Autism Spectrum Disorders: Identification of Novel Microdeletions and Microduplications and their Associated Phenotypes

by

Noémie Riendeau

BSc. Biology, McGill University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES
(Medical Genetics)

The University Of British Columbia
(Vancouver)

October 2009

Abstract

Background: Autism Spectrum Disorders (ASDs) are common, heritable neurobiologic conditions of unknown etiology confounded by significant clinical and genetic heterogeneity. In recent years, array CGH technology has been used to rapidly screen the genome for pathogenic copy number variants (PCNVs) associated with ASDs and data from 6 studies suggests that PCNVs contribute to ASD pathogenesis in 6-27% of cases. However, the role of PCNVs in ASDs remains poorly understood, due to the absence of comprehensive phenotyping of ASD subjects.

Methods: To address this, we collected detailed clinical, medical, physical and morphologic information on all subjects and investigated whether these phenotypes would be good indicators of PCNV risk. We studied somatic phenotypes, as opposed to behavioural indices that change over time and with treatment, in an attempt to provide better evidence for the biological/embryological origin of ASDs and help define new ASD syndromes.

Results: Seven disease-specific and potentially pathogenic CNVs were uncovered in 6/40 patients (15%). Two changes were *de novo* and 5 were inherited from normal parents, but had never been reported in normal populations before. All PCNVs were discovered in individuals without family history of autism, ranged in sizes from 175kb to 2.5Mb, and revealed 9 good candidate genes. Our results suggest that whilst no single phenotypic feature investigated associates with PCNV risk, there is an indication that the presence of phenotypic abnormalities involving multiple body areas may be a better indicator of PCNVs in ASDs than the presence or number of minor physical anomalies alone. In addition, our findings lend support to the idea that complex autism, involving significant dysmorphology, is etiologically distinct from essential autism, with an

increased prevalence of ID, seizures and health problems, and a higher proportion of individuals without family history of ASDs.

Conclusion: We identified novel areas of chromosomal imbalance associated with ASDs and provide detailed phenotypic information for every subject for which these new PCNVs were detected. The extensive phenotyping of affected individuals carrying clinically relevant CNVs is needed in order to understand their role in the etiology of autism and ultimately provide earlier and more reliable means for ASD diagnosis and treatment.

Table of Contents

List of Tables

List of Figures

Glossary

ACRD Autism Chromosome Rearrangement Database

AD Autistic Disorder

ADHD Attention Deficit and Hyperactivity Disorder

ADM Autism Dysmorphology Measurement

AGP Autism Genome Consortium Project

array **CGH** array-based comparative genomic hybridization

AS Asperger's Syndrome

ASD Autism Spectrum Disorder

BAC bacterial artificial chromosome

BAP broader autism phenotype

BCNV benign CNV

BSID Bailey Scales of Infant Development

CDD Child Disintegrative Disorder

CNS central nervous system

CNV copy number variant

C-Section Cesarian section

CT computed tomography

DD Developmental delay

DGV Database of Genomic Variants

DMR differentially methylated region

DZ dizygotic

EEG electro-encephalogram

EMG electromyogram

FISH fluorescent *in situ* hybridization

FRAX Fragile X syndrome

GABA gamma-aminobutyric acid

GERD gastro-esophageal reflux

ID intellectual disability

IQ intellectual quotient

MCA major congenital anomalies

MPA minor physical anomalies

MPX multiplex

MPX-E multiplex-extended:>first degree relationship

MPX-I multiplex-immediate:first degree relationship

MRI magnetic resonance imaging

MZ monozygotic

NMDA N-methyl-D-aspartic acid

OC obstetric complication

OFC orbito-frontal circumference

PCNV pathogenic CNV

PDA patent ductus arteriosus

PDD Pervasive Developmental Disorder

PDD-NOS Pervasive Developmental Disorder-not-otherwise-specified

PNS peripheral nervous system

PWS Prader-Willi syndrome

QTL Quantitative Trait Loci

ROMA representational oligonucleotide microarray analysis

SD standard deviation

SMS Smith-Magenis syndrome

SNP single nucleotide polymorphism

SPX simplex

TSC Tuberous sclerosis complex

VABS Vineland Adaptive Behaviour Scale

UTR untranslated region

WBS Williams-Beuren syndrome

WIAT Wechsler Individual Achievement Test

WPPSI Wechsler Preschool and Primary Scale of Intelligence

Gene list

A2BP1 ataxin 2-binding protein 1

ABAT 4-aminobutyrate aminotransferase

ABL1 c-abl oncogene 1, receptor tyrosine kinase

BNDF brain-derived neurotrophic factor

CACNA1C calcium channel, voltage-dependent, L type, alpha 1C subunit

CDH9 cadherin 9

CDH10 cadherin 10

CNTN3 contactin 3 (plasmacytoma associated)

CNTNAP2 contactin(CTNT)-associated-protein 2

DACH1 dachshund homolog 1

DACH2 dachshund homolog 2

DIA1 deleted-in-autism 1

DOC2A double C2-like domains, alpha

EHMT1 euchromatic histone-lysine N-methyltransferase 1

EN2 engrailed homeobox 2

FMR1 Fragile X mental retardation 1

GABRB3 gamma amino butyric acid (GABA) A receptor beta 3

GRIK2 glutamate receptor ionotropic kainate 2 precursor

GRM7 glutamate receptor, metabotropic 7

IKKB IKK(beta)

ITGB3 integrin beta 3

KCNH7 potassium voltage-gated channel, subfamily H (eag-related), member 7

KCNK3 potassium channel, subfamily K, member 3

KCNK9 potassium channel, subfamily K, member 9

KLHL4 kelch-like 4

MAZ myc-associated zinc finger protein

MECP2 methyl CpG binding protein 2

MET met proto-oncogene

MVP major vault protein

NF1 neurofibromin 1

NFKB NF-kappaB

NGF nerve growth factor

NIBP NIK and IKK(beta) binding protein

NLGN3 neuroligin 3

NLGN4X neuroligin 4, X-linked

NR1 glutamate receptor, ionotropic, N-methyl D-aspartate 1

NRXN1 neurexin 1

OXTR oxytocin receptor

PCDH10 protocadherin 10

PEG13 paternally expressed gene 13

PRIP1 phospholipase C-like 1

PTEN phosphatase and tensin homolog

QPRT quinolinate phosphoribosyltransferase

RELN reelin

RNF8 ring finger protein 8

SCN7A sodium channel, voltage-gated, type VII, alpha

SEZ6 seizure-like 6

SEZ6L2 seizure-related 6 homolog (mouse)-like 2

SHANK3 SH3 and multiple ankyrin repeat domains 3

SLC6A4 solute carrier family 6 (neurotransmitter transporter, serotonin) member 4

SLC9A6 solute carrier family 9 (sodium/hydrogen exchanger), member 6

SLC9A9 solute carrier family 9 (sodium/hydrogen exchanger), member 9

SLC25A12 solute carrier family 25 (mitochondrial carrier, Aralar) member 12

SPN sialophorin

TOR1A torsinA

TOR1B torsinB

TNFA tumor necrosis factor alpha

TRAPPC9 trafficking protein particle complex 9

TSC1 hamartin

TSC2 tuberin

UBE3A ubiquitin protein ligase E3A

Acknowledgments

First of all I would like to thank Dr. Suzanne Lewis and Dr. Evica Rajcan-Separovic for their constant support, guidance and encouragement throughout all stages of this project. It has been a truly inspiring experience to work with such dedicated and passionate people, and the experience would not have been such an interesting and challenging one without their ideas, help and understanding.

