Molecular Genetic Study of Autism and Intellectual Disability genes on the X-Chromosome

by

Abdul Noor

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Institute of Medical Science
University of Toronto

© Copyright by Abdul Noor 2012
Abstract

Autism is a neurodevelopmental disorder with an estimated prevalence of 1 in 150 children which makes it more common than childhood cancer and juvenile diabetes. It is estimated that there are more than 100,000 individuals affected by autism in Canada and tens of millions worldwide. It is well established that genetic factors play important role in the pathophysiology of autism; still, our current understanding of these genetic factors is limited and cause of autism remains an important question. During the past decade, after completion of human genome, several new high throughput genome scan technologies have been developed such as microarrays. In the present study, we undertook the challenge of identifying X-chromosomal genes involved in autism by performing genome-wide copy number variation analysis of more than 400 probands with autism using Affymetrix 500K single nucleotide polymorphism (SNP) microarrays. We identified copy number variants implicating several genes on the chromosome X such as PTCHD1, IL1RAPL1, IL1RAPL2 and TSPAN7 as autism candidate genes. We also demonstrated that autism and intellectual disability may share some of these genes as etiologic factors. We performed a comprehensive analysis of PTCHD1 locus and showed that mutations at this locus are associated with autism in ~1 % of the cases. This study also demonstrated that
*PTCHD1* mutations can cause intellectually disability with or without autism, and that the *PTCHD1* protein may act as a receptor in Hedgehog signaling pathway. We have also carried out a detailed analysis of *TSPAN7* and *IL1RAPL1* to explore the contributions of these genes in autism. We identified one family with intronic deletion of *IL1RAPL1* and another case with a missense mutation in this gene, thus implicating this known intellectual disability gene in autism. Our findings highlight the importance of the X chromosome in the etiology of autism, and demonstrate the power of copy number variation analysis coupled with other technologies in identification of disease genes, in particular for complex genetic disorders such as autism.
Acknowledgments

First and foremost, I offer my sincerest gratitude to my supervisor Dr. John Vincent for his support, guidance and encouragement which allowed me to complete this journey of completion of my PhD thesis. Without his excellent mentorship, I would not have achieved this task.

I would like to thank my supervisory committee members, Dr Stephen Scherer and Dr. Lucy Osborne for providing guidance over the years. The valuable directions from my advisory committee allowed me to successfully execute my project and polished my scientific skills.

I would like to thank all the past and present members of Dr. Vincent’s Lab. I would especially like to thank Chris Harvey, Beata Stachowiak and Anna Mikhailov for their help, support and for sharing their knowledge and skills with me. I would also like to thank Muhammad Arshad, Liana Kaufman and Peter Gianakopoulos for their contributions in my project. Working with all of you was a great learning and sharing experience for me.

Special thanks to members of Dr. Scherer’s Lab, in particular, to Dr. Christian Marshall and Jennifer Howe. I must acknowledge that my entire thesis project branched out from data generated by Dr. Marshall at Dr. Scherer’s lab.

I am also thankful to my family and friends for their continued support of my career goals, especially; I thank my parents and my wife, Ayeshah for supporting me to achieve my career goals. I also acknowledge my son, Bilal for bringing lot of joy and happiness to our family which provided my strength to focus towards my goals even when times were difficult. I am also very excited to welcome the new member of our family, my daughter, Iman.

Last but not least, I am very thankful to the families and patients who participated in our study. This endeavor was not possible without their commitment and support. I am also grateful to the Canadian Institutes of Health Research (CIHR) for awarding me the Canada Graduate Scholarship. I also acknowledge funding from Genome Canada and Autism Speaks which enabled us to perform this study.
Table of Contents

Acknowledgments.......................................................................................................................... iv

Table of Contents ............................................................................................................................ v

List of Tables ........................................................................................................................................ ix

List of Figures ........................................................................................................................................ x

Abbreviations ....................................................................................................................................... xiii

Chapter 1. Background...................................................................................................................... 1

1.1 Autism ........................................................................................................................................... 2

1.1.1 Endophenotypes ............................................................................................................. 5

1.1.2 Autism in other single gene disorders ..................................................................... 7

1.1.3 Causes of Autism .................................................................................................... 9

1.1.4 Identification of Causative Genetic Factors .......................................................... 14

1.2 Intellectual Disability (ID) ................................................................................................. 25

1.3 The X-Chromosome ............................................................................................................. 30

1.3.1 Evolution of the X-Chromosome .......................................................................... 30

1.3.2 Autism and X-chromosomal genes: ...................................................................... 34

1.3.3 Autism and Non-syndromic XLID genes: ............................................................ 36

1.3.4 Evidence of X-chromosomal involvement in autism from measures of skewed X-inactivation ....................................................................................................................... 37

1.4 Thesis Objectives .............................................................................................................. 37

1.4.1 Chapter 2 ............................................................................................................... 37

1.4.2 Chapter 3 ............................................................................................................... 38

1.4.3 Chapter 4 ............................................................................................................... 38

Chapter 2. Chromosome X CNVs in Autism ......................................................................... 39

2.1 Introduction .................................................................................................................... 40

2.2 Methods: ....................................................................................................................... 41
2.2.1 DNA Samples ........................................................................................................ 41
2.2.2 Microarray and karyotyping experiments ............................................................ 42
2.2.3 CNV Analysis ....................................................................................................... 44
2.2.4 Validation of CNVs ............................................................................................ 44
2.2.5 Identification of Candidate Loci ......................................................................... 44

2.3 Results ...................................................................................................................... 46
2.3.1 Chromosome X CNVs in ASD cases ................................................................. 46
2.3.2 Validation of Interesting Loci and Identification of autism candidate genes ...... 52

2.4 Discussion ............................................................................................................... 62

Chapter 3. Analysis of X-Linked Autism Candidate Genes TSPAN7 & IL1RAPL1 ....... 65
3.1 Introduction ............................................................................................................... 66
3.2 Methods .................................................................................................................. 68
3.2.1 Samples ............................................................................................................. 68
3.2.2 PCR and Sequencing ...................................................................................... 68
3.2.3 Expression studies ......................................................................................... 68
3.3 Results ................................................................................................................... 72
3.3.1 TSPAN7 .......................................................................................................... 72
3.3.2 IL1RAPL1 ....................................................................................................... 74

3.4 Discussion ............................................................................................................... 77

Chapter 4. Disruption at the PTCHD1 locus on Xp22.11 in autism spectrum disorder and intellectual disability ................................................................. 80
4.1 Abstract .................................................................................................................. 84
4.2 Introduction ............................................................................................................ 84
4.3 Results .................................................................................................................. 85
4.3.1 CNV Analysis of PTCHD1 ........................................................................... 85
4.3.2 Mutation Screening of PTCHD1 ................................................................. 86
4.3.3 CNVs upstream of PTCHD1 (PTCHD1AS1/PTCHD1AS2 locus) ....................... 87
4.3.4 Expression and Functional Studies of PTCHD1 ............................................. 88
4.4 Discussion ............................................................................................................. 89
4.5 Methods .................................................................................................................. 92
  4.5.1 Source of Subjects ............................................................................................ 92
  4.5.2 Copy Number Variation Analysis .................................................................... 93
  4.5.3 DNA Sequencing and Mutation Screening ..................................................... 94
  4.5.4 X-Inactivation Studies ...................................................................................... 95
  4.5.5 Expression Analysis and Protein Localization ............................................... 95
  4.5.6 Luciferase Assays ........................................................................................... 96
4.6 Supplementary Information ................................................................................... 103
  4.6.1 Cytogenetic and CNV analysis of proband from Family 9 .............................. 103
  4.6.2 RT-PCR failed to find evidence for a shortened 3’ PTCHD1 transcript from individual with PTCHD1 exon 1 deletion ...................................................... 103
  4.6.3 Consensus Sequence for PTCHD1AS1: ......................................................... 104
  4.6.4 Consensus Sequence for PTCHD1AS2: ......................................................... 104
  4.6.5 RT-PCR and 5’ RACE (Rapid Amplification of cDNA Ends) analysis of the ncRNAs, PTCHD1AS1 and PTCHD1AS2 and the PTCHD1 gene .................................. 105
  4.6.6 Alternative 5’ exons for PTCHD1AS1, identified by 5’RACE: ....................... 107
  4.6.7 Putative promoter and enhancer sequences in intergenic region between DDX53 and PTCHD1 ........................................................................................................ 108
  4.6.8 eQTL at PTCHD1 locus .................................................................................. 109
Chapter 5. Future Directions ....................................................................................... 140
  5.1 Screening of additional autism families for CNVs using improved microarray platforms ............................................................................................................. 140
  5.2 Analysis of splice variation and isoforms at the PTCHD1 locus ......................... 140
  5.3 Induced pluripotent stem (iPS) studies ............................................................... 141
  5.4 Zebrafish knockdown experiments ..................................................................... 141
5.5 Mouse models ................................................................................................................. 142
5.6 Next Generation Sequencing of the entire chromosome X ........................................ 142
5.7 Development of Potential Therapies ........................................................................... 142

Chapter 6. References ........................................................................................................... 144
List of Appendices ...................................................................................................................... 176
List of Tables

Table 1-1 Categories of intellectual disability by IQ and ability to function in society as defined by the DSM-IV-TR. ................................................................. 25

Table 2-1 Chromosome X CNVs identified in 427 ASD cases................................................. 47

Table 2-2 Autism Specific CNVs are listed. All genomic coordinates are based on hg17: Build 35................................................................................................................................. 48

Table 2-3 Autism Specific Stringent CNVs............................................................................. 51

Table 2-4 CNV regions validated with qPCR........................................................................ 55

Table 3-1 Primers used to amplify the coding regions and splice sites of TSPAN7............. 69

Table 3-2 Primers used to amplify the coding regions and splice sites of IL1RAPL1......... 70

Table 3-3 Primer sequences used for qPCR validation and cDNA amplification of TSPAN7 duplication........................................................................................................ 71

Table 3-4 Primer sequences used for amplification of IL1RAPL1 cDNA............................. 72

Table 4-1 Primers used to amplify all three exons of PTCHD1 ............................................. 94

Table S 1 Clinical description of cases with disruptions at the PTCHD1 locus on Xp22.11..... 110

Table S 2 Breakpoint of deletions at the PTCHD1 locus: ..................................................... 117

Table S 3 Additional CNVs in 9 subjects with upstream deletions.......................................... 119

Table S 4 Gene co-expressed with PTCHD1....................................................................... 123

Table S 5 Summary of Samples Analyzed in the Study:....................................................... 127
List of Figures

Figure 2-1 Major steps involved in processing of Affymetrix 500K microarrays .................. 43

Figure 2-2 Workflow for mapping autism susceptibility genes using CNVs on the X chromosome is outlined. .......................................................................................................................... 45

Figure 2-3 Genomic region showing a 167 Kb deletion (blue line) at Xp22.11 which involves Exon 1 of the \textit{PTCHD1} gene. .................................................................................................................. 56

Figure 2-4 Pedigree showing the segregation of deletion at the \textit{PTCHD1} locus ............... 56

Figure 2-5 Genomic region showing an 82 Kb deletion (blue line) at Xp21.3 in intron 5 of \textit{IL1RAPL1} gene. ..................................................................................................................... 57

Figure 2-6 Genomic region showing a 485 Kb duplication (red line) at Xq22.3 which involves the \textit{IL1RAPL2} gene. .................................................................................................................. 58

Figure 2-7 Genomic region showing a 172 Kb duplication (red line) at Xq23 which involves the \textit{IL13RA2} and \textit{LRCH2} genes. ........................................................................................................ 58

Figure 2-8 Genomic region showing a 121 Kb duplication (red line) at Xp11.4 which involves the \textit{TSPAN7} gene. .................................................................................................................. 59

Figure 2-9 Genomic region showing a 505 Kb duplication (red line) at \textit{IDS} locus (Xq28) ....... 60

Figure 2-10 Genomic region showing a 5.8 Mb deletion (blue line) at X22.33-p22.31 which involves the \textit{NLGN4} gene. ......................................................................................................................... 60

Figure 2-11 Genomic region showing a 4.6 Mb duplication (red line) at Xp11.23-p11.22 which involves more than 50 RefSeq genes. ......................................................................................................................... 61

Figure 3-1 Agarose gel shows the amplification of \textit{TSPAN7} cDNA (bands) using lymphoblast RNA of patient with intragenic duplication of \textit{TSPAN7}. (A) 732 bp PCR product shows the amplification of \textit{TSPAN7} cDNA containing Exons 1-7. (B) 381 bp PCR product shows the amplification of Exons 1-3 of \textit{TSPAN7} cDNA ...................................................................................................................... 73
Figure 3-2 PCR amplification using primers in the deleted region. ............................................. 75

Figure 3-3 Chromatogram shows a C to T substitution at cDNA nucleotide position 349. Pedigree shows the maternal inheritance of the variant to the male proband. ................................. 75

Figure 3-4 Conservation of Alanine residue at position 113 in the IL1RAPL1 protein. .............. 76

Figure 4-1 Genomic organization of the PTCHD1 locus. ............................................................ 97

Figure 4-2 Pedigrees of families showing segregation of PTCHD1 mutations. ......................... 99

Figure 4-3 Expression analysis ................................................................................................... 101

Figure 4-4 Expression and functional studies of PTCHD1 ........................................................ 102

Figure S 4-1 PTCHD1 missense variants. Electropherograms indicate the nucleotide substitutions within PTCHD1 in six unrelated ASD families and two ID families. ................................. 133

Figure S 4-2 PTCHD1 domain structure .................................................................................... 134

Figure S 4-3 Quantitative RT-PCR for PTCHD1 in human brain regions. PTCHD1 expression in 24 regions of human adult brain is shown. Relatively higher expression was observed in the cerebellum. .................................................................................................................................. 136

Figure S 4-4 PTCHD1 functional analysis. 10T1/2 cells were transiently transfected with β-galactosidase to normalize for transfection efficiency, and Gli2, PTCH1, PTCH2 or PTCHD1. PTCHD1 exerted a statistically significant inhibitory effect on Gli-dependent transcription, similar to PTCH1 and PTCH2 (** PTCHD1: \( p = 0.0061 \); PTCH1: \( p = 0.0024 \); PTCH2: \( p = 0.0010 \)). Statistical significance (\( p \) below 0.05) was calculated using the Student’s \( t \)-test. Standard error bars are shown. .................................................................................................... 137

Figure S 4-5 Comparative and phylogenetic analysis of human Patched-related proteins. Phylogram of Patched-related homologues: (Homo sapiens), created using CLUSTALW 2.0.12 (www.ebi.ac.uk), with N-J tree–type. The phylogram is assumed to be an estimate of a
phylogeny, where branch lengths are proportional to the amount of inferred evolutionary change.

Figure S 4-6 SNP coverage at the PTCHD1 locus across different genotyping platforms. The SNP\CNV probes on Affymetrix 500K, Affymetrix 6.0 and Illumina 1M arrays are shown. ...
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>Autism Spectrum Disorder</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>Pervasive developmental disorder not otherwise specified</td>
</tr>
<tr>
<td>BAP</td>
<td>Broad autism phenotype</td>
</tr>
<tr>
<td>ADOS</td>
<td>Autism Diagnostic Observation Schedule</td>
</tr>
<tr>
<td>ADI-R</td>
<td>Autism Diagnostic Interview-Revised</td>
</tr>
<tr>
<td>M-CHAT</td>
<td>Checklist for Autism in Toddlers-modified</td>
</tr>
<tr>
<td>ID</td>
<td>Intellectual Disability</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligent quotient</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>ADM</td>
<td>Autism Dysmorphology Measure</td>
</tr>
<tr>
<td>FXS</td>
<td>Fragile X syndrome</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>TSC</td>
<td>Tuberous Sclerosis Complex</td>
</tr>
<tr>
<td>MIM</td>
<td>Mendelian Inheritance in Man</td>
</tr>
<tr>
<td>JSRD</td>
<td>Joubert Syndrome Related Disorder</td>
</tr>
<tr>
<td>PKU</td>
<td>Phenylketonuria</td>
</tr>
<tr>
<td>MZ</td>
<td>Monozygotic</td>
</tr>
<tr>
<td>DZ</td>
<td>Dizygotic</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>LOD</td>
<td>logarithm of odds</td>
</tr>
<tr>
<td>NPL</td>
<td>Non-parametric LOD</td>
</tr>
<tr>
<td>AGPC</td>
<td>Autism Genome Project Consortium</td>
</tr>
<tr>
<td>SNP</td>
<td>Nucleotide polymorphism (SNP)</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait loci</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>AGP</td>
<td>Autism Genome Project</td>
</tr>
<tr>
<td>aCGH</td>
<td>Array-based comparative genomic hybridization</td>
</tr>
<tr>
<td>MR</td>
<td>Mental Retardation</td>
</tr>
<tr>
<td>S-ID</td>
<td>Syndromic intellectual disability</td>
</tr>
<tr>
<td>NS-ID</td>
<td>Non-syndromic intellectual disability</td>
</tr>
<tr>
<td>XLID</td>
<td>X-linked intellectual disability</td>
</tr>
<tr>
<td>XCI</td>
<td>X-chromosome inactivation</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>DGV</td>
<td>Database of Genomic Variants</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>AA</td>
<td>Amino acid</td>
</tr>
<tr>
<td>BP</td>
<td>Base pair</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
</tbody>
</table>
Chapter 1. Background

The review of literature presented in this chapter is partly published in;

1. Book chapter

Common Genetic Etiologies and Biological Pathways Shared Between Autism Spectrum Disorders and Intellectual Disabilities.

Liana Kaufman, Abdul Noor, Muhammad Ayub and John B. Vincent

Autism Spectrum Disorders: The Role of Genetics in Diagnosis and Treatment,


2. Review

The genetic basis of non-syndromic intellectual disability: a review.

Kaufman L, Ayub M, Vincent JB.

1.1 Autism

Autism (MIM 209850) is a severe, neurodevelopmental disorder characterized by impairments in communication, socialization, and repetitive behavior. It was first outlined by Leo Kanner in 1943 as a disorder of innate inability of formation of the usual, biologically provided, affective contact with people. The Autism Spectrum Disorder (ASD) includes autistic disorder, Asperger syndrome, pervasive developmental disorder not otherwise specified (PDD-NOS) and Rett syndrome (DSM-IV, 1994; ICD-10, 1992). These disorders differ from each other with regard to severity of symptoms and early development of language, cognitive and social behavior. Individuals with autism show deficits in all three domains and an abnormal development before age 3 years. Individuals with Broad autism phenotype (BAP) have some symptoms of autism, but do not meet the full criteria for autism or ASD (Hurley, Losh, Parlier, Reznick, & Piven, 2007). Asperger syndrome is characterized by qualitative impairment in social interaction and restricted repetitive and stereotyped patterns of behavior, interests and activities, however, the language and cognitive development is relatively unaffected (McConachie & Diggle, 2007). Individuals with PDD-NOS meet autism criteria and these individuals may also show severe and pervasive impairment in one or two of the three core areas with or without cognitive or language delay. Rett syndrome almost exclusively occurs in females and it is characterized by developmental arrest between 6 and 18 months of age, followed by loss of speech, stereotypical movements, microcephaly, seizures, and intellectual disability (Hagberg, Aicardi, Dias, & Ramos, 1983).

The symptoms of autism spectrum disorders (ASDs) are usually present by age of three years and may persist throughout the life. According to DSM-IV, a child meets the diagnostic criteria for autism if at least six of the 12 behaviors, at least two from (1), and one each from (2) and (3) described in the three domains are being documented;
1. **Qualitative impairment in social interaction, as manifested by at least two of the following:**
   a. marked impairment in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction.
   b. failure to develop peer relationships appropriate to developmental level.
   c. a lack of spontaneous seeking to share enjoyment, interests or achievements with other people (e.g., by a lack of showing, bringing or pointing out objects of interest.
   d. lack of social or emotional reciprocity.

2. **Qualitative impairments in communication as manifested by at least one of the following:**
   a. delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gesture or mime).
   b. in individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others.
   c. stereotyped and repetitive use of language or idiosyncratic language.
   d. lack of varied, spontaneous, make-believe play or social imitative play appropriate to developmental level.

3. **Restricted, repetitive and stereotyped patterns of behavior, interests and activities, as manifested by at least one of the following:**
   a. encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus.
   b. apparently inflexible adherence to specific nonfunctional routines or rituals.
   c. stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting, or complex whole-body movements).
   d. persistent preoccupation with parts of objects.

The Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) are two widely accepted instruments used for diagnosis of ASD in both clinical and research settings. ADI-R, a revised version of Autism Diagnostic Interview (ADI), is a semi-structured, investigator-based interview for the caregivers of children with autism and adults for whom autism or ASD is a possible diagnosis (Lord, Rutter, & Le, 1994). The ADOS is a semi-
structured, standardized assessment of social interactions, communication, play, and imaginative use of objects for children suspected of having ASD (Lord et al., 2000). It is an observational assessment of the child’s behavior, often performed by a psychologist or another trained professional. There are some disadvantages of ADI-R and ADOS, for instance, these are lengthy, require elaborate training and are suitable for use in more specialized settings. However, in recent years, ADI-R and ADOS have been adapted to make them more appropriate for use in clinical settings, as well as for diagnosis of toddlers and patients with intellectual disabilities. In particular, the shorter version of ADOS is becoming increasingly popular in clinics.

Checklist tools such as The Checklist for Autism in Toddlers-modified (M-CHAT) are also widely used in clinical practice because of their ease and efficiency. M-CHAT includes a checklist of 23 items to be filled out by parents and it can be administered at a much earlier stage to identify toddlers who are at the risk of autism. A recent study has confirmed the validity of this instrument in detecting possible ASD at 16-30 months of age (Kleinman et al., 2008). However, the high sensitivity of this checklist means that some children without autism will fail the screening. It has been suggested that children who fail and do not have autism are at increased risk for other developmental disorders or delays and should be monitored accordingly (Kleinman et al., 2008).

In published literature, a wide range in the incidence of autism has been reported, with a worldwide trend consistently showing a steady increase in prevalence. During 1980s, ASDs were thought to be rare, with a prevalence of less than 5 per 10,000 persons (Gillberg, Steffenburg, & Schaumann, 1991) and were not categorized as major public health problem. During 1990s, the prevalence of autism was estimated to be 21 to 31 per 10,000 in preschool children (Fombonne, 1999). Afterward, an epidemiologic study conducted in United Kingdom report a prevalence rate of 16.8 per 10,000 for autism and 63 per 10,000 for all ASDs in children under the age of five years (Chakrabarti & Fombonne, 2001). A recent study in United States reported the incidence of ASD in ~1% of children age three to seventeen years, furthermore, it was estimated that 643,000 children in United States have ASD (Kogan et al., 2009). It is noteworthy that this study was based on parents’ reporting of ASD, and it could be argued that these estimates might be falsely high. However, at least two other recently published studies also report the prevalence of ASD to
be more than 1% (Baird et al., 2006; Baron-Cohen et al., 2009). On the other hand, it has been argued that the increasing incidence of autism might be due to increased awareness of public and professionals coupled with the broadening of the diagnostic criteria (Fombonne, Zakarian, Bennett, Meng, & McLean-Heywood, 2006). Today, the prevalence of ASDs is believed to be very high and this condition is now thought to be second only to Intellectual Disability (ID) among the most common developmental disabilities in the United States (Yeargin-Allsopp et al., 2003).

1.1.1 Endophenotypes

ASDs are clinically and etiologically complex neurodevelopmental disorders. It is has been emphasized that delineation of the clinical heterogeneity of ASD may help in identification of etiological factors and predict the outcome and treatment choices. ASD can be sub-grouped on the basis of presence or absence of certain clinical features, also termed as endophenotypes, such as intelligent quotient (IQ), seizures, brain malformations, dysmorphology and head circumference (Viding & Blakemore, 2007). Historically, based on non-verbal IQ testing, up to 70% of autistic children have reported to have some form of ID (Fombonne, 2003) and DSM-IV estimates that ~75% of children with autism have some degree of intellectual disability, usually within the moderate range (IQ 35-50). However, this is likely an over-estimate and more recent estimates propose a more modest level of ~60% (Chakrabarti & Fombonne, 2005). (Chakrabarti & Fombonne, 2001; Bertrand et al., 2001; Baird et al., 2001) Furthermore, in a study performed on an ID population, 28% met the criteria for an autism diagnosis on the ADI-R scale and only half of these had been previously diagnosed (Bryson, Bradley, Thompson, & Wainwright, 2008). These findings have been replicated repeatedly in the past, showing that within ID populations, the prevalence of autism is 8-20%, with more individuals with severe ID meeting criteria for ASD (Bryson et al., 2008; de et al., 2005; Stromme & Diseth, 2000; Nordin & Gillberg, 1996). ID and autism have several overlapping phenotypic domains. The three major phenotypes that are present in autism; language abnormalities, social deficits and stereotypies can all be present to varying degrees in some ID individuals. Individuals with ID often display stereotypies, which have a tendency to become more obvious, and often self-injurious, as IQ decreases. Studies have shown that 30-60% of individuals with ID exhibit some form of stereotypy (Bodfish, Powell,
Golden, & Lewis, 1995; Bodfish, Symons, Parker, & Lewis, 2000; Goldman et al., 2009) and language abnormalities are often particularly severe in individuals with severe-profound ID. Previously published longitudinal studies report that IQ scores can strongly predict the long-term outcomes and are directly associated with the psychopathology of autism, even in young children (Howlin, Goode, Hutton, & Rutter, 2004). Also, preschool cognitive functioning has been found to be a strong predictor of School-age functioning and high IQ has been shown to be necessary but not sufficient for optimal outcome in the presence of severe language impairment (Stevens et al., 2000).

Another, Central Nervous System (CNS) dysfunction associated with autism is the high risk of epilepsy (Spence & Schneider, 2009). The prevalence of seizures in autism is estimated to be up to 46% (Hughes & Melyn, 2005) and it has been estimated that as many as 32% of epilepsy patients may meet the diagnostic criteria for ASD (Clarke et al., 2005). Notably, the prevalence of seizures is higher among individuals with moderate to severe ID and those with motor abnormalities (Tuchman & Rapin, 2002). Furthermore, the individuals with autism plus epilepsy have, on average, lower IQs, and presence of epilepsy is a negative factors on cognitive, adaptive and behavioral/emotional outcomes for autistic individuals (Hara, 2007).

Structural brain malformations, including accentuated Virchow–Robin space, acrocallosal syndrome and polymicrogyria have been reported be associated with autism (Steiner, Guerreiro, & Marques-de-Faria, 2004; Schifter et al., 1994; Zeegers et al., 2006), however, until recently, MRI has been judged to be of insufficient value and it is not included in the standard clinical evaluation of autism. A recent study has revealed an unexpectedly high prevalence of brain abnormalities (48%) in autism patients, among these, the white-matter signal abnormalities, severely dilated Vicrchow-Robin spaces and temporal lobe structural abnormalities were the most common (Boddaert et al., 2009).

Generalized dysmorphology, an insult to early development has been reported in 15 to 20 % of individuals with autism (Miles & Hillman, 2000) and it has been suggested to be a predictor of a poor response to early intensive behavioral therapy. According to the Autism Dysmorphology Measure (ADM) guidelines, the 12 body areas assessed for dysmorphology are: height, hair growth pattern, structure and size of ear, nose size and shape, face size and structure, philtrum,
mouth and lips, teeth, hand size, fingers and thumbs, nails and feet. Besides generalized
dysmorphology, the head size abnormalities (microcephaly and macrocephaly) have also been
found in autistic individuals. Microcephaly, head circumference <2\textsuperscript{nd} centile, occurs in 5 to 15%
of children with autism and is a predictor of poor outcome (Miles, Hadden, Takahashi, &
Hillman, 2000; Miles et al., 2005). On the other hand, macrocephaly, head circumference >97\textsuperscript{th}
centile has been observed in ~30\% of children with autism (Miles et al., 2000). Generalized
dysmorphology and head circumference are proposed as good predictors of the clinical outcome
and may be used to classify the autism phenotype into subgroups, namely the complex autism
and essential autism. Complex autism consists of autistic individuals with the evidence of some
abnormality of early morphogenesis, manifested by either significant dysmorphology or
abnormal head size. The remainders, without dysmorphic features or head size abnormalities are
classified as essential autism (Miles et al., 2005), it is estimated that 70 to 80\% of children with
autism have essential autism. Another study compared the facial morphology of 72 boys with
ASD and their 128 first-degree relatives to that of 254 unrelated controls and found that
asymmetry of the supraorbital and periorbital regions anterior to the frontal cerebral pole has
been shown to be associated with autism in male patients (Hammond et al., 2008).

1.1.2 Autism in other single gene disorders

In many single gene disorders, clinical features of autism are also observed. For example, more
than 30\% of children with fragile X syndrome (FXS) have some autistic features (Macedoni-
Luksic et al., 2009) and 1 to 3\% of children ascertained on the basis of diagnosis of autism have
FXS (Harris et al., 2008). FXS (MIM #300624) is the most common cause of hereditary ID and
it results from an expansion of CGG trinucleotide repeats in the \textit{FMR1} (Fragile X Intellectual
disability-1) gene at Xq27-3. The degree of ID is usually mild to severe in males with FXS while
females usually demonstrate a mild degree of learning disability, although this can be more
severe in a small percentage (~25\%) of females (Cornish, Turk, & Hagerman, 2008). Molecular
studies indicate that the disruption of \textit{FMR1} gene may cause autism phenotype by RNA toxicity
to neurons and by silencing of genes involved in neuronal connectivity (Hagerman, 2006).
Tuberous Sclerosis Complex (TSC) is another genetic disorder which has an overlaps with
autism. It is an autosomal dominant disorder which involves multiple systems and it is
characterized by hamartomas in the brain, skin, heart, kidneys, lung and other organs. CNS symptoms may include epilepsy, learning difficulties, behavioral issues, and autism. Mutations in TSC1 and TSC2 cause this disease. Interestingly, individuals with TSC along with ID are more likely to have autism. By ADOS evaluation of TSC children, a recent study shows that 66% of infants meet criteria for autism or ASD at the age of 18 months, 54% at the age of 24 months, 46% at the age of 36 months and 50% at the age of 60 months (Jeste, Sahin, Bolton, Ploubidis, & Humphrey, 2008).

Autistic patients have a significantly higher (100- to 190-fold) risk of Neurofibromatosis type I (NF1) compared to the general population which suggests that the two diseases may share some common etiological factors (Marui et al., 2004). NF1 (MIM #162200) is caused by mutation in the neurofibromin gene (NF1) on chromosome 17q11.2. Clinical features of NF1 include the cafe-au-lait spots, Lisch nodules in the eye, and skin fibromatous, also, the individuals with NF1 have higher susceptibility to the development of benign and malignant tumors. It is unclear whether these relatively common childhood disorders have a true association or they co-occur by chance (Plank et al., 2001).

