TOWARDS AN UNDERSTANDING OF THE ROLE OF CHROMOSOME 15Q11-Q13 IN IDIOPATHIC AUTISM

By

Ryan J. Delahanty

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

Human Genetics

May, 2010

Nashville, Tennessee

Approved:

Professor James S. Sutcliffe

Professor Chun Li

Professor Lawrence T. Reiter

Professor John A. Phillips III

Professor Scott M. Williams

I humbly dedicate this thesis:

In the memory of my grandfather, James L. Delahanty, Sr.; whose heart so inspired a love in his children and grandchildren;

To my wife, Katty, for encouraging to always aspire to be more;

To my wonderful children, Ainay and Aiden, the reason I wake each morning optimistic about mankind;

To my parents, James and Linda Delahanty, whose love, support, and encouragement shepherded me through my darkest days; and

To my early teacher, Tom Dale, for igniting in me an interest in biology and genetics that will burn throughout my days.

ACKNOWLEDGEMENTS

This work has been made possible by financial support from NIH grant MH061009-08.

I would first like to thank members of my thesis committee, Dr. Scott Williams, (chairmen), Dr. Chun Li, Dr. Lawrence Reiter, and Dr. John Phillips, whose guidance and assistance were indispensable to my development and the evolution of the work presented herein. Their friendship and support has been tremendous. I would like to thank my mentor, Dr. James Sutcliffe, for his patience, encouragement, and the opportunity to work on an important disease with a world-class team. In addition to supporting my work and our team, Jim always gave me latitude to pursue new ideas, which has been a wonderful fuel for my curiosity. Through thick and thin, small lab or large lab, with little or plentiful funding, we have persisted and now grown. His faith in me during challenging times has been tremendously important in the course of this journey.

I would like to acknowledge the staff support in the Vanderbilt Center for Human Genetics Research, most notably Lan Jiang, Kim Luci, and Melissa Potter, without whose support, none of this work would be possible. Ascertainment work by Genea Crocket and Sharlet Anderson has been important for increasing our sample size and increasing outreach to the autism community.

Vanderbilt's excellent DNA Resources core, particularly Cara Sutcliffe, Ping Mayo, and Jessica Blair have been extremely helpful for their use of

equipment and helpful access to genotyping reagents and resources, cell line storage, and DNA preparation.

I am greatly thankful to the current and past members of the Sutcliffe lab who have added so much light to my life with their friendship and collegiality. These people include Jacob McCauley, Edward Organ, Taneem Amin, Joshua Smart, Jackie Recktenwald, Rachel Game, Nicholas Campbell, Jude McElroy, and Emily Crawford. I have been fortunate to work with such a sharp, enriching, and very colorful group of individuals.

None of this work would be possible without the work of scientists and philosophers on whose shoulders we all stand. The philosophers of science, the basic scientists, specialists in psychiatric genetics, and the clinicians who contribute have constructed foundations of autism genetics on which we stand and from which we proceed. It is through all of them, collectively, that we move forward as a people and as a field.

Autism genetics relies most importantly on patients, parents, and caregivers who, with those mentioned above, come together and aspire to understand, treat, and possibly one day cure autism as we understand it.

My close friends from Nashville and beyond have been a beacon of light in their friendship, emotional, and spiritual support. This includes our spiritual family in the Church, especially Edward and Sharon Short. My thanks extend to William Walker, my classmate, for challenging me both mentally and physically, in the lab and in the gym. I am also thankful to my great friends from before graduate school, who wondered if I would ever graduate, and encouraged me so long to

finish, notably Charles Merz, Holly Ranney, Brooke Dyer, Liz Carry, Jack Crancer, and Walter Wedan.

I wish to thank my family. My mother and father, who have given me so much more than chromosomes, but also never judged (too much) and always encouraged me to be not just more, but to also be happy. Their love and support has stretched far beyond Michigan. Without them, I would no doubt have not completed this journey. My sister, Sarah Delahanty, has been an anchor to me and a model of what an unaffected sib with fifty percent of a common genome can do. My grandparents, Francis and Anna Carrico; James and Maggielee Delahanty, with my parents, have shown me love and wisdom about what is important, eternal, and persistent.

Finally, my wife and children, I thank them the most. I started graduate school much a boy, but they have matured me to become a man, a husband, and a father. Katty took a chance on me, and married a lesser man, someone not her equal. I have not always been easy, quite often the opposite, but her life is an example love, discipline, and aspiration. In our children, Ainay and Aiden, I see the best of both of us and together it makes me want to be better than I am.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADHD..... attention deficit hyperactivity disorder

ADI-R.....Autism Diagnostic Interview-Revised

ADOS..... Autism Diagnostic Observation Schedule

AbD.....Applied Biosystems Assays by Design (Taqman)

AoD..... Applied Biosystems Assays on Demand (Taqman)

AS.....Asperger syndrome

ASD .....autism spectrum disorders

BAP..... broader autism phenotype

bp..... base pair

cDNA ..... complementary deoxyribonucleic acid

CEPH.....Centre d'Etude du Polymorphism Humain

CLSA .....Collaborative Linkage Study of Autism

cM..... centiMorgan

CpG ..... cytosine phosphate guanine

CNV .....copy number variation

CNVR.....copy number variable region

dNTPs..... deoxyribonucleotides

DNA .....Deoxyribonucleic Acid

dup(15) .....interstitial or isodicentric duplications of 15q11-q13

DZ.....dizygotic twins

ECT2..... epithelial cell transforming sequence 2 oncogene

FBAT ..... Family Based Association Test

GABA ..... Gamma (γ)-aminobutyric acid

GABRA5 ..... $GABA_A$ receptor, alpha 5

GABARB3 ..... $GABA_A$ receptor, beta 3

GABARG3 ..... $GABA_A$ receptor, gamma 3

GCH1 ..... GTP cyclohydrolase 1

HapMap ..... International HapMap Project

HEK ..... Human Embryonic Kidney 293 cells

HLOD ..... heterogeneity LOD score

htSNP ..... haplotype tag SNP

HWD ..... Hardy-Weinberg Disequilibrium

HWE ..... Hardy-Weinberg Equilibrium

ICD ..... International Classification of Disease

Idic(15) ..... isodicentric 15 or inverted duplicate region of chromosome 15

kb ..... kilobase

LD ..... linkage disequilibrium

LOD ..... logarithm of the odds

MAF ..... minor allele frequency

Mb ..... megabase

MECP2 ..... methyl CpG binding protein 2

MLPA ..... multiplex ligation probe analysis

MRI ..... magnetic resonance imaging

mRNA ..... messenger ribonucleic acid

MZ.....monozygotic twins

NIMH..... National Institute of Mental Health

ng.....nanogram

OMIM.....Online Mendelian Inheritance in Man

OR ..... odds ratio

OSA ..... ordered subset analysis

P11S..... Proline 11 to Serine substitution in *GABRB3*

PCR .....Polymerase Chain Reaction

PDD .....pervasive developmental disorders

PDD-NOS ..... pervasive developmental disorder-not otherwise specified

PDT ..... Pedigree Disequilibrium Test

PET..... Positron Emission Tomography

PWS.....Prader-Willi syndrome

QTL..... quantitative trait locus

ROH..... runs of homozygosity

siRNA..... small interfering RNA

SNP ..... single nucleotide polymorphism

SSRI .....selective serotonin reuptake inhibitor

Taq..... Thermus aquaticus polymerase

TDT..... Transmission Disequilibrium Test

TFO..... to-female-only transmission analysis

TMO..... to-male-only transmission analysis

UBE3A ..... Ubiquitin Protein Ligase E3A

UPD ..... uniparental disomy

UTR ..... untranslated region

CHAPTER I

INTRODUCTION TO AUTISM: HISTORY, EPIDEMIOLOGY, AND GENETICS

Overview

Classically defined, autism (OMIM 209850) is a neurodevelopmental disorder characterized by deficits in language and reciprocal social interaction, as well as repetitive, restricted behaviors and interests. Narrowly defined autism or autistic disorder is a specific diagnosis in DSM-IV category termed Pervasive Developmental Disorders (PDDs). PDDs include autism, Asperger syndrome, Rett syndrome, PDD not otherwise specified (PDD-NOS), and the very rare childhood disintegrative disorder. Autistic disorder, PDD-NOS, and Asperger syndrome are collectively termed autism spectrum disorders or ASDs, though "autism" is often used interchangeably to describe those with all of these ASDs and not necessarily narrowly defined autism. Autism has a prevalence of about 1 in 150 individuals, with males affected at four times the rate of females [1]. ASDs typically manifest before the age of three, and span the life of the individual. While autism is typically a devastating disorder, a spectrum of functioning exists ranging from high functioning individuals with normal IQ to those with severe cognitive impairment requiring permanent institutional care. The estimated lifetime cost of care for a person with autism is estimated at $3.2 million dollars, and the total cost to society in the United States is upwards of $35 billion dollars a year [2]. Despite the evidence for strong heritability in autism and the

consensus that its etiology is primarily genetic, the nature of that underlying genetic etiology of the majority of cases remains elusive [3].

A Brief History of Autism

Swiss psychiatrist Eugen Bleuler, originally coined the term “autism” around 1912 to refer to *“*an escape from reality” *[4]*. Hans Asperger first used the term in a psychiatric context to describe “autistic psychopaths” in 1938 [5]. Asperger described a milder syndrome (now called Asperger syndrome) resembling autism, but that leaves language intact. In his original patients, he noticed a lack of empathy, little interest in forming friendships, one-sided conversations, and often clumsy movements. He also noted that these children frequently showed highly circumscribed interests and thus he termed them “little professors”.

In 1943, a contemporary of Asperger’s, Leo Kanner, published “Autistic Disturbances of Affective Contact” in the journal *Nervous Child*. Kanner adapted this term to describe children he believed to be afflicted with a syndrome not previously described, labeling it infantile autism that was sometimes referred to as Kanner’s syndrome. Kanner noted that affected individuals lacked the typical enjoyment of social interactions, frequently preferring solitude instead. His patients had specific features that included this autistic aloneness and the desire for the maintenance of sameness. His original cohort also exhibited abnormal speech with echolalia, pronominal reversal, literalness and inability to use language for communication; these are language deficits frequently noted in

autism as it is described today. Kanner suggests that any intrusion into the outer or inner world of these children often caused extreme distress often presenting in temper-tantrums [6,7]. Kanner anecdotally noted that these children were virtually all the offspring of university faculty and that, among the parents, there were very few "*warmhearted fathers and mothers".* He further noted that the parents and grandparents were often "*preoccupied with abstractions of literary, scientific, or artistic nature,"* and showed limited interest in socializing with others [6]. The observations of subclinical ASD traits in these family members, later termed the "broader autism phenotype," would prove to be valuable in studying the genetics of the disorder.

From Kanner's time and through the 1970s, 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. So called "refrigerator mothers" were claimed to lack maternal warmth or emotion were blamed for autism. This view persisted as late as 1976 as exhibited in a review by Hanson and Gottesman entitled "The genetics, if any, of infantile autism and childhood schizophrenia," in which they concluded that that no strong evidence exists implicating genetics [8]. The first twin study by Rutter and Folstein in 1977 was extremely influential in moving the field to understand that autism had a significantly genetic etiology [9,10]. The current concept of autism as a biomedical disorder was driven largely by these and other twin and family studies that document its heritability and association with a variety of genetic syndromes

[9,11,12,13]. Today, the definition of autism, understanding of its epidemiology and genetics, have all been greatly refined.

The Autism Phenotype

In the 65 years since Kanner's description, the concept and diagnosis of autism has changed dramatically. As the understanding and definition of autism has evolved, autism is now placed under the heading of pervasive developmental disorder (PDD) with Asperger syndrome, Rett syndrome, and pervasive developmental disorder not otherwise specified (PDD-NOS) in the Diagnostic and Statistical Manual of Mental Disorders IV [14].

As a phenotype, what is termed "autism" is often meant to include not just autism in a classical sense, but also autism spectrum disorders (ASDs). Autism is characterized by problems in three domains: (1) deficits in language and communication, (2) difficulties with social interaction and reciprocity, and (3) patterns of stereotypic behaviors and restricted interests [14]. The presentation of autism can vary dramatically from mild to severe onset is typically before three years of age [14]. If threshold deficits in these core domains are met, the diagnosis of autistic disorder, sometimes termed "classical autism" is given. If criteria for autism are not met, one may be diagnosed with one of two other Autism Spectrum Disorders (ASDs), Asperger syndrome or PDD-NOS. PDD-NOS makes up the majority of diagnosed ASDs. The disturbance must not be better accounted for by Rett syndrome or the very rare childhood disintegrative

disorder. The ICD-10 uses essentially the same definition and criteria [15]. The domains affected in each of the three major ASDs are shown below in Table 1-1.

Table 1-1 Domains of impairment in the major ASDs[1].

Domain	Autistic disorder	Asperger	PDD-NOS	All ASDs
Social communication	All	All	All	
language	All	-	Variable	
Repetitive and restrictive behaviors	All	All	Variable	
Sensory abnormalities	>90%	80%	Variable	94%
Developmental regression [2]	15%-40%	?	?	15%-40%
Motor signs [3]	60%-80%	60%	60%	60%-80%
Sleep disturbance	55%	5%-10%	40%	50%
Gastrointestinal abnormalities [4]	45%	4%	50%	4%-50%
Epilepsy [5]	10%-60%	0%-5%	5%-40%	6%-60%
Comorbid psychiatric diagnosis [6]	70%	60%	>25%	25%-70%
Cognitive impairment (IQ<70)	70%	-	>70%	70%

[1] If required for all diagnosis, this is indicated by the word "All"; if not observed, a – is used. Estimates used are from references cited [16,17,18,19,20,21,22,23]. The far right column gives the estimate of frequency in all common ASDs. [2] Reduction of abilities in language and social skills. [3] Hypotonia, gait problems, toe walking, apraxia. [4] Six months or more of diarrhea, constipation, reflux, or bloating. [5] Range of epilepsy estimates reflects presence of other comorbid features, such as intellectual disability and CP, which may dramatically increase epilepsy risk. [6] Mood and conduct disorders, aggression, ADHD.

In addition, a number of comorbidities are common in autism and include sensory abnormalities, developmental regression, motor signs and symptoms, sleep disorders, seizures/epilepsy, gastrointestinal abnormalities, and comorbid psychiatric diagnoses [16,17,18,19,20,21,23,24]. Approximately 70% of those with autism also meet criteria for mental retardation, now termed intellectual

disability [16]. The frequency of the comorbidities is also estimated for the three major ASDs in table 1-1. Comorbidity of ASDs with other neuropsychiatric disorder (such as anxiety, depression, ADHD) further underscores the need to accurately parse quantitative endophenotypes that may shed light on pathways common to comorbid diseases.

Diagnosis and measurement of ASDs core components and comorbidities has improved dramatically with the development of a number of diagnostic instruments. The Autism Diagnostic Interview current revision (ADI-R) along with the Autism Diagnostic Interview (ADOS) are the two most commonly used research instruments for the careful measurement of the core and co-morbid features of autism [25,26,27,28].

The ADI-R is an interview given to the child's primary caregiver or caregivers and consists of 93 questions assessing these features [27]. The ADOS is an age-adjusted interactive interview with the child which consists of four modules that gauge responses to tasks or activities [26,28]. Each of these tests measures deficits in categories of language and social skills and patterns of repetitive and restrictive behavior. Using these instruments, the extent of autistic impairment and other co-morbidities may be indexed in a semi-quantitative manner. If scoring thresholds are met for domains in a given diagnostic instrument, a child may meet criteria for autism. Sub-threshold measurements in any one category may still earn the subject a label of autism spectrum disorder. These tools, along with a clinician's best judgment, can thus help to distinguish

autism from other Pervasive Developmental Disorders (PDDs), including Asperger syndrome, and PDD not otherwise specified (PDD-NOS) [29].

The clinical heterogeneity of the disorder across the core domains makes ASDs challenging to analyze. Children may differ widely in intellectual ability. Two patients with a diagnosis of autism may present quite differently: one may have above average IQ, the other with severe intellectual disability, which is common in 30%-70% of those with autism [1,16]. The developmental trajectory of those with ASDs can often vary widely with some showing dramatic improvement over time and others making little or no improvements, though higher IQ appears to be associated with more favorable outcomes [30]. Since ASDs are not thought to be single conditions based on pathophysiology or etiology, they are considered rather to be clinical syndromes. Some recent progress in genetics has permitted the identification of subsets of ASDs based on etiology.

Autism Epidemiology and Risk Factors

ASDs, as defined above, affect 1.5 million Americans of all ancestral, socioeconomic, and geographic backgrounds and carries a prevalence rate of approximately 1 in 150, though classic autism (i.e. autistic disorder) afflicts just 1 in 500 [1,31,32]. This means that ASDs are more prevalent than many other "common" childhood disorders [3].

Advanced maternal and paternal ages have emerged as replicable risk factors for ASDs [33,34,35]. It has been postulated that this might be due to (1) increased risk for chromosomal non-disjunction or non-alleleic homologous

recombination during maternal oogenesis, (2) increased mutation burden in older sperm, and (3) parents with features of the broader phenotype often marry later [35,36]. However the extent to which pathogenic CNVs and point mutations arise maternally or paternally and the true risk that advanced maternal and paternal age is a risk factor in autism has yet to be determined.

A few of other events appear to be correlated with increased risk of autism. These include a family history of psychiatric conditions (such as schizophrenia), multiple births, cerebral palsy, and cognitive impairment.

While the number of diagnosed cases of autism has increased several fold in the past decade, the rapid increase appears to be most critically driven by awareness of the disorder, the changing definition of autism, increased diagnosis by physicians, and the availability of diagnostic tools [14,37,38].

Further, since educational and other resources, medical coverage, and reduced stigma may be associated with a diagnosis of autism compared to “mental retardation” (now called intellectual disability), a number of cases of diagnostic substitution appear to have played a role in the increased prevalence of autism [39]. States or regions which offer greater resources to those with autism as a diagnosis instead of alternative diagnoses will therefore have higher rates of autism ascribed, though such diagnoses may not be accurate. Therefore, there has been a marked decrease in the number of diagnosed cases of mental retardation and a concordant increase in the number of diagnoses of ASDs [40,41].

The Genetics of Autism Spectrum Disorders

Twin and family studies and large epidemiological studies have shown that autism is primarily a genetic condition. It is now widely appreciated that ASDs are genetic disorders and that the genetics of these disorders is complex and confounded by locus and allelic heterogeneity and potentially by oligogenic inheritance, epigenetics, gene-gene interactions [23,24].

Twin and family studies indicate that autism is a primarily genetic disorder with about 90% of the disorder explainable by genetics (i.e. broad sense heritability of about 0.9) and a concordance ranging from 60%-92% for monozygotic twins (MZ) and 0%-10% for dizygotic twins (DZ) depending on the phenotypic definition [42,43]. Males are affected four times as frequently as females, and epidemiological studies indicate a sibling risk ratio (sibling recurrence risk / population prevalence) in the range of 25-67 [44,45]. This makes autism the most heritable of psychiatric disorders.

The search for autism susceptibility alleles has included linkage studies, CNV studies, and genome-wide association and candidate gene studies, each briefly reviewed below. It will be shown how these lines of evidence, while each contributing to our understanding of autism, complement one another and support the proposed aims of the thesis.

Family studies indicate that family members of those with idiopathic autism often show subclinical traits of autism, collectively called the broader autism phenotype, and it is thought that autism may manifest through the accumulation of genes that increase the degree to which these traits are

expressed [46,47]. Studies of families with a history of ASDs have shown that first-degree relatives (i.e. parents and siblings) have higher frequency of characteristics common to ASD phenotypes (e.g. language deficits, social responsiveness differences, and presence of repetitive or restrictive behaviors) than in non-autism family controls and the general population [48,49,50,51]. Thus, it seems reasonable that a measurable proportion of the genetic liability to ASDs in some cases might take the form of risk alleles that increase risk for some or all of the component traits of autism (i.e. the broader autism phenotype).

Also clear is that the genetic architecture of autism is complex since, aside from a growing number of copy number variations (e.g. 15q11-q13, 16p11.2, 22q11.2) that can carry with them a frequently clear diagnosis of autism, the majority of cases appear to be spontaneous and likely polygenic in nature [52]. Approximately 10% of individuals presenting clinically with autistic features can be can be diagnosed with common single-gene disorders with an etiologically distinct cause. These include fragile X syndrome (OMIM #300624), tuberous sclerosis complex (OMIM #191100) [24,53], Angelman syndrome (OMIM #105830) [54] and Rett syndrome (OMIM #312750), [55] and others [3].

Twin- and family-based studies suggest that for most idiopathic cases, the most parsimonious model involves multiple genes and potential epistasis that collectively give rise to ASD phenotypes [56]. Given the high heritability, it was thought that identifying autism genes would be a relatively straightforward; however it is currently only possible to identify the etiological cause of a approximately 10-20% of all cases [57,58,59]. Genetic abnormalities known to be

associated with autism include a number of copy number variants (CNVs), gene mutations (*MECP2*, *SHANK3*, *NLGN3*, *NLGN4X*), and cytogenetic abnormalities [60,61,62]. The majority of cases, however, have no apparent etiology.

Given the former findings of increased rates of the broader autism phenotype in some families without obvious molecular etiology in combination with the 10-20% of cases which have an identifiable etiology, a model emerges that may accommodate both of these observations. This model may involve the applied force of CNVs, point mutations, and common variation with epistatic forces to account for the phenotypic variation seen in autism and the finding of subclinical BAP traits in family members. It is apparent that there are two emergent themes in recent analysis of the data: a case for highly penetrant rare variants that dramatically increase risk, and a case for more common variants that collectively increase risk for presentation of each of the core phenotypic domains.

The genes selected for study, namely *UBE3A* and *GABRB3* of the 15q11-q13 region, were selected not just because of mechanism, but because of multiple convergent lines of evidence supporting their selection. The evolution of

Molecular and Statistical Genetics: Linkage Analysis

Genetic linkage analysis in human populations is used to analyze multiplex family pedigrees for linkage between a chromosomal region and a trait. The statistic most frequently used is the logarithm of the odds (LOD) score or some variant thereof. In brief, the LOD and its related linkage statistics measure

the likelihood that affected family members share a chromosomal region more often than would be expected by chance alone [63]. The power of linkage is that it does not presuppose a *common disease-common variant* model of a given disease. Rather, linkage should be able to detect both highly penetrant rare variants and, to lesser extent, variants of more modest effect sizes.

Despite the benefits of linkage, in the search for autism susceptibility loci, linkage signals have appeared on every chromosome, reflecting significant locus heterogeneity [64]. The reasons for this are incompletely understood. It may be that sampling variation or ascertainment biases do not identify comparable samples. This may be due in part to the 'winner's curse' that may result in association and linkage for the reporting of a first instance to reach a suggestive or significant LOD score or P-value [65]. As a result, few replication studies find the same effects in independent data sets. Replication studies may need to be far larger than initial studies to detect loci conferring minor effects [66]. From recent reviews [64,67] and search of Pubmed, it appears nearly 20 genome-wide linkage studies have been conducted to date. The few replicated findings common between are studies are reviewed in Table 1-2.

Table 1-2 Chromosomal regions showing linkage in more than one study.

Region	Candidate genes	References
2q24-q31	*SLC25A12, CENTG2*	[68,69,70]
5p13-p14	-	[45,71]
7q22-q32	*RELN, MET, CADPS2*	[70,72,73,74]
7q34-q36	*CNTNAP2, EN2*	[72,75,76]
9q33-q34	*TSC1*	[45,77]
15q11-q13	*GABRB3, UBE3A*	[78,79,80]
17q11-q21	*SLC6A4, ITGB3*	[71,81,82,83]

Some of the findings above may result from regions containing multiple genes that could increase risk for autism which would make detection and replication easier for the above regions.

One noteworthy point from review of linkage studies is that many of the significant findings (as well as part of the reason for failure to replicate) have come from not analyzing autism as a categorical trait, but rather analyzing the individual components of autism such as measures of language. In fact, methods such as ordered subset analysis (OSA), and quantitative trait linkage (QTL) mapping have shown greater promise for identifying genes with alleles that increase severity of language deficits or rigid compulsive behaviors, as examples [75,84]. Several studies showed by QTL mapping for “age of first word” that a language locus existed on 7q34-q36 [75,76,85]. Thus restricting analysis to the component traits in autism (see Table 1-1), there may be greater sensitivity to detect loci through use of linkage studies.

These results seem to support an oligogenic model, in which ASDs may result from perturbations of many different genes. There is some evidence that single deleterious mutations in such genes are sufficient to cause autism, but also evidence that many such genes may contribute in some manner to the deficits characteristic of the autism phenotype. These results support the view that multiple risk alleles may act independently or together through additive or epistatic mechanisms to modulate risk for autism [86].

Some loci, such as 17q11.2, which harbors the serotonin transporter gene (*SLC6A4)*, show linkage in families with only affected males, but little or no linkage in families with affected females [82]. Sex-chromosome linkage analysis, too, has been disappointing despite the 4:1 difference in male:female prevalence, with most studies assessing linkage of the X chromosome not yielding significant results [87]. This could be explained in part by reduced power to analyze allele sharing on the X chromosome in a collection of autism families. The absence of consistent linkage signals on the X chromosome may be due to rare or *de novo* variants in multiple X chromosome loci are too numerous. There may be simply too many X chromosome genes which when upset by mutation in hemizygous males can result in cognitive impairment or autism. It is also possible that skewed X chromosome inactivation in females could provide a challenge to identifying risk loci or the presence of two X chromosomes could buffer females from the higher frequency of ASDs. Another possibility is that autosomal loci show differential penetrance in males compared to females as the results of male-only family analysis on chromosome 17q11.2 and other studies seem to suggest [82,88,89].

Locus heterogeneity, allelic heterogeneity, phenotypic heterogeneity and epistatic interactions may also confound the ability to replicate linkage signals. Despite these difficulties, several genes showing evidence for linkage replicated across two or more linkage studies have indicated the involvement of rare and/or common variants with ASDs [88,90,91,92,93,94].

Molecular Genetics: Candidate Gene Studies

A number of candidate genes, some underlying linkage peaks and others not, have been studied for the presence of rare or common variation. Candidate gene studies have most typically taken the form of resequencing of exons in candidate genes or association studies. The former seeks to identify rare variants that result in rare coding changes that impact the resultant protein. The latter seeks to identify common variants and alleles that are over-transmitted to affected individuals.

Mutation in a number of single genes can result in syndromes that present with features of autism, but are clinically distinct. Examples include Angelman syndrome (*UBE3A*), Fragile X syndrome (*FMR1*), Rett syndrome (*MECP2*), Timothy syndrome (*CACNA1C*), and tuberous sclerosis (*TSC1* and *TSC2*) [95,96,97,98,99]. While mutations in these provide a molecular diagnosis of a relevant disorder, these conditions nevertheless provided strong hypotheses regarding the types of genes, proteins, and pathways that might contribute to what we would consider idiopathic autism and that common pathways may be affected in autism and these disorders.

The X-linked neuroligins and SHANK3, the respective proteins which interact to guide glutamatergic synapse formation, have recently been shown to be associated with autism by virtue of rare deletions or point mutations [60,62,100]. Sanger sequencing identified mutations in two neuroligins (*NLGN3*, *NLGN4X*, and a gene which interacts with the neuroligins, *SHANK3*, all genes involved in the formation of synapses, were early victories in sequencing which

indicated that potential failure to form synapses might result in ASDs [60,62]. Neurexin 1 (*NRXN1*) was identified as a deleted gene in a 2007 paper by the Autism Genome Project [45]. By extension of these findings, resequencing of *NRXN1*, which also interacts with the neuroligins, identified rare coding variants in this gene [101].

Other genes with autism-associated variants include *SLC6A4*, the serotonin transporter. These variants have been shown to confer one of two distinct gains of function on the activity of the protein [88]. *CNTNAP2* has been shown to have both rare and common variants associated with autism [90,102]. Some genes, such as Engrailed 2 (*EN2*), have shown to have not just coding variants but also a functional intronic variant shown to be associated with autism that have been identified by resequencing [92,103].

One difficult obstacle here is that the rarer the variant and the lower the penetrance, the greater is the threshold of samples required to demonstrate statistical significance or to evaluate the total contribution to ASD risk. The availability of next generation sequencing methods will allow the rapid resequencing of entire genes, not just coding regions.

By contrast to sequencing studies which seek to find more highly penetrant *rare* alleles, association studies typically seek to find more *common* alleles associated with autism typically through case-control analysis or family-based association tests. A summary of genes with evidence of common alleles associated with autism is presented in Table 1-3.

Table 1-3 Common variants associated with ASDs.

Gene	Function (position)	Type[1]	References
CNTNAP2	Neurexin family, important in potassium channel clustering (7q35)	C/R	[85,90]
EN2	Involved in cerebellar formation(7q36)	C	[91,92]
GABRB3	Neurotransmitter receptor (15q12)	C	[104,105,106,107]
ITGB3	QTL for blood serotonin levels (17q21)	C	[94,108]
MET	Proto-oncogene receptor (7q31)	C	[109,110]
NLGN3/NLGN4X	Neuroligins involved in synapse formation (Xq13/Xp22)	R	[62,111]
NRXN1	Neurexins interact with neuroligins to permit cell adhesion (2p16)	R	[101]
SHANK3	Synaptic adaptor which interacts with neuroligins	R	[60,61]
SLC6A4	Transports serotonin into cells (17q11)	C/R	[88,112,113,114]
UBE3A	Angelman syndrome gene and ubiquitin ligase important for synapse formation (15q11)	C/R	[115,116]

[1] C = common variants, R = rare variants

Association studies have become increasingly frequent in the past ten years as the cost of genotyping has decreased and both more samples can be genotyped for a single or small number of SNPS (e.g. Taqman, Sequenom) or a single sample can be genotyped for multiple SNPs (e.g. Affymetrix 5.0 and 6.0, Illumina 610S and1M chips). This in combination with awareness of the haplotype structure of the genome, made possible most notably by the International HapMap Project, has made it possible to more thoroughly index most common alleles of various human populations. As the 1000 Genomes project races toward its completion, our awareness of rare and common SNPs and their corresponding haplotypes will continue to increase.

In contrast to the findings of most sequencing efforts, most associated SNPs appear to be intronic, with *MET* and *ITGB3* being notable exceptions

[93,109]. Most of these genes do not have any functional variants in linkage disequilibrium to explain the observed association. Frequently groups find association in the same gene, but in different regions or with different alleles.

The disconnect between association and apparent functional effects is the source of much consternation in the field of autism genetics and other gene or genome-wide association studies. One of the more promising avenues for explaining variants at great distance from candidate genes is that these variants have an impact on gene expression in cis- or in trans-. The tools to investigate these possibilities are in early stages of development, but tools such as SCAN may allow us to investigate the possibility that associated SNPs (and those in strong LD) may act as quantitative trait loci which influence gene expression [117].

Using endophenotypes or specific traits within the autism spectrum to select more homogenous populations or to define quantitative traits for analysis of allelic association may offer greater sensitivity and specificity in detecting loci or specific alleles in some cases. But the lesson from recent studies is that both common and rare variant studies should be pursued in parallel.

One more point that emerges from the study of candidate genes is the importance of genes involved in synapse formation (see Table 1-3). This notion of autism, posited by Zoghbi 2003, as potentially a synaptopathology, or disorder of the synapse, seems supported in many ways by these studies and others described below [118].

Molecular Genetics: Copy Number Variation

Maternal interstitial duplication of 15q11-q13 was the first CNV (excluding the clinically well-defined Fragile X disorder) found to be frequently associated in autism [119]. It remains today the most frequent recurrent CNV in autism.

Since this time, copy number variation (CNV) has been identified to be an important source of variation in the genome. Copy number variation frequently arises through non-allelic homologous recombination events where there is unequal crossing over and joining between chromosomal regions with high levels of homology. Typically, the greater the homology, the greater is the likelihood of non-allelic homologous recombination. Two major papers in late 2004 one by Sebat et al. and another by Iafrate et al. indicated that copy number variation was a common element in human genomes [120,121]. It was later shown by Conrad et al., in the HapMap samples, that genotyped SNPs could be mined for tracks of Mendelian inconsistencies or homozygosity to determine if CNVs existed [122]. However, it was not until 2007 when the Affymetrix 500K chips became available that measures of intensity, homozygosity, and Mendelian inconsistency could be used together to map common copy number variation in the human genome. By the estimate of one influential paper by Redon et al., there were over 1,400 copy number variable regions (CNVRs) in the 270 HapMap samples [123]. These CNVRs spanned over 360Mb or 12% the human genome and notably encompass a greater number of total bases than are estimated to be SNPs in the genome.

Vorstman et al. showed in 2006 that novel structural variation observed in previously published reports, often overlapping with regions demonstrating linkage with autism, were associated with autism [124]. A new confirmation of the importance of CNVs in autism was reported by Sebat et al. in which the authors document a highly significant association of *de novo* CNVs in 10% of 118 patients with sporadic autism vs. 1% of controls [125]. The finding is striking since those with syndromic forms of autism (severe MR, dysmorphologies, or known cytogenetic abnormalities) were excluded from the study, highlighting the importance of CNVs in the genomic architecture of autism.

A number of studies since this time have increased resolution and now the importance of CNVs in causing autism is known to be greater than previously expected. Estimates range between 6%-44% of all cases which bear potentially pathogenic CNVs [3,45,125,126,127,128]. The highest estimates are for cases where all syndromic forms of autism are included (i.e. non-idiopathic cases). The most notable of these CNVs include 15q11-q13 duplication, 16p11.2 deletion and duplication, and 22q11.2 (*SHANK3)* deletions [127]. These and other major CNVs are shown in Table 1-4.

Table 1-4 CNVs known to be associated with autism.

Region (loss/gain)	Genes (candidate)	Size (kb)	Transmitted/ *De novo*	References
2q37.3 (loss)	>40	>2000	*De novo* > Trans	[124,129,130]
5p15.2 (loss>gain)	>50	<15000	Both	[124,131,132]
15q11-q13 (gain)	8-35 (*UBE3A*, *GABRB3*, others)	2000-6000	Both	[119,133]
16p11.2 (both)	24-31	>750	*De novo* > Trans	[128,134,135]
22q11.2 (loss)	8-61 (SHANK3)	750-4500	*De novo* > Trans	[60,124,136,137]
Xp22 (loss>gain)	>20	>5000	Both	[100,124]

[1] C = common variants, R = rare variants

Smaller structural variation, in the form of CNVs, is the next logical step in the progression from macroscale to microscale detection of genomic variation. Even the most common CNV known to cause autism, 15q11-q13, is difficult to see by G-banding and often requires FISH or a specific focus on this region with high quality metaphase spreads [138]. As resolution has improved in the past several years, a number of microdeletions and microduplications within or flanking 15q11-q13 have been identified and shown associated with autism, schizophrenia, and epilepsy among other disorders [139,140,141,142]. The march from G-banding to array-based methods and SNP chips with increased coverage dramatically increases our ability to detect duplication and deletion,

though not necessarily inversion events. Since mutations ranging from point mutation to megabase deletions have been shown to often cause the same phenotype, the line between calling one syndromic and the other idiopathic is increasingly blurred. SHANK3 or the X-linked neuroligins, are examples of genes where both copy number variations or point mutations can result in autism. As technology moves from arrays to sequence level data, our ability to detect single base pair changes and rare potentially pathogenic changes will likely be the next frontier [143].

Generally speaking, CNVs are many are also seen in controls suggesting they are not significant risk factors (see table 1-4), as many are transmitted from non-autistic parents. This incomplete penetrance suggests a model where other factors (environment, modifier genes, sex, parent-of-origin, etc) may modify whether or not the autism phenotype is expressed in events of these CNVs.

These *de novo* events may be more common in simplex ASD families (7-10%) than in multiplex ASD families (2-3%) [58,128]. However, it is important to note that given emerging estimates of sibling recurrence risk as high as 20% and the tendency of families with an autistic child to stop reproducing, a significant fraction of “simplex” families are multiplex families that are not realized because of limited sibling number. Nevertheless, if you accept the simplex vs. multiplex trend and if you consider common variant findings two models may be suggested: one where new highly penetrant mutations are occur, and another where multiple risk alleles act in conjunction with one another to increase risk. These models are not mutually exclusive and typical autism may present as the

result of a collection of both lower penetrance and higher penetrance variants. By this model, first degree relatives of probands in multiplex families would be more likely to harbor traits of the broader autism phenotype, as has been reported [144]. Indeed, it appears that in multiplex families, parents are more likely to display elements of the broader autism phenotype where in simplex or "sporadic autism" families the rates of the broader autism phenotype are nearer that observed in the population [144].

Maternal Duplication of 15q11-q13 in Autism

As mentioned above, the first and most common CNV seen in autism is duplication of 15q11-q13, which is estimated to occur in 1%-3% of all individuals with autism [145]. This genomic region contains a large number of low copy repeats with high levels of homology which allow for non-allelic homologous recombination at its breakpoints, thus making the region subject to a number of genomic disorders including Angelman syndrome/Prader-Willi syndrome and autism. The region on which my studies focus has been refined by the mapping of these breakpoints. Duplication of the region (dup(15)) can take two major forms, which include the more common isodicentric chromosomes (idic(15)) and the less common interstitial duplications spanning 4-5 Mb of the region (int dup(15)) (Figure 1-1). It now appears that the major forms of the interstitial duplication and the idic(15) have recurrent breakpoints, the with a common interval spanning BP2-BP3 (see Figure 1-1, panel B) [146]. The recurrent nature of the most common breakpoints (i.e. BP1, BP2; see Figure 1-1) has helped to

identify the genes between BP2 and BP3 as the most commonly duplicated genes in this genomic interval [147]. Typically, duplications of greater size (i.e. idic(15) or larger interstitial CNVs) are associated with more severe phenotypes.

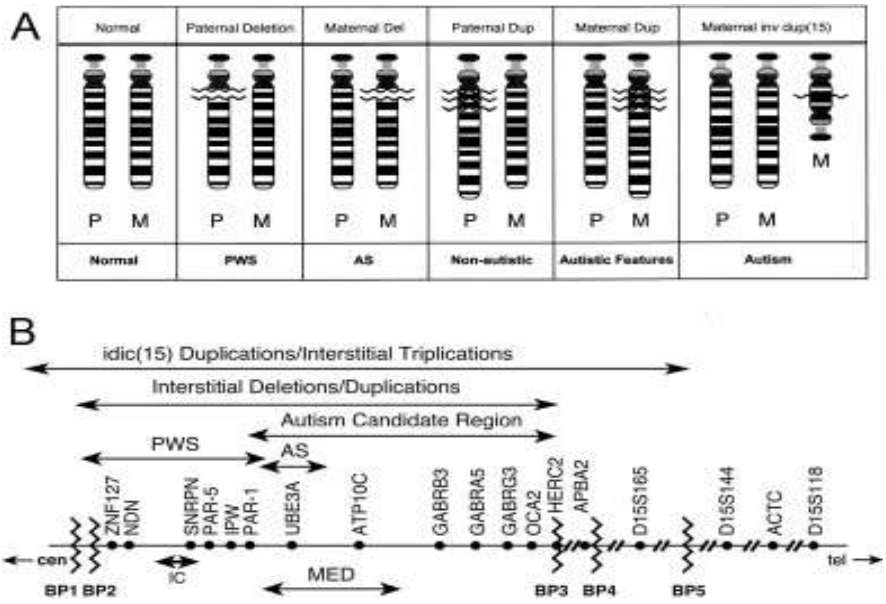


FIGURE 1-1 Genomic disorders of 15q11-q13 and genes of the interval.
A. Deletions and duplication disorders of 15q11-q13.
Genomic disorders resulting from deletion or duplication of 15q11-q13 are shown. In this figure, P indicates chromosomes of *paternal* origin, and M indicates chromosomes of *maternal* origin. From left to right can be seen (1) karyotypically normal individual with two copies of chromosome 15; (2) paternal deletion of 15q11-q13 resulting in Prader-Willi syndrome (PWS); (3) maternal deletion of the 15q11-q13 interval resulting in Angelman syndrome (AS); (4) paternal interstitial duplication of the region which confers much lower risk for ASD phenotypes; (5) maternal interstitial duplication of the region which typically results in autism; and (6) the isodicentric chromosome containing four copies of 15q11-q13 is typically associated with the most severe autistic phenotypes.
B. Chromosome 15q11-q13 and the autism candidate region.
The chromosomal region indicated in panel A that is subject to duplication and deletion is shown here. Genes involved in Prader-Willi syndrome and Angelman syndrome are indicated (PWS / AS). The autism candidate region is identified and contains the genes central to the studies reported here: *UBE3A* and *GABRB3*. Jagged lines indicate the recurrent deletion/duplication breakpoints (BPs). All forms of the duplication contain, at a minimum, the BP2-BP3 region. The region contains maternally and paternally expressed genes (PWS / AS respectively) under epigenetic control of the imprinting center (IC).