I would also like to thank the genetic counselors, Jeanette Hildebrand and Jane Hurlburt and the members of the clinical team, Simone Dharmaratne, Levina Kasmara and Eleanor Chow, for helping me collect the clinical and phenotypic information on patients included in the study and for their complete dedication to the patients and their families.

I also want to thank the members of the Separovic Laboratory, Chansonette Harvard, Dr. Ying Qiao and Celina Fawcett, for training me and helping me with array experiments and FISH experiments and most importantly for their friendship.

I want to thank the members of my advisory committee, Dr. Wendy Robinson and Dr. Margot Van Allen, for sharing their ideas and always making themselves available despite their busy schedules. Finally I want to thank my boyfriend, whose help and support was necessary to write this thesis using a software that felt more like a foreign language at first.

Chapter 1

Introduction

1.1 Autism Spectrum Disorders

Autism Spectrum Disorders (ASDs) are a group of Pervasive Developmental Disorders (PDDs) characterized by significant qualitative changes in three core behavioural domains: deficits in reciprocal social interactions, impaired language and communication abilities and restricted repetitive and stereotyped patterns of behaviour, interests and activities [2]. ASDs encompass several clinically defined conditions (Autistic Disorder (AD), Asperger's Syndrome (AS), Pervasive Developmental Disorder-not-otherwise-specified (PDD-NOS)), which differ with regards to symptom severity, early development of language and cognitive abilities and social behaviours. Individuals affected by AD show impairments in all 3 domains with an abnormal development before the age of 3 years. In contrast, individuals suffering from AS show an apparently normal language and cognitive development before 3 years but are affected by deficits in social interaction and demonstrate restricted and stereotyped behaviours and interests. A diagnosis of atypical autism, or PDD-NOS is given to individuals who may show a late age of onset, or who show severe impairment in one or two domains, with or without language and cognitive delays. Two other, less common PDDs are listed in the current version of DSM-IV: Rett syndrome and Child Disintegrative Disorder (CDD).

ASDs are a significant cause of lifetime neuropsychiatric morbidity, affecting around 60/10000 individuals according to recent estimates [3, 4]. AD prevalence is estimated at 20/10000, while PDD-NOS is estimated at 30/10000 and AS is the lowest at 10/10000 [4]. Males are 4 times more likely to be diagnosed with autistic disorder, and this ratio goes down to about 2 males per one female in cases with severe intellectual disability (ID), and up to 6-8:1 in high-functioning individuals [3]. Empirical sibling recurrence risks of approximately 5% for all ASDs have been reported in the literature [5, 6].

ASDs are extremely variable in their phenotypic presentation, even within the narrow clinical definitions, and range in severity along many cognitive and behavioural dimensions [7, 8]. Because of this immense clinical and subclinical heterogeneity in phenotype manifested by individuals diagnosed with autism, the term ASD was coined, to emphasize the full scope of its signs and symptoms. ASDs are also phenotypically heterogeneous in that they can be described as *essential* or *complex*, differentiating simple from more syndromic forms of autism that may differ in outcome and etiology, whether environmental, epigenetic or genetic in nature.

Complex autism is associated with other clinical features such as facial dysmorphisms and/or other minor physical, systemic or growth anomalies, with or without ID [9]. This subtype of idiopathic autism accounts for 20% of cases, and individuals have poorer outcomes with lower intellectual quotients (IQs), more seizures, more abnormal electro-encephalograms (EEGs) (46% vs. 30%) and more brain abnormalities on magnetic resonance imaging (MRI) (28% vs. 13%). The remainder have *essential* autism, which is the more heritable group devoid of outwardly dysmorphic features, with higher sibling recurrence (4% vs. 0%), more relatives with autism (20% vs. 9%) and a higher male to female ratio (6.5:1 vs. 3.2:1) [9, 10]. Clinical genetics assessment can identify causes in up to 40% of cases, in association with new and known genomic disorders, single gene disorders or environmental insults [11, 12]. For the vast majority of cases presenting with essential non-syndromic autism (60-80%), a conclusive underlying etiology has yet to be identified, although a number of extremely rare single gene mutations (e.g. neuroligin 3 (*NLGN3*) and contactin(CTNT)-associated-protein 2 (*CNTNAP2*)) have just recently been associated with some cases of isolated autism [13].

Complex or essential, with a known etiology or not, autism still lacks any clear unifying pathology at the molecular, cellular or systemic level [8]. Nevertheless, progress has been made in recent years, with the great diversity of methodologies used to uncover genes involved in ASD susceptibility starting to provide us with clues about biological processes and pathways at play. In the next sections of the introduction, I will discuss previous research pertaining to the genetics of ASDs and their associated phenotypes. First, I will discuss the evidence for the involvement of genetic factors in the etiology of ASDs (Section 1.2). In Section 1.3, I will point out some medical co-morbidities and physical anomalies sometimes reported with ASDs and underscore the importance of phenotypic subgrouping and genotype-phenotype correlations in the study of genetic factors for autism. Following this, in Section 1.4, I will describe different methodologies used to identify ASD-related genes and use selected examples to illustrate recent research contributions in each area. The last part of the introduction (Section 1.5) will be reserved for the statement of the specific aims, goals and hypotheses of this work.

1.2 The Genetics of Autism Spectrum Disorders

Multiple epidemiological lines of evidence support the predominant role of genetic factors in the etiology of autism. First, three twin studies conducted in the last three decades indicate that concordance rates for monozygotic (MZ) twins are several fold higher than those for dizygotic (DZ) twins. Reported concordance rates for AD average 70% for MZ twins and 0% for DZ twins [14–16]. This observed rate of 0% is undoubtedly an artifact that results from the small number of pairs studied, which totalled only 66, with 36 MZ and 30 DZ pairs [17]. In fact, DZ pair concordance would be expected to be at least as high as sibling recurrence rates, around 5%. Interestingly, when concordance is defined to include behavioural autistic-like phenotypes, such as milder social deficits or language-related cognitive disorders, concordance rates can be as high as 92% for MZ and 30% for DZ [15, 16].

Family studies provide us with two additional lines of evidence for the involvement of genetic factors in the etiology of ASDs. First, the fact that the relative risk of a child of being diagnosed with ASD is increased many times above the population risk in families with one affected child further supports a strong genetic liability for autism [5, 6].

Second, taken as a whole, studies of non-autistic parents and siblings have clearly shown that they are more likely to show subtle cognitive or behavioral features reminiscent of those observed in probands [18, 19]. These qualitatively similar behavioural and cognitive phenotypes include social reticence, communication difficulties and preference for routine. These traits are conceptually the same as those used to diagnose autism, but are much milder and thus have come to be known as ”the broader autism phenotype” [17]. According to Bolton et al. 1994 [18], as many as 20.4% of siblings and 11.3% of parents exhibit this broader phenotype, defined as abnormalities in either of the three core domains. This familial aggregation of the broad phenotype is not associated with lower cognitive functioning in unaffected relatives [17].

Even though available family and twin studies convincingly point to ASDs as being some of the most heritable neuropsychiatric disorders, their underlying genetic etiology is undeniably complex, involving many genes and potential interactions with environmental factors. In some cases (up to 40%)[11, 20], ASD behaviours are a component of specific genetic or chromosomal disorders, and/or occur in association with specific and identifiable environmental factors. However in most cases, we do not know the cause of the ASD and despite national and international efforts, researchers have not been able to identify any definitive neuropathologic theory that resolves the biologic basis of ASD behaviours [13, 21–23]. Part of the problem is that the high level of genetic complexity is added onto the complexity and broadness of the diagnosis itself, which integrates complicated human behaviours that can change over the time and trajectory of a child’s development and in response to various therapies or treatments.