Sotos syndrome is characterized by congenital macrocephaly, a prominent forehead with an apparently receding hairline, overgrowth and mild to severe learning disability. Autosomal dominant mutations in the NSD1 gene on chromosome 5 are responsible for 80 to 90% of cases of Sotos syndrome (Buxbaum et al., 2007). The prevalence of disorder is estimated to be between 1/10,000 and 1/50,000 and its prevalence in autism was recently reported to be 0.5% (Zafeiriou, Ververi, & Vargiami, 2007). Although cases of autism and Sotos syndrome comorbidity have been reported, the genetic bases of this overlap are not clear.

Joubert Syndrome (JS) is an autosomal recessive disorder characterized by partial or complete agenesis of the cerebellar vermis which appears as the “molar tooth sign” on MRI, breathing difficulties, abnormal eye movement, ID, and behavioral problems. Mutations in nine genes (NPHP1, AHI1, CEP290, RPGRIP1L, TMEM67, ARL13B, CC2D2A, INPP5E and TMEM216) have been associated with JS. In our own study, we have identified a splice mutation in CC2D2A gene in a family with Joubert Syndrome Related Disorder (JSRD) (Noor et al., 2008). Subsequently, several mutations in this gene have been reported in JS, COACH syndrome and
Meckel syndrome (Tallila, Jakkula, Peltonen, Salonen, & Kestila, 2008; Doherty et al., 2010). In a previous study, 11 children with JS were assessed for diagnosis of autism and three out of 11 children with JS met diagnostic criteria for autism and one out of 11 for PDD-NOS (Ozonoff, Williams, Gale, & Miller, 1999). In another study, three sisters affected with JS were reported. Among these two sisters were monozygotic twins and the twin with the more severe cerebellar abnormality had autism (Raynes, Shanske, Goldberg, Burde, & Rapin, 1999). Both these studies were done on a very small population, therefore, the overlap between JS and autism remains inconclusive.

Autistic features have also been reported in several metabolic disorders including phenylketonuria (PKU), adenylosuccinate lyase deficiency and creatine deficiency syndromes. Co-morbidity of autism and untreated PKU has been described; however, due to severe ID in these children, it is difficult to assess these patients for autistic features. A systematic study investigated 243 PKU patients and showed that none of 62 early diagnosed and treated PKU patients met diagnostic criteria for autism, whereas two of 35 (5.7%) late diagnosed patients fulfilled the diagnostic criteria for ASD. These finding indicated that the classical PKU is one of the causes of autism with very low prevalence (Baieli, Pavone, Meli, Fiumara, & Coleman, 2003). Adenylosuccinate lyase deficiency is a rare autosomal disorder of purine synthesis which results in the increase of succinylpurines in body fluids. The clinical picture is heterogeneous; clinical manifestations include developmental delay, seizures, and autistic symptoms including failure to make eye contact, repetitive behavior, agitation, temper tantrums, and aggression. Approximately 50% of patients with adenylosuccinate lyase deficiency show an autistic-like phenotype (Stone et al., 1992).

As reviewed above, there are several genetic disorders which have co-morbidity with ASD. These clinical findings suggest that some disorders may have the potential to cause the autism phenotype by affecting the so far unknown autism regions or circuits of the brain.

1.1.3 Causes of Autism

1.1.3.1 Genetic

Epidemiological evidence suggests that genetic factors contribute significantly to the etiology of autism. The evidence of involvement of genetic factors in etiology of autism comes primarily
from the family and twin studies and this is further supported by the cytogenetic and molecular studies. A study in 1985, assessed the biological siblings of 29 autistic probands with severe ID and found a significant clustering of autism and nonspecific ID in the siblings of autistic individuals with severe ID (Baird & August, 1985). Another study evaluated the developmental, social, and psychiatric histories of the 67 adult siblings of 37 autistic probands and found that two out of 67 siblings (3.0%) were autistic, three out of 67 siblings (4.4%) had severe social problems, 10 out of 67 (15%) had cognitive disorders, and 10 out 67 (15%) had affective disorder (Piven et al., 1990). This report was the first to investigate the frequency of neuropsychiatric disorder in the adult siblings of autistic patients, and it demonstrated the enrichment of a variety of neuropsychiatric conditions among the siblings of autistic individuals. Bolton et al reviewed the family histories of 99 autistic patients and 36 probands with Down's syndrome and confirmed an increased familial aggregation of both autism and broader ASD in the siblings, and a similar trend was not observed among Down's syndrome families who were studied as a comparison group (Bolton et al., 1994). A recent study examined the cognitive, adaptive, social, imitation, play, and language abilities of 42 non-autistic siblings and 20 toddlers with no family history of autism. The siblings were below average in expressive language abilities and IQ; also, they had lower mean receptive language, adaptive behavior, and social communication skills. They used fewer words, distal gestures, and social smiles than children with no history of autism (Toth, Dawson, Meltzoff, Greenson, & Fein, 2007).

Collectively, the family studies show ~3% occurrence of autism among siblings of autistic probands. However, a recent study has estimated a significantly higher recurrence rate of 18.7% (Ozonoff et al., 2011). The occurrence rate supports familiality because the frequency is more than 100 times the occurrence in the general population, and this estimated relative risk is higher than for any other major psychiatric disease. On the other hand, the familiality of autism does not mean that genetic factors are exclusively responsible for the disease, and the role of the environmental factors which are also shared by family members who live together cannot be excluded as having a role in this observed familiality.

The twin studies provide an alternate approach to investigate the relative magnitude of genetic and/or environmental factors on the autism phenotype and penetrance. The higher the monozygotic (MZ) concordance rate, the more important the genetic contribution while
phenotypic differences between dizygotic (DZ) twins are due to non-genetic factors, including environmental factors. Several twin studies provide major evidence of genetic etiology of autism. In 1977, a landmark study by Folstein and Rutter showed a significant difference between monozygotic (N = 11) and dizygotic (N = 10) twins in their concordance for autism. Based on presence of absence of autism and ID, they observed a concordance rate of 82% and 10% for MZ twins and DZ twins, respectively (Folstein & Rutter, 1977). The concordance difference between MZ and DZ suggested a major role for genes in the etiology of autism and it was confirmed by subsequent studies (Ritvo et al., 1985; Steffenburg et al., 1989). Ritvo et al studied 40 pairs of affected twins and show a concordance rate of 95.7% in the monozygotic twins (22 of 23) and 23.5% in the dizygotic twins (Ritvo, Freeman, Mason-Brothers, Mo, & Ritvo, 1985). Another study of 21 twins under the age of 25 years and matched for gender reported a concordance rate of 91% in the monozygotic and 0% in the dizygotic twin pairs (Steffenburg et al., 1989).

Recently, a large scale study of 277 twin pairs (210 DZ) and 67 MZ) reported 88% concordance MZ twins and 31% concordance for DZ twins. In MZ twins, the authors also observed a higher prevalence of bipolar disorder and Asperger syndrome with a higher concordance of the latter (Rosenberg et al., 2009).

In conclusion, for autism, the twin studies show a higher concordance rate (60-90%) in MZ compared to concordance rate 0-30% in DZ and the heritability is estimated to be ~90%. These findings provide strong but indirect evidence of the role of genetic factors in the etiology of autism. Also, the twin and family studies have significantly contributed to our understanding of the causes of autism and triggered the search for causative genetic factors. Nevertheless, a recently published study have shown that the environmental factors may account for 55% of the liability to autism, which suggest a much lower contribution of genetics factors compared to previously published twin studies (Hallmayer et al., 2011).

1.1.3.2 Epigenetic

It is widely accepted that genetic factors play a major role in etiology of ASD; however, the involvement of epigenetic factors cannot be excluded. Epigenetic modifications include cytosine
methylation and post-translational modification of histones may act as a mechanism for control of gene expression (Samaco, Hogart, & LaSalle, 2005). The epigenetic modulation of gene expression can be influenced by exposure to the environmental factors and it can show the parent of origin effects. Notably, epigenetic factors play a central role in pathogenesis of two single gene disorders, Rett syndrome and fragile X syndrome (FXS) that are commonly associated with autism (Gillberg, 1986; Brown et al., 1982). Rett syndrome, a progressive neurodevelopmental disorder is classified among ASDs and it is caused by mutation in the MeCP2 gene that encodes the methyl-CpG-binding protein 2 which is involved in epigenetic regulation of gene expression (Amir et al., 1999). FXS is caused by expansion of a CGG repeats in the 5’ untranslated region of the FMR1 gene. This expansion results in epigenetic silencing of the region, causing the loss of expression of the gene, thus, FXS is caused by a genetic mutation resulting in epigenetic dysregulation (Hagerman, Ono, & Hagerman, 2005). RELN gene is another interesting example of possible contributions of epigenetic factors in ASD. RELN gene encodes a large extracellular matrix protein that organizes neuronal positioning during corticogenesis, one of the most notable effects of disruption of RELN is the abnormal formation of the cerebral cortex and inversion of cells in the horizontal laminations in mice (Caviness, Jr., 1976). Several independent studies have shown an association between RELN and ASD (Skaar et al., 2005; Ashley-Koch et al., 2007; Holt et al., 2010). Interestingly, reduced levels of reelin and its isoforms have been previously shown in autistic twins and their first degree relatives (Fatemi et al., 2005). RELN is not an imprinted gene, however, studies have shown a potential regulation of RELN expression by DNA methyltransferase 1 in mouse primary cortical cultures (Noh et al., 2005). In conclusion, the RELN gene has been shown to be associated with ASD, and its expression is possibly regulated by epigenetic mechanisms which further highlight the importance of epigenetic factors in pathophysiology of ASD. Genomic imprinting is another mode of regulation of gene expression by epigenetic modifications and it results in the parent of origin-specific gene expression. Interestingly, genomic duplications of an imprinted region on the proximal long arm of chromosome 15 (15q11-q13) are associated 0.5-3.0% of autism (Hogart, Wu, LaSalle, & Schanen, 2010). Therefore, genomic importing may also play a role in etiology of autism.
1.1.3.3 Other Factors

Support for the possibility of environmental contribution to causation of autism comes from the incomplete concordance in monozygotic twins, which cannot be accounted for by standard genetic mechanisms. Furthermore, there is evidence that in-utero exposure to valproic acid or thalidomide may increase the risk of ASD (Arndt, Stodgell, & Rodier, 2005). Recently, a long-term study of 632 children exposed to antiepileptic drugs during gestation found that 6.3% of the children in-utero exposed to valproic acid had ASD or some features of ASD. This incidence is seven times higher than the control group (0.9%) (Bromley, Mawer, Clayton-Smith, & Baker, 2008). Similarly, a higher incidence of autism has been reported among children prenatally exposed to thalidomide. In a population of 100 Swedish thalidomide embryopathy cases, at least four met full diagnostic criteria for autism (Stromland, Nordin, Miller, Akerstrom, & Gillberg, 1994; Rodier, 2002). Animal models have also demonstrated that in-utero exposure to thalidomide or valproic acid can carry an increased risk for the development of ASD. It has been shown that early serotonergic neural development is disrupted in rats exposed to thalidomide or valproic acid on the ninth day of gestation (Narita et al., 2010).

Mercury (Hg), because of its known neurotoxicity has drawn particular attention in relation to deficits in neurodevelopment of autism patients and a number of studies have compared the level of Hg in blood, hair, or urine in children with autism versus without autism. However, none of these studies has shown any substantial evidence of involvement of Hg in autism. Recently, a study conducted on 452 autism patients failed to demonstrate any difference in blood Hg level of autism patients compared to controls (Hertz-Picciotto et al., 2010).

Childhood immunization is an environmental factor that has been popularized in the media as a potential cause of autism. The use of mercury in vaccines has been one of the prime sources of concern surrounding vaccines and their role in autism (Baker, 2008). However, there is no consistent evidence in support of the theory that vaccines are related to the etiology of autism. In the late 1990s, a link between vaccines and autism was reported by clinical observation of onset of autism soon after vaccination of children (Wakefield & Montgomery, 1999). These observations triggered a series of studies in the US, UK, Europe and Japan, however, none of these studies found any compelling evidence for a link between vaccines and autism.
Although the majority of research to date has focused on genetic factors involved in the etiology of autism, non-genetic factors are also likely to contribute. Our knowledge of these factors is currently very limited. It has been suggested that distinct genetic features/pathways may cause distinct domains of autistic behavior, but this has yet to be tested at the molecular level (Happe & Ronald, 2008). It does, however, resonate with the idea that autism is a genetically heterogeneous spectrum, and that multiple genetic aberrations may be necessary to reach the autism phenotype threshold (Cook, Jr. & Scherer, 2008). The threshold theory postulates that the cumulative effect of several genetic aberrations, for instance a copy number variant together with one or more single nucleotide variants, and possibly in combination with environmental factors, in a single individual, may result in an autism phenotype. These genetic aberrations may include chromosomal, single nucleotide or epigenetic abnormalities. It has also been noted that some genetic aberrations are more penetrant than others and may be more likely to result in a phenotype. In contrast with ID genetics, which are relatively straightforward, autism presents us with a convoluted, likely multigenic/multifactorial disorder for which it may be more difficult to delineate causes.

1.1.4 Identification of Causative Genetic Factors

1.1.4.1 Linkage and Association Studies

Analysis of pedigrees with multiple incidences of autism would suggest that the mode of inheritance is neither clearly dominant nor recessive or X-linked, and is frequently described as non-Mendelian or complex. A high degree of genetic heterogeneity is anticipated. Genetic linkage analysis may be used to identify the regions of the genome that are shared between affected members in a family, and may be performed by using gene or locus specific markers or a set of markers covering the entire genome. The two commonly used models for linkage analysis are parametric and non-parametric. The results of Linkage analysis are reported as a logarithm of odds (LOD) score. For complex genetic disorders such as autism, in general, a LOD score greater than 3.6 is considered to be significant at the genome-wide level, while a score greater than 2.2 is considered as suggestive of linkage (Lander & Kruglyak, 1995). Because of the complex mode of inheritance in ASD, non-parametric linkage may be considered a more appropriate model for analysis. Relatively few parametric linkage studies have been reported. For example, Laumonnier et al used parametric linkage to study a large pedigree with 13 males
with autism and a statistically significant linkage was achieved at Xp22.33. By sequencing the NLGN4 gene in this region a 2 base pair (bp) deletion was identified in all affected patients (Laumonnier et al., 2004). Parametric linkage analysis has also successfully led to the identification of additional ASD loci, including the MECP2 gene, which is implicated in Rett syndrome (Amir et al., 1999), also the CNTNAP2 gene, which is involved in epilepsy, intellectual disability, and autism (Strauss et al., 2006).

A number of studies have used nonparametric or “model-free” linkage analysis to search for ASD genes. These studies have reported suggestive to significant evidence of linkage on almost every chromosome, however, only a few of these loci have been replicated (Abrahams & Geschwind, 2008). Among these replicated loci, 2q24-2q31, 7q, and 17q11-17q21 have been shown by at least two studies to have genomewide significant LOD scores (Abrahams & Geschwind, 2008). In a study from the Vincent and Gurling labs using markers on the X-chromosome, a maximum parametric LOD score of 1.7 was reported for the narrowest diagnostic category of the typical autism/severe autism spectrum, and nonparametric analysis produced a maximum non-parametric LOD (NPL) score of 2.1 for markers on Xq27-q28, encompassing the FMR1 (FXS) and MECP2 (Rett) genes, thus provided evidence of modest linkage to this region (Vincent et al., 2005). By using 10,000 single nucleotide polymorphism (SNP) markers, members of the Autism Genome Project Consortium (AGPC) performed a genome scan of about 1160 multiplex families with autism and reported suggestive evidence of linkage at 11p12-p13 (Szatmari et al., 2007). Recently, a linkage and association study was performed by using ~500,000 SNP markers in set of 1,031 multiplex families with autism, and suggestive linkage at 6q27 and a significant linkage at 20p13 was reported (Weiss, Arking, Daly, & Chakravarti, 2009). The linkage analysis studies have demonstrated that many loci may underlie the risk of autism; these findings are consistent with the well-accepted supposition that many genes may be associated with autism. This genetic heterogeneity may result in weak linkage signals and along with variation in clinical ascertainment and other diagnostic issues, may contribute to failure of replication of linkage signals. Several studies have attempted to overcome this issue by increasing the sample homogeneity, or by focusing on selected sub-phenotypes of autism (endophenotypes) such as large head circumference, language, developmental milestones and stereotyped patterns of behavior and interests. This approach
facilitates the identification of quantitative trait loci (QTL) for a trait that contributes to the overall phenotype. For example, in the study by Alarcon et al, a QTL for language delay was mapped at 7q34-7q36 (Alarcon, Cantor, Liu, Gilliam, & Geschwind, 2002).

Linkage analysis is a powerful method for identifying high-risk disease alleles, however, in the case of complex genetic diseases such as ASD in which it is anticipated that contributions of several loci result in the expression of phenotype, association studies may be more suitable to search for disease genes/loci. Family-based association studies compare the frequency of transmitted versus non-transmitted parental alleles of polymorphic markers in the affected individuals, whereas case-control association studies compare the frequencies of alleles in affected individuals with unaffected controls in order to detect alleles that differ significantly in frequency between the two groups (Carlson, Eberle, Kruglyak, & Nickerson, 2004). One advantage of the family-based association studies over the case-control association studies is that the former avoids the possible effects of population stratification, thus the possibility of false positive results are reduced. Frequently, association studies are performed using genetic markers selected from small chromosomal regions or from single genes. In most such studies, the selection of candidate region/gene for autism is based on its location within a significant autism linkage region, a known cytogenetic abnormality, or is based on its role in a biological pathway hypothesized to be involved in ASDs or its known function in brain and Central Nervous System (CNS) development (Alarcon et al., 2002).

In previously published association studies, over 100 ASD candidate genes/regions have been tested. Notably, most of these studies were inconclusive and failed to find significant and replicable genetic association with ASD. On the other hand, this approach has also resulted in the identification of some interesting candidate genes that may be involved in a sub-set of ASD patients. Among these, SLC25A12, RELN and CNTNAP2 are of interest.

At least two studies have mapped an ASD susceptibility locus at chromosome 2q (AUTS5). The first study analyzed 95 affected-relative-pair families and found a maximum multipoint heterogeneity LOD score of 1.96 and a maximum multipoint nonparametric linkage (NPL) score of 2.39 (Buxbaum et al., 2001). The second study was performed by the genome-wide linkage analysis of 152 autistic sib pairs and showed further support for this locus at 2q (IMGSAC,
One of the interesting candidate genes in this region is \textit{SLC25A12}, which encodes a calcium-binding mitochondrial carrier protein. The protein localizes to the mitochondria and is potentially involved in the exchange of aspartate for glutamate across the inner mitochondrial membrane (Palmieri et al., 1997). By genotyping two SNPs (rs2056202, rs2292813) within \textit{SLC25A12} gene in 411 autistic families, an association between autistic disorder and these two SNPs was reported (Ramoz et al., 2004). This association was investigated by another independent study - 158 Irish affected child-parent trios were genotyped for the rs2056202 and rs2292813 SNPs. The Transmission Disequilibrium Test was applied to examine these markers for association with autism. Interestingly, in agreement with the previous study, the authors found significant association and provided the replication of the association between autism and \textit{SLC25A12} (Segurado et al., 2005). These studies strongly associated this region with ASD, following which several other groups started further exploration of this region. Contrarily, at least three independent studies failed to find any association for this gene. Blasi et al used a collection of families from the International Molecular Genetic Study of Autism Consortium (IMGSAC) and failed to find any significant association for the SNPs tested at \textit{SLC25A12} locus, suggesting that the variants at this locus are unlikely to play a major role in genetic susceptibility to autism in their samples (Blasi et al., 2006a). In another study, Rabionet et al attempted to test for association in \textit{SLC25A12} in an independent data set of 327 families with autistic offspring (Rabionet et al., 2006). Again, this study was also unable to find an association between \textit{SLC25A12} and autism, which suggests that \textit{SLC25A12} is not a major contributor to autism risk in these families. A recent study of Han Chinese samples from Taiwan using a population-based case-control approach was also unable to find any evidence of association of the \textit{SLC25A12} gene with autism (Chien et al., 2010). The discrepant results of these studies may be due to the clinical heterogeneity of cases analyzed in these studies and the genetic heterogeneity of autism.

Several independent studies have mapped an autism susceptibility locus at chromosome 7q (Philippe et al., 1999; Barrett et al., 1999; IMGSAC, 2001). The Reelin gene, \textit{RELN}, maps to 7q and plays an important role in the migration of several neuronal cell types and in the development of neural connections (Del Rio et al., 1997; Ogawa et al., 1995). The physical position and function of this gene make it a good candidate and several studies have attempted to explore possible contribution of this gene in autism phenotype. Persico et al tested the association and linkage to this gene in 95 Italian patients in comparison with 186 ethnically-
matched controls, and using the transmission/disequilibrium test and haplotype-based haplotype relative risk was assessed in 172 complete trios from 165 families collected in Italy and in the USA. Both case-control and family-based analyses found a significant association between autistic disorder and a polymorphic GGC repeat located immediately 5’ of the RELN gene (Persico et al., 2001). A study by Skaar et al replicated the findings of association of 5'-UTR repeat of RELN with autism (Skaar et al., 2005). These two studies show the potential of RELN as an important contributor to genetic risk in autism. The RELN gene includes a polymorphic GGC repeat region in the 5’UTR. In a more recent study, authors transfected mammalian cells with constructs encompassing the RELN 5'UTR with 4-to-13 GGC repeats upstream of a luciferase reporter gene and demonstrated decreasing luciferase activity with increasing GGC repeat number, which further highlights the potential importance of the RELN GGC repeats in autism (Persico, Levitt, & Pimenta, 2006). An interesting study by Fatemi et al tested the expression levels of Reelin protein and mRNA, as well as mRNA levels for the Reelin receptor gene, VLDLR, and downstream markers of the Reelin pathway, Dab-1 and GSK3, in brain regions from autistic versus control individuals. Interestingly, reduced levels of the Reelin protein and mRNA and Dab 1 mRNA were observed, these findings emphasize the impairments in the Reelin signaling system in autism and may account for some of the brain structural and cognitive deficits observed in this disorder (Fatemi et al., 2005). Contrary to these findings, several other reports have not found any association between autism and RELN trinucleotide repeats (Li et al., 2004; Bonora et al., 2003; Krebs et al., 2002).

In a recent genome-wide association study from our AGP collaborative group, by testing ~1 million SNP markers for association with ASD, a strong association of rs4141463 in MACROD2 with autism was identified ($P < 5 \times 10^{-8}$). Strong association signals were also observed for several other genes such as KIAA0564, PLD5, POU6F2, ST8SIA2 and TAF1C (Anney et al., 2010).

1.1.4.2 Cytogenetic Analysis
The identification of the candidate genes for autism by linkage and association studies has proven to be a complex endeavor, and the analysis of cytogenetic abnormalities associated with autism has been proven to be a relatively successful alternative. It is estimated that
cytogenetically identified abrasions are present in 6-7% of ASD cases and this proportion is even higher in patients with intellectual disability and dysmorphic features (Marshall et al., 2008). These chromosomal abnormalities are widely distributed across the genome and none of these aberrations account for major fraction of individuals with ASDs, with the exception of a duplication on 15q11-q13, which is the most frequently (0.5 to 3.0 %) and consistently found chromosomal abnormality in ASD patients (Hogart et al., 2010). This region is of particular interest because of its relationship with other neurodevelopmental and behavioral syndromes. Deletion of the maternal copy of this interval, as well as mutations in the *UBE3A* gene within this interval, lead to Angelman syndrome (MIM# 105830), a disorder that has some overlap with severe autism phenotypes. Conversely, deletion of the paternal copy leads to Prader-Willi syndrome (MIM# 176270) characterized by prominent behavioral features including perseveration, obsessive-compulsive phenomena, and impulsive behavior (Vorstman et al., 2006). Duplications of this 15q11–13 region, mostly though not exclusively involving the maternal copy, have been reported in patients with ASD (Miller et al., 2009). This gene rich region encompasses several candidate genes such as *GABRB3* (GABA A receptor beta-3) and *UBE3A* (ubiquitin protein ligase E3A)- both highly expressed in CNS. However, no mutations have been identified in these genes in cytogenetically normal patients with ASDs, and no association of common alleles has been conclusively demonstrated. Chromosomal abnormalities have also been reported on many chromosomes, including chromosomes for which positive linkage or association has been established such as chromosome 2 and 7.

Characterization of cytogenetic aberrations has resulted in the identification of several ASD candidate genes including *RAY1/ST7* (suppression of tumorigenicity 7), *AUTS2* (autism susceptibility gene 2), *MMP16* (matrix metalloproteinase 16), *NBEA* (neurobeachin), *GRPR* (gastrin-releasing peptide receptor), *GABRG1* (GABA-A receptor gamma-1), *DSC1* (Desmocollin 1) and *DSC2* (Desmocollin 2). *RAY1/ST7* was identified by characterizing a translocation (t(7;13)(q31.3;q21)) disrupting this gene in an ASD patient (Vincent et al., 2000). *AUTS2* was implicated by characterization of the translocation t(7;20) (q11.2; p11.2) in a monozygotic twin pair concordant for autism (Sultana et al., 2002). To date, there are at least seven reports of translocations disrupting the *AUTS2* gene in autistic kids (Huang, Zou, Maher, Newton, & Milunsky, 2010). The *MMP16* gene was identified by breakpoint mapping of a de
novo, apparently balanced t(2;8)(q35;q21.2) translocation in a boy with developmental delay and autism. The 8q21.2 breakpoint was mapped within MMP16, which encodes the matrix metalloproteinase 16 protein (Borg et al., 2002). In another patient with idiopathic autism and no family history of autism, the NBEA gene was disrupted by a de novo translocation t(5;13)(q12.1;q13.2) (Castermans et al., 2003). In two brothers diagnosed with autism, Vincent et al reported a paracentric inversion of the short arm of chromosome 4 (46,XY, inv(4)(p12-p15.3)), the proximal breakpoint (4p12) mapped within a cluster of gamma-aminobutyric acid A (GABA(A)) receptor genes and directly disrupted the GABRG1 gene (Vincent et al., 2006). In a recent study from our group, we have reported a de novo balanced translocation t(5;18)(q33.1;q12.1) in a boy with autism. Further molecular characterization revealed that the 5 breakpoint lies at the 3' end of the SH3TC2 gene and distal to beta-adrenergic receptor gene ADRB2 and serotonin receptor gene HTR4, while the 18q breakpoint lies between desmocollin genes DSC1 and DSC2. We attempted to check the possibility of mono-allelic expression of these genes due to a position effect of the translocation interfering with gene expression; interestingly, the DSC1 and DSC2 were only transcribed from the normal chromosome 18 in lymphocytes from the proband. DSC1 and DSC2 play important role in cell adhesion and desmosome formation, disruption of these genes may contribute to etiology of autism (Vincent et al., 2009).

1.1.4.3 Copy Number Variation

For many years, the single nucleotide polymorphism (SNP), small insertion-deletions, variable numbers of repetitive sequences, microsatellite and minisatellite were the known forms of genetic variations. However, with the completion of human genome project and availability of human reference sequence coupled with advances in the microarray technology, a new form of genomic structural variation, the submicroscopic copy number variants (CNVs) were revealed in 2004 (Iafrate et al., 2004; Sebat et al., 2004). The term CNV was introduced in 2006 and a CNV was defined as ‘a segment of DNA that is 1 kb or larger and is present at a variable copy number in comparison with a reference genome’ (Feuk, Carson, & Scherer, 2006). These genomic structural variants can be genomic copy number gains, insertions, or losses relative to a designated reference genome sequence. In a landmark study, Iafrate and colleagues used array-based comparative genomic hybridization (array CGH) to analyze the genomes of 55 unrelated
individuals. Using this approach, they identified 255 loci across the human genome that had genomic imbalances among unrelated individuals. Among these, 24 variants were identified in > 10% of the examined individuals. Interestingly, 50% these variants encompassed annotated genes and many coincided with segmental duplications or gaps in the human genome assembly (Iafrate et al., 2004). Another report published at the same time identified 221 copy number differences among 20 unrelated individuals analyzed on oligonucleotide microarrays. Out of these 221 structural variants, 76 were unique and on average 11 variants per genome were identified with an average length of a 465 kb (Sebat et al., 2004). Similar to Iafrate et al, this study also found that many of these variants encompassed known genes, including genes involved in neurological function, regulation of cell growth, regulation of metabolism, and disease associated genes (Sebat et al., 2004). Both these reports uncovered the dynamic nature of the human genome and suggested the possible involvement of these variants in normal human variation and in genetic defects. Since then, microarrays and other tools for discovery of CNVs have been substantially improved and this approach has been very successful in the identification of genomic regions\genes involved in different genetic disorders, and in particular, complex genetic disorders such as autism.

The first comprehensive CNV study of autism was published in 2007, through a genome scan of 715 families (1,109 samples) using 10K SNP arrays, in which 2,788 putative CNVs were identified (Szatmari et al., 2007). Further quality controls measures resulted in the identification of 624 stringent CNVs from 350 different families. This study reported several autism susceptibility genes\loci including NRXN1, 1q21 and 22q11.2 (Szatmari et al., 2007).

Using comparative genomic hybridization (CGH) on the genomic DNA of 264 families, Sebat et al tested the hypothesis that de novo CNVs are associated with autism. Interestingly, prevalence of de novo CNVs was significantly higher in cases (10%) compared to controls (1%) which confirmed the strong association of de novo CNVs with autism (P = 0.0005) (Sebat et al., 2007).

In a study involving our collaborative group, using Affymetrix 500K microarray, a genome scan was performed on 427 ASD families, and 277 unbalanced CNVs in 44% of ASD families were identified. These CNVs were not present in ~1600 controls (Marshall et al., 2008). Amongst the patients, 11% had de novo CNVs, which is comparable to Sebat et al (10%). Our data further
implicated some of the previously known ASD susceptibility genes such as \textit{SHANK3}, \textit{NLGN4} and \textit{NRXN1}. Furthermore, we identified several novel ASD candidate loci\ genes such as \textit{DPP6}, \textit{DPP10}, \textit{PCDH9}, \textit{ANKRD11}, \textit{DPYD}, \textit{PTCHD1}, and 15q24. Most importantly, we showed that CNVs at 16p11.2 are associated with autism in \(\sim 1\)% of the cases (\(p = 0.002\)). Subsequently, the association of the recurrent 16p11.2 CNVs (micro-deletions and reciprocal micro-duplications) with 1\% ASD cases was further confirmed by several independent studies (Weiss et al., 2008). We also highlighted several CNV events spanning previously known ID genes, indicating possible etiological overlap between autism and ID (Marshall et al., 2008). The detailed characterization of chromosome-X CNVs identified in this study is the primary focus of this dissertation and will be elaborated on in detail in the data chapters.

Another CNV study of 859 ASD cases and 1,409 healthy children identified several new CNV regions potentially involved in etiology of autism. The authors validated their positive findings in an independent cohort of 1,336 ASD cases and 1,110 controls (Glessner et al., 2009). This data supported the involvement of some of the previously reported ASD candidate genes, such as \textit{NRXN1} and \textit{CNTN4} and identified many new ASD susceptibility genes such as \textit{NLGN1}, \textit{ASTN2}, \textit{UBE3A}, \textit{PARK2}, \textit{RFWD2}, \textit{FBXO40} and \textit{AK123120}. It was noticed that several of the genes disrupted by CNVs were involved in neuronal cell-adhesion or ubiquitin degradation, which suggested that genes involved in these biological processes may play a key role in the susceptibility to ASD (Glessner et al., 2009). CNVs at 15q11-q13 (including \textit{UBE3A}) and other previously know loci were also confirmed in another study of exonic CNVs in autism patients (Bucan et al., 2009).