Mutation Spectrum and Phenotypic Overlap Between Autism, PWS, and AS

The 15q11-q13 interval is subject to genomic imprinting with paternal deletion of the region resulting in Prader-Willi syndrome (PWS) (OMIM 176270) and maternal deletion resulting in Angelman syndrome (AS) (OMIM 105830) [95,148]. The influence of epigenetics in the etiology of the disorder is evident in the broad array of disorders which can result from such these structural variations (Figure 1-1). Several genes including *SNRPN, NDN*, and many smaller snoRNAs are transcriptionally silenced on the maternal allele [149,150]. *UBE3A* is subject to maternal-specific expression in selected regions of the brain and such gene regulation is essential to normal neuronal development [151]. More recent work has shown allele specific expression or allelic exclusion is more common than previously thought [152].

AS is caused by both maternal deletion of 15q11-q13 (70%), or point mutations in *UBE3A* (10%), defects in the IC (i.e. “imprinting mutations”) (5%), and paternal UPD (2%), with the remaining patients with a clinical diagnosis not having any identifiable molecular defects in 15q11-q13 [95,153]. Since the vast majority of dup(15) forms which result in autism are of maternal origin, this makes UBE3A of particular interest. PWS cases are caused by paternal deletions of the 15q11-q13 interval (70%), maternal uniparental disomy (UPD) (25%), and imprinting mutations caused by small deletions in the IC (5%) [154].

While these conditions are very distinct, it is worth comparing them as there is some phenotypic overlap between autism, PWS, and AS, with the genomic disorders sharing some unique and some common characteristics.

PWS presents neonatally with hypotonia, failure to thrive and a poor suck reflex. As patients grow older, the most prominent symptom is hyperphagia and resulting obesity. Most patients show intellectual disability and various behavioral problems [155]. One study reported that 10 out of 56 PWS cases had a history of seizures [156]. One report suggests that individuals with an extra maternal chromosome, but without a paternal chromosome (maternal UPD) may be more likely to express symptoms of autism compared to those with the paternal deletion [157]. This is intriguing particularly since those with maternal duplications of 15q11-q13 are very likely to manifest autism and such patients have increased expression of maternally-expressed genes [158]. Individuals with PWS often have compulsive, repetitive, or ritualistic behaviors analogous to those seen in autism cases [159]. It is also appears that those with the maternal UPD form of PWS have social deficits similar to those in autism [160]. Veltman et al found elevated scores on the Autism Screening Questionnaire and their review of the literature shows that those with UPD show overexpression of the maternally-expressed genes increasing the risk of having an ASD diagnosis [161,162].

AS is a severe neurological disorder that presents with profound intellectual disability, absent speech, epilepsy, ataxia, hand-flapping, and inappropriate laughter [163]. Many of these features are seen (though typically

with less severity) in those with maternal duplication of 15q11-q13. Deletion AS patients tend to have more profound abnormalities than those with UPD and imprinting defects [164,165]. Those with point mutations in *UBE3A* have characteristic seizures, absent speech, and microcephaly, but they still tend to develop better motor skills and the ability to follow simple commands, unlike most with the typical deletion [165]. Mutations in 3′ exons of *UBE3A* tend not to be associated with autistic traits, while those in 5′ exons are more often associated with ASD traits [153].

Approximately 50% of individuals with maternal interstitial duplications (int dup(15)) and 88% of individuals with the isodicentric chromosome idic(15) meet criteria for autism [146][166]. For those meeting criteria for autism, ADI-R domain scores are indistinguishable from apparently chromosomally normal individuals with idiopathic autism. That 50% of individuals with the interstitial deletion do not meet criteria for autism (though they may have other problems) suggests that other factors may modulate the expression of the phenotype in those with the duplication. Idic(15) cases are more profoundly affected and accompanied by frequent physical findings (hypotonia, cleft palate, etc.) All show delay in physical developmental milestones, mental retardation, with epilepsy a virtually constant finding [167,168,169].

Selection of Candidate Genes *UBE3A* and *GABRB3*

The region most commonly duplicated or deleted in 15q11-q13 is the BP2-BP3 interval (Figure 1-1). I hypothesized that the two best candidate genes in the

interval are *UBE3A and GABRB3.* The former shows maternal-specific expression and the latter is a developmentally critical subunit of the $GABA_A$ receptor. The major reasons for selection of candidates centered on these two genes are as follows: (1) Maternal bias (or maternal-specificity) of dup(15) association with autism *directly implicates UBE3A.* (2) *UBE3A* localizes to the synapse and nucleus and causes alterations in synapse formation (3) Multiple genetic studies have implicated *GABRB3*, in particular, as being associated with non-dup(15) autism. (4) GABA is he major inhibitory neurotransmitter in the brain. (5) Elevated blood plasma levels and reduced numbers of GABA receptors are observed in those with autism. (6) *GABRB3* is developmentally critical and abundantly expressed in the embryonic and neonatal brains. (7) The *Gabrb3* knockout mouse has many features of autism. For all these reasons I hypothesize that *UBE3A* and *GABRB3* harbor common and/or rare alleles that confer increased risk of autism.

In this dissertation, I report on my studies of these genes and genes encoding proteins regulated by *UBE3A* as well as *MECP2* which may regulate gene expression for both *UBE3A* and *GABRB3*. The motivation for studying the substrates of the UBE3A protein are (1) Action of *UBE3A* on these substrates is likely the major means by which AS (and possibly some cases of autism) are caused and (2) The substrates of UBE3A and how this gene causes AS are poorly understood. *MECP2* is (1) known to cause Rett syndrome, (2) expressed at lower levels in autism brains, and (3) correlated with expression of *UBE3A and GABRB3.* The complete list of genes studied includes: *MECP2, (chapter 3);*

UBE3A, ECT2, and GCH1 (chapter 4), and *GABRB3 (chapter 5)*. The model illustrating the relationship between the genes proposed for study is shown below in Figure 1-2.

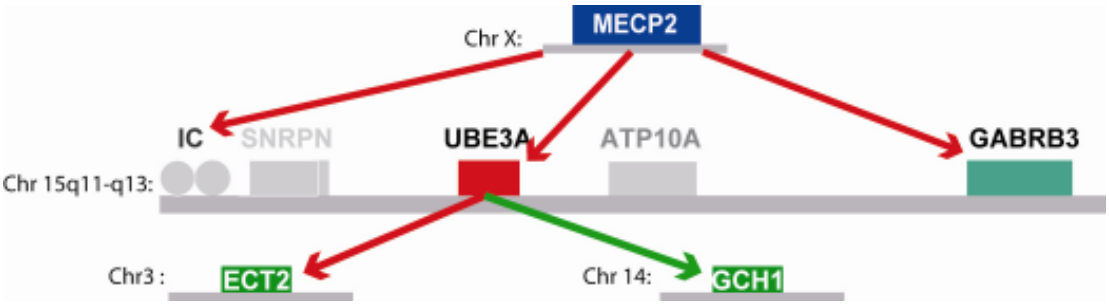


Figure 1-2. Schematic of the relationship between genes proposed for study. Mutations in *MECP2* cause Rett syndrome and studies have suggested that *MECP2* deficiency results in reduced *UBE3A* and *GABRB3* expression. It may do this by acting directly on *UBE3A* and *GABRB3* or at the imprinting center locus, which controls the imprinting status of the genes of 15q11-q13. Maternal deletions of *UBE3A* cause AS, which exhibits some phenotypic similarities to autism. Since *UBE3A* is maternally-expressed, and maternal duplications of 15q11-q13 are associated with autism, *UBE3A* is a strong candidate. *ECT2*, involved in cell motility and migration, is downregulated by *UBE3A*. *GCH1*, which is involved in catecholamine production, is upregulated by increased *UBE3A* and the dysregulation of these proteins by *UBE3A* may lead to synaptic phenotypes such as those seen in autism. *GABRB3* has been observed to be dysregulated in autism, critical in development, and a number of genetic studies have shown it to be associated with ASDs.

UBE3A and related loci *ECT2* and *GCH1*: selection and studies

The UBE3A protein acts as an E3 ubiquitin-ligase which adds a ubiquitin molecule to proteins, tagging them for regulation or, more commonly, degradation by the ubiquitin-proteasome system. The widely-held view is that *UBE3A* will mediate the phenotypic effects seen in AS (and potentially autism) by virtue of its abnormal gene expression and subsequent action on its substrates.

In addition to its role in protein degradation, UBE3A also may function as a transcriptional activator in the nucleus [170]. It is known to stimulate transcription

of steroid hormone receptors in the nucleus among other genes and so dysregulation of its expression could lead more broadly to additional changes in gene expression [171].

In addition, our group previously published evidence supporting association to the imprinted *UBE3A* gene at a peak marker, D15S122, located at the 5′ end of the *UBE3A* gene [115]. A recent autism genome-wide association study has provided strong evidence supporting a role for the ubiquitin pathway in the pathogenesis of ASDs [172].

Additionally, we and other have hypothesized that the effect of loss of *UBE3A* in causing AS and dup(15) autism is a function of UBE3A acting on other loci/proteins. A major limitation of other studies of *UBE3A* and its role in AS and potentially autism, is that no one has studied the protein substrates that are regulated by UBE3A and which may underlie these phenotypes.

My hypothesis (and one generally accepted in the field) is that over-expression of UBE3A in maternal dup(15) cases and corresponding excess protein contributes to the ASD phenotype as a result of altered regulation of UBE3A substrates. Thus, the crux of the hypothesis is that *genes encoding UBE3A substrates and/or genes regulated by transcriptional co-activation of UBE3A are candidate loci for harboring autism susceptibility alleles.*

Our collaborator, Dr. Lawrence Reiter, has used a proteomic strategy to identify two UBE3A-regulated candidates: *ECT2* and *GCH1*, which, following the logic above, I proposed may be good candidates for autism [173][174]. Briefly, their approach involves over-expressing *UBE3A* in *Drosophila melanogaster*,

extracting total protein from heads, separating the protein by 2D gel, and identifying proteins as up- or down-regulated in response to increased levels of UBE3A in experimental vs. wild-type flies(Figure 1-3).

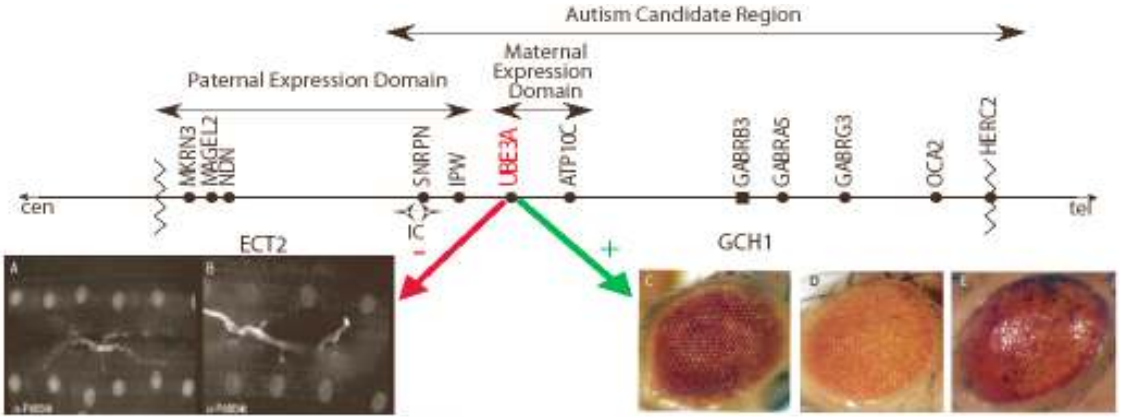


FIGURE 1-3. *UBE3A* and its relationship with *ECT2* and *GCH1*. Since maternal duplication of the region including *UBE3A* puts a carrier at high risk for autism, *UBE3A* and related loci are attractive candidates for investigation. UBE3A downregulates ECT2 and upregulates GCH1. *Drosophila* synaptic boutons are shown stained for *ECT2* ortholog *pbl* in wild-type and UBE3A overexpressing flies in A and B. Notice the loss of signal from synaptic boutons. Mutations in the *GCH1* ortholog, *Punch*, enhance a Dube3A over-expression rough eye phenotype as shown in C-E. The wild type eye in C does not result in a rough eye phenotype while it does when overexpressing the *Drosophila UBE3A* ortholog, D. When expressed with a mutation to the *GCH1* ortholog as in E, offspring have a more dramatically affected rough eye phenotype with fewer bristles.

Dr. Reiter's group previously reported that *pebble* the *Drosophila* ortholog of *ECT2* exhibits extensive down-regulation concurrent with the over-expression of UBE3A [173]. They also showed that loss of Ube3a protein in the brains of *Ube3a* knockout animals caused an increase in Ect2 protein expression and distribution in the hippocampus and cerebellum [173]. Both regions of the brain have been implicated in the pathogenesis of autism and Angelman syndrome [175]. *UBE3A* is imprinted and expressed preferentially from the maternal allele

in central neurons in most regions of the brain with strong evidence for its expression in the cerebellum [176,177]. The cerebellar findings may explain the motor stereotypies common to Angelman Syndrome and autism [153]. There is also evidence that cerebellar defects may explain some emotion recognition and language problems observed in individuals with ASDs [178]. In light of this and other findings on *ECT2*, I hypothesized that the control of levels of *ECT2* by *UBE3A* has a role in synapse formation, as mutations in *ECT2*'s *Drosophila* homolog upset outgrowth in post-mitotic cells as well as GABAergic cell differentiation in *C. elegans* [179,180]. Dysregulation of ubiquitin pathways has also been shown to have consequences for synapse formation [181].

The *Drosophila* protein *punch*, corresponding to human GTP cyclohydrolase 1, or *GCH1*, is a second *UBE3A*-regulated locus and protein that is also currently under investigation. *GCH1* is neuronally expressed and was identified as a significantly upregulated protein in response to *UBE3A* overexpression. Mutations of *Drosophila* ortholog *punch* enhance a rough eye phenotype, a marker of neuronal development, when *Drosophila UBE3A* is overexpressed (Figure 1-3). GCH1 catalyzes the hydrolysis reaction of GTP and H2O, forming a precursor of tetrahydrobiopterin (THB), which is involved in the production of serotonin and other catecholamine neurotransmitters. It is interesting to note that serotonin is often present at increased levels in people with autism [182].

Only modest association results have been demonstrated near this locus known to be critical in Angelman syndrome, and a prime suspect in autism. To

this end, the novel proteomic strategy above has uncovered the first known UBE3A substrates whose dysregulation results in a clear neurological phenotype. In the simplest model of *UBE3A* dysregulation, it would be expected that proteins such as ECT2 will be present at reduced levels and GCH1 at increased levels due to an overexpression of UBE3A. It should be noted that this would only be of importance in cells where dosage regulation is abnormal. Chapter 4 reports on studies of the relationship between *UBE3A*, *ECT2*, and *GCH1*.

GABRB3: Role in Development and Guilt by Association

While the strong maternal bias in dup(15) association with autism make *UBE3A* perhaps the most attractive candidate for contributing to the phenotype, genetic studies within the autism candidate region most strongly implicate the cluster of GABA receptor subunit genes and, in particular, *GABRB3*. A number of reports (though not all positive) have documented association of common alleles at *GABRB3* with autism [104,105,107,183,184,185,186,187,188,189]. Few association studies have implicated the other subunits of the GABA receptor cluster at 15q11-q13. The repeated findings at *GABRB3* suggest this subunit harbors alleles that increase autism risk. McCauley et al. proposed that multiple regions of *GABRB3* could harbor susceptibility alleles, thus complicating consistent detection of association.

The $GABA_A$ receptor subunit genes represent very attractive functional candidate genes. $GABA_A$ receptors are ligand-gated chloride channels that

mediate the majority of fast synaptic inhibition in the brain. Mutations in $GABA_A$ receptor subunits have been linked to generalized epilepsies, common in autism. Neuropathological studies have shown that $GABA_A$ receptor expression is significantly reduced in the brains of children with autism [190,191]. Elevated plasma levels of GABA have been observed in autistic subjects [192]. *GABRB3* is expressed extensively during the late embryonic to early postnatal period of brain development. A deficiency in this subunit during this critical period could have significant effects on synaptogenesis.

The imprinting status of and potential epigenetic dysregulation of*GABRB3*, particularly in autism, is not clear. Studies by Hogart et al. have shown that while expression of *GABRB3* is biallelic in normal brain samples, a subset of autism samples show monoallelic expression or allelic bias, suggesting that epigenetic control of may also extend to this gene [193,194].

Despite the abundance of functional, genetic, and epigenetic evidence implicating *GABRB3,* a thorough indexing of all major alleles and potential rare variants of this gene has not yet been carried out. Chapter 5 covers my analysis of common and rare variants of *GABRB3* in autism.

<u>*MECP2* a Potential Regulator of *UBE3A* and *GABRB3*</u>

Finally, I propose that *MECP2*, encoding the methyl CpG binding protein, responsible for virtually all Rett syndrome (OMIM 312750), shows sufficient evidence to warrant investigation. The major reasons for this are as follows: (1) Mutations in *MECP2* are associated with Rett syndrome and in some cases

autism [195]. (2) Rett syndrome shares a number of phenotypic features with autism including deficits in language, social skills, and repetitive and restricted behavior [196]. (3) *MECP2* expression levels have been shown to be lower in autism (and AS) and to potentially regulate gene expression of *UBE3A* and *GABRB3.*

Since females carry two X-chromosomes, females who carry disease-causing mutations in *MECP2* typically have a normal copy of the gene on the other chromosome. As a result, females often make it to term and present with Rett syndrome. By contrast, males who carry only one X-chromosome are thought to not survive if carrying equal or less pathological mutations.

So similar are the two disorders and so common is the misdiagnosis of autism in young females that girls with a suspected diagnosis of autism are routinely referred for Rett syndrome testing [197]. A number of documented cases of rare polymorphisms are located in the Rettbase database (http://mecp2.chw.edu.au/mecp2/) and indicate that in many cases these patients meet criteria for autism, but not Rett syndrome.

Samaco et al. showed reduced expression of *MECP2* in the brains of 4 of 5 autistic individuals without *MECP2* mutations, suggesting that more modest effects on gene expression in *MECP2* may be a feature of autism and contribute to the presentation of the phenotype [198]. The study by Samaco et al. suggests that *MECP2*, *UBE3A*, and *GABRB3* dysregulation may exist in autism, Angelman, and Rett syndrome and interaction between these genes is borne out by animal models and patient samples [198].

Finally, in the course of the work reported, another group has published evidence supporting an association of common alleles of *MECP2* with autism [199].

Given the above, it is plausible that common allelic forms of *MECP2* may predispose some individuals to the development of autism and may be more broadly involved in the disorder.

Summary

In short, functional relevance, linkage, association, and CNV analysis all support the study of the 15q11-q13 region and specifically the *UBE3A* and *GABRB3* genes. The selection of the Rett syndrome gene, *MECP2*, is supported given the evidence for its regulation of *UBE3A* and *GABRB3*. Finally, study of two of the genes, *ECT2* and *GCH1*, dysregulated by overexpression of the Angelman syndrome gene *UBE3A* may yield new genetic targets important in synapse formation and relevant to the pathogenesis of both Angelman syndrome and autism.

CHAPTER II

HYPOTHESIS AND SPECIFIC AIMS

The purpose of this project was to identify both common and rare alleles in chromosome 15q11-q13 that predispose chromosomally normal individuals to developing autism or autistic traits by investigating selected loci within 15q11-q13 as well as loci that may regulate or be regulated by the products of these genes (see Figure 1-2). To clarify the latter point, I hypothesized that alleles for two recently discovered UBE3A substrates, ECT2 and GCH1, may increase susceptibility to autism and that common alleles of *MECP2* may lead to increased risk for autism.

This dissertation describes the detailed molecular and genetic analyses focused around two candidate genes of the 15q11-q13 region, *UBE3A* and *GABRB3*. In addition, it also sought to study the role of *MECP2*, *ECT2*, and *GCH1* in idiopathic autism.

The genetic analysis included here focus on common variants and their potential role in autism, rare variants with potential function in autism, and an investigation of the region using genotype data and other methods to assess the possibility of smaller genomic deletions or duplications in *UBE3A* and *GABRB3* which could potentially perturb cellular homeostasis and lead to an autism phenotype. In addition, biochemistry and gene expression studies are used to assess the nature of the relationship between *UBE3A* and *ECT2* and *GCH1*.

Hypotheses

Specifically, I hypothesized the following**:**

1. Common and/or rare alleles of *GABRB3* and/or *UBE3A* may contribute to autism susceptibility.

2. Upstream or downstream loci which regulate or are regulated by *GABRB3* and *UBE3A*, namely *MECP2*, *ECT2*, and *GCH1*, may act to increase autism susceptibility.

Toward this end, the following aims were undertaken:

Specific Aim 1: Common Variant Association in Autism

This aim seeks to undertake association studies of SNPs in selected 15q11-13 loci (*GABRB3*, *UBE3A*) as well as upstream (*MECP2*) and downstream (*ECT2* and *GCH1*) interacting loci to identify *common* autism risk alleles that may act alone or in concert to increase disease risk.

Rationale: This approach attempts to capture common variants in chromosomally normal individuals, and to examine potential association of selected common variants and haplotypes with autism or autistic traits. For the loci of interest, it seeks to localize the source and particular alleles that contribute to autism susceptibility.

These results are presented in chapter 3 (*MECP2*), chapter 4 (*UBE3A*, *ECT2*, and *GCH1*), and chapter 5 (*GABRB3*).

Specific Aim 2: Rare Variant Association in Autism

This aim involves sequencing and functional analysis of *GABRB3* and *UBE3A* for potentially functional variants which may predispose individuals to autism.

Rationale: This aim sought to find rare variants that the I hypothesized would act to increase risk for autism or autistic traits. This is done since *GABRB3* point mutations may cause epilepsies and *UBE3A* mutations early exons are sufficient to cause Angelman syndrome with autistic features [153].

I identified and functionally characterized one *GABRB3* variant, P11S , covered in the first part of chapter 5.

Specific Aim 3: **15q11-q13 Small CNVs in Autism**

This aim seeks to evaluate sites within *UBE3A* and *GABRB3* for potential copy number variation and association with disease.

Rationale**:** The 15q11-q13 region has a large number segmental copy number gains, typically resulting in recurrent duplication/deletion events. The clinical features of idic(15) are distinct enough and the level of developmental delay profound enough that it is unlikely many of these individuals exist in our sample. However, dysmorphologies and developmental delay is less striking in the (int dup(15)) patients and some are known to exist in out sample. CNV analysis is expected to uncover interstitial duplications as well as potentially

novel submicroscopic CNVs (microdeletions/microduplications) or association with known CNVs within 15q11-q13 potentially further refining the susceptibility loci within 15q11-q13.

Phenotype and genotype information were used together in chapter 6 to evaluate the possibility of smaller deletions and duplications in *UBE3A* and *GABRB3*.

Chapter 3 covers the studies documenting association of a common allelic form of *MECP2* with idiopathic autism. Chapter 4 explores the association of *UBE3A*, *ECT2*, and *GCH1* loci with autism and relationship between *UBE3A* and its interacting proteins in autism. Chapter 5 is a study of *GABRB3* and association of common alleles in autism and epilepsy as well as a detailed functional study of one of the major variants identified in *GABRB3*. Chapter 6 leverages genotype data and phenotype data in a group of autism samples to determine if small previously undetected CNVs might exist in the exons of *UBE3A* or *GABRB3*. Chapter 7 closes with a summary of findings and suggested future analysis based on this work and the evolution of autism genetics over the course of this project.

CHAPTER III

STUDIES OF *MECP2* IN IDIOPATHIC AUTISM

Introduction

Autism (OMIM 209850) is a neurodevelopmental disorder in which individuals present with deficits in social reciprocity and language, and additionally exhibit features of repetitive behaviors and restricted interests. When narrowly defined, the prevalence of autism is estimated to be approximately 1/500, but as high as 1/150 when all autism spectrum disorders (ASDs) are included [1]. Rett syndrome (OMIM 312750) is a severe neurodevelopmental disorder with features of autism, but which also includes deceleration of head growth, ataxia, and stereotypical hand movements. Autism and Rett syndrome are both classified as pervasive developmental disorders, but considered distinct conditions based on clinical and etiological differences. While the features of Rett often make the disease clinically distinct from autism, a number of overlapping traits are common to both disorders [196].

The genetic architecture of autism is comparatively complex, with strong evidence for both locus and allelic heterogeneity [24,52]. The genetic etiology of Rett syndrome, by contrast, is more straightforward. Approximately 95% of cases are caused by loss-of-function mutations of the X-linked *MECP2* gene, which leads to defective forms or reduced levels of the Methyl CpG binding protein 2(MeCP2) [97,200] .

MeCP2 encodes the Methyl CpG binding protein, which selectively binds to methylated DNA altering gene expression across genomic loci, including the autism associated 15q11-q13 cluster. Depletion of MeCP2 is associated with changes in the histone modification profile to a more active conformation and reduced promoter methylation and thus increased gene expression [201,202]. Other work indicates that MeCP2 has a more complicated affect on gene expression binding to non-methylated promoters or in active regions of expression [203]. MeCP2 may act as a chromatin organizer in genes such as *GABRB3* allowing optimal expression of such genes and absence or reduction of MeCP2 thus leads to misexpression and possibly the phenotypic consequences of Rett syndrome and related disorders [194]. These and a number of other studies point to the importance of the *MECP2* gene in proper brain development and synapse formation [204,205].

The sex difference in for Rett syndrome and autism is notable. In autism, males are affected four times as frequently as females and epidemiology indicates a risk ratio (sibling recurrence risk / population prevalence) of 25-67 [44,45]. All PDDs are expressed at a higher rate in males with the possible exception of Rett syndrome. As such, the potential involvement of X-chromosome loci is tantalizing and though few linkage studies support involvement of the X-chromosome, though there is evidence that mutations of X-chromosome genes (including NLGN3 and NLGN4X) often segregates with autism [62,87,100,206]. Rett syndrome, by contrast to other PDDs, is almost

exclusively present in females and is the second most frequent cause of female mental retardation [195].

Rett syndrome is typically lethal in males, while female mutation carriers may not express features of Rett syndrome due random X-linked inactivation leading to mosaic expression of mutant copies X-linked *MECP2* transcript. The stochastic nature of X-inactivation may result in silencing of the mutant allele thus avoiding the neurological consequences of Rett syndrome. In this manner, female mosaicism is probably often protective in that it ameliorates expression of deleterious alleles for X-linked loci which otherwise might be lethal or present with greater severity [207]. Further, it is thought that the reason fewer males present with Rett syndrome is that sporadic cases probably have their origin in the paternal germline and thus can only be transmitted to females [208]. There are reports of *MECP2* mutations in male patients, but severity is typically greater with patients presenting phenotypes including fatal encephalopathy , non-specific X-linked mental retardation, and autism [209,210,211].

Since the original report of *MECP2* mutations resulting in Rett syndrome, a number of reports have documented *MECP2* variants in cases of autism, atypical Angelman syndrome, and nonspecific mental retardation [55,212,213,214,215,216,217]. While autistic cases harboring *MECP2* mutations are known to exist, they appear to be very rare [212,215,218]. Cases of Rett syndrome previously diagnosed with autism often progress to a less severe Rett phenotype without the typical motor difficulties. This misdiagnosis does indicate that some *MECP2* mutations may be tolerated with fewer consequences than

other mutations [197]. In some cases of *MECP2* mutations, Rett features were not noted until a later age [197]. For this reason, it seems possible that some *MECP2* mutations are incompletely penetrant or may present with a lesser phenotype, such as autism.

Samaco et al. showed reduced expression of *MECP2* in the brains of 4 of 5 autistic individuals without *MECP2* mutations, suggesting that more modest effects on gene expression in *MECP2* may be a feature in some cases of autism [198].

More recent work in genetic association studies has shown that a common allele of *MECP2* is associated with ASDs [219]. Loat. et al. reported on a common haplotype, accounting for more than 80% of haplotypes in their sample, which was overtransmitted to affected individuals in a group of 219 trios [219]. More recent work reported that the less common allele, accounting for about 17% of haplotypes, was associated with decreased cortical surface area in two different samples, one with psychiatric disorders and another with Alzheimer's disease and cognitive impairment. This is very provocative given the evidence for increased head circumference in many cases of autism [220].

Given the above evidence, we thought it plausible that common allelic forms of *MECP2* may predispose some individuals to the development of autism. This chapter describes our effort to determine if common allelic forms of *MECP2* are involved in the etiology of autism.

Subjects and Methods

Sample Description and Phenotype Definition

The study sample consisted of 965 families recruited at Vanderbilt or Tufts-New England Medical Center or obtained from the NIMH Center for Collaborative Genetic Studies on Mental Health Disorders (Table 3-1). Families were excluded if they were found to have a non-idiopathic autism (e.g. fragile-X, dysmorphic features, birth trauma) or gross chromosomal abnormalities. To determine the ancestry of families genotyped, we examined genome-wide SNP data available for the majority of our sample. This data is described in two previous studies [45,221]. Ancestry determinations used the software STRUCTURE available from http://pritch.bsd.uchicago.edu/structure.html in combination with HapMap SNP genotype data for known ancestral groups to correctly classify founders (Table 3-1). In cases without genotyping data, self-report was used. Remaining families were classified as unknown. Autism diagnosis was assigned using the Autism Diagnostic Interview (ADI or ADI-R), Autism Diagnostic Observation Schedule (ADOS) or both [25,26,27,28].

Table 3-1. Sample description by ancestry.

	ALL	AA	ASIAN	CAUC	HISP	UNK
Families	965	31	26	796	39	104
Individuals	4032	116	133	3308	169	415
Number with autism	1624	49	42	1318	69	158
Female	306	8	9	246	20	28
Male	1318	41	46	1072	49	130
Type of Families:						
Simplex	301	7	1	253	13	27
Multiplex	664	24	25	543	26	77

Stratification of Families

Families were further stratified into two diagnostic categories, spectrum or strict, based on ADI and/or ADOS measures according to Table 3-2. These classifications are defined as indicated in Table 3-2. NA indicates the instrument was not available or not administered.

Table 3-2. Phenotypic categories as defined by ADI-R and ADOS.

Individual category	Phenotype classifications:	
	ADI-R	ADOS
(1) Strict	Autism	Autism
(2) Spectrum	Autism	NA
	ASD	ASD
	NA	Autism

Tag SNP Selection

SNPs were selected using the HapMap release #22, build 35 Caucasian genotypes and inter-marker linkage disequilibrium (LD) was measured in the computer package Haploview [222]. The Haploview implementation of Tagger facilitates selection of tag SNPs. To capture common alleles, we used settings corresponding to a pairwise of $r^2 > 0.8$ and MAF ≥5% were used. Potential tag SNPs are selected and analyzed using a BLAT query to control for interfering SNPs and repetitive elements. Due to high levels of LD in the region, only two tag SNPs were required to capture the common alleles of the 75.9kb *MECP2* gene using these criteria. However, to err on the side of caution and include

SNPs flanking exons, 5 SNPs were selected. These SNPs allowed capture of all (Caucasian) alleles with a mean max r^2 of 0.98. As Figure 3-1 shows, the 5 SNPs selected spanned 100.3kb encompassing *MECP2* and flanking sequence between rs4898375 and rs5945397.

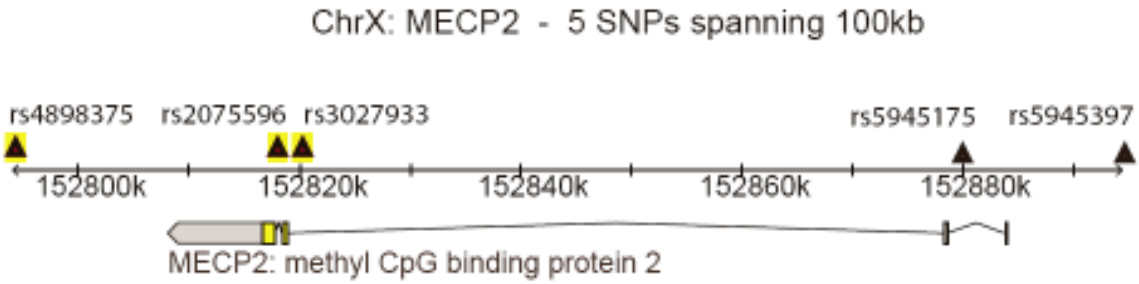


Figure 3-1. *MECP2* Gene Structure and Tag SNP positions.

The *MECP2* gene spans 75.9kb and the tag SNPs 100.3kb on chromosome X. SNPs genotyped are identified with triangles and the three significant SNPs are highlighted. All five SNPs appear to be in a single block of linkage disequilibrium with a high level of r-squared between the three significant SNPs located near the terminal (3′) exons.

Genotyping

Markers meeting criteria were ordered as Assays-on-Demand (AoDs) from Applied Biosystems (Foster City, CA, USA). Genotyping was carried out by ABI TaqMan reactions were performed in a 5-µl volume according to the manufacturer's recommendations using 5ng of genomic DNA. Cycling conditions included an initial denaturation at 95°C for 7 min, followed by 50 cycles of 92°C for 15 s and 60°C for 1 min. Samples were analyzed using an ABI 7900HT

Sequence Detection System. Genotypes were automatically called if there was 95% confidence by the SDS software calling algorithm that the call was correct.

Quality Control

Inter- and intra-plate genotype controls were included in each 96-well plate provide for quality control measure. The Pedcheck script was used to check for Mendelian inconsistencies and a modification of this script was used to check for heterozygous males for the X-chromosome locus [223]. Five families were excluded from analysis due to irresolvable Mendelian inconsistencies or because genotyping showed heterozygous males. Genotyping efficiencies for all five of the markers genotyped were between 98 and 99 percent. Hardy-Weinberg equilibrium (HWE) P-values were calculated for each SNP using Haploview [222]. Since appreciable differences in allele frequencies were found to exist between HapMap populations, HWE values were checked in the overall sample and within ancestral groups. Without doing so large deviations from HWE may exist as an artifact.

LD and Family-based Association Analysis

Haploview was used to calculate pair-wise LD (r^2) for each pair of SNPs in our dataset [222]. Analyses of allelic transmission were performed in both spectrum and strict using the transmission disequilibrium test (TDT) as implemented in PLINK , association was tested in all families and in the spectrum and strict subsets [224,225]. In addition, I tested for overtransmission to affected

males and females, and the four major ancestral groups which compose the sample: African American, Asian, Caucasian, and Hispanic groups.

Factor Score Analysis in Males and Females

Factor scores based on ADI-R inputs have values from 0-1 and index separate components of the autism phenotype including: spoken language, social intent, compulsions, developmental milestones, savant skills and sensory aversions. These scores are calculated by principal components analysis as described by an earlier publication [226]. Males and females, respectively hemizygous or homozygous for the major haplotypes, were tested by parametric methods (t-test and ANOVA) for factor scores with normal distributions and by nonparametric methods for factor scores with non-normal distributions to see if a difference in means was observed between genotype groups.

Results

Hardy-Weinberg Equilibrium

While violation of HWE was present, as expected, in the overall sample, no violation of HWE appeared present within ancestral groups. Minor allele frequencies and Hardy-Weinberg P-value for the overall sample and each ancestral are shown in Table 3-3. This was sufficient to provide deviation from HWE in accordance with Wahlund's principle [227], which indicates that an abundance of homozygotes will be observed in situations of underlying subpopulations.

Table 3-3. Minor allele frequencies and Hardy Weinberg Equilibrium P-values in overall sample and ancestral groups.

			All		African American		Asian		Caucasian		Hispanic	
SNP	Name	Position	MAF	HWpval	MAF	HWpval	MAF	HWpval	MAF	HWpval	MAF	HWpval
1	RS4898375	152926420	0.199	0.002	0.052	1	0.192	1	0.158	0.5437	0.383	0.9631
2	RS2075596	152950586	0.191	6.0E-04	0.065	1	0.22	0.9324	0.149	0.6085	0.358	0.5276
3	RS3027933	152952068	0.198	0.0034	0.169	1	0.228	0.9841	0.154	0.628	0.36	0.5276
4	RS5945175	153011951	0.044	1	0	1	0	1	0.049	1	0.009	1
5	RS5945397	153026720	0.055	1	0	1	0	1	0.061	1	0.037	1

LD Analysis

LD analysis of the region showed a pattern of LD in our sample largely consistent with HapMap with all 5 of the genotyped SNPs existing in a single block of LD with high levels of D´ between all SNPs tested. Further, we tested to see if LD patterns were consistent between ancestral groups. Figure 3-2 shows LD patterns in the four ancestral groups. For each ancestral group, with the exception of the African American, r^2 levels between SNPs 1-3 were greater than 0.8. In African American and Asian groups, SNPs four and five were not polymorphic. In all ancestral groups, D´ values were estimated to be equal to 1) for all polymorphic SNPs suggesting no recombination of haplotypes within the markers genotyped.

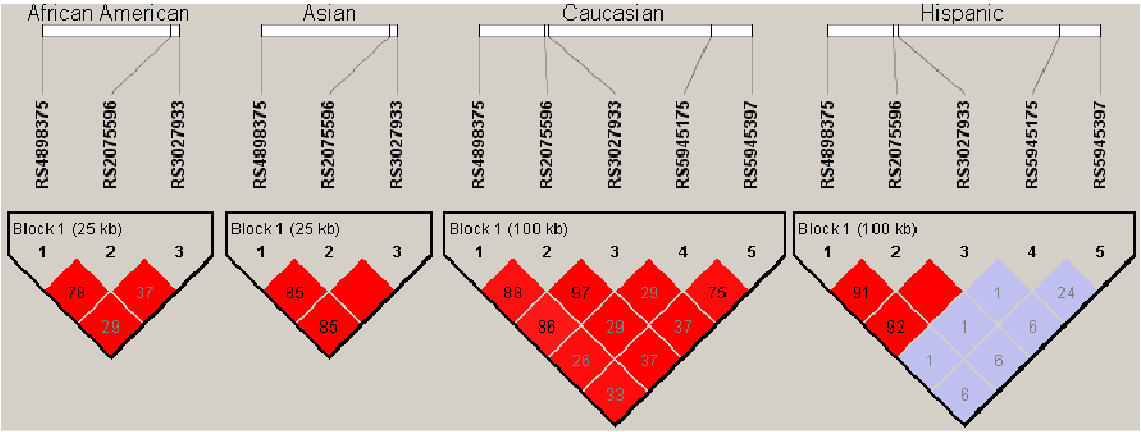


Figure 3-2. Linkage Disequilibrium Patterns in the Four Major Ancestral Groups. LD patterns are shown for African American, Asian, Caucasian, and Hispanic families. All five genotyped SNPs lie within one block of LD as measured by D´=1 between all polymorphic SNPs in all ancestral groups. Levels of r^2 were high between the first three SNPs in most groups. In the figure above, if r^2=1, the blocks are red and there is no numeric value given in the square. If r^2 is less than 1, and confidence in D´ is high, (as measured by a LOD>2), red squares will give their value of r^2 (times 100). Purple squares indicate D´ of 1, but lower confidence in the value of D´ (LOD<2), frequently due to smaller sample size and/or low frequency of some alleles within different ancestral groups. Only three SNPs are shown for African American and Asian groups as SNPs 4 and 5 are not polymorphic in these groups.

Single Marker Association Results

Single marker analysis testing for over-transmission of alleles to affected individuals was done using the TDT implementation in PLINK. The results are shown in Table 3-4. The five SNPs genotyped in *MECP2* are listed with their genomic positions and relative position in the gene. A1 indicates the minor allele and A2 the major allele. Transmission counts are given for the minor allele (T=Transmitted, U=Untransmitted). P-values, TDT odds ratios (OR) with 95% confidence intervals, and TDT chi-square statistics are given only *for significant SNPs*. In this case it appears the major allele correlates with increased risk.

Table 3-4. Single marker association results for the entire 965 family dataset.

SNP	Name	A1	A2	T	U	CHISQ	OR	P
1	RS4898375	A	G	162	222	9.375	0.73 (0.6-0.89)	0.0022
2	RS2075596	A	G	162	209	5.954	0.78 (0.63-0.95)	0.0147
3	RS3027933	G	C	162	217	7.982	0.75 (0.61-0.92)	0.0047
4	RS5945175	C	T	57	58			
5	RS5945397	A	G	70	76			

Since results for SNPs 1-3 produced significant evidence in support of association, we wanted to examine the potential effect of disease classification (see Table 3-2) on association for these SNPs. The results of our analysis of SNPs 1-3 in considering transmissions to individuals affected under the strict model are shown in Table 3-5. SNPs 1-3 were very significant, surviving Bonferroni correction. Families with probands that met strict criteria for autism (see Table 3-2) showed the greatest support for association. Therefore, this strict group, contains probands that meet criteria for classical autism by both ADI and ADOS-measures and does not appear to indicate that the association is driven by Asperger syndrome or PDD-NOS.

Table 3-5. Single marker association results under the strict phenotype model.