1.3 The Phenotypes of Autism Spectrum Disorders

In any genetic analysis, defining a valid and accurate phenotype is an essential condition if one wishes to identify genes and causal pathways involved. As noted above, there is compelling evidence that the autism behavioural phenotype is highly genetically determined, yet no matter how rigorous or clinically reproducible, the clinical diagnosis of autism only represents the ”end points” of extremely complex biological systems [24], and is not an etiologically based diagnosis. In other words, an ASD diagnosis is clinically useful but not necessarily genetically valid [25] and it has been proposed that progress

in identifying the underlying etiology and pathogenesis of autism has been inhibited by focusing on the behavioural phenotypes permitting its diagnosis [25].

One approach to improving diagnosis is to identify phenotypic variables (endophenotypes) that occur in a significant percentage of individuals with autism and that can be used to separate the ASDs into etiologically discrete disorders [26]. Endophenotypes are defined by Gottesman and Gould as "measurable components unseen by the unaided eye along the pathway between disease and distal genotype" [27]. A phenotype that is closer to the genetic origin of the disorder is expected to be a stronger etiologic indicator than a more distal symptom such as a behavioural trait. This is why the pursuit of etiologic clues that define ASDs physiologically are more likely to advance our understanding of the biological/embryologic pathways involved in the pathogenesis of autism, help us provide more accurate and timely diagnosis and improve our understanding of the natural history of the disorder and the provision of appropriate treatment strategies.

In this work, I will be investigating biologically-based phenotypes found consistently in ASD individuals for their association with deletions and duplications of genetic material found in our subjects. I will thus briefly introduce some of the work that has been done previously regarding phenomic predictors of ASD susceptibility and etiology.

1.3.1 ASD-associated Minor Physical Anomalies

Autism was first described by Leo Kanner who originally reported it as a disorder of children who were well formed and free of obvious defects [28]. However in the 1970s and 1980s, several studies reported that autistic children often had physical features outside the norm [29, 30]. Similar findings have been documented in other psychiatric disorders like schizophrenia and Attention Deficit and Hyperactivity Disorder (ADHD) [31].

Minor physical anomalies (MPAs) are mild errors of morphogenesis with a prenatal origin that have no serious medical or cosmetic significance to the individual [32]. Nevertheless, they are of great value to the clinician as indicators of disturbed development, as they are found to be more common in individuals with an obvious major birth defect [33]. An excess of MPAs in autistic children has received sustained attention through

the years, and recent studies are in keeping with older ones in that they consistently find an increased number of MPAs in individuals with ASDs [34, 35]. A recent meta-analysis showed higher MPA scores for children on the autism spectrum in all 7 studies examined [32]. Although the discovery of increased dysmorphology scores is robust across studies, findings are inconsistent when it comes to identifying specific anomalies that would distinguish autistic individuals from typically developing children or unaffected siblings [29, 34, 35].

The number of studies investigating major congenital anomalies (MCAs) is much lower, and results are sometimes contradictory, with most studies suggesting a higher prevalence of MCAs in autistic children compared to controls [36–38] with one study finding no significant difference [39]. Specific anomalies reported to be different between control and ASD populations are again inconsistent. The most recent of these studies found MCAs in 10.8% of ASD versus 6.2% of control individuals and this remained significant after adjustment for gestational age, plurality, birth order and maternal age, ethnicity and education [38].

In 2000, Miles et al. proposed that a subset of children with ASD can be identified with features suggestive of abnormal processes occurring during embryogenesis, and that this subgroup would be etiologically distinct [9]. They performed comprehensive clinical morphology examinations to classify children into phenotypically abnormal and phenotypically normal subgroups. Those children with 3 or less minor structural anomalies (e.g. single palmar crease, ear helix abnormality) or measurement anomalies (e.g.hypertelorism, macrocephaly) and 3 or less descriptive features (e.g.deep-set eyes, prominent forehead) were placed in the phenotypically normal subgroup. Those with 6 or more minor structural anomalies or measurement anomalies and 6 or more descriptive features were placed in the phenotypically abnormal subgroup while those in between were termed equivocal. Their initial comparisons suggested that the group with significant dysmorphology had a lower male-to-female ratio and lower IQs, although there was a significant overlap in cognitive abilities reported for each group.

In 2005, the same investigators showed that the group they called ”complex”, comprising individuals with significant dysmorphology and/or microcephaly, had lower recurrence rates, a higher number of cases without family history and a higher prevalence

of seizures and brain anomalies, in addition to the previously reported lower male-to-female ratio and lower intelligence [10]. All 11 individuals diagnosed with genetic syndromes (e.g. Tuberous sclerosis complex (TSC), Sotos syndrome, Fetal valproate syndrome) were in the complex group, suggesting that this subset of cases is very genetically heterogeneous.

Later in 2008, the same investigators developed a scoring algorithm they called the Autism Dysmorphology Measurement (ADM), that directs the clinician to score 12 body areas as normal or abnormal sequentially, to arrive at a determination of dysmorphic or non-dysmorphic [26]. To do this, they determined which dysmorphisms (standardized using the London Dysmophology Database (LDDB) codes) and which body areas would be most useful to distinguish between dysmorphic and non-dysmorphic individuals. Their results highlighted 20 representative feature codes (e.g. low-set ears) and 25 representative body areas (e.g. eyes) that were highly significantly different between the two groups.

In conclusion, autism research will benefit from the use of MPAs and MCAs as a tool to delineate more homogeneous subtypes within the core diagnosis of ASD [32]. The definition of such homogeneous subgroups will provide increased power for linkage and association studies and facilitate future investigations of the pathophysiology of the disorder [9, 35]. Subphenotyping might be the key to accurate prognosis, genetic counselling and the development of tailored treatment and prevention strategies [9].

1.3.2 ASD-associated Obstetric Complications

In this work I will also investigate whether obstetric complications (OCs) are good predictors of an underlying chromosomal imbalance in ASD subjects. This is why I briefly introduce here some of the previous work that has been done regarding OCs as etiologic indicators of ASD susceptibility.

Numerous studies have compared the incidence of prenatal, perinatal and neonatal complications in children with PDDs and their unaffected siblings or matched controls [36, 39–45]. Generally, children diagnosed with ASD show an increase in the number of OCs during pregnancy, labor and neonatal periods [44]. A number of these studies have

looked at the association of particular risk factors and complications, however no single complication emerges as a consistent risk factor, and findings are often contradictory. A recent systematic review of studies that used large population-based epidemiological samples found that 2 broadly defined OCs appear to be associated with a diagnosis of ASD [45]. The first category is small birth weight (<2500g), premature delivery (<35 weeks gestation) and small-for-gestational age which have been looked at in isolation or in combination. The other category contains several obstetric variables that may act as surrogates of fetal hypoxia, including low Apgar scores, fetal distress, C-section, threatened abortion and vaginal bleeding. Maternal hypertension, gestational diabetes, cord around the neck at birth and prolonged labor can also be associated with fetal hypoxia [45].

Alternatively, adverse events have been assessed using composite measures of obstetric optimality, with the argument that detection of liability is not dependent on specific items and that factors associated with autism might be non-specific [41]. Investigators who have evaluated OCs in that way have consistently found that affected children have lower optimality scores [43].