Our collaborative group recently reported a comprehensive map of CNVs in 996 ASD individuals of European ancestry and \(\sim 5000\) ethnically matched controls (Pinto et al., 2010). Our finding confirmed a higher global burden of rare, genic CNVs (1.19 fold, \(P = 0.012\)) in ASD cases, especially, for the previously implicated loci. Furthermore, similar to our finding in Marshall et al, we found significant enrichment of CNVs at known ID loci (1.69 fold, \(P = 3.4 \times 10^{-4}\)). We identified numerous \textit{de novo} and inherited events which resulted in identification of new ASD susceptibility loci such as \textit{SHANK2}, \textit{SYNGAP1}, \textit{DLGAP2} and the X-linked \textit{DDX53-PTCHD1} locus. We have also grouped genes disrupted by these CNVs on the basis of their
function, which indicated that they are mainly involved in cellular proliferation, projection and motility, and GTPase/Ras signaling (Pinto et al., 2010).

CNV analysis coupled with direct sequencing has proved to be a powerful approach to uncover disease causing mutations in several genes. For example, in case of *SHANK2* (synaptic scaffolding gene) gene, Berkel et al initially found de novo CNVs in this gene in two unrelated individuals with ASD and ID. Subsequent sequencing of coding regions of *SHANK2* in 396 ASD cases and 184 ID cases identified a de novo nonsense mutation and seven rare inherited changes. *SHANK2* is another example, where mutations in a gene can cause ID or Autism or both (Berkel et al., 2010). Using similar strategy, we identified *PTCHD1* mutations in ASD and/or ID cases, discussed in detail in chapter 4.

The study of CNVs in ASD has proven to be a successful approach in delineating a portion of the etiology of autism. These studies have accelerated the pace of discovery of causative genetic factors for autism and resulted in increasing our understanding of this complex disorder. Several CNV studies have highlighted some critical biological pathways, and further investigation of these pathways may yield novel targets for therapeutic interventions.

### 1.1.4.4 Candidate Gene Sequencing

Direct sequencing of candidate genes is another rapid approach for identification of disease genes. The candidate genes can be prioritized on the basis of protein function, expression in brain regions or physical position of gene such as genes in previously linked susceptibility regions.

Using this approach, causative mutations have been previously reported in *NLGN3* and *NLGN4* genes (Jamain et al., 2003). These genes map within previously linked regions of chromosome X (Auranen et al., 2002) and these are involved in cell-adhesion with important function in synaptogenesis during brain development and in connection of pre and postsynaptic membranes. A frameshift mutation (1186insT) in *NLGN4* has been reported in a family with two affected brothers, one with typical autism and the other with ASD and a missense mutation (R451C) in *NLGN3* in another family (Jamain et al., 2003). In another large family, a two base pair deletion
within NLGN4 was reported, the mutation segregated with phenotype, X-linked intellectual disability and three males with ASD (Laumonnier et al., 2004). Another study that involved performing extensive mutation screening of NLGN3 and NLGN4 revealed four new missense mutations (Yan et al., 2005). Recently, a de novo 1 base pair (335G>A) substitution located in the promoter region of NLGN4 has been reported in a patient with autism and nonsyndromic profound ID (Daoud et al., 2009). However, several other studies failed to identify any causative variants in NLGN3\NLGN4, and therefore the mutations in these genes are proven to be very rare (Vincent et al., 2004; Wermter, Kamp-Becker, Strauch, Schulte-Korne, & Remschmidt, 2008).

By chromosomal analysis and DNA sequencing several mutations in SHANK3 (also known as ProSAP2) has been recently reported. In one the families, a heterozygous, 1bp insertion has been reported in two brothers with autism, both brothers had severely impaired speech and severe ID (Durand et al., 2007). A follow up study performed the sequencing of 400 cases and identified a de novo mutation. Additionally, by CNV analysis two gene deletions were also discovered. Combining the sequence and CNV mutations, the frequency of SHANK3 mutation was estimated to be 0.75% in the study cohort (Moessner et al., 2007).

By systematic sequencing of X chromosomal synaptic genes, IL1RAPL1 (Interleukin-1 Receptor Accessory Protein-Like 1) has been recently implicated in autism (Piton et al., 2008). All coding exons of IL1RAPL1 were sequenced in a cohort of 142 subjects (20 females and 122 males) and a seven base pair deletion was identified in a French-Canadian girl diagnosed with ASD without language delay. It is noteworthy that this gene was initially implicated in non-syndromic ID (Bhat et al., 2008a), however, the girl reported here did not have ID.

In our own attempt to sequence candidate genes, we have identified several disease associated mutations in PTCHD1 and IL1RAPL1 genes, which will be elaborated in detail later in this dissertation.

In conclusion, numerous studies have attempted to find causative mutation by sequencing candidate genes but with a limited success, possibly, due to the high degree of genetic heterogeneity of this disorder. Nevertheless, this approach has found many causative variants in different genes, as reviewed above. The study of these genes will help in the diagnosis of more
cases in the future, and will help us understand the function these gene and the biological pathways in which they play a role.

### 1.2 Intellectual Disability (ID)

Intellectual disability (ID) or Mental Retardation (MR) is characterized by an intelligence quotient (IQ) of 70 or below coupled with deficits in at least two adaptive behaviors, which include skills related to everyday life (American Psychiatric Association 2000). On the basis of IQ, ID is divided into five subtypes: Mild, moderate, severe, profound and unable to classify (DSM IV) and the prevalence of ID is estimated to be up to 3% in general population (Leonard & Wen, 2002a).

<table>
<thead>
<tr>
<th>Severity</th>
<th>IQ</th>
<th>Proportion of ID</th>
<th>Functional Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borderline</td>
<td>70-84</td>
<td>N/A</td>
<td>Normal</td>
</tr>
<tr>
<td>Mild</td>
<td>50-69</td>
<td>85%</td>
<td>Can often live independently with social support</td>
</tr>
<tr>
<td>Moderate</td>
<td>35-49</td>
<td>10%</td>
<td>Acquire some communication and self-help skills,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>require moderate supervision</td>
</tr>
<tr>
<td>Severe</td>
<td>20-34</td>
<td>3-4%</td>
<td>Acquire only basic self-help and communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>skills, require supervision</td>
</tr>
<tr>
<td>Profound/Unspecified</td>
<td>20-34</td>
<td>1-2%</td>
<td>Require highly structured and supervised living</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conditions</td>
</tr>
</tbody>
</table>

Table 1-1 Categories of intellectual disability by IQ and ability to function in society as defined by the DSM-IV-TR.

While the prevalence of severe ID is relatively stable, the prevalence of mild ID is variable and often depends heavily on external environmental factors such as level of maternal education,
access to education/opportunity and access to healthcare (Leonard & Wen, 2002b; Drews, Yeargin-Allsopp, Decoufle, & Murphy, 1995; Roeleveld, Zielhuis, & Gabreeels, 1997). Study design, age of subjects, and the catchment population for various epidemiological studies may also contribute to the variability seen across mild ID prevalence studies (Leonard & Wen, 2002b; Roeleveld et al., 1997).

The new DSM-V will be changing the severity criteria to encompass behavioral deficits and the level of impact that these have in the lives of affected individuals (DSM5.org). The manual will also be changing the wording of the ID diagnostic criteria to encourage cultural sensitivity and relevance, and to ensure that culturally validated psychometric tests are used to evaluate IQ and level of functioning.

In addition to categorization by severity/IQ level, ID can also be grouped into syndromic intellectual disability (S-ID) and non-syndromic intellectual disability (NS-ID). In S-ID, individuals present with one or multiple clinical features or co-morbidities in addition to ID. While S-ID has a clear definition, there is debate over the classification of NS-ID. Traditionally, NS-ID has been defined by the presence of intellectual disability as the sole clinical feature. However, it has been a challenge to rule out the presence of more subtle physical signs, neurological anomalies and psychiatric disorders in these individuals, as they may be less apparent, or difficult to diagnose due to the cognitive impairment. Additionally, the symptoms of some syndromes may be so subtle that they are extremely difficult to diagnose unless the features are looked for specifically in the context of a known genetic defect previously associated with these features (Ropers, 2006a). Thus the distinction between S-ID and NS-ID is often blurred.

Despite its universal occurrence, there tends to be higher prevalence of ID in areas of lower socioeconomic status and developing countries, particularly for mild cases (Drews et al., 1995; Roeleveld et al., 1997; Durkin, Hasan, & Hasan, 1998; Emerson, 2007). As previously mentioned, variability of prevalence is more pronounced for mild ID. It has been suggested that this discrepancy is likely due to environmental factors (Drews et al., 1995; Roeleveld et al., 1997; Durkin et al., 1998; Emerson, 2007).
Approximately 30% more males are diagnosed with ID than females (American Psychiatric Association, 2000; McLaren & Bryson, 1987). However, despite a higher ratio of males to females among milder cases of ID, the ratio decreases as IQ decreases (American Psychiatric Association, 2000; McLaren & Bryson, 1987). Some studies suggest that severe ID may be more prevalent among females (Katusic et al., 1996; Bradley, Thompson, & Bryson, 2002), however these studies were performed in quite specific communities, and may not necessarily be generalizable to other regions. Some of this gender bias can be accounted for by mutations on the X-chromosome. In most cases of X-linked ID (i.e. X-linked intellectual disability; XLID) or X-linked autism, more males are affected due to hemizygosity. However, in some disorders, such as Rett syndrome, this ratio is reversed because mutations in the Rett syndrome gene, MeCP2, are generally lethal in haploid genomes, or in female-restricted epilepsy and intellectual disability (EFID), in which heterozygous mutations in the gene PCDH19 cause the disease in females and in which there is “reprieve in males” with hemizygous PCDH19 mutations (Dibbens et al., 2008; Hynes et al., 2009).

It is believed that genetic factors play important role in etiology of ID, however, the contributions of environment and other factors cannot be excluded. ID is a clinically and genetically heterogeneous disorder, to date, at least, 215 X-linked ID (XLID) conditions have been reported and more than 85 genes have been linked to this disorder. Based on the mode of inheritance, ID is classified into autosomal recessive, autosomal dominant, and X-linked. During the past decade, many genes\loci causing ID have been identified, however, the cause of ID in up to 60% of the cases still remains unknown (Rauch et al., 2006). Chromosomal aberrations are identified in ~15% of all ID cases (Leonard & Wen, 2002a). Approximately, 30% more males are diagnosed with ID than females. To date, more than 80 X-linked genes are known to cause ID, however, mutations in these genes are thought to be responsible for only 10% of the ID found in males, which suggests that there must be many other unknown genes or other contributing factors which result in this gender bias (McLaren & Bryson, 1987; Ropers, 2006b).

The discovery of XLID genes began in the early 1980s, when mutations in the HPRT, PGK1 and PLP genes were identified, since then, the list of XLID genes has grown rapidly, and, to date, more than 85 genes have been linked to XLID (Ropers, 2008). During 2006-2008, in just two years, 16 new genes were implicated in XLID. A regularly updated table of all ID genes is
available at http://www.ggc.org/XLID.htm. Among known XLID genes, mutations in the \textit{FMR1} gene which result in the Fragile X syndrome (MIM#300624) are the most frequent cause of XLID which may account for 25\% of all families with XLID (Fishburn, Turner, Daniel, & Brookwell, 1983). Second to \textit{FMR1} is \textit{ARX} – mutations in which may cause syndromic or nonsyndromic XLID in >5\% of the families (Gecz, Cloosterman, & Partington, 2006). Other frequently implicated genes include \textit{CUL4B}, \textit{JARID1C} and \textit{SLC6A8}, mutations in each of these genes account for 2–3\% of the families, whereas mutations in the rest of the known XLID genes are rare (< 1\%) (Ropers, 2008).

Most of the known XLID genes have been identified by candidate gene sequencing and/or characterization of chromosomal aberrations. For instance, characterization of a deletion resulted in identification of \textit{IL1RAPL1} as an XLID gene. Subsequently, several other studies further confirmed the role of this gene in XLID (Carrie et al., 1999a; Piton et al., 2008; Nawara et al., 2008). Disruption of this has also been shown to cause autism with or without ID (Bhat et al., 2008b; Marshall et al., 2008; Piton et al., 2008).

Recently, in the largest study of chromosome X gene sequencing to date, most of the annotated X-chromosomal genes (~900) were sequenced in 208 families with ID. This study discovered nine new XLID genes including \textit{SYP}, \textit{ZNF711} and \textit{CASK} (Tarpey et al., 2009a). Hundreds of novel variants were also identified, however, the contributions of many of these variants to XLID remains unclear. Another interesting finding was the observation of loss of function mutations in ~1\% of X-chromosomal genes without any apparent associated phenotype (Tarpey et al., 2009a).

Also, in recent candidate gene sequencing studies, a nonsense and a splice site mutation in \textit{RAB39B} have been recently shown to cause ID with autism, epilepsy, and macrocephaly in unrelated families (Giannandrea et al., 2010). Another study has reported the disruption of \textit{IQSEC2} gene in first family published as NS-XLID (MRX1) in 1988. In addition to this family, disease causing mutations were also identified in three other XLID families (Shoubridge et al., 2010).

Because of the haploid status of most X chromosome genes in males, mutations are much more severe in males, at least in genes that do not escape X-inactivation, leading to a distinct gender bias in X-linked disorders. Hence, mutations in X-linked genes account for the observation of
more males affected with intellectual disability than females. Interestingly, a similar direction of gender bias has been observed in autism; the ratio of affected males to females is estimated to be 4:1 (Volkmar, Szatmari, & Sparrow, 1993). As reviewed above, the majority of ID genes identified to date map to the chromosome-X, and many forms of XLID, both syndromic and non-syndromic, have very broad spectra of phenotype and severity, and often include features that overlap significantly with autism. DSM-IV ascribes ~5% of intellectual disability individuals to known genetic conditions such as Tay-Sachs, tuberous sclerosis or fragile X syndrome, and ~30% to recognized predisposing factors affecting early embryonic development such as trisomy 21 or maternal alcohol consumption, and ~15-20% to environmental influences and other mental disorders such as autistic disorder. DSM-IV also estimates that ~75% of children with autistic disorder have ID, usually within the moderate range (IQ 35-50). This is likely to be an overestimate, and more recent estimates suggest a more modest level of 50-60% (Chakrabarti & Fombonne, 2001; Bertrand et al., 2001; Baird et al., 2000). Therefore, these factors suggest a possible overlap between ID and autism and these disorders may share some genetic etiologies. In our own study of CNVs in autism, we have identified CNVs disrupting a number of ID genes- the first time such observations have been reported for autism (Marshall et al., 2008). Details of these findings are illustrated in chapter 2 of this dissertation.
1.3 The X-Chromosome

The X chromosome is one of two sex-determining chromosomes in humans and many other mammals. The first finished sequence of the human X chromosome was published in 2005 by an international team of more than 250 genomic researchers led by the Wellcome Trust Sanger Institute. The X chromosome was reported to be ~155 Mb in length having 1098 genes, of which at least 800 genes are protein coding. Sequence annotation revealed that the X chromosome is a relatively gene poor chromosome, and has a lower GC content (39%) compared to the genome average (41%) (Ross et al., 2005).

The human X chromosome has many interesting features that are unique in the human genome, for example, females inherit one X chromosome from each parent while males inherit a single, maternal X chromosome. At most of the X chromosomal loci the gene dosage in females is equaled to males by the process of X-chromosome inactivation (XCI) during early development, and one chromosome remains inactive in somatic tissues thereafter. In the female germ line, the inactive chromosome is reactivated and undergoes meiotic recombination with the second X chromosome. Furthermore, male X chromosomes do not recombine along its entire length during meiosis but recombination is restricted to short regions at the tips of the X chromosome arms that recombine with equivalent segments on the Y chromosome. Genes inside these regions are shared between the X and Y chromosomes, and they are described as pseudoautosomal genes. Genes outside the pseudoautosomal regions of the X chromosome are strictly X-linked, and the vast majority is hemizygous in the male genome. Interestingly, the X chromosome acquired these unique features as consequence of its evolutionary process.

1.3.1 Evolution of the X-Chromosome

In 1967, Ohno proposed that the mammalian sex chromosomes evolved from a pair of autosomes following their selection into a chromosomal system for sex determination, within the last 300 million years. Ohno’s law states that the establishment of a dosage compensation mechanism had a stabilizing effect on mammalian X chromosomal genes, and the mammalian X chromosome length and gene content is extraordinarily conserved (Graves, 1996). During the process of evolution, a barrier to recombination developed between these ‘proto’ sex chromosomes, isolating the sex-determining regions and eventually spreading throughout the
two homologues. As a result of absence of recombination and accumulation of mutation events the Y chromosome was subsequently degenerated. Ancestral chromosome blocks that fused to form the X-Y pair can be identified by comparing the gene content of the X and Y in humans and distantly related mammals like marsupials and monotremes, which diverged 180 and 210 million years ago respectively, and the gene contents of orthologous regions in other vertebrates. Richardson et al reported that the human sex-linked enzyme loci such as are also sex-linked in marsupials (Richardson, Czuppon, & Sharman, 1971). A more recent study by Johnston et al also supported these finding (Johnston & Robinson, 1986).

Spencer et al studied ten human genes located on the long arm of the X chromosome (Xq) and by somatic cell analysis and in situ hybridization, they found that these genes were located on the X chromosome in marsupial species. The results of this study indicated that the long arm of the human X chromosome represents a highly conserved region that formed part of the X chromosome in a therian ancestor 120-150 million years ago, before the mammalian infraclasses diverged (Spencer, Watson, & Graves, 1991). However, many genes located on the short arm (Xp) of the human X chromosome were shown to be absent from the X chromosome in marsupials, in which they appear to be autosomal (Sinclair, Wrigley, & Marshall Graves, 1987; Watson, Spencer, Riggs, & Graves, 1991). The fusion point for this appears to be around Xp11.23 (homo sapiens)- from here to the centromere, genes are X chromosomal in eutherian and non-eutherian mammals (Wilcox et al, 1996).

Watson et al proposed two alternative hypotheses to explain the finding of human Xp genes on marsupial autosomes. First, the human Xp region was part of the X chromosome in the common therian ancestor of the eutherians and marsupials and was translocated to an autosome in the marsupial lineage. Second, the human Xp region was originally autosomal and was translocated to the X chromosome in the eutherian lineage (Watson, Spencer, Riggs, & Graves, 1991). Experimental testing of these hypotheses using comparative chromosome painting techniques revealed that the human Xp region was originally autosomal and was translocated to the X chromosome in the eutherian lineage (Glas, Marshall Graves, Toder, Ferguson-Smith, & O'Brien, 1999). Another study proposed four evolutionary strata on the human X chromosome and suggested that the human sex chromosome evolution was punctuated by at least four events in which X-Y recombination was suppressed in one stratum and the gene order was conserved.
Furthermore, it was proposed that the first event, which marked the beginnings of X-Y differentiation, occurred about 240 to 320 million years ago, shortly after divergence of the mammalian and avian lineages (Lahn & Page, 1999). The placental mammalian X is extraordinarily conserved in size, gene content and even gene-order. However, only part of the human X (Xqter-p11.2) is shared with the marsupial and monotreme X chromosomes, and the rest is autosomal. This defines an ancient X-conserved region (XCR) and an X-added region (XAR), which was added to the placental X after the marsupial–placental divergence but before the placental radiation (Graves, 1995).

Although, conventional mapping techniques provided insight into human X chromosome evolution, with the availability of whole genome sequence of human and other species, a comprehensive reconstruction of evolutionary events became possible by comparing genomes from species that diverged from mammals early in the history of vertebrates. Mammals diverged from birds and reptiles ~310 million years ago (Hedges & Kumar, 2004) and from fish ~450 million years ago (Vandepoele, De Vos, Taylor, Meyer, & Van de Peer, 2004). Interestingly, the sex chromosome system of birds evolved independently during the last 300 million years and is not homologous to those of mammals. The evolution of sex chromosomes of birds gave rise to homogametic (ZZ) male birds and heterogametic (ZW) female birds, in contrast to the mammalian system of XY males and XX females. The alignment of the human X chromosome and chicken whole genome sequences supports the autosomal origin of mammalian sex chromosomes. Orthologues of some human X chromosome genes were previously mapped to different chicken chromosomal loci including 1q13-q21 and 4p11-p14 (Schmid et al., 2000). The comparison of draft human genome sequence and chicken sequence confirmed and extended the previous model of human X chromosome evolution. However, the fourth stratum on human Xp22.3 described earlier by Lahn and Page, which resulted from a recent inversion of the Y chromosome between prosimians and simians, cannot be reconstructed by this approach. Additionally, the genome sequence comparison confirmed that X chromosome is strikingly conserved during evolution with exception of few genes in Xp11.2 and Xq28 (Kohn, Kehrer-Sawatzki, Vogel, Graves, & Hameister, 2004).

With availability of finished human chromosome X sequence, using genomic sequence alignment, Ross et al identified ~30 regions of homology that together cover most of human Xq
and are confined to a single section of approximately 20Mb at the end of chicken chromosome 4p. In contrast, most of the short arm of X chromosome including the pseudoautosomal region PAR1, aligned to a single block of chicken chromosome 1q, while the emergence of the rest of the short arm remained unclear. This analysis supported the previously described X-conserved region (XCR) and an X-added region (XAR). However, in contrast to earlier hypotheses, it was proposed that much of the proximal short arm (Xcen–p11.3) should no longer be considered part of an XCR. Furthermore, this study investigated the precise order of genes in XAR and proposed new model that the XAR was acquired by recombination between the X chromosome and a ring chromosome in which the ancestral PLCXD1, RGN and RGN2 sequences were neighbors. The recent patterns of evolution were examined by comparison of the human X chromosome with other mammalian sequences. Nine major blocks of sequence homology between human and mouse X chromosomes, and eleven between human and rat were identified. The homology blocks occupy most of the X chromosome, confirming the remarkable degree of conserved synteny of this chromosome within the eutherian mammalian lineage (Ross et al., 2005).

Another interesting aspect of the X chromosomal evolution is the special selection process that forced some X-linked genes to develop an extended functional spectrum. One such example is genes involved in development of cognitive abilities. It is well established that an excess of genes responsible for cognitive abilities are mapped on X chromosome and this phenomenon is also described as ‘the large X-chromosome effect’ (Turelli & Orr, 1995). However, Zechner et al, proposed a large X-chromosome effect for general cognitive abilities in humans and defined the evolution of enhanced cognitive abilities as a specifically human trait. A large X-chromosome effect influencing the development of a specific character like fertility or cognitive ability implies that this character is selected in the species (Zechner et al., 2001).

In conclusion, the origin and evolution of the X chromosome is very interesting for several reasons, including its unique dosage compensation mechanism by XCI, its role in evolution of chromosome Y and its large chromosome effect for development of cognitive capabilities in humans. It is also notable that the X chromosome is the most conserved chromosome in the human genome. Particularly, the Xq is strikingly conserved during evolutionary processes spanning millions of years, and even much of the gene order is conserved across diverse species.
Further investigation of X chromosome evolution will increase our understanding of how genes acquired extended functions during human evolution.

1.3.2 Autism and X-chromosomal genes:

The striking observation that males are nearly four times as likely to be affected by autism as females (Volkmar et al., 1993) may implicate the sex chromosomes having some direct or indirect etiological effect. In order to test the hypotheses that loci on the X chromosome may increase susceptibility to autism, various studies have used X-chromosome-specific approaches or genome-wide approaches to look for X-linkage to autism. Linkage analysis performed by Dr. Vincent’s research group using 23 multiplex families with autism from the UK found that linkage to autism for markers in the Xq27-Xqter region could not be excluded, for a variety of phenotypic definitions of affected status. For the most inclusive phenotypic definition, a non-parametric LOD score of 2.1 was computed (Vincent et al., 2005). Our continuing follow-up of this finding using the CANAGEN study families has also identified suggestive linkage around Xq27 (unpublished data).

A genome-wide screen by Liu et al (2001), using the AGRE multiplex families with autism, also demonstrated suggestive linkage to the region encompassing Xq25-qter, with a MLS of 2.67 at DXS1047 (Liu et al., 2001). Both studies used relatively low density of markers on the X chromosome (1 marker per 12.6cM (AGRE), and 5.6cM (UK)), and hence only a small proportion of the meiotic information has been extracted from the families genotyped. A recent study of French-Canadian males with autism also showed association for markers and haplotypes within this region, with p values as low as p=0.00001 for marker DXS8043 at Xq27.3 (Gauthier et al., 2006). These results show that further investigation, either by genetic linkage analysis or by the direct detection of mutations in coding and regulatory regions of DNA is required if this region is to be excluded as a candidate locus. Another study used the linkage and association data pointing to the distal Xq region to search for potential candidate genes for autism, and appear to have identified mutations within the ribosomal protein encoding gene (RPL10) on Xq28 (Klauck et al., 2006). Although the importance of this gene in autism needs to be confirmed, this study highlights the interest in X chromosome genes in autism.
In a study by Thomas et al, de novo deletions at Xp22.3 locus were reported in three autistic females (Thomas et al., 1999). These finding prompted a comprehensive analysis of this region which lead to the identification of NLGN4 gene, a member of Neuroligin family (Jamain et al., 2003). Neuroligins constitute a family of proteins thought to mediate cell-to-cell interactions between neurons (see HNL1; MIM 600568). Neuroligins function as ligands for the neurexin family of cell surface receptors (see NRX1; MIM 600565), and are thought to act as a trans-neuronal signal that triggers synapse formation (Scheiffele, Fan, Choi, Fetter, & Serafini, 2000). Several stop mutations have been identified and segregate with disease in three families with autism in the X-chromosomal orthologs NLGN3 and 4 (Jamain et al., 2003). A number of studies including our own (Vincent et al., 2004) have screened autism populations for mutations in NLGN3 and 4, but have failed to find any evidence of mutations in their autism patients (Ylisaukko-oja et al., 2005; Gauthier et al., 2005; Blasi et al., 2006b). One study reported four putative missense changes in NLGN4 in four families with autism (Yan et al., 2005). Another study describes a large family where a frameshift mutation in NLGN4 is present in 10 members with non-specific x-linked intellectual disability, two with autism and one with pervasive developmental disorder (Laumonnier et al., 2004). More recently, the gene encoding ribosomal protein 10 (RPL10) on Xq28 was screened in autism patients, and missense mutations identified in 2 families (Klauck et al., 2006). Mutations in the intellectual disability genes FMR1, MECP2, the X-linked creatine transporter SLC6A8, and ARX have all been linked to occasional cases of autism, as outlined in the following paragraphs. In addition, the first study to identify a gene disrupted by a translocation in a patient with autism (Ishikawa-Brush et al., 1997) identified the gastrin releasing peptide receptor (GRPR) on Xp22.2. Although there are no further reports linking this gene to autism, GRPR blockade in rats resulted in impaired social behavior, and thus it should not be excluded from a possible etiologic role. Another X-linked gene that may play a role in autism is EFHC2- a recent study has identified this gene as a quantitative trait locus for fear-recognition in Turner’s syndrome. This is relevant to autism, because impaired recognition of emotion is present in autism, and a large proportion of Turner’s syndrome patients (~30%) have autism (Weiss et al., 2007).
1.3.3 Autism and Non-syndromic XLID genes:

There are at least 34 genes which have been implicated in non-syndromic (NS) XLID, including ACSL4, AFF2/FMR2, AGTR2, AP1S2, ARHGEF6, ARX, ATRX, BRWD3, CASK, DLG3, FGD1, FTSJ1, GDI1, HUWE1, IL1RAPL1, JARID1C (KDM5C), MAGT1, MECP2, NLGN4, OPHN1, PAK3, PQBP1, PTCHD1, RPS6KA3, SHROOM4, UPF3B, ZNF41, ZNF674, ZNF711, ZNF81, SLC6A8, SYP and TSPAN7 (Kaufman, Ayub, & Vincent, 2010). NS-XLID disorders appear to have an extremely broad phenotypic range. Furthermore, given the potential mutation spectrum along with wide scope for protein regulation of these NS-XLID genes, it is anticipated that some mutations within some of these genes will lead to symptoms within the autistic spectrum.

The identification of genes for either syndromic or non-syndromic forms of XLID has indicated some surprising pathways which, when disrupted, can lead to impaired cognitive development. For instance, the fatty acid-coA ligase 4 gene, FACL4, has recently been identified as the causative factor in both syndromic (Alport syndrome) and non-syndromic XLID (Meloni et al., 2002). OPHN1, PAK3, ARHGEF6 and FGD1 genes are involved in the RhoGTPase cycle, which mediates cytoskeletal organization and cell motility, and is involved in outgrowth of axons and dendrites (Ramakers, 2002). Similarly, other XLID genes have shown that proteins involved in ERK/MAPK pathway, cell cycle regulation, transcriptional regulation, chromatin remodeling, cell adhesion are important for normal cognitive development (Kaufman et al., 2010).

Interestingly, it is also apparent that the X-chromosomal neuroligin genes, NLGN3 and NLGN4, which were recently identified as carrying mutations in families with autism (Jamain et al., 2003), may also be involved in NS-XLID. Mutations in NLGN4 were identified among members of a large pedigree with NS-XLID either with or without autism or pervasive developmental disorders (Laumonnier et al., 2004). Similarly, mutations in IL1RAPL1 and PTCHD1 may cause autism, ID or both.
1.3.4 Evidence of X-chromosomal involvement in autism from measures of skewed X-inactivation

In females, who carry two X-chromosomes, gene dosage for most X-chromosome genes is suppressed to equal that of the male by the inactivation of one of the two X chromosomes. Genes that escape X-inactivation are believed to be responsible for gender-specific traits (sexual dimorphism). X-inactivation randomly inactivates either the paternally-derived or maternally-derived X-chromosome in somatic cells. Skewed X-inactivation is a common feature among XLID disorders (Plenge, Stevenson, Lubs, Schwartz, & Willard, 2002) as well as recurrent pregnancy loss. Skewed X-inactivation would be able to explain differences in severity in affected families for X-linked intellectual disability (XLID) and autism, and can explain lack of phenotype in mothers transmitting disease alleles to offspring. In a recent study 33% of 35 autistic females tested showed clearly skewed X-inactivation, compared with 11% of unaffected females (N=42), further implicating the X chromosome’s involvement in autism (Talebizadeh, Bittel, Veatch, Kibiryeva, & Butler, 2005).

To summarize, a number of converging lines of evidence have already linked X-chromosomal genes to a proportion of cases of autism, where the gene mutation is the major contribution to the disease, and there is very good reason to believe that there will be more such genes on the X chromosome. Linkage and association studies suggest that there may also be X-chromosomal loci that are relatively common risk factors for autism. Use of high resolution SNP data to characterize CNVs may help the identification of new genes for autism on the X chromosome.