SNP	Name	A1	A2	T	U	CHISQ	OR	P
1	RS4898375	A	G	88	144	13.52	0.61 (0.47-0.8)	0.0002
2	RS2075596	A	G	87	132	9.247	0.66 (0.5-0.86)	0.0024
3	RS3027933	G	C	91	140	10.39	0.65 (0.5-0.85)	0.0013

Since *MECP2* is X-linked and might be expected to have differential impact depending on whether it was transmitted to males or females, we sought to examine transmissions to only affected males and to females independently.

Analysis of transmissions to-males-only (TMO) and to-females-only (TFO) are shown for these three SNPs in the strict disease classification in Table 3-6. While transmissions TMO was significant for all SNPs and TFO for only SNP 3, the direction of transmission bias and estimated OR were comparable suggesting that this is most likely the impact of a smaller number of transmissions to consider for affected females since the ratio of males to females in our sample is nearly 4:1.

Table 3-6.Transmission to Affected Males and Females under strict model.

				TMO					TFO				
SNP	Name	A1	A2	T	U	CHISQ	OR	P	T	U	CHISQ	OR	P
1	RS4898375	A	G	70	114	10.52	0.61 (0.46-0.83)	0.0012	18	30	3.00	0.60 (0.33-1.08	0.0833
2	RS2075596	A	G	71	107	7.28	0.66 (0.49-0.9)	0.0070	16	25	1.98	0.64 (0.34-1.20	0.1599
3	RS3027933	G	C	76	112	6.89	0.68 (0.51-0.91)	0.0087	15	28	3.93	0.54 (0.29-1.00	0.0474

Multimarker Haplotype Analysis

The PLINK program allows haplotype-based TDT association tests. The only major haplotypes with frequencies greater than 0.05, h1 (GGCTG) and h2 (AAGTG), were tested. As table 3-7 indicates, both showed the major haplotype significantly overtransmitted to affected individuals in each of the disease strata.

Table 3-7. Haplotype-based TDT association test.

Model	Halotype	Afreq	T	U	CHISQ	P
SPEC	h1	0.79	230.70	171.40	8.74	0.00311
	h2	0.13	90.02	145.40	13.02	0.00031
STRICT	h1	0.79	148.60	94.47	12.06	0.00002
	h2	0.14	47.02	97.77	17.79	0.00051

Factor Scores in Associated and Unassociated Males and Females

Spoken Language, Social Intent, Developmental Milestones, Savant Skills, Rigid-Compulsive Behavior, and Sensory Aversions factor scores were calculated based on ADI-R items. Depending on the normality of the data distribution, either ANOVA or the Kruskall-Wallis test to determine if those homozygous (or hemizygous in the case of males) for the risk haplotype were significantly different to those with the protective haplotype. The tests were carried out controlling for sex and ancestry. Ancestry and sex were predictive of Spoken Language score, but genotype was not. Sex, but not ancestry or genotype, was significantly associated with Social Intent scores, with males having scores indicating less severe dysfunction in social domains. Milestones did not appear to be impacted. Rigid-compulsive behavior, as measured by the ADI, was significantly impacted by sex, but not by ancestry or genotype.

Only the Savant Skills factor score appeared impacted by genotype. Using Kruskall-Wallis, savant skills was significant by both ancestry and genotype group, with those bearing the risk variant having lower scores on measures of Savant Skills. This indicates that these individuals are less likely to show high function in savant domains which include visuospatial ability, computational

ability, and memory skill ($p<0.0001$) [226]. To the extent that genotype group is accounted for by ancestry, we repeated the test within ancestral groups using the non-parametric two sample Wilcoxon rank sum test to determine if genotype was significantly associated with Savant Scores within each group. We found that for within Asian ($p=0.038$), Caucasian ($p=0.002$), and Hispanic ($p=0.016$) groups that genotype was associated with savant score, but not in African American group. For those three groups, the savant score was significantly lower with the associated haplotype (h1) than for the unassociated haplotype (h2) $p=0.0009$.

Discussion

Mutations in *MECP2* have been observed in subjects clinically diagnosed with Rett syndrome, autism, and atypical Angelman syndrome among other conditions [55,212,213,214,215,216,217]. However, mutations in *MECP2* are also present in phenotypically normal female individuals which may be due to the mosaic nature of X-inactivation in females or preferential inactivation of mutant alleles [228,229]. By extension, we have hypothesized that common alleles may contribute to the risk profile in the development of idiopathic autism. The mechanism I propose is minor differences in *MECP2* expression levels as a function of the major haplotypes which thereby impact other gene and protein networks though to lesser extent than in cases of Rett syndrome.

One other report by Loat et al. has also shown common alleles of *MECP2* to be associated with autism [219]. Preliminary evidence in this large sample of families with autism is a replication of this earlier finding. SNPs genotyped and

shown associated are in high LD as indicated by levels of r^2 indicating that the same effect is being observed. The same risk haplotype (though different SNPs were genotyped) appears to be overtransmitted in both studies. Incidentally, SNPs 1-3 also show strong r^2 with SNP rs2239464, the rare allele that has been shown associated with reduced cortical surface area [230]. This is notable since the common allele is in strong r^2 with our SNPs 1-3 and macrocephally is a common finding in a number of autism cases [220]. In this case, the major allele is overtransmitted to those with autism in our study, and in the Joyner et al. study the minor allele is associated with reduced cortical surface area.

Given these findings, it is possible that *MECP2* may be a QTL for brain size and potentially IQ as well as a risk variant for autism. This is supported by the finding of the transmission bias of the major allele to those with classical autism and the minor allele of their SNP (which is on the less common unassociated haplotype, h2) with reduced cortical surface area. Impact on IQ might be reflected in part in the Savant Skills factor score findings where those with the associated h1 haplotype were less likely to have lower savant skills, reflecting lower ability in visuospatial ability, computational ability, memory skill, and musical ability. However the absence of finding of impact on developmental milestones suggests that the effect might be more subtle, since in the original factor score analysis this was the only factor score associated with IQ [226].

The association was driven almost completely by those with strict affection status. This suggests that those carrying the variant are more likely to manifest with classical autism rather than Asperger syndrome or PDD. This was, in part,

the motivation for the factor score analysis. Since fewer transmissions are considered in “strict” families, (as fewer individuals meet this criteria,) it is surprising that both single-marker and haplotype-based TDT tests identified the strict disease category as more significantly impacted. TDT based tests are not only valid in the face of population heterogeneity, but can gain power [231].

Previously a significant reduction in MeCP2 expression in frontal cortex has been shown for cases of Rett syndrome, autism, and Angelman syndrome compared to age matched control autopsy specimens [202,232]. A study by Samaco et al. suggests that a *MECP2-UBE3A-GABRB3* axis of dysregulation may exist in autism, Angelman, and Rett syndrome and this is supported by animal models and subject samples [198]. It has been suggested that patients clinically diagnosed with Angelman syndrome, but without identifiable 15q11-q13 molecular be screened for *MECP2* mutations as such mutations have been observed in these cases [233,234]. Brain architecture in MeCP2 null mice shows a slight decrease in neuron size and an increase in packing density in the hippocampus, cerebral cortex, and cerebellum [235].

It has previously been established that *MECP2* is subject to X-inactivation and that not all mutations are equally damaging [236,237,238]. This may explain the variability of the phenotype observed in those mutation carriers with clinical diagnoses as diverse as encephalopathy, Rett syndrome, autism, and Angelman as well as phenotypically normal carriers. There does appear to be a trend or potential mechanism whereby the mutant *MECP2* alleles are selectively inactivated [239,240]. The MeCP2 protein has been shown to bind to and repress

or activate transcription by binding tightly to methylated DNA sequence which it does so at many sites in the genome [241].

Localization of the MECP2 protein has been shown to be dependent on DNA methylation where MECP2 and others can be recruited to neurologically important genes such as *FMR1* and *BDNF* and cause their repression [242,243,244]. In the case of *BDNF*, it has been shown that neuronal membrane depolarization and calcium increases may lead to phosphorylation and release of MeCP2 from the *BDNF* promoter [244]. In this manner, MeCP2 is involved in activity-dependent gene regulation [245].

During development Mecp2 levels are high in adult mouse brain though there is not a strong correlation between protein levels and RNA levels, suggesting translation may be regulated at a post-transcriptional level and its levels may correlate with neurological maturity [246]. Restoration of expression of a functional copy of *MECP2* has been shown to completely rescue the phenotype in mice, while overexpression can be lethal [247,248].

There are two major splice isoforms for the *MECP2* transcript (*MECP2*A and the more abundant *MECP2*B) with exon 2 occasionally excluded so the two isoforms differ only in their N-terminus[249,250]. *MECP2*B is more abundant in ES cells and *MECP2*A increases as differentiation proceeds [249]. It remains unknown whether or not the two different isoforms have different function. If h1 and h2 haplotypes are associated with differential isoform expression, this might be another means by which the association could be explained.

MeCP2-mediated modulation of local chromatin structure might paradoxically facilitate rather than suppress expression of *UBE3A* or *GABRB3*, so that null alleles cause concordant reduction in these transcripts [198].

Cheadle et al. found that most mutations in Rett syndrome patients were in the methyl-CpG-binding domain or the transcription repression domain of MeCP2 [251]. The group also characterized nine recurrent mutations in 33 unrelated cases (73% of all cases with *MECP2* mutations). Milder disease was a characteristic of patients with missense mutations compared to those with nonsense mutations. Disease was also milder the later the truncating mutation. Bienvenu found 30 mutations in 46 RTT patients including 12 novel mutations most of which were in exon 3 [252]. R270Xand frameshift deletions in a (CCACC)n-rich region were found with multiple recurrences [252]. Hupke et al. found mutations in 24/31 Rett patients with 20/24 being *de novo.* Most mutations were truncating and some females having the same mutation manifesting with varying phenotypes, suggesting other factors may influence phenotype [253]. Amano et al. identified 12 different mutations in the *MECP2* gene, 8 of which were novel [254]. De Bona et al. noted 2 hotspots in Rett patients R270X and R294X [255].

My data represents a second report of association of common alleles at *MECP2* with autism probands. We have shown that this association is strongest in families with probands that meet strict criteria for autism, and that it is the major allele of *MECP2* which is overtransmitted to those with autism. Given the vast number of genes with transcription impacted by *MECP2* and the relative

frequency of the two major haplotypes, determining the molecular action by which risk alleles of *MECP2* impact gene expression leading to autism is a tractable, though complex, problem.

CHAPTER IV

ASSOCIATION, GENE AND PROTEIN EXPRESSION STUDIES OF TWO *UBE3A* NETWORK GENES IN AUTISM

Introduction

The most common chromosomal abnormality observed in ASD is maternal duplication of chromosome 15q11-q13 which accounts by some estimates for as many as 3% of all autism cases. This duplication region contains a number of genes including *UBE3A*, maternal deficiency of which results in the mental retardation disorder Angelman syndrome (AS) [116,256]. The relatively high frequency of maternal 15q duplications resulting in an ASD phenotype implicate the genes that exhibit maternal-specific expression in this region (*ATP10* and *UBE3A*) as the most likely contributors to the ASD phenotype in individuals with 15q duplications.

UBE3A is subject to epigenetic control and shows maternal-biased (maternal > paternal) and maternal-specific expression in some brain regions. I hypothesize that alleles at *UBE3A* that act to alter gene expression, imprinting regulation, or protein function could increase risk for autism. *UBE3A* encodes the E6-AP ubiquitin protein ligase, an E3 ubiquitin ligase. I hypothesized that dup(15) and AS phenotypes result in part or entirely as a result of *UBE3A* dysregulation. As such, both AS and autism may could result from the action of too much or too little *UBE3A* and its protein on target protein substrates or by defects directly in

those genes and proteins regulated by, or known to regulate *UBE3A*, (e.g. *MECP2*). Both of these phenotypes may be a direct result of the dysregulation of a number of *UBE3A* protein substrates or transcriptionally regulated genes. I hypothesized that changes in gene expression or protein stability of *UBE3A* targets may also result in an autism phenotype on their own. Dr. Lawrence Reiter, identified two proteins that are affected by increased *UBE3A* expression: epithelial cell transforming sequence 2 oncogene (*ECT2*), a Rho-GTPase involved in actin cytoskeletal remodeling and axon guidance and GTP cyclohydrolase I (*GCH1*), the rate limiting enzyme in the synthesis of a variety of neurotransmitters [257,258].

As an E3 ubiquitin ligase, the primary function of *UBE3A* is to tag proteins with ubiquitin for subsequent degradation or cellular re-localization [259]. *UBE3A* has also been shown to be a potent transcriptional activator for estrogen and androgen receptors among others [170]. In addition, it may act indirectly to increase or decrease levels of proteins by ubiquitinating and thus degrading transcriptional regulatory proteins for these genes. Reiter and colleagues showed that *ECT2*, which remodels the actin cytoskeleton physically interacts with Ube3a in cultured cells and is down-regulated by increased expression of Ube3a in *Drosophila* [173]. Likewise, loss of Ube3a protein in the brains of *Ube3a* null mice has a profound effect on *ECT2* protein expression in both the hippocampus and cerebellum, two regions of the brain implicated in the pathogenesis of both AS and autism [175]. *ECT2* is important in the creation of the cleavage furrow during cytokinesis [260]. However, it also plays a critical role later in development where

it is involved in the migration of neuronal P-cells, the progenitors of GABAergic interneurons in *C. elegans* [180].

Another protein identified in the screen for *UBE3A* regulated proteins/genes is GTP cyclohydrolase 1 (*GCH1*), which appears to be up-regulated by the over-expression of *UBE3A* in both fly and human (manuscript in preparation). Mutations (typically loss-of-function) in *GCH1* cause dopa-responsive dystonia and Parkinson's disease since this protein is the rate limiting enzyme in tetrahydrobiopterin (BH4) biosynthesis [261,262]. With *ECT2* involved in synaptic plasticity and *GCH1* in monoamine synthesis, both are excellent candidate genes for autism since they regulate synaptic stability and function.

Here I show that in addition to *UBE3A*, both *ECT2* and *GCH1* genes are strong candidates for harboring susceptibility alleles for idiopathic autism as anticipated above. To test this hypothesis, I conducted an association study in 694 combined autism families using tag SNPs to index all major alleles for these three genes: *ECT2*, *GCH1*, and *UBE3A*. In addition, I analyzed gene expression and protein levels in lymphoblastoid cell lines from patients with both AS and dup(15q) autism patients to determine if the levels of *ECT2* and *GCH1* were correlated with *UBE3A* dosage. I then used lymphoblastoid cell lines from patients that carry associated alleles and unassociated alleles for *ECT2* and *GCH1* in our analysis to test for differences in gene expression as a result of the specific SNP haplotypes. To test the response of these genes to changes in *UBE3A* protein expression, I transfected *UBE3A* and *UBE3A*-siRNA constructs into mammalian cells to evaluate changes in *ECT2* and *GCH1* expression and

stability. Finally, I attempted to validate our findings by examining gene expression for the HapMap CEPH trio offspring. Our findings support the need for further study of *UBE3A* and associated loci *ECT2* and *GCH1* as candidates for involvement in the pathogenesis of autism and AS.

Subjects and Methods

Subjects

The cohort used for the association studies contained 694 families with one or more ASD affected offspring. Families were ascertained through 5 different centers shown in Table 4-1. These included families recruited at Vanderbilt (VAN) or Tufts-New England Medical Center (TUF). Samples for the Autism Genetic Resource Exchange (AGRE), Iowa, and Stanford samples were obtained from the NIMH Center for Collaborative Genetic Studies on Mental Health Disorders. Families were excluded if they were found to have a non-idiopathic autism (e.g. fragile-X, dysmorphic features, birth trauma) or gross chromosomal abnormalities. Autism was assessed using the Autism Diagnostic Interview (ADI or ADI-R), Autism Diagnostic Observation Schedule (ADOS) or both. For the purposes of this study, those who met criteria for autism or ASD by either or both of these measures were considered “affected” in the association study. In terms of ethnicity, 4% percent of the families were of African-American origin, 7% of Latino origin, 4% of Asian origin, 83% of Caucasian origin, and 2% of unknown ethnic background. 78% of those with autism in our cohort were

male. Several individuals for whom cell lines were available were chosen for additional gene and protein expression studies.

Table 4-1 – Characteristics of genotyped families with autism stratified by center

	ALL	AGR	IOW	STA	TUF	VAN
Families	694	327	85	131	98	53
Individuals	2823	1420	361	522	353	167
Number with autism	1256	623	162	259	153	59
Female	268	147	26	52	34	9
Male	988	476	136	207	119	50
Family type						
Simplex	143	36	19	16	27	45
Multiplex	551	291	66	115	71	8
ADI version		ADI / ADI-R	ADI	ADI*	ADI	ADI / ADI-R

*Short form ADI

Genotyping and Association Analysis

The *ECT2*, *GCH1*, and *UBE3A* genes span 67kb, 61kb, and 102kb, respectively. *ECT2, GCH1*, and *UBE3A* contain 24, 7, and 17 exons respectively. HapMap data and Haploview's implementation of the Tagger application, was used to select tag SNPs with minor allele frequencies (MAFs) greater than 0.05 (http://www.broad.mit.edu/mpg/haploview/). Given the predominantly Caucasian makeup of our sample, 30 CEPH parent-offspring trios were used for the estimation of haplotype blocks and tag SNPs were selected so that each had a minimum pairwise r^2 of 0.8 with other HapMap genotyped SNPs with MAF>0.05.

Genotyping was done using TaqMan-MGB probes designed by the Assay-on-Demand service of Applied Biosystems (Applied Biosystems, Foster City, CA,

USA). All PCR was carried out in 384-well plates in a 5µl volume containing 2.5µl 2x TaqMan Universal PCR Mastermix with 0.125µl 20x Assay-on-Demand probe and 5ng of genomic DNA according to manufacturer's recommendations. Post-PCR scanning was done on the ABI 7900HT apparatus. Genotyping quality control measures included 95% genotyping efficiency for all SNPs, test of Hardy-Weinberg Equilibrium, and inter-plate and intra-plate QC replicate samples, as well as checks for within-family Mendelian inconsistency using Pedcheck [223].

The Family-based Association Test (FBAT) was used to test the transmission bias of alleles from heterozygous parents to affected offspring [263]. Empirical analysis of LD in our cohort was carried out on genotype data using the Haploview program [222]. The HBAT implementation of the FBAT program was used to measure overtransmission of haplotypes from parents to affected offspring [264]. I examined the results from these analyses to identify associated and unassociated alleles and haplotypes of *ECT2* and *GCH1* for selection of patient cells for gene expression analysis described below.

Cell culture

Lymphoblastoid cell lines (LCLs) for subjects diagnosed with idiopathic autism were previously made from Epstein Barr virus-immortalized patient lymphocytes. Samples for affected subjects identified as homozygous for *ECT2* and *GCH1* associated or unassociated alleles were selected from these cell lines to determine if gene expression changes existed between different allelic forms of *ECT2* and *GCH1*. Lymphoblastoid cell lines for AS, CEPH, and maternal

dup(15q) samples were acquired from the Coriell Cell Repository (http://ccr.coriell.org/). AS lines were all known deletions of the 15q11-q13 interval and dup(15q) cell lines were known to be isodicentric (idic) duplications, containing four copies of the 15q11-q13 region. Cell cultures were maintained in RPMI-1640 with 15% BCS, 5% L-glutamine, and 5% pen-strep, at a density of 200,000-500,000 cells/ml at 37° C with 5% CO_2.

RNA isolation

RNA was isolated from $5x10^5$ cells in 1ml of Trizol (Invitrogen, Carlsbad, CA, USA) according to manufacturer instructions for using suspended cells. RNA was quantified by measuring absorbance at 260nm with an ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Concentrations were adjusted to 200ng/µl for preparation of cDNA.

Preparation of cDNA

cDNA was prepared using Applied Biosystems TaqMan Reverse Transcription Reagents (cat #N808-0234, Applied Biosystems, Foster City, CA, USA). Reagent volumes were adjusted from manufacturer specifications appropriate for 10µl samples each with 400ng of RNA added. One duplicate sample and one blank (no RNA) were used as controls.

Expression analysis by Quantitative Real-time PCR

From each sample, I diluted cDNA 1:50 and then transferred triplicate 2µl samples to 384-well plates along with 2.5µl Applied Biosystems Universal Mastermix (cat no. 4305719) and 0.25µl VIC-conjugated RNase P (cat no. 4316844), and 0.25µl FAM-conjugated *ECT2*, *GCH1*, or *UBE3A* real time PCR assays (cat # Hs00978168_m1, Hs00609198_m1, and Hs00963668_g1, respectively). Gene expression was assessed in triplicate for each sample in 5µl reactions normalizing the target genes to RNase P expression and using the $2^{-\Delta\Delta Ct}$ method with a relative quantification $(RQ)_{min}/RQ_{max}$ confidence set at 95% [265]. Amplification efficiency for each gene product was calculated by serial dilutions for each assay.

Transfection with *UBE3A* and siRNA constructs

Since expression profiles might be cell-type dependent, I used both immortalized CEPH lymphoblastoid cell lines and HEK-293T cells for transfection of *UBE3A* and *UBE3A*-siRNA constructs. A pcDNA3 vector (Invitrogen) containing the *UBE3A* cDNA was used for transfection of *UBE3A* into cells. A substitution in the same vector at nucleotide 833 (C to A) formed a catalytically inactive form of *UBE3A* which cannot tag substrates, including itself, for degradation was also used for transfection [266]. In addition, empty pcDNA3 vector was used. An siRNA construct against *UBE3A* was used (cat # AM16706, Ambion, Austin, TX) and transfected using Oligofectamine (Invitrogen). All other

constructs named above were transfected using Fugene (Invitrogen), in accordance with the manufacturer's recommendations.

Protein extraction, antibodies and Western blot

To harvest protein, I collected 5mL of cells with a density of ~$5x10^5$ cells/mL. Cells were washed with PBS, then combined with 200µl RIPA buffer (50mM Tris-HCL, pH 7.4, 150mM NaCl, 1% NP-40, 0.25% Na-Deoxycholate, 1mM EDTA) per sample. Cells were lysed by shaking the tubes at 4°C for 30 min, before collecting the lysates into 1.5ml Eppendorf tubes. Tubes were rotated for another 20 min, centrifuged at maximum speed for 5 min and the supernatant taken. Protein concentrations were determined using a Bicinchoninic Acid (BCA) assay measuring samples in triplicate with a BMG Labtech, PolarStar Optima and Omega software (FLUOstar 403; BMG LabTechnologies, Durham, NC).

Primary antibodies used for Western blot analysis were α-*ECT2* (cat #sc-25637, Santa Cruz, CA); a chicken polyclonal α-*GCH1* generated against a mixture of two C-terminal peptide fragments (Nterm-CFSRRLQVQERLTK-Cterm and Nterm-HDLELDHKPPTREC-Cterm); α-*UBE3A* (cat #611416 BD Biosciences, San Jose, CA); and for a loading control, α-GAPDH was used (cat 2-RGM2, Advanced ImmunoChemical Inc, Long Beach, CA). Secondary antibodies were as follows: Goat Anti-Mouse IgG, Donkey Anti-Chicken IgG, and Goat Anti-Rabbit IgG (cat #115-035-062, cat #703-035-155, and cat #111-035-003, Jackson Immuno Research, West Grove, PA).

A molecular weight standard, SeeBlue Plus2, pre-stained molecular mass standard (Invitrogen) was loaded in adjacent lanes for protein size determination. After adding 4X SDS sample buffer, samples containing equal protein concentration were heated for 5 min at 100 °C and resolved on a 10% Tris-HCl denaturing ready-gel (Bio-Rad, Hercules, CA) for detection of *UBE3A*, *ECT2*, *GCH1*, and GAPDH. Proteins were transferred to Amersham nitrocellulose membrane at 4°C for 1 hour at 0.20uA, blocked with Tris Buffered Saline Tween (TBST) + 5% milk, followed by incubation with primary antibodies at 4°C overnight. The following day membranes were washed with TBST followed by a 1 hour treatment with secondary antibodies, and additional washes with TBST. Western blots were developed with ECL reagents (Detection reagent 1 and 2, Product #1859701 and #1859698, Thermo Scientific, Rockford, IL). Chemiluminescence was recorded on an Omega 12iC Molecular Imaging System and was quantified using UltraQuant software (UltraLum, Claremont, CA, USA).

Analysis of HapMap genotype and gene expression

Genotypes were downloaded from the HapMap's Hapmart website (http://hapmart.hapmap.org/BioMart/martview) for all SNPs shown in table 4-2 for the CEPH trio offspring. Gene expression data for these samples was downloaded from GENEVAR (http://www.sanger.ac.uk/humgen/genevar/) and the data was imported into Stata (version 9.0; Stata Corp, College Station, Tex)

for analysis of differences in gene expression differences in *UBE3A*, *ECT2*, and *GCH1* genes by genotype and expression level of each gene.

Statistical Analysis

Comparison of gene and protein expression levels was performed by ANOVA or Student's *t*-test, where appropriate. The distribution of expression data across samples was checked for normality. Where appropriate, adjustments were made to control for sex and genotype. Linear regression was used to assess the degree to which gene and/or protein expression levels were correlated and the strength of that correlation.

Results

Association Studies

Tag SNPs designed to index all common alleles with frequency greater than 0.05 were identified for *ECT2*, *GCH1*, and *UBE3A* using Haploview and Tagger with an r^2 >0.8. Tag SNPs covered 109.3kb, 67.4kb, and 143kb, respectively for *ECT2*, *GCH1*, and *UBE3A*. The respective transcriptional units for these loci are 67kb (*ECT2*), 60.8kb (*GCH1*), and 99.4kb (*UBE3A*). Since the majority of our sample is of European ancestry, the CEPH HapMap phase II data provided the basis for defining LD relationships for tag SNP selection. In all, six informative SNPs were selected for *ECT2*, eight for *GCH1*, and eight for *UBE3A*. I genotyped these SNPs using Taqman AoD assays in a sample of 694 ASD families described above in Table 4-1 and shown below in Figure 4-1.

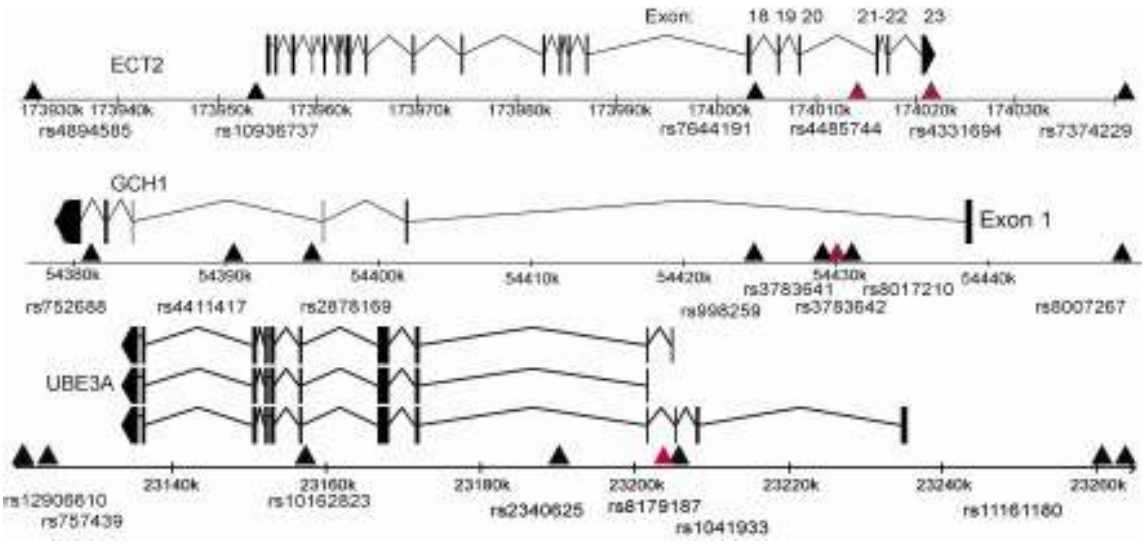


Figure 4-1- **Association results for *ECT2*, *GCH1*, and *UBE3A* in the 694 autism family sample.** Intron/exon gene structure for each of the three genes is shown. Tag SNPs (triangles) were selected using Tagger and the HapMap CEPH samples to select SNPs which would index the common alleles of each gene with frequency of greater than 0.05. SNPs significantly associated by the FBAT statistic are shown in red. These SNPs formed the basis for the "associated" (+) and "unassociated" (-) selection of samples used for gene and protein expression studies in Figures 4-2 through 4-5.

Results from association analysis are shown in Table 4-2. The Family Based Association Test (FBAT) was used to test allelic transmission from heterozygous parents to affected offspring and to generate a statistic based on what would be observed under the null hypothesis of no association. As shown in Table 4-2, both *UBE3A* substrates *ECT2* and *GCH1*, showed nominal evidence for allelic association; *ECT2* for two SNPs (P=0.029 and P=0.022) and *GCH1* for one SNP (P=0.022). In addition *UBE3A* showed association with one SNP (P=0.005). An additive model of affection is assumed.

Table 4-2. Association results for *ECT2*, *GCH1*, and *UBE3A.**

	SNP	Position	Relative position	Major Allele	Minor Allele	MAF	T	NT	P (-e)	RR
ECT2	rs4894585	173931157	-23.8kb 5' exon 1	G	**T**	0.22	284	264	0.32	
	rs10936737	173954018	-981bp 5' exon 1	C	**T**	0.44	428	390	0.19	
	rs7644191	174003520	intron 18	**C**	G	0.08	144	134	0.40	
	rs4485744	174014362	intron 20	G	**A**	0.40	**459**	**397**	**0.028**	**1.15 (1.05-1.27)**
	rs4331694	174021052	3'UTR exon 23	T	**C**	0.44	**440**	**375**	**0.022**	**1.17 (1.06-1.29)**
	rs7374229	174040526	18.6kb 3' exon 23	**C**	T	0.26	349	345	0.65	
GCH1	rs752688	54381319	intron 5	C	**T**	0.19	162	148	0.56	
	rs4411417	54390313	intron 3	T	**C**	0.2	164	142	0.41	
	rs2878169	54395743	intron 3	G	**T**	0.08	51	49	0.54	
	rs998259	54424781	intron 1	**C**	T	0.24	205	180	0.18	
	rs3783641	54429889	intron 1	T	**A**	0.2	165	141	0.14	
	rs3783642	54429953	intron 1	T	**C**	0.4	228	186	**0.022**	**1.22 (1.06-1.40)**
	rs8017210	54431586	intron 1	G	**A**	0.19	154	144	0.51	
	rs8007267	54448741	9.4kb 5' exon 1	**C**	T	0.18	93	121	0.07	
UBE3A	rs12906610	23120693	14.6kb 3' exon 15	T	**C**	0.13	186	148	0.10	
	rs757439	23123915	11.4kb 3' exon 15	**T**	C	0.29	345	306	0.14	
	rs10162823	23157263	intron 8	G	**A**	0.14	189	153	0.12	
	rs2340625	23190278	intron 6	**C**	G	0.23	281	256	0.28	
	rs8179187	23203419	exon 5	T	**G**	0.08	134	91	**0.005**	**1.47 (1.21-1.78)**
	rs1041933	23205842	intron 1	A	**G**	0.26	289	272	0.70	
	rs11161180	23260682	25.9 kb 5' exon 1	G	**A**	0.24	290	264	0.40	
	rs714900	23263676	28.9 kb. 5' exon 1	**C**	T	0.06	93	80	0.22	

* SNPs listed are given for the forward strand of each chromosome.

Selection of Associated and Unassociated Haplotypes

The haplotypes were tested, using the haplotype-based association test (HBAT) program. The HBAT -p option (haplotype permutation test) was used to compute the "exact" P-value via a Monte Carlo method. The HBAT -e option was used to compute the empirical P-value for association in the presence of linkage. The option "–e" of HBAT was used because it is a test of association given linkage.

In order to determine the nature of the association in the *ECT2* and *GCH1* genes, I used associated (and flanking) SNPs in both genes to identify the most

common associated haplotype and most common unassociated haplotype for each gene. For example, in the *ECT2* gene the minor alleles A and C for SNPs rs4485744 and rs4331694, respectively, are over-transmitted to affected individuals. The A and the C are on the same haplotype background and therefore the most common haplotype containing A-C was deemed the "*ECT2* associated" (*ECT2*+) haplotype. By contrast, the G-T alleles for the same two SNPs were under-transmitted to affected individuals. Therefore the most common haplotype background containing G-T was identified as the "*ECT2* unassociated" (*ECT2*-) haplotype. The same was done for *GCH1* to identify its associated and unassociated haplotypes based on the most frequent haplotypes bearing the C or T allele. The frequency of all *ECT2*, *GCH1*, and *UBE3A* haplotypes with frequency > 5 percent are shown in Table 4-3 along with those deemed "associated" and "unassociated". Major *ECT2* and *GCH1* haplotypes and their frequencies are given below. LD was empirically analyzed using the SNPs genotyped in our dataset by Haploview. Associated cell lines were those available who were homozygous for the associated and unassociated haplotypes, respectively, as determined by the most common haplotype bearing the associated and unassociated alleles, respectively.

Table 4-3 – Haplotypes selected for *ECT2* and *GCH1* associated and unassociated cell lines.

ECT2 Haplotypes - rs4485744 rs4331694

Alleles	Haplotype	freq	fam#	Haplotypes	S-E(S) [1]	Var(S) [2]	Z [3]	P	Global P
ECT2 -	GT	0.56	370	2	-25.06	196.68	-1.79	0.07	0.08
ECT2+	AC	0.387	366		29.88	188.47	2.18	0.03	

GCH1 Haplotypes - rs4331694

GCH1-	T	0.58	216	2	-25.83	127.86	-2.29	0.02	0.02
GCH1+	C	0.416	216		25.83	127.86	2.29	0.02	

[1-2] The S statistic in HBAT is calculated using the distribution of offspring genotypes conditional on affection status and parental genotype. E(S) and Var(S) are calculated under the null hypothesis conditioned on the parental genotypes.
[3] The Z statistic equals $(S-E(S)/(Var(S)^{(1/2)})$.

I then identified individuals who were homozygous for the associated and unassociated *ECT2* and *GCH1* haplotypes and cross-referenced them against our collection of available cell lines. Four associated individuals were chosen for each of four groups: *ECT2* associated (*ECT2*+), *ECT2* unassociated (*ECT2*-), *GCH1* associated (*GCH1*+), and *GCH1* unassociated (*GCH1*-).

In addition to the sixteen cell lines representing the above haplotypes, I also acquired lymphoblastoid cell lines from individuals with AS (n=4), dup(15q) autism (n=5), and CEPH controls (n=3) from Coriell Cell Repositories (see methods). These samples were included in gene and protein expression studies to best determine the effect of decreased, normal, and increased *UBE3A* expression on *ECT2* and *GCH1* expression and protein levels.

Gene Expression Findings

Comparisons in gene expression levels for multiple group means were done by ANOVA or by T-test when only two groups were compared. *UBE3A* transcript levels were did not differ between idiopathic autism groups associated

or unassociated for particular *ECT2* and *GCH1* haplotypes (Figure 4-2). As expected, AS samples showed significantly lower levels of *UBE3A* than all idiopathic autism groups (P=0.002), significantly lower levels than CEPH (P=0.027), and significantly lower levels than Dup(15q) autism (P<0.0001). In addition, *UBE3A* expression levels were higher in dup(15q) autism than idiopathic autism samples (P<0.0001).

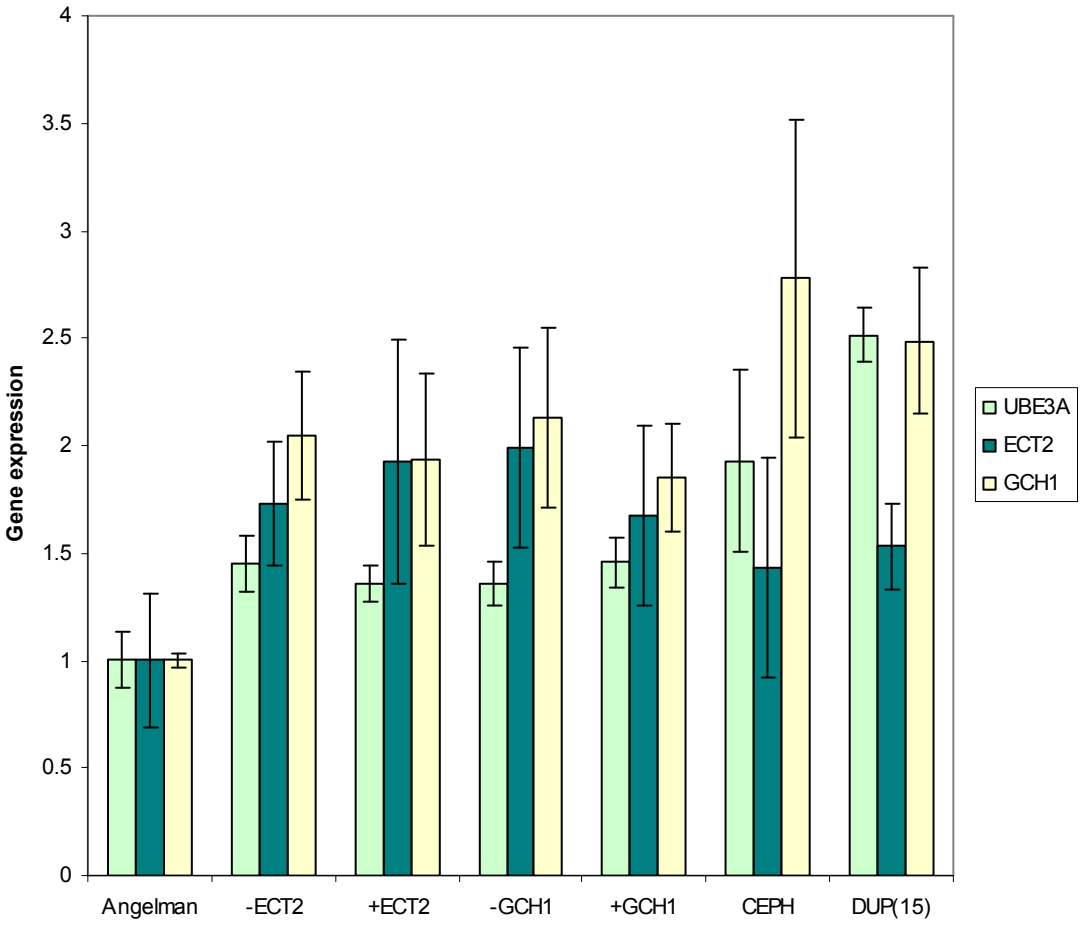


Figure 4-2 – Expression analysis for *UBE3A*, *ECT2*, and *GCH1* in control and patient cell lines. qRT-PCR was carried out to measure gene expression in lymphoblastoid cell lines from individuals with deletions of *UBE3A* (Angelman) and duplications of *UBE3A* (dup(15)). Expression analysis was also carried out on cell lines from individuals with autism who carried the associated (+) or unassociated (-) forms of *ECT2* and *GCH1*. While no significant differences were observed for the associated vs unassociated alleles of *ECT2* and *GCH1*, a trend is seen with *GCH1* transcript levels increasing with *UBE3A* expression as predicted by our model. Gene expression analysis was done in triplicate for each sample from each group (e.g. Angelman n=4 has triplicates for each for a total of 12 reactions). Error bars show standard error.

ECT2 gene expression levels did not differ significantly between *ECT2*+ and *ECT2*- groups. Expression levels showed no sign of difference in all idiopathic autism groups and dup(15) autism. This is not surprising given that

UBE3A and ECT2 *proteins* physically interact and that the ECT2 protein is a target for UBE3A ubiquitination. In other words, the ECT2 *protein* and not the *mRNA transcript* would not be expected to be impacted by increased UBE3A gene and protein expression. There is no evidence that the *ECT2 gene* is regulated by the UBE3A protein at the transcript level [173].

As with *ECT2*, *GCH1* transcript levels are similar comparing between *GCH1*+ and *GCH1*- groups. However, *GCH1* transcript levels were significantly different between AS and idiopathic autism groups (P = 0.008) but not between dup(15) and idiopathic autism groups (P = 0.1647). *GCH1* expression levels thus were higher in groups with higher levels of *UBE3A* (i.e. Dup(15q) and CEPH controls to lesser extent) and lower in groups with lower levels of *UBE3A* (i.e. AS).

Since *GCH1* transcript levels did not differ between *GCH1*+ and *GCH*- groups, but were positively correlated with increasing *UBE3A* levels, I analyzed *individual* samples (as opposed to groups) to determine how *GCH1* transcript levels might vary as a function of *UBE3A* gene expression levels. I found that across samples as there was a strong positive correlation between *UBE3A* and *GCH1* expression levels (p = 0.001). In order to rule out a phenomenon of global gene over-expression, I tested to see if there were also a relationship between *UBE3A* and *ECT2* gene expression and found that the two were not correlated (p = 0.294).