The evidence for a strong genetic component in the etiology of autism has to be reconciled with the evidence supporting an association between autism and obstetric complications [42]. According to Bolton et al. 1997, 3 main hypotheses could account for the apparent conflict. The first hypothesis, the "heterogeneity model", proposes that autism is heterogeneous and is sometimes caused by genetic abnormalities and sometimes by pregnancy and birth complications [15]. The second hypothesis, the "birth-order effect", argues that the association simply reflects the fact that autistic individuals are more likely to be born in first, or fourth and later birth order positions [41], for which there is a well established association with a higher frequency of OCs [46]. The third hypothesis is the "epiphenomenon model", which postulates that OCs are an epiphenomenon of ASD or familial factors that predispose to it [42], rather than playing a causative role.

Support for the "birth-order effect" hypothesis has come from a study showing that the association was no longer evident after adjustment for birth order [41]. However, the birth order effect in autism might come from decisions of parents against having

additional children after the birth of a child with special needs [47], such that the association may be of no etiological significance. Support against this second hypothesis comes from 2 studies that found that lower optimality scores in autistic individuals were not adequately explained by birth-order effects [42, 43].

Excluding the second hypothesis, the ”birth-order effect”, the first and third hypotheses predict opposing relationships between OCs and indices of genetic liability (specifically the amount of familial loading for the broader autism phenotype (BAP)). The ”heterogeneity model” (first hypothesis) forecasts that OCs will be greatest in individuals without family history of autism, whereas the ”epiphenomenon model” (third hypothesis) predicts that OCs will either occur regardless of familial recurrence of ASD or be increased as familial loading increases [42, 43]. Both Bolton et al. 1997 and Zwaigenbaum et al. 2002 found no evidence for the first hypothesis, in other words no evidence that minor obstetric complications are a significant cause of autism. Support for the ”epiphenomenon model”, in other words the notion that intrinsic disorder in the fetus can predispose to complications, is found in studies of Down syndrome [48] and Prader-Willi syndrome (PWS) [49].

Although the ”epiphenomenon model” has the most support, distinguishing whether the association stems from impairments in the fetal contribution to the maintenance of a normal pregnancy, or derives from some shared genetic risk factors is difficult [42]. Zwaigenbaum et al. 2002 attempted to resolve this question by looking at the relationship between OCs and familial loading for the BAP in unaffected siblings for simplex (SPX) and multiplex (MPX) pedigrees. Since they found lower optimality scores in unaffected siblings with higher family loading, regardless of SPX or MPX status, they argue that their results support the conclusion that OCs in ASDs are an epiphenomenon relating to familial genetic factors that lead to the expression of milder autistic traits in relatives.

1.4 Strategies to Uncover ASD Genes

The conceptual trend for more than a decade has been that of a polygenic or oligogenic mode of inheritance in most ASD cases, meaning that many common gene variants of

small effect at different loci are required to interact together to modulate phenotype and bring an individual above a certain behavioural diagnostic threshold [7]. The current view also acknowledges the involvement of single or a few rare gene variants, but with relatively high effect magnitudes in a higher proportion of cases than what was previously anticipated [8].

In their very nice review of the genetics of autism published in 2008, Abrahams and Geschwind [13] point out that rare and highly penetrant gene variants and common variants of small effects both play important roles in the etiology of ASDs and consolidate two contrasting but valid and compatible paradigms for the role of gene variants in ASD susceptibility. They further point out that since the relative proportion of ASDs explained by either rare mutations, or common variants (or both) is still unknown, multiple approaches are needed to advance our understanding of the genes and networks involved in ASD pathophysiology. These different methodologies are discussed in this section.

This section is not intended as a systematic review of findings from studies applying such methodologies, but rather serves to illustrate, through selected examples, how each strategy has contributed to our current knowledge of genes involved in the pathophysiology of autism. What becomes evident in reviewing the literature is that while approaches targeted at identifying common and rare variants should be pursued in parallel, it is also often necessary to integrate results from different methodologies to identify ASD culprit genes.

1.4.1 Linkage Studies

Numerous whole-genome scans and fine-mapping linkage studies, which search for linkage of autism to shared genetic markers in affected sibling pairs and MPX families, have been conducted over the years and the accumulated evidence has implicated at least one region on almost every chromosome (see [21] for review). Unfortunately, most studies have identified regions of suggestive linkage at best and only a handful have been independently replicated (reviewed in [13]).

This lack of reproducibility between studies is probably attributable to the genetically heterogeneous nature of autism, and the small magnitude of risk imparted by any individual gene variant. To provide adequate power to detect genes of small to moderate effect, a genome-wide linkage study would possibly need to investigate thousands of families. Even the largest genome-wide study to date, with 1168 families, only lead to the identification of one suggestive peak on chromosome 11p12p13[50]. It is even possible that by increasing the number of families, investigators are only "diluting" the signals even more, by increasing the genetic heterogeneity of the population under study [13].

One approach that seems to increase power is to study qualitative behavioural endophenotypes of autism instead of the categorical diagnosis of ASD. Building on findings from family studies and the heritability of the broader autism phenotype, some linkage studies have linked chromosome 15q11-q13 to insistence on sameness. Quantitative Trait Loci (QTL) mapping has also shown some success and it has the advantage of looking at the whole range of phenotypic variation. For example, linkage of the quantitative phenotype "age-at-first-word" was linked to chromosome 2q and chromosome 7q, and a "social responsiveness score" was linked to 11p13 (reviewed in [13]). And since many aspects of social behaviour, language and repetitive behaviours have been shown to be heritable, and continuously distributed in the general population, these studies could be extended to include non-autistic families [51, 52].

Whilst behavioural endophenotype studies indicate how phenotypic heterogeneity can be responsible for some of the discrepant results found from whole genome scans, they too are confounded by a high degree of variability that can blur their validity over the time and trajectory of a child's development, especially since they can change in response to various therapies or treatments. Instead, well-defined physical endophenotypes of ASDs should be used to subgroup subjects participating in genetic studies of autism, to increase the likelihood of discovering the associated susceptibility genes. In this work I will use this approach to subgroup subjects undergoing array-based comparative genomic hybridization (array CGH) studies (see Section 1.4.4), but linkage studies and association studies (see Section 1.4.2) would also benefit from such biologically-based endophenotyping of ASD subjects.

Although of great interest, the data from linkage studies usually implicate entire chromosome bands, if not entire chromosome arms, and thus the linked regions are always too large to directly incriminate a specific gene of interest, much less a specific susceptibility allele or causal variant. Additional studies, using different methodologies such as candidate gene analysis or mutation analysis are needed to identify ASD susceptibility genes responsible for linkage signals. The identification of one such gene, called *CNTNAP2*, was a direct result of replicated linkage in the 7q34-q36 region [53].

Even back in 2004, the region of 7q34-q36 was considered to be one of the strongest regions for linkage [22]. In 2008, common variation in *CNTNAP2* was associated with age-at-first-word, a quantitative ASD endophenotype [53]. In the same study, Alarcón et al. identified a deletion in the gene in one proband [53]. A rare *de novo* chromosomal inversion disrupting the gene, and point mutations predicted to be deleterious were also discovered in ASD probands around the same time [54], while another common variation was associated with the categorical ASD diagnosis [55].

In addition, a single rare recessive mutation in an Amish family had previously been shown to cause a syndrome involving focal epilepsy and neuronal migration deficits, as well as ASD in some two-thirds of affected children [56]. These multiple converging lines of evidence for the involvement of *CNTNAP2* in ASD might never have been discovered if the region had not been highlighted as being of great interest through linkage and QTL mapping studies.