1.4 Thesis Objectives

1.4.1 Chapter 2

a) Although it is well established that genetic factors play an important role in the etiology of autism, conventional methods such as linkage analysis and association studies have shown a limited success in identification of autism genes. Our **objective** was to identify autism candidate genes by mapping the CNV data.
1.4.2 Chapter 3

a) While the CNV analysis highlighted several genes which may contribute to autism, still, the biological consequences, if any, were not clear for many loci. The objective was to study additional autism cases using sequence analysis and to perform qPCR assays to establish the clinical significance of candidate genes.

1.4.3 Chapter 4

a) By characterizing CNVs, we identified several autism candidate genes including the \textit{PTCHD1}. The objective was to execute a comprehensive genomic and functional study of \textit{PTCHD1} gene, in order to explore the involvement of this gene in autism and to understand the biological function of protein encoded by this gene.
Chapter 2. Chromosome X CNVs in Autism

The data presented in this chapter is included in “Structural variation of chromosomes in autism spectrum disorder” originally published in The American Journal of Human Genetics (Appendix 1).


Structural variation of chromosomes in autism spectrum disorder (2008).

Contributions: I analyzed all chromosome X CNV calls and filtered out the CNVs present in controls. Further, I identified all CNVs mapped at previously known ID loci and prioritized the potentially interesting loci. I designed and performed all experiments for the validation of selected CNVs and mapped the breakpoints of deletions. I also performed the segregation analysis of the CNVs. I helped in presentation of this data in the manuscript and edited the manuscript.
2.1 Introduction

The contribution of genetic factors to autism is well established, but the mode of genetic transmission is unclear. However, it is apparent that autism is a complex non-Mendelian disorder, and it is anticipated that genetic heterogeneity and oligo/polygenic inheritance are involved.

There are several lines of evidence such as gender bias, positive linkage findings, role of X-chromosomal genes in cognition and skewed inactivation which suggest the possible involvement of X-chromosomal loci in the etiology of autism. Although the X chromosome is generally a gene poor chromosome, majority of the known ID genes map to the X chromosome. To date, more than 80 X-linked have been implicated in various forms of ID (Ropers, 2008). Because of the haploid status of most sex chromosomal genes in males, mutations of chromosome X or Y are much more severe in male, which may explain in part the gender bias in autism and ID.

Cytogenetically-detectable chromosome abnormalities may be identified in up to 7.4% of ASD cases (Marshall et al., 2008). The occurrence of genomic imbalances is higher in syndromic forms of ASD (Vorstman et al., 2006). Balanced translocations and inversions account for 17% of the genomic rearrangements and the most frequent chromosomal anomaly observed is maternally-derived duplication of chromosome 15q11-q13 in 1-3% of cases (Veenstra-Vanderweele, Christian, & Cook, Jr., 2004). By using chromosomal abnormalities to identity ASD candidate genes, mutations have been identified in \textit{SHANK3} on chromosome 22q13 (Moessner et al., 2007; Durand et al., 2007), \textit{NLGN3} and \textit{NLGN4} genes on the X-chromosome (Jamain et al., 2003), and the neurexin 1 gene (\textit{NRXN1}) on chromosome 2p16 (Szatmari et al., 2007). It has been recently shown that the sub-microscopic CNVs may contribute to etiology of autism and \textit{de novo} CNVs seem to be an even more significant risk factor in sporadic compared with familial forms of ASD (Sebat et al., 2007).

Epidemiological studies have shown a gender bias in autism as well as in ID, where the number of males affected with autism is 3-4 times greater than females, and for ID the ratio is ~1.3:1, although this is believed to decrease with decreasing IQ (American Psychiatric Association 2000; McLaren and Bryson 1987) and some studies suggest that severe ID may be more
prevalent among females (Katusic et al. 1996; Bradley et al. 2002). Also, some form of ID has been reported in up to 70% of autistic children (Fombonne, 2003). This overlap between autism and ID warrants the search for common etiological factors and the sex ratios in these disorders make the X chromosome an excellent candidate. Thus, the present study hypothesized that a number of autism candidate genes are present on chromosome X, and some of these can be identified by the analysis of X chromosome CNVs and through the direct sequencing of candidate genes among probands with autism.

In the present study we took advantage of the technological advances in the field of DNA microarrays. The genome scan of over 400 hundred probands with autism was performed using the Affymetrix 500k SNP array. We attempted to identify autism candidate genes by mapping the CNVs on the X chromosome, coupled with the sequencing of relevant candidate genes that map to these CNVs, in order to identify additional patients with sequence mutations.

2.2 Methods:

2.2.1 DNA Samples

The study included 427 families with ASD. Among these 228 families were recruited at The Hospital for Sick Children, 99 families at Memorial University, 86 families at McMaster University and another 14 families were recruited at other sites. ADOS and ADI-R tools were used for assessments and all probands met the DSM-IV diagnostic criteria for autism. Within this cohort, 32 patients carried a cytogenetic chromosome rearrangement and 18 of these 32 chromosomal rearrangements had been detected by previous karyotyping. Siblings of the proband were also assessed for ASD. 236 families had one ASD child (simplex) and 189 families had more than one ASD child (multiplex). This high ratio of simplex to multiplex is likely due to bias in the ascertainment of the families. Approximately 75% of cases were screened for FRX mutations, and families with FRX mutations were excluded from the study. Most experiments were performed on genomic DNA extracted from lymphocytes (80%), and for the remainder the DNA was extracted from lymphoblastoid cell lines. Using multi-locus SNP genotype data, the population ancestry was inferred by STRUCTURE (Falush, Stephens, & Pritchard, 2003). Analysis by STRUCTURE software revealed that ~90% (386/427) of probands were of European origin, 4.5% (19/427) were of European-mixed origin, 4.5% (19/427) were of Asian
origin and 0.07% (3/427) were of African origin. Probands were clustered without considering their original geographical origin using 780 unlinked SNPs, assuming three ancestral populations. In the same clustering, 209 unrelated HapMap individuals (African, European and Asian) were used as reference.

CNVs were also assessed in 500 European control individuals from Northern Germany, who were ascertained through the PopGen project (Krawczak et al., 2006). Additionally, 1152 non-disease control individuals of European origin who were recruited from the province of Ontario were also included in the study. Details of these control individuals are published elsewhere (Zogopoulos et al., 2007).

2.2.2 Microarray and karyotyping experiments

For each sample, approximately 500,000 SNPs were genotyped using the Affymetrix GeneChip Human Mapping 500K Array. The 500K Human Mapping microarrays are comprised of two-chips, the Affymetrix NspI array and StyI array. Also, some of the samples were analyzed on Human Mapping 500K Early Access Arrays. All arrays were processed at the microarray facility at The Centre for Applied Genomics (TCAG) according to the manufacturer’s instructions. The basic steps in processing of these arrays are outlined in Figure 2-1.
Figure 2-1 Major steps involved in processing of Affymetrix 500K microarrays

Adapted from [www.affymetrix.com](http://www.affymetrix.com). The basic steps involved in the processing of Affymetrix 500K microarray are outlined. Briefly, 250 ng genomic DNA is digested by NspI or Styl restriction enzymes. Standard adapters are ligated to the fragments of DNA followed by a PCR amplification and End-labeling. These labeled fragments are then hybridized to chips and unligated fragments are subsequently washed. The chips are then scanned and fluorescent colors and intensities are used to infer the genotyping and CNVs.
2.2.3 **CNV Analysis**

CNVs, i.e. either gain or loss of genomic material, were initially inferred by comparative analysis of hybridization intensities using dChip analyzer (Li & Wong, 2001). After normalization, we used Hidden Markov Model (HMM) to infer the DNA copy number from the raw signal data. Additionally, the data was also analyzed using the CNAG software which is an improved algorithm for copy number analysis of the human genome (Nannya et al., 2005). The advantages of this algorithm are the improvement of signal-to-noise (S/N) ratios and the use of an optimized reference. To maximize the CNV discovery and for cross validation the microarray data was also analyzed with another software called GEMCA (Komura et al., 2006). The analysis of NspI and StyI chips data was performed separately and also by combining both chips together in the analysis.

2.2.4 **Validation of CNVs**

Validation of CNV calls was performed using the SYBR-Green I based real-time quantitative PCR (qPCR), using the FOXP2 locus as a control amplicon, as CNVs of this region are not seen in general population. At least two independent assays were used for CNV confirmations. For hemizygous deletions of chromosome X loci in males, standard PCR reactions were performed using the primers within the deletion region and products were analyzed on agarose gel.

2.2.5 **Identification of Candidate Loci**

In order to identify candidate autism susceptibility genes\loci, 500K SNP microarray data were used to infer CNVs on the X chromosome in 427 autism patients. To exclude copy number polymorphisms (CNPs), the presence of these CNVs was checked in 1652 controls and the Database of Genomic Variants (DGV). The CNVs present in controls or in the DGV were discarded. The rest of the CNVs were classified as autism specific events, and were marked for validation using SYBR Green-I quantitative PCR (qPCR), in the probands and their family members. CNVs that were either *de novo* or transmitted from unaffected mothers to affected sons were further followed up. Any deletions in male patients for which there was known overlap with genomic variants in controls, were also considered, under the hypothesis that the CNV at a X-linked recessive gene may be present (and harmless) in control females yet etiologically relevant in hemizygous males. Within the autism specific CNV regions the
candidate genes were selected based on the function, expression profile or physical position of a gene. An overview of the workflow in outlined in Figure 2-2.

**Figure 2-2** Workflow for mapping autism susceptibility genes using CNVs on the X chromosome is outlined.
2.3 Results

2.3.1 Chromosome X CNVs in ASD cases

We used the Affymetrix GeneChip® Human Mapping 500K SNP array (NspI and StyI Chips) to scan the genomes of 427 ASD cases to assess genomic imbalances that may be associated with the disease. The entire data set is published in Marshall et al, 2008. This thesis will only focus on X chromosomal CNVs. Using high-density SNP microarray ~10,000 SNP markers across chromosome X, with average marker density of ~15 kb, were genotyped which enabled us to call CNVs and map the boundaries of these CNVs with relatively high precision. For selection of appropriate controls, we used the SNP genotypes to categorize the ancestry of the samples, and showed that over 90% of samples had European ancestry.

A total of N=116 genomic X chromosomal CNVs were detected among the 427 ASD samples (Table 2-1). Among these, 97 CNVs were gain (duplications) and 19 were loss (deletions). These CNVs were further analyzed for overlap with polymorphic CNVs in 500 PopGen controls, 1152 Ontario controls and CNVs annotated in the DGV. It was observed that among these variants, N=54 CNVs were specific to autism cases, as they were not identified in any of the controls (Table 2-2). The autism specific CNVs were further classified as stringent autism specific if a CNV was picked by more than one of the three algorithms (dCHIP, CNAG or GEMCA), or on both NspI and StyI arrays. With this criteria, N=15 stringent autism specific CNVs were indentified (Table 2-3).

The CNVs identified range from a few kilobases (Kb) to several megabases (Mb) in size, and some of them encompass functionally important candidate genes such as *IL1RAPL1*, *TSPAN7*, *NLGN4* and *FMRI*, or genes of unknown function yet of interest due to homology with genes of known function, e.g. *PTCHD1* and *IL1RAPL2*. Two autism specific CNVs, a 104 Kb deletion at Xq21.31 and a 121 Kb deletion at Xq27.3 do not involve any known RefSeq genes. However, it is possible that these regions may contain some novel and uncharacterized genes or regulatory regions that may play a role in autism susceptibly.

Other interesting regions include the Xq28, Xq25 and Xp22.31, which have CNVs in unrelated cases. Duplications within Xq28 were found in N=4 autism probands -- one of these gains encompasses the *FMRI* and *FMRI*NB genes. This region has been previously reported to have
modest linkage to autism, and expansion/hypermethylation of the FMR1 gene is a known cause of intellectual disability and occasionally autism. Therefore, our CNV findings at this locus further reinforce the importance of detailed analysis of this CNV.

Table 2-1 Chromosome X CNVs identified in 427 ASD cases.

<table>
<thead>
<tr>
<th>Type of CNVs</th>
<th>Analysis Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total CNVs</td>
</tr>
<tr>
<td>All Chromosome X CNVs</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism Specific</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism Specific Stringent</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2-2 Autism Specific CNVs are listed. All genomic coordinates are based on hg17: Build 35.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Start</th>
<th>Stop</th>
<th>Length</th>
<th>Call</th>
<th>Method_Chip</th>
<th>Cytoband</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>34,419</td>
<td>108,567</td>
<td>74,148</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp22.33</td>
</tr>
<tr>
<td>M</td>
<td>34,419</td>
<td>159,978</td>
<td>125,559</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp22.33</td>
</tr>
<tr>
<td>M</td>
<td>34,419</td>
<td>375,007</td>
<td>340,588</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp22.33</td>
</tr>
<tr>
<td>M</td>
<td>34,419</td>
<td>1,016,559</td>
<td>982,140</td>
<td>gain</td>
<td>dChip_Sty</td>
<td>Xp22.33</td>
</tr>
<tr>
<td>F</td>
<td>34,419</td>
<td>5,859,730</td>
<td>5,825,311</td>
<td>loss</td>
<td>CNAG_Sty, dChip_Sty, CNAG_Nsp, dChip_Nsp</td>
<td>Xp22.33, Xp22.32, Xp22.31</td>
</tr>
<tr>
<td>M</td>
<td>11,496,818</td>
<td>13,241,465</td>
<td>1,744,647</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp22.2</td>
</tr>
<tr>
<td>M</td>
<td>12,512,339</td>
<td>12,848,932</td>
<td>336,593</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp22.2</td>
</tr>
<tr>
<td>M</td>
<td>16,784,446</td>
<td>16,973,300</td>
<td>188,854</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp22.2</td>
</tr>
<tr>
<td>M</td>
<td>19,943,694</td>
<td>21,501,963</td>
<td>1,558,269</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp22.12</td>
</tr>
<tr>
<td>M</td>
<td>20,399,305</td>
<td>20,579,609</td>
<td>180,304</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp22.12</td>
</tr>
<tr>
<td>M</td>
<td>22,793,958</td>
<td>24,734,469</td>
<td>1,940,511</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp22.11, Xp21.3</td>
</tr>
<tr>
<td>M</td>
<td>22,962,800</td>
<td>23,119,000</td>
<td>156,200</td>
<td>loss</td>
<td>dChip_Nsp_EA, dChip_Sty_EA</td>
<td>Xp22.11</td>
</tr>
<tr>
<td>M</td>
<td>23,397,818</td>
<td>23,847,382</td>
<td>449,564</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp22.11</td>
</tr>
<tr>
<td>M</td>
<td>25,516,263</td>
<td>25,620,400</td>
<td>104,137</td>
<td>loss</td>
<td>CNAG_Nsp, dChip_Nsp, CNAG_Sty, dChip_Sty</td>
<td>Xp21.3</td>
</tr>
<tr>
<td>M</td>
<td>32,917,766</td>
<td>33,917,680</td>
<td>999,914</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp21.1</td>
</tr>
<tr>
<td>M</td>
<td>36,222,896</td>
<td>36,319,718</td>
<td>96,822</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp21.1</td>
</tr>
<tr>
<td>M</td>
<td>38,250,331</td>
<td>38,371,333</td>
<td>121,002</td>
<td>gain</td>
<td>dChip_Nsp, CNAG_Nsp</td>
<td>Xp11.4</td>
</tr>
<tr>
<td>M</td>
<td>38,490,148</td>
<td>41,130,651</td>
<td>2,640,503</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp11.4</td>
</tr>
<tr>
<td>M</td>
<td>39,837,594</td>
<td>40,393,087</td>
<td>555,493</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp11.4</td>
</tr>
<tr>
<td>M</td>
<td>42,060,666</td>
<td>42,634,113</td>
<td>573,447</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp11.4, Xp11.3</td>
</tr>
<tr>
<td>M</td>
<td>44,395,900</td>
<td>45,060,800</td>
<td>664,900</td>
<td>gain</td>
<td>dChip_Sty, CNAG_Nsp, CNAG_Sty, dChip_Nsp</td>
<td>Xp11.3</td>
</tr>
<tr>
<td>M</td>
<td>44,779,891</td>
<td>51,230,762</td>
<td>6,450,871</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xp11.3, Xp11.23, Xp11.22</td>
</tr>
<tr>
<td>Sex</td>
<td>Chr Start</td>
<td>Chr End</td>
<td>Length</td>
<td>Status</td>
<td>Gene Gain/Loss</td>
<td>Location</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>F</td>
<td>48,073,600</td>
<td>52,716,966</td>
<td>4,643,366</td>
<td>gain</td>
<td>dChip_Sty,CNAG_Sty,GEMC_A_Nsp+Sty,CNAG_Nsp,dChip_Nsp</td>
<td>Xp11.23, Xp11.22</td>
</tr>
<tr>
<td>M</td>
<td>51,459,178</td>
<td>53,166,194</td>
<td>1,707,016</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xp11.22</td>
</tr>
<tr>
<td>F</td>
<td>65,488,795</td>
<td>65,801,228</td>
<td>312,433</td>
<td>loss</td>
<td>CNAG_Nsp</td>
<td>Xq12</td>
</tr>
<tr>
<td>M</td>
<td>65,591,616</td>
<td>65,696,889</td>
<td>105,273</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq12</td>
</tr>
<tr>
<td>M</td>
<td>78,334,183</td>
<td>79,401,607</td>
<td>1,067,424</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq11.23</td>
</tr>
<tr>
<td>M</td>
<td>79,200,454</td>
<td>83,135,188</td>
<td>3,934,734</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xq21.1</td>
</tr>
<tr>
<td>M</td>
<td>80,371,599</td>
<td>80,887,402</td>
<td>515,803</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xq21.1</td>
</tr>
<tr>
<td>M</td>
<td>81,444,763</td>
<td>82,211,554</td>
<td>766,791</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq21.1</td>
</tr>
<tr>
<td>F</td>
<td>83,866,300</td>
<td>92,175,100</td>
<td>8,308,800</td>
<td>loss</td>
<td>dChip_Nsp_EA,dChip_Sty_EA</td>
<td>Xq21.1, Xq21.2, Xq21.31, Xq21.32</td>
</tr>
<tr>
<td>M</td>
<td>85,144,362</td>
<td>86,633,742</td>
<td>1,489,380</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq21.2, Xq21.31</td>
</tr>
<tr>
<td>M</td>
<td>87,452,050</td>
<td>87,595,200</td>
<td>143,150</td>
<td>gain</td>
<td>CNAG_Sty,CNAG_Nsp,dChip_Nsp</td>
<td>Xq21.31</td>
</tr>
<tr>
<td>M</td>
<td>87,941,538</td>
<td>87,986,051</td>
<td>44,513</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq21.31</td>
</tr>
<tr>
<td>M</td>
<td>90,370,143</td>
<td>92,382,892</td>
<td>2,012,749</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xq21.31, Xq21.32</td>
</tr>
<tr>
<td>M</td>
<td>90,965,080</td>
<td>91,396,910</td>
<td>431,830</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq21.31</td>
</tr>
<tr>
<td>M</td>
<td>95,626,658</td>
<td>97,930,570</td>
<td>2,303,912</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xq21.33</td>
</tr>
<tr>
<td>M</td>
<td>104,153,000</td>
<td>104,638,000</td>
<td>485,000</td>
<td>gain</td>
<td>dChip_Nsp_EA,dChip_Sty_EA</td>
<td>Xq21.33</td>
</tr>
<tr>
<td>M</td>
<td>109,904,707</td>
<td>109,994,757</td>
<td>90,050</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq22.3</td>
</tr>
<tr>
<td>M</td>
<td>112,066,070</td>
<td>112,473,368</td>
<td>407,298</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq23</td>
</tr>
<tr>
<td>F</td>
<td>112,325,000</td>
<td>113,213,000</td>
<td>888,000</td>
<td>loss</td>
<td>dChip_Sty_EA</td>
<td>Xq23</td>
</tr>
<tr>
<td>M</td>
<td>114,042,922</td>
<td>114,215,435</td>
<td>172,513</td>
<td>gain</td>
<td>dChip_Nsp,CNAG_Sty,CNAG_Nsp,dChip_Sty</td>
<td>Xq23</td>
</tr>
<tr>
<td>M</td>
<td>125,043,881</td>
<td>125,731,339</td>
<td>687,458</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq25</td>
</tr>
<tr>
<td>M</td>
<td>130,406,000</td>
<td>130,695,499</td>
<td>289,499</td>
<td>gain</td>
<td>dChip_Nsp,CNAG_Nsp,CNAAG_Sty</td>
<td>Xq26.2</td>
</tr>
<tr>
<td>M</td>
<td>130,637,670</td>
<td>133,185,428</td>
<td>2,547,758</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xq26.2</td>
</tr>
<tr>
<td>F</td>
<td>140,600,370</td>
<td>140,907,495</td>
<td>307,125</td>
<td>gain</td>
<td>GEMCA_Nsp+Sty,CNAG_Ns</td>
<td>Xq27.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------------------</td>
<td>------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>M</td>
<td>142,561,000</td>
<td>142,682,000</td>
<td>121,000</td>
<td>loss</td>
<td>CNAG_Sty,dChip_Nsp,CNA G_Nsp</td>
<td>Xq27.3</td>
</tr>
<tr>
<td>M</td>
<td>143,059,574</td>
<td>143,399,300</td>
<td>339,726</td>
<td>gain</td>
<td>dChip_Nsp,CNA G_Nsp</td>
<td>Xq27.3</td>
</tr>
<tr>
<td>M</td>
<td>145,322,000</td>
<td>145,431,000</td>
<td>109,000</td>
<td>gain</td>
<td>dChip_Sty</td>
<td>Xq27.3</td>
</tr>
<tr>
<td>M</td>
<td>146,494,249</td>
<td>147,235,753</td>
<td>741,504</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq27.3, Xq28</td>
</tr>
<tr>
<td>F</td>
<td>147,697,891</td>
<td>147,754,812</td>
<td>56,921</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq28</td>
</tr>
<tr>
<td>M</td>
<td>147,974,000</td>
<td>148,479,449</td>
<td>505,449</td>
<td>gain</td>
<td>CNAG_Sty,dChip_Nsp,dChip_Sty</td>
<td>Xq28</td>
</tr>
<tr>
<td>M</td>
<td>148,391,695</td>
<td>148,440,293</td>
<td>48,598</td>
<td>gain</td>
<td>CNAG_Sty</td>
<td>Xq28</td>
</tr>
<tr>
<td>M</td>
<td>152,846,293</td>
<td>154,411,193</td>
<td>1,564,900</td>
<td>gain</td>
<td>CNAG_Nsp</td>
<td>Xq28</td>
</tr>
</tbody>
</table>
**Table 2-3 Autism Specific Stringent CNVs.**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Start</th>
<th>Stop</th>
<th>Length</th>
<th>Call</th>
<th>Method_Chip</th>
<th>Cytoband</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>34,419</td>
<td>5,859,730</td>
<td>5,825,311</td>
<td>loss</td>
<td>CNAG_Sty,dChip_Sty,CNAG_Nsp,dChip_Nsp</td>
<td>Xp22.33, Xp22.32, Xp22.31</td>
</tr>
<tr>
<td>M</td>
<td>22,962,800</td>
<td>23,119,000</td>
<td>156,200</td>
<td>loss</td>
<td>dChip_Nsp_EA,dChip_Sty_EA</td>
<td>Xp22.11</td>
</tr>
<tr>
<td>M</td>
<td>25,516,263</td>
<td>25,620,400</td>
<td>104,137</td>
<td>loss</td>
<td>CNAG_Nsp,dChip_Nsp,CNAG_Sty,dChip_Sty</td>
<td>Xp21.3</td>
</tr>
<tr>
<td>M</td>
<td>38,250,331</td>
<td>38,371,333</td>
<td>121,002</td>
<td>gain</td>
<td>dChip_Nsp,CNAG_Nsp</td>
<td>Xp11.4</td>
</tr>
<tr>
<td>M</td>
<td>44,395,900</td>
<td>45,060,800</td>
<td>664,900</td>
<td>gain</td>
<td>dChip_Sty,CNAG_Nsp,CNAG_Sty,dChip_Nsp</td>
<td>Xp11.3</td>
</tr>
<tr>
<td>F</td>
<td>48,073,600</td>
<td>52,716,966</td>
<td>4,643,366</td>
<td>gain</td>
<td>dChip_Sty,CNAG_Sty,GE MCA_Nsp+Sty,CNAG_Nsp,dChip_Nsp</td>
<td>Xp11.23, Xp11.22</td>
</tr>
<tr>
<td>F</td>
<td>83,866,300</td>
<td>92,175,100</td>
<td>8,308,800</td>
<td>loss</td>
<td>dChip_Nsp_EA,dChip_Sty_EA</td>
<td>Qx21.1, Qx21.2, Qx21.31, Qx21.32</td>
</tr>
<tr>
<td>M</td>
<td>87,452,050</td>
<td>87,595,200</td>
<td>143,150</td>
<td>gain</td>
<td>CNAG_Sty,CNAG_Nsp,dChip_Nsp</td>
<td>Qx21.31</td>
</tr>
<tr>
<td>M</td>
<td>104,153,000</td>
<td>104,638,000</td>
<td>485,000</td>
<td>gain</td>
<td>dChip_Nsp_EA,dChip_Sty_EA</td>
<td>Qx22.3</td>
</tr>
<tr>
<td>M</td>
<td>114,042,922</td>
<td>114,215,435</td>
<td>172,513</td>
<td>gain</td>
<td>dChip_Nsp,CNAG_Sty,CNAG_Nsp,dChip_Sty</td>
<td>Qx23</td>
</tr>
<tr>
<td>M</td>
<td>130,406,000</td>
<td>130,695,499</td>
<td>289,499</td>
<td>gain</td>
<td>dChip_Nsp,CNAG_Nsp,CNAG_Sty</td>
<td>Qx26.2</td>
</tr>
<tr>
<td>F</td>
<td>140,600,370</td>
<td>140,907,495</td>
<td>307,125</td>
<td>gain</td>
<td>GEMCA_Nsp+Sty,CNAG_Nsp</td>
<td>Qx27.2</td>
</tr>
<tr>
<td>M</td>
<td>142,561,000</td>
<td>142,682,000</td>
<td>121,000</td>
<td>loss</td>
<td>CNAG_Sty,dChip_Nsp,CNAG_Nsp</td>
<td>Qx27.3</td>
</tr>
<tr>
<td>M</td>
<td>143,059,574</td>
<td>143,399,300</td>
<td>339,726</td>
<td>gain</td>
<td>dChip_Nsp,CNAG_Nsp</td>
<td>Qx27.3</td>
</tr>
<tr>
<td>M</td>
<td>147,974,000</td>
<td>148,479,449</td>
<td>505,449</td>
<td>gain</td>
<td>CNAG_Sty,dChip_Nsp,dChip_Sty</td>
<td>Qx28</td>
</tr>
</tbody>
</table>
2.3.2 Validation of Interesting Loci and Identification of autism candidate genes

N=11 CNV regions of interest spanning positional and/or functional candidate genes were selected for follow up (Table 2-4). Among these, six CNVs were gains and five CNVs were losses. We attempted to validate these CNVs by qPCR, SNP calls and standard PCR (for deletions in male probands). 10 out 11 CNVs were successfully confirmed; however, a 741kb gain at Xp27.3-q28 (at FMR1) was not validated by qPCR. This CNV was possibly a false call. We discovered a 167 Kb (chrX:23,114,179-23,281,723; NCBI Build 36) deletion spanning the exon 1 and upstream region of patched domain containing 1 gene (PTCHD1, NM_173495.2) at Xp22.11 (Figure 2-3). The deletion also spans a previously uncharacterized small non-coding RNAs (nc-RNA), DA355362 (PTCHD1AS2). Initially, the deletion was validated by PCR amplification of primers spanning the exon 1 sequence. PCR reaction failed to amplify in the proband as well as in the affected dizygotic twin brother. To check the presence of this deletion in other family members, qPCR assays were performed. Segregation analysis revealed that the CNV was inherited from a carrier mother, leading to null PTCHD1 in the proband and his dizygotic twin affected brother, an unaffected sister was a carrier (Figure 2-4). In silico analysis suggests that PTCHD1 is an 888 amino acid long transmembrane protein containing a patched-related domain with twelve transmembrane helices, highly related to the Hedgehog (Hh) receptors PATCHED1 (PTCH1) and PTCH2 as well as to Niemann-Pick Type C1 protein (NPC1). Hh is one of the key signaling pathways involved in the formation of the neural tube and brain, specifically the differentiation of motor neurons ventrally and commissural interneurons dorsally (11, 12). Mutations in Sonic Hedgehog, SHH (MIM 600725), have been reported in patients with developmental abnormalities, delay in speech acquisition and learning disabilities (13). Niemann-Pick disease type C1 is a disorder of cholesterol transport and esterification, and involves neurological and intellectual deficits (MIM 257220). This led us to investigate a possible role for PTCHD1 as a candidate gene for ASD and ID.

We therefore decided to explore a possible role for the PTCHD1 gene in the etiology of ASD and ID; these studies and results are reported in Chapter 4 of this thesis.

In another autism patient, we identified an 82 Kb deletion within intron 5 of the IL1RAPL1 gene which encodes the IL1RAPL1 protein, a member of the interleukin 1 receptor family (Figure 2-
and known XLID gene (MIM #300143). The PCR primers from the deletion region failed to amplify in the proband and an unaffected male sibling. The qPCR results showed that the deletion was maternally inherited to male proband and an affected female sibling, as well as the unaffected male sibling. The *IL1RAPL1* gene has been previously implicated in NS-ID (Bahi et al., 2003; Tabolacci et al., 2006). As reviewed above, there is a significant overlap between ID and autism which makes it a good candidate gene for autism. Therefore, we explored this gene in our autism cases, and the data is presented in Chapter 3 of this dissertation.

We also identified CNVs spanning two other members of the interleukin 1 receptor family, namely the *IL1RAPL2* and *IL13RA2*. A 485 Kb gain (chrX:104,153,000-104,638,000; NCBI Build 34) spanned the exons 3-6 of *IL1RAPL2* (Figure 2-6). We confirmed this CNV by qPCR and the segregation analysis revealed that it was maternally inherited to the proband. We also validated a 172 Kb duplication (chrX:114,042,922-114,215,435; NCBI Build 35) spanning the entire *IL13RA2* gene and at least 7 exons of *LRCH2* (Figure 2-7). The segregation analysis confirmed the maternal inheritance of this CNV.

Another interesting CNV, a 121 Kb duplication spanned the exons 2-7 of tetraspanin 7 (*TSPAN7*; also known as *TM4SF2*) gene (chrX:38,250,331-38,371,333; NCBI Build 35). The distal breakpoint of this duplication mapped between the exon 1 and 2. We confirmed this CNV by qPCR assays and segregation was found to be maternal. In addition to these CNV findings, mutations in this gene have been previously reported to cause NS-ID (MRX58; MIM #300096) (Abidi et al., 2002) which makes it a good candidate gene for ASD. We studied the possible effect of this CNV on gene expression and performed the mutation screening of this gene in ASD cases. The data is presented in the Chapter 3 of this dissertation.