Although transcript levels in the associated and unassociated haplotype groups for *GCH1* were similar, I wanted to examine *UBE3A* levels *within these*

groups to explore any differences. I found that for the *GCH1 associated* group, *GCH1* levels increased with *UBE3A* levels ($p = 0.016$). In the *GCH1 unassociated* group, *GCH1* levels appeared to show a trend to decrease with increasing *UBE3A* levels, though the correlation was not significant ($p = 0.581$). To exclude the possibility that this provocative difference in correlation was the result of artifact, three independent technical replicates for RNA extraction were performed and gene expression analyzed for *UBE3A* and *GCH1* for both *GCH1*+ and *GCH1*- groups. The results were consistent with the trend that *GCH1* transcript expression increases as *UBE3A* levels increase in *GCH1*+ groups ($P=0.003$) with *GCH1*- apparently uncorrelated to *UBE3A* levels ($P=0.447$). In the larger experiment, the slope of the *GCH1*- group was not negative. I adjusted *UBE3A* levels for group and regression indicating no overall effect for groups on the response of *GCH1* to *UBE3A* (i.e. there was no significant difference in slope between groups.) The relationship between *UBE3A* and *GCH1* expression in the associated and unassociated groups is plotted in Figure 4-3.

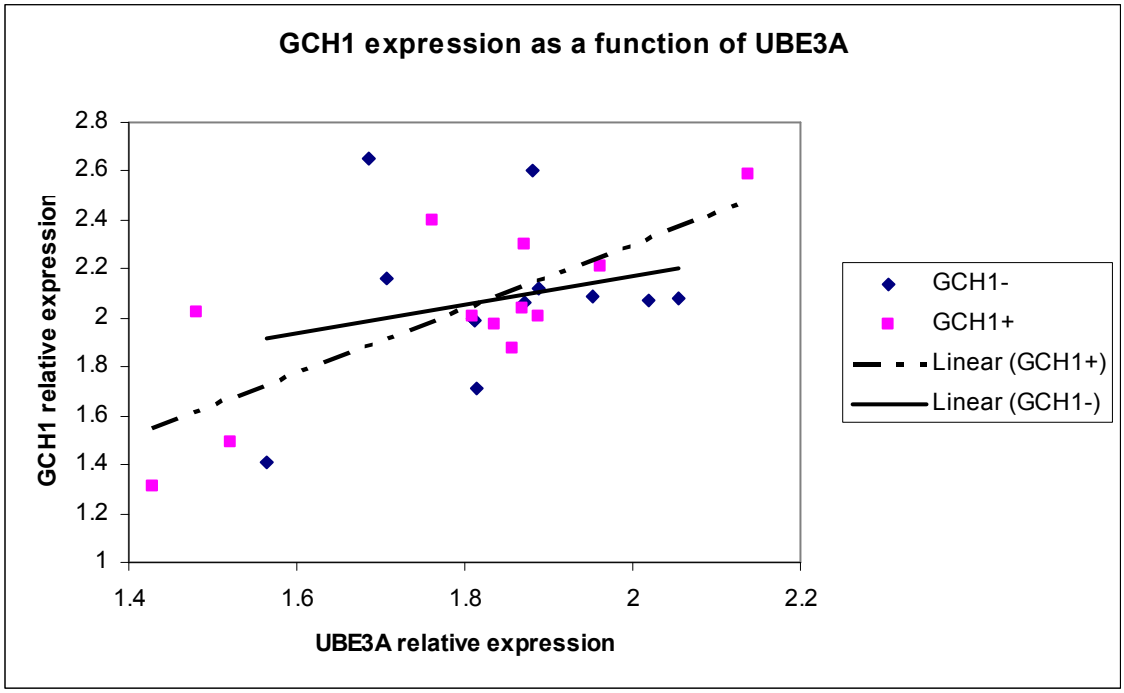


Figure 4-3 – *GCH1* levels as a function of *UBE3A* between *GCH1*+ and *GCH1*- groups.

Since *GCH1* transcript levels (in the *GCH1*+ group) were positively correlated with *UBE3A* expression, I further probed this relationship by manipulating *UBE3A* expression levels in two different cell types (HEK 293T and immortalized human lymphoblasts) to validate that this relationship was consistent. I transfected two different cell types, HEK 293T cells and a CEPH lymphoblastoid cell line with either (1) siRNA against *UBE3A* (2) empty pcDNA vector, (3) mutant *UBE3A*-Cys833Ala, or (4) wild type *UBE3A*. The *UBE3A*-C833A construct encodes an enzymatically inactive form of *UBE3A* which cannot catalyze the addition of ubiquitin to its substrate proteins [259].

UBE3A transcript levels were appropriately elevated in mutant and wild-type transfectants. In both cell lines, *UBE3A* and *GCH1* expression levels were

positively correlated. The strength of this correlation appeared high in both CEPH cells (p = 0.042, R-squared = 0.4692) and HEK-293T cells (p = 0.043, R-squared = 0.3803). Further, *t*-test of *GCH1* showed *GCH1* significantly higher in the *UBE3A* group than the siRNA group for the overall group (P=0.0049) and for both cell groups (HEK293-T, P=0.0266; CEPH, P=0.0262).

Protein Expression Studies

In order to verify that *UBE3A* protein levels were different between patient groups (AS, CEPH, and dup(15)), I performed Western blots on the three patient groups, shown in Figure 4-4, panel A. Quantification of *UBE3A* signal between groups was carried out and significant difference in *UBE3A* expression was found between groups by ANOVA (P=0.0005).

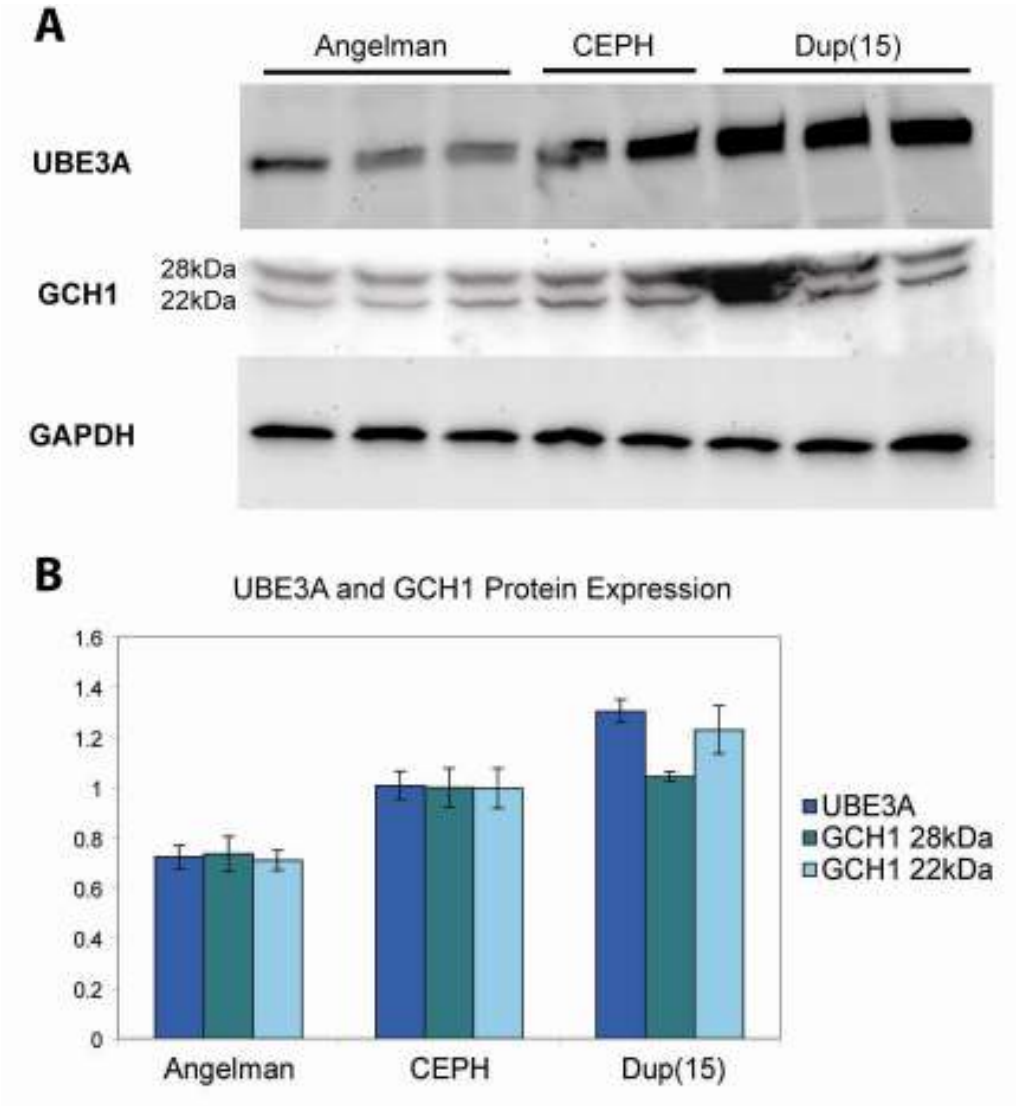


Figure 4-4 A- Western blot of *UBE3A* and *GCH1* in AS deletion, CEPH, and autism families. As expected, *UBE3A* protein expression is lower in AS than either CEPH or dup(15q) autism samples. *GCH1* levels also vary with AS samples showing the lowest levels of *GCH1* protein bands at 28 kDa and 22kDa and dup(15q) samples showing greater levels of *GCH1* protein expression. GAPDH was used as a loading control and did not change significantly among the cell lines. **4B- *GCH1* protein levels increase with increasing levels of *UBE3A* in patient lymphoblasts**. Relative expression levels measured by X on Western blots from patients with AS deletion, CEPH controls and interstitial duplication 15q. Note that although the levels of *UBE3A* and the *GCH1* isoforms are approximately equal in any given group they also increase together as the copy number for *UBE3A* increases, with the exception of the 28 kDa *GCH1* band in dup15q samples. The 22 kDa band for *GCH1* does increase in the dup15q samples, however.

I initially examined both ECT2 and GCH1 protein levels and found ECT2 levels to be surprisingly unchanged across all groups (data not shown). For GCH1, however, I noticed small differences between subject groups. Protein quantification was done for both GCH1 isoforms (Figure 4-4) and showed a significant difference between groups for both the larger isoform (28kDa, P=0.0496) and the smaller isoform (20kDa, P=0.0112). Quantification of protein is shown in Figure 4-4 panel B. To determine if there was a relationship between UBE3A and GCH1 protein levels, Linear regression was carried out. This showed a positive correlation between UBE3A protein levels and GCH1 28 kDa isoform (P=0.003) and 20 kDa isoform (P=0.001).

To determine whether the UBE3A-GCH1 correlation would persist for protein in an over-expression transfection mode, I transfected HEK293-T cells with (1) an siRNA against *UBE3A*, (2)empty vector, (3) *UBE3A*-C833A, and (4) wild type *UBE3A* and then measured UBE3A and GCH1 protein levels in samples from these four groups (Figure 4-5). UBE3A levels differed significantly among groups by ANOVA (P=0.0067) as predicted from the lymphoblast studies. In particular, the UBE3A-C833A form shows a much higher level of stable protein compared with wild type UBE3A. The UBE3A protein can self-ubiquitinate and the higher molecular weight smear may reflect some poly-ubiquitinated protein on the blot (Figure 4-5). GCH1 levels were significantly different for the 28kDa isoform (P=0.0045), but not 22 kDa isoform (0.1383) between the siRNA and UBE3A groups.

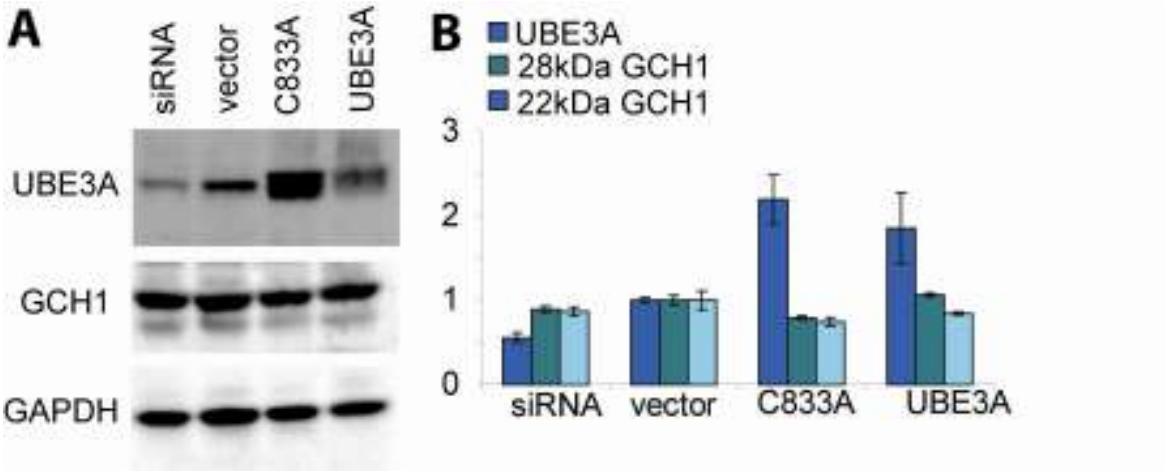


Figure 4-5A – **Transfection of *UBE3A* constructs do not change *GCH1* protein levels in HEK cells**. HEK 293T cells were transfected with either siRNA against *UBE3A*, empty vector, enzymatically defective *UBE3A*-C833A, or wild type *UBE3A*. Although changes in *UBE3A* protein levels were clear, they did not correspond to analogous changes in the amount of *GCH1* protein detected. **5B- Quantification of banding patterns.**

Linear regression on protein levels indicated that the linear relationship between *UBE3A* and *GCH1* transcript levels observed in lymphoblastoid cell lines was not recapitulated in this system, since neither isoform proved to significantly correlate with *UBE3A* levels. Between three replicate experiments (n=3 measurements for each treatment), the 28kDa isoform in *GCH1* did appear to be significantly stronger in the *UBE3A* group than in the siRNA group (P=0.0146). This was not true for the 22 kDa isoform (P=0.6985). Protein level quantification between experiments is shown in Figure 4-5.

To determine if gene expression levels by *UBE3A* were mirrored by protein levels, I looked at gene expression levels from mRNA and protein drawn from the same cell harvests between our patient groups (AS, CEPH, and dup(15)). Regression showed a positive correlation between all gene expression

and protein levels for *UBE3A* (P<0.0001) and *GCH1* (22 kDa isoform, P=0.012; 28 kDa isoform, P=0.043.

Gene Expression in HapMap Samples

To further test the co-directionality in expression levels of *UBE3A* and *GCH1* using an alternative strategy, I used HapMap genotypes and gene expression data from HapMap CEPH lymphoblastoid cell lines (available from GeneVar: http://www.sanger.ac.uk/humgen/genevar/) as the majority of our sample is Caucasian. We chose the 30 offspring of the HapMap trios. *UBE3A* and *GCH1* gene expression levels were evaluated for gender effects on transcription. There were no differences in *GCH1* transcript levels by association group in 8 *GCH1*- and 6 *GCH1*+ homozygous individuals (P=0.7409) However, there was again a relationship between the *GCH1*+ group and increasing levels of *UBE3A*, but there was no significant difference in the slopes between *GCH1*+ and *GCH1*- groups. There was, however, a strong correlation in the samples between *UBE3A* and *GCH1* expression levels, consistent with our earlier data (P=0.0007). This relationship is shown in Figure 4-6. To be sure that this might not be a common phenomenon, ten random gene expression probes were evaluated for correlation with *UBE3A* and none of these showed significant correlation, increasing the confidence in the *UBE3A*-*GCH1* relationship.

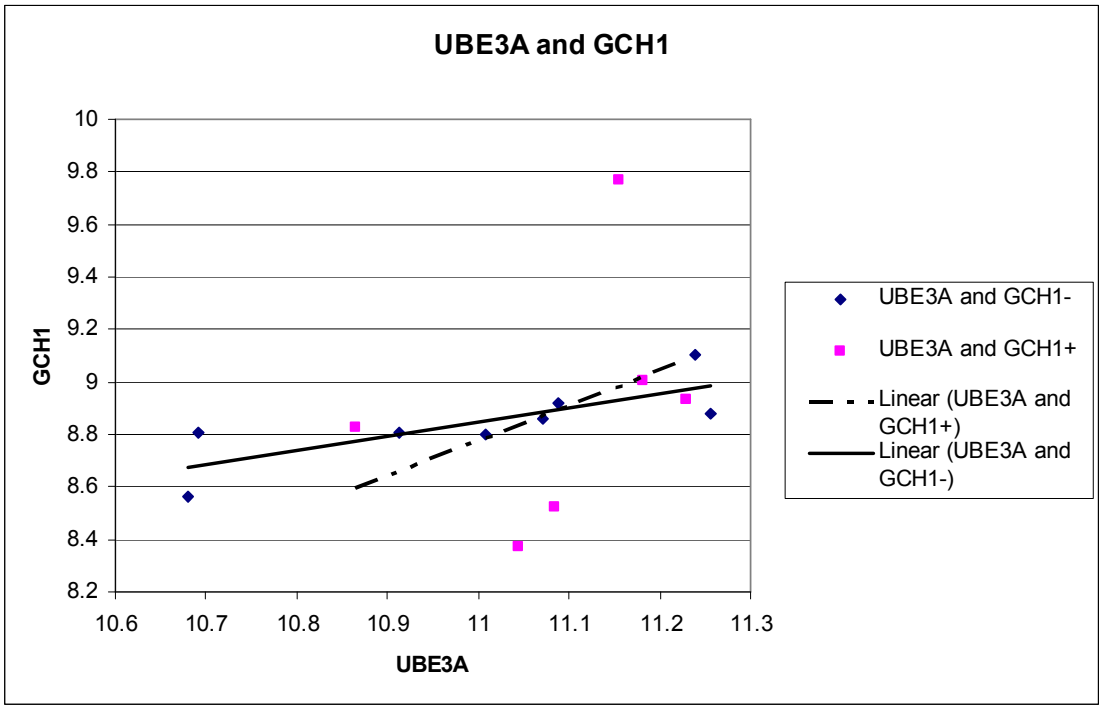


Figure 4-6 – HapMap CEPH trio offspring relative expression levels of *UBE3A* and *GCH1* in associated (*GCH1*+) and unassociated (*GCH1*-) groups.

Discussion

Maternal duplications of chromosome 15q11-q13 are the most common chromosomal abnormalities found in autism. I postulated that disruption or dysregulation of the UBE3A network confers risk for autism. In *Drosophila melanogaster*, over-expression of human *UBE3A* leads to dramatic down-regulation of the *ECT2* ortholog and up-regulation of the *GCH1* ortholog at the protein level. Here I have shown that there are particular alleles in this gene network associated with increased autism risk. Additionally, I validated the relationship between *UBE3A* and *GCH1*. It is possible that *ECT2* may be more involved in the AS phenotype than inherited autism [173].

A number of studies have indicated that maternal deficiency for *UBE3A* is necessary and sufficient to cause AS [95,116,256]. In addition, maternally inherited duplications of the chromosome 15q11-q13 region containing *UBE3A* are typically associated with an autism phenotype [119,267,268]. Although there are two genes in the autism candidate region, *UBE3A* and *ATP10A*, that exhibit maternal-specific expression, our data support the hypothesis that *UBE3A* regulated genes are promising autism candidate genes. Several studies in both animal models and autopsy brain samples have indicated reduced expression of *UBE3A* in Rett syndrome and idiopathic autism [194,198]. }. In fact, I did observe lower levels of *UBE3A* gene expression in idiopathic autism groups (non-dup(15)) relative to CEPH controls (P=0.0178). Given the significant phenotypic overlap between autism and Rett syndrome there is further support for the idea too little or too much *UBE3A* causes a synaptic phenotype by its action on other proteins in the brain.

It has been proposed that autism may be a disorder of synaptic modulation and maintenance, possibly through the maintenance of synaptic partners (e.g. *SHANK3* and the neuroligins) by *UBE3A* [118]. Recent studies have shown that *UBE3A* may also localize to the nucleus and the synapse [177]. In fact, the premise that subtle changes in synaptic plasticity and function may play a significant role in autism is bolstered by the observation that otherwise normal appearing neurons from Ube3a deficient mice have a reduced number of dendritic spines [177]. The protein product of *UBE3A* not only enhances the degradation of p53 but also regulates neuronal cell growth [259,266]. This

suggests that *UBE3A*, as with a growing list of other autism associated genes that regulate synaptic plasticity (e.g. neuroligins and neurexins), may function locally to regulate spine development or synaptic function. Individuals with the autism associated disorders such as fragile X and Rett syndrome have also been shown to exhibit altered spine morphology [269,270,271].

Transfection of *UBE3A* constructs into mammalian cells causes significant (though non-linear) increases in *GCH1* protein levels with over-expression of either wild type *UBE3A* or the catalytically inactive *UBE3A*-C833A. These data are consistent with recent findings in *Drosophila* that indicate an increase in the *GCH1* product tetrahydrobiopterin when either Dube3a or Dube3a-C/A is over-expressed in neurons (L. Reiter manuscript in preparation). The finding that *UBE3A* may regulate transcription of *GCH1* in the nucleus is not unprecedented, since it has been known for some time that this particular E3 ligase also has a transcriptional co-activation function during the regulation of steroid hormone receptors [272].

Previous work by Nurmi et al. has shown association of a microsatellite in the *UBE3A* gene with idiopathic autism [273]. Some studies have shown no associations with functional variants in *UBE3A* [274]. Most studies have found associations of only nominal significance either in *UBE3A* or the imprinting center and maternal expression domain and replication of identical associations in independent datasets is lacking [185,275,276]. The most recent genome-wide screens for association has indicated *UBE3A* among the top candidates by virtue of its frequent duplication status in autism patients [127,172]. Here I have

undertaken to capture *all* major allelic forms of *UBE3A* therefore more comprehensively indexing its common alleles than in earlier studies. More importantly, the association findings, while nominal, are provocative in that they show significant association in a large family dataset in *UBE3A* regulated genes (Table 4-2). The finding that both *ECT2* and *GCH1* show positive association suggests that variants within these genes may potentially increase risk for autism and that other *UBE3A* regulated genes may also represent excellent autism candidate genes.

According to one study, *UBE3A* mutations occur in 5.4% of cases of Angelman syndrome and loss of function mutations in *UBE3A* are known to be sufficient to cause AS [277]. The rate of *UBE3A* mutations is significantly higher in familial compared with sporadic subsets of AS [278]. Individuals with *UBE3A* mutations are typically more severely affected than those with imprinting center mutations or subjects with uniparental disomy (UPD), but less severely affected than those with 15q11-q13 deletions (*UBE3A* and surrounding genes) [165].

Using information from association analysis (Table 4-3) I selected autism subject cell lines homozygous for risk haplotypes and then examined gene expression. Our studies indicated, comparing associated and unassociated *groups*, there were no significant differences in *UBE3A*, *ECT2* or *GCH1* gene expression levels. This is not necessarily surprising since gene expression patterns may be dependent on cell types and specific factors which could be different in our lymphoblastoid cell lines.

However, when I investigated the relationship between *UBE3A* and *GCH1* expression, the relationship between these two genes observed in *Drosophila* was confirmed. Also interesting is that there appears to be some modest difference in the strength of correlation of *GCH1*+ compared to *GCH1*- cells to increasing levels of *UBE3A* with regard to gene expression. That is, the *GCH1*+ cells showed a trend toward increases in *GCH1* levels than the *GCH1*- group. In transfection experiments *GCH1* levels increased in two cell types as a function of increasing *UBE3A* levels. This includes cells transfected with the *UBE3A*-C833A construct which cannot catalyze ubiquitination, suggesting that the transcriptional co-activation function of *UBE3A* may act to increase *GCH1* levels.

Protein studies mirrored gene expression studies closely in this regard. In our patient groups, *GCH1* levels increased with *UBE3A* levels and this was largely true in our transfected cells. The exception to this was with transfection of the *UBE3A*-C833A group, which appeared to have lower levels of *GCH1* protein than the siRNA group. This is a harder phenomenon to explain. Increased *UBE3A*, however, only produced modest increases in *GCH1* protein expression so it appears at least in these two cell types (HEK-293T and lymphoblastoid cell lines) that the relationship between *UBE3A* is present, but not profound, so additional experiments in animal models in the appropriate cell type (neurons) will be required to confirm this relationship is relevant to autism.

ECT2 and *GCH1* orthologs were initially identified using proteomic profiling in *Drosophila* head extracts as two proteins up-regulated or down-regulated, respectively, as a result of *UBE3A* over-expression [173].

The finding of genetic association in both of these genes, as well as with *UBE3A* indicates that the network of *UBE3A* and its downstream targets (and possibly upstream regulators) can individually act as risk factors in autism. I have biochemically confirmed the positive correlation between *UBE3A* and *GCH1* protein expression (but not for *UBE3A* and *ECT2*). Some caveats should be considered in the interpretation of these data. First, our analysis was performed on immortalized lymphoblastoid cell lines, rather than from brain tissue or neuronal cell lines; and the relationship among these genes may by depend on tissue specific factors. Furthermore, while there is some evidence that elevating *UBE3A* levels increases *GCH1* protein levels, this was not the case for over-expression of *UBE3A*-C833A (Figure 4-5). It might be that a catalytically inactive form of *UBE3A* increases gene expression of *GCH1* (as I observed) but because it cannot ubiquitinate targets it may not properly regulate intermediates in the regulation of *GCH1* protein activity and turnover.

Given the evidence presented here and published elsewhere, *UBE3A* and genes and proteins it regulates by *UBE3A* appear to be valid candidate loci for harboring alleles that confer risk for autism. We hypothesize that dysregulation of these genes in idiopathic autism may lead to a diminished ability to form functional synapses.

CHAPTER V

STUDIES OF *GABRB3* IN AUTISM

Introduction

The most common chromosomal abnormalities in autism are 15q11-q13 duplications (OMIM: 608636; a.k.a. *AUTS4*) that occur as an interstitial gain or supernumerary idic(15) marker chromosomes with two additional copies of 15q11-q13 [3,119,268,279]. While paternal duplications are observed, autistic phenotypes are almost always associated with duplications of maternal origin. Maternal deletion of this region leads to Angelman syndrome (OMIM: 105830), and paternal deletion to Prader-Willi syndrome (OMIM: 176270), both of which share features of ASDs [119,154,256,280]. As a consequence, the genes in this interval have become candidates for investigation of their potential contribution to *idiopathic* autism susceptibility. A cluster of $GABA_A$ receptor subunit genes lies within this interval. $GABA_A$ receptors are ligand-gated chloride channels, which mediate the majority of fast synaptic inhibition in the brain. Functional $GABA_A$ receptors are composed of five subunits that form a chloride ion channel and are typically composed of two α, two β, and a γ or δ subunit.

Multiple lines of evidence have pointed to the potential involvement of GABAergic systems in autism [190,192,281,282]. One gene in the 15q11-q13 cluster encodes the β3 subunit, *GABRB3*. Several reports have documented association of common alleles at *GABRB3* with autism. Data from autism studies

(not all of which are positive) show some evidence for replication, but also suggest allelic heterogeneity [104,105,106,187,283,284]. Recent epigenetic studies of the 15q $GABA_A$ receptor subunit cluster by Hogart and colleagues indicate predominantly biallelic expression of these genes in normal brain samples [193,194]. A subset of autism samples, however, showed monoallelic or allelic bias in expression suggesting epigenetic dysregulation.

GABRB3 is known to play a significant role in development of the CNS. β3 is the major β isoform present in a number of brain regions in the prenatal and neonate brain [285,286]. β3-containing receptors are critical for stem cell proliferation as knock down of *GABRB3* by siRNA has been shown to cause blastocyst-stage embryos to develop faster than controls and contain smaller cells and a larger blastocoel [287]. In adult brain *GABRB3* exhibits a far more restricted pattern of expression, indicating a potentially critical role for β3 in early brain development [285,288,289]. The importance of *GABRB3* in development is underscored by work with knockout mice. There is a 90% mortality rate in *Gabrb3* null animals, but the 5-10% of β3−/− mice that survive to adulthood, despite multiple behavioral and physiological abnormalities, achieve normal weight and are reproductive by adulthood, though they die prematurely [290]. *Gabrb3* knockout mice show evidence of seizures, hypersensitive behavior, both common to autism [290]. The *Gabrb3* knockout mouse also exhibits cleft palate, an occasional finding in those with chromosome 15q11-q13 duplications or other syndromic ASDs [290]. *Gabrb3* null mice also show fewer functional $GABA_A$ receptors and pharmacological evidence indicates that other β subunits are not

adequately replacing the absent β3 subunit [291]. Behavioral findings include diminished interest in social activity, reduced exploratory behavior, and fewer nurturing tendencies than wild-type littermates [292]. These impairments suggest that the Gabrb3 knock-out mouse may be a model for studying ASD behaviors in the mouse [292,293].

Based on the above evidence for *GABRB3* as a candidate gene for the idiopathic autism condition, I sought to determine if rare and/or common variants and common variants in this gene contributed to susceptibility. This chapter is divided into two parts. In the part 1, I describe our efforts to interrogate *GABRB3* for rare variants and present a detailed report of one of those variants, P11S. In part 2, I describe our use of genotyping tag SNPs to index all common alleles of *GABRB3.*

Chapter V - Part 1: Analysis of a Rare Variant P11S in Autism

Subjects and Methods

Subjects

Families included in this study were recruited at Vanderbilt University, the University of Chicago, the University of California at San Diego, or they were obtained from the NIMH Repository (http://nimhgenetics.org). Affected individuals were subject to a research diagnosis based on scoring of assessment using the Autism Diagnostic Interview – Revised and the Autism Diagnostic Observation Schedule. All subjects provided informed consent and this work was conducted under approvals from the Institutional Review Boards at recruiting institutions.

Genetics

Genotyping was conducted using Applied Biosystems (ABI, Foster City, CA, USA) TaqMan Assays-on-Demand (AoD). PCR for the P11S variant (rs25409; ABI assay C__44811455_10) was carried out in 384-well plates in 5µl reactions containing 0.125 µl 20x AoD probe/primer mix, 5ng of genomic DNA and TaqMan Universal PCR Mastermix according to manufacturer's recommendations. Products were scanned on the ABI 7900HT instrument to call genotyped. Genotyping efficiency was 98% and genotyped conformed to expectations under Hardy-Weinberg Equilibrium. Quality control also included inter-plate and intra-plate replicate samples, as well as checks for within-family Mendelian

inconsistency using PEDCHECK [223]. Family-based association tests and genotype relative risks were calculated using the GenAssoc module for STATA (v9.2) provided by David Clayton (http://www-gene.cimr.cam.ac.uk/clayton/software/) [294]. Log-linear models were constructed using SAS version 9.1.

Phenotype

Subjects were classified as affected under a "strict" diagnostic classification if they met criteria for autism on the ADI-R, while "broad" classification also includes individuals who met ASD1 or ASD2 criteria according to Risi et al [295]. Items from the ADI-R were the basis for comparing S11 carriers and P11 homozygotes. Subjects with missing data were excluded from analysis. Effect of genotype on seizures was conducted using a Fisher's Exact test comparing numbers of subjects with definite seizures (score of 2) to those with no history of attacks (score of 0) on the "Faints/Fits/Blackouts" item. Individuals with a history of attacks without a diagnosis of epilepsy or suspected attacks (score of 1) or febrile seizures only (score of 7) were treated as missing. For exploratory analyses, core behavior domain and subdomain scores were compared using univariate analysis of variance (ANOVA) with score as the dependent variable, genotype group as the independent variable, and age at ADI-R as a covariate. Algorithm item scores were compared using Mahon's chi-square test to account for the ordinality of the ADI-R scores which were intended to be scored qualitatively (from 0 = "absence of behavior specified" to 2 =

"behavior definitely present"). When present, individual item scores of 3 were down-coded to 2 as instructed on the algorithm. The "ever" or most severe scores for the following items were also down-coded and compared when available: "Loss of Skills," "Overall Level of Language," Sensitivity to Noise," "Difficulties with Minor Changes in Subject's Own Routines or Personal Environment," "Resistance to Trivial Changes in the Environment,, "Gait," "Coordination," "Aggression toward Caregivers or Family Members," "Aggression toward Noncaregivers or Nonfamily Members," "Self-Injury," "Overactivity," and "Special Isolated Skills." Ages for developmental milestones (including 1st steps, 1st words, 1st phrases, and continence) were also compared between groups using ANOVA with age as the dependent variable, genotype group as the independent variable, and age at ADI-R as a covariate; when codes rather than specific ages were given (e.g., 997 for "not known, but apparently delayed") the subject's data was treated as missing.

Ancestry Analysis

Classical multidimensional scaling (MDS) was conducted using PLINK [225] for the sample of AGRE parents (founders). MDS dimensions were estimated from genome-wide average proportion of alleles shared identical by state for each possible pair of founders in the sample. Graphical representation of the first two dimensions were used to identify population substructure and ancestry clusters.

Electrophysiology

Expression of recombinant $GABA_A$ receptors and subsequent whole cell recordings from lifted cells were conducted as previously described [296]. Human embryonic kidney (HEK) 293-T cells were co-transfected with 2 µg of each subunit-encoding plasmid and 1 µg of the pHook-1 cDNA (Invitrogen, Carlsbad, CA) using a modified calcium phosphate precipitation method and subsequently selected 24 hours after transfection using magnetic hapten-coated beads.

Biotinylation and Western Blot Analysis

Cell surface receptor biotinylation and western blot procedures were modified from a previous protocol [296]. For cell surface receptor biotinylation, live, transfected cells were washed with phosphate buffered saline (PBS) containing 0.1 mM $CaCl_2$ and 1 mM $MgCl_2$ (pH 7.4) followed by incubation with sulfo-NHS biotin for 1 hour at 4º C. Sulfo-NHS biotin was quenched with PBS containing 0.1 mM glycine. Cells were lysed in RIPA buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 1mM EGTA, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate), supplemented with protein inhibitor (Roche) for 1 hour at 4º C. The extracted supernatant was then incubated with immobilized streptavidin overnight at 4º C. Biotinylated proteins were eluted from the streptavidin by incubation with 1× NEB glycoprotein protein denaturing buffer (5% SDS, 0.4 M DTT) at room temperature for 30 minutes. The supernatant was then either undigested or digested with Endo-H or PNGase F prior to fractionation by 10% SDS-PAGE.

^{35}S radiolabeling Metabolic Pulse-Chase Assays

^{35}S methionine pulse-chase experiments were conducted with modifications from a previously published protocol [297]. Briefly, 48h hours following transfection, cells were replenished with starving medium that lacked methionine and cysteine (Invitrogen), and incubated at 37°C for 30 min. Starving medium was then replaced by 1.5 ml ^{35}S radionuclide methionine (100-250 μCi/ml (1 Ci = 37GBq); PerkinElmer, Wellesley, MA) labeling medium for a series of different time points at 37°C. FLAG-tagged human β3 subunits were then immunoprecipitated from radio-labeled lysates with an anti-FLAG M2-agarose affinity gel by rotating at 4°C overnight. Immunoprecipitated products were then eluted from the beads with FLAG peptide (Sigma-Aldrich)I, and immunopurified subunits were then analyzed by 10 % SDS-PAGE and exposed on a digital PhosphorImager (GE Healthcare, Piscataway, NJ).

Data analysis

Macroscopic currents were low pass filtered at 2 kHz, digitized at 10 kHz, and analyzed using pClamp9 software suite (Axon Instruments). Except for the pulse-chase assays, proteins were quantified by ChemiImager AlphaEaseFC software. Data from pulse-chase experiments were quantified using Quantity One software (Bio-Rad, Hercules, CA). Numerical data were expressed as mean ± SEM. When wild-type data were arbitrarily taken as 1, column statistics were

used. Statistical significance, using Student's unpaired t test (GraphPad Prism), was taken as $P < 0.05$.

Results

Association of *GABRB3* P11S with Autism

To screen for novel functional variation, *GABRB3* exons were sequenced in a discovery sample of 100 unrelated probands, and from this effort a single nonsynonymous variant (C87T, Pro11Ser) was identified in exon 1a, one of two alternative initiating exons for *GABRB3* [298]. This variant, deposited into dbSNP (rs25409), was determined to be inherited and of maternal origin in two unrelated cases in an initial study cohort. We subsequently genotyped this variant in a sample of 1,152 combined simplex and multiplex families to determine its frequency and potential association with autism in a larger population (Table 5-1). Seventeen families (1.47%) were found to harbor the rare S11 variant (Figure 5-1), corresponding to an allele frequency of 0.40%. Given substantial precedent for parent-of-origin effects in the region, parental transmissions were examined separately, and we observed a maternal (but not paternal) S11 over-transmission ($P = 0.045$; Table 5-2). Indeed, 12 of 16 maternal transmissions resulted in a broad autism phenotype (see Methods), compared with only 3 of 8 paternal transmissions. For the strict autism phenotype the maternal over-transmission was more pronounced ($P = 0.008$), with 12 of 14 transmissions resulting in a strict autism phenotype. The S11 variant confers a genotype relative risk (GRR) of 3.00 (95% CI: 1.26-7.12, $P = 0.013$) for the broad autism phenotype when the

transmission is maternal in origin. Similarly, the S11 variant confers a GRR of 6.00 (95% CI: 1.62-22.16, *P* = 0.007) for the strict autism phenotype when transmission is maternal in origin. Genotyping Caucasian controls identified a single S11 carrier from a sample of 584 chromosomes, corresponding to a 0.17% allele frequency, compared with 0.40% observed in the overall autism sample.

Table 5-1. Characteristics of families with autism genotyped for the Pro11Ser variant (P11S) stratified by ancestry.

	ALL	AA	ASIAN	CAUC	HISP	UNK
Families	1152	40	40	922	49	101
Individuals	4603	145	168	3719	201	370
Number with autism	1776	58	68	1433	82	135
Female	331	9	11	247	23	41
Male	1465	49	57	1186	59	114
Number without diagnosis	2807	87	100	2266	119	235
Female	1419	50	54	1144	64	107
Male	1388	37	46	1122	55	128
Number of Families with:						
0 affected	38	2	1	19	0	16
1 affected	498	18	14	406	22	38
2 affected	553	20	22	449	23	39
3 affected	56	0	2	43	4	7
4 affected	6	0	1	4	0	1
5 affected	1	0	0	1	0	0

AA=African American, CAUC=Caucasian, HISP=Hispanic, OTHER=other, UNK=unknown, or more than one ethnicity.

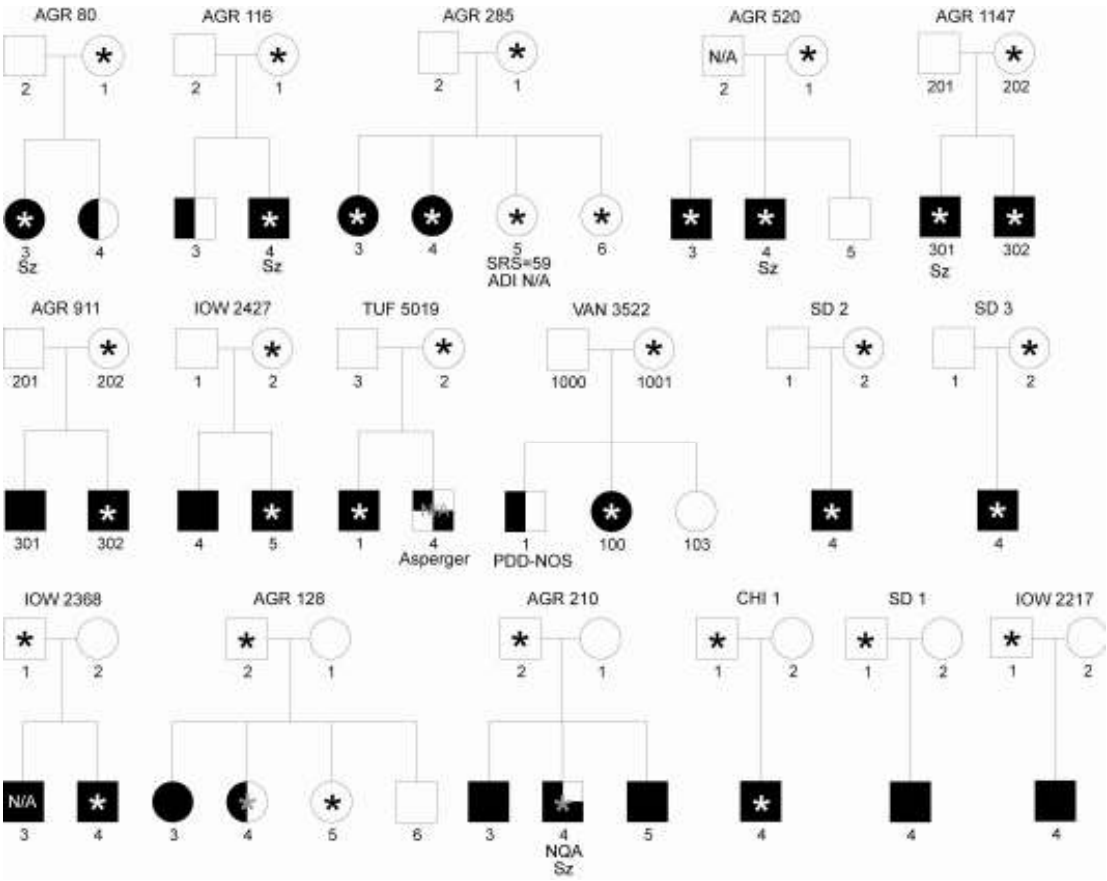


Figure 5-1 - Pedigree structure of ASD families bearing the P11S variant. 17 ASD families in total were identified to harbor the variant. Individuals carrying the variant are marked with asterisks. Individuals for whom DNA was not available are marked N/A. Our strict analyses considered individuals affected if they met criteria for autism on the ADI-R diagnostic algorithm (completely filled black). Our broad analyses considered these individuals and individuals who met the AGP criteria ASD1 or ASD2 as affected (half filled black) (see [295] for details); note no individuals met ASD2. Unfilled individuals are considered unknown. For individuals who did not meet broad criteria Social Responsiveness Scale (SRS) t-scores from teacher (/parent) report are provided when available. SRS scores ranging from 60-75 are considered mild to moderate range for ASD; children with high functioning autism may score in the t-score range of 55-59. AGR 80-4 had significant language delay and impairment per the ADI-R, but did not meet our broad criteria or have a SRS. Individuals with definite seizures per the ADI-R are marked SZ = 2; those with suspected seizures per the ADI-R are marked SZ = 1.