1.4.2 Association Studies

Whereas linkage studies detect physical association between a phenotype and a certain polymorphic marker (e.g. SNP or microsatellite) within families, association studies detect associations between phenotypes and specific alleles of polymorphic markers in the general population [57]. In a linkage study, which polymorphic marker allele is segregating with the disease within pedigrees is not relevant. In one family, allele A might segregate with the disease while in another family allele B will be linked to the phenotype. In contrast, association studies are designed to detect when a specific allele

at a locus is associated more frequently with the phenotype under study than what would be expected by chance [57].

Association studies can either be genome-wide or geared towards identifying common variants in specific regions. Only one genome-wide association of autism has been published to date [58], and its results implicate neuronal cell-adhesion molecules, more specifically cadherin 9 (*CDH9*) and cadherin 10 (*CDH10*). Several dozen positional and functional candidate genes have also been examined for their potential association with ASDs[13]. All candidates investigated have yielded contradictory results, with some studies reporting associations and others not. Nonetheless, some candidate associations have been independently replicated, (see [13]) including integrin beta 3 (*ITGB3*) and solute carrier family 6 (neurotransmitter transporter, serotonin) member 4 (*SLC6A4*), two genes that interact together [59] and are located in a replicated linkage region on chromosome 17q11q21 [60]. Other replicated associations include glutamate receptor ionotropic kainate 2 precursor (*GRIK2*) at 6q21, solute carrier family 25 (mitochondrial carrier, Aralar) member 12 (*SLC25A12*) on 2q24, oxytocin receptor (*OXTR*) on 3p25 and gamma amino butyric acid (GABA) A receptor beta 3 (*GABRB3*) on 15q11q13, a region found to be duplicated in 1-3% of ASD subjects [23].

Of all chromosome regions implicated by linkage studies, only chromosome 7q has been linked to ASD at genome-wide significance in meta-analyses of ASD genome scans [61, 62]. Furthermore, several individual genome scans have reported positive findings on much of the long arm of chromosome 7 (from band q22 to band q36) [21], and several cytogenetic reports provide supportive evidence for the involvement of one or more genes on the long arm of chromosome 7, notably 7q22 and 7q31[22]. These lines of evidence have prompted association studies of more than a dozen genes located on 7q. Some of them have been independently replicated, including engrailed homeobox 2 (*EN2*), met proto-oncogene (*MET*) and reelin (*RELN*) [13]. Given the enormous degree of genetic heterogeneity in ASDs, it is not surprising that most genes investigated as candidates through associations have not been independently replicated in different populations.

Another complication in interpreting results from association studies is that in most cases where a positive associations has been found, the true functional variant remains

unidentified. Two exceptions to this rule are *MET* and *SLC6A4* [63–66]. Transcriptional activity of *SLC6A4* is modulated by a polymorphic repetitive element upstream of the transcription start site. This polymorphism has a long (L) and a short (S) allele, and the S allele is associated with reduced expression of *SLC6A4* [63] and increased volume of cerebral cortical grey matter [64]. The *SLC6A4* gene is a classic example of conflicting results from association studies. While some studies report positive association with the S allele (e.g. [67]), a roughly equal number of studies report positive association of the ASD phenotype with the L allele (e.g. [68]), while other studies have found no evidence supporting the involvement of either allele [69].

In the case of *MET*, a common allele of a single nucleotide polymorphism (SNP) in the promoter region of the gene was associated with ASD in one study, and found to result in a 2-fold decrease in transcription activity [65]. A follow-up study by the same group found reduced expression of *MET* in the cerebral cortex of cases versus controls [66]. Because *MET* is a pleiotropic receptor tyrosine-kinase that functions in both brain development and gastrointestinal repair [65], investigators recently assessed whether the functional promoter variant would be associated more strongly with phenotype in a subset of families with probands suffering both from autism and gastrointestinal conditions. They reported a positive association for families with co-occurring conditions, but the association disappeared in families where the co-occurrence was not present [70].

Dysregulation of gene expression in cases versus controls has been reported for other genes with replicated association findings. Examples include *SLC25A12*, which is upregulated in prefrontal cortex of autistic patients [71] and the oxytocin receptor (*OXTR*). Oxytocin is important in animal social behaviour, and is found at reduced level in plasma of ASD subjects when compared to controls [72].

These confusing results echo findings in other diseases with a complex genetic basis. Many thousands of cases might need to be analyzed in order to find variants conferring mild to moderate increases in susceptibility to ASDs, but association studies are without a doubt the method of choice for evaluating the role of a specific common variant in disease.

1.4.3 Genetic Syndromes and Cytogenetically Detectable Lesions Associated with ASDs

Considerable insight into the neurobiological basis of autism has been obtained by studying monogenic disorders in which ASD is observed as part of the phenotype at higher than expected frequencies. Mutations in the Fragile X mental retardation 1 (*FMR1*) gene at Xq27 that causes Fragile X syndrome (FRAX) is the single most common mutation in ASDs, with 3% [22] of individuals with ASD carrying an expansion of a CGG repeat in the 5'untranslated region (UTR) of the gene, resulting in hypermethylation and a reduced transcription of *FMR1* [73].

Another single-gene disorder associated with an increased risk of ASD is Tuberous sclerosis complex (TSC), which is caused by heterozygous mutations in hamartin (*TSC1*) on 9q34 or tuberin (*TSC2*) on 16p13 [74]. Between 25-50% of people suffering from TSC are also diagnosed with autism and 1-4% of people with ASDs carry a mutation in either gene [75]. Other monogenic syndromes associated with high rates of ASDs include Neurofibromatosis type 1 (neurofibromin 1 (*NF1*)), PTEN hamartoma syndrome (phosphatase and tensin homolog (*PTEN*)), Timothy's syndrome (calcium channel, voltage-dependent, L type, alpha 1C subunit (*CACNA1C*)) and Rett's syndrome (methyl CpG binding protein 2 (*MECP2*)) [76].

In addition to those syndromes caused by single-gene alterations, several genomic disorders are also associated with an autism phenotype, as well as numerous karyotypically detectable structural chromosomal aberrations, including some unique and rare recurring ones for which a defined syndrome has not been delineated. A recent review of 15 worldwide studies found that cytogenetic abnormalities are found in an average of 7.4% of ASD subjects [22]. Such abnormalities can lead to functional genetic changes in four ways [77]: 1) altered dosage of a genetic region, 2) the direct disruption of a gene at the breakpoint, 3) positional effects leading to dysregulation of a gene or genes and 4) unmasking of recessive mutations.

Cytogenetic studies in autism have implicated most chromosomes and both numerical and structural aberrations have been reported, including translocations, inversions, duplications, deletions, marker and ring chromosomes and even mosaicism [23].

Interestingly, apparently balanced translocations and classic chromosomal syndromes such as Trisomy 21 and Monosomy X have both been documented in around 1% of ASD cases, suggesting that certain chromosome aberrations may act as susceptibility factors for autism [22] and further highlighting the genetic heterogeneity of ASDs.

Maternally-derived duplications of 15q11-q13 is the most consistent chromosomal anomaly detected in ASD patients, occurring in 1-3% of cases, with interstitial duplications and isodicentric marker chromosomes observed in most cases [23]. Proximal 15q is a hotspot for chromosomal rearrangements, and deletions of this same imprinted region cause PWS or Angelman's syndrome depending on the parent-of-origin of the deletion [22]. This region is gene rich and harbours good biological candidates for ASD, including ubiquitin protein ligase E3A (*UBE3A*) and *GABRB3* which are currently thought to have a central role [13]. As discussed in Section 1.4.2, positive associations have been independently replicated for *GABRB3*. In addition, both genes show reduced expression in a high proportion of autistic brains according to one small study [78].

