano, J. J., Mamourian, A. C., Hawk, A. B., Belloni, D. R., Noll, W. W., and Moeschler, J. B., Paternally derived de novo interstitial duplication of proximal 15q in a

patient with developmental delay. *Am. J. Med. Genet.*, 1999. **82**: p. 294-300.

136. Roberts, S. E., Dennis, N. R., Browne, C. E., Willatt, L., Woods, G., Cross, I., Jacobs, P. A., and Thomas, S., Characterisation of interstitial duplications and triplications of chromosome 15q11-q13. *Human Genetics.*, 2002. **110**: p. 227-34.

137. Robinson, W. P., Binkert, F., Gine, R., Vazquez, C., Muller, W., Rosenkranz, W., and Schinzel, A., Clinical and molecular analysis of five inv dup(15) patients. *Eur. J. Hum. Genet.*, 1993. **1**: p. 37-50.

138. Flejter, W. L., Bennett-Baker, P. E., Ghaziuddin, M., McDonald, M., Sheldon, S., and Gorski, J. L., Cytogenetic and molecular analysis of inv dup(15) chromosomes observed in two patients with autistic disorder and mental retardation. *Am. J. Med. Genet.*, 1996. **61**: p. 182-7.

139. Bundey, S., Hardy, C., Vickers, S., Kilpatrick, M. W., and Corbett, J. A., Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. *Developmental Medicine & Child Neurology*, 1994. **36**: p. 736-42.

140. Cheng, S. D., Spinner, N. B., Zackai, E. H., and Knoll, J. H., Cytogenetic and molecular characterization of inverted duplicated chromosomes 15 from 11 patients. *Am. J. Hum. Genet.*, 1994. **55**: p. 753-9.

141. Webb, T., Hardy, C. A., King, M., Watkiss, E., Mitchell, C., and Cole, T., A Clinical, Cytogenetic and Molecular Study of Ten Probands With Supernumerary Inv Dup (15) Marker Chromosomes. *Clinical Genetics*, 1998. **53**: p. 34-43.

142. Jiang, Y. H., Tsai, T. F., Bressler, J., and Beaudet, A. L., Imprinting in Angelman and Prader-Willi Syndromes. *Curr. Opin. Genet. Dev.*, 1998. **8**: p. 334-342.

143. Nurmi, E. L., Bradford, Y., Chen, Y., Hall, J., Arnone, B., Gardiner, M. B., Hutcheson, H. B., Gilbert, J. R., Pericak-Vance, M. A., Copeland-Yates, S. A., Michaelis, R. C., Wassink, T. H., Santangelo, S. L., Sheffield, V. C., Piven, J., Folstein, S. E., Haines, J. L., and Sutcliffe, J. S., Linkage

disequilibrium at the Angelman syndrome gene UBE3A in autism families. *Genomics*, 2001. **77**: p. 105-13.

144. Menold, M. M., Shao, Y., Wolpert, C. M., Donnelly, S. L., Raiford, K. L., Martin, E. R., Ravan, S. A., Abramson, R. K., Wright, H. H., Delong, G. R., Cuccaro, M. L., Pericak-Vance, M. A., and Gilbert, J. R., Association analysis of chromosome 15 gabaa receptor subunit genes in autistic disorder. *J Neurogenet*, 2001. **15**: p. 245-59.

145. Glatt, K., Glatt, H., and Lalande, M., Structure and Organization Of Gabrb3 and Gabra5. *Genomics*, 1997. **41**: p. 63-69.

146. Kirkness, E. F. and Fraser, C. M., A strong promoter element is located between alternative exons of a gene encoding the human gamma-aminobutyric acid-type A receptor beta 3 subunit (GABRB3). *J. Biol. Chem.*, 1993. **268**: p. 4420-8.

147. Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., and Altshuler, D., The structure of haplotype blocks in the human genome. *Science*, 2002. **296**: p. 2225-9.

148. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., and Lander, E. S., High-resolution haplotype structure in the human genome. *Nat Genet*, 2001. **29**: p. 229-32.

149. Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganske, R., Lohmussaar, E., Zernant, J., Tonisson, N., Remm, M., Magi, R., Puurand, T., Vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D. R., Cardon, L. R., and Dunham, I., A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, 2002. **418**: p. 544-8.

150. Jeffreys, A. J., Kauppi, L., and Neumann, R., Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet*, 2001. **29**: p. 217-22.

151. Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., Nguyen, B. T., Norris, M. C., Sheehan, J. B., Shen, N., Stern, D., Stokowski, R. P., Thomas, D. J., Trulson, M. O., Vyas, K. R., Frazer, K. A., Fodor, S. P., and Cox, D. R., Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, 2001. **294**: p. 1719-23.

152. Clark, A. G., Nielsen, R., Signorovitch, J., Matise, T. C., Glanowski, S., Heil, J., Winn-Deen, E. S., Holden, A. L., and Lai, E., Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. *Am J Hum Genet*, 2003. **73**: p. 285-300.

153. Couzin, J., Genomics. New mapping project splits the community. *Science*, 2002. **296**: p. 1391-3.

154. Cardon, L. R. and Abecasis, G. R., Using haplotype blocks to map human complex trait loci. *Trends Genet*, 2003. **19**: p. 135-40.

155. Sutcliffe, J. S., Nurmi, E. L., and Lombroso, P. J., Genetics of childhood disorders: XLVII. Autism, part 6: duplication and inherited susceptibility of chromosome 15q11-q13 genes in autism. *Journal of the American Academy of Child & Adolescent Psychiatry.*, 2003. **42**: p. 253-6.

156. Martin, E. R., Kaplan, N. L., and Weir, B. S., Tests for linkage and association in nuclear families. *Am. J. Hum. Genet.*, 1997. **61**: p. 439-48.

157. Abecasis, G. R., Cardon, L. R., and Cookson, W. O., A general test of association for quantitative traits in nuclear families. *Am J Hum Genet*, 2000. **66**: p. 279-92.

158. Nicholls, R. D. and Knepper, J. L., Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annual Review of Genomics & Human Genetics*, 2001. **2**: p. 153-75.

159. Amos-Landgraf, J. M., Ji, Y. G., Gottlieb, W., Depinet, T., Wandstrat, A. E., Cassidy, S. B., Driscoll, D. J., Rogan, P. K., Schwartz, S., and Nicholls, R. D., Chromosome breakage in the Prader-Willi and Angelman syndromes

involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am. J. Hum. Genet.*, 1999. **65**: p. 370-386.

160. Christian, S. L., Fantes, J. A., Mewborn, S. K., Huang, B., and Ledbetter, D. H., Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). *Hum. Mol. Genet.*, 1999. **8**: p. 1025-1037.

161. Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D., and Horsthemke, B., Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat. Genet.*, 1995. **9**: p. 395-400.

162. Sutcliffe, J. S., Nakao, M., Christian, S., Orstavik, K. H., Tommerup, N., Ledbetter, D. H., and Beaudet, A. L., Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nat. Genet.*, 1994. **8**: p. 52-8.

163. Buiting, K., Barnicoat, A., Lich, C., Pembrey, M., Malcolm, S., and Horsthemke, B., Disruption of the bipartite imprinting center in a family with Angelman syndrome. *Am. J. Hum. Genet.*, 2001. **68**: p. 1290-4.

164. Sutcliffe, J. S., Han, M., Christian, S. L., and Ledbetter, D. H., Neuronally-Expressed Necdin Gene - an Imprinted Candidate Gene In Prader-Willi-Syndrome. *Lancet*, 1997. **350**: p. 1520-1521.

165. Ren, J., Lee, S., Pagliardini, S., Gerard, M., Stewart, C. L., Greer, J. J., and Wevrick, R., Absence of Ndn, encoding the Prader-Willi syndrome-deleted gene necdin, results in congenital deficiency of central respiratory drive in neonatal mice. *J Neurosci*, 2003. **23**: p. 1569-73.

166. Lee, S., Walker, C. L., Karten, B., Kuny, S. L., Tennese, A. A., O'Neill M, A., and Wevrick, R., Essential role for the Prader-Willi syndrome protein necdin in axonal outgrowth. *Hum Mol Genet*, 2005.

167. Kishino, T., Lalande, M., and Wagstaff, J., UBE3A/E6-AP mutations cause Angelman syndrome. *Nat. Genet.*, 1997. **15**: p. 70-3.

168. Matsuura, T., Sutcliffe, J. S., Fang, P., Galjaard, R. J., Jiang, Y. H., Benton, C. S., Rommens, J. M., and Beaudet, A. L., De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat. Genet.*, 1997. **15**: p. 74-7.