Table 5-2. **Transmission statistics, parental origin and genotype relative risk estimates are provided for the Pro11Ser (C87T from NM_021912) variant rs25409.**

Affection	Parental origin	Fams.	Informative Trans.	P11 (C) Observed Trans.	P11 (C) Expected Trans.	S11 (T) Observed Trans.	S11 (T) Expected Trans.	χ^2	P	Genotype Relative Risk
Spectrum	Both parents	17	24	9	12	15	12	1.5	0.220	1.67 (0.73-3.81)
	Maternal	11	16	4	8	12	8	4.0	**0.046**	**3.00 (0.97-9.30)**
	Paternal	6	8	5	4	3	4	0.5	1.000	0.60 (0.14-2.51)
Strict	Both parents	17	21	7	10.5	14	10.5	2.3	0.130	2.57 (1.01-5.86)
	Maternal	11	14	2	7	12	7	7.1	**0.008**	**6.00 (1.34-26.81)**
	Paternal	6	7	5	3.5	2	3.5	1.3	0.250	0.40 (0.08-2.06)

Exploratory parent-of-origin tests were also conducted using a log-linear model framework allowing for a maternal genetic effect [299]. Application of this approach allowed us to exclude the possibility that a simple effect of maternal genotype, in which offspring of mothers with the S11 allele were at increased risk of developing autism regardless of whether the S11 allele was transmitted to the offspring, accounted for our observations, since there was clearly preferential transmission of the S11 allele to affected offspring of heterozygous mothers. Using either a broad or a strict autism phenotype, this parent-of-origin effect was statistically significant (p=0.040 and p=0.004 respectively). Although it is theoretically possible that there could be a direct effect of maternal genotype on offspring risk of autism in addition to the parent-of-origin effect that I have established, we have little power to test for a maternal effect in the presence of a parent-of-origin effect. A fully developed test for such joint effects would require a more fully characterized sample with ascertainment of all affected and unaffected offspring from both mothers and fathers carrying this risk allele.

The *GABRB3* P11S mutation was recently identified in 2 independent families segregating for CAE in a Hispanic/Mexican-American cohort of 48 families [300]. Self report data for the families genotyped in our study suggested that the S11

variant was on a non-Hispanic Caucasian background. To resolve uncertainty about ancestry, and to ensure that an appropriate control sample was used, we examined genome-wide SNP data available for a subset of subjects from the Autism Genetics Resource Exchange (AGRE) collection, which was previously genotyped using the Affymetrix 5.0 (500k) SNP platform [135]. Multidimensional scaling (MDS) of SNP genotype data for all AGRE parents was conducted using PLINK to estimate dimensions of population genetic variation [225]. Figure 5-2 shows a graphical representation of the first two dimensions from this analysis that identified population substructure and ancestral clusters for founders (parents). Our analysis found that the S11 variant was present in Caucasian parents. The rare S11 variant had a frequency of 0.94% (7/743) and 0% (0/105) in Caucasian and Hispanic founders, respectively, though only 75 families screened were determined to have Hispanic ancestry. Given the mixed ancestral history of Hispanic populations with Spanish Caucasians and Native Americans, it is not surprising that the variant was identified previously in a Hispanic sample. The total S11 frequency was 0.48% in the overall sample.

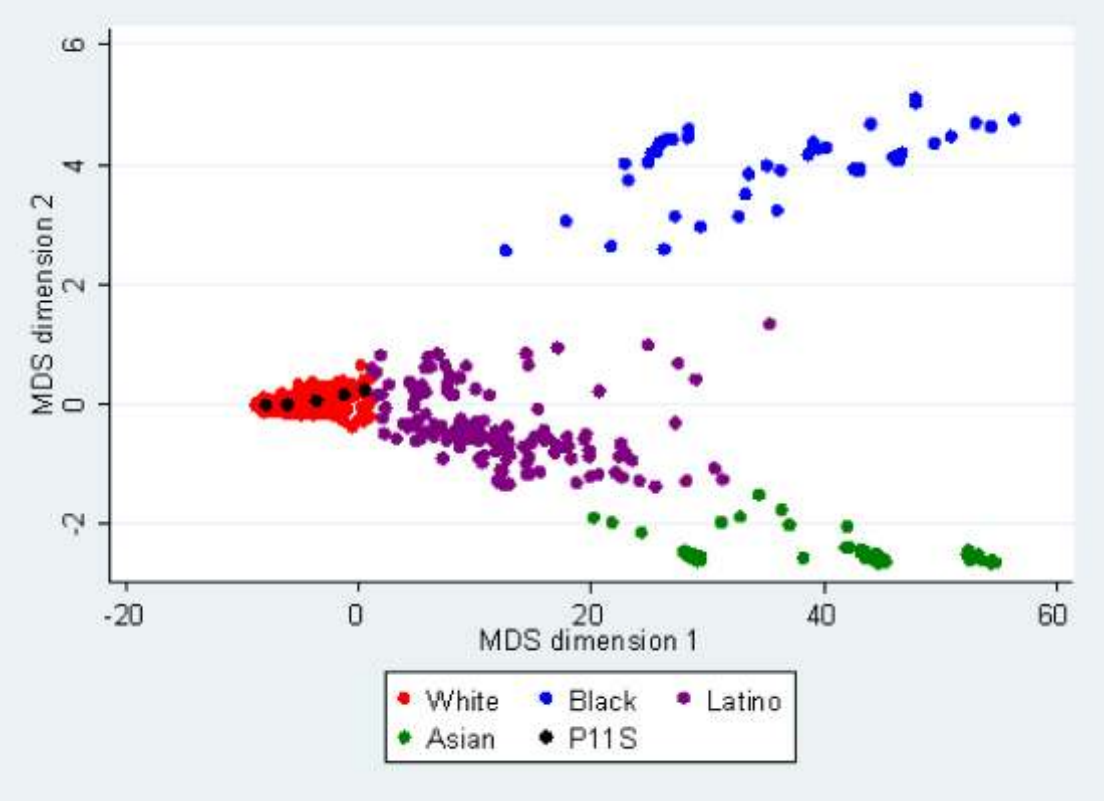


Figure 5-2 – MDS plot of all AGRE parents carrying the S11 risk allele
Black dots represent parents with p11s mutation. Out of 22 AGRE samples with rare variation, 7 are founders with AGRE Affy 5.0 genome-wide data available. Classical multidimensional scaling (MDS) was conducted in PLINK using the total sample of AGRE parents in order to estimate dimensions of population genetic variation. These dimensions are estimated from genome-wide average proportion of alleles shared by state for each possible pair of individuals in the sample. Graphical representation of the first two dimensions is used to identify population substructure and ancestry clusters.

Phenotypic Correlates of P11S in Autism

Given the association of the S11 mutation with CAE, I examined item level data from the Autism Diagnostic Interview-Revised (ADI-R)[27] on seizure history. The rate of confirmed non-febrile seizures was 16.7% (2/12) in affected probands with the maternal S11 compared to 4.9% (55/973) for those homozygous for the P11 allele, but this difference was not significant ($P = 0.10$; Table 5-3). Seizure rates in autism increase with age, however, age did not

significantly differ between the maternal S11 (M = 97 mos ± 45) and homozygous P11 (M = 94 mos ± 52) groups, t(1137) = -0.18, *P* = 0.86.

Table 5-3. Frequency (% within group) of seizures in probands with and without S11.

Rs25409 genotype	None (0)[1]	Possible (1)	Definite (2)	Febrile seizures (7)	Total	Mean age (sd) (mos)
Homozygous P11	973 (86)	76 (7)	55 (5)	23 (2)	1127	94 (52)
P11/S11 (maternal)	9 (75)	1 (8)	2 (17)	0 (0)	12	97 (45)
P11/S11 (paternal)	1 (50)	1 (50)	0 (0)	0 (0)	2	48 (28)

[1] Score on ADI-R item "Faints/fits/seizures." A score of 7 means "Febrile convulsions only, with no continuing daily medication outside the period of fever." Individuals with definite seizures with S11 were from different families. Seven sibling pairs homozygous P11 reported definite seizures; the remaining 41 individuals homozygous P11 with definite seizures were from different families.

Tables 5-4 and 5-5 provide characteristics of probands with maternally- and paternally-derived variants, respectively. Inspection of S11 pedigrees indicates that maternal transmission results in a narrowly-defined autism phenotype based on standard diagnostic algorithms including the ADI-R. To further explore P11S phenotypic correlates, I compared ADI-R dimensions between the maternal S11 to the homozygous P11 cases. Individuals with the paternal S11 were considered unknown for these analyses and omitted since the variant was not significantly over-transmitted from fathers to individuals with ASD. Table 5-6 shows the results for the comparison of groups' domain and subdomain scores. The maternal S11 and homozygous P11 groups did not significantly differ on any domain score. The maternal S11 group scored more severely than the homozygous P11 group on the subdomain "Stereotyped, repetitive, or idiosyncratic speech" and less severely on "Preoccupations with part of objects or non-functional elements of material." There was no significant

difference in ADI-R item level data between groups where the maternal S11 cases were more severe at the $P < 0.01$ level.

Table 5-4. Description of affected individuals with maternally inherited S11. (See legend on next page.)

	AGR 80-3	AGR 116-4	AGR 285-3	AGR 285-4	AGR 520-3	AGR 520-4	AGR 1147-301	AGR 1147-302
Sex	F	M	F	F	M	M	M	M
Epilepsy	Possible	Possible	No	No	No	Confirmed	Confirmed	No
Age head circumference (yrs)	8	--	12	9	--	--	--	9
Head circumference (%ile)	82	--	98	98	--	--	--	90
Autism phenotype	Strict	Strict	Strict	Strict	Strict	Strict	Strict	Strict
ADOS classification	Autism	--	Autism	Autism	--	--	Autism	Autism
Age IQ test (yrs)	--	--	12	9	--	--	--	--
Verbal IQ	--	--	72	43	--	--	--	--
Non-verbal IQ	--	--	83	83	--	--	--	--
ADI-R								
Age ADI-R (yrs)	6	17	9	7	11	6	8	6
Classification	Autism	Autism	Autism	Autism	Autism	Autism	Autism	Autism
Social domain	22	28	16	25	28	27	28	26
Communication domain verbal			9	16	26	19		
Communication domain non-verbal	13	14					14	14
Restricted, repetitive behavior domain	5	8	4	4	11	5	4	4
Age 1st words	Regression	96	11	14	38	38	16	32
Age 1st phrases	DNA	114	13	33	38	48	18	38
Any language regression	SP,W,S,A	No	No	No	No	No	SP,W,A	No
Any regression of other skills	C,S,P,SH,M	No	No	No	No	No	S	No
Anxiety	Absent	Absent	+	+	++	++	--	--
Aggression	++	++	+	++	++	--	++	++
Overactivity	++	+++	Absent	Absent	++	Absent	--	--
Savant skills	M	MEM,R	MEM,C	Absent	D	Absent	Absent	Absent
Age 1st steps	18	12	11	11	12	15	12	10
Odd gait	+	+	No	+	+++	++	++	++
Gross or fine motor difficulties	GF	None	F	None	GF	GF	--	--
Age urinary continence	--	--	--	--	--	--	DNA	DNA
Age bowel continence	DNA	DNA	45	45	24	52	DNA	DNA

Sex (M: male, F: female); Epilepsy designation comes from the ADI-R item “history of seizures” and the AGRE physical examination when available; Classification “Autism” met autism cut-offs on the ADI-R algorithm; Level of language (NV: non-verbal or fewer than 5 words used/day, V: verbal, fluent with daily phrase speech); Language regression (SP: speech, W: 5+ words, S: syntax, A: articulation); Regression of other skills (C: communication, S: social, P: physical, SH: self-help, M: motor); Anxiety includes social anxiety and fears (+ mild to +++ most severe); Aggression includes to family members, others, and self-injury (+ mild to +++ most severe); Savant skills (M: music, MEM: memory, R: reading, C: computational, D: drawing); Gait abnormality (+ mild to +++ most noticeable); Bowel: age of continence (DNA – did not achieve continence at time of interview); ADOS classification of “Autism” met autism cut-offs on the ADOS algorithm.

Table 5-4 (continued). Description of affected individuals with maternally inherited S11. (See legend on page 109.)

	AGR 911-3	IOW 2427-5	TUF 5019-1	VAN 3522-100	SD2	SD3
Sex	M	M	M	F	M	M
Epilepsy	No	No	No	No	No	Yes
Age head circumference (yrs)	--	--	--	--	4.5	4.9
Head circumference (%ile)	--	--	--	--	99	84
Autism phenotype	Strict	Strict	Strict	Strict	Strict	Strict
ADOS classification	--	--	--	Autism	--	--
Age IQ test (yrs)	--	--	--	--	8	6
Verbal IQ	--	--	--	--	89	49
Non-verbal IQ	--	--	--	--	117	46
ADI-R						
Age ADI-R (yrs)	4	13	7	5	12	6
Classification	Autism	Autism	Autism	Autism	Autism	Autism
Social domain	20	20	15	26	24	28
Communication domain verbal		18	20		21	20
Communication domain non-verbal	11			13		
Restricted, repetitive behavior domain	3	5	7	3	13	9
Age 1st words	15	12	--	Regression	24	30
Age 1st phrases	DNA	16	--	Regression	36	60
Any language regression	No	SP,W,S,A	No	SP,W,S,A	--	--
Any regression of other skills	No	No	No	No	--	--
Anxiety	--	--	++	++	--	--
Aggression	++	++	++	Absent	--	--
Overactivity	++	++	+++	Absent	--	--
Savant skills	--	Absent	M	M	--	--
Age 1st steps	12	Typical	14	13	--	--
Odd gait	+	++	No	+	--	--
Gross or fine motor difficulties	--	F	GF	GF	--	--
Age urinary continence	--	54	39	DNA	--	--
Age bowel continence	--	54	51	DNA	--	--

Table 5-5. Description of affected individuals with paternally inherited S11. **(See legend page 107.)**

	IOW 2368-4	AGR 128-4	AGR 210-4	CHI 1
Sex	M	F	M	M
Epilepsy	--	No	Possible	--
Age head circumference (yrs)	--	--	--	--
Head circumference (%ile)	--	--	--	--
Autism Phenotype	Strict	Unknown	Broad	Strict
ADOS classification	--	Not spectrum or autism	Not spectrum or autism	--
Age IQ test (yrs)	--	14	--	9
Verbal IQ	--	117	--	77
Non-verbal IQ	--	97	--	95
ADI-R				
Age ADI-R (yrs)	--	6	2	9
Classification	Autism	Broad Spectrum	Not Quite Autism	Autism
Social domain	--	3	23	24
Communication domain verbal	--	3		14
Communication domain non-verbal	--		11	
Restricted, repetitive behavior domain	--	6	2	5
Age 1st words	--	18	DNA	15
Age 1st phrases	--	21	DNA	Normal
Any language regression	--	No	--	--
Any regression of other skills	--	No	--	--
Anxiety	--	+	Absent	--
Aggression	--	Absent	--	--
Overactivity	--	Absent	Absent	--
Savant skills	--	MEM,R,C	--	--
Age 1st steps	--	11	--	--
Odd gait	--	No	++	--
Gross or fine motor difficulties	--	No	No	--
Age urinary continence	--	--	--	--
Age bowel continence	--	36	DNA	--

Table 5-6. Comparison of Autism Diagnostic Interview-Revised domain and subdomain scores for individuals with maternally inherited S11 and homozygous P11.

	S11 (maternal)/P11			P11/P11			ANOVA[1]		
Measure	Mean	SD	n	Mean	SD	n	F	df	*P*
Domain (WPS label[2])									
Qualitative Impairments in Reciprocal Social Interaction (A)	23.4	4.7	12	23.0	5.0	1448	0.08	1, 1457	0.78
Communication: Verbal (B)	18.0	5.5	6	17.4	3.9	916	0.16	1, 919	0.69
Communication: Non-verbal (B)	13.2	1.2	6	12.6	1.7	532	0.49	1, 535	0.49
Repetitive Behaviors and Stereotyped Patterns (C)	5.3	2.3	12	6.1	2.2	1448	2.04	1, 1457	0.15
Subdomain (WPS label)									
Failure to use nonverbal behaviors to regulate social interaction (A1)	4.4	1.4	12	4.3	1.6	1448	0.04	1, 1457	0.84
Failure to develop peer relationships (A2)	6.8	1.5	12	6.5	1.6	1448	0.32	1, 1457	0.57
Lack of shared enjoyment (A3)	4.8	1.5	12	5.1	1.3	1448	0.35	1, 1457	0.56
Lack of socioemotional reciprocity (A4)	7.4	2.1	12	7.2	2.1	1448	0.09	1, 1457	0.77
Lack of, or delay in, spoken language and failure to compensate through gesture (B1)	6.5	2.1	12	6.1	2.1	1448	0.35	1, 1457	0.56
Lack of varied spontaneous make-believe or social imitative play (B4)	4.8	1.6	12	5.1	1.2	1448	0.78	1, 1457	0.38
Relative failure to initiate or sustain conversational interchange (B2V)	3.7	0.8	7	3.6	0.8	887	0.08	1, 890	0.78
Stereotyped, repetitive or idiosyncratic speech (B3V)	5.7	1.5	6	4.1	1.9	916	4.13	1, 919	0.04
Encompassing preoccupation or circumscribed pattern of interest (C1)	1.0	1.3	12	1.5	1.3	1446	1.93	1, 1455	0.17
Apparently compulsive adherence to nonfunctional routines or rituals (C2)[3]	1.2	1.5	12	1.4	1.3	1352	0.10	1, 1360	0.75
Stereotyped and repetitive motor mannerisms (C3)	1.7	0.7	12	1.6	0.7	1447	0.27	1, 1456	0.60
Preoccupations with part of objects or non-functional elements of material (C4)	1.4	0.7	12	1.7	0.5	1448	4.52	1, 1457	0.03

[1] Test statistic for independent variable of genotype group (maternally inherited S11/P11 vs. P11/P11). Age was included in as a covariate in all ANOVAs, and was significant ($p<.05$) for all but the following domains and subdomains: BV, A2, A4, B1, B2V, and B3V. The maternally inherited S11 group (n = 12, M = 97 mos, SD = 45) did not significantly differ from the homozygous P11 group (n = 1448, M = 95, SD = 57) in age, $t(1458) = 0.15$, $P = 0.89$.

[2] WPS label refers to the abbreviation of the domain or subdomain on the scoring algorithm of the Western Psychological Services (2003) version of the Autism Diagnostic Interview-Revised.

3 Some individuals were missing data for this subdomain.

Functional Analysis of the P11S Variant

To identify the molecular defect underlying the genetic association, a P11 and S11-encoding human β3 subunit cDNAs were engineered for *in vitro* studies in HEK293T cells. β3 subunits are incorporated into hetero-pentameric complexes at the cell surface (e.g. synapses), and the most common combination involves co-assembly of β3 with γ2 and α3 subunits during development and/or α1 subunits in the adult brain. Therefore, I co-expressed both wild-type and mutant β3 subunits with either α1 or α3 and γ2S subunits. The γ2S subunit was used instead of the γ2L subunit since it is much more abundant in the brain. To mimic the heterozygous condition seen in patients, equal amounts of wild-type β3(P11) and mutant β3(S11) subunits were co-transfected with α3 and γ2S subunits (Figure 5-3a and 5-3b). Compared with wild-type α3β3γ2S receptors, mutant receptors displayed reduced peak current amplitudes with either mixed β3(P11)/β3(S11) subunit expression or with only mutant β3(S11) subunit expression, and the amplitude reduction was greater with expression of only mutant β3(S11) subunits than with the mixed condition. Reduced receptor function in the context of the more developmentally relevant α3β3γ2S subunit combination is consistent with recent findings by Tanaka and colleagues [300].

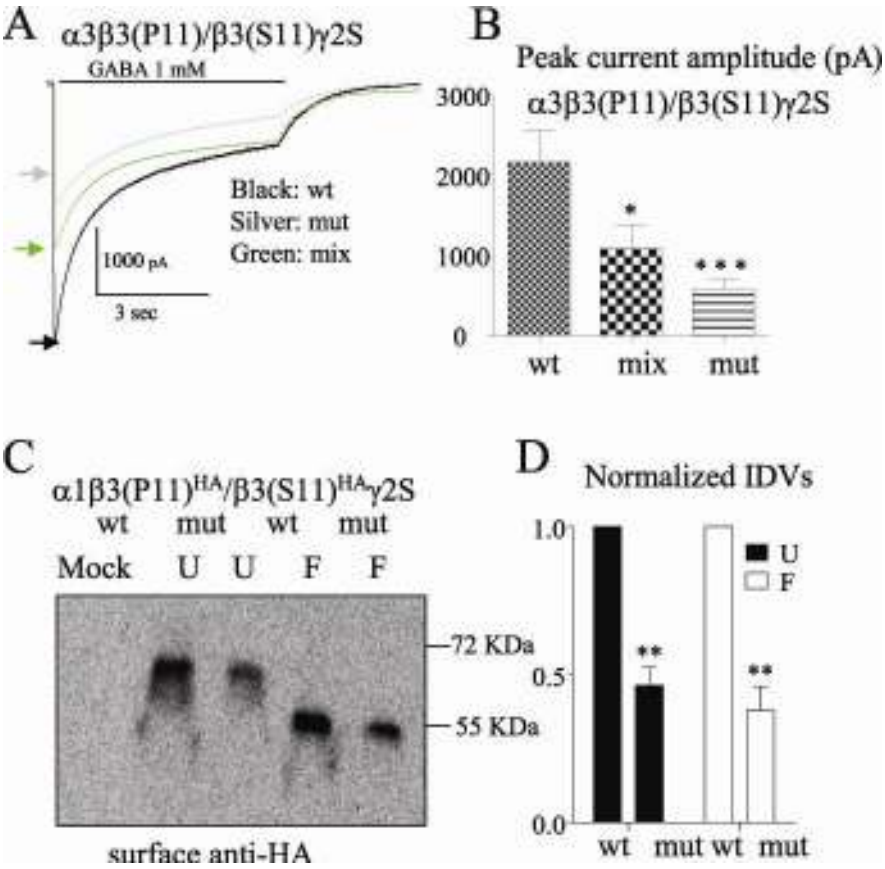


Figure 5-3 - Mutant β3(S11) subunit harboring receptors had reduced current and subunit surface expression. (A) Human $GABA_A$ receptor currents were obtained from HEK 293T cells co-transfected with α3 and γ2S subunit cDNAs and wild-type β3(P11) and the mutant β3(S11) subunit for wild-type (wt α3β3γ2S 1:1:1 cDNA ratio, black), mixed of the wild-type β3(P11) and mutant β3(S11) (1:0.5:0.5:1, mix, green) or for mutant (mut α3β3(S11)γ2S) and evoked with 1 mM GABA for 6 sec (A). In A arrows indicates the peak of each actual trace. (B) The mean peak amplitude of each group was plotted (n = 10 for wt, n = 15 for mix, n = 13 for mutant from three different transfections). (C) HEK 293T cells co-transfected with α1, $β3^{HA}$ (wt) or $β3(P11S)^{HA}$ (mut) and γ2S subunit cDNAs. Equal amounts of membrane-bound protein from live cells cell biotinylation, were pulled down with immobilized streptavidin, eluted with 1X NEB glycoprotein protein denaturing buffer (5% SDS, 0.4 M DTT) at room temperature for 30 min. The eluted products were then incubated in absence (U) or presence of PNGAse F (F) for 1hr at 37°C before fractionated by 10% SDS-PAGE and probed with monoclonal anti-HA antibody. (D) The relative amount of surface $β3^{HA}$ subunit protein of wild-type and mutant receptors from C was plotted (n = 4). In B and D, the data were plotted as mean ± SEM, **P* < 0.05, ***P* < 0.01, ****P* < 0.001 vs. wild-type).

The reduction of mutant receptor channel current could be due to a reduced number of functional receptors. I expressed α1 and γ2S subunits with β3(P11)HA and β3(S11)HA subunits and used surface biotinylation to determine the relative expression of wild-type and mutant β3 subunits. Cell surface membrane proteins were isolated following biotinylation of live cells expressing wild-type α1β3(P11)HAγ2S receptors or mutant α1β3(S11)HAγ2S receptors and treated with PNGase F to remove all carbohydrates attached in the ER and trans-Golgi. Products were then fractionated by SDS-PAGE and immunoblotted with anti-HA antibody. Both mutant β3(S11)HA and wild-type β3(P11)HA subunits on the cell surface appeared to the same size both before and after PNGase treatment, indicating likely signal peptide cleavage for β3(S11) subunits (Figure 5-3C) as for wild-type subunits. Confirmation of signal peptide cleavage, however, will require verification by protein sequencing. While mutant β3(S11) subunits were trafficked to the cell surface, when compared to wild-type β3(P11)HA subunits, mutant $\beta3(S11)^{HA}$ subunit levels were reduced both before (U) (1 vs. 0.4637 ± 0.062; $P < 0.0017$) and after PNGase F treatment (F) (1 vs. 0.3799; $P < 0.0092$) (Figure 5-3D).

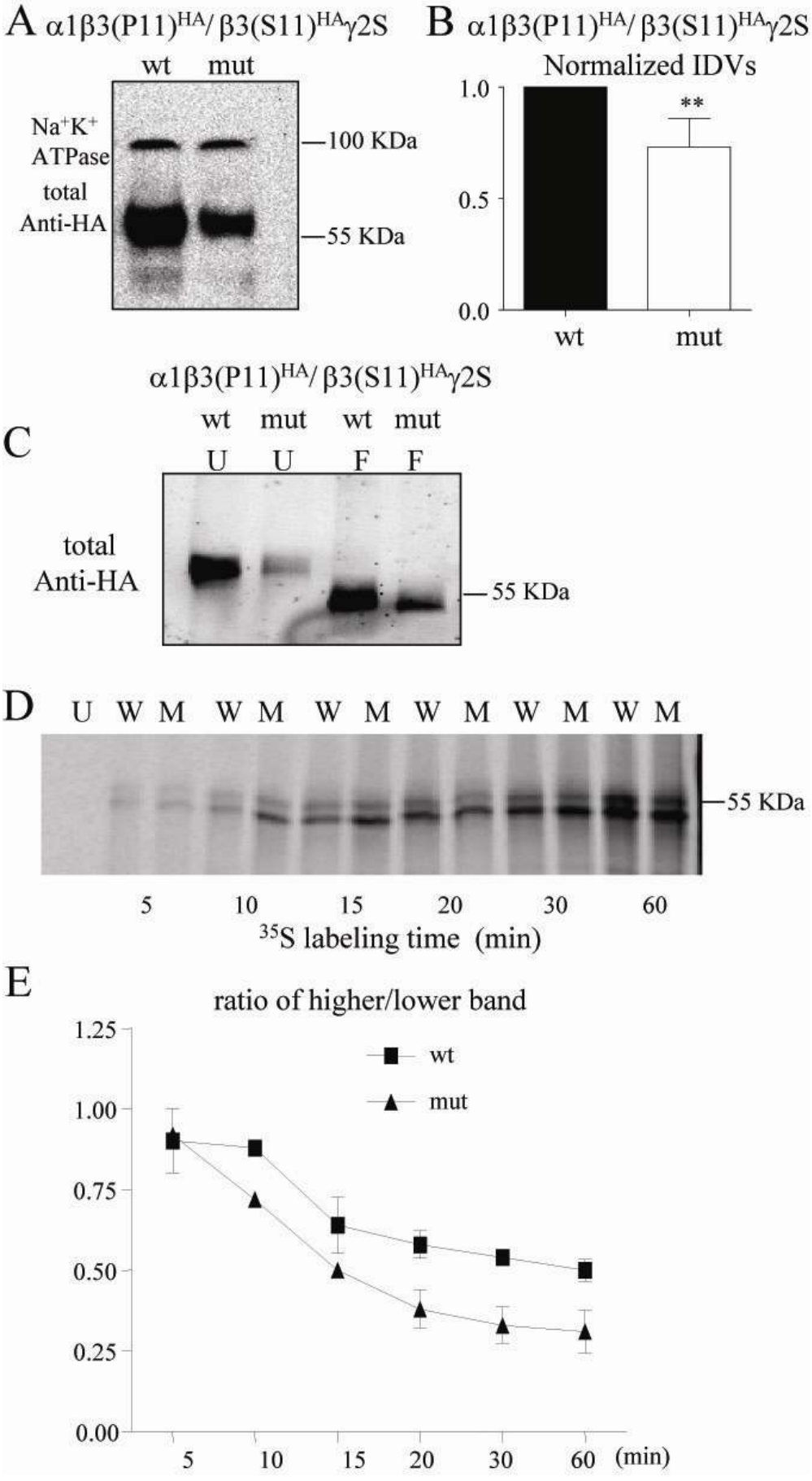
Our data are consistent with a molecular defect produced by the P11S substitution that occurs at a post-translational level, with mutant subunits exhibiting abnormal intracellular processing. Wild-type β3(P11)HA or mutant β3(S11)HA subunits were co-expressed in HEK293T cells with α1 and γ2S subunits. Compared to wild-type β3(P11)HA subunits, mutant β3(S11)HA subunits had a small but significant reduction in protein intensity (1 vs. 0.7316 ±

0.1270; *P* = 0.0012, N = 7). Although surface β3(S11) subunits appeared to have the same molecular mass as wild-type β3(P11) subunits, it was possible that if the signal peptides attached to mutant β3(S11) subunits were not cleaved, the subunits might not fold and oligomerize properly, and thus, would be retained inside cells instead of trafficking to the surface. I thus determined the molecular mass of mutant β3(S11) subunit from total cell lysates. Total wild-type (α1β3(P11)HAγ2S) or mutant (α1β3(S11)HAγ2S) receptors were digested with PNGase-F, which removes all carbohydrates attached in both ER and trans-Golgi regions, Without treatment (U), both wild-type β3(P11)HA and the mutant β3(S11)HA subunits migrated in a main band at ~58-60 kDa, although a faint band with a lower molecular mass representing different glycosylation form was also observed. With PNGaseF (F) treatment, both the wild-type β3(P11)HA and mutant β3(S11)HA subunits migrated in a single main band with the same molecular mass about 52 KDa consistent with previous reports [301,302], suggesting that the majority of wild-type and mutant β3 subunits had successful signal peptide cleavage (Figure 5-4C). I also used 35S methionine metabolic labeling to characterize biogenesis of wild-type β3FLAG and mutant β3(S11) FLAG subunits. When expressed alone, both wild-type and mutant subunits migrated in two bands, with the higher molecular mass band representing a more mature form and the lower band representing a less mature form (Figure 5-4D). Compared to wild-type subunits, the ratio of the higher molecular mass band to the lower band was lower for mutant subunits as soon as 10 min after translation (Figure 5-4E). The same molecular mass of wild-type β3(P11) and the mutant

β3(S11) subunits after PNGase F digestion suggested that the signal peptides were both cleaved. The slightly reduced total amount of mutant β3 (S11) subunit protein suggested that mutant subunits were not as stable as wild-type subunits.

Figure 5-4 – (next page) Mutant β3(P11S) subunit protein had impaired intracellular processing.

(A) HEK 293T cells co-transfected with α1 and γ2S subunits with HA tagged $\beta3(P11)^{HA}$ (Wt) or $\beta3(S11)^{HA}$ (mut) subunit cDNAs. Equal amounts of total lysates protein were analyzed by 10% SDS-PAGE and probed with monoclonal anti-HA and with monoclonal anti-Na^+ K^+ ATPase antibody as internal loading control. **(B)** The relative amount of surface $\beta3^{HA}$ subunit protein versus loading control of wild-type and mutant receptors from A was plotted (n = 7) **(C)** Equal amount of total cell lysates from A were undigested (U) or PNGase-F (F**)** at 37°C for 3 hr. **(D-E)** HEK 293T cells containing pulse-chase ^{35}S methionine radio-labeled wild-type $\beta3^{FLAG}$ (W) and mutant $\beta3(S11)^{FLAG}$ (M) subunits were pulse-labeled for a series of time points. The cells were lysed and the same amount of protein for each sample was used for immunopurification and SDS-PAGE (C). The relative ratio of radioactivity of the upper versus lower band is plotted at each time point for either the wild-type or mutant subunits (D, n = 4).


A α1β3(P11)HA/ β3(S11)HAγ2S
wt
mut
Na+K+
ATPase
total
Anti-HA
100 KDa
55 KDa
B α1β3(P11)HA/ β3(S11)HAγ2S
Normalized IDVs
1.0
0.5
0.0
**
wt
mut
α1β3(P11)HA/ β3(S11)HAγ2S
C
wt mut wt mut
U U F F
total
Anti-HA
55 KDa
D
U W M W M W M W M W M W M
55 KDa
5 10 15 20 30 60
35S labeling time (min)
E
ratio of higher/lower band
1.25
1.00
0.75
0.50
0.25
0.00
wt
mut
5 10 15 20 30 60 (min)

Part 2: Investigation of Common Allelic Association of *GABRB3* with Autism

Subjects and Methods

Subjects

The sample used for association analysis contained 961 families with evidence of ASDs. Families were excluded from analysis if there was a diagnosis or significant indicator of non-idiopathic autism (e.g. fragile-X syndrome, dysmorphic features, neonatal trauma). The majority of samples were obtained from the NIMH Center for Collaborative Genetic Studies on Mental Health Disorders. The balance were ascertained through the Tufts-Vanderbilt University consortium. In Autism was assessed by the Autism Diagnostic Interview (ADI or ADI-R), Autism Diagnostic Observation Schedule (ADOS) or both. For purposes of this study, those who met criteria for autism or ASD by either or both of these measures were considered “affected” in the association study. Of our families, 4% percent of families were of African-American origin, 7% of Latino origin, 4% of Asian origin, 83% of Caucasian origin, and 2% of unknown origin. In our sample, 81% of those with autism were male (Table 5-7).

Table 5-7. Characteristics of families genotyped by ascertainment center.

	ALL	AGRE	IOW	STA	TUF	VAN
Families	961	538	90	128	113	92
Individuals	4039	2524	377	515	401	222
Number with autism	1629	996	158	243	157	75
Female	305	207	18	44	25	11
Male	1324	789	140	199	132	64
Type of families:						
Simplex	298	131	27	22	55	63
Multiplex	663	407	63	106	58	29

AGRE=Autism Genetics Resource Exchange, IOW=Iowa, STA=Stanford, TUF=Tufts, VAN=Vanderbilt

Tag SNP Selection

Tag SNPs were selected from markers spanning all 10 exons of *GABRB3* using HapMap data and Haploview's implementation of Tagger (http://www.broad.mit.edu/mpg/haploview/) (Table 5-8 and Figure 5-5). Given the primarily Caucasian makeup of our sample, 30 CEPH parent-offspring trios were used for the estimation of haplotype blocks, and tag SNPs were selected so that each had a minimum pairwise r^2 of 0.7 with other HapMap genotyped SNPs with MAF greater than 0.05.

Genotyping

Markers meeting criteria were ordered from Applied Biosystems (Applied Biosystem, Foster City, CA, USA) as Taqman Assays-on-Demand (AoDs). Taqman genotyping reactions were performed in a 5-µl volume in 384 well plates in accordance with manufacturer's recommendations. Cycling conditions included an initial denaturation at 95°C for 7 min, followed by 50 cycles of 92°C

for 15 s and 60°C for 1 min. Samples were analyzed using an ABI 7900HT (Applied Biosystems) Sequence Detection System. Genotypes were automatically called if there was 95% confidence by the software calling algorithm that the call was correct.

Quality control

Pairs of three control samples were included per 96-well sample plate; concordant genotypes were required for each of the plates to pass QC. The Pedcheck script was used to check for Mendelian inconsistencies [223]. Genotyping efficiencies for all 60 markers were between 95 and 99 percent. Tests for conformity of the data to expectations under Hardy-Weinberg equilibrium (HWE) P-values were conducted for each SNP using Haploview [222].

LD and transmission analysis

Pair-wise linkage disequilibrium (r^2) was calculated for each SNP pair using Haploview [222]. FBAT was employed to test for association by considering transmissions to probands [263]. We carried out analysis in four groupings of our dataset: the overall dataset, the AGRE family dataset, the non-AGRE family dataset, and a “replication” dataset which consisted of the entire dataset with the 189 families previously genotyped by our group in an earlier study removed [105].

Results

Association results for *GABRB3* SNPs with autism are shown in Table 5-8, which also lists the SNPs genotyped, position, alleles and FBAT P values detailed. The Family Based Association Test with the –e option was used under an additive model. This test measures overtransmission of alleles from heterozygous parents to affected offspring and generates a statistic based on what would be observed under a null hypothesis of no association. This was done on four groups: the overall dataset (ALL), the AGRE sample (AGR), the non-AGRE sample (NONAGR) and the independent replication sample (REPLICATION). Those SNPs that were previously significant in our previous study (McCauley et al.) are highlighted in purple [105]. SNPs significant in recent genotyping by Curran et al. are highlighted in light blue [107]. SNPs flanking the 155CA-2 microsatellite are highlighted in green [106,283]. Since only two SNPs (14 and 23) met significance levels of $P<0.05$ in the overall dataset, eight SNPs in the AGR only set (10,12-14,27, 31, 32, and 55), no SNPs in the NONAGR set, and one SNP in the ALLEXJAKE set (23), I decided that I would pursue these SNPs further. In addition, since four of these nine SNPs fall within the first haplotype block (SNPs 2-14) [see Figure 6 for LD), in the 3´ terminal of *GABRB3* containing all of the *GABRB3* terminal exons 4-9, it was decided that the SNPs composing this block would be a focus of closer investigation.

Table 5-8. *GABRB3* association results in ALL families, AGRE, non-AGRE, and the Replication dataset.