We also detected a 505 Kb (chrX:147,974,000-148,479,449; NCBI Build 35) gain at Xq28 spanning the entire iduronate 2-sulfatase (*IDS*) gene and several other genes (Figure 2-9). The qPCR assays confirmed this CNV and it was found to be maternally transmitted to the male proband. The mutations in *IDS* cause mucopolysaccharidosis type 2 (MPS2) (MIM#309900) also known as Hunter syndrome. Patients with Hunter syndrome present with skeletal defects, cardiac abnormalities, hyperactivity, developmental delay and ID.
Interestingly, we also discovered two large genomic imbalances, more than 4 Mb in size. A 5.8 Mb (chrX:34,419,5,859,730: NCBI Build 35) deletion mapped to Xp22.31-p22.33 (Figure 2-10). This deletion spanned the \textit{NLGN4} and several other genes. The qPCR assay confirmed this CNV and familial segregation revealed that it was a \textit{de novo} loss in a female proband. The proximal break point of this deletion mapped between the exon 2 and 3 of \textit{NLGN4}. Mutations in \textit{NLGN4} gene have been reported to cause ID and/or ASD (Jamain et al., 2003), also Laumonnier et al, 2004).

A second large CNV, a 4.6 Mb (chrX:48,073,600-52,716,966; NCBI Build 35) gain mapped to Xp11.22-p11.23 (Figure 2-11). The CNV was validated by qPCR and was found to be a \textit{de novo} imbalance in a female proband. The deleted region spanned more than 50 annotated genes.

We also analyzed two autism specific CNVs which do not encompass any known gene. The first CNV, a 104 Kb deletion, mapped to Xp21.3 No annotated gene was directly disrupted by this CNV. The PCR amplification of sequence within this CNV region failed in the male proband. Subsequent, qPCR assays showed that deletion was inherited from mother to the affected son. The second CNV was a 121 Kb loss at Xq27.3. This deletion was also maternally inherited to a male proband and did not directly disrupt any known gene. Both these CNVs were not seen in more than 1600 controls included in this study.

In conclusion, the present study has identified several candidate genes for autism. Our findings underscore the importance of Chromosome X genes in the etiology of autism. More importantly, we have demonstrated for the first time that several ID genes may also cause autism with or without ID. These finding also confirm that in some cases autism and ID may share common etiologies.
Table 2-4 CNV regions validated with qPCR

<table>
<thead>
<tr>
<th>Length</th>
<th>Call</th>
<th>Autism specific</th>
<th>Cytoband</th>
<th>Gene</th>
<th>Inheritance</th>
</tr>
</thead>
<tbody>
<tr>
<td>167 Kb</td>
<td>loss</td>
<td>YES</td>
<td>Xp22.11</td>
<td>PTCHD1</td>
<td>Maternal</td>
</tr>
<tr>
<td>82 Kb</td>
<td>loss</td>
<td>NO</td>
<td>Xp21.2</td>
<td>IL1RAFL1</td>
<td>Maternal</td>
</tr>
<tr>
<td>485 Kb</td>
<td>gain</td>
<td>YES</td>
<td>Xq22.3</td>
<td>IL1RAFL2</td>
<td>Maternal</td>
</tr>
<tr>
<td>104 Kb</td>
<td>loss</td>
<td>YES</td>
<td>Xp21.3</td>
<td>No Gene</td>
<td>Maternal</td>
</tr>
<tr>
<td>121 Kb</td>
<td>loss</td>
<td>YES</td>
<td>Xq27.3</td>
<td>No Gene</td>
<td>Maternal</td>
</tr>
<tr>
<td>172 Kb</td>
<td>gain</td>
<td>YES</td>
<td>Xq23</td>
<td>IL13RA2</td>
<td>Maternal</td>
</tr>
<tr>
<td>121 Kb</td>
<td>gain</td>
<td>YES</td>
<td>Xp11.4</td>
<td>TSPAN7</td>
<td>Maternal</td>
</tr>
<tr>
<td>505, Kb</td>
<td>gain</td>
<td>YES</td>
<td>Xq28</td>
<td>IDS</td>
<td>Maternal</td>
</tr>
<tr>
<td>5.8 Mb</td>
<td>Loss</td>
<td>YES</td>
<td>Xp22.33-p22.31</td>
<td>NLGN4</td>
<td>De Novo</td>
</tr>
<tr>
<td>4.6 Mb</td>
<td>Gain</td>
<td>Yes</td>
<td>Xp11.23-p11.22</td>
<td>&gt;50 genes</td>
<td>De Novo</td>
</tr>
</tbody>
</table>
Figure 2-3 Genomic region showing a 167 Kb deletion (blue line) at Xp22.11 which involves Exon 1 of the \textit{PTCHD1} gene.

Figure 2-4 Pedigree showing the segregation of deletion at the \textit{PTCHD1} locus
Figure 2-5 Genomic region showing an 82 Kb deletion (blue line) at Xp21.3 in intron 5 of *IL1RAPL1* gene.
Figure 2-6 Genomic region showing a 485 Kb duplication (red line) at Xq22.3 which involves the *IL1RAPL2* gene.

Figure 2-7 Genomic region showing a 172 Kb duplication (red line) at Xq23 which involves the *IL13RA2* and *LRCH2* genes.
Figure 2-8 Genomic region showing a 121 Kb duplication (red line) at Xp11.4 which involves the TSPAN7 gene.
Figure 2-9 Genomic region showing a 505 Kb duplication (red line) at IDS locus (Xq28)

Figure 2-10 Genomic region showing a 5.8 Mb deletion (blue line) at X22.33-p22.31 which involves the NLGN4 gene.
Figure 2-11 Genomic region showing a 4.6 Mb duplication (red line) at Xp11.23-p11.22 which involves more than 50 RefSeq genes.
2.4 Discussion

In present study, we characterized the CNVs on the X chromosome to identify autism susceptibility genes. The X chromosome is of particular interest for mapping genes involved in the etiology of autism and ID. Using high density SNP arrays, we identified 116 chromosome X CNVs in 427 unrelated probands with autism. In contrast to genome-wide CNVs, where the number of deletions and duplications were almost equal, on the chromosome X, number of deletions was significantly lower than that of duplications (97 duplications and 19 deletions). This may be due to the fact that males are hemizygous for chromosome X and deletions might not be tolerated. After filtering out the CNVs present in controls, N=54 CNV were found to be specific to autism. Among autism specific CNVs, N=15 were inferred by more than algorithm.

We attempted to identify autism candidate genes by characterizing the disease associated CNV regions. One of our compelling findings was the identification of several CNVs directly involving previously known syndromic or non-syndromic ID loci. For example, we identified an 85 kb intergenic deletion of \textit{IL1RAPL1}, a member of interleukin 1 receptor family, previously implicated in non-syndromic ID (Tabolacci et al., 2006). We also identified 485 kb duplication in a closely related gene, \textit{IL1RAPL2}. Another, 175 kb duplication was identified in \textit{IL13RA2}, another member of interleukin family. Based on these findings, we proposed the possible involvement of these genes in autism (Marshall et al., 2008). Later studies further supported our findings by reporting \textit{IL1RAPL1} mutations in autism. A study by Bhat et al reported a pericentromeric inversion resulting in disruption of \textit{IL1RAPL1} gene in a patient with Autism and ID (Bhat et al., 2008b). Piton et al sequenced the coding regions of \textit{IL1RAPL1} and \textit{IL1RAPL2} in autism patients and identified a frameshift mutation in \textit{IL1RAPL1} in an autism patient. However, they failed to identify any coding mutations in \textit{IL1RAPL2} (Piton et al., 2008).

We also identified a 121 Kb duplication spanning 5 exons of \textit{TSPAN7} gene. The protein encoded by this gene is a cell surface glycoprotein which may control neurite outgrowth. This gene has been previously implicated in non-syndromic XLID. First, Abidi et al reported a 2 bp deletion in this gene in an ID patient (Abidi et al., 2002). Secondly, mutations in this gene was reported in 4 male patients with non-specific XLID (De, Frints, Borghgraef, & Fryns, 2002). Thus, \textit{TSPAN7} is another gene originally implicated in XLID, and in our study we found a disease associated CNV.
also involving this gene. Based our findings, we proposed TSPAN7 as an autism candidate gene (Marshall et al., 2008). Recently, a large scale sequencing study by Piton et al identified a missense mutation in two autism patients in this gene, further supporting our findings (Piton et al., 2010). The authors also sequenced 190 ethnically matched controls and could not identify this variant in any of the controls. Interestingly, an identical missense variant has also been reported by another group (Maranduba et al., 2004). In our study, we also sought to further explore the contribution of this gene in etiology of autism. The details of experiments and results are discussed in Chapter 3 of this thesis.

Another interesting CNV spanned exons 1 and 2 of the NLGN4 gene. This CNV was 5.8 Mb in size and the proximal breakpoint of the CNV mapped between the exon 2 and 3 of NLGN4. The CNV spanned more than 50 annotated genes. Although, NLGN4 is an obvious candidate, the loss of other genes within this region may also contribute to the phenotype. Neuroligins are cell adhesion molecules and play a crucial role in the synaptogenesis. NLGN4 was initially indentified by characterizing a deletion breakpoint and further analysis discovered causative mutations in this gene in ASD cases (Jamain et al., 2003). A later study identified a 2 bp deletion in NLGN4 segregating with phenotype in a large French family. Interestingly, the mutation was reported to cause ASD or ID or both in the affected individuals (Laumonnier et al., 2004). In our own study, we were unable to identify any coding mutations among 196 probands with autism (Vincent et al., 2004). Another, follow up study failed to find any NLGN4 mutation in French-Canadian male autism patients (Gauthier et al., 2005). These findings may suggest that NLGN4 mutations are very rare, and may account for only a minor fraction of autism cases.

In addition to CNVs involving non-syndromic ID, we also uncovered genomic structural variants spanning genes involved in syndromic forms of ID. For example, we detected a 505 kb duplication spanning entire IDS gene. The IDS mutations are known to cause Hunter syndrome (MIM#309900), also known MPS2. The Hunter syndrome involves multiple systems and patients present with skeletal abnormalities, cardiac problems, hyperactivity, developmental delay and ID (Beck, 2011).

Importantly, using our CNV data we also identified novel autism candidate genes such as PTCHD1. We identified a 167 Kb deletion spanning the exon 1 and upstream regulatory regions
of *PTCHD1* gene. Additional analysis confirmed the maternal inheritance of this deletion to two dizygotic twin brothers (Marshall et al., 2008). This was the first CNV mutation of this gene. We also found the disruption of this can cause ID as well as ASD phenotype (Noor et al., 2010). The details of our analysis of *PTCHD1* are included in the Chapter 4 of this dissertation. Recently, a study by Filges et al has identified a submicroscopic deletion of entire *PTCHD1* gene in two boys with ID (Filges et al., 2011). These findings further strengthen our claim that disruption of *PTCHD1* may cause autism or ID or both. Thus, *PTCHD1* has emerged as another gene which may cause autism and/or ID.

We have demonstrated here that a number of XLID genes may also play a role in etiology of autism. Although, the overlap between autism and ID is very well established, it is unclear to what degree they share common genetic etiologies.
Chapter 3. Analysis of X-Linked Autism Candidate Genes *TSPAN7* & *IL1RAPL1*

Results of TSPAN7 CNV analysis and mutation screening are originally published in Psychiatric Genetic (Appendix 2)


Copy number variation analysis and sequencing of the X-linked mental retardation gene TSPAN7/TM4SF2 in patients with autism spectrum disorder.


**Contributions:** For this manuscript, I performed the CNV validations, expression analysis and mutation screening. I also prepared the initial draft of manuscript and performed subsequent revisions. Gianakopoulos PJ helped with sequencing and preparation of manuscript.
3.1 Introduction

CNV data analysis has allowed us to identify several autism candidate genes on the X chromosome (discussed in Chapter 2). A substantial number of CNVs present in the human genome are likely to affect gene expression if a gene(s) and/or regulatory region(s) is duplicated, deleted or interrupted. This in turn, could contribute to human phenotypic differences, and may cause disease or even embryonic lethality. However, many CNVs may result in loss of intronic regions or duplication of a gene but have little or no effect on human phenotypic variation or disease. Furthermore, CNV studies cannot determine if an intronic deletion or partial gene duplication results in loss of normal expression of a gene. Therefore, we undertook further characterization of some of the candidate loci to determine the effect of CNVs on gene expression as well as attempting to sequence coding regions of these candidate genes to identify additional cases with sequence mutations of these genes.

In this study, by performing a genome-wide scan of 427 unrelated probands with autism using the 500K SNP microarray (Affymetrix), we reported several autism candidate genes, including TSPAN7 and ILIRAPLI (Marshall et al., 2008). We selected these two genes to further explore the possible contribution of these genes in the etiology of autism.

TSPAN7 maps to Xp11.4 and spans 127.4 Kb genomic region. It has eight exons and encodes a 249 AA protein which is a member of the transmembrane 4 superfamily, also known as the tetraspanin family. Members of this protein family are cell-surface proteins that are characterized by the presence of four hydrophobic domains. The TSPAN7 protein is a cell surface glycoprotein that may have a role in the control of neurite outgrowth, and is also known to complex with integrins (Berditchevski, 2001).

Previously, by characterizing the X chromosome breakpoint of an X;2 balanced translocation, TSPAN7 (also known as TM4SF2) was found to be disrupted in a female patient with ID and some autistic features (Zemni et al., 2000). Zemni et al also sequenced the coding regions of TSPAN7 gene in additional families with ID and identified mutations in two families. A G218X mutation was identified in one family which resulted in premature truncation of protein, thus removing the fourth transmembrane segment and the carboxy-terminal domain. The second mutation, P172H, resulted in substitution of a non-conserved amino acid. Additional studies have
also reported several mutations in \textit{TSPAN7} in NS-XLID patients. For example, 2 bp deletion (564delGT) in \textit{TSPAN7} was reported to segregate with ID in a family with NS-XLID (Abidi et al., 2002).

Based on these observations and the identification of 121 Kb duplication involving exonic sequences of this gene in our own study, we hypothesized that mutations in \textit{TSPAN7} might also be present in a number of autistic patients. To test our hypothesis we sequenced the entire coding region and splice sites of this gene in 250 unrelated probands with autism. We also attempted to check the effect of this duplication on the expression of \textit{TSPAN7} mRNA.

The second autism candidate gene we have tested here is \textit{IL1RAPL1}, which spans 1,368 Kb of genomic DNA, contains 10 coding exons and encodes a protein of 696 AA. The \textit{IL1RAPL1} gene has been previously implicated in XLID (Carrie et al., 1999a; Tabolacci et al., 2006). In an extended pedigree with NS-XLID a W487X mutation was reported to cause premature truncation which resulted in loss 210 amino acids of the cytoplasmic domain (Tabolacci et al., 2006). In another patient with ID and autism, a pericentromeric inversion of the chromosome X was reported to potentially disrupt the expression of \textit{IL1RAPL1} gene (Bhat et al., 2008a). This study reported some autistic features associated with disruption of \textit{IL1RAPL1}. At roughly the same time, we also reported an intronic deletion in \textit{IL1RAPL1} in a patient and affected sibling with autism (Marshall et al., 2008). To further investigate the involvement of \textit{IL1RAPL1} gene in autism, we undertook the sequencing of all coding exons of this gene in 250 unrelated probands with autism. We also attempted to test if the intronic deletion which we identified in the proband with autism results in loss of expression of this gene.
3.2 Methods

3.2.1 Samples

250 unrelated probands diagnosed with autism were checked for the coding mutations of \textit{TSPAN7} and \textit{IL1RAPL1}. These samples were from the same cohort used in the CNV study. These families were recruited at three sites, namely, The Hospital for Sick Children, Memorial University and McMaster University. ADOS and ADI-R Assessments were performed and all probands met diagnostic criteria for ASD. Details of these samples are included in Chapter 2 of this thesis.

3.2.2 PCR and Sequencing

PCR primers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/). Primer sequences are provided in Table 3-1 for \textit{TSPAN7} and Table 3-2 for \textit{IL1RAPL1}. These primers were designed to amplify all coding exons and exon-intron boundaries, extending >50 bp into intronic sequences both 5' and 3' to each exon.

PCR amplifications were performed using Qiagen Hot StarTaq Master Mix Kit (Cat. # 203446) in a final volume of 12 μl. PCR cycling conditions consisted of an initial denaturation step at 95°C for 15 min, followed by 30 cycles of 95°C for 45 sec, annealing at 57°C for 45 sec and elongation at 72°C for 45 sec followed by final extension for 10 min at 72°C.

DNA sequencing was performed in a 10 μL final volume using Big Dye Terminator Ready Reaction mix (Applied Biosystems), and data was generated using the ABI 3730 Genetic Analyzer. Sequence chromatograms were checked for any sequence changes using the DNASTAR SeqMan software (http://www.dnastar.com/).

3.2.3 Expression studies

To check the effect of CNVs in expression of \textit{TSPAN7} and \textit{IL1RAPL1}, RNA was extracted from the lymphoblastoid cells of the appropriate patients, and first-strand cDNA was synthesized by standard methods. cDNA primers were designed to amplify exonic sequences. We first tested these cDNA primers using human brain cDNA to ensure the PCR conditions were optimum for the successful amplification. Next, we tested the expression of \textit{TSPAN7} and \textit{IL1RAPL1} genes in
lymphoblast cDNA. Primer sequences for TSPAN7 are listed in Table 3-3 and those for IL1RAPL1 are listed in Table 3-4.

**Table 3-1 Primers used to amplify the coding regions and splice sites of TSPAN7.**

<table>
<thead>
<tr>
<th>EXON1-F</th>
<th>CCCGGCTTTTTTCAGTAGGAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXON1-R</td>
<td>GAGGGGTCCCATTTTGATT</td>
</tr>
<tr>
<td>EXON2-F</td>
<td>TTTCTCGTTGCTCAGGAAT</td>
</tr>
<tr>
<td>EXON2-R</td>
<td>GCTGTGTGGGTTTTGTTTT</td>
</tr>
<tr>
<td>EXON3-F</td>
<td>CCAGATTTCCTCCAGGTAGC</td>
</tr>
<tr>
<td>EXON3-R</td>
<td>CCCAAACACCCCTAACTCA</td>
</tr>
<tr>
<td>EXON4-F</td>
<td>CCCCTCCAGTAGGTACATTCA</td>
</tr>
<tr>
<td>EXON4-R</td>
<td>CCAACCTACAGGCAGTCCAT</td>
</tr>
<tr>
<td>EXON5-F</td>
<td>TTAGCTCACAAAGGTGCAC</td>
</tr>
<tr>
<td>EXON5-R</td>
<td>AAAGCCAAAGCTGCTGCT</td>
</tr>
<tr>
<td>EXON6-F</td>
<td>TGGGGAAGGTCATGTGTC</td>
</tr>
<tr>
<td>EXON6-R</td>
<td>TATGCCACGAGTTCTCCAG</td>
</tr>
<tr>
<td>EXON7-F</td>
<td>AGTGCCCTTCCCATTTAC</td>
</tr>
<tr>
<td>EXON7-R</td>
<td>CCTTCCAGAACCACAGA</td>
</tr>
</tbody>
</table>
Table 3-2 Primers used to amplify the coding regions and splice sites of *IL1RAPL1*.

<table>
<thead>
<tr>
<th>EXON</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXON1</td>
<td>NON-CODING, NOT SEQUENCED</td>
</tr>
<tr>
<td>EXON2-F</td>
<td>ATTGCACCGATCATGTTTGA</td>
</tr>
<tr>
<td>EXON2-R</td>
<td>CCACATGGCAGACATGTTAGA</td>
</tr>
<tr>
<td>EXON3-F</td>
<td>TGACCCCAATATGGATGCTCA</td>
</tr>
<tr>
<td>EXON3-R</td>
<td>AGACAACGCTTTTAGGCAA</td>
</tr>
<tr>
<td>EXON4-F</td>
<td>CCCTGTGGAATAAGTCAAGATACC</td>
</tr>
<tr>
<td>EXON4-R</td>
<td>TGCCCCAGAATATAAGGCACAA</td>
</tr>
<tr>
<td>EXON5-F</td>
<td>TGCAATTTTAAGCTTTTTTTTTT</td>
</tr>
<tr>
<td>EXON5-R</td>
<td>TCACCTATAGGAATCCACTTAGCA</td>
</tr>
<tr>
<td>EXON6-F</td>
<td>TGAAAGTGAAAAATATTGGGAAAA</td>
</tr>
<tr>
<td>EXON6-R</td>
<td>CAAATGGATTTAGCTGCAGT</td>
</tr>
<tr>
<td>EXON7-F</td>
<td>TGTTACCTGTCAGTTTGCCAATAAA</td>
</tr>
<tr>
<td>EXON7-R</td>
<td>AACAGTGTTTNTGCTTTTATCATT</td>
</tr>
<tr>
<td>EXON8-F</td>
<td>CATCAGATTCCGATTACCATTACA</td>
</tr>
<tr>
<td>EXON8-R</td>
<td>TCGGTGGCTCTAATGCAAAT</td>
</tr>
<tr>
<td>EXON9-F</td>
<td>ACCCGTAAACCCACATCTGA</td>
</tr>
<tr>
<td>EXON9-R</td>
<td>TATACGAGCTGCTGCCATTG</td>
</tr>
<tr>
<td>EXON10-F</td>
<td>AAATGGGACATTTGGAGACG</td>
</tr>
<tr>
<td>EXON10-R</td>
<td>TTCATGTGAACACACAAAGACG</td>
</tr>
<tr>
<td>EXON11A-F</td>
<td>AGGAGAAGCAAGTCCCAAACT</td>
</tr>
<tr>
<td>EXON11A-R</td>
<td>GTGGCTAGAGCTGTGGAGGT</td>
</tr>
<tr>
<td>EXON11B-F</td>
<td>GAACTCCAAGTTCTGGAAACG</td>
</tr>
<tr>
<td>EXON11B-R</td>
<td>ACAGCAGCAGTCGAGGATTT</td>
</tr>
</tbody>
</table>

Table 3-3 Primer sequences used for qPCR validation and cDNA amplification of \textit{TSPAN7} duplication.

| \textit{TSPAN7}-QPCR1-F | TGGGCAGTCAGTCTCTGTGG |
| \textit{TSPAN7}-QPCR1-R | GCATCAGCCTCTGTATGGT |
| \textit{TSPAN7}-QPCR2-F | AAAGCCCAGTGGCATCATA |
| \textit{TSPAN7}-QPCR2-R | CTGGGAGCAGGTCTCAAGTC |
| \textit{TSPAN7}-CDNA1-F | CGAGGAGAATGGGAGACCAAA |
| \textit{TSPAN7}-CDNA1-R | CATACTGATTGGCCTGATG |
| \textit{TSPAN7}-CDNA2-F | CGAGGAGAATGGGAGACCAAA |
| \textit{TSPAN7}-CDNA2-R | TGACGAACACAAACCCTGA |
Table 3-4 Primer sequences used for amplification of \textit{IL1RAPL1} cDNA.

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Forward Sequence</th>
<th>Reverse Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{IL1RAPL1}-CNV1-F</td>
<td>TGGGTGAAAATGACACTGGA</td>
<td>TGGGTGAAAATGACACTGGA</td>
</tr>
<tr>
<td>\textit{IL1RAPL1}-CNV1-R</td>
<td>TCCGCTGTACCCAAAGAAAG</td>
<td>TCCGCTGTACCCAAAGAAAG</td>
</tr>
<tr>
<td>\textit{IL1RAPL1}-CNV2-F</td>
<td>GCTCCGATTCCACACTTGAT</td>
<td>GCTCCGATTCCACACTTGAT</td>
</tr>
<tr>
<td>\textit{IL1RAPL1}-CNV2-R</td>
<td>AGGAGCTTGATGGTGCTT</td>
<td>AGGAGCTTGATGGTGCTT</td>
</tr>
<tr>
<td>\textit{IL1RAPL1}-CNV3-F</td>
<td>GAAGCCACCATTCTTTTTGG</td>
<td>GAAGCCACCATTCTTTTTGG</td>
</tr>
<tr>
<td>\textit{IL1RAPL1}-CNV3-R</td>
<td>TGACCACCAGCAGTACAAGG</td>
<td>TGACCACCAGCAGTACAAGG</td>
</tr>
</tbody>
</table>

3.3 Results

3.3.1 \textit{TSPAN7}

We identified a ~121 Kb duplication (ChrX 38,250,331 to 38,371,333 (UCSC 2004) flanked by SNPs rs5917211 and rs5917628. The duplication included 12 SNPs on the NspI and StyI Affymetrix microarray, spanned exons 2-8 of \textit{TSPAN7} and was not present in a control cohort of 1652 samples. This CNV call was validated by performing SYBR Green-I based quantitative Real-Time PCR and it was observed that this CNV gain was inherited in the affected male proband from his unaffected mother. To ascertain a possible effect of this CNV on gene expression of \textit{TSPAN7}, we first checked the expression of \textit{TSPAN7} mRNA in lymphoblast cDNA from normal individuals and found a moderate expression. Next, we amplified and sequenced \textit{TSPAN7} mRNA procured from Epstein Barr virus-transformed lymphoblasts from the proband. RT-PCR revealed a normal size transcript was expressed in the patient, and no change in mRNA sequence was observed (Figure 3-1). These results indicated that this CNV gain does not disrupt the expression or coding sequence of the \textit{TSPAN7} gene. To further investigate the
potential involvement of *TSPAN7* in the etiology of autism, we sequenced the entire coding region and exon-intron boundaries of *TSPAN7* in a cohort of 250 (210 male and 40 female) unrelated autistic probands. Analysis of sequencing data revealed no coding mutations in any of the 250 subjects.

**Figure 3-1** Agarose gel shows the amplification of *TSPAN7* cDNA (bands) using lymphoblast RNA of patient with intragenic duplication of *TSPAN7*. (A) 732 bp PCR product shows the amplification of *TSPAN7* cDNA containing Exons 1-7. (B) 381 bp PCR product shows the amplification of Exons 1-3 of *TSPAN7* cDNA
3.3.2 **IL1RAPL1**

In another male autism patient, we identified an intronic 82 Kb deletion in the *IL1RAPL1* gene. Deletion was initially validated by performing PCR using primers within the deleted region and it failed to amplify in the proband and unaffected male sibling (Figure 3-2). To investigate the segregation of this CNV, qPCR was performed. qPCR results confirmed that the deletion was maternally inherited to the male proband, female affected sibling and a male unaffected sibling. We attempted to investigate if this intronic deletion resulted in the loss of normal expression of this gene, however, the expression of this gene was not detected in the lymphoblast. Therefore, we were unable to check the effect of *IL1RAPL1* deletion in this family.

Mutation screening of all coding regions of *IL1RAPL1* was performed using DNA from 250 unrelated probands with autism. The mutation analysis unveiled three rare variants, one coding non-synonymous variant, one coding synonymous variant and an intronic variant.

The non-synonymous variant c.G349T resulted in the substitution of Alanine to Serine at amino acid position 113. Segregation analysis revealed that the mutation was maternally inherited to male proband and was not transmitted to unaffected male sibling (Figure 3-3). The Alanine residue at position 113 is highly conserved (Figure 3-4) and this variant was not seen in 48 unrelated controls sequenced in the ExoSeq project ([http://www.sanger.ac.uk/cgi-bin/humgen/exoseq/exoseqview](http://www.sanger.ac.uk/cgi-bin/humgen/exoseq/exoseqview)). Furthermore, this variant was not identified in additional 276 control chromosomes (Piton et al., 2008).

A synonymous C to T variant was also identified in the proband with autism. Additionally, a G to A change in intron 5’ was also observed in two unrelated probands. This variant was 30 bp upstream of exon 6 and is not predicted to have any effect on splicing.
Figure 3-2 PCR amplification using primers in the deleted region.

Figure 3-3 Chromatogram shows a C to T substitution at cDNA nucleotide position 349. Pedigree shows the maternal inheritance of the variant to the male proband.
Figure 3-4 Conservation of Alanine residue at position 113 in the IL1RAPL1 protein.
3.4 Discussion

In this chapter, we studied two autism candidate genes *TSPAN7* and *IL1RAPL1* by employing two different approaches, namely the investigation of effect of a CNV at the mRNA level, and screening for point mutations by direct DNA sequencing to identify additional cases with mutation within these genes.

Disruption of the *TSPAN7* by a balanced translocation and sequence mutations of this gene implicated it in the etiology of XLID (Zemni et al., 2000). Another study reported a two bp deletion in this gene in patients with XLID, thus, further supporting the role of this gene ID (Abidi et al., 2002). In our own study, we identified a 121 Kb gain partially spanning the *TSPAN7* (Marshall et al., 2008). In a recent study, duplication of similar region was reported in two patients, one with syndromic and the other with non-syndromic intellectual disability (Froyen et al., 2007). We hypothesized that *TSPAN7* may play a role in the etiology of autism and the partial gene duplication identified in proband with autism may result in disruption of this gene. To test this hypothesis, expression analysis of *TSPAN7* was performed on lymphoblast cDNA of the proband with partial gene duplication of this gene. However, our PCR and sequencing results confirmed the expression of the normal transcript. Thus, the CNV does not directly disrupt the expression of this gene in our proband and may represent a rare benign CNV. A possible reason could be that the duplicated region may not necessarily be directly in tandem with the normal genomic region. Recently, duplication of this region has also been reported in male and female healthy individuals, thus supporting our findings that this CNV is likely a rare, benign variant which is not associated with a phenotype (Cai et al., 2008). Nonetheless, the null result of the *TSPAN7* CNV further emphasizes the importance of validating possible CNV effects at the mRNA level to establish a contribution to disease.

Furthermore, by direct DNA sequencing, we were unable to identify any sequence mutations in 250 unrelated probands with autism. Our results indicate that coding mutations in *TSPAN7* are not associated with our cohort of autism patients. However, the involvement of *TSPAN7* mutations in a very small fraction of autism patients cannot be excluded and further studies are required to investigate if genetic variants in non-coding regions of this gene are involved in the increased risk or etiology of autism. Altogether, our findings exclude the involvement of *TSPAN7* sequence variants in most cases of autism.
In this study, we also investigated the role of *ILIRAPLI* in the pathogenesis of autism by studying the effect of an intronic deletion in this gene, as well as by sequencing the coding regions of this gene in 250 unrelated probands with autism. Large deletions resulting in contiguous deletion syndromes initially suggested that deletion of the *ILIRAPLI* gene might be responsible for the ID phenotype in those patients (Jin, Gardner, Viswesvaraiah, Muntoni, & Roberts, 2000). In 2006, a truncating mutation in *ILIRAPLI* gene was reported in four affected males, thus, confirming the role of this gene in etiology of XL-NSID (Tabolacci et al., 2006). These studies established that mutations in this gene are associated with XL-NSID. We discovered a 82 Kb deletion in intron 5 of *ILIRAPLI* and proposed that this gene may also play a role in etiology of autism (Marshall et al., 2008). At the same time, another study reported a pericentromeric inversion of the X chromosome in a patient with intellectual disability and autism which resulted in disruption of *ILIRAPLI* (Bhat et al., 2008a).