Another example of a genomic disorder associated with ASDs is the microdeletion 22q11.2 syndrome or DiGeorge\Velocardiofacial syndrome [79]. The clinical presentation can be quite variable between patients, with some individuals being essentially normal, while the most severe cases have life-threatening problems [80]. Based on a study of around 100 subjects with the deletions, in as many as 14%, the clinical presentation includes a diagnosis of ASD [79]. More recently it has been recognized that the reciprocal duplications of both Williams-Beuren syndrome (WBS)(del(7)(q11.23)) and Smith-Magenis syndrome (SMS)(del(17)(p11.2)) are associated with autism as well [81, 82].

Recurrent deletions in gene-rich subtelomeric regions have also been associated with autism. These include del(2)(q37), del(X)(p22.3) and del(22)(q13.3) [77]. Deletions of 2q37 have been observed in more than 70 cases [13] with breakpoints at 2q37.1, 2q37.2 and 2q37.3 [83]. Autism or autistic features are reportedly found in 17%(2q37.1), 50%(2q37.2) and 32%(2q37.3) of cases depending on the breakpoint [84]. In the case of del(2)(q37), the isolation of the contributory molecule(s) has remained elusive, while both discoveries of recurrent deletions of Xp22.3 and 22q13.3 have lead to the identification of causal point mutations in a small number of ASD cases [85, 86].

In the case of Xp22.3, the resequencing of neuroligin 4, X-linked (*NLGN4X*) identified a maternally inherited frameshift mutation in two affected brothers, but not in 350 controls, nor in an unaffected sibling [85]. In the same study, mutation analysis of another gene, *NLGN3*, located in a linkage region on Xq13q21, lead to the identification of a transition of a highly conserved residue (an amino acid highly conserved across distant species-suggesting functional importance) in two affected brothers in another family. The mutation was inherited from a normal mother, but was not found in 200 controls.

Neuroligins are cell-adhesion molecules localized post-synaptically at glutamatergic synapses [87]. Mutations in X-linked *NLGN3* and *NLGN4X* have been identified in individuals with AD, AS and ID [85, 88]. Although coding variants in these genes do not seem to be common in ASDs [89], they have provided crucial information on the synaptic abnormalities possibly present in autism. Functional studies of the *NLGN3* R451C and *NLGN4X* D396X mutations indicate defective trafficking and synapse induction properties of the mutated proteins [90] and mice harbouring this same Neuroligin 3 mutation exhibit increased inhibitory synaptic transmission [91].

Mutations in SH3 and multiple ankyrin repeat domains 3 (*SHANK3*), located within the 22q13.3 deletion, the product of which interacts with Neuroligins [92], were identified via sequencing as well and provide further evidence for the potential role of defective synaptogenesis in ASDs [86]. Mutations in *SHANK3* are believed to be the cause of autism in about 1% of cases [93]. Both these examples demonstrate the successful resequencing of candidate genes based on previous knowledge of chromosomal abnormalities associated with autism.

Although these rare variants lack a significant effect at the population level, they provide essential clues regarding biological pathways and genetic networks at play. The identification of rare mutations, as opposed to common variants, has the advantage of allowing clearer assessment of causality, and greatly facilitates the generation of mouse models and conduction of downstream functional studies that are necessary to unambiguously link genotype to phenotype. Molecular connections between different syndromes and rare variants are indeed starting to emerge, pointing to dysregulated protein translation [76] and aberrant activity-dependent changes in neuronal function [94].

1.4.4 Copy Number Variants and Autism

Chromosomal anomalies found by karyotyping or fluorescent *in situ* hybridization (FISH) offered the first glimpse at potential roles for rare variants in ASDs [77]. With the advent of whole-genome array CGH, it is now possible to detect DNA copy number changes at a much higher resolution than previously achieved by conventional cytogenetic approaches, and the analysis of the whole genome is done in one experiment. A copy number variant (CNV) is defined as a segment of DNA that is 1kb or larger and is present at a variable copy number in comparison with a reference genome [95]. Microarrays can detect such CNVs at varying resolutions, depending on the type, number and genome coverage of the segment of DNA (oligonucleotides or bacterial artificial chromosomes (BACs)) sampled on the array.

The detection of copy number variation below the threshold for traditional cytogenetics opened the door to the discovery of what are known as submicroscopic deletions and duplications, and several large CNV studies in autism (Table 1.1) have thus far identified hundreds of important candidate loci [96]. Most CNVs identified are unique (being detected in a single individual), suggesting that *de novo* and inherited CNVs might be important causes of ASDs, either as rare variants that strongly modulate risk, or as new syndromes linked to autism [13].

Genome-wide array CGH studies of autism published to date report the discovery of potentially pathogenic CNVs in 7.2 to 27.5% of individuals (Table 1.1). The higher detection rate of 27.5% is explained by the different patient selection method used, as this study by Jacquemont et al. 2006 investigated CNVs in individuals presenting with syndromic forms of autism that included significant physical dysmorphology [97]. Since most genome-wide array CGH studies made efforts to exclude individuals with known karyotypic or molecular abnormalities, their results provide a strong indication that the percentage of idiopathic cases attributable to rare structural variants is much higher than the 7.4% [22] identifiable by traditional cytogenetic methods, and could be increased to 15-20%, with 10% only detectable by arrays (Table 1.1).

There have been 6 large scale CNV studies of autism published to date [50, 97–101], with one additional study of 100 patients from our group to be published in the near

future [102]. A summary of findings from 6/7 of these can be found in Table 1.1. We did not include the study by the Autism Genome Consortium Project (AGP) [50] in the table because we felt the results were not directly comparable to those of other studies for the following reason. All studies but the one from the AGP compare the CNVs found in their autism cohort to those found in control populations to exclude any CNVs reported in such "normal" populations. They then use additional criteria to report on what they believe to be putative pathogenic CNVs (PCNVs). The study by the AGP reports their results differently, as they do not exclude CNVs reported previously in controls, and thus their numbers should not be compared directly to those of other studies.

The AGP study reported 254 CNVs in 196 ASD individuals from MPX families and divided them into four categories. The first category contained 10 *de novo* changes in 16/196 ASD individuals (8.2%). The second group was composed of 18 CNVs in unrelated individuals that overlap with previously published regions in ASDs (CNVs published in other genome-wide array CGH studies or the Autism Chromosome Rearrangement Database (ACRD)). The third group of CNVs were those 126 with overlapping or recurrent boundaries (overlapping CNVs are those where one boundary overlapped with another CNV in an unrelated individual while recurrent CNVs were those with identical boundaries in unrelated individuals). Finally the last category consisted of seven 15q11q13 gains from 3 different families. Taking this total as the number of putative PCNVs would give us a discovery rate of *de novo* changes in MPX families of 8.2% and a total discovery rate of 85%, both much higher than any other studies (see Table 1.1), probably because most of these have been reported in control populations. The authors of this study propose that potentially clinically relevant CNVs are those in 17 regions where CNVs (inherited or *de novo*) are found in two or more affected siblings of the same family. This gives us a discovery rate of 17 distinct regions from 173 families (10%).