						FBATA p -e SPEC			
No.	MARKER	Position	MAF	Alleles	Block	ALL	AGR	NONAGR	REPLICATION
1	rs8025575	24339176		C G	1	0.087	0.111	0.462	0.387
2	rs2081648	24349292		T C	1	0.983	0.931	0.930	0.732
3	rs1432007	24361782		A G	1	0.333	0.959	0.124	0.798
4	52615	24363873		C T	1	0.181	0.369	0.319	0.155
5	rs7173108	24364412		G A		0.755	0.857	0.766	0.976
6	rs1426217	24372218		G A		0.872	0.784	0.906	0.902
7	rs10519563	24374497		C G		0.579	0.329	0.653	0.514
8	rs11633705	24383984		G C		0.137	0.081	0.942	0.126
9	rs8037653	24395693		T G		0.806	0.832	0.510	0.976
10	rs4906683	24405714		A G		0.074	0.041	0.848	0.156
11	rs1897356	24416628		T C		0.821	0.939	0.774	0.532
12	rs2873027	24418502		T C		0.064	0.029	0.955	0.195
13	rs4542636	24419024		C T		0.056	0.029	0.898	0.166
14	rs768899	24428917		C T		0.021	0.024	0.439	0.075
15	rs754186	24438907		C T		0.626	0.632	0.872	0.603
16	rs11161324	24440090		A G		0.675	0.134	0.203	0.533
17	rs11636966	24444121		C T		0.594	0.106	0.217	0.516
18	rs12442543	24444672		G T		0.197	0.174	0.779	0.180
19	rs11629819	24446174		C A		0.721	0.521	0.157	0.828
20	rs2194958	24458502		T C		0.320	0.794	0.189	0.347
21	rs1863463	24463613		G C		0.396	0.511	0.591	0.581
22	rs17646800	24463770		T C		0.888	0.516	0.370	0.923
23	rs1863455	24469262		T C		0.026	0.078	0.174	0.013
24	rs17561112	24473371		A G		0.163	0.397	0.238	0.267
25	rs7494844	24475173		T C		0.592	0.832	0.312	0.415
26	rs11161329	24477803		A T		0.918	0.610	0.375	0.647
27	rs10873637	24477858		C A		0.150	0.048	0.789	0.217
28	rs11634050	24479765		C T		0.152	0.260	0.373	0.134
29	rs878960	24480029		T C		0.654	0.741	0.263	0.897
30	rs878959	24480389		C A		0.515	0.702	0.092	0.419
31	rs878961	24480541		A G		0.214	0.044	0.222	0.138
32	rs7178595	24482695		T C		0.090	0.036	0.981	0.085
33	rs4906687	24483273		A T		0.532	0.623	0.165	0.355
34	rs12593415	24484321		G C		0.888	0.521	0.512	0.960
35	rs1863464	24489581		G A		0.184	0.663	0.135	0.076
36	rs11631421	24502431		C T		0.982	0.977	0.941	0.885
37	rs981778	24508333		G A		0.753	0.920	0.700	0.486
38	rs11632969	24509870		G A		0.936	0.955	0.819	0.882
39	rs2114217	24511122		A T		0.545	0.477	0.991	0.601
40	rs11630462	24511505		A C		0.601	0.731	0.678	0.337
41	rs17561473	24512954		T C		0.162	0.404	0.215	0.341
42	rs12592816	24516890		C G		0.888	0.897	0.955	0.801
43	rs737098	24518263		A G		0.887	0.232	0.203	0.834
44	rs7179279	24532939		G C		0.580	0.943	0.374	0.805
45	rs11629495	24547101		C T		0.657	0.203	0.298	0.922
46	rs17117335	24547219		C A		0.792	0.474	0.178	0.883
47	rs7178713	24553979		A T		0.703	0.853	0.382	0.397
48	rs4453447	24554565		G C		0.798	0.964	0.693	0.981
49	rs8041610	24559357		A C		0.528	0.738	0.127	0.388
50	rs3212343	24560619		G A		0.237	0.226	0.759	0.271
51	rs3212337	24562204		G A		0.279	0.238	0.791	0.108
52	rs3212334	24563936		T G		0.848	0.657	0.314	0.917
53	rs3212331	24565862		A G		0.408	0.380	0.862	0.797
54	rs25409	24569934		G A		.	.	.	.
55	rs20317	24570030		C G		0.069	0.026	0.982	0.191
56	rs4906902	24570861		A G		0.603	0.558	0.947	0.238
57	rs7497801	24584480		G A		0.995	0.779	0.717	0.970
58	rs12440086	24589585		A C		0.438	0.214	0.702	0.833
59	rs35586628	24683233		T C		0.580	0.889	0.276	0.967
60	rs9745027			T G		0.721	0.474	0.723	0.764

McCauley et. al.
Curran et. al.
155CA-2

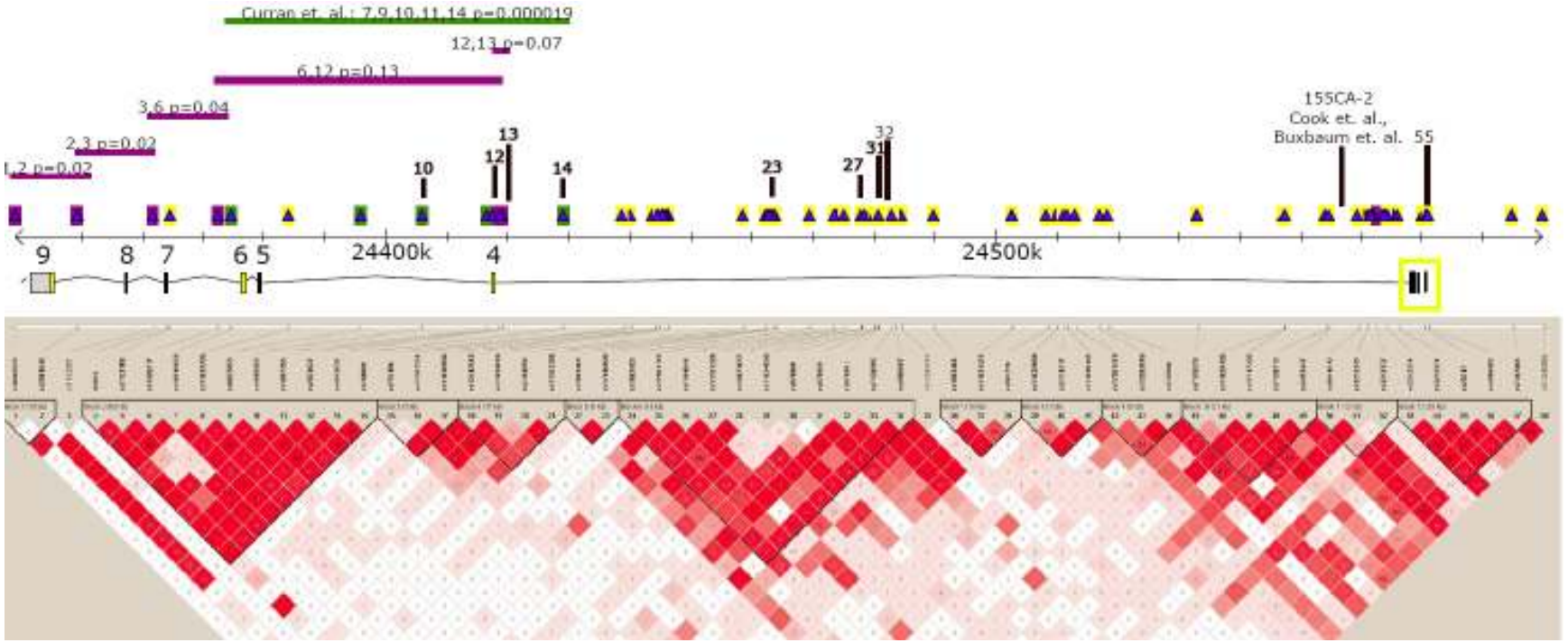


Figure 5-5. *GABRB3* locus, genomic structure, and genotyping markers. 60 markers were genotyped over *GABRB3*, 58 of which are pictured above (2 others are far 5´ near GABRA5). The region is characterized by low levels of linkage disequilibrium which is indicated by r^2 values. The first 14 SNPs in the gene show high levels of LD and are the site of earlier association by our group in a set of 189 families with significant two-locus haplotypes are shown in purple. Later work by Curran et. al found a 5 marker haplotype highly significant in a heterogeneous cohort and this is indicated in green. Two studies by Cook et al. and Buxbaum et al. also showed microsatellite 155CA-2 between two 5´ SNPs significant. In the figure above, if r^2 is equal to 1, the block blocks are red and there is no numeric value given in the square. If r^2 is less than 1, and confidence in D´ is high, (as measured by a LOD>2), red squares will give their value of r2 (times 100). SNPs significant by FBAT –e analysis are indicated by numbers and black lines.

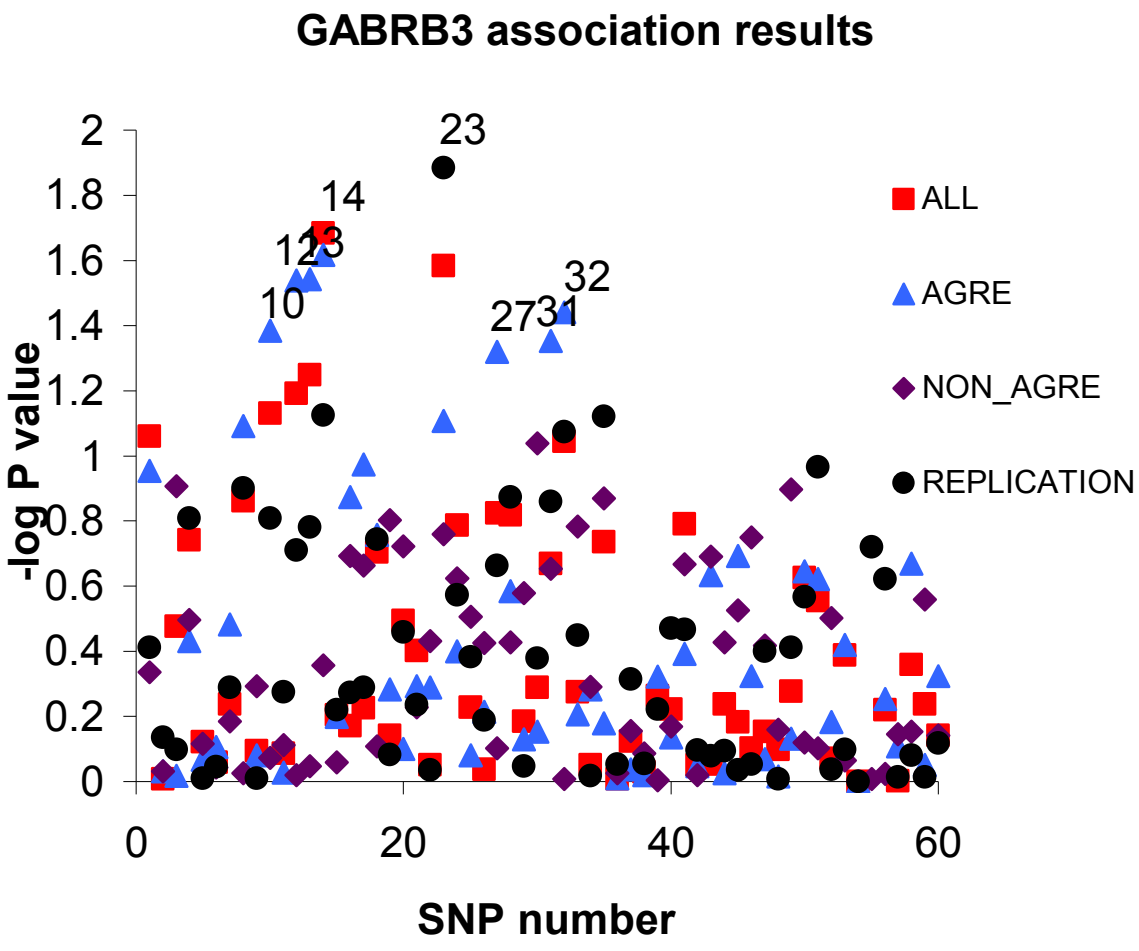


Figure 5-6 Plot of FBAT –e log10 P values for overtransmitted alleles. 60 SNPs genotyped in *GABRB3* are shown under the FBAT –e test. The –e option estimates the empirical variance and is more reliable when linkage has been established and provides a more conservative estimate of association. Results here are shown for all SNPs under the additive model. SNPs indicated above with numbers (see table 5-8 for marker names) showed nominal evidence of transmission ($P < 0.05$) in one of four subsets of the data. A Haplotype block which includes SNPs 1-14 showed prior evidence of transmission bias in an earlier study. Only SNPs in this block and those indicated by number outside of the block were selected for further analysis.

The SNPs in the first block (1-14) and SNPs 23, 27, 31, 32, and 55 listed above as significant in one of the four groups were investigated in the same four groups more closely under a dominant model. The results are shown below in figure 5-7.

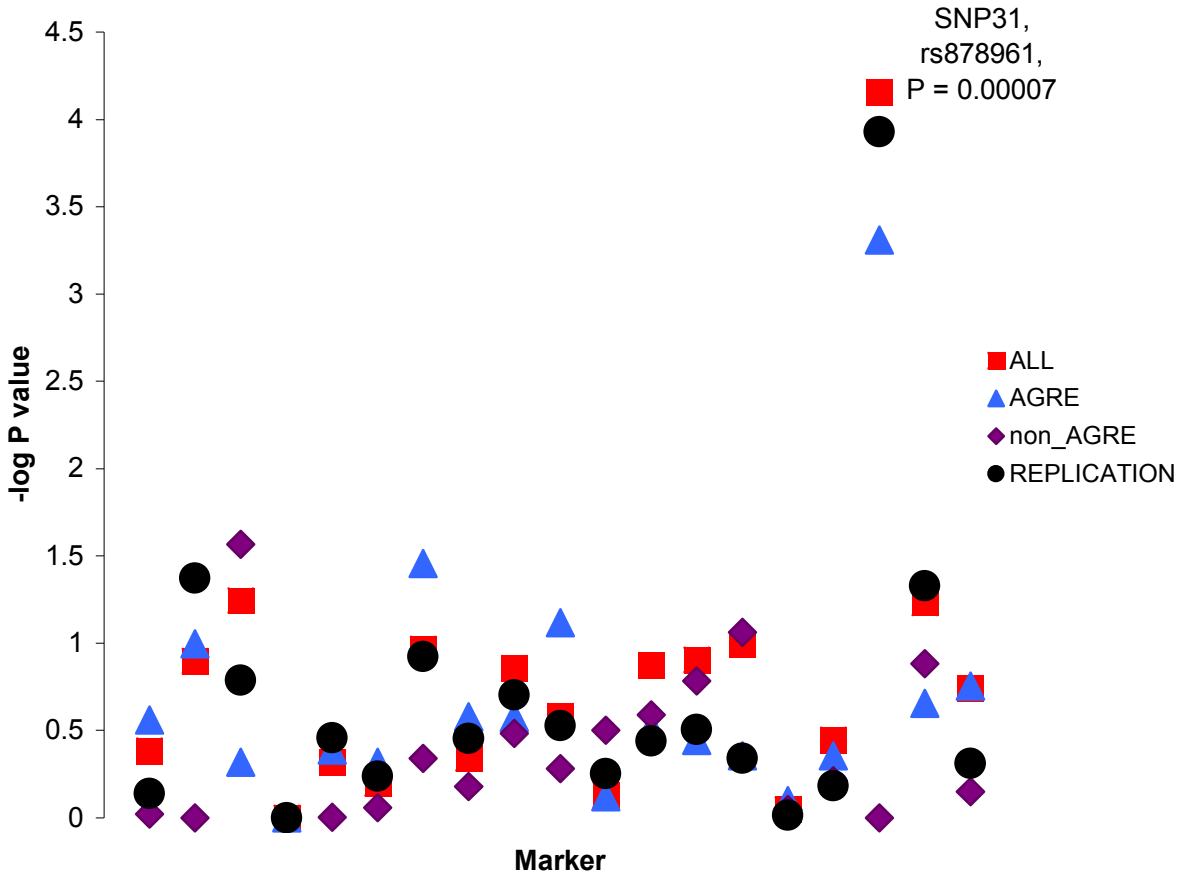


Figure 5-7 Plot of FBAT –e log10 P values under a dominant model. The 14 SNPs in the first haplotype block, as well as SNPs 23, 27, 31, 32, and 55 which showed evidence of association under the additive model were analyzed for transmission disequilibrium using the FBAT –e option under a dominant model. As can be seen, one SNP, SNP 31 showed strong evidence of association in the overall, AGRE, and replication dataset under the dominant model.

Discussion

Rare Variant P11S and its Role in Autism

I have shown that a rare coding variant of the *GABRB3* gene is associated with autism when transmitted maternally. Statistical association is explained by an intracellular processing defect imparted by the N-terminal P11S substitution in the signal peptide. While the S11-encoded signal peptide is cleaved, molecular evidence indicates that the S11 signal peptide defect results in reduced surface expression and resultant decreased receptor current, probably due to abnormal intracellular processing, intracellular retention and fast degradation of the mutant subunit. I suggest that impaired GABAergic signaling during a critical time window in brain development could lead to both autism and CAE, and possibly other more subtle neurological phenotypes.

A problem inherent in the study of rare variants is the difficulty in amassing a sufficiently large clinical sample to detect risk effects in the absence of full penetrance. Such is the case for P11S. While maternal association is significant, it is nevertheless a modestly-significant statistical finding for an allelic association. It is important therefore to consider these results in the context of recent and emerging rare variant findings in autism. By and large, these take the form of CNVs, and there is a growing list of genes or regions (e.g. *NRXN1*, *SHANK3*, 16p11.2) identified as “autism loci” based upon observed *de novo* variants, but for which other cases show *inherited* gene disrupting variants [61,101,111,128,135]. Such cases reveal incomplete penetrance of such (e.g.) loss-of-function mutations for a given gene. Indeed families may show

unexpected segregation patterns with only one of two affected sibs receiving the variant or both an affected and unaffected sib inheriting the variant. The pattern is much the same for P11S. The fact that even maternal duplication of the *GABRB3* containing 15q11-q13 region does not always manifest with autism, along with the complex epigenetic regulation of the genes in the region, may provide a lens through which to view a potential mechanism [279,290].

GABRB3 is known to play a major role in development of the CNS, being the major β isoform in a number of regions in prenatal and neonatal brain [285,286,289]. The result of *Gabrb3* disruption in the mouse is a useful comparison, given the effect of P11S substitution in families segregating autism and epilepsy. In the 10% of null animals that survive to maturity, epilepsy and hypersensitivity and various other behavioral and physiological abnormalities are noted [290]. Null mice produce fewer functional $GABA_A$ receptors, and pharmacological evidence indicates that other β subunits do not compensate for the absence of β3 [291]. The *Gabrb3* knock-out has been proposed as an autism model because of reduced social and exploratory behaviors and a tendency for diminished nurturing behaviors.

In light of the epigenetic regulation of 15q11-q13 genes, and that it is typically maternal, but not paternal duplication of the interval leads to autism, the S11 association is provocative. Maternal bias implicates *UBE3A* in the dup(15) autism phenotype, however, a contiguous gene effect including *GABRB3* is very likely [303]. Supporting potential contribution of *GABRB3* in causing dup(15) autism, several genetic and epigenetic studies have independently implicated the

GABRB3 gene in autism in the absence of duplication [104,105,106,107,194,198,283]. While genetic studies of common allele effects are largely positive, the data suggest allelic heterogeneity, with associations seen in two or more regions at the *GABRB3* locus [105,106,107,283]. Maternal inheritance has also been noted in genetic epilepsies [304,305], and given the maternal specificity of dup(15) autism [119,279], we speculate that epigenetic effects at *GABRB3* may tilt the balance to risk for autism when the origin of the S11 variant is maternal. Other studies, including one by Nakatani et al. indicate that duplication the syntentic region in mouse is associated with traits common in autism, though they observe this paternally and acknowledge that some epigenetic controls may be different between human and mouse [306]. Further, theirs is a contiguous gene duplication effect where he I report a single point mutation in just one of the genes of the region.

I also speculate that the presence of the mutation in both mother and fetus may lead to greater impact at some key developmental stage. To formally test for maternal genotypic effects, analytic approaches must assess any increased association due to the sex of the non-transmitting parent. This would be consistent with maternal genotype-created environment interacting with fetal genotype to increase risk even more in offspring, as opposed to a classic parent-of-origin effect. If S11 effects on maternal environment are important, I would expect to see increased risk in S11-carrier offspring of carrier mothers, relative to offspring of non-carrier mothers (and carrier fathers). Tests of maternal and parent-of-origin effects have been explored in this sample, allowing us to

establish a parent-of-origin effect; I were, however, unable to conduct an adequately powered test of whether there are effects of maternal genotype on offspring risk of autism in addition to established parent-of-origin effects.

Epigenetic and expression studies of *GABRB3* provide some context for considering the observed association, although the picture vis-à-vis parent-of-origin effects on gene expression is far from clear From their analysis of cerebral cortex (BA9) brain samples, Samaco and colleagues reported that levels of *GABRB3* protein were significantly reduced in five of nine autism samples but not in controls [198]. Another study by this group revealed monoallelic expression of *GABRB3* in one of seven autism frontal cortex samples informative for an exon 1a SNP, and evidence from deletion PWS and AS samples suggesting the allelic bias favors paternal expression [194]. These reports cannot tell us, however, whether similar effects might occur during fetal development and in other, perhaps more phenotypically-relevant, brain regions. Nevertheless, the possibility that imprinting or allelic bias in gene expression might, like with the serotonin 2A receptor gene (*HTR2A*) be a polymorphic trait, leads us to speculate that incomplete penetrance of S11 relates to inter-individual differences in gene expression and/or epigenetic regulation of the region [307].

Spatiotemporal expression of $GABA_A$ receptor subunits is controlled through the use of alternative initiating exons. The P11S variant is in exon 1a, which like exon 1, expresses an alternative signal peptide. In adult, exon 1-containing transcripts are more abundant than those containing exon 1a. However, the relative abundance of exon 1a transcripts is significantly higher in

fetal brain than in adult hippocampus [298]. Thus, an exon 1a variant is more likely to exert its effects developmentally, prior to the shift in balance of exon 1a to exon1 transcripts. Indeed, age effects are common in epilepsy, with the majority of CAE patients having absence seizures that lessen or abate completely with age [308].

In our phenotype analysis, we asked initially whether or not epilepsy was associated with the S11 variant, given the observation of S11 segregating in families with CAE[300]. The rate was higher in P11S carriers (14.3%) compared with P11 homozygotes (5.2%), but not greater than overall estimates of epilepsy in autism (5-38.3%) [309]. Aside from ascertainment biases, the broad range of epilepsy estimates may reflect the inherent difficulty in identifying more subtle epilepsy and/or EEG abnormalities in autism. The S11 variant displays incomplete penetrance in both autism and published CAE families, and thus milder phenotypes may be present in S11 carriers [310]. As functional data indicate that S11 reduces, but does not eliminate, functional $GABA_A$ receptors, this would not be surprising.

Molecular defects produced by the P11S substitution in *GABRB3* likely include impaired GABAergic signaling and cellular homeostasis. The present study provides a direct link between $GABA_A$ receptor β3 subunit dysfunction and autism, and explains the observed association with epilepsy. Precedent already exists for $GABA_A$ receptor subunit dysfunction leading to epilepsy, with mutations in the γ2 subunit also associated with CAE [296]. Given the nature of its

dysfunction, association of this specific, albeit infrequent, allele with autism and epilepsy suggests a potential for individualized treatment in these cases.

In summary, we report (1) the first example of a $GABA_A$ receptor subunit gene coding variant statistically associated with autism; (2) the first signal peptide mutation associated with autism; and (3) the first evidence for maternal over-transmission of a coding variant within this known imprinted, autism-associated region. Since *maternal* duplication of 15q11-q13 is the most frequent chromosomal and copy number abnormality known to cause autism, it is provocative that *maternal* over-transmission of the more discrete *GABRB3* S11 variant may increase risk for autism and epilepsy. These findings complement other common-allele linkage and association studies and provide further support for involvement of *GABRB3* in autism[105,106,107,185,187,283,284]. It appears then that allelic heterogeneity at *GABRB3*, both common and P11S, act as genetic risk factors. We predict for autism, as found for other complex diseases, that rare variants possessing greater effect sizes will emerge in genes showing association of common alleles conferring more modest effects [88,311].

Common Variant Association with Autism

Despite the relatively high levels of coverage employed by our tagging strategy, we failed to demonstrate significant association in our region of prior association at the 3´ end of *GABRB3* in SNPs 1-14. Only SNP 7 showed evidence of association in the AGRE dataset under the dominant model. SNPs 10,12, 13, and 14 showed significance in the AGRE dataset under the additive

model, but not in the dominant model. The one SNP which appeared associated in the AGRE model for which significance dramatically improved under the dominant model was SNP 31, rs878961.

Results for SNP31 were most significant under a dominant model, which is of importance since the most significant result, this SNP31, has a minor allele frequency of just 5%. This means that a single copy of the SNP significantly increases risk for autism. This SNP was found significant in the overall dataset, AGRE dataset, and replication dataset, that is the dataset with the original McCauley et al. families removed. The highly significant P-value is the only to survive correction for multiple testing. It may be that we have detected a new signal in GABRB3.

The meaning of this finding is hard to interpret. *GABRB3* has a very large intron which spans over 150kb of the 230kb transcript. RS878961 lies almost precisely in the middle of this massive intron in a region of very low levels of conservation and linkage disequilibrium. Nearby are repeat structures, but little evidence for other functional regions of DNA and the block of linkage disequilibrium is quite short. There are no pre-computed siRNA binding sites apparent here. Further, this variant is more than 80kb from the 155CA-2 microsatellite and over 50kb from SNPs found significant in the studies by Curran et al. and McCauley et. al [105,106,107]. This is often the case in so many association studies where the variant associated is intronic and the biological meaning is thus difficult to interpret.

Since this variant is in a region that, from review of the literature, appears to be untagged by previous association studies, it will be interesting to see if other groups replicate this signal in independent samples.

The finding of modest association in the 3´ haplotype block restricted to the AGRE set appears to be the same finding reported by Curran et al. and is most likely a phenomenon of multiplex families. The AGRE set, by its very nature, is predominantly multiplex families. When our initial sample from the McCauley et al. paper is removed, we find four SNPs 2, 7, 13, and 14 in the 3´ haplotype block remain significant (dominant not additive model). This seems to indicate that the effect in this block is being driven either A) by the AGRE families or B) by multiplex families. Whatever the case, the effect size appears to be modest.

Previously Curran et al. showed association of a haplotype consisting of our SNPs 7-9-10-11-14 which was overtransmitted to affected offspring with a P-value of 0.000019 [107]. Our test for overtransmission of the same haplotype indicated this was not a significant overtransmission in our own dataset (P = 0.8097) using the PLINK haplotype TDT test for the same five SNPs. The sample used by Curran et al. was highly heterogeneous including only 55% meeting criteria for autism and the rest consisting of Asperger syndrome and PDD-NOS. In addition, the subjects in their sample had widely varying IQs.

Here I have undertaken the task of exhaustively indexing all major alleles in the *GABRB3* locus. It is reasonable to conclude from the above studies that there is little evidence for replication of the association in the 3´ haplotype block

in which McCauley et al. and Curran et al. observed prior association [105,107]. There was no evidence of association near the 5´ end of the gene in the vicinity of microsatellite 155CA-2. However, here we have shown an additional SNP, rs878961, with a low minor allele frequency and reasonably strong evidence for association with *GABRB3* under a dominant model. While interpretation of the meaning of the association of this intronic SNP is difficult and the finding of a new signal in *GABRB3* is more frustrating than exciting, it will be interesting to see if the association is replicated in other datasets.

CHAPTER VI

ANALYSIS OF SMALLER CNVS WITHIN *GABRB3* AND *UBE3A*

Introduction

As understanding of autism has evolved and the available technology matured, it has become increasingly clear that much of the disorder is genomic in nature. Copy number variation (CNV) has emerged in the past few years, driven largely by technological advances in genotyping methods which capture not only SNP genotype, but also intensity information at adjacent SNPs throughout the genome. In 2004, while the vast majority of the genetics world was focused on SNPs and ramping up for the coming explosion of genome wide association studies, two papers emerged suggesting copy number variation (CNV) may play a larger role in the genomic architecture of humans than previously appreciated [120,121]. The early HapMap data was mined looking for runs of Mendelian inconsistencies or runs of homozygosity, both of which could suggest CNVs [312]. And as Illumina and Affymetrix genotyping technologies matured allowing not only the collection of more genotypes, but also the analysis of intensity data for genotyped SNPs, an entire field quickly began to mature where CNVs now began to be more widely appreciated as having a role in a number of diseases previously unappreciated. By one estimate, copy number variable regions comprise approximately 12% of the genome [313].

(CNV) has emerged as a previously unknown mechanism, which appears to explain a minimum of 10% of diagnosed cases of autism [58,126]. One analysis of sporadic autism in trios showed a high rate of de-novo CNVs in 10% of 118 patients with autism compared to 1% of controls [58]. This was after the exclusion of those with severe MR, dysmorphologies, or other known cytogenetic abnormalities. Other studies by the Autism Genome Project and others identified NRXN1, SHANK3, 16p11, and other CNVs in ASDs [61,101,111,128,135].

The largest genome wide association study in autism to-date over 550,000 SNPs for 780 families consisting of 1204 cases and 6,491 controls showed very few SNPs reaching genome-wide significance, but CNV analysis strongly implicated the role of ubiquitin pathway genes in the pathogenesis of the disease [172,221]. One of the most commonly identified genes identified in the analysis was *UBE3A* [172].

Since 15q11-q13 contains *UBE3A* as well as *GABRB3*, the two candidate genes under study, and is the most frequent CNV seen in autism, it seemed reasonable to investigate both of these genes for potential submicroscopic CNVs that might impact these genes and their respective proteins. The 15q11-q13 region has very high levels of repetitive sequences and has been subject to a large number of genomic disorders, many of which involve new microdeletions and microduplications which manifest with autism, epilepsy, and schizophrenia and are frequently characterized by incomplete penetrance [140,141,314]. Partial duplication of the *GABRB3* gene has previously been reported in a patient, as well [315].

Validation of findings based on inferences made from genotyping intensity data, ROH analysis, or Mendelian inconsistencies is often labor intensive and can involve cytogenetic approaches, multiplex ligation probe amplification (MLPA), or quantitative polymerase chain reaction (Q-PCR).

The MLPA method was described by Schouten et al. in 2002 and is a method to specifically identify copy number in a large set of genomic targets [316]. It involves the placement of two probes adjacent to one another on genomic target sequence, followed by their ligation and then amplification by Q-PCR. Since this is done in multiplex and using multiple reference probes (as opposed to a single housekeeping gene) the method is an excellent way to test multiple sites in the genome for potential CNVs.

In the present study, we used the MLPA method, in combination with genotype data for *UBE3A* and *GABRB3* and phenotype data from the ADI-R identifying those with autism and epilepsy, to determine the extent to which small CNVs at these loci may be involved in autism or autism with epilepsy.

Subjects and Methods

Sample Selection

Samples were selected from a set from the 961 families genotyped for the *GABRB3* markers listed in chapter 5, table 5-7; and those genotyped for the *UBE3A* markers listed in chapter 4, table 4-2. In total, I chose 48 samples for MLPA analysis. I selected samples using three criteria: (1) ROH scores for *UBE3A* and *GABRB3*, (2) high levels of inconsistent genotypes, (3) epilepsy, (4)

IBD2 for SNPs in *GABRB3* and *UBE3A*. I excluded samples for which low genotyping efficiency was achieved at unrelated loci in addition to *UBE3A* and *GABRB3* loci.

ROH Analysis

ROH analysis used the –homozyg function of PLINK and examined both *UBE3A* SNPs independently, *GABRB3* SNPs independently, and both sets of SNPs together [225].

In this study, I used Taqman genotype data for the eight SNPs genotyped over *UBE3A* and 60 SNPs genotyped over *GABRB3* and subjected that data to ROH analysis to identify samples most likely to be homozygous at SNPs across one or both of these genes. SNPs that were homozygous across 50 or more SNPs (permitting up to two missing genotypes) across 50 or more SNPs in *UBE3A*+*GABRB3*, all 8 SNPs in *UBE3A*, or across 50 SNPs in *GABRB3* were selected. Since *UBE3A* exists in a single block of LD and hundreds of samples were homozygous for all 8 SNPs, this was less useful criteria. The SNPs genotyped were those genotyped in chapters 4 and 5. Table 6-2 lists those SNPs in *UBE3A* and *GABRB3* that were genotyped along with their genomic position. In total, these SNPs span 1.5Mb.

Genotype Inconsistency

Dirty flat files with Mendelian inconsistencies identified by the Pedcheck program were used to count up the number of Mendelian inconsistencies over

GABRB3 and *UBE3A* loci. This could indicate cases of UPD that might escape ROH analysis which tended to look at SNPs over larger distances.

Identification of Epilepsy

Since dozens of papers have been reported on a number of 15q13 microdeletions which result in epilepsy, I used samples with potential epilepsy in our analysis. Samples with likely diagnosis of epilepsy were identified based on available ADI items as described in the previous chapter 5. The two items of the ADI-R used were the answers to the following questions:

Has the subject ever fainted of had a fit/seizure/convulsion?

Has the subject ever had medicine to control fits?

The answers to both are 0-no attacks, 1-history suggesting epilepsy, 2-definite diagnosis of epilepsy, 7-febrile convulsions only. Any family with an individual who reported 1, 2, or 7 for either question was considered in one set with suspected epilepsy while those with 2 for both answers were also assigned to confirmed epilepsy. Those individuals in the “confirmed” category took precedence over those in the “suspected category”.

Synthesis

The top 48 samples by (1) ROH scores for *UBE3A* and *GABRB3*, (2) high levels of inconsistent genotypes, AND (3) epilepsy shown in table 6-1. I tried, when possible, to select individuals who met as many of the criteria cutoffs as possible and further those affected sib pairs who showed evidence of IBD2 and

concordance for diagnoses of autism and epilepsy. The rationale being that chromosome 15 is more likely a source of the genetic etiology of said subjects' disease in cases where sibs are identical by phenotype and genotype. I elected, but in rare cases, to genotype only one sib per family.

Table 6-1. Samples selected for MLPA analysis.

Combined ID	GABA+*UBE3A* homozyg	*UBE3A* homozyg	*GABRB3* homozyg	Incons	Broad Epilepsy	Narrow Epilepsy
AGR_138_4	X	X	X	17		
AGR_567_5	X	X	X		X	X
AGR_922_301	X	X	X		X	X
AGR_922_302	X	X	X		X	X
AGR_543_3	X	X	X		X	
AGR_293_3	X	X			X	
AGR_1072_301	X	X	X			
AGR_1072_303	X	X	X			
AGR_111_5	X	X	X			
AGR_1376_301	X	X	X			
AGR_1376_302	X	X	X			
AGR_356_3	X	X	X			
AGR_450_3	X	X	X			
AGR_835_5	X	X	X			
IOW_2843_5	X	X	X			
IOW_3505_3	X	X	X			
STA_447_7894	X	X	X			
TUF_1072_8	X	X	X			
TUF_6028_103	X	X	X			
VAN_3507_1	X	X	X			
VAN_3531_1	X	X	X			
AGR_1038_303	X	X				
AGR_1214_302				19	X	X
VAN_3599_101				17		
STA_235_4508				16		
AGR_240_3				15		
TUF_2004_3				15		
TUF_3057_4				13		
TUF_5012_1				13		
AGR_138_3				12		
AGR_959_301				11		
IOW_2836_4				11		
IOW_2836_5				10		
TUF_1092_4				9		
TUF_2004_1				9		
TUF_2010_1				9		
AGR_1270_301				8		
AGR_1323_301				8		
AGR_521_4				7	X	
AGR_862_302				4	X	X
AGR_548_5			X		X	X
AGR_520_4			X		X	X
AGR_939_303			X		X	X
IOW_2806_5			X		X	X
AGR_1249_302			X		X	
AGR_994_301			X		X	
AGR_1312_302			X			
AGR_164_4			X			

MLPA

Since MLPA is a relatively new method and validating its reliability was essential, I used CEPH control samples, Angelman syndrome samples known to be deleted for both *UBE3A* and *GABRB3*, as well as samples with duplication of chromosome 15q11-q13 as controls. All samples used were checked for DNA concentration by use of Nanodrop. The MLPA kit was a product TS-047 custom manufactured by MRC-Holland (Amsterdam, Netherlands) and contains 13 probes for 11 exons of *UBE3A*, 13 probes for 9 exons of *GABRB3* (excluding exon 1a) as well as 14 independent reference probes which target independent genomic loci. These are shown in table 6-2 along with the length of each amplicon, the target probe gene, the right probe oligo (RPO), the left probe oligo (LPO) the mapview coordinates of the probe junction, relative chromosome position, and targeted exon in the case of non-reference loci. The MLPA protocol was performed according to the manufacturer's protocol available in SALSA MLPA Kit TS-047. Probe amplification products were run on an ABI 3730 DNA Analyzer using the GS500 size standard (Applied Biosystem, Foster City, CA, USA).

Table 6-2. Probes for MLPA analysis of 9 *GABRB3* and 11 *UBE3A* exons as well as 14 independent reference probes for unrelated genomic loci.

length ampl.	probe	RPO	LPO	mapview	chr. pos.	exon
409	GABRB3	M-02038	LIG-02330	15-024.569103	15q12	1
226	GABRB3	M-10865	LIG-11535	15-024.568953	15q11	2
250	GABRB3	M-10866	LIG-11536	15-024.568896	15q11	2
355	GABRB3	M-10867	LIG-11537	15-024.568655	15q11	3
184	GABRB3	M-10868	LIG-11538	15-024.417665	15q11	4
274	GABRB3	M-10869	LIG-11539	15-024.417552	15q11	4
319	GABRB3	M-10870	LIG-11540	15-024.379640	15q11	5
148	GABRB3	M-10872	LIG-11542	15-024.376510	15q11	6
202	GABRB3	M-10873	LIG-11543	15-024.363971	15q11	7
382	GABRB3	M-10874	LIG-11544	15-024.363881	15q11	7
292	GABRB3	M-10875	LIG-11545	15-024.357380	15q11	8
220	GABRB3	M-01315	LIG-09339	15-024.344242	15q12	9
436	GABRB3	M-10876	LIG-11546	15-024.344314	15q11	9
166	UBE3A	M-10877	LIG-11547	15-023.204824	15q11	1
283	UBE3A	M-10879	LIG-11549	15-023.201527	15q11	2
373	UBE3A	M-10878	LIG-11548	15-023.201674	15q11	2
196	UBE3A	M-10880	LIG-11550	15-023.171919	15q11	3
160	UBE3A	M-04620	LIG-00863	15-023.167740	15q12	4
427	UBE3A	M-10881	LIG-11551	15-023.156677	15q11	5
142	UBE3A	M-10883	LIG-11553	15-023.152935	15q11	6
178	UBE3A	M-10882	LIG-11552	15-023.153027	15q11	6
310	UBE3A	M-10884	LIG-11554	15-023.152177	15q11	7
214	UBE3A	M-10885	LIG-11555	15-023.150798	15q11	8
244	UBE3A	M-10886	LIG-11556	15-023.150555	15q11	9
256	UBE3A	M-01317	LIG-00864	15-023.136395	15q12	10
346	UBE3A	M-10887	LIG-11557	15-023.135402	15q11	11
264	Reference probe	M-10243	LIG-02149	01-247.075309	01q44	
328	Reference probe	M-05297	LIG-04685	03-095.102326	03q11.2	
364	Reference probe	M-02601	LIG-02072	05-176.651646	05q35.3	
301	Reference probe	M-09986	LIG-10445	07-098.341749	07q21	
238	Reference probe	M-07642	LIG-07327	08-011.651883	08p23.1	
337	Reference probe	M-09937	LIG-12248	08-071.291427	08q13	
154	Reference probe	M-09431	LIG-09680	11-066.148469	11q13	
136	Reference probe	M-09285	LIG-09516	11-067.950073	11q13.4	
418	Reference probe	M-08682	LIG-08694	13-099.607548	13q32.3	
445	Reference probe	M-08404	LIG-08258	15-029.149319	15q13.3	
400	Reference probe	M-02989	LIG-09994	16-016.188508	16p13.1	
172	Reference probe	M-03087	LIG-02487	16-003.869877	16p13.3	
190	Reference probe	M-09979	LIG-10438	19-010.998005	19p13.2	
391	Reference probe	M-09929	LIG-10388	22-039.866992	22q13.2	

MLPA data was analyzed using the GeneMarker software (SoftGenetics, State College, PA, USA) for MLPA analysis. After population normalization, data were compared to two different controls: 1) a single control sample, representing

the sample with the fewest abnormal calls in each experiment; and, 2) a synthetic control sample, which represents the median of all normal (CEPH) samples in each experiment, though these methods typically produced virtually identical results. A threshold of dosage change <0.75 marked samples as potential deletions and changes greater than 1.3 marked potential duplications.

Results

Validation of Target Amplification in Control Samples

Since amplicon sizes differ between 6bp and 12bp, each sample produce peaks representing the molecular weight of the product and an amplitude concordant with its amplification, scaled against the control samples as described above. The resultant plots appear as those below in figure 6-1.

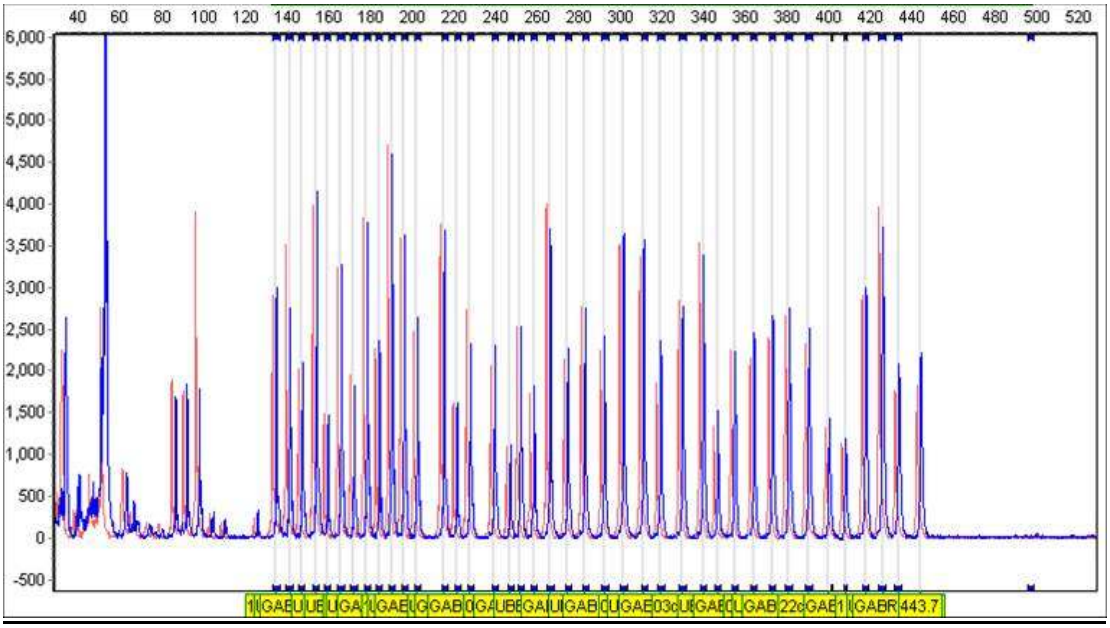


Figure 6-1. MLPA products analysis for a single sample. Amplicons are separated by capillary electrophoresis. Each peak represents the amplification product of a specific probe. Here the red peaks are those being tested and the blue peak is the synthetic control, which is the average of copy number known (CEPH) samples. On the X axis above is amplicon size. On the Y axis is relative fluorescent units (RFU) intensity values.