169. Halleck, M. S., Lawler, J. J., Blackshaw, S., Gao, L., Nagarajan, P., Hacker, C., Pyle, S., Newman, J. T., Nakanishi, Y., Ando, H., Weinstock, D., Williamson, P., and Schlegel, R. A., Differential expression of putative transbilayer amphipath transporters. *Physiological Genomics*, 1999. **1**: p. 139-50.

170. Cassidy, S. B., Dykens, E., and Williams, C. A., Prader-Willi and Angelman syndromes: sister imprinted disorders. *Am. J. Med. Genet.*, 2000. **97**: p. 136-46.

171. Veltman, M. W., Thompson, R. J., Roberts, S. E., Thomas, N. S., Whittington, J., and Bolton, P. F., Prader-Willi syndrome--a study comparing deletion and uniparental disomy cases with reference to autism spectrum disorders. *Eur Child Adolesc Psychiatry*, 2004. **13**: p. 42-50.

172. Samaco, R. C., Nagarajan, R. P., Braunschweig, D., and LaSalle, J. M., Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. *Hum Mol Genet*, 2004. **13**: p. 629-39.

173. Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, L., and Lalande, M., An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat. Genet.*, 1998. **19**: p. 15-6.

174. International Molecular Genetic Study of Autism, C., Further characterization of the autism susceptibility locus AUTS1 on chromosome 7q. *Hum. Mol. Genet.*, 2001. **10**: p. 973-82.

175. Buxbaum, J. D., Silverman, J., Keddache, M., Smith, C. J., Hollander, E., Ramoz, N., and Reichert, J. G., 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*, 2004. **9**: p. 144-50.

176. Ashley-Koch, A., Wolpert, C. M., Menold, M. M., Zaeem, L., Basu, S., Donnelly, S. L., Ravan, S. A., Powell, C. M., Qumsiyeh, M. B., Aylsworth, A. S., Vance, J. M., Gilbert, J. R., Wright, H. H., Abramson, R. K., DeLong, G. R., Cuccaro, M. L., and Pericak-Vance, M. A., Genetic studies of autistic disorder and chromosome 7. *Genomics*, 1999. **61**: p. 227-36.

177. Folstein, S. E. and Mankoski, R. E., Chromosome 7q: where autism meets language disorder? *Am. J. Hum. Genet.*, 2000. **67**: p. 278-81.

178. CLSA, Incorporating language phenotypes strengthens evidence of linkage to autism. *Am. J. Med. Genet.*, 2001. **105**: p. 539-47.

179. Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., Fombonne, E., Leboyer, M., and Minshew, N., Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J. Autism Dev. Disord.*, 1997. **27**: p. 501-17.

180. DiLavore, P. C., Lord, C., and Rutter, M., The pre-linguistic autism diagnostic observation schedule. *J. Autism Dev. Disord.*, 1995. **25**: p. 355-79.

181. Sparrow, S. S. and Cicchetti, D. V., Diagnostic uses of the Vineland Adaptive Behavior Scales. *J Pediatr Psychol*, 1985. **10**: p. 215-25.

182. Volkmar, F. R., Sparrow, S. S., Goudreau, D., Cicchetti, D. V., Paul, R., and Cohen, D. J., Social deficits in autism: an operational approach using the Vineland Adaptive Behavior Scales. *J Am Acad Child Adolesc Psychiatry*, 1987. **26**: p. 156-61.

183. Geschwind, D. H., Sowinski, J., Lord, C., Iversen, P., Shestack, J., Jones, P., Ducat, L., Spence, S. J., and The, A. S. C., The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.*, 2001. **69**: p. 463-6.

184. O'Connell, J. R. and Weeks, D. E., PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*, 1998. **63**: p. 259-66.

185. Weeks, D. E., Sobel, E., O'Connell, J. R., and Lange, K., Computer programs for multilocus haplotyping of general pedigrees. *Am J Hum Genet*, 1995. **56**: p. 1506-7.

186. Haynes, C. S., Speer, M. C., Peedin, M., Roses, A. D., Haines, J. L., Vance, J. M., and Pericak-Vance, M. A., A comprehensive data management system to facilitate efficient and rapid disease gene mapping. *Am. J. Hum. Genet.*, 1995. **57**: p. A193.

187. Gudbjartsson, D. F., Jonasson, K., Frigge, M. L., and Kong, A., Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.*, 2000. **25**: p. 12-3.

188. McPeek, M. S., Optimal allele-sharing statistics for genetic mapping using affected relatives. *Genet Epidemiol*, 1999. **16**: p. 225-49.