It is important to note that none of the studies in Table 1.1 relied on the same combination of detection/confirmation methods and criteria to judge possible clinical relevance. Nevertheless, similar trends as discussed below can be seen in the data upon comparison, trends that could impact how future studies are planned and analyzed. One such interesting trend in the data was first reported by Sebat et al. 2007 [98] who found that the contribution of *de novo* CNVs to the etiology of ASD was much lower in MPX

families than SPX families. While the study reported an overall *de novo* discovery rate of 7.2%, this type of aberration was only found in 2.6% of MPX cases compared to 10% of SPX cases, a rate that is still notably higher than the 1% rate they reported in the control group. Another study has also reported a lower rate of *de novo* aberrations in multiplex families [99], while the numbers from Christian et al. 2008 [100] do not support this difference, however they studied very few simplex families. When inherited CNVs are not discarded as having no clinical relevance, as was the case in the earlier studies by Jacquemont et al. 2006 and Sebat et al. 2007 [97, 98], the discovery rate of autism-specific PCNVs is actually much more similar, between 9 and 12% for both multiplex and simplex families [99, 100].

The fact that such rare, autism-specific, and sometimes quite large CNVs can be inherited suggests that they have incomplete penetrance. This can be due to several factors including interactions with other genetic or epigenetic determinants, environmental agents or the unmasking of a recessive allele in the case of an inherited deletion [101]. Another possibility, and this is also true of some *de novo* changes as well, is that these aberrations are merely ”passengers”, chance findings of no phenotypic consequence. One argument against that would be the one raised by Sebat et al. 2007 [98], in light of their findings of a reduced male-to-female ratio in subjects with *de novo* changes. They proposed that a more equal contribution to disease across gender is evidence of increased penetrance. Such a reduced M:F ratio is the second trend that can be seen across the studies, as it is found in every single investigation, whether inherited changes are included or not.

One important point to emphasize about inherited CNVs is that their contribution as a source of more common genetic variation underlying ASDs has just recently gained research attention, with only one genome-wide association study of CNVs published so far [103]. In this recent paper, Glessner et al. 2009 reported on the association of certain rare but not unique, and mostly inherited CNVs with the autism phenotype, and their findings implicate both neuronal cell-adhesion molecules and protein ubiquitination (protein turnover) pathways [103]. These studies will be important because some more common heritable CNVs may have subtle phenotypic effects affecting cognition and behaviour [8]. Just as certain SNPs have been postulated to modulate risk in various candidate gene studies, and more recently in the first genome-wide SNP association

study in autism [58], so-called benign CNVs (BCNVs) (i.e. those seen in unaffected individuals) could also play a role in the etiology of autism.

Establishing the clinical relevance of detected CNVs in disease population is indeed complicated by our inability to fully discriminate between benign and pathogenic CNVs. In the last few years, hundreds of CNVs with no obvious phenotypic consequence have been reported in the human genome [104–107]. These CNVs are collected in the Database of Genomic Variants (http://projects.tcag.ca/variation/) and are usually referred to as "common" or "benign" copy number variants. Criteria that strongly suggest pathogenicity of a CNV in disease are the following: the CNV 1) associates with a known clinical syndrome; 2) has a *de novo* origin (or a maternally inherited X-linked origin in males); 3) contains multiple genes not known to show copy number variation according to current databases; 4) involves a gene for which a dosage effect is known; or 5) is $>$1 Mb and encompasses multiple well-defined genes, even when the inheritance of the CNV cannot be determined [108, 109].

Whether *de novo* or inherited, disease-specific or benign, each individual CNV needs multiple follow-up analyses to assess its role and to determine the subset of genes that could be causally related to autism. Such follow-up studies include analyses of very large samples (thousands, possibly tens of thousands, see discussion of del and dup(16)(p11.2) below) to evaluate the frequency of the CNV in the autistic populations. Other types of analyses include assessment of concordance between affected siblings and phenotype-genotype correlation studies [100]. For some of the rare, virtually unique deletions and duplications, even large sample sizes might not be sufficient to demonstrate significant associations, even though the biological significance might become clear through functional studies of individual genes and molecules [13].

While most of the CNVs in autistic individuals reported to date were unique, some regions of overlap between studies have been reported as well as overlap with regions previously identified through cytogenetic methods and some examples will be discussed here. In our laboratory, the discovery of two overlapping deletions of 2p15p16.1 in two unrelated individuals with similar clinical phenotypes led to the identification of a new microdeletion syndrome associated with ASDs [110]. Careful phenotypic examinations of both subjects suggested that this syndrome includes, but is not restricted to, moderate

to severe ID, autism/autistic features, microcephaly and a distinctive pattern of craniofacial features. One more case with a similar size deletion has been published so far [111], with a strikingly similar phenotype, except that the presence or the absence of autistic features was not noted or objectively studied. One other case with a much smaller aberration shared many craniofacial features, but again objective testing for autism was not reported [112]. This absence of specificity for autism is not surprising, since ASDs are a constellation of variable behavioural symptoms of an underlying disorder of genetic, genomic, epigenetic or environmental cause and not a single, biologically-based homogenous diagnostic entity.

One region that has been the subject of follow-up studies is 16p11.2, a region found to be deleted and duplicated in autistic subjects in multiple studies [98–100, 113, 114]. The deletion region spans 600kb and is flanked by segmental duplications that are >99% identical. This predisposes the region to unequal crossing-over during meiosis, and consequently to the microdeletion or microduplication of the unique sequence between the duplicons [115]. Two large scale studies looked at the frequency of the deletion and reciprocal duplication in autism compared to the general population.

The first study investigating 16p11.2 in ASDs [113] found the deletion in 4/712 (0.6%) of autistic individuals compared to 0/837 controls. All 4 cases were *de novo*, and in only one of these 4 MPX families, the deletion was found in multiple affected persons. Parental mosaicism probably explained the recurrence in the last family. None of the unaffected children carried the deletion. As for the duplication, it was found only in one proband, her unaffected mother and 2 controls. The mother was depressive and anxious, while the two controls reported several minor behavioural abnormalities such as compulsions, anxiety, phobias and panic attacks. No distinct behavioural subtype could be identified by studying the 4 deletions cases, only a trend towards aggression and hyperactivity.

The largest study investigating 16p11.2 [114] looked at 1441 subjects from MPX families, as well as 299 additional cases from Iceland and 512 children referred to the clinic for Developmental delay (DD), ID or suspected autism and different cohorts of controls totalling 20688 individuals. The deletion was found in 13/2252 cases (0.6%) and 2/20688 controls (0.01%). Ten were *de novo*, 2 were inherited (1 parent with mild

ID and 1 parent with ADHD) while 1 was of unknown origin. Again some affected siblings did not carry the deletion. As for the duplication, it was found in 11/2252 (0.49%) of subjects, but only 7/20688 (0.03%) of controls. There were 6 inherited duplications, 1 *de novo* and 4 of unknown origin. The inherited duplications were found in 6/6 affected individuals from 2 families, as well as one unaffected child, while the *de novo* duplication was found in 1/2 affected siblings.

Taken together, the 16p11.2 deletion and reciprocal duplication are present in approximately 1% of cases, but only 0.04% of controls, an increase in individuals with autism by a factor of 25. Since the same deletion was also found at a 0.1% rate in other psychiatric disorders in a large Icelandic population (bipolar disorder, ADHD, schizophrenia, panic disorder, anxiety, depression, addiction) and most controls are not systematically screened for all these conditions, the 25-fold increase is probably an underestimate [114]. As with other CNVs or mutations reported to be associated with ASDs, the 16p11.2 aberrations are not autism-specific. The same microdeletion has been reported in twins presenting with mild ID, seizure disorder and aortic valve anomalies [116]. Two individuals with ID out of approximately 100 have been found to carry the duplication in our laboratory (unpublished results), one paternally inherited and the other of unknown origin. Perhaps both del(16)(p11.2) and its reciprocal duplication are predisposing factors for ASDs, ID and other psychiatric disorders, just like deletions and duplications at 1q21.1 are predisposing factors for ID and ASDs [117] and deletions at 15q13.3 may predispose to ID, ASDs, schizophrenia and idiopathic generalized epilepsy [118].