Here, we explored further the association of this gene with autism. We first attempted to observe the effect of the intronic deletion in our patient on the expression of *ILIRAPLI*, however we were unable to amplify *ILIRAPLI* mRNA from the lymphoblast cDNA due to the extremely low level of expression of the gene in this tissue. Segregation analysis showed the CNV was maternally transmitted to the male proband and an affected female proband. However, deletion was also seen in unaffected male sibling, therefore, did not segregate with the phenotype, consequently, it is likely to represent a benign rare variant. On the other hand, the presence of *ILIRAPLI* deletion in the male sibling without a phenotype could also be explained by incomplete penetrance. Alternately, it can also be explained by the Threshold Model of relative contribution for neuropsychiatric disorders (Cook, Jr. & Scherer, 2008), where, this CNV alone may not have substantial effect to express phenotype and may require additional genetic mutations at other loci to cross the threshold to demonstrate phenotype.

Mutation screening of *ILIRAPLI* discovered three new variants among 250 autism patients. One synonymous variant is likely a rare benign variant as it does not alter the amino acid. The intronic variant identified in another proband is 30 bp upstream of exon 5, and hence, does not directly involve the splice site. The best way to check the effect of this variant, if any, on the splicing of *ILIRAPLI* mRNA would the expression analysis but our ability to perform such experiments was limited by the lack of expression of this gene in lymphoblasts.
The third variant resulted in substitution of a conserved Alanine residue at position 113 to Serine. The mutation was maternally inherited to male proband but was not inherited to an unaffected male sibling, therefore segregated with the phenotype. Moreover, this mutation is within the extracellular domain of protein and was not observed in more than 300 control chromosome which further strengthened the causal role of this variant. However, further functional studies are required to determine if this variant is pathogenic. Interestingly, in this proband, in addition to the *IL1RAPL1* sequence mutation, we also identified a *de novo* duplication partially spanning the *DLGAP2* (Marshall et al., 2008). DLGAP2 protein is predicted to be involved in the organization of synapses and in neuronal cell signaling, and this gene has recently been reported as an autism candidate (Pinto et al., 2010). These finding can be explained by the model of relative contribution to the susceptibility to autism where more than one genetic variant may be required to express the phenotype (Cook, Jr. & Scherer, 2008).

In summary, our data indicate that the *TSPAN7* duplication and sequence mutations are not associated with cases of autism; however, contribution of this gene in a small fraction of cases cannot be excluded due to limitation of sample size of this study. We also provide further evidence for involvement of *IL1RAPL1* in the pathogenicity of autism. Our data highlights the genetic overlap between autism and ID, therefore, studies of other ID genes in autism may imply some of these genes in autism phenotype.
Chapter 4. Disruption at the *PTCHD1* locus on Xp22.11 in autism spectrum disorder and intellectual disability

*Originally published in Science Translational Medicine (Appendix 3)*


**Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability.**


**Contributions:** For this manuscript, I performed the CNV validations, segregation analysis, sequencing, multiple tissue expression analysis and cloning experiments. I also prepared the initial draft of manuscript and performed subsequent revisions. Microarray analysis, sequencing of additional cohorts, phenotypic analysis and functional studies of Hedgehog Signaling pathway were performed by our collaborators.
Running title: **PTCHD1**, mutations in autism and intellectual disability.

**Abdul Noor**¹, Annabel Whibley², Christian R. Marshall³, Peter J. Gianakopoulos¹, Amelie Piton⁴, Andrew R. Carson³, Marija Orlic-Milacic¹, Anath Lionel⁵, Daisuke Sato³, Dalila Pinto³, Irene Drmic⁵, Carolyn Noakes⁵, Lili Senman⁵, Xiaoyun Zhang⁶, Rong Mo⁶, Julie Gauthier⁴, Jennifer Crosbie⁷, Alistair T. Pagnamenta⁸, Jeffrey Munson⁹, Annette M. Estes¹⁰, Andreas Fiebig¹¹, Andre Franke¹¹, Stefan Schreiber¹¹, Alexandre F.R. Stewart¹³, Robert Roberts¹³, Ruth McPherson¹³, Stephen J. Guter¹⁴, Edwin H. Cook Jr¹⁴, Geraldine Dawson¹⁵, Gerard D. Schellenberg¹⁶, Agatino Battaglia¹⁷, Elena Maestrini¹⁸, Autism Genome Project Consortium†, Linda Jeng¹⁹, Terry Hutchison²⁰, Evica Rajcan-Separovic²¹, Albert E. Chudley²², Suzanne M.E. Lewis²³, Xudong Liu²⁴, Jeanette Holden²⁴, Bridget Fernandez²⁵, Lonnie Zwaigenbaum²⁶, Susan E. Bryson²⁷, Wendy Roberts⁵, Peter Szatmari²⁸, Louise Gallagher²⁹, Michael R. Stratton³⁰, Jozef Gecz³¹, Angela F. Brady³², Charles E. Schwartz³³, Russell J. Schachar⁷, Anthony P. Monaco⁸, Guy A. Rouleau⁴, Chi-chung Hui⁶,³⁴, F. Lucy Raymond², Stephen W. Scherer³,³⁴, John B. Vincent¹,³⁵

¹Neurogenetics Section, Centre for Addiction and Mental Health, Toronto, Ontario, Canada;

²Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK;

³Program in Genetics and Genome Biology and The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada;

⁴Center of Excellence in Neuromics, Centre Hospitalier de l'Université de Montréal, and Department of Medicine, University of Montreal, Montreal, Quebec, Canada;

⁵Autism Research Unit, The Hospital for Sick Children, Toronto, Ontario, Canada;

⁶Program in Developmental & Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario, Canada;

⁷Department of Psychiatry, Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, Ontario, Canada;

⁸Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK;
9Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA;

10Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington, USA;

11Institute for Clinical Molecular Biology, Christian-Albrechts-University and Biobank popgen, Kiel, Germany;

12Dept. of Internal Medicine, University of Kiel, Kiel, Germany;

13University of Ottawa Heart Institute, Ottawa, Ontario, Canada;

14Laboratory of Developmental Neuroscience, University of Illinois at Chicago, Chicago, Illinois, USA;

15Autism Speaks, New York, NY, and Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA;

16Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA;

17Stella Maris Institute for Child and Adolescent Neuropsychiatry, Calambrone, Pisa, Italy

18Department of Biology, University of Bologna, Bologna, Italy;

19Department of Laboratory Medicine, University of California, San Francisco, San Francisco, California, USA;

20Department of Neurology, University of California San Francisco, San Francisco, California, USA

21Department of Pathology (Cytogenetics), Children's and Women's Health Centre of BC, Vancouver, British Columbia, Canada;

22Program in Genetics and Metabolism, Children's Hospital, Winnipeg, Manitoba, Canada;
23Department of Medical Genetics, UBC, Vancouver, BC, Canada

24Depts. Psychiatry and Physiology, Queen's University, Ontario, Canada;

25Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John’s, Newfoundland, Canada;

26Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada;

27Departments of Pediatrics and Psychology, Dalhousie University, Halifax, Nova Scotia, Canada;

28Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada;

29Neuropsychiatric Genetics Research Group, Trinity Centre for Health Sciences, Trinity College Dublin, Dublin, Ireland;

30Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus Hinxton, Cambridge, UK;

31SA Pathology, Women’s and Children’s Hospital, North Adelaide and The University of Adelaide, Australia;

32North West Thames Regional Genetic Centre, Northwick Park Hospital, Harrow, UK;

33JC Self Research Institute, Greenwood Genetic Center, Greenwood, South Carolina, USA;

34Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada;

35Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada;

†See supplementary text for full list of names and affiliations;

Correspondence should be addressed to: J.B.V. (john_vincent@camh.net); or to S.W.S. (stephen.scherer@sickkids.ca).
4.1 Abstract

Autism is a common neurodevelopmental disorder with a complex mode of inheritance. It is one of the most highly heritable of the complex disorders, however, the underlying genetic factors remain largely unknown. Here, we report mutations in the X-chromosome PTCHD1 (patched-related) gene, in seven families with autism spectrum disorder (ASD) and in three families with intellectual disability (ID). A 167 Kb microdeletion spanning exon 1 was found in two brothers, one with ASD the other with learning disability and ASD features, and a 90 Kb microdeletion spanning the entire gene was found in three males with ID in a second family. In 900 ASD and 208 ID male probands we identified seven different missense changes in eight probands, all male and inherited from unaffected mothers, and not found in controls. Two of the ASD individuals with missense changes also carried a de novo deletion at another ASD-susceptibility locus (DPYD and DPP6), suggesting complex genetic contributions. In additional males with ASD, we identified deletions in the 5’ flanking region of PTCHD1 disrupting a complex non-coding RNA and potential regulatory elements; equivalent changes were not found in male control individuals (p=1.2 x10^{-5}). Systematic screening at PTCHD1 and 5’-flanking regions, suggests involvement of this locus in ~1% of ASD and ID individuals.

4.2 Introduction

Autism (MIM 209850) is a severe, lifelong neurodevelopmental disorder characterized by impairments in communication and socialization, and by repetitive behavior. Recent studies of sub-microscopic genomic copy number variation (CNV) have identified several loci associated with Autism Spectrum Disorder (ASD; MIM 209850) (Szatmari et al., 2007; Sebat et al., 2007). De novo CNVs associated with ASD have been reported in ~7% of simplex families and ~2% of multiplex families (Sebat et al., 2007; Marshall et al., 2008). CNV studies have also led to the identification of autism candidate genes such as SHANK3 (MIM 606230) and NRXN1 (MIM 600565) (Szatmari et al., 2007; Durand et al., 2007; Moessner et al., 2007). Intellectual disability (ID) is frequently associated with autism (in up to ~30% of cases for ASD, and ~67% for autism) (Chakrabarti & Fombonne, 2005). Moreover, mutations in several X-linked ID (XLID) genes (e.g. NLGN4 and ILIRAPLI) have been shown to result in an autistic phenotype, which suggests that autism and ID may often share a common genetic etiology (Marshall et al., 2008; Jamain et
al., 2003; Laumonnier et al., 2004; Bhat et al., 2008a; Piton et al., 2008). We previously reported a 167 Kb microdeletion of exon 1 of $PTCHD1$ (NM_ 173495.2) on chromosome Xp22.11 (Marshall et al., 2008). $PTCHD1$ has three exons spanning ~62 Kb and it is predicted to encode a protein of 888 amino acids. *In silico* analysis suggests that $PTCHD1$ is a transmembrane protein containing a patched-related domain with twelve transmembrane helices, highly related to the Hedgehog (Hh) receptors PATCHED1 (PTCH1) and PTCH2 as well as to Niemann-Pick Type C1 protein (NPC1). Hh is one of the key signaling pathways involved in the formation of the neural tube and brain, specifically the differentiation of motor neurons ventrally and commissural interneurons dorsally (Jessell, 2000; Jacob & Briscoe, 2003). Mutations in Sonic Hedgehog, $SHH$ (MIM 600725), have been reported in patients with developmental abnormalities, delay in speech acquisition and learning disabilities (Hehr et al., 2004). Niemann-Pick disease type C1 also involves neurological and intellectual deficits (MIM 257220). This led us to investigate a possible role for $PTCHD1$ as a candidate gene for ASD and ID. 

Further to the initial CNV-screening ASD cohort (Marshall et al, 2008), we have now analyzed CNV screening data for a cohort of ID subjects, as well as cohorts of unaffected subjects, and, where CNVs have been identified at the $PTCHD1$ locus, we have validated and characterized the CNVs and their inheritance in the families. This screening identified a second deletion at $PTCHD1$, segregating among males in a family with ID. This finding also prompted screening of additional ASD cohorts for CNVs at the $PTCHD1$ locus. We also screened a proportion of the cases and controls for coding mutations within $PTCHD1$ (see table S5 for details on cohorts studied). Preliminary functional evidence for the $PTCHD1$ protein is consistent with a role in Hh signaling.

### 4.3 Results

#### 4.3.1 CNV Analysis of $PTCHD1$

We characterized the precise breakpoints of the 167 Kb deletion at $PTCHD1$ identified in the male proband from Family 1. This CNV also disrupts long, spliced non-coding RNAs (ncRNAs) on the opposite strand, but no other coding genes were interrupted (Figure 4-1). The deletion was validated in the family using both PCR and SYBR-Green I-based real-time quantitative PCR (qPCR) and was found to be transmitted from a heterozygous unaffected
mother to two affected dizygotic twin sons, also to an unaffected daughter (Figure 4-2). X-chromosome inactivation (XCI) analysis of the mother, carrier of the PTCHD1 deletion, revealed a highly skewed allelic ratio of 94:6.

To assess the possible involvement of CNVs disrupting X-chromosomal genes in the etiology of ID, we initially screened 246 males with intellectual disability and probable X-linked inheritance using a custom-designed NimbleGen 385K array with probes targeting the X chromosome. A 90 Kb deletion encompassing the entire PTCHD1 gene and 5’ exons of the ncRNAs (but no other known coding genes) was found in a male ID patient (Family 2). The deletion was validated using qPCR revealing that the deletion was maternally inherited in two affected brothers and their affected uncle (Figure 4-2). XCI analysis revealed allele ratios of 51:49 and 75:25 from lymphocytes of 2 obligate female carriers.

Subsequent to the ascertainment of these cohorts, an additional case of ID with dysmorphic features was referred to us through cytogenetic services at the University of California, San Francisco, California, USA. CNV analysis with a custom designed 105K microarray identified a 146 kb deletion in this patient which spans PTCHD1 exon 1 and upstream regions (chrX:23,146,927-23,293,273, hg18).

4.3.2 Mutation Screening of PTCHD1

In order to identify additional cases with PTCHD1 mutations, we sequenced the coding regions in 900 (M=723; F=177) unrelated ASD cases and 225 unrelated male ID cases. Seven missense changes were identified in six unrelated probands with ASD and two ID probands (Figure 4-2; Figure S 4-1 & Figure S 4-2; table S1). All of these variants, which resulted in the substitution of highly conserved amino acids, were inherited from unaffected carrier mothers (Figure S 4-1). In six of the eight families the missense variants appear to segregate with the phenotype, however in Family 6 L73F did not segregate, and in Family 7 the A470D did not segregate in different loops (not shown) of the extended pedigree (see Figure 4-2 and table S1 for details).

We sequenced the entire coding region of PTCHD1 in 700 control individuals (M=531 F=169), and none of the missense changes identified from among the ASD and ID patient cohorts has been detected. Only two missense changes have been identified: P252L from amongst our
controls, and N497K reported in the SNP database (rs35880456, in 1 out of 39 screened; NCBI) (2010b), both in females who were heterozygotes. Altogether, absence of PTCHD1 missense variants indicates that these variants are significantly enriched in the males with ASD (6/723 male ASD versus 0/531 male control: Fisher’s exact test: $p = 0.042$) and may contribute to the phenotype.

Additional controls were sequenced for the exons in which missense mutations were identified. We tested control chromosomes for the sequence underlying the I173V and V195I mutations ($N=1101$ chromosomes), the ML336_337II mutation ($N=1193$), and the L73F, E479G, A470D and H359R mutations ($N=869$) and detected none of these variants.

4.3.3 CNVs upstream of PTCHD1 (PTCHD1AS1/PTCHD1AS2 locus)

Additionally, from a study of 996 ASD families examined with the Illumina 1M BeadChip (Pinto et al., 2010), we identified eight deletions in probands or affected siblings, and a ninth in a father with a diagnosis of Broad Autism Phenotype (BAP) (Hurley et al., 2007; Constantino & Todd, 2005), all occurring 5' of PTCHD1, and overlapping with an anti-sense non-coding RNA, PTCHD1AS1/PTCHD1AS2 (Figure 4-1). A tenth deletion at this upstream locus was identified in a patient from a CNV study of 167 unrelated attention deficit-hyperactivity disorder (ADHD) patients. The ADHD proband with the deletion also has a BAP diagnosis. These deletions were validated with qPCR and exact breakpoints were mapped (table S2). Additional CNV data for these 10 individuals are included in table S3.

We analyzed SNP microarray data from 10,246 control individuals (4,829 male; 5,417 female), for CNVs at PTCHD1 and the upstream region. In a 1.4-Mb region spanning from PTCHD1 to adjacent genes PRDX4 (proximal) and ZNF645 (proximal), we identified 15 CNVs (7 duplications and 8 deletions); however, it is notable that only 1 male control with a deletion was identified, which was 20.6 Kb in length and did not disrupt any known exons of any genes or non-coding RNAs, or any of the identified conserved or putative regulatory sequences. The remaining 7 deletions were all identified among female controls, consistent with the X-linked recessive inheritance observed for the PTCHD1 mutations. Thus, PTCHD1 and upstream
deletions were not observed in 4,829 male controls, or in the Database of Genomic Variants (Iafrate et al., 2004), which suggests that the CNV directly disrupting PTCHD1 and the 6 CNVs located just upstream in unrelated ASD probands are associated with autism (male ASD cases \(N=7\), out of 1,185; male controls \(N=0\) out of 4,829; Fisher’s exact test: \(p =1.2\times10^{-5}\)).

Subsequently, an additional case with ID and with a number of other clinical features was referred to us from Dr. A.E. Chudley, Children’s Hospital, Winnipeg and Dr. Evica Rajcan-Separovic at Children's and Women's Health Centre of BC, Vancouver, BC. This family has a maternally inherited 112 kb deletion upstream of PTCHD1, including DDX53 and at least one exon of PTCHD1AS1 (chrX:22,819,116-22,931,588, hg18).

4.3.4 Expression and Functional Studies of PTCHD1

Expression analysis for the PTCHD1 and the ncRNA transcripts suggests that they are transcribed in brain regions, notably the cerebellum, as well as in other tissues (Figure 4-3 and Figure S 4-3). RNA in situ hybridization of Ptchd1 in mouse showed widespread expression in the developing brain from E9.5/10.5 to P1 (Figure 4-4a), as well as broad expression in the adult mouse brain (6 months), with highest density in the cerebellum (see Allen brain atlas online (2010a)).

To investigate its function, we studied the sub-cellular localization of PTCHD1 and found that a PTCHD1-GFP fusion protein predominantly localizes to the cell membrane (Figure 4-4b). We further hypothesized that PTCHD1 may function in the Hh-signaling pathway and have similar functional attributes as PTCH1 and PTCH2. We performed a Gli-dependent transcription assay in Hh-responsive 10T1/2 cells to test whether PTCHD1 could interfere with Hh signaling. In 10T1/2 cells, overexpression of PTCH1 or PTCH2 inhibits transcription from a Gli-luciferase reporter containing multiple copies of the Gli protein-binding site in the presence of Smoothened agonist purmorphamine (Sinha & Chen, 2006) (Figure 4-4c) or Gli2 (Figure S 4-4). Similar to PTCH proteins, PTCHD1 also exerted a statistically significant inhibitory effect in these assays suggesting that PTCHD1 functions in the Hedgehog signalling pathway.
4.4 Discussion

We have identified microdeletions that directly disrupt the *PTCHD1* gene in males in three families affected with either ASD, ID or learning disability. These deletions are maternally inherited and were not observed in more than 10,000 controls, indicating that these alterations are associated with ASD and ID. We also report seven maternally inherited missense mutations in eight male probands. These variants were not seen in more than 500 controls, further supporting a possible role of this gene in autism and ID.

In addition, we have found another 11 deletions that map to regions upstream of *PTCHD1*. The region 5’ and distal to *PTCHD1* is relatively gene poor. Within this upstream region, a coding gene, *DDX53*, encoding DEAD Box 53, lies ~335 Kb 5’ to *PTCHD1*. Five of the 11 upstream deletions span *DDX53*. However, based on the function of the DDX53 protein and the expression pattern of this gene (which is restricted mainly to testis and tumor cells (Cho et al., 2002)), it is unlikely to contribute to the ASD or ID phenotype. Additionally, within the gene-poor region between *PTCHD1* and *DDX53*, there is a putative pseudogene of *FAM3C*, *FAM3C2*, which is disrupted by five of the 10 upstream deletions. *FAM3C*, a cytokine-like gene on 7q31.31, consists of 10 exons (Zhu et al., 2002) whereas *FAM3C2*, although 99% identical, has no intron/exon structure and is interrupted by a short interspersed nuclear element (SINE). It appears to have inserted on Xp22 after human/chimp evolutionary divergence. Since no mRNA or EST matches exactly to *FAM3C2*, it is most likely an untranscribed processed pseudogene.

We examined the region just distal to *PTCHD1* in detail and identified a number of putative enhancer and promoter sequences, as well as conserved (and putative regulatory) elements (Figure 4-1). We also identified several overlapping spliced long (>200nt) non-coding (nc) RNAs (*PTCHD1AS1* from cDNA clone IMAGE:1560626; BX115199) and *PTCHD1AS2* (from cDNA clone BRSTN2000219; DA355362), which map to the opposite strand and distal to *PTCHD1* (see Figure 4-1). 5’RACE (Rapid Amplification of cDNA Ends) shows that a number of splice variants of these transcripts originate at the CpG island just upstream of *PTCHD1*, encompassing its putative promoter. Similar antisense transcripts are present at syntenic loci in other mammalian species, at least two exons of which appear to be conserved between rat, mouse and humans (Figure 4-1).
Although these ncRNAs do not appear to encode protein, they may serve as regulators for other coding genes, particularly for *PTCHD1*, since the 5’ exons are adjacent on opposite strands. Such ncRNAs may regulate expression of a coding transcript on the opposite strand through a number of mechanisms, including modification of chromatin, transcriptional regulation and post-transcriptional modification (Mercer, Dinger, & Mattick, 2009; Kleinjan & van, V, 2005).

All 11 of the upstream deletions as well as the three *PTCHD1* deletions (Families 1 and 2) disrupt conserved (and putative regulatory) sequences and/or exons of these ncRNAs (see Figure 4-1). These deletions were not inherited by a subset of the affected family members; also, the missense variants do not segregate with disease in two families (Families 6 & 7) (Figure 4-2). These findings are similar to other previously reported major affect ASD loci such as 16p11.2 (Weiss et al., 2008) and are also consistent with the complex, non-Mendelian inheritance believed to control the etiology of autism. As discussed in a recently proposed threshold model of relative contribution in ASD (Cook, Jr. & Scherer, 2008), it is anticipated that multiple common and rare variants may act in concert to generate the phenotype. For instance, under this model, some *de novo* CNVs may be solely sufficient to cause ASD. Conversely, other *de novo* CNVs may have weaker effects, requiring contributions from additional loci (for example additional risk haplotypes, or other CNVs), or environmental risk factors, for the burden of contributory factors to cross a risk threshold and result in an ASD phenotype. In three of the eight families (6 ASD and 2 ID) that carry putative *PTCHD1* missense mutations (Families 8, 9 and 10), we have identified other CNVs involving genes that may also contribute to the phenotype. In Family 9, in addition to the I173V substitution, we found a *de novo* ~1.1 Mb loss at 1p21.3 resulting in deletion of the entire *DPYD* gene (MIM 274270), encoding dihydropyrimidine dehydrogenase (DPD) (Marshall et al., 2008). Complete DPD deficiency results in highly variable clinical outcomes, with convulsive disorders, motor retardation, and intellectual disability being the most frequent manifestations, and autistic features occasionally reported (van Kuilenburg et al., 1999). In this family, a balanced translocation, t(19;21)(p13.2; q22.12) is also present in the proband, but is inherited from the unaffected mother and shared with an unaffected sister (see Supplementary Materials). In Family 10, which shows the V195I substitution in *PTCHD1*, we have previously reported a 66 Kb *de novo* loss at 7q36.2 that results in deletion of the third exon of *DPP6* (MIM 126141) – previously reported as a positional and functional candidate gene for autism (Marshall et al., 2008). In ID Family 8, we have identified a
H359R substitution in *PTCHD1* and a 2 Kb deletion spanning the last exon of *SLC16A2*, both variants are maternally inherited. The phenotype in this family was severe ID compatible with Allan-Herndon-Dudley syndrome (MIM 300523) (Friesema et al., 2004; Schwartz et al., 2005), for which mutations in *SLC16A2* have previously been reported.

Thus, in two ASD individuals we have evidence for the possible involvement of more than one locus in the disease, and these findings may support the threshold model of relative contribution in ASD described above (Cook, Jr. & Scherer, 2008) and polygenic inheritance in autism. As such, some *de novo* CNVs may be highly penetrant in causing ASD susceptibility (e.g. disruption of *PTCHD1* in Family 1). Conversely, other *de novo* CNVs (e.g. *DPP6* and *DPYD* deletions) may have more subtle effects, requiring contributions of additional loci (e.g. *PTCHD1* missense mutations in the case of Families 9 & 10) for ASD to be phenotypically evident. This scenario may also apply to the ID families with *PTCHD1* mutations, although for Family 8 the *PTCHD1* missense variant contribution is likely overwhelmed by the phenotypic effect of a whole exon deletion of *SLC16A2*.

*PTCHD1* gene expression showed high correlation with expression of other cerebellar genes such as *ZIC1*, *CADPS2*, *EN2*, *CBLN1*, and with synaptic genes such as *PCLO*, *NRXN3*, *SNAP25*, *SYT2*, *DPP6* and *DPP10* (see table S4). Cerebellar abnormalities have frequently been linked to autism, including recent magnetic resonance imaging (MRI) studies showing significant decrease in cerebellar grey matter (Courchesne et al., 2001; Toal et al., 2009), and decreased cerebellar connectivity and activity (Mostofsky et al., 2009).

*PTCHD1* encodes a Patched-related protein with 12 transmembrane domains and a sterol-sensing domain, structurally similar to the Hh receptors PTCH1 and PTCH2, as well as the Niemann-Pick Type C1 protein (NPC1) and several others (Figure S 4-5). Many Patched-related genes have been found in various organisms, from nematodes to humans, and they appear to play diverse biological functions, including cytokinesis, growth and pattern formation (Zugasti, Rajan, & Kuwabara, 2005). For instance, there are just seven patched-related genes in humans (*PTCH1*, *PTCH2*, *PTCHD1*, *PTCHD2*, *PTCHD3*, *NPC1* and *c6orf138*) (see Figure S 4-5), whereas in *C. elegans* there are at least 26 patched-related genes, with diverse roles in development in addition to Hh signaling, including cytokinesis, growth and pattern formation.
In 10T1/2 cells, we have demonstrated an inhibitory effect of PTCHD1 on Gli-dependent transcription. Although these results suggest that PTCHD1 exhibits biochemical activity in Hh-dependent processes similar to that of PTCH1 and 2, other functions or roles for PTCHD1 cannot be excluded at this point.

In summary, our data indicate that mutations at the PTCHD1 locus are highly penetrant and strongly associated with ASD (including BAP) and ID in ~1.1% and ~1.3% of the individuals analyzed, respectively (based on probands for whom comprehensive mutation screening, for both CNVs and sequence variants, has been performed (4 out of 353 ASD, and 3 out of 225 ID). Overall, our finding are reminiscent of genetic findings for several other X chromosome genes, including NLGN4 (Jamain et al., 2003; Laumonnier et al., 2004) and IL1RAPL1 (Bhat et al., 2008a; Piton et al., 2008; Carrie et al., 1999b), in that mutations can apparently cause either ASD or ID (or both), and thus PTCHD1 may be a gene for both. IL1RAPL1, for example, was initially reported as a gene for non-syndromic X-linked ID (34), and then subsequently was also found to harbor mutations in ASD pedigrees (9, 10). We have also identified two families in whom at least two loci may be contributing to the pathogenesis of ASD, and another seven families bearing upstream microdeletions that disrupt a complex non-coding RNA, providing possible genetic explanations for the clinical heterogeneity of these disorders. Finally, our results raise the possibility that Hh signaling may be perturbed in these conditions. This discovery may help provide possible targets for therapeutics in individuals with mutations at this locus.

4.5 Methods

4.5.1 Source of Subjects

CNVs at the PTCHD1 locus were initially assessed in 427 ASD patients, as described (Marshall et al., 2008). DNA samples from 900 individuals diagnosed with ASD were sequenced for PTCHD1 mutations. Among these, 400 samples were collected at three sites, namely The Hospital for Sick Children (HSC) in Toronto and child diagnostic centers in Hamilton, Ontario and St, John’s, Newfoundland. Details of these samples are published elsewhere (Moessner et al., 2007). 420 ASD cases were recruited at Montreal, details of these samples are published elsewhere (Gauthier et al., 2006). Another 80 ASD probands from the Autism Genetic Resource Exchange (AGRE) were also included. The second cohort of 996 autism probands was recruited
at different sites as a part of the Autism Genome Project (AGP); ascertainment is described elsewhere (Pinto et al., 2010). 246 male patients with intellectual disability were recruited from the UK, United States, Australia, Europe and South Africa as the IGOLD study. A subset of 225 from this cohort were also used for sequence analysis of \textit{PTCHD1}. Details of these samples are published elsewhere (Tarpey et al., 2009b). 167 unrelated patients diagnosed with ADHD were recruited through the Department of Psychiatry at the Hospital for Sick Children, Toronto. Microarray data from controls included 1,123 (M=623, F=500) controls recruited from northern Germany as a part of the PopGen project, 1,234 (M=586, F=648) healthy controls of European origin recruited from the province of Ontario, Canada, 1,287 (M=383, F=904) controls from the Study of Addiction: Genetics and Environment (SAGE), 1,320 (M=589, F=1320) controls from Children’s Hospital of Philadelphia (CHOP), 4783 (M=2460, F=2323) controls were recruited by the Wellcome Trust Case Control Consortium, 440 (M=158, F=282) controls were recruited by The Centre of Addiction and Mental Health (CAMH) and GlaxoSmithKline (GSK), and 59 (M=30, F=29) from the Centre d’Etude Polymorphisme Humaine (CEPH) HapMap controls (total \textit{N}=5,023). We sequenced more than 650 Ontario controls obtained from The Centre for Applied Genomics (TCAG) and The Centre for Addiction and Mental Health (CAMH). Details of all samples included in the study are summarized in table S5. Institutional ethical review board approval (CAMH, HSC, CHOP and all other collaborating institutions) was obtained for the study, and informed written consent was obtained for each family. Details of the clinical findings in families with \textit{PTCHD1} mutations or CNVs are summarized in table S1.

4.5.2 \hspace{1cm} \textbf{Copy Number Variation Analysis}

We used Affymetrix 500K SNP arrays to assess CNVs in a cohort of 427 ASD cases. Details on the methods of copy number analysis and complete results are published elsewhere (Marshall et al., 2008). Only the CNV result at \textit{PTCHD1} is described here. Another cohort of 996 autism probands was analyzed on 1M BeadChips (Illumina) (Pinto et al., 2010). 246 male patients with ID were analyzed on a custom designed NimbleGen 385K array. Genomic DNA samples were sent to NimbleGen for the hybridizations to be performed. Each patient sample (Cy5-labelled) was co-hybridised with DNA from the reference sample NA10851 (Cy3-labelled; obtained from Coriell Cell Repository). After data normalisation, the ADM-1 algorithm (CGH Analytics 3.4, Agilent) was used for CNV discovery. The ADHD cohort was analyzed on Affymetrix 6.0
arrays. Three algorithms (Birdsuite, iPattern and Affymetrix Genotyping console (GTC)) were used to infer CNVs. The CEPH, PopGen and Ontario controls were analyzed on Affymetrix 6.0 arrays, SAGE controls were analyzed 1M BeadChips (Illumina) and Illumina 550K arrays were used for the CHOP and CAMH/GSK controls. Similar methods were used to infer CNVs in controls. The probe density of different microarray platforms at the *PTCHD1* locus is shown in Figure S 4-6. Fisher’s Exact Test was used to calculate the two-tailed *p* value.