189. Cottingham, R. W., Jr., Idury, R. M., and Schaffer, A. A., Faster sequential genetic linkage computations. *Am. J. Hum. Genet.*, 1993. **53**: p. 252-63.

190. Sawcer, S., Jones, H. B., Judge, D., Visser, F., Compston, A., Goodfellow, P. N., and Clayton, D., Empirical genomewide significance levels established by whole genome simulations. *Genet Epidemiol*, 1997. **14**: p. 223-9.

191. Kruglyak, L. and Daly, M. J., Linkage thresholds for two-stage genome scans. *Am J Hum Genet*, 1998. **62**: p. 994-7.

192. Abecasis, G. R., Cherny, S. S., Cookson, W. O., and Cardon, L. R., Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*, 2002. **30**: p. 97-101.

193. Hauser, E. R., Watanabe, R. M., Duren, W. L., Bass, M. P., Langefeld, C. D., and Boehnke, M., Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol*, 2004. **27**: p. 53-63.

194. Scott, W. K., Hauser, E. R., Schmechel, D. E., Welsh-Bohmer, K. A., Small, G. W., Roses, A. D., Saunders, A. M., Gilbert, J. R., Vance, J. M., Haines, J. L., and Pericak-Vance, M. A., Ordered-subsets linkage analysis

detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22. *Am J Hum Genet*, 2003. **73**: p. 1041-51.

195. Schmidt, S., Scott, W. K., Postel, E. A., Agarwal, A., Hauser, E. R., De La Paz, M. A., Gilbert, J. R., Weeks, D. E., Gorin, M. B., Haines, J. L., and Pericak-Vance, M. A., Ordered subset linkage analysis supports a susceptibility locus for age-related macular degeneration on chromosome 16p12. *BMC Genet*, 2004. **5**: p. 18.

196. Cuccaro, M. L., Shao, Y., Grubber, J., Slifer, M., Wolpert, C. M., Donnelly, S. L., Abramson, R. K., Ravan, S. A., Wright, H. H., DeLong, G. R., and Pericak-Vance, M. A., Factor analysis of restricted and repetitive behaviors in autism using the Autism Diagnostic Interview-R. *Child Psychiatry Hum Dev*, 2003. **34**: p. 3-17.

197. Ogdie, M. N., Macphie, I. L., Minassian, S. L., Yang, M., Fisher, S. E., Francks, C., Cantor, R. M., McCracken, J. T., McGough, J. J., Nelson, S. F., Monaco, A. P., and Smalley, S. L., A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: suggestive linkage on 17p11. *Am J Hum Genet*, 2003. **72**: p. 1268-79.

198. McCauley, J. L., Olson, L. M., Dowd, M., Amin, T., Steele, A., Blakely, R. D., Folstein, S. E., Haines, J. L., and Sutcliffe, J. S., Linkage and association analysis at the serotonin transporter (SLC6A4) locus in a rigid-compulsive subset of autism. *Am J Med Genet*, 2004. **127B**: p. 104-12.

199. Weiss, L. A., Veenstra-VanderWeele, J., Newman, D. L., Kim, S. J., Dytch, H., McPeek, M. S., Cheng, S., Ober, C., Cook, E. H., and Abney, M., Genome-wide association study identifies ITGB3 as a QTL for whole blood serotonin. *Eur J Hum Genet*, 2004.

200. John, S., Shephard, N., Liu, G., Zeggini, E., Cao, M., Chen, W., Vasavda, N., Mills, T., Barton, A., Hinks, A., Eyre, S., Jones, K. W., Ollier, W., Silman, A., Gibson, N., Worthington, J., and Kennedy, G. C., Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. *Am J Hum Genet*, 2004. **75**: p. 54-64.

201. Sandelin, A., Wasserman, W. W., and Lenhard, B., ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res*, 2004. **32**: p. W249-52.

202. Naar, A. M., Taatjes, D. J., Zhai, W., Nogales, E., and Tjian, R., Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes Dev*, 2002. **16**: p. 1339-44.

203. Laplante, J. M., O'Rourke, F., Lu, X., Fein, A., Olsen, A., and Feinstein, M. B., Cloning of human Ca2+ homoeostasis endoplasmic reticulum protein (CHERP): regulated expression of antisense cDNA depletes CHERP, inhibits intracellular Ca2+ mobilization and decreases cell proliferation. *Biochem J*, 2000. **348 Pt 1**: p. 189-99.