The pilot experiment used only samples with Angelman syndrome, duplications of 15q11-q13, and CEPH controls. GeneMarker software uses peak ratios to determine gains or losses in a sample. Since the first sample used a three Angelman and three dup(15) samples, it would be expected to be easy to distinguish the two from one another. In the graph below, one can see the clear discrimination of Angelman samples from dup(15) autism samples.

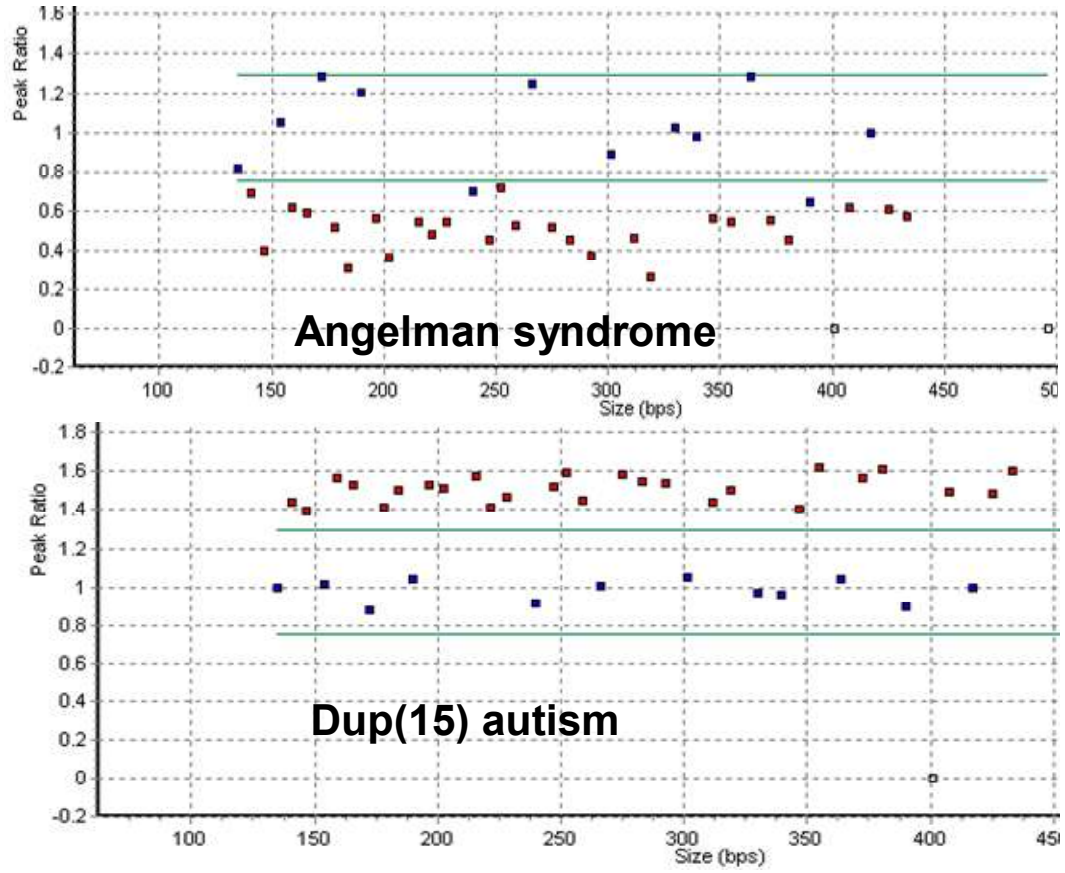


Figure 6-2. MLPA peak ratios for Angelman and Dup(15) samples. Here the blue dots between the green lines represent a normal copy number. Blue dots represent reference probes located at genomic loci on chromosomes separate from 15q. The red dots are the probes for *UBE3A* and *GABRB3* exons. In the top panel, those probes all fall below the green line at 0.75, indicating their loss, consistent with Angelman syndrome. In The lower panel, those probes are above the green line at 1.3, indicating their duplication status for this isodicentric duplication. The X-axis shows the amplicon sizes of the amplicons listed in table 6-2 and the Y axis shows the peak ratio between the measured relative fluorescence of the synthetic control sample and the sample being measured.

Absence of Fine Copy Number Variation in ASD Samples

The success of MLPA in correctly determining the copy number of Angelman, CEPH, and dup(15) samples across *UBE3A* and *GABRB3* exons gave us confidence to try running those samples listed in table 6.2. The samples were selected as described above by their ROH status at *UBE3A* and *GABRB3*,

abundance of Mendelian inconsistencies, or epilepsy status. To be sure that the protocol worked as before, Angelman, dup(15), and CEPH samples were also run again.

Results appeared largely negative for the samples of Table 6.1. Figure 6-3 shows two representative plots from the MLPA analysis of this group, neither of which suggested any exonic losses or gains in the *UBE3A* or *GABRB3* genes.

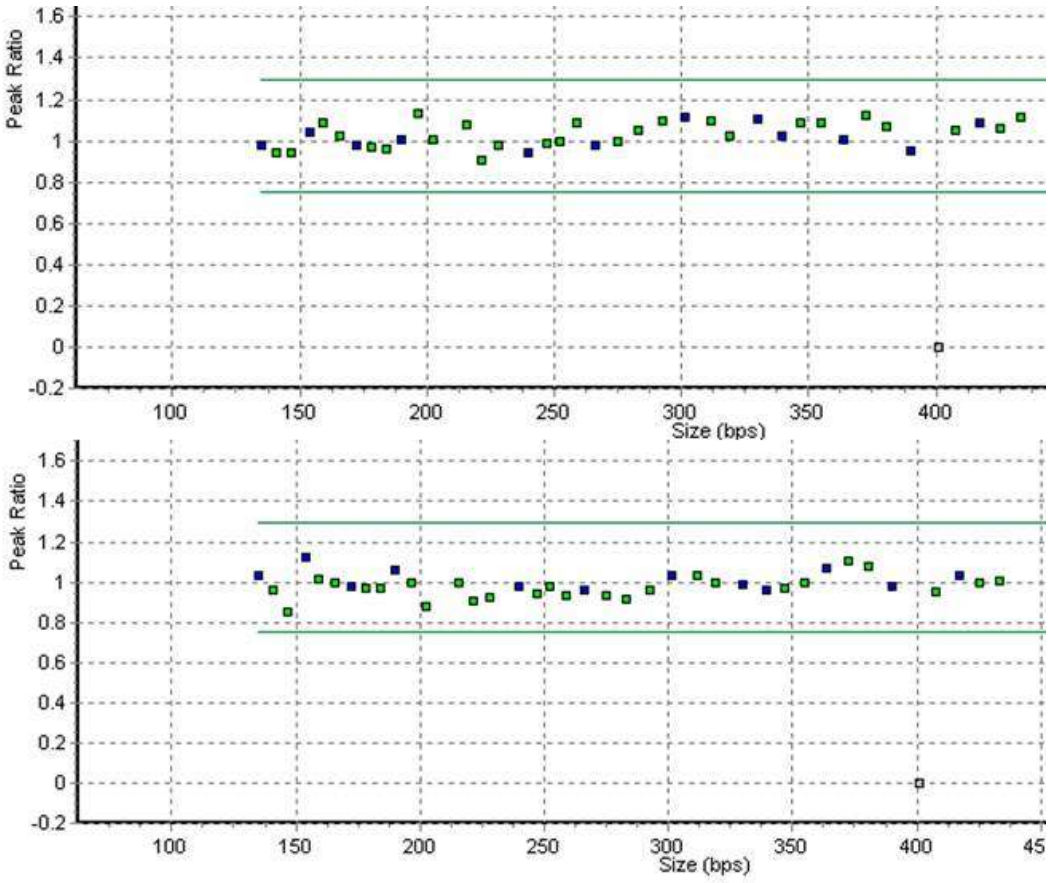


Figure 6-3. MLPA peak ratios for idiopathic autism samples. Here, in both panels blue dots represent reference probes and green dots are the probes for *UBE3A* and *GABRB3* exons, all within the normal peak ratio range. The X-axis shows the amplicon sizes of the amplicons listed in table 6-2 and the Y axis shows the peak ratio between the measured relative fluorescence of the synthetic control sample and the sample being measured.

Discussion

In this study, I investigated 48 samples with apparently idiopathic autism for the presence of microdeletion or microduplication in the *UBE3A* and *GABRB3* genes.

MLPA appears to be a powerful and rapid technique for interrogating multiple genomic loci for potential copy number variation. Its correct identification of duplication or deletion in dup(15) and Angelman samples, respectively, for all *UBE3A* and *GABRB3* exons is proof of principle that the method can correctly identify small deletions and duplications that may fall beyond the threshold of cytogenetics. Further, the built-in independent reference sequences at neutral genomic loci offer assurance that the phenomenon of gains or losses at these loci is not simply an artifact of a problematic sample.

Here, I correctly identified the loss of 9 exons of *GABRB3* and 11 exons of *UBE3A* from Angelman samples and their respective gains in dup(15) samples. Idiopathic autism samples, however, appeared chromosomally normal despite our best efforts to enrich for samples I thought would prove most promising for deletions or duplications in these genes. There are several reasons this could be so.

ROH analysis is typically performed on genome-wide association datasets. Its utility might be still better suited for detecting much larger deletions. Since the window of 50 SNPs required meant that it would span most of *UBE3A*

or *GABRB3*, it would not pick up smaller deletions such as those of single exons. As such, it may have lent little or no value in the selection of initial samples.

Taqman genotyping does not provide intensity data in a manner that permits the calling of copy number from genotype plots in an effective manner. The absence of this intensity data, a feature of the newer SNP-chip based genotyping platforms available from companies such as Illumina or Affymetrix, makes inferring gains or losses from Taqman data impractical if not impossible. It should be noted, however, that MLPA would be useful in detecting small deletions below the threshold of both intensity and ROH algorithms to pick up, particularly in regions of poor coverage or within exons.

The high levels of LD present in the *UBE3A* gene, which exists in a single block of LD, along with the relatively few number of common haplotypes, means that hundreds of samples appeared homozygous across *UBE3A*, giving little value to the use of its SNPs in ROH analysis. *GABRB3*, with its relatively low levels of LD, seemed a better target since homozygosity across multiple LD blocks becomes increasingly less probable.

Mendelian inconsistencies are often compatible with a deletion. Occasionally, manual inspection of the pedigrees indicated parents homozygous for the opposite allele present in homozygous offspring, suggesting hemizygosity at the locus. Other times, both parents appeared homozygous over a number of loci. It is possible that two parents both chromosomally identical over some number of loci could pass on some kind of recessive gene to chromosomally

normal children and that this could results in autism. Again, this type of scenario would not be detected by MLPA.

While epilepsy is associated with most of the 15q microdeletion disorders, our sample included only relatively small number with potential epilepsy. As this sample was ascertained for autism and possible epilepsy, and status was garnered just from two questions on the ADI-R, it is difficult to draw inferences from no findings of exonic CNVs in *GABRB3* and *UBE3A* for those samples with epilepsy. We can conclude, however, that such CNV events are likely to be very rare.

I restricted our analysis to two genes. A broader assay testing exons centered in more candidate genes in our sample may have been more likely to yield positive findings, but such an experiment is outside of the scope of this project.

Finally, my sample size, while dictated by clinical criteria, may also have been an issue. With only 48 samples tested and the apparent infrequency of 15q11.2 microdeletions and duplications, our power to detect a rare event was very low [315].

In summary, MLPA appears to be a powerful method to rapidly interrogate a large number of loci simultaneously. It appears to be accurate and while I identified no CNVs in *UBE3A* and *GABRB3*, it appears to have correctly and reliably detected the presence or absence of these exons in controls. Future use of this method should potentially be informed by alternate selection measures,

probe spacing calculations, sample size considerations, and cost of alternative platforms.

CHAPTER VII

CONCLUSIONS AND FUTURE DIRECTIONS

Autism in Context

The Blessing (and Curse) of Genetic Technology

The development of the publicly financed genome projects, most notably the HapMap, in combination with the parallel maturation of genotyping and sequencing technologies has revolutionized genetics. In just the past six years, we have gone from a single and now multiple finished genome sequences, to having human variation catalogued in several major world populations, now being expanded by the 1000 Genomes project. As “next generation sequencing” proliferates and improves in cost and performance, investigators will be able to capture vast sequences of DNA in single experiments that would have previously taken months or years. The explosion in the field mirrors Moore’s law as the cost of sequencing and genotyping is reduced by a factor of two or three each year. This technology-driven revolution in genetics makes it evident that human genetics is approaching an era where complete genome information (structural variation, common and rare SNPs, and complete genome sequence,) will be available in the future.

These developments and the maturation of the field of autism genetics has meant that researchers have moved from cytogenetics to array based CGH studies, from linkage studies and candidate gene studies to much larger scans

for genome-wide association and CNVs. But the wealth of information carries with it its own problems. Not every discovered functional, disease-associated, or even highly penetrant variant is causal in all cases, as CNVs such as 15q11-q13 and 16p11.2 illustrate. How do we account for this? In genome-wide association studies, how will we correct for 1 million tests and sub-significant results? What is the meaning of common allelic variants with odds ratios on the order of 1.2-1.6? New genome-wide methods for interpreting association results and integrating both common and rare risk alleles (see tables 1-3 and 1-4) must be developed.

Overcoming Heterogeneity

Despite many advances and the high heritability and relative frequency of autism and the ASDs, the genetic etiology of the vast majority of cases remains unknown. However, the past decade has witnessed major advances in defining aspects of the genetic etiology of autism. As indicated in the introduction, a large number of genomic disorders and syndromes can be easily identified. Just as the declination of intellectual disability (i.e. mental retardation) diagnoses has been correlated with increased ASD diagnoses, I would speculate that autism will—at least in part—resolve into some distinct genomic disorders as the genetics and biology are better understood.

Many of the samples in existing autism datasets may already have an identified molecular defect, such as a highly penetrant CNV, which may be sufficient to be causal in a given case. As indicated in the introduction (see table 1-4), a number of CNVs are now understood to be sufficient to cause an ASD

phenotype. Identifying individuals who have causal or highly penetrant defects or those whose disorder is caused by a combination of less penetrant alleles will likely become key to better understanding the mixed genetic architecture of the disease and knowing how to appropriately council families carrying autism-related variants. If the CNV burden in a family appears high, this might be sufficient to warrant meaningful genetic counseling to families facing difficult decisions.

Decomposing the Autism Phenotype

In addition to parsing genetic heterogeneity, it is also important to relate the underlying genetics to ASD clinical heterogeneity. ASDs have highly variable clinical presentations (see table 1-1) with wide ranging differences in IQ, medical co-morbidities, developmental trajectory, and other phenotypic dimensions. We must move beyond the categorical diagnostic category of autism to more carefully dissect the phenotype.

In addition to approaches which analyze autism as a whole, understanding the genes that may subtly increase risk for individual autistic traits (e.g. language deficits, rigid/repetitive behaviors, and impairments in social interaction) may also help to unravel the disorder. Such traits have been measured in the general population and people who do not meet diagnostic criteria for ASDs may exhibit significant impairment for an individual trait [317,318]. Further, there are low correlations between impairments in these three domains in the general population [319,320]. Perhaps considering how impaired both autistic individuals

and those in the general population are in these domains or studying the tails of these distributions may yield insight into the genes which increase risk for each of the component traits of autism.

Sporadic vs. Familial Autism

It will become increasingly important to understand any differences between cases of sporadic autism versus familial autism. More importantly, it may become important to determine a way to distinguish them from one another. Since there is an established risk of increased autism in children born to older mothers and to older fathers, this suggests that for sporadic cases there may be a strong case for de-novo mutation. However, it is not so straightforward since the birth of a single autistic child will often dissuade the parents from having additional children, so that a family which may be theoretically “multiplex” in its genetic architecture of autism risk alleles may present as a “simplex” family, that is, with just one affected child [321].

I would posit that most chromosomal forms that are known to be inherited are often due to CNVs originating in the mother or to point mutations originating in the father. Age is a known risk in mothers for chromosomal abnormalities. It seems possible that this could be the case for smaller structural variation. Since we do not yet have extensive ability to sequence whole genomes in a cost-efficient manner, the impact of older fathers and their contribution to de-novo point mutation remains to be seen. Nevertheless, parental age at conception is one means by which researchers may begin to determine if it is likely that autism

cases may be sporadic or familial. The knowledge that there is increased likelihood of de-novo events in sporadic vs. familial forms of autism is the motivation for the development of the Simons Simplex Collection (SSC) which, unlike the Autism Genetics Resource Exchange (AGRE), features simplex families whereas AGRE is almost exclusively multiplex families. Perhaps these collections will soon help to reveal if this distinction between simplex and multiplex families and their supposed genetic architecture of autism risk alleles is a meaningful distinction.

Another means to distinguish sporadic or *de novo* events from events that may be familial or polygenic may be by investigating unaffected family members for co-morbid phenotypes (see table 1). Elements of the BAP may be evident in "unaffected" siblings or in parents. Identifying families that have either traits that are part of the core phenotype or comorbidities common in autism may help to separate the sporadic/*de novo* cases from the familial/polygenic cases and to genetically map loci associated with subclinical traits of the BAP.

The Importance of 15q11-q13 to Idiopathic Autism

In the 1980s and 1990s the chromosome 15q11-q13 region was identified to be deleted in Prader-Willi and Angelman syndromes and maternally-duplicated in some cases of autism [119,322,323]. Since that time, it is now understood that a number of other rare deletions and duplications in or near the 15q11-q13 region (e.g. 15q13.3) can also dramatically increase risk for autism [324]. Other work has shown rare, but recurrent submicroscopic deletions and duplications in

and around this region with phenotypes including autism, seizures, schizophrenia, and mental retardation [139,141,314,325]. Now that the methods for assaying CNVs have matured, it will be possible to quantitate the extent to which the collected *genomic disorders in and flanking the 15q11-q13 interval contribute to total autism risk*. The *common* 15q11-q13 duplications (interstitial and idic(15)) alone account for 1-3% of autism cases, so the total fraction of autism explained by *additional* CNVs in and around this region is likely to be slightly higher.

In my study, analysis of genes believed to contribute to the phenotype and associated with genomic disorders in this region was restricted to *UBE3A* and *GABRB3*. While MLPA assay was shown to work correctly in identifying deletion and duplications of this region, of the 48 autism samples screened, none appeared to have had exonic duplications of *UBE3A* or *GABRB3*.

Several points here are worth re-stating, however. First, since most of our overall sample (nearly 1000 families) were karyotyped, they should be largely devoid of any gross chromosomal abnormalities including idic(15), the most common form of the duplication. This is due largely to the severity presentation which often includes dysmorphologies which would be more likely to come to the attention of a physician or clinical geneticist. If any duplications were present, it would most likely have only been interstitial duplications, because these are more likely to be missed by standard karyotyping. Second, given the low frequency of “common” interstitial duplications of 15q11-q13, 48 samples is certainly a small number to screen and it is unlikely that we would detect smaller

CNVs in so few samples. A clear future direction would include screening more samples. Third, while my strategy was to enrich for samples with CNVs by selecting those demonstrating of homozygosity, no exon level CNVs were detected and this method would be unlikely to enrich for small duplications. I also hypothesized that selection of subjects with a history of epilepsy and those whose samples showed short runs of Mendelian inconsistency might enrich the sample for CNV discovery. Perhaps better methods (e.g. CNV calling using more high density SNP or CGH arrays) will identify samples with CNVs in this region not detected by the limited MLPA pilot. Also characterization of the phenotype of those found to have CNVs in these genes might better help to identify those candidates which would merit such screening. Fourth, we were restricted not just to *UBE3A* and *GABRB3*, but specifically to exons of those genes. CNVs outside of exons would thus be missed.

In this limited pilot experiment, while sensitive to CNV differences experimentally, only gives a very limited snapshot of copy number status of a narrowly defined portion of the 15q11-q13 interval. Acknowledging that other CNVs are being discovered outside of this region this further suggests that CNVs involving *UBE3A* and *GABRB3* play an incomplete role in explaining the autism risk conferred by all CNVs possible in the 15q11-q13 and surrounding region.

Turning attention to common variation, *GABRB3* has been studied in numerous family-based association studies as a potential autism candidate gene. While results for *GABRB3* studies are mixed, most studies show positive results and Curran et. al. replicated association that we published in McCauley et. al.

We have speculated that there may be multiple susceptibility alleles which have aggregated at *GABRB3,* since multiple regions—not in LD—show evidence for association. It is important to consider that while the single replication of association at *GABRB3* by McCauley and Curran supports a real risk effect at this locus, many explanations exist for failed replication. The simplest is that the initial finding was a false positive. It may also be that several risk variants come to accumulate in different populations, as is supported by the rare variant, P11S, which makes results inconsistent and associations difficult to interpret.

Fewer studies have been conducted for *UBE3A,* in part because most findings have been negative or mixed. Clearly the positive association I show with *UBE3A* needs to be replicated in independent samples. Our association study of *UBE3A* (chapter 4) showed association with *UBE3A* and two genes which code for proteins that are regulated by *UBE3A* in an a *Drosophila* experimental system. That genetic association showed association of alleles of *UBE3A*, *ECT2*, *and GCH1* is very encouraging and suggests that model organisms and known biology may be able to correctly guide us to additional candidate genes.

While allelic association was modest at *UBE3A*, *ECT2*, and *GCH1,* the associated allele for *ECT2* as for *GCH1* were not good predictors of mRNA or protein expression level. Despite this, the positive relationship between *UBE3A* and *GCH1* gene expression was consistent with the correlations of protein expression observed in *Drosophila*.

Future studies should seek to verify or refute association in *UBE3A*, *ECT2*, and *GCH1*. Ideally, functional studies in appropriate cell lines (i.e. neuronal) or tissue samples (patient and control brain samples) could help to corroborate whether or not the gene and protein expression relationships observed in Drosophila extend to humans, and similarly perturbed in autism as predicted for Angelman syndrome. Such studies are underway in other labs. There is strong evidence that in chromosomally-normal from subjects with autism that *UBE3A* and *GABRB3* expression are perturbed [198,326]. This suggests that despite the absence of high rates of CNVs or strong association in the *UBE3A* and *GABRB3* genes that these genes may be subject to dysregulation in idiopathic autism. Determining what proportion of autism cases show such a profile could be important for unraveling a major pathway that may result in autism.

Association results were also modest in *GABRB3* given the number of tests performed. In addition, the site of association for the strongest signal observed was not in LD with that observed near the 155CA-2 region by two other groups, nor with that in the 3′ region of the gene previously observed by our group and Curran et. al [105,106,107,283].

The top result was identified for a relatively rare SNP in the large intron 3 region. The dominant model under which this SNP appears significant could imply that a single copy is all that is required for increased risk of autism. Encouragingly, this SNP does not appear to be in an LD block captured by previous association studies. As such, it will be interesting to see if the

association replicates in an independent dataset. Even so, there is virtually no apparent functional sequence (conserved non-coding regions, miRNA or other binding sites) around the variant to explain the observed association.

GABRB3 is a very attractive functional candidate. Despite these disparate results, positive association has been found on several occasions that the potential involvement of this gene and others encoding GABAA receptor subunits in autism is quite plausible. However, aside from this it is fair to ask what, if anything, these collected findings of association are contributing to our knowledge of the role of *GABRB3* in autism. Such disparate findings suggest involvement, but do not offer clear guidance on a *single* allele of the gene which may be associated. Perhaps other studies which identify QTL for *GABRB3* (and UBE3A) expression may shed light on other genes which may regulate expression of these two genes which appear to play an important role in the pathology of autism. Until then, it seems unlikely that additional association studies in *GABRB3* by themselves will yield much useful information.

While rare, the P11S variant provides more direct evidence for a role of *GABRB3* in autism by virtue of the functional evidence implicating its affect on receptor function and association with autism. Perhaps more importantly, it illustrates two phenomena. First, it shows that a rare coding variant in a gene that has hitherto shown only evidence of common allele association with autism may also have rare variants of higher penetrance. Such has also been shown for *CNTNAP2* and *SLC6A4* [88,90,102]. Secondly, it demonstrates that maternal and not paternal overtransmission of the S11 risk allele is associated with autism.

This is extremely interesting in light of the strong maternal bias of 15q11-q13 in dup(15) autism. It provides further evidence that parental origin of not just a genomic variant, but also a single point mutation, may play a role in disease. Such is the case in Angelman syndrome in which maternal deletion of the region was first identified to cause the disorder and later point mutations on the maternal copy of *UBE3A* found to be sufficient [256,327].

By contrast to the modest case for association of common alleles of *GABRB3* and *UBE3A* with autism, the case for common alleles of the X-chromosome gene *MECP2* showing association with autism is more consistent. Loat et al. showed association of SNPs in strong LD with SNPs that I genotyped in this project [219]. Work not shown here indicates that the same allele is associated in yet another independent set of families confirming this pattern of association.

Since the major allele appears associated, it will be interesting to determine if the major allele increases risk for any particular subphenotypic trait or is associated with other perturbations in gene expression in males or females. This gene could hypothetically represent a very attractive QTL which increases risk for some autism-related trait. However, given the high frequency of the associated form, by itself this will likely give little information about the likelihood of developing autism. The new finding of association of *MECP2* with idiopathic autism is potentially very important in light of the genes role in Rett syndrome and strong evidence suggesting involvement in the regulation of *UBE3A* and *GABRB3*.

In addition to *MECP2*, a number of other X-linked loci including *FMR1* (fragile X), *NLGN3*, and *NLGN4X* have shown rare alleles that segregate with ASD phenotypes. More than 22 genes on the X-chromosome may give rise to X-linked forms of mental retardation [328]. X-linked mental retardation is, of course, far more common in males than in females. The male:female ratio of X-linked mental retardation is 1.4:1 for profound MR and approximately 1.9:1 for moderate MR [329]. That a number of other X-linked loci are being shown to contribute to autism and emerging data continues to support this trend will make investigation of both common and rare alleles on the X-chromosome an exciting avenue for research.

Making Sense of the Model

While specific genomic disorders, point mutations, and common alleles may, at first blush, collectively only implicate 15q11-q13 genes in just a small fraction of autism, it is important to consider the larger picture (see figure 1-2). While such variants in this region may be present in only a small number of those ascertained to have idiopathic autism, dysfunction of the genes in this system and those regulating this system—most notably *MECP2, UBE3A*, and *GABRB3*—might contribute in a greater fraction of autism. While genomic and coding variants of this region are relatively uncommon and may offer explanation of just a small percentage of the phenotype, this network of genes may be perturbed in a much wider set of individuals. We still have only a very limited knowledge of the extent of this “gene network”.

Genes act within networks and multiple genes and proteins (in addition to *MECP2*) are likely to regulate the expression of *UBE3A* and *GABRB3*. Moving beyond what CNVs and coding mutations are present to understand more comprehensively the networks which are perturbed with such dysregulation is central to understanding how these genes may collectively contribute to the phenotype.

Work frequently cited in this thesis by Samaco et al. and Hogart et al. is perhaps the best indicator of the relationship between *UBE3A, GABRB3*, and *MECP2* [194,198,326]. In this thesis I have shown in the case of the *GABRB3* mutation, P11S, evidence that suggests deviation from biallelic expression of *GABRB3* which supports earlier observations [194]. An alternate explanation, not mutually exclusive, is that there is an effect based on maternal genotype.

I predict that the net effect of dysregulation of *UBE3A*, *GABRB3*, and *MECP2* is to contribute to a disruption of synapse development that may contribute to autism. Syndromic conditions with autism phenotypes have indicated that autism is often caused by genes important in playing a role in synapse formation and/or function. Evidence is accumulating that autism may in part be a synaptopathology as Zoghbi proposed in 2003 [118]. *SHANK3*, neuroligins, neurexins, *CNTNAP2*, *UBE3A*, *SLC6A4*, *GABRB3*, and others all have a role in formation of or signaling at the synapse. Examination of other genes in these pathways may uncover further candidate genes for which dysfunction may lead to autism.

This project has sought to bridge and garner evidence to build an understanding of the role of two 15q11-q13genes, and two associated network genes, in the development of autism. It has sought to further our understanding of whether and how two genes in the most commonly-observed CNV in autism, maternal duplication of 15q11-q13, play a role in autism. We have found the following: (1) We have shown association at *MECP2*, whose product shows evidence for regulating epigenetic control of this region. (2) We have shown association with *UBE3A*, the Angelman syndrome gene; and two loci in the *UBE3A* network, *ECT2* and *GCH1*; (3) weak evidence for association in the 3´ region of *GABRB3*; (4) a new, undocumented region of significance in intron 3; and (5) a functional *GABRB3* variant (P11S), in the signal peptide that confers increased autism risk when maternally-transmitted. To complete this profile it will be essential to understand the depth and extent to which dysregulation of the genes in this model contribute to autism susceptibility or even potentially cause autism, in the case of highly penetrant alleles.

Concluding Remarks

The last ten years have been an exciting time in human genetics, particularly for autism genetics where our understanding of the underlying disease risk has increased by leaps and bounds. As researchers work toward more closely dissecting the phenotype and the component rare and common variants that contribute to it, new methods will be required to integrate this information. The evidence for a synaptic pathology in autism has increased and

hopefully emerging clues may also lead us to better understand the types of synaptopathologies and other developmental defects that might cause autism. It is hoped that understanding the dysregulation of networks that lead to autism—including those involving UBE3A, GABRB3, and MECP2--may ultimately permit early detection and suggest medical treatments.

REFERENCES

1. (2007) Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002. MMWR Surveill Summ 56: 12-28.

2. Ganz ML (2007) The lifetime distribution of the incremental societal costs of autism. Arch Pediatr Adolesc Med 161: 343-349.

3. Abrahams BS, Geschwind DH (2008) Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet 9: 341-355.

4. Kuhn R (2004) Eugen Bleuler's concepts of psychopathology. Hist Psychiatry 15: 361-366.

5. Asperger H (1944) Die „Autistischen Psychopathen" im Kindesalter. Eur Arch Psychiatry Clin Neurosci.

6. Kanner L (1943) Autistic disturbance of affective contact. Nervous Child 2: 217-250.

7. Wolff S (2004) The history of autism. Eur Child Adolesc Psychiatry 13: 201-208.

8. Hanson DR, Gottesman, II (1976) The genetics, if any, of infantile autism and childhood schizophrenia. J Autism Child Schizophr 6: 209-234.

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

10. Folstein S, Rutter M (1977) Genetic influences and infantile autism. Nature 265: 726-728.

11. Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM (1985) Concordance for the syndrome of autism in 40 pairs of afflicted twins. Am J Psychiatry 142: 74-77.

12. Gillberg C, Wahlstrom J (1985) Chromosome abnormalities in infantile autism and other childhood psychoses: a population study of 66 cases. Dev Med Child Neurol 27: 293-304.

13. Wahlstrom J, Gillberg C, Gustavson KH, Holmgren G (1986) Infantile autism and the fragile X. A Swedish multicenter study. Am J Med Genet 23: 403-408.

14. American Psychiatric Association., American Psychiatric Association. Task Force on DSM-IV. (2000) Diagnostic and statistical manual of mental disorders : DSM-IV-TR. Washington, DC: American Psychiatric Association. xxxvii, 943 p. p.

15. World Health Organization. (2004) International statistical classification of diseases and related health problems. Geneva: World Health Organization.

16. Gillberg C, Coleman M (2000) The biology of the autistic syndromes. booksgooglecom.

17. Richdale AL, Prior MR (1995) The sleep/wake rhythm in children with autism. European child & adolescent psychiatry 4: 175-186.

18. Wiggs L, Stores G (1996) Severe sleep disturbance and daytime challenging behaviour in children with severe learning disabilities. Journal of intellectual disability research : JIDR 40 (Pt 6): 518-528.

19. Wiggs L, Stores G (2004) Sleep patterns and sleep disorders in children with autistic spectrum disorders: insights using parent report and actigraphy. Developmental medicine and child neurology 46: 372-380.

20. Tuchman R, Moshé SL, Rapin I (2009) Convulsing toward the pathophysiology of autism. Brain Dev 31: 95-103.

21. Munesue T, Ono Y, Mutoh K, Shimoda K, Nakatani H, et al. (2008) High prevalence of bipolar disorder comorbidity in adolescents and young adults with high-functioning autism spectrum disorder: a preliminary study of 44 outpatients. Journal of affective disorders 111: 170-175.

22. Folstein SE, Rosen-Sheidley B (2001) Genetics of autism: complex aetiology for a heterogeneous disorder. Nature Reviews Genetics 2: 943-955.

23. Veenstra-VanderWeele J, Christian SL, Cook JEH (2004) Autism as a Paradigmatic Complex Genetic Disorder. Annual Review of Genomics and Human Genetics 2003 Jan 01;5(1):379 5: 407-441.

24. Folstein SE, Rosen-Sheidley B (2001) Genetics of autism: complex aetiology for a heterogeneous disorder. Nat Rev Genet 2: 943-955.

25. Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, et al. (1989) Autism diagnostic interview: a standardized investigator-based instrument. J Autism Dev Disord 19: 363-387.

26. Gotham K, Risi S, Pickles A, Lord C (2007) The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. J Autism Dev Disord 37: 613-627.

27. Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24: 659-685.

28. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, et al. (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord 30: 205-223.

29. Lord C, Leventhal BL, Cook EH, Jr. (2001) Quantifying the phenotype in autism spectrum disorders. Am J Med Genet 105: 36-38.

30. Howlin P, Goode S, Hutton J, Rutter M (2004) Adult outcome for children with autism. J Child Psychol Psychiatry 45: 212-229.

31. Kuehn BM (2007) CDC: autism spectrum disorders common. JAMA 297: 940.

32. Van Naarden Braun K, Pettygrove S, Daniels J, Miller L, Nicholas J, et al. (2007) Evaluation of a methodology for a collaborative multiple source surveillance network for autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2002. MMWR Surveill Summ 56: 29-40.

33. Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, et al. (2006) Advancing paternal age and autism. Arch Gen Psychiatry 63: 1026-1032.

34. Cantor RM, Yoon JL, Furr J, Lajonchere CM (2007) Paternal age and autism are associated in a family-based sample. Mol Psychiatry 12: 419-421.

35. Croen LA, Najjar DV, Fireman B, Grether JK (2007) Maternal and paternal age and risk of autism spectrum disorders. Arch Pediatr Adolesc Med 161: 334-340.

36. Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet 1: 40-47.

37. Fombonne E (2003) The prevalence of autism. Jama 289: 87-89.

38. Rutter M (2005) Incidence of autism spectrum disorders: changes over time and their meaning. Acta Paediatr 94: 2-15.

39. Shattuck PT (2006) The contribution of diagnostic substitution to the growing administrative prevalence of autism in US special education. Pediatrics 117: 1028-1037.

40. Croen LA, Grether JK, Selvin S (2001) The epidemiology of mental retardation of unknown cause. Pediatrics 107: E86.

41. Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, et al. (2007) The epidemiology of autism spectrum disorders. Annu Rev Public Health 28: 235-258.

42. Skuse DH, Mandy WP, Scourfield J (2005) Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. The British journal of psychiatry : the journal of mental science 187: 568-572.

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

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

45. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39: 319-328.

46. Piven J, Palmer P, Jacobi D, Childress D, Arndt S (1997) Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. The American journal of psychiatry 154: 185-190.

47. Dawson G, Estes A, Munson J, Schellenberg G, Bernier R, et al. (2007) Quantitative assessment of autism symptom-related traits in probands and parents: Broader Phenotype Autism Symptom Scale. J Autism Dev Disord 37: 523-536.

48. Briskman J, Happe F, Frith U (2001) Exploring the cognitive phenotype of autism: weak "central coherence" in parents and siblings of children with autism: II. Real-life skills and preferences. J Child Psychol Psychiatry 42: 309-316.

49. Piven J, Wzorek M, Landa R, Lainhart J, Bolton P, et al. (1994) Personality characteristics of the parents of autistic individuals. Psychol Med 24: 783-795.

50. Abramson RK, Ravan SA, Wright HH, Wieduwilt K, Wolpert CM, et al. (2005) The relationship between restrictive and repetitive behaviors in individuals with autism and obsessive compulsive symptoms in parents. Child Psychiatry Hum Dev 36: 155-165.

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

52. Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, et al. (1999) A genomic screen of autism: evidence for a multilocus etiology. Am J Hum Genet 65: 493-507.

53. Gutierrez GC, Smalley SL, Tanguay PE (1998) Autism in tuberous sclerosis complex. Journal of Autism & Developmental Disorders 28: 97-103.

54. Williams CA, Angelman H, Clayton-Smith J, Driscoll DJ, Hendrickson JE, et al. (1995) Angelman syndrome: consensus for diagnostic criteria.Angelman Syndrome Foundation. American Journal of Medical Genetics 56: 237-238.

55. Zappella M, Meloni I, Longo I, Canitano R, Hayek G, et al. (2003) Study of MECP2 gene in Rett syndrome variants and autistic girls. Am J Med Genet 119B: 102-107.

56. Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, et al. (1995) 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 57: 717-726.

57. Battaglia A, Carey JC (2006) Etiologic yield of autistic spectrum disorders: a prospective study. American journal of medical genetics Part C, Seminars in medical genetics 142C: 3-7.

58. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. (2007) Strong association of de novo copy number mutations with autism. Science 316: 445-449.

59. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, et al. (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. J Med Genet 43: 843-849.

60. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, et al. (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39: 25-27.

61. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, et al. (2007) Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet 81: 1289-1297.

62. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34: 27-29.

63. Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265: 2037-2048.

64. Yang MS, Gill M (2007) A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence. Int J Dev Neurosci 25: 69-85.

65. Goring HH, Terwilliger JD, Blangero J (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 69: 1357-1369.

66. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33: 177-182.

67. Freitag CM (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. Mol Psychiatry 12: 2-22.

68. Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, et al. (2001) Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. Am J Hum Genet 68: 1514-1520.

69. Shao Y, Raiford KL, Wolpert CM, Cope HA, Ravan SA, et al. (2002) Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. Am J Hum Genet 70: 1058-1061.

70. (2001) A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. Am J Hum Genet 69: 570-581.

71. Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, et al. (2003) A genomewide screen of 345 families for autism-susceptibility loci. Am J Hum Genet 73: 886-897.

72. Trikalinos TA, Karvouni A, Zintzaras E, Ylisaukko-oja T, Peltonen L, et al. (2006) A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. Mol Psychiatry 11: 29-36.

73. (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. Hum Mol Genet 7: 571-578.

74. (2001) Further characterization of the autism susceptibility locus AUTS1 on chromosome 7q. Hum Mol Genet 10: 973-982.

75. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH (2002) Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. Am J Hum Genet 70: 60-71.

76. Alarcon M, Yonan AL, Gilliam TC, Cantor RM, Geschwind DH (2005) Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs. Mol Psychiatry 10: 747-757.

77. Schellenberg GD, Dawson G, Sung YJ, Estes A, Munson J, et al. (2006) Evidence for multiple loci from a genome scan of autism kindreds. Mol Psychiatry 11: 1049-1060, 1979.

78. Shao Y, Cuccaro ML, Hauser ER, Raiford KL, Menold MM, et al. (2003) Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. Am J Hum Genet 72: 539-548.

79. Nurmi EL, Dowd M, Tadevosyan-Leyfer O, Haines JL, Folstein SE, et al. (2003) Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. J Am Acad Child Adolesc Psychiatry 42: 856-863.

80. Kilpinen H, Ylisaukko-oja T, Rehnstrom K, Gaal E, Turunen JA, et al. (2009) Linkage and linkage disequilibrium scan for autism loci in an extended pedigree from Finland. Hum Mol Genet 18: 2912-2921.

81. Cantor RM, Kono N, Duvall JA, Alvarez-Retuerto A, Stone JL, et al. (2005) Replication of autism linkage: fine-mapping peak at 17q21. Am J Hum Genet 76: 1050-1056.

82. Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, et al. (2004) Evidence for sex-specific risk alleles in autism spectrum disorder. Am J Hum Genet 75: 1117-1123.