4.5.3 DNA Sequencing and Mutation Screening

PCR primers were designed with Primer 3 (v. 0.3.0) to amplify all three exons and intron-exon boundaries (Table 4-1). PCR were performed under standard conditions, and products were purified and sequenced directly with the BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems).

**Table 4-1 Primers used to amplify all three exons of *PTCHD1***

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1F</td>
<td>AGA GCT CAG GGT CTC GCC</td>
</tr>
<tr>
<td>Exon 1R</td>
<td>CTA GGA GAG GTG GCG CTC T</td>
</tr>
<tr>
<td>Exon 2F</td>
<td>GAA TGT CCA CCC TCT CCA AA</td>
</tr>
<tr>
<td>Exon 2R</td>
<td>AAG GCT ACT CCT GGC CTT TT</td>
</tr>
<tr>
<td>Exon 3aF</td>
<td>CTT TGA CCC AGT AGT CCC TCA</td>
</tr>
<tr>
<td>Exon 3aR</td>
<td>GCA CAA ACC CCT TGG TGT A</td>
</tr>
<tr>
<td>Exon 3bF</td>
<td>TGT GAT TGG GTT TTA CAT ATA TGA GTC</td>
</tr>
<tr>
<td>Exon 3bR</td>
<td>AGG TCA GAT TTG AAG GCA CAG</td>
</tr>
<tr>
<td>Exon 3cF</td>
<td>AAA AAT GCC CTG GAA GTG C</td>
</tr>
<tr>
<td>Exon 3cR</td>
<td>TGT GTG AAT TCT CAT AAC AAC TCC T</td>
</tr>
</tbody>
</table>
4.5.4  X-Inactivation Studies

The X Chromosome Inactivation assay was performed on genomic DNA extracted from peripheral blood as described (Allen, Zoghbi, Moseley, Rosenblatt, & Belmont, 1992). Briefly, X Chromosome Inactivation was measured by the analysis of the (CAG)n repeat in the androgen receptor gene at Xq11-q12 before and after digestion with methylation sensitive restriction enzymes HhaI and HpaII. Quantitative PCR amplification of androgen receptor gene repeat alleles was compared, with and without restriction digestion, to determine the ratio of X-active/inactive alleles.

4.5.5  Expression Analysis and Protein Localization

Expression analysis and tissue distribution for PTCHD1, PTCHD1AS1 and PTCHD1AS2 was performed by RT-PCR, with a multiple tissue panel of first strand cDNA. The housekeeping gene G3PDH was used as a control. Origene human adult brain tissue panel was used to check the expression of PTCHD mRNA in different regions of the brain. qRT-PCR was performed with TaqMan Gene Expression assay Hs00288486, and samples were pre-normalized to GAPDH expression. Northern blot analysis was performed with a six tissue mRNA blot (BioChain). The BioChain FastHyb solution was used to hybridize the probe according to manufacturer’s instructions. RNA in situ hybridization was performed on paraffin sections and whole-mounted fetal mouse and adult mouse brain using a 411 bp (chrX:152,008,934-152,009,344, UCSC Mouse July, 2007 (2010c)) digoxigenin-labeled mouse antisense probe (and sense probe as negative control), using standard methods. To examine cellular localization of PTCHD1 protein, full-length human fetal brain PTCHD1 cDNA was PCR amplified and cloned into the pcDNA3.1/CT-GFP-TOPO expression vector (Invitrogen). After confirming the correct sequence and orientation of the insert, we transiently transfected COS-7 and SK-N-SH cells with 2 μg of purified construct DNA with SuperFect (Qiagen). 24 hours after transfection, we visualized the PTCHD1-GFP fusion protein in transfected cells using a Zeiss Axioplan 2 imaging microscope, equipped with the LSM510 array confocal laser scanning system, and the Zeiss LSM510 version 3.2 SP2 software package.
4.5.6 **Luciferase Assays**

A luciferase assay was performed to compare the effect of PTCH1, PTCH2 and *PTCHD1* on Gli-dependent transcription with a previously described method (Nieuwenhuis et al., 2006). Briefly, the 10T1/2 cells were transiently transfected with mixtures containing 0.1 µg β-galactosidase to normalize for transfection efficiency, 1 µg reporter plasmid (8xGlipro) encoding multimerized Gli binding sites fused to the luciferase gene and up to 1 µg of Gli2, PTCH1 or PTCH2 or *PTCHD1*. Gli-dependent transcription was measured and normalized by β-galactosidase. Data were replicated in independent experiments performed in triplicates. In another assay, 10T1/2 cells were transiently transfected with mixtures containing 0.1 µg β-galactosidase, 1 µg 8xGlipro reporter plasmid and puromorphamine, PTCH1 or PTCH2 or *PTCHD1*. The effect of PTCH1, PTCH2 and *PTCHD1* on the endogenous Gli-dependent transcription was measured. Statistical significance was calculated as *p* below 0.05, using the Student’s *t*-test.
Figure 4-1 Genomic organization of the PTCHD1 locus.

Detailed genomic organization of the PTCHD1 locus. The known genes, predicted CpG islands (>300 bp), predicted promoters (ElDorado Suite from Genomatix) and conserved sequences (>75% identity with chicken, >90% identity with opossum or 100% identity with dog or horse) are shown. Putative non-coding RNA transcripts PTCHD1AS1 (from cDNA clone IMAGE:1560626; BX115199) and PTCHD1AS2 (cDNA clone BRSTN2000219; DA355362) from human, mouse and rat genomes are also shown, with the transcripts assembled from RT-PCR and 5’ RACE (PTCHD1AS3) (see Supplementary Material). The dotted line between the two exons in transcript PTCHD1AS1 indicates that this is a putative exon, identified through clone sequencing. This exon is putative because, although this location represents its best genomic hit, it only partially matches the 5’ end of the clone sequence. Black boxes within the spliced transcripts indicate homologous exons between the sequences. White bars with black borders indicate CNV losses within this locus that have been identified in patients with ASD and controls. Cross-hatched or grey bars indicate CNV losses identified in patients with ADHD and ID, respectively. Colored lines within these bars indicate overlap with exons of known transcripts (blue) or ncRNA (red). The breakpoints of the deletions for all families that are reported here were mapped by sequencing the junction (see table S2 for coordinates). Breakpoints for all CNVs in controls were mapped by using the physical positions of microarray probe fragments.
Figure 4-2 Pedigrees of families showing segregation of *PTCHD1* mutations.

Pedigrees of families. (A) Pedigrees showing *PTCHD1* mutations. (B) Pedigrees showing deletions at the *PTCHD1/PTCHD1AS1-3* locus. The third male in Family 18 was assessed at age 4 and had speech and language problems, but was not available for further assessment. The father in Family 19 has a broader autism phenotype (BAP). The proband in Family 20 (hatched) has ADHD plus BAP. A diamond symbol represents siblings who were not tested as part of the study, and with gender not indicated.
A  PTCHD1 mutations

Family 1 (S0240)
- 167kb del (PTCHD1)

Family 2 (GOLD540)
- 90kb del (PTCHD1)

Family 3 (S01407)
- 1172V

Family 4 (S01433)
- ML336-71I

Family 5 (S01385)
- E419G

Family 6 (AU0501)
- L73F

Family 7 (GOLD243)
- AA10D

Family 8 (GOLD398)
- 2kb del (SLC10A2)

Family 9 (1-0215)
- N\(172V^*\)

Family 10 (3-0002)
- V196I

B  PTCHD1AS1/PTCHD1AS2/PTCHD1 locus

Family 11 (5298)
- 125kb del

Family 12 (5065)
- 65kb del

Family 13 (3424)
- 104kb del

Family 14 (5111)
- 59kb del

Family 15 (3253)
- 6kb del

Family 16 (13047)
- 36kb del

Family 17 (8273)
- 101kb del

Family 18 (8013)
- 66kb del

Family 19 (3387)
- 213kb del

Family 20 (1-27075)
- 36kb del

ID  ASD  ADHD/BAP  Proband  ? — Phenotype unclear  NA — Genotype/Phenotype not available
Figure 4-3 Expression analysis

Transcription analysis. (A) RT-PCR expression analysis of \textit{PTCHD1} transcript in 30 different adult tissues. The housekeeping gene \textit{G3PDH} was used as a control. (B) Northern hybridization analysis of \textit{PTCHD1} showing a $\sim$4.1Kb band in all lanes. Current RefSeq annotation of \textit{PTCHD1} describes a $\sim$5.3Kb transcript; however, the only polyadenylation site predicted for the mRNA sequence (NM_173495) by POLYA is at 4.379 bp. RT-PCR expression analysis of (C) \textit{PTCHD1AS1} and (D) \textit{PTCHD1AS2} expression in seven human tissues, also with \textit{G3PDH} as a control. Northern analysis of the ncRNAs did not give sufficient signal for detection.
Figure 4-4 Expression and functional studies of PTCHD1

Expression and functional studies. (A) Whole-mount RNA in situ hybridization showing expression of PTCHD1 in mouse embryo E9 and E14. (B) Localization of PTCHD1 protein in COS7, SK-N-SH and control cells shows that the PTCHD1-GFP protein is predominantly localized in the cell membrane. (C) PTCHD1 exerted a statistically significant inhibitory effect on endogenous Gli-dependent transcription, similar to PTCH1 and PTCH2, when transfected in Hedgehog-responsive 10T1/2 cells (PTCHD1: \( p = 0.0101 \); PTCH1: \( p = 0.0096 \); PTCH2: \( p = 0.0159 \)). Statistical significance was calculated using the Student’s t-test. Absolute expression of reporter gene normalized to β-gal expression is shown. Standard error bars are shown.
4.6 Supplementary Information

4.6.1 Cytogenetic and CNV analysis of proband from Family 9

Localization of translocation breakpoints was performed by fluorescence in situ hybridization (FISH; performed in accordance with standard procedures) initially using bacterial artificial chromosome (BAC) clones across the suspected breakpoint regions, and then narrowing the search using fosmid clones. BAC clones were obtained from the RP11 human genomic library, and fosmid clones from the Whitehead fosmid library WIBR2. For the chromosome 19 locus, the clone G248P85500F11 was translocated, and thus distal to the breakpoint, while clone G248P85559B4 was not translocated, and thus proximal to the breakpoint. The breakpoint therefore lies within a 32 Kb region between these two clones (UCSC March 2006: Chr19: 7,843,511-7,874,724. This region encompasses just two genes: FLJ22184, LRRC8E. At the chromosome 21 translocation site, fosmid clone G248P87249E2 was translocated, and G248P89542E9 was not translocated, and the breakpoint thus lies within a ~14.5 Kb region between these two clones, within an intron of the RUNX1 gene.

Whole-genome SNP analysis was performed using the Affymetrix 260K NspI SNP microarray. Analysis using the dCHIP and CNAG programs indicated a loss of heterozygosity from SNPs rs10875047 at Chr1:97,367,581 and rs822559 at Chr1:98,424,675 (inclusive; UCSC March 2006). This apparent deletion spans from intron 20 of the gene DPYD to include the first 20 DPYD exons, as well as two proximal putative genes, AK094607 and AX747691.

4.6.2 RT-PCR failed to find evidence for a shortened 3’ PTCHD1 transcript from individual with PTCHD1 exon 1 deletion

We speculated that the difference in phenotype between the two PTCHD1 deletion families (#1 and #2), could be explained by residual PTCHD1 protein function in relevant brain regions in Family 1 due to downstream transcription and translation of a shorter isoform, possibly driven by a secondary promoter just upstream of exon 2, resulting in the milder ASD symptoms, rather than the severer ID with the full deletion. However, RT-PCR did not detect any evidence of shorter downstream transcripts. Alternatively, genetic factors elsewhere may modify the effect of the deletion.
4.6.3 **Consensus Sequence for** _PTCHD1AS1_:  
TCTACACAAA CCAGATGAAC CTCTCAATCT CCTGCCCTCGA GTATTGAAGC  
CTGGCTACTG TGACTGAGGA GAAGGGATTAA ATGGTCTCAG CATTCAGCCA  
ACAACAATAC CTGCTCACTA TAAGCATTCA GAAAAACAGAA AAGTTTCAG  
AAGCAGGAAG AAAAGACTCA CCTATGATCC CAACACCCAG AGATAAGAGT  
CCTGAAGCTC AGATGACACA GCTGATAACA GGGAAGCCAG GACAGAATCT  
CATTGTCTTG AACACAAAA CCCGTCTCCCT TGACAACTTG GCTATACTAC  
ACTATTGGA TGTTGCAGAT ACTGTGGTCA CATTTCAGAG GCCAGATCTT  
TCCCAAGGTCT TAAGCTGTTC CTTGGATACT TTTGGTAAAGT CATTTATCCA  
CTAATCATTT AGTAATCGTC TCTGACATGC CAAACACCCT GTCAGGGCT  
GGAAATGCAG AACCTGGGAA GCCACTGGCC TTGTCTCTAA GATCTCTCTC  
TGGCTCCCTT TGAATTGTGCT ATTCAGACT TTCACATTTCC CCCAGAAAA  
AATCATAGG ACCAAATCAT ATCCGTTTTC TCAAATGGCT TCAAAGACCC  
ATGTCATCGT TTGGCATCAT GTAATTCTTT ACTGATGTAC TTTAAGAGTC  
ACGTATTTATT CTCTTTATGC AGCTGTCAAG GACAGACACA AAGAGGGGGG  
GGGNGGCTCT CCTACTAAAA TACTTTTCCCA ACAACA

4.6.4 **Consensus Sequence for** _PTCHD1AS2_:  
ACAACTGCAG CGAGAGAAGA GGCTGGCAGC ATGGGTGGCA GGAGGCTTGG  
CAGCCTCACA GGATGCTTCGAA AAATACCTTT CACTTATGCA GTTTGGCAGT  
AGCAGCTGTTGACATGGACAG CTGCTTGGGC CTGGCACCCCA CAGtCACTTA
RT-PCR and 5' RACE (Rapid Amplification of cDNA Ends) analysis of the ncRNAs, \textit{PTCHD1AS1} and \textit{PTCHD1AS2} and the \textit{PTCHD1} gene

By RT-PCR, the annotated exons of \textit{PTCHD1AS1} and \textit{PTCHD1AS2} were amplified from human cerebellum cDNA. Sequencing of RT-PCR product confirmed the current annotation of the ncRNAs. Additionally, we verified the annotation of \textit{PTCHD1AS1} by re-sequencing of the IMAGE clone 1560626.

We attempted to identify additional 5' sequence of the ncRNAs and \textit{PTCHD1} by 5' RACE analysis using the Clontech Marathon-Ready™ fetal brain cDNA (Cat. No. 639300). According to the manufacturer instructions the gene specific primers were designed for \textit{PTCHD1AS1}, \textit{PTCHD1AS2} and \textit{PTCHD1} and RT-PCR was performed. The PCR products were cloned into the Promega pGEM®-T Easy Vector and the clones were sequenced using standard methods. We were unable to find any additional upstream sequence for \textit{PTCHD1}. However, for the \textit{PTCHD1AS1} we found at least two additional exons. One of these exons completely overlaps
with the \textit{PTCHD1AS2} exon 2 (chrX:23,198,089-23,198,215), while the second exon mapped further upstream at chrX:23,261,313-23,261,767 (UCSC 2006). RT-PCR also identified another splice variant with an initial exon at ChrX:23,262,967-23,262,009, which skips to exon 2 in the current annotation of \textit{PTCHD1AS1}. It is possible that the extremely GC-rich nature of the 5’ region of \textit{PTCHD1} prevented us finding additional upstream sequence.

<table>
<thead>
<tr>
<th>NCRNA</th>
<th>Exon</th>
<th>Size (bp)</th>
<th>Coordinates</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{PTCHD1AS1}</td>
<td>1\textsuperscript{I}</td>
<td>126</td>
<td>chrX:23,198,089-23,198,214</td>
<td>This exon is alternatively spliced and completely overlaps with the exon 2 of the NCRNA355362.</td>
</tr>
<tr>
<td>\textit{PTCHD1AS1}</td>
<td>1\textsuperscript{II}</td>
<td>455</td>
<td>chrX:23,261,313-23,261,767</td>
<td>Starts 1.1Kb upstream of \textit{PTCHD1} and overlaps with the exon 1 of mouse transcript AK028243 and the \textit{PTCHD1} CpG island.</td>
</tr>
<tr>
<td>\textit{PTCHD1AS1}</td>
<td>1\textsuperscript{III}</td>
<td>43</td>
<td>chrX:23,261,967-23,262,009</td>
<td>Starts ~900 bp upstream of \textit{PTCHD1} and overlaps with the \textit{PTCHD1} CpG island. The transcript starting from this exon skips the Exon 1\textsuperscript{II}, 1\textsuperscript{I} and exon 1.</td>
</tr>
</tbody>
</table>
4.6.6. Alternative 5’ exons for \textit{PTCHD1AS1}, identified by 5’RACE:

4.6.6.1 Sequence of exon 1 I

\begin{verbatim}
CAATTGGTAGACATCTGGGTAGCTTCCACTTTTCCTGAACCAACTTTTAC
TGCAATTGTGACAGCTAGTTGTCCACGTCTGTGTTCCTCTCCAGGACT
CCACTTCTAAGTGCTGTGGGTGC
\end{verbatim}

4.6.6.2 Sequence of exon 1 II

\begin{verbatim}
ACCTGTGCGTGCCGCTTCCCGCCGCGCCGAGGTCTATCCCGGAGGCGA
AGCCGAGCGCCTTCGGGGAAATTCTCCGGAGGAGGAGTGCAGGAGGG
AACCACGGATTGCAGCTAGCTCAGGCTGGCGACACAGCACGACGC
CCACTTTGCCCAAGCAGCGTCATTGCTCTCTGAGATAAGGTTC
CCTCCACTCCTCACCTTTGACAGAGATCTATATCTCTGTAGTAGATTTC
GGCTCAACACGCTCGGAGACTGTATCTCCTCTGTCTGAGCTAAGGTTC
CCTCCACCTCCACTCCTAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGG
AGAGAGGAGGAGGTTATGCTACTGACTCCCTGGCCAGCCTTTCTCAAAC
TCTACGCCGAGGCCGCGAGCCGCCGCCGCGGCCGAGGACAGGCCACGAC
GACCTGCTGCGACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GAACC
\end{verbatim}
4.6.6.3 Sequence of exon 1\textsuperscript{III}

CTTTTGAGTGGACGTGCTCCAGACACACCCGGACCCCGTGG

4.6.7 Putative promoter and enhancer sequences in intergenic region between \textit{DDX53} and \textit{PTCHD1}

The identification of predicted promoter sequences may indicate the presence of an alternative upstream transcription start site for \textit{PTCHD1} (or possibly another unknown gene), that may be disrupted by the CNVs identified upstream of \textit{PTCHD1} in ASD families (see Figure 4-1). We have used the Genomatix ElDorado suite to predict promoter sequences. In addition to promoter sequences at the 5’ ends of \textit{DDX53} and \textit{PTCHD1}, on the plus strand we identified a putative promoter sequence in the intergenic region, from ChrX:22,927,508-22,928,108. This putative promoter lies ahead of ENSEMBL predicted non-coding transcript ENST00000407873. On the minus strand we identified a putative promoter sequence in the intergenic region, from ChrX:chrX:23,022,123-23,022,723, which lies just ahead of ENSEMBL predicted non-coding transcript ENST00000356867 and an EST clone (AU118198). We also used the ElDorado Suite from Genomatix, as well as the FPROM algorithm from the Softberry suite, which predicted promoter/enhancer sequences just upstream of the \textit{FAM3C2} predicted pseudogene.

Comparative sequence analysis indicates a number of regions located in the gene desert upstream of \textit{PTCHD1} and between \textit{DDX53} where nucleotide sequence conservation is relatively high through vertebrate evolution or through mammalian evolution. Such conserved regions may represent functional regions, possibly cis-regulatory sequences for \textit{PTCHD1}. Regions were selected through the Vertebrate Multiz Alignment & PhastCons Conservation (28 Species) track on the UCSC (March 2006 build) browser. Results are shown in Supplementary Table S 1, and indicate which conserved elements overlap with CNV losses upstream of \textit{PTCHD1}. 
4.6.8 eQTL at *PTCHD1* locus

The SNP rs7878766, located within *PTCHD1* intron 1, has been reported as a quantitative trait locus for expression of mRNA levels of *MAP8KIP2* in control brain cortex (http://eqtl.uchicago.edu), with a QTL score of 5.3. RefSeq Summary reports this to encode a scaffold protein involved in the c-Jun N-terminal kinase signaling pathway, and is thus thought to act as a regulator of signal transduction. Using mRNA by SNP Browser 1.0.1 (Dixon et al., 2007), other SNPs at the *PTCHD1* locus that show as suggestive QTLs for mRNAs include rs5925800 (*ACSM2A; LOD= 5.039, p=1.5 x 10^{-6}; GALNT4, LOD=5.095, p=1.3 X 10^{-6}; PIK3C2G, LOD= 5.27, p=8.4 x 10^{-7}), rs868659 (*DLEU2, LOD= 5.427, p=5.8 x 10^{-7}), and rs6526278 (*SGCG, LOD= 5.248, p=8.8 x 10^{-7}).
Table S 1 Clinical description of cases with disruptions at the *PTCHD1* locus on Xp22.11

<table>
<thead>
<tr>
<th>Family ID</th>
<th>Genes; Mutation</th>
<th>#Chromosomes Tested in Controls</th>
<th>Clinical Details in Proband‡</th>
<th>Family Segregation Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 1</td>
<td><em>PTCHD1</em>; <em>PTCHD1</em>AS2/3 167 Kb del</td>
<td>15,663 (M=4,829 F=10,834)</td>
<td>Proband (deletion) = Autism (based on ADI &amp; ADOS-Module 1) &amp; ADHD. Leiter-R brief IQ: 97 (42%); PLS-3: 86 (18%); VABS: COM=88 (21%); DLS=79 (8%), SOC=80 (9%), MOT=75 (5%), ABC=74 (4%).</td>
<td>Simplex family. Proband’s brother DZ twin (deletion) = ASD features and Learning Disability. WASI: Non-Verbal IQ=67 (1%), Verbal IQ=86 (18%); VABS: COM=84 (14%), DLS=95 (37%), SOC=104 (61%), ABC=92 (30%) Proband’s sister (heterozygous deletion) = non-ASD</td>
</tr>
<tr>
<td>Family 2</td>
<td><em>PTCHD1</em>; <em>PTCHD1</em>AS2/3 90Kb del</td>
<td>15,663 (M=4,829 F=10,834)</td>
<td>Proband (deletion) = non-ASD, moderate to severe ID.</td>
<td>Multiplex family. Proband’s brother (deletion) = moderate to severe ID. Proband’s maternal uncle (deletion) = moderate to severe ID</td>
</tr>
<tr>
<td>Study ID</td>
<td>Genotype</td>
<td>Age</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-----</td>
<td>----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>(S01407)</td>
<td>I173V</td>
<td>(M=613 F=488)</td>
<td>Non-Verbal IQ=95, Verbal IQ=85.</td>
<td></td>
</tr>
<tr>
<td>Family 4</td>
<td>PTCHD1</td>
<td>1193*</td>
<td>Proband (mutation) = Autism (based on ADI &amp; ADOS-Module 1). Some traits were observed that might be related to schizophrenia. Simplex family. No other siblings.</td>
<td></td>
</tr>
<tr>
<td>(S01433)</td>
<td>ML336-7II</td>
<td>(M=643 F=550)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family 5</td>
<td>PTCHD1</td>
<td>869</td>
<td>Proband (mutation) = High Functioning Autism Simplex family. Proband’s brother (no genotype data) = non-ASD</td>
<td></td>
</tr>
<tr>
<td>(S01355)</td>
<td>E479G</td>
<td>(M=531 F=338)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family 6</td>
<td>PTCHD1</td>
<td>869</td>
<td>Proband (mutation) = Autism Multiplex family. Proband’s brother #1 (no mutation) = ASD Proband’s brother #2 (mutation) = phenotype is currently unclear.</td>
<td></td>
</tr>
<tr>
<td>(AU0501)</td>
<td>L73F</td>
<td>(M=531 F=338)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family 7</td>
<td>PTCHD1</td>
<td>869</td>
<td>Proband=non-ASD; mild ID Clinodactyly, 3rd finger; camptodactyly. Proband’s siblings (no genotype data) = no ID.</td>
<td></td>
</tr>
<tr>
<td>(GOLD243)</td>
<td>A470D</td>
<td>(M=531 F=338)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Family 8 | PTCHD1 | 869 | Proband (mutation) = non-ASD, severe ID, seizures. | Multiplex family.  
Proband’s brother (no genotype data) = ID |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(GOLD398)</td>
<td>H359R and 2 Kb deletion spanning last exon of SLC16A2</td>
<td>(M=531 F=338)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Family 9 | PTCHD1 | 1101 | Proband (mutation) = Autism (based on ADI & ADOS-Module 1), intellectual disability, hyperactive, poor motor coordination. Leiter-R Brief IQ = 38. OWLS = 40 (<1%). VABS: COM=36 (<1%); DLS=<20 (<1%), SOC=31 (<1%), ABC=26 (<1%). | Simplex family.  
Proband’s sister (mutation) = non-ASD |
| (1-0215) | I173V and de novo ~1.1 Mb loss at DPYD | (M=613 F=488) |
| Family 10 | PTCHD1 | 1101 | Proband (mutation) = Autism (based on ADI & ADOS-Module 1). Severe expressive/receptive language delay. CT head=Normal. | Simplex family. No other siblings |
| (3-0002) | V195I and 66 Kb de novo loss at DPP6 | (M=613 F=488) |
Proband’s sister (heterozygous deletion) = non-ASD |
<p>| (5298) | 125 Kb del | (M=4,829 F=10,834) |
|-----------|---------------|-----------|--------------------------|--------------------------------------------------------------------------------|
|           |               |           |                          | Leiter IQ: 71 (3%). VABS: COM=68 (2%), DLS=45 (&lt;1%), SOC=58 (&lt;1%), ABC=52 (&lt;1%). |
|           |               |           |                          | Multilplex family. Paternal family history of ASD. Proband’s brother (no deletion) = Autism (based on ADI &amp; ADOS-Module 4). Verbally Fluent. VABS: COM=71 (3%), DLS=38 (&lt;1%), SOC=51 (&lt;1%), ABC=49 (&lt;1%). |
| Family 13 | 104 Kb del    | 15,663 (M=4,829 F=10,834) | Proband (deletion) = Autism (based on ADI &amp; ADOS). |
|           |               |           | WISC-R: Non-Verbal IQ=58, Verbal IQ=50, Total IQ=50 |
|           |               |           | Simplex family. Proband’s brother (no deletion) = non-ASD |
| Family 14 | <em>PTCHD1/AS1</em>  | 59 Kb del | 15,663 (M=4,829 F=10,834) | Proband (deletion) = Autism (based on ADI-R &amp; ADOS-Module 1). Uses single words. MRI = normal. |
|           |               |           |                          | Leiter IQ: 46 (&lt;1%). VABS: COM=37 (&lt;1%), DLS=31 (&lt;1%), SOC=52 (&lt;1%), ABC=37 (&lt;1%). |</p>
<table>
<thead>
<tr>
<th>Family</th>
<th>PTCHD1/AS1</th>
<th>PTCHD1/AS1-3, DDX53</th>
<th>PTCHD1/AS1-3, DDX53</th>
<th>(5) 8 (\text{Kb del})</th>
<th>(38.9) (\text{Kb del})</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>(3523)</td>
<td></td>
<td></td>
<td>15,663</td>
<td>(M=4,829, F=10,834)</td>
</tr>
<tr>
<td>16</td>
<td>(13047)</td>
<td></td>
<td></td>
<td>15,663</td>
<td>(M=4,829, F=10,834)</td>
</tr>
<tr>
<td>17</td>
<td>(8273)</td>
<td></td>
<td></td>
<td>15,663</td>
<td>(M=4,829, F=10,834)</td>
</tr>
</tbody>
</table>

Proband's father (no deletion) = non-ASD, OCD.

Proband's brother (no deletion) = ASD

Proband's sister (no deletion) = non-ASD.

Proband (deletion) = Autism (based on ADI & ADOS), Non-Verbal IQ=75, Verbal IQ=56

Multiplex family.

Proband (deletion) = Autism (based on ADI & ADOS), No epilepsy, history of language delay followed by a rapid language learning progression. Average to above average Non-Verbal and Verbal IQ.

Proband's brother #1 (no deletion) = Autism (based on ADI & ADOS), IQ=average to above average

Proband's brother #2 (no deletion) = ASD

Proband's sister (no CNV data) = non-ASD, semantic-pragmatic language disorder.

Proband's brother (deletion) = ASD

Proband's sister #1 (deletion) = ASD

Proband's brother (deletion) = ASD

Proband's sister #1 (deletion) = ASD
<table>
<thead>
<tr>
<th>Family 18</th>
<th><strong>PTCHD1/AS1</strong>&lt;br&gt;65 Kb del</th>
<th>15,663&lt;br&gt;(M=4,829 F=10,834)</th>
<th>Proband (no deletion) = Autism (based on ADI &amp; ADOS-Module 3).&lt;br&gt;WISC-III: Non-Verbal IQ=139 (&gt;99%), Verbal IQ=89 (23%).&lt;br&gt;VABS: SOC=76 (5%).</th>
<th>Multiplex family.&lt;br&gt;Proband’s brother #1(deletion) = Autism (based on ADI &amp; ADOS-Module 3). WISC III: Total IQ=44 (1%).&lt;br&gt;Proband’s brother #2 (deletion) = non-ASD.&lt;br&gt;WPPSI-R: Verbal IQ=89 (23%), non-verbal=100 (50%).</th>
<th>Proband’s sister #2 (deletion) = ASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 19</td>
<td><strong>PTCHD1/AS1-3,</strong>&lt;br&gt;<strong>DDX53</strong>&lt;br&gt;213 Kb del</td>
<td>15,663&lt;br&gt;(M=4,829 F=10,834)</td>
<td>Proband (no deletion) = ASD</td>
<td>Multiplex family.&lt;br&gt;Proband’s father (deletion) = Broad Autism Phenotype&lt;br&gt;Proband’s brother (no deletion) = ASD&lt;br&gt;Proband’s sister (deletion) = non-ASD.</td>
<td></td>
</tr>
<tr>
<td>Family 20</td>
<td><strong>PTCHD1/AS1-3,</strong>&lt;br&gt;<strong>DDX53</strong>&lt;br&gt;388 Kb del</td>
<td>15,663&lt;br&gt;(M=4,829 F=10,834)</td>
<td>Proband (deletion) = ADHD, NVLD&lt;br&gt;Verbal IQ =131, Performance IQ =113.&lt;br&gt;Proband has some ASD spectrum features (disinterest in</td>
<td>Simplex family.&lt;br&gt;Proband’s sister#1 (genotype unknown) = non-ASD</td>
<td></td>
</tr>
</tbody>
</table>
social relationships, preference for being alone, difficulty with change and over-adherence to structure and rules, difficulty with reading nonverbal cues resulting in social difficulties) but no evidence of restricted, repetitive, or stereotyped behaviour.