83. Strom SP, Stone JL, Ten Bosch JR, Merriman B, Cantor RM, et al. (2009) High-density SNP association study of the 17q21 chromosomal region linked to autism identifies CACNA1G as a novel candidate gene. Mol Psychiatry.

84. McCauley JL, Li C, Jiang L, Olson LM, Crockett G, et al. (2005) Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. BMC Med Genet 6: 1.

85. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, et al. (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 82: 150-159.

86. Jorde LB, Hasstedt SJ, Ritvo ER, Mason-Brothers A, Freeman BJ, et al. (1991) Complex segregation analysis of autism. Am J Hum Genet 49: 932-938.

87. Hallmayer J, Pintado E, Lotspeich L, Spiker D, McMahon W, et al. (1994) Molecular analysis and test of linkage between the FMR-1 gene and infantile autism in multiplex families. Am J Hum Genet 55: 951-959.

88. Sutcliffe JS, Delahanty RJ, Prasad HC, McCauley JL, Han Q, et al. (2005) Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. Am J Hum Genet 77: 265-279.

89. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, et al. (2007) A unified genetic theory for sporadic and inherited autism. Proc Natl Acad Sci U S A 104: 12831-12836.

90. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, et al. (2008) A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am J Hum Genet 82: 160-164.

91. Gharani N, Benayed R, Mancuso V, Brzustowicz LM, Millonig JH (2004) Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. Mol Psychiatry 9: 474-484.

92. Benayed R, Gharani N, Rossman I, Mancuso V, Lazar G, et al. (2005) Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. Am J Hum Genet 77: 851-868.

93. Weiss LA, Kosova G, Delahanty RJ, Jiang L, Cook EH, et al. (2006) Variation in ITGB3 is associated with whole-blood serotonin level and autism susceptibility. Eur J Hum Genet 14: 923-931.

94. Coutinho AM, Sousa I, Martins M, Correia C, Morgadinho T, et al. (2007) Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. Hum Genet 121: 243-256.

95. Fang P, Lev-Lehman E, Tsai TF, Matsuura T, Benton CS, et al. (1999) The spectrum of mutations in UBE3A causing Angelman syndrome. Hum Mol Genet 8: 129-135.

96. Hatton DD, Sideris J, Skinner M, Mankowski J, Bailey DB, Jr., et al. (2006) Autistic behavior in children with fragile X syndrome: prevalence, stability, and the impact of FMRP. Am J Med Genet A 140A: 1804-1813.

97. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, et al. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23: 185-188.

98. Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, et al. (2004) Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell 119: 19-31.

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

100. Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, et al. (2004) X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. Am J Hum Genet 74: 552-557.

101. Feng J, Schroer R, Yan J, Song W, Yang C, et al. (2006) High frequency of neurexin 1beta signal peptide structural variants in patients with autism. Neurosci Lett 409: 10-13.

102. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, et al. (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet 82: 165-173.

103. Benayed R, Choi J, Matteson PG, Gharani N, Kamdar S, et al. (2009) Autism-associated haplotype affects the regulation of the homeobox gene, ENGRAILED 2. Biol Psychiatry 66: 911-917.

104. Curran S, Roberts S, Thomas S, Veltman M, Browne J, et al. (2005) An association analysis of microsatellite markers across the Prader-Willi/Angelman critical region on chromosome 15 (q11-13) and autism spectrum disorder. Am J Med Genet B Neuropsychiatr Genet 137: 25-28.

105. McCauley JL, Olson LM, Delahanty R, Amin T, Nurmi EL, et al. (2004) A linkage disequilibrium map of the 1-Mb 15q12 GABA(A) receptor subunit

cluster and association to autism. Am J Med Genet B Neuropsychiatr Genet 131: 51-59.

106. Buxbaum JD, Silverman JM, Smith CJ, Greenberg DA, Kilifarski M, et al. (2002) Association between a GABRB3 polymorphism and autism. Mol Psychiatry 7: 311-316.

107. Curran S, Powell J, Neale BM, Dworzynski K, Li T, et al. (2006) An association analysis of candidate genes on chromosome 15 q11-13 and autism spectrum disorder. Mol Psychiatry 11: 709-713.

108. Weiss LA, Ober C, Cook EH, Jr. (2006) ITGB3 shows genetic and expression interaction with SLC6A4. Hum Genet 120: 93-100.

109. Campbell DB, Sutcliffe JS, Ebert PJ, Militerni R, Bravaccio C, et al. (2006) A genetic variant that disrupts MET transcription is associated with autism. Proc Natl Acad Sci U S A 103: 16834-16839.

110. Sousa I, Clark TG, Toma C, Kobayashi K, Choma M, et al. (2009) MET and autism susceptibility: family and case-control studies. Eur J Hum Genet 17: 749-758.

111. Yan J, Oliveira G, Coutinho A, Yang C, Feng J, et al. (2005) Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. Mol Psychiatry 10: 329-332.

112. Coutinho AM, Oliveira G, Morgadinho T, Fesel C, Macedo TR, et al. (2004) Variants of the serotonin transporter gene (SLC6A4) significantly contribute to hyperserotonemia in autism. Mol Psychiatry 9: 264-271.

113. Kim SJ, Cox N, Courchesne R, Lord C, Corsello C, et al. (2002) Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder. Mol Psychiatry 7: 278-288.

114. Anderson GM, Gutknecht L, Cohen DJ, Brailly-Tabard S, Cohen JH, et al. (2002) Serotonin transporter promoter variants in autism: functional effects and relationship to platelet hyperserotonemia. Mol Psychiatry 7: 831-836.

115. Nurmi EL, Bradford Y, Chen YH, Hall J, Arnone B, et al. (2001) Linkage disequilibrium at the Angelman syndrome gene UBE3A in autism families. Genomics 77: 105-113.

116. Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang YH, et al. (1997) De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat Genet 15: 74-77.

117. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, et al. (2010) SCAN: SNP and copy number annotation. Bioinformatics 26: 259-262.

118. Zoghbi HY (2003) Postnatal neurodevelopmental disorders: meeting at the synapse? Science 302: 826-830.

119. Cook EH, Jr., Lindgren V, Leventhal BL, Courchesne R, Lincoln A, et al. (1997) Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. Am J Hum Genet 60: 928-934.

120. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949-951.

121. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, et al. (2004) Large-scale copy number polymorphism in the human genome. Science 305: 525-528.

122. Conrad DF, R., Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) A high-resolution survey of deletion polymorphismGlobal variation in copy number in the human genome. Nat GenetNature 38444: 75-81444-81454.

123. Redon R, J. A., Ishikawa SS, W. G., Fitch KRD, E., Feuk L, Perry GvE, H., et al. (2006) Global variation in copy number in the human genomeIdentification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. NatureMol Psychiatry 44411: 444-541, 418-428.

124. Vorstman JA, Staal WG, van Daalen E, van Engeland H, Hochstenbach PF, et al. (2006) Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. Mol Psychiatry 11: 1, 18-28.

125. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. (2007) Strong association of de novo copy number mutations with autism. Science 316: 445-449.

126. Cook EH, Scherer SW (2008) Copy-number variations associated with neuropsychiatric conditions. Nature 455: 919-923.

127. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, et al. (2009) Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 5: e1000536.

128. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al. (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82: 477-488.

129. Sherr EH, Owen R, Albertson DG, Pinkel D, Cotter PD, et al. (2005) Genomic microarray analysis identifies candidate loci in patients with corpus callosum anomalies. Neurology 65: 1496-1498.

130. Burd L, Martsolf JT, Kerbeshian J, Jalal SM (1988) Partial 6p trisomy associated with infantile autism. Clin Genet 33: 356-359.

131. Vostanis P, Harrington R, Prendergast M, Farndon P (1994) Case reports of autism with interstitial deletion of chromosome 17 (p11.2 p11.2) and monosomy of chromosome 5 (5pter-->5p15.3). Psychiatr Genet 4: 109-111.

132. Ylisaukko-oja T, Alarcon M, Cantor RM, Auranen M, Vanhala R, et al. (2006) Search for autism loci by combined analysis of Autism Genetic Resource Exchange and Finnish families. Ann Neurol 59: 145-155.

133. Bundey S, Hardy C, Vickers S, Kilpatrick MW, Corbett JA (1994) Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. Dev Med Child Neurol 36: 736-742.

134. Barnby G, Abbott A, Sykes N, Morris A, Weeks DE, et al. (2005) Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. Am J Hum Genet 76: 950-966.

135. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 358: 667-675.

136. Manning MA, Cassidy SB, Clericuzio C, Cherry AM, Schwartz S, et al. (2004) Terminal 22q deletion syndrome: a newly recognized cause of speech and language disability in the autism spectrum. Pediatrics 114: 451-457.

137. Prasad C, Prasad AN, Chodirker BN, Lee C, Dawson AK, et al. (2000) Genetic evaluation of pervasive developmental disorders: the terminal 22q13 deletion syndrome may represent a recognizable phenotype. Clin Genet 57: 103-109.

138. Sahoo T, Shaw CA, Young AS, Whitehouse NL, Schroer RJ, et al. (2005) Array-based comparative genomic hybridization analysis of recurrent chromosome 15q rearrangements. Am J Med Genet A 139A: 106-113.

139. Pagnamenta AT, Wing K, Akha ES, Knight SJ, Bolte S, et al. (2009) A 15q13.3 microdeletion segregating with autism. Eur J Hum Genet 17: 687-692.

140. Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, et al. (2009) Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. J Med Genet 46: 242-248.

141. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, et al. (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 41: 160-162.

142. Dibbens LM, Mullen S, Helbig I, Mefford HC, Bayly MA, et al. (2009) Familial and sporadic 15q13.3 microdeletions in Idiopathic Generalized Epilepsy: Precedent for Disorders with Complex Inheritance. Hum Mol Genet.

143. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet.

144. Losh M, Childress D, Lam K, Piven J (2008) Defining key features of the broad autism phenotype: a comparison across parents of multiple- and single-incidence autism families. Am J Med Genet B Neuropsychiatr Genet 147B: 424-433.

145. Veenstra-Vanderweele J, Christian SL, Cook EH, Jr. (2004) Autism as a paradigmatic complex genetic disorder. Annu Rev Genomics Hum Genet 5: 379-405.

146. Wang NJ, Liu D, Parokonny AS, Schanen NC (2004) High-resolution molecular characterization of 15q11-q13 rearrangements by array comparative genomic hybridization (array CGH) with detection of gene dosage. Am J Hum Genet 75: 267-281.

147. Christian SL, Fantes JA, Mewborn SK, Huang B, Ledbetter DH (1999) Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). Hum Mol Genet 8: 1025-1037.

148. Horsthemke B, Buiting K (2006) Imprinting defects on human chromosome 15. Cytogenet Genome Res 113: 292-299.

149. Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. Nature 366: 362-365.

150. Neumann B, Kubicka P, Barlow DP (1995) Characteristics of imprinted genes. Nat Genet 9: 12-13.

151. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2: 21-32.

152. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. Science 318: 1136-1140.

153. Bonati MT, Russo S, Finelli P, Valsecchi MR, Cogliati F, et al. (2007) Evaluation of autism traits in Angelman syndrome: a resource to unfold autism genes. Neurogenetics.

154. Dimitropoulos A, Schultz RT (2007) Autistic-like symptomatology in Prader-Willi syndrome: a review of recent findings. Curr Psychiatry Rep 9: 159-164.

155. Borgatti R, Piccinelli P, Passoni D, Dalpra L, Miozzo M, et al. (2001) Relationship between clinical and genetic features in "inverted duplicated chromosome 15" patients. Pediatr Neurol 24: 111-116.

156. Fan Z, Greenwood R, Fisher A, Pendyal S, Powell CM (2009) Characteristics and frequency of seizure disorder in 56 patients with Prader-Willi syndrome. Am J Med Genet A 149A: 1581-1584.

157. Descheemaeker MJ, Govers V, Vermeulen P, Fryns JP (2006) Pervasive developmental disorders in Prader-Willi syndrome: the Leuven experience in 59 subjects and controls. Am J Med Genet A 140: 1136-1142.

158. Herzing LB, Cook EH, Jr., Ledbetter DH (2002) Allele-specific expression analysis by RNA-FISH demonstrates preferential maternal expression of UBE3A and imprint maintenance within 15q11- q13 duplications. Hum Mol Genet 11: 1707-1718.

159. Milner KM, Craig EE, Thompson RJ, Veltman MW, Thomas NS, et al. (2005) Prader-Willi syndrome: intellectual abilities and behavioural features by genetic subtype. J Child Psychol Psychiatry 46: 1089-1096.

160. Koenig K, Klin A, Schultz R (2004) Deficits in social attribution ability in Prader-Willi syndrome. J Autism Dev Disord 34: 573-582.

161. Veltman MW, Craig EE, Bolton PF (2005) Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review. Psychiatr Genet 15: 243-254.

162. Veltman MW, Thompson RJ, Roberts SE, Thomas NS, Whittington J, et al. (2004) Prader-Willi syndrome--a study comparing deletion and uniparental disomy cases with reference to autism spectrum disorders. Eur Child Adolesc Psychiatry 13: 42-50.

163. Clayton-Smith J, Laan L (2003) Angelman syndrome: a review of the clinical and genetic aspects. J Med Genet 40: 87-95.

164. Varela MC, Kok F, Otto PA, Koiffmann CP (2004) Phenotypic variability in Angelman syndrome: comparison among different deletion classes and between deletion and UPD subjects. Eur J Hum Genet 12: 987-992.

165. Lossie AC, Whitney MM, Amidon D, Dong HJ, Chen P, et al. (2001) Distinct phenotypes distinguish the molecular classes of Angelman syndrome. J Med Genet 38: 834-845.

166. Dr. Carolyn Schanen`, personal communication.

167. Battaglia A (2005) The inv dup(15) or idic(15) syndrome: a clinically recognisable neurogenetic disorder. Brain Dev 27: 365-369.

168. Crolla JA, Harvey JF, Sitch FL, Dennis NR (1995) Supernumerary marker 15 chromosomes: a clinical, molecular and FISH approach to diagnosis and prognosis. Hum Genet 95: 161-170.

169. Battaglia A, Gurrieri F, Bertini E, Bellacosa A, Pomponi MG, et al. (1997) The inv dup(15) syndrome: a clinically recognizable syndrome with altered behavior, mental retardation, and epilepsy. Neurology 48: 1081-1086.

170. Nawaz Z, Lonard DM, Smith CL, Lev-Lehman E, Tsai SY, et al. (1999) The Angelman syndrome-associated protein, E6-AP, is a coactivator for the nuclear hormone receptor superfamily. Mol Cell Biol 19: 1182-1189.

171. Verma S, Ismail A, Gao X, Fu G, Li X, et al. (2004) The ubiquitin-conjugating enzyme UBCH7 acts as a coactivator for steroid hormone receptors. Mol Cell Biol 24: 8716-8726.

172. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, et al. (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 459: 569-573.

173. Reiter LT, Seagroves TN, Bowers M, Bier E (2006) Expression of the Rho-GEF Pbl/ECT2 is regulated by the UBE3A E3 ubiquitin ligase. Hum Mol Genet 15: 2825-2835.

174. Unpublished data.

175. Bauman ML, Kemper TL (2005) Neuroanatomic observations of the brain in autism: a review and future directions. Int J Dev Neurosci 23: 183-187.

176. Albrecht U, Sutcliffe JS, Cattanach BM, Beechey CV, Armstrong D, et al. (1997) Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. Nat Genet 17: 75-78.

177. Dindot SV, Antalffy BA, Bhattacharjee MB, Beaudet AL (2008) The Angelman syndrome ubiquitin ligase localizes to the synapse and nucleus, and maternal deficiency results in abnormal dendritic spine morphology. Hum Mol Genet 17: 111-118.

178. Dyck MJ, Piek JP, Hay D, Smith L, Hallmayer J (2006) Are abilities abnormally interdependent in children with autism? J Clin Child Adolesc Psychol 35: 20-33.

179. Settleman J (2001) Rac 'n Rho: the music that shapes a developing embryo. Dev Cell 1: 321-331.

180. Morita K, Hirono K, Han M (2005) The Caenorhabditis elegans ect-2 RhoGEF gene regulates cytokinesis and migration of epidermal P cells. EMBO Rep 6: 1163-1168.

181. DiAntonio A, Haghighi AP, Portman SL, Lee JD, Amaranto AM, et al. (2001) Ubiquitination-dependent mechanisms regulate synaptic growth and function. Nature 412: 449-452.

182. Cook EH, Jr., Arora RC, Anderson GM, Berry-Kravis EM, Yan SY, et al. (1993) Platelet serotonin studies in hyperserotonemic relatives of children with autistic disorder. Life Sci 52: 2005-2015.

183. Buxbaum JD, Silverman JM, Smith CJ, Greenberg DA, Kilifarski M, et al. (2002) Association between a GABRB3 polymorphism and autism. Molecular Psychiatry 7: 311-316.

184. Cook EH, Courchesne RY, Cox NJ, Lord C, Gonen D, et al. (1998) Linkage-Disequilibrium Mapping of Autistic Disorder, With 15q11-13 Markers. American Journal of Human Genetics 62: 1077-1083.

185. Kim SJ, Brune CW, Kistner EO, Christian SL, Courchesne EH, et al. (2008) Transmission disequilibrium testing of the chromosome 15q11-q13 region in autism. Am J Med Genet B Neuropsychiatr Genet 147B: 1116-1125.

186. Salmon B, Hallmayer J, Rogers T, Kalaydjieva L, Petersen PB, et al. (1999) Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism. American Journal of Medical Genetics 88: 551-556.

187. Maestrini E, Lai C, Marlow A, Matthews N, Wallace S, et al. (1999) 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. American Journal of Medical Genetics 88: 492-496.

188. Martin ER, Menold MM, Wolpert CM, Bass MP, Donnelly SL, et al. (2000) Analysis of linkage disequilibrium in gamma-aminobutyric acid receptor subunit genes in autistic disorder. Am J Med Genet 96: 43-48.

189. Kim SA, Kim JH, Park M, Cho IH, Yoo HJ (2006) Association of GABRB3 polymorphisms with autism spectrum disorders in Korean trios. Neuropsychobiology 54: 160-165.

190. Blatt GJ, Fitzgerald CM, Guptill JT, Booker AB, Kemper TL, et al. (2001) Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study. J Autism Dev Disord 31: 537-543.

191. Oblak A, Gibbs TT, Blatt GJ (2009) Decreased GABAA receptors and benzodiazepine binding sites in the anterior cingulate cortex in autism. Autism Res 2: 205-219.

192. Dhossche D, Applegate H, Abraham A, Maertens P, Bland L, et al. (2002) Elevated plasma gamma-aminobutyric acid (GABA) levels in autistic youngsters: stimulus for a GABA hypothesis of autism. Med Sci Monit 8: PR1-6.

193. Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, et al. (2008) Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. J Med Genet.

194. Hogart A, Nagarajan RP, Patzel KA, Yasui DH, Lasalle JM (2007) 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. Hum Mol Genet 16: 691-703.

195. Ballestar E, Ropero S, Alaminos M, Armstrong J, Setien F, et al. (2005) The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. Hum Genet 116: 91-104.

196. Jedele KB (2007) The overlapping spectrum of rett and angelman syndromes: a clinical review. Semin Pediatr Neurol 14: 108-117.

197. Young DJ, Bebbington A, Anderson A, Ravine D, Ellaway C, et al. (2007) The diagnosis of autism in a female: could it be Rett syndrome? Eur J Pediatr.

198. Samaco RC, Hogart A, LaSalle JM (2005) Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. Hum Mol Genet 14: 483-492.

199. Loat C, Curran S, Lewis C, Abrahams B, Duvall J, et al. (2008) Methyl - CpG - binding protein (MECP2) polymorphisms and vulnerability to autism. Genes Brain Behav.

200. LaSalle JM (2007) The odyssey of MeCP2 and parental imprinting. Epigenetics 2: 5-10.

201. Abuhatzira L, Makedonski K, Kaufman Y, Razin A, Shemer R (2007) MeCP2 deficiency in the brain decreases BDNF levels by REST/CoREST-mediated repression and increases TRKB production. Epigenetics 2: 214-222.

202. Nagarajan RP, Hogart AR, Gwye Y, Martin MR, LaSalle JM (2006) Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. Epigenetics 1: e1-11.

203. Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, et al. (2007) Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A 104: 19416-19421.

204. Armstrong DD (2005) Can we relate MeCP2 deficiency to the structural and chemical abnormalities in the Rett brain? Brain Dev 27 Suppl 1: S72-S76.

205. Armstrong DD, Deguchi K, Antallfy B (2003) Survey of MeCP2 in the Rett syndrome and the non-Rett syndrome brain. J Child Neurol 18: 683-687.

206. Vincent JB, Melmer G, Bolton PF, Hodgkinson S, Holmes D, et al. (2005) Genetic linkage analysis of the X chromosome in autism, with emphasis on the fragile X region. Psychiatr Genet 15: 83-90.

207. Moretti P, Levenson JM, Battaglia F, Atkinson R, Teague R, et al. (2006) Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. J Neurosci 26: 319-327.

208. Venancio M, Santos M, Pereira SA, Maciel P, Saraiva JM (2007) An explanation for another familial case of Rett syndrome: maternal germline mosaicism. Eur J Hum Genet 15: 902-904.

209. Masuyama T, Matsuo M, Jing JJ, Tabara Y, Kitsuki K, et al. (2005) Classic Rett syndrome in a boy with R133C mutation of MECP2. Brain Dev 27: 439-442.

210. Moog U, Smeets EE, van Roozendaal KE, Schoenmakers S, Herbergs J, et al. (2003) Neurodevelopmental disorders in males related to the gene causing Rett syndrome in females (MECP2). Eur J Paediatr Neurol 7: 5-12.

211. Neul JL, Zoghbi HY (2004) Rett syndrome: a prototypical neurodevelopmental disorder. Neuroscientist 10: 118-128.

212. Beyer KS, Blasi F, Bacchelli E, Klauck SM, Maestrini E, et al. (2002) Mutation analysis of the coding sequence of the MECP2 gene in infantile autism. Hum Genet 111: 305-309.

213. Carney RM, Wolpert CM, Ravan SA, Shahbazian M, Ashley-Koch A, et al. (2003) Identification of MeCP2 mutations in a series of females with autistic disorder. Pediatr Neurol 28: 205-211.

214. Lam CW, Yeung WL, Ko CH, Poon PM, Tong SF, et al. (2000) Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. J Med Genet 37: E41.

215. Lobo-Menendez F, Sossey-Alaoui K, Bell JM, Copeland-Yates SA, Plank SM, et al. (2003) Absence of MeCP2 mutations in patients from the South Carolina autism project. Am J Med Genet B Neuropsychiatr Genet 117: 97-101.

216. van Karnebeek CD, van Gelderen I, Nijhof GJ, Abeling NG, Vreken P, et al. (2002) An aetiological study of 25 mentally retarded adults with autism. J Med Genet 39: 205-213.

217. Ylisaukko-Oja T, Rehnstrom K, Vanhala R, Kempas E, von Koskull H, et al. (2005) MECP2 mutation analysis in patients with mental retardation. Am J Med Genet A 132: 121-124.

218. Coutinho AM, Oliveira G, Katz C, Feng J, Yan J, et al. (2007) MECP2 coding sequence and 3'UTR variation in 172 unrelated autistic patients. Am J Med Genet B Neuropsychiatr Genet 144: 475-483.

219. Loat CS, Curran S, Lewis CM, Duvall J, Geschwind D, et al. (2008) Methyl-CpG-binding protein 2 polymorphisms and vulnerability to autism. Genes Brain Behav 7: 754-760.

220. Courchesne E (2004) Brain development in autism: early overgrowth followed by premature arrest of growth. Ment Retard Dev Disabil Res Rev 10: 106-111.

221. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459: 528-533.

222. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265.

223. O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63: 259-266.

224. Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506-516.

225. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

226. Tadevosyan-Leyfer O, Dowd M, Mankoski R, Winklosky B, Putnam S, et al. (2003) A principal components analysis of the Autism Diagnostic Interview-Revised. J Am Acad Child Adolesc Psychiatry 42: 864-872.

227. Hartl DL, Clark AG (2007) Principles of population genetics. Sunderland, Mass.: Sinauer Associates. xv, 652 p. p.

228. Hammer S, Dorrani N, Dragich J, Kudo S, Schanen C (2002) The phenotypic consequences of MECP2 mutations extend beyond Rett syndrome. Ment Retard Dev Disabil Res Rev 8: 94-98.

229. Wan M, Lee SS, Zhang X, Houwink-Manville I, Song HR, et al. (1999) Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. Am J Hum Genet 65: 1520-1529.

230. Joyner AH, Roddey JC, Bloss CS, Bakken TE, Rimol LM, et al. (2009) A common MECP2 haplotype associates with reduced cortical surface area in humans in two independent populations. Proc Natl Acad Sci U S A.

231. Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57: 455-464.

232. Samaco RC, Nagarajan RP, Braunschweig D, LaSalle JM (2004) Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. Hum Mol Genet 13: 629-639.

233. Watson P, Black G, Ramsden S, Barrow M, Super M, et al. (2001) Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. J Med Genet 38: 224-228.

234. Milani D, Pantaleoni C, D'Arrigo S, Selicorni A, Riva D (2005) Another patient with MECP2 mutation without classic Rett syndrome phenotype. Pediatr Neurol 32: 355-357.

235. Chen RZ, Akbarian S, Tudor M, Jaenisch R (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet 27: 327-331.

236. Adler DA, Quaderi NA, Brown SD, Chapman VM, Moore J, et al. (1995) The X-linked methylated DNA binding protein, Mecp2, is subject to X inactivation in the mouse. Mamm Genome 6: 491-492.

237. D'Esposito M, Quaderi NA, Ciccodicola A, Bruni P, Esposito T, et al. (1996) Isolation, physical mapping, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. Mamm Genome 7: 533-535.

238. Tate P, Skarnes W, Bird A (1996) The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Nat Genet 12: 205-208.

239. Braunschweig D, Simcox T, Samaco RC, LaSalle JM (2004) X-Chromosome inactivation ratios affect wild-type MeCP2 expression within mosaic Rett syndrome and Mecp2-/+ mouse brain. Hum Mol Genet 13: 1275-1286.

240. Young JI, Zoghbi HY (2004) X-chromosome inactivation patterns are unbalanced and affect the phenotypic outcome in a mouse model of rett syndrome. Am J Hum Genet 74: 511-520.

241. Nan X, Campoy FJ, Bird A (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell 88: 471-481.

242. Nan X, Tate P, Li E, Bird A (1996) DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol 16: 414-421.

243. Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, et al. (2005) Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet 37: 254-264.

244. Chen WG, Chang Q, Lin Y, Meissner A, West AE, et al. (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science 302: 885-889.

245. Martinowich K, Hattori D, Wu H, Fouse S, He F, et al. (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science 302: 890-893.

246. Shahbazian MD, Antalffy B, Armstrong DL, Zoghbi HY (2002) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. Hum Mol Genet 11: 115-124.

247. Luikenhuis S, Giacometti E, Beard CF, Jaenisch R (2004) Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. Proc Natl Acad Sci U S A 101: 6033-6038.

248. Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, et al. (2004) Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. Hum Mol Genet 13: 2679-2689.

249. Kriaucionis S, Bird A (2004) The major form of MeCP2 has a novel N-terminus generated by alternative splicing. Nucleic Acids Res 32: 1818-1823.

250. Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, Yamada T, et al. (2004) A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. Nat Genet 36: 339-341.

251. Cheadle JP, Gill H, Fleming N, Maynard J, Kerr A, et al. (2000) Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location. Hum Mol Genet 9: 1119-1129.

252. Bienvenu T, Carrie A, de Roux N, Vinet MC, Jonveaux P, et al. (2000) MECP2 mutations account for most cases of typical forms of Rett syndrome. Hum Mol Genet 9: 1377-1384.

253. Huppke P, Laccone F, Kramer N, Engel W, Hanefeld F (2000) Rett syndrome: analysis of MECP2 and clinical characterization of 31 patients. Hum Mol Genet 9: 1369-1375.

254. Amano K, Nomura Y, Segawa M, Yamakawa K (2000) Mutational analysis of the MECP2 gene in Japanese patients with Rett syndrome. J Hum Genet 45: 231-236.

255. De Bona C, Zappella M, Hayek G, Meloni I, Vitelli F, et al. (2000) Preserved speech variant is allelic of classic Rett syndrome. Eur J Hum Genet 8: 325-330.

256. Kishino T, Lalande M, Wagstaff J (1997) UBE3A/E6-AP mutations cause Angelman syndrome. Nat Genet 15: 70-73.

257. Tatsumoto T, Xie X, Blumenthal R, Okamoto I, Miki T (1999) Human ECT2 is an exchange factor for Rho GTPases, phosphorylated in G2/M phases, and involved in cytokinesis. J Cell Biol 147: 921-928.

258. Thony B, Auerbach G, Blau N (2000) Tetrahydrobiopterin biosynthesis, regeneration and functions. Biochem J 347 Pt 1: 1-16.

259. Huibregtse JM, Scheffner M, Howley PM (1993) Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. Mol Cell Biol 13: 4918-4927.

260. O'Keefe L, Somers WG, Harley A, Saint R (2001) The pebble GTP exchange factor and the control of cytokinesis. Cell Struct Funct 26: 619-626.

261. Souza CP, Valadares ER, Trindade AL, Rocha VL, Oliveira LR, et al. (2008) Mutation in intron 5 of GTP cyclohydrolase 1 gene causes dopa-responsive dystonia (Segawa syndrome) in a Brazilian family. Genet Mol Res 7: 687-694.

262. Wu ZY, Lin Y, Chen WJ, Zhao GX, Xie H, et al. (2008) Molecular analyses of GCH-1, TH and parkin genes in Chinese dopa-responsive dystonia families. Clin Genet 74: 513-521.

263. Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 9: 301-306.

264. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, et al. (2004) Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 26: 61-69.

265. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408.

266. Talis AL, Huibregtse JM, Howley PM (1998) The role of E6AP in the regulation of p53 protein levels in human papillomavirus (HPV)-positive and HPV-negative cells. J Biol Chem 273: 6439-6445.

267. Jacobsen J, King BH, Leventhal BL, Christian SL, Ledbetter DH, et al. (1998) Molecular screening for proximal 15q abnormalities in a mentally retarded population. J Med Genet 35: 534-538.

268. Gurrieri F, Battaglia A, Torrisi L, Tancredi R, Cavallaro C, et al. (1999) Pervasive developmental disorder and epilepsy due to maternally derived duplication of 15q11-q13. Neurology 52: 1694-1697.

269. Koekkoek SK, Yamaguchi K, Milojkovic BA, Dortland BR, Ruigrok TJ, et al. (2005) Deletion of FMR1 in Purkinje cells enhances parallel fiber LTD, enlarges spines, and attenuates cerebellar eyelid conditioning in Fragile X syndrome. Neuron 47: 339-352.

270. Comery TA, Harris JB, Willems PJ, Oostra BA, Irwin SA, et al. (1997) Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. Proc Natl Acad Sci U S A 94: 5401-5404.

271. Armstrong DD (2001) Rett syndrome neuropathology review 2000. Brain Dev 23 Suppl 1: S72-76.

272. Ramamoorthy S, Nawaz Z (2008) E6-associated protein (E6-AP) is a dual function coactivator of steroid hormone receptors. Nucl Recept Signal 6: e006.

273. Nurmi EL, Amin T, Olson LM, Jacobs MM, McCauley JL, et al. (2003) Dense linkage disequilibrium mapping in the 15q11-q13 maternal expression domain yields evidence for association in autism. Mol Psychiatry 8: 624-634, 570.

274. Veenstra-VanderWeele J, Gonen D, Leventhal BL, Cook EH, Jr. (1999) Mutation screening of the UBE3A/E6-AP gene in autistic disorder. Mol Psychiatry 4: 64-67.

275. Vourc'h P, Martin I, Bonnet-Brilhault F, Marouillat S, Barthelemy C, et al. (2003) Mutation screening and association study of the UBE2H gene on chromosome 7q32 in autistic disorder. Psychiatr Genet 13: 221-225.

276. Kato C, Tochigi M, Ohashi J, Koishi S, Kawakubo Y, et al. (2008) Association study of the 15q11-q13 maternal expression domain in Japanese autistic patients. Am J Med Genet B Neuropsychiatr Genet 147B: 1008-1012.

277. Molfetta GA, Silva-Jr WA, Pina-Neto JM (2003) Clinical, cytogenetical and molecular analyses of Angelman syndrome. Genet Couns 14: 45-56.

278. Hitchins MP, Rickard S, Dhalla F, Fairbrother UL, de Vries BB, et al. (2004) Investigation of UBE3A and MECP2 in Angelman syndrome (AS) and patients with features of AS. Am J Med Genet A 125A: 167-172.

279. Schanen NC (2006) Epigenetics of autism spectrum disorders. Hum Mol Genet 15 Spec No 2: R138-150.

280. Buiting K, Dittrich B, Gross S, Greger V, Lalande M, et al. (1993) Molecular definition of the Prader-Willi syndrome chromosome region and orientation of the SNRPN gene. Hum Mol Genet 2: 1991-1994.

281. Blatt GJ (2005) GABAergic cerebellar system in autism: a neuropathological and developmental perspective. Int Rev Neurobiol 71: 167-178.

282. Fatemi SH, Reutiman TJ, Folsom TD, Thuras PD (2008) GABA(A) Receptor Downregulation in Brains of Subjects with Autism. J Autism Dev Disord.

283. Cook EH, Jr., Courchesne RY, Cox NJ, Lord C, Gonen D, et al. (1998) Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers. Am J Hum Genet 62: 1077-1083.

284. Salmon B, Hallmayer J, Rogers T, Kalaydjieva L, Petersen PB, et al. (1999) Absence of linkage and linkage disequilibrium to chromosome 15q11-q13

markers in 139 multiplex families with autism. Am J Med Genet 88: 551-556.

285. Laurie DJ, Wisden W, Seeburg PH (1992) The distribution of thirteen GABAA receptor subunit mRNAs in the rat brain. III. Embryonic and postnatal development. J Neurosci 12: 4151-4172.

286. Wisden W, Laurie DJ, Monyer H, Seeburg PH (1992) The distribution of 13 GABAA receptor subunit mRNAs in the rat brain. I. Telencephalon, diencephalon, mesencephalon. J Neurosci 12: 1040-1062.

287. Andang M, Hjerling-Leffler J, Moliner A, Lundgren TK, Castelo-Branco G, et al. (2008) Histone H2AX-dependent GABA(A) receptor regulation of stem cell proliferation. Nature 451: 460-464.

288. Brooks-Kayal AR, Pritchett DB (1993) Developmental changes in human gamma-aminobutyric acidA receptor subunit composition. Ann Neurol 34: 687-693.

289. Ma W, Saunders PA, Somogyi R, Poulter MO, Barker JL (1993) Ontogeny of GABAA receptor subunit mRNAs in rat spinal cord and dorsal root ganglia. J Comp Neurol 338: 337-359.

290. Homanics GE, DeLorey TM, Firestone LL, Quinlan JJ, Handforth A, et al. (1997) Mice devoid of gamma-aminobutyrate type A receptor beta3 subunit have epilepsy, cleft palate, and hypersensitive behavior. Proc Natl Acad Sci U S A 94: 4143-4148.

291. Krasowski MD, Rick CE, Harrison NL, Firestone LL, Homanics GE (1998) A deficit of functional GABA(A) receptors in neurons of beta 3 subunit knockout mice. Neurosci Lett 240: 81-84.

292. DeLorey TM, Sahbaie P, Hashemi E, Homanics GE, Clark JD (2008) Gabrb3 gene deficient mice exhibit impaired social and exploratory behaviors, deficits in non-selective attention and hypoplasia of cerebellar vermal lobules: a potential model of autism spectrum disorder. Behav Brain Res 187: 207-220.

293. DeLorey TM (2005) GABRB3 gene deficient mice: a potential model of autism spectrum disorder. Int Rev Neurobiol 71: 359-382.

294. Cordell HJ, Clayton DG (2005) Genetic association studies. Lancet 366: 1121-1131.

295. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, et al. (2006) Combining information from multiple sources in the diagnosis of autism spectrum disorders. J Am Acad Child Adolesc Psychiatry 45: 1094-1103.

296. Kang JQ, Macdonald RL (2004) The GABAA receptor gamma2 subunit R43Q mutation linked to childhood absence epilepsy and febrile seizures causes retention of alpha1beta2gamma2S receptors in the endoplasmic reticulum. J Neurosci 24: 8672-8677.

297. Gallagher MJ, Shen W, Song L, Macdonald RL (2005) Endoplasmic reticulum retention and associated degradation of a GABAA receptor epilepsy mutation that inserts an aspartate in the M3 transmembrane segment of the alpha1 subunit. J Biol Chem 280: 37995-38004.

298. Kirkness EF, Fraser CM (1993) 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 268: 4420-4428.

299. Weinberg CR (1999) Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. Am J Hum Genet 65: 229-235.

300. Tanaka M, Olsen RW, Medina MT, Schwartz E, Alonso ME, et al. (2008) Hyperglycosylation and reduced GABA currents of mutated GABRB3 polypeptide in remitting childhood absence epilepsy. Am J Hum Genet 82: 1249-1261.

301. Taylor PM, Thomas P, Gorrie GH, Connolly CN, Smart TG, et al. (1999) Identification of amino acid residues within GABA(A) receptor beta subunits that mediate both homomeric and heteromeric receptor expression. J Neurosci 19: 6360-6371.

302. McDonald BJ, Amato A, Connolly CN, Benke D, Moss SJ, et al. (1998) Adjacent phosphorylation sites on GABAA receptor beta subunits determine regulation by cAMP-dependent protein kinase. Nat Neurosci 1: 23-28.

303. Sutcliffe JS, Nurmi EL, Lombroso PJ (2003) Genetics of childhood disorders: XLVII. Autism, part 6: duplication and inherited susceptibility of chromosome 15q11-q13 genes in autism. J Am Acad Child Adolesc Psychiatry 42: 253-256.

304. Tsuboi T, Endo S (1977) Incidence of seizures and EEG abnormalities among offspring of epileptic patients. Hum Genet 36: 173-189.

305. Ottman R, Annegers JF, Hauser WA, Kurland LT (1988) Higher risk of seizures in offspring of mothers than of fathers with epilepsy. Am J Hum Genet 43: 257-264.

306. Nakatani J, Tamada K, Hatanaka F, Ise S, Ohta H, et al. (2009) Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. Cell 137: 1235-1246.

307. Bunzel R, Blumcke I, Cichon S, Normann S, Schramm J, et al. (1998) Polymorphic imprinting of the serotonin-2A (5-HT2A) receptor gene in human adult brain. Brain Res Mol Brain Res 59: 90-92.

308. Loiseau P, Duche B, Pedespan JM (1995) Absence epilepsies. Epilepsia 36: 1182-1186.

309. Tuchman R, Rapin I (2002) Epilepsy in autism. Lancet Neurol 1: 352-358.

310. Levisohn PM (2007) The autism-epilepsy connection. Epilepsia 48 Suppl 9: 33-35.

311. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324: 387-389.

312. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75-81.

313. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444: 444-454.

314. van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, et al. (2009) Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. J Med Genet 46: 511-523.

315. Cai G, Edelmann L, Goldsmith JE, Cohen N, Nakamine A, et al. (2008) Multiplex ligation-dependent probe amplification for genetic screening in autism spectrum disorders: efficient identification of known microduplications and identification of a novel microduplication in ASMT. BMC Med Genomics 1: 50.

316. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, et al. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30: e57.

317. Constantino JN, Przybeck T, Friesen D, Todd RD (2000) Reciprocal social behavior in children with and without pervasive developmental disorders. J Dev Behav Pediatr 21: 2-11.

318. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E (2001) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord 31: 5-17.

319. Ronald A, Happe F, Bolton P, Butcher LM, Price TS, et al. (2006) Genetic heterogeneity between the three components of the autism spectrum: a twin study. J Am Acad Child Adolesc Psychiatry 45: 691-699.

320. Happe F, Ronald A (2008) The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research. Neuropsychol Rev 18: 287-304.

321. Jones MB, Szatmari P (1988) Stoppage rules and genetic studies of autism. J Autism Dev Disord 18: 31-40.

322. Reish O, King RA (1995) Angelman syndrome at an older age. Am J Med Genet 57: 510-511.

323. (1987) Second Annual Prader-Willi Syndrome Scientific Conference. Houston, June 17, 1987. Proceedings and abstracts. Am J Med Genet 28: 779-924.

324. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63: 1111-1117.

325. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nat Genet 40: 322-328.

326. Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, et al. (2009) Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. J Med Genet 46: 86-93.

327. Williams CA, Zori RT, Stone JW, Gray BA, Cantu ES, et al. (1990) Maternal origin of 15q11-13 deletions in Angelman syndrome suggests a role for genomic imprinting. Am J Med Genet 35: 350-353.

328. Ropers HH, Hamel BC (2005) X-linked mental retardation. Nat Rev Genet 6: 46-57.

329. Leonard H, Wen X (2002) The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev 8: 117-134.