Proband’s sister#2 (genotype unknown) = non-ASD

§All probands are male and are of European ancestry except for those in family 9 (Mixed European), family 4 (East Asian), and families 6 and 7 (Not available). The referring diagnosis for all probands is Autism Spectrum Disorder (ASD) except for Families 2, 7, 8 (intellectual disability; ID) and Family 20 (ADHD)


†Standard Score 100 ±15 (percentile)

*Controls included N=92 of Asian ancestry
Table S 2 Breakpoint of deletions at the *PTCHDI* locus:

<table>
<thead>
<tr>
<th>Family</th>
<th>Breakpoints</th>
<th>Deletion size (bp)</th>
<th>Method used to map the breakpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 1(5240)</td>
<td>chrX:23,114,179-23,281,723</td>
<td>167,543</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 2 (GOLD540)</td>
<td>chrX:23,239,008-23,329,210</td>
<td>90,203</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 11 (5298)</td>
<td>chrX:22,890,415-23,015,667</td>
<td>125,253</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 12 (5065)</td>
<td>chrX:22,859,294-22,924,136</td>
<td>64,843</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 13 (3424)</td>
<td>chrX:23,011,719-23,116,212</td>
<td>104,494</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 14 (5111)</td>
<td>chrX:22,841,534-22,900,490</td>
<td>58,957</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 15 (3253)</td>
<td>chrX:22,853,977-22,908,345</td>
<td>54,367</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 16 (13047)</td>
<td>chrX:22,826,477-23,215,032</td>
<td>388,556</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 17 (8273)</td>
<td>chrX:22,989,332-23,091,080</td>
<td>101,749</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 18 (8013)</td>
<td>chrX:22,859,294-22,924,136</td>
<td>64,843</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 19 (3387)</td>
<td>chrX:22,824,496-23,037,508</td>
<td>213,013</td>
<td>Sequencing of junction fragment.</td>
</tr>
<tr>
<td>Family 20 (1-27075)</td>
<td>chrX: 22,678,814–23,066,819</td>
<td>388,006</td>
<td>Sequencing of junction fragment.</td>
</tr>
</tbody>
</table>
Table S 3 Additional CNVs in 9 subjects with upstream deletions

<table>
<thead>
<tr>
<th>Family</th>
<th>Gender</th>
<th>Inheritance</th>
<th>Physical Position</th>
<th>Size (bp)</th>
<th>CNV</th>
<th>Cytoband</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 1</td>
<td>M</td>
<td>Maternal</td>
<td>2:236932539_236990050</td>
<td>57,512</td>
<td>3</td>
<td>2q37.2</td>
<td>IQCA1</td>
</tr>
<tr>
<td>(5240)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5298)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16p13.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>Maternal</td>
<td>16:18153166_18699648</td>
<td>546,483</td>
<td>3</td>
<td>16p12.3</td>
<td>ABCC6P1, NOMO2, LOC339047,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RPS15A</td>
</tr>
<tr>
<td>Family 12</td>
<td>M</td>
<td>Maternal</td>
<td>1:17079505_17140083</td>
<td>60,579</td>
<td>1</td>
<td>1p36.13</td>
<td>CROCC</td>
</tr>
<tr>
<td>(5065)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family 13</td>
<td>M paternal</td>
<td>5:98798044_98836932</td>
<td>38,889</td>
<td>1</td>
<td>5q21.1</td>
<td>No gene.</td>
<td></td>
</tr>
<tr>
<td>Family 13 (3424)</td>
<td>M paternal</td>
<td>7:149089061_149159195</td>
<td>70,135</td>
<td>3</td>
<td>7q36.1</td>
<td>SSPO,ZNF467</td>
<td></td>
</tr>
<tr>
<td>M paternal</td>
<td>3:1719782_1786952</td>
<td>67,171</td>
<td>3</td>
<td>3p26.3</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>3:17494057_17542224</td>
<td>48,168</td>
<td>1</td>
<td>3p24.3</td>
<td>TBC1D5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>3:197219312_197527449</td>
<td>308,138</td>
<td>3</td>
<td>3q29</td>
<td>PCYT1A,TCTEX1D2,TFRC,ZDHHC19,OSTalpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>4:22488002_22620537</td>
<td>132,536</td>
<td>3</td>
<td>4p15.31</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>10:68138586_68227559</td>
<td>88,974</td>
<td>1</td>
<td>10q21.3</td>
<td>CTNNA3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M paternal</td>
<td>11:61516315_61632187</td>
<td>115,873</td>
<td>3</td>
<td>11q12.3</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>16:21506626_21647775</td>
<td>141,150</td>
<td>3</td>
<td>16p12.2</td>
<td>METTL9,IGSF6,OTOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M paternal</td>
<td>3:17494057_17542224</td>
<td>48,168</td>
<td>1</td>
<td>3p24.3</td>
<td>TBC1D5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>3:197219312_197527449</td>
<td>308,138</td>
<td>3</td>
<td>3q29</td>
<td>PCYT1A,TCTEX1D2,TFRC,ZDHHC19,OSTalpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>4:22488002_22620537</td>
<td>132,536</td>
<td>3</td>
<td>4p15.31</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>10:68138586_68227559</td>
<td>88,974</td>
<td>1</td>
<td>10q21.3</td>
<td>CTNNA3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M paternal</td>
<td>11:61516315_61632187</td>
<td>115,873</td>
<td>3</td>
<td>11q12.3</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>16:21506626_21647775</td>
<td>141,150</td>
<td>3</td>
<td>16p12.2</td>
<td>METTL9,IGSF6,OTOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M paternal</td>
<td>3:17494057_17542224</td>
<td>48,168</td>
<td>1</td>
<td>3p24.3</td>
<td>TBC1D5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>3:197219312_197527449</td>
<td>308,138</td>
<td>3</td>
<td>3q29</td>
<td>PCYT1A,TCTEX1D2,TFRC,ZDHHC19,OSTalpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>4:22488002_22620537</td>
<td>132,536</td>
<td>3</td>
<td>4p15.31</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>10:68138586_68227559</td>
<td>88,974</td>
<td>1</td>
<td>10q21.3</td>
<td>CTNNA3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M paternal</td>
<td>11:61516315_61632187</td>
<td>115,873</td>
<td>3</td>
<td>11q12.3</td>
<td>No gene.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M Maternal</td>
<td>16:21506626_21647775</td>
<td>141,150</td>
<td>3</td>
<td>16p12.2</td>
<td>METTL9,IGSF6,OTOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family 13 (3424)</td>
<td>M paternal</td>
<td>5:98798044_98836932</td>
<td>38,889</td>
<td>1</td>
<td>5q21.1</td>
<td>No gene.</td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>(ID)</td>
<td>Sex</td>
<td>Relationship</td>
<td>Chromosome</td>
<td>Length (bp)</td>
<td>Count</td>
<td>Chromosome Region</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-----</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>NA</td>
<td>7:109552072_109593909</td>
<td>41,838</td>
<td>1</td>
<td>7q31.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>NA</td>
<td>9:11936421_12032535</td>
<td>96,115</td>
<td>1</td>
<td>9p23</td>
</tr>
<tr>
<td>16</td>
<td>(13047)</td>
<td>M</td>
<td>Maternal</td>
<td>1:244036261_245191978</td>
<td>1,160,000</td>
<td>1</td>
<td>1q44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>Maternal</td>
<td>9:24652558_24705098</td>
<td>52,541</td>
<td>1</td>
<td>9p21.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>Maternal</td>
<td>18:67894269_67931021</td>
<td>36,753</td>
<td>1</td>
<td>18q22.3</td>
</tr>
<tr>
<td>17</td>
<td>(8273)</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8013)</td>
<td>Family 19 (3387)</td>
<td>NA</td>
<td>Family 20 (1-27075)</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>----</td>
<td>---------------------</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table S4 Gene co-expressed with *PTCHD1*

Genes co-expressed with *PTCHD1*, from gene Affymetrix gene expression microarray analysis from A. BioGPS (Gene Atlas U133A, gcrma; http://biogps.gnf.org); B. UCLA Gene Expression Tool (UGET: http://genome.ucla.edu/~jdong/GeneCorr.html; using human HG-U133_Plus_2 microarrays (Day, Carlson, Dong, O'Connor, & Nelson, 2007), and C. correlation with mouse *PTCHD1* using UGET and Mouse430_2 microarrays. These algorithms correlate expression based on banked Affymetrix gene microarray data, and is not tissue specific. Ranking counts multiple probes as single hits, and excludes hypothetical proteins.

<p>| A. BioGPS co-expression data for <em>PTCHD1</em> from Gene Atlas, U133A |
|---------------------|-----------------|---------|-------|</p>
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Correlation</th>
<th>Rank#</th>
<th>OMIM #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>PTCHD1</em></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZIC1</td>
<td>0.7564</td>
<td>2</td>
<td>600470</td>
<td>Zinc finger protein in cerebellum; homologue of Gli</td>
</tr>
<tr>
<td>GABRD</td>
<td>0.7064</td>
<td>12</td>
<td>137163</td>
<td>Receptor subunit (delta) for GABA neurotransmitter</td>
</tr>
<tr>
<td>MAB21L1</td>
<td>0.6916</td>
<td>17</td>
<td>601280</td>
<td>Autism susceptibility locus, AUTS3, candidate gene</td>
</tr>
<tr>
<td>Gene</td>
<td>Log2 Fold Change</td>
<td>Chr</td>
<td>Gene ID</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>------------------</td>
<td>-----</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CBLN1</td>
<td>0.6832</td>
<td>21</td>
<td>600432</td>
<td>Precerebellin 1</td>
</tr>
<tr>
<td>CADPS2</td>
<td>0.6827</td>
<td>22</td>
<td>609978</td>
<td>Cerebellar gene; involved in vesicular trafficking; autism candidate gene</td>
</tr>
<tr>
<td>CACNA1A</td>
<td>0.6801</td>
<td>23</td>
<td>601011</td>
<td>Gene for spinocerebellar ataxia 6</td>
</tr>
<tr>
<td>CALN1</td>
<td>0.6675</td>
<td>26</td>
<td>607176</td>
<td>Calneurin 1; cerebellar homologue of calmodulin</td>
</tr>
<tr>
<td>NRXN3</td>
<td>0.6041</td>
<td>42</td>
<td>600567</td>
<td>Neurexin 3; synaptic adhesion and presynaptic voltage-gated Ca(^{2+}) signalling</td>
</tr>
<tr>
<td>EN2</td>
<td>0.5799</td>
<td>50</td>
<td>131310</td>
<td>Engrailed 2; candidate gene at autism locus, AUTS10</td>
</tr>
<tr>
<td>SYT2</td>
<td>0.5782</td>
<td>51</td>
<td>600104</td>
<td>Synaptotagmin 2; synaptic vesicle associated protein, CA(^{2+}) sensor</td>
</tr>
<tr>
<td>GRM1</td>
<td>0.5747</td>
<td>52</td>
<td>604473</td>
<td>Metabotropic glutamate neurotransmitter receptor</td>
</tr>
<tr>
<td>GABRA6</td>
<td>0.5171</td>
<td>77</td>
<td>137143</td>
<td>Receptor subunit (alpha-6) for GABA neurotransmitter</td>
</tr>
<tr>
<td>SNAP25</td>
<td>0.5034</td>
<td>87</td>
<td>600322</td>
<td>Synaptosomal-associated protein</td>
</tr>
</tbody>
</table>

UGET co-expression data for *PTCHD1* from HG-U133_Plus_2 platform

<table>
<thead>
<tr>
<th>Gene</th>
<th>Log2 Fold Change</th>
<th>Chr</th>
<th>Gene ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTCHD1</td>
<td>0.85455</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SNAP25</strong></td>
<td>0.5389</td>
<td>7</td>
<td>600322</td>
<td>Synaptosomal-associated protein</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>----</td>
<td>--------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>CACNA1A</strong></td>
<td>0.52815</td>
<td>10</td>
<td>601011</td>
<td>Gene for spinocerebellar ataxia 6</td>
</tr>
<tr>
<td><strong>NRXN3</strong></td>
<td>0.514</td>
<td>13</td>
<td>600567</td>
<td>Neurexin 3; synaptic adhesion and presynaptic voltage-gated Ca(^{2+}) signalling</td>
</tr>
<tr>
<td><strong>GABRA6</strong></td>
<td>0.50935</td>
<td>15</td>
<td>137143</td>
<td>Receptor subunit (alpha-6) for GABA neurotransmitter</td>
</tr>
<tr>
<td><strong>GRM1</strong></td>
<td>0.50555</td>
<td>19</td>
<td>604473</td>
<td>Metabotropic glutamate neurotransmitter receptor</td>
</tr>
<tr>
<td><strong>GABRD</strong></td>
<td>0.4958</td>
<td>24</td>
<td>137163</td>
<td>Receptor subunit (delta) for GABA neurotransmitter</td>
</tr>
<tr>
<td><strong>KCNC1</strong></td>
<td>0.4935</td>
<td>25</td>
<td>176258</td>
<td>Voltage-gated K+ channel, Shaw-related, Kv3.1</td>
</tr>
<tr>
<td><strong>SYT4</strong></td>
<td>0.4934</td>
<td>26</td>
<td>600103</td>
<td>Synaptotagmin 4; synaptic vesicle associated protein, CA(^{2+}) sensor</td>
</tr>
<tr>
<td><strong>CBLN3</strong></td>
<td>0.4867</td>
<td>32</td>
<td>612978</td>
<td>Precerebellin 3</td>
</tr>
<tr>
<td><strong>DPP6</strong></td>
<td>0.4771</td>
<td>45</td>
<td>126141</td>
<td>Dipeptidyl peptidase 6: forms complex with Kv4.2 channels at synapse</td>
</tr>
<tr>
<td><strong>CADPS2</strong></td>
<td>0.4699</td>
<td>54</td>
<td>609978</td>
<td>Cerebellar gene; involved in vesicular trafficking; autism candidate gene</td>
</tr>
</tbody>
</table>

UGET co-expression data for mouse *PTCHD1* from Mouse430_2 platform
<table>
<thead>
<tr>
<th>Gene</th>
<th>Score</th>
<th>Rank</th>
<th>Gene ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTCHD1</td>
<td>0.7053</td>
<td>1</td>
<td>607567</td>
<td></td>
</tr>
<tr>
<td>Olfm3</td>
<td>0.4714</td>
<td>2</td>
<td>607567</td>
<td>Olfactomedin 3</td>
</tr>
<tr>
<td>Gria4</td>
<td>0.4397</td>
<td>3</td>
<td>138246</td>
<td>Glutamate receptor (AMPA); L-glutamate-gated ion channel</td>
</tr>
<tr>
<td>Pclo</td>
<td>0.4235</td>
<td>5</td>
<td>604918</td>
<td>Piccolo; presynaptic cytoskeletal matrix component</td>
</tr>
<tr>
<td>Dpp10</td>
<td>0.4165</td>
<td>9</td>
<td>608209</td>
<td>Dipeptidyl peptidase 10; forms complex with Kv4.2 channels at synapse</td>
</tr>
<tr>
<td>Cadps2</td>
<td>0.39</td>
<td>19</td>
<td>609978</td>
<td>Cerebellar gene; involved in vesicular trafficking; autism candidate gene</td>
</tr>
<tr>
<td>Nrxn3</td>
<td>0.3879</td>
<td>21</td>
<td>600567</td>
<td>Neurexin 3; synaptic adhesion and presynaptic voltage-gated Ca^{2+} signalling</td>
</tr>
<tr>
<td>En2</td>
<td>0.3816</td>
<td>30</td>
<td>131310</td>
<td>Engrailed 2; candidate gene at autism locus, AUTS10</td>
</tr>
</tbody>
</table>
Table S 5 Summary of Samples Analyzed in the Study:

<table>
<thead>
<tr>
<th>Samples</th>
<th>Gender</th>
<th>Ancestry</th>
<th>Platform used</th>
<th>Analysis</th>
<th>Major Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>427 ASD cases</td>
<td>M = 346 F = 81</td>
<td>&gt;90% have European ancestry</td>
<td>Affymetrix 500K SNP arrays</td>
<td>CNV analysis and sequencing</td>
<td>PTCHD1 exonic deletion and 3 missense mutations</td>
</tr>
<tr>
<td>420 ASD cases recruited at Montreal</td>
<td>M = 364 F = 56</td>
<td>Europeans (French Canadian)</td>
<td>Not applicable</td>
<td>Sequencing</td>
<td>3 missense mutations</td>
</tr>
<tr>
<td>996 autism probands from Autism Genome Project (AGP)</td>
<td>M = 839 F = 157</td>
<td>996 have European ancestry</td>
<td>Illumina 1M BeadChips</td>
<td>CNV analysis</td>
<td>Upstream PTCHD1 deletions</td>
</tr>
<tr>
<td>246 patients with ID</td>
<td>M = 246 F = 0</td>
<td>Not available</td>
<td>NimbleGen 385K array</td>
<td>CNV analysis and sequencing of 200 cases</td>
<td>PTCHD1 exonic deletion and 2 missense mutations</td>
</tr>
<tr>
<td>167 patients diagnosed with ADHD</td>
<td>M = 114 F = 53</td>
<td>88 % have European ancestry</td>
<td>Affymetrix 6.0 arrays</td>
<td>CNV Analysis</td>
<td>Upstream PTCHD1 deletion</td>
</tr>
<tr>
<td>1,123 PopGen controls</td>
<td>M = 623 F = 500</td>
<td>European</td>
<td>Affymetrix 6.0 arrays</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Study</td>
<td>Sex Distribution</td>
<td>Ancestry</td>
<td>Array Type</td>
<td>Analysis Type</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>1,234 Ontario controls</td>
<td>M = 586, F = 648</td>
<td>European</td>
<td>Affymetrix 6.0 arrays</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>59 CEPH controls</td>
<td>M = 30, F = 29</td>
<td>European</td>
<td>Affymetrix 6.0 arrays</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>1,320 CHOP</td>
<td>M = 589, F = 731</td>
<td>European</td>
<td>Illumina 550K arrays</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>440 CAMH\GSK controls</td>
<td>M = 158, F = 282</td>
<td>European</td>
<td>Illumina 550K arrays</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>1,287 SAGE</td>
<td>M = 383, F = 904</td>
<td>73% have European ancestry</td>
<td>Illumina 1M BeadChips</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>4,783 Welcome Trust controls</td>
<td>M = 2460, F = 2323</td>
<td>European</td>
<td>Affymetrix 6.0 arrays</td>
<td>CNV Analysis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>650</td>
<td>M = 650, &gt;95%</td>
<td>Not applicable</td>
<td>Sequencing</td>
<td>Not applicable</td>
<td></td>
</tr>
<tr>
<td>Ontario/CAMH Controls</td>
<td>F = 0</td>
<td>have European ancestry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Names and Affiliations of Autism Genome Project (AGP) Consortium Authors:


1 Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh 19104-6100, Pennsylvania, USA. 2 Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College Dublin 8, Ireland. 3 Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto. 4 School of Medicine and Medical Science University College, Dublin 4, Ireland. 5 Instituto Nacional de Saude Dr Ricardo Jorge and Instituto Gulbenkian de Ciencia Lisbon, Portugal. 6 Department of Neurology, University of California - Los Angeles School of Medicine, Los Angeles, California 90095, USA. 7 Hospital Pediatrico de Coimbra, Coimbra, Portugal. 8 Department of Biology, University of Bologna, 40126 Bologna, Italy. 9 Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford, OX3 7JX, UK. 10 Newcomen Centre, Guy's Hospital, London, SE1 9RT, UK. 11 Child and Adolescent Mental Health, University of Newcastle, Sir James Spence Institute, Newcastle upon Tyne, NE1 4LP, UK. 12 Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, 60528 Frankfurt, Germany. 13 Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London, SE5 8AF, UK. 14 Human Genetics and Cognitive Functions, Institut Pasteur; University Paris Diderot-Paris 7, Fondation FondaMental, 75015 Paris, France. 15 Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario, M5G 1Z8, Canada. 16 The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Ontario, M5G 1L7, Canada. 17 Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, Michigan, USA. 18 Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy
Hospital, 115 27 Athens, Greece.  

Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester, M13 9PT, UK.  

INSERM U952 and CNRS UMR 7224 and UPMC Univ Paris 06, UMR-S 952, Paris 75005, France.  

Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, North Carolina 27599-3366, USA.  

MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom.  

Social, Genetic and Developmental Psychiatry Centre, Institute Of Psychiatry, London, SE5 8AF, UK.  

Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri 63130, USA.  

Department of Pediatrics and Howard Hughes Medical Institute Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA.  

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio 43205, USA.  

Neuropsichiatria Infantile, Ospedale Santa Croce,61032 Fano, Italy.  

Child Study Centre, Yale University, New Haven, Connecticut 06520, USA.  

Department of Psychiatry, Carver College of Medicine, Iowa City, Iowa 52242, USA.  

Center for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, Tennessee 37232, USA.  

Departments of Biostatistics and Medicine, University of Washington, Seattle, Washington 98195, USA.

* Lead AGP investigators who contributed equally to this project.

X Deceased
Figure S 4-1 *PTCHD1* missense variants. Electropherograms indicate the nucleotide substitutions within *PTCHD1* in six unrelated ASD families and two ID families.
**Figure S 4-2 PTCHD1 domain structure**

*PTCHD1* domain structure and protein sequence conservation. The protein structure of the transmembrane protein *PTCHD1* is illustrated. (A) Twelve transmembrane domains (blue cylinders) and Patched-domain (red line) were identified using the SMART tool (http://smart.embl-heidelberg.de/) with the PFAM domain option selected. In addition, the locations of seven missense sequence variants discovered among ASD (black) and ID (blue) probands are shown. (B) CLUSTAL 2.0 alignments for *PTCHD1* showing position of missense mutations among ASD and ID probands. Amino acid positions given are relative to the human *PTCHD1* sequence (NP_775766). Other sequences used include mouse (NP_001087219), opossum (XP_001366520), platypus (XP_001512040), chicken (XP_425565), zebrafish (XP_690754), sea urchin (XP_001199849) and nematode (*C. elegans*) (NP_499380). (C) CLUSTAL 2.0 alignments for PTCH1, showing missense mutations reported for holoprosencephaly (Ming et al, 2002; Ribeiro et al, 2006), and including sequences from human PTCH1 (NP_000255), mouse (NP_032983), opossum (XP_001368370), chicken (NP_990291), *Xenopus laevis* (NP_001082082), zebrafish (XP_001922161), fruitfly (NP_523661) and nematode (*C. elegans*; NP_495662). In these organisms, the closest matching homologs were used, although this does not infer proof that these are the true ancestral forms.
Figure S 4-3 Quantitative RT-PCR for *PTCHD1* in human brain regions. *PTCHD1* expression in 24 regions of human adult brain is shown. Relatively higher expression was observed in the cerebellum.
Figure S 4-4 *PTCHD1* functional analysis. 10T1/2 cells were transiently transfected with γ-galactosidase to normalize for transfection efficiency, and Gli2, PTCH1, PTCH2 or *PTCHD1*. *PTCHD1* exerted a statistically significant inhibitory effect on Gli-dependent transcription, similar to PTCH1 and PTCH2 (** *PTCHD1*: $p = 0.0061$; PTCH1: $p = 0.0024$; PTCH2: $p = 0.0010$). Statistical significance ($p$ below 0.05) was calculated using the Student’s $t$-test. Standard error bars are shown.
Figure S 4-5 Comparative and phylogenetic analysis of human Patched-related proteins. Phylogram of Patched-related homologues: (Homo sapiens), created using CLUSTALW 2.0.12 (www.ebi.ac.uk), with N-J tree–type. The phylogram is assumed to be an estimate of a phylogeny, where branch lengths are proportional to the amount of inferred evolutionary change.
Figure S 4-6 SNP coverage at the *PTCHD1* locus across different genotyping platforms. The SNP\CNV probes on Affymetrix 500K, Affymetrix 6.0 and Illumina 1M arrays are shown.
Chapter 5. Future Directions

5.1 Screening of additional autism families for CNVs using improved microarray platforms

Since the start of this study the microarray platforms and CNV calling algorithms have improved immensely. The CNV data presented in this study was generated using the Affymetrix GeneChip® Human Mapping 500K Array set, which provides a coverage of ~500,000 SNP probes across the genome and ~154 Mb genomic region of Chromosome X is being covered by only ~10,000 probes. Therefore, the possibility of missing some important CNVs is very likely. Although, at the time of start of this study, this microarray platform was the best available, currently, the new microarrays offer substantially higher genomic coverage. For example, the new Affymetrix Genome-Wide Human SNP Array 6.0 offers 1.8 million genetic markers which include ~906,000 SNP markers, and ~ 946,000 probes for the detection of genome-wide CNVs. Recently, Illumina has introduced HumanOmni 5 microarray which will enable analysis of ~5.0 million markers across the genome. Furthermore, the CNV calling algorithms have also been improved over time, and several new commercial software packages are now available for the detection of copy number variants. Thus, additional autism families should be screened using improved microarrays and CNV calling software to uncover CNVs which may have been missed previously. Moreover, at this time, the coverage of clinical microarray platforms for genes identified in this study is very poor. For example, the mostly widely used microarray, Agilent 4x180K covers PTCHD1 locus with only four probes. Hence, any CNVs at this locus will be missed in samples processed in clinical laboratories. Therefore, the future version of clinical microarrays should have an improved coverage of this locus.

5.2 Analysis of splice variation and isoforms at the PTCHD1 locus

As a precursor to more detailed investigations into the function of PTCHD1, it is important to know whether the currently studied, 888 amino acid protein is the sole isoform, and if not then which isoform/isoforms are predominant in tissues relevant to autism and ID. The only commercially available antibodies raised against PTCHD1 (raised against an epitope located in a region encoded by exon 2) shows two bands upon western hybridization- one at ~100KDa,
corresponding closely to the anticipated size for the 888 amino acid protein, and another, much stronger signal at ~50KDa. Recent preliminary investigations undertaken in the Vincent lab suggest that in brain tissue a major splice variant exists in which exon 2 is spliced out. This variant is predicted to result in an mRNA size that correlates better with that shown by northern blot analysis. However, this isoform cannot explain the western data, but goes some way to demonstrate the importance investigate alternative splicing thoroughly. Use of alternative polyadenylation sites should also be investigated.

5.3 Induced pluripotent stem (iPS) studies

To better understand the biological role of candidate genes discovered in this study, iPS cells can be generated from the fibroblasts of patients and the carrier mothers, if the CNV is maternally inherited. These iPS cells can be reprogrammed to neurons and other cell lines to explore the effect of a CNV at cellular level. If a neurophysiological, or neuroanatomical phenotype can be determined in these iPS cells, we may be able to assay the efficiency of recovering the normal phenotype after introducing constructs carrying either wild-type PTCHD1 or carrying missense variants such as those identified as part of the studies reported in this thesis. In addition, these cells may be used to help further delineate the function or functions of PTCHD1, and these functions may also be amenable to assaying the effects of missense mutations in PTCHD1.

5.4 Zebrafish knockdown experiments

The generation of morpholino gene knockdowns in zebrafish has become a standard and rapid way to test the effects of a gene in a mammalian system. In such a system, an antisense molecule to the gene of interest is introduced into the developing fish embryos, which disrupts the gene expression, thus allowing the effects of the gene on early development to be assessed. In addition to the rapidity of such a system (in comparison to mouse knockouts, for instance), another advantage is that constructs carrying the wild type or mutant gene can then be introduced to see if they can recover the normal phenotype. Thus, this system could also be a powerful tool to assess whether missense variants, such as those identified in PTCHD1 or IL1RAPL1 in this study, are likely to be disease-related.
5.5 Mouse models

Another possible avenue for future research is the development and characterization of mouse models for autism candidate genes reported in this thesis. In particular, development of a murine knock-out (KO) of PTCHD1 will be of great interest. This approach will help us understand the pathogenic nature of these candidate genes and their role in neurodevelopment and neurophysiology, also it will help us learn if other systems are also affected. A future study may include performing detailed neurocognitive and behavioral testing of a PTCHD1 mouse model to determine to what degree Ptchd1 disruption contributes to the autistic and cognitive impairment seen in the human PTCHD1 –ve individuals. In particular, it would be interesting to see if PTCHD1 KO mice exhibit the deficits in social interactions and show behavioral and motor disturbances. Also, immunohistochemical studies of neurons of KO mice will be help to elucidate any morphological defects. Furthermore, neurophysiological and/or electrophysiological studies of WT and KO mice will be helpful in identification of any functional abnormalities of synapses.

5.6 Next Generation Sequencing of the entire chromosome X

The present study has further highlighted the importance of X-chromosomal genes in the ID and Autism. In fact, to date more than 95 chromosome X genes have been shown to cause some form of ID. Consequently, future studies may utilize the Next Generation Sequencing (NGS) technologies to search the sequence variants associated with autism. In recent years, NGS technologies have rapidly emerged. Deep sequencing of the entire X chromosome has great potential to uncover new autism candidate genes, and discover new biological pathways, important in the etiology of autism. In addition, NGS approaches will enable us to screen many more patients for coding mutations in PTCHD1.

5.7 Development of Potential Therapies

Another most important research direction may be the development of potential therapies for the candidate genes. Monogenic causes of autism are the most suitable targets for development of therapies, as the rescue of only a single protein would be required. Importantly, in this study, we have identified deletion mutations of PTCHD1 as a monogenic cause of autism and/or ID, albeit in a very small proportion of cases. In a larger number of cases we see missense variants in
PTCHD1 and CNVs upstream of PTCHD1 contributing to an oligogenic form of ASD. Hence, development of any potential therapies for this target only may have enormous medical implications. Nevertheless, development of therapies for other autism genes is also warranted. This goal can be achieved through several different approaches. For example, a mouse model can be generated by disrupting the gene of interest and then phenotype can be rescued by restoring the functional gene.
Chapter 6. References


UCSC Genome Browser: http://genome.ucsc.edu/.


Berditchevski, F. (2001). Complexes of tetraspanins with integrins: more than meets the eye. 

*Nat.Genet.*, 42, 489-491.


List of Appendices

Appendix 1: Manuscript entitled “Structural variation of chromosomes in autism spectrum disorder” *originally published in The American Journal of Human Genetics.*

Appendix 2: Manuscript entitled "Copy number variation analysis and sequencing of the X-linked mental retardation gene TSPAN7/TM4SF2 in patients with autism spectrum disorder" *published in Psychiatric Genetics.*

Appendix 3: Manuscript entitled "Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability" *published in Science: Translational Medicine.*