In the past, the assumption was that a large, *de novo* chromosomal abnormality would be associated with disease, but the higher level of resolution provided by arrays increases the complexity of interpreting the significance of autism-specific CNVs [100]. Smaller, *de novo* submicroscopic deletions and duplications may have more subtle effects and incomplete penetrance, inherited changes even more so. Further studies with large numbers of subjects and controls are needed, along with detailed phenotypic evaluations and functional and expression studies of individual genes in order to fully understand what each reported CNV is contributing to the autism phenotype.

1.5 Specific Aims and Hypothesis

GOAL

- Identify genomic regions and hidden "culprit" genes, comprising the clinical and gene "signatures" of autism, using array CGH screening for ASD-related microdeletions and microduplications, coupled with extensive phenotypic characterization of the subjects.

OBJECTIVES

1. Screen 40 subjects with ASD using array CGH to find submicroscopic gains and losses of genomic material.
2. Confirm novel, disease-specific microdeletions and microduplications, determine the origin of the changes (inherited or *de novo*) and establish breakpoints as precisely as possible.
3. Investigate genes contained in the regions and the possible overlapping of confirmed, disease-specific CNVs with other reported cases in the literature in order to establish the clinical relevance of these changes and identify putative candidate genes.
4. Gather detailed phenotypic and medical data on all 40 patients, in order to identify predictive factors for the presence of clinically relevant deletions and duplications, and to provide detailed phenome-genome correlations for novel microdeletions and microduplications identified in our cohort.

HYPOTHESES

1. Deletions and duplications found in individuals with ASDs signal the locations of ASD-related culprit genes. Our own findings suggest that array CGH screening

will lead to the identification of at least 4 genomic regions to characterize from 40 subjects, based on an average 10% detection frequency [102].

2. The study of somatic phenotypes and their relationship with CNVs will identify biological/embryonic predictive factors for the presence of clinically relevant microdeletions and microduplications in idiopathic autism.

Table 1.1: Comparison of Genome-wide Copy Number Variation Studies in Autism

Study		Orig. Sample Cohort			Subjects with PCNVs					
Name (Ref.)	Array	Subjects	(%) Families MPX | SPX	M:F	Frequency in subjects(%) (total CNVs)	Inheritance	(%)PCNVs freq. in MPX | SPX (*de novo*)	M:F	Del:Dup	Size
Jacquemont ([97])	1Mb BAC	29 Syndromic	N/A	1.4:1	8/29(27.5) (8)	7 *de novo* 1 Unknown[a]	N/A	1:1	3:1	1.4-16Mb
Sebat ([98])	85K ROMA	195	39 | 61	5:1	14/195(7.2) (15)	15 *de novo*[b]	(2.6 | 10)	1.8:1	4:1	99kb-12Mb
Christian ([100])	19K Tiling BAC	397	91 | 9	1.4:1	46/397(11.6) (51)	9 *de novo* 42 inherited	11.9 | 8.6 (3.0 | 2.8)	1.2:1	0.4:1	190kb-6.1Mb
Marshall ([99])	500K SNP	395	46 | 54	N/A	34/395[c](8.6) (37)	14 *de novo* 16 inherited 7 Unknown	7.7 | 9.4 (2 | 7)	2.4:1	0.7:1	16kb-18.2Mb
Cusco ([101])	Homemade BAC[d]	96	N/A	7:1	12/96(12.5) (13)	5 inherited 8 Unknown	N/A	5:1	3.3:1	90kb-2.4Mb
Qiao ([102])	1Mb BAC	100 Complex[e]	SPX:31 MPX-I:45 MPX-E:24	3.2:1	12/100(12) (12)	7 *de novo* 3 inherited 2 Unknown	SPX:16 (6.5) MPX-I:8.9 (2.2) MPX-E:12.5 (12.5)	2:1	3:1	200kb-10Mb

[a] 2 inherited CNVs discarded

[b] inherited CNVs ignored

[c] cases with CNVs detected by karyotype have been excluded from frequency calculation

[d] Array contains 5442 clones, with higher probe density in regions presumed to be hot spots for genomic rearrangements

[e] With phenotype score ≥ 3 (see Section 2.1 and Table 2.1)

Chapter 2

Materials and Methods

Note: All families who participated in this study did so voluntarily. Consent forms approved by the Clinical Research Board of the University of British Columbia and the BC Child and Family Research Institute was signed by each participating family member. For probands under the age of 7, a consent form were signed by the parents or guardians on behalf of the proband. In all cases where a recognizable etiology was recognized, genetic counselling for the individual and their family was provided.

2.1 Patient Selection

In the course of systematic medical genetic evaluation of subjects with an ASD recruited through the research registry of the Autism Spectrum Disorders-Canadian American Research Consortium (ASD-CARC; www.AutismResearch.com), 40 subjects with idiopathic ASD (34 males and 6 females) were selected for array CGH analysis. Autism diagnoses for all subjects were based on standardized DSM-IV-TR criteria using ADI-R and/or ADOS-G measures [119, 120].

Among the 40 subjects studied, 24 were from simplex families, 10 from multiplex-immediate:first degree relationship (MPX-I) and 6 from multiplex-extended:>first degree relationship (MPX-E) families. All subjects had normal karyotypes, including tar-

geted 22q11/22q13 and 15q11-q13 FISH, subtelomeric FISH studies, negative Fragile X and clinical chemistry screening (serum lactate, ammonia, creatine phosphokinase, lead, complete blood cell count and microscopy, uric acid, TSH, urine purine/pyrimidine and creatine metabolites).

Subjects were selected non-randomly to include at least 10 individuals with essential autism, with phenotype scores $\leq$3 excluding points for family history (see Table 2.1). The 40 subjects studied included 14 individuals with essential autism and 26 individuals with complex autism (with phenotype scores $\geq$4), with 6 females (3 complex, 3 essential) and 34 males (23 complex and 11 essential). All subjects had comprehensive, systematic evaluation of multi-generation family pedigrees, prenatal, medical and developmental histories and complete physical and dysmorphology examinations by one medical geneticist (Dr. MES Lewis), incorporating anthropometric craniofacial assessments. A number of obstetric complications were recorded for further analysis, including:

1. pregnancy complication: poly- or oligohydramnios, fetal abnormalities, growth retardation detected via ultrasound examination, vaginal bleeding, decreased fetal movements, maternal hypertension, eclampsia , gestational diabetes

2. labor complications: fetal distress/variable heart rate during labor, cord around neck (or knots), induction/failure to progress, premature labor, prematurity, vacuum extraction, blood incompatibility, exchange transfusion, cephalo-pelvic disproportion, breech or transverse position

3. postnatal complications: respiratory distress, feeding difficulties (poor suck, poor latch), failure to thrive, seizures, resuscitation, hypo- hypercalcemia, hypo- hyperglycemia

A five item checklist developed by de Vries et al. 2001 [1] was used to phenotypically categorize and score each proband for further analysis (Table 2.1). I adapted the checklist to this study population by incorporating family history of ASDs into the checklist.

Table 2.1: Five Item Checklist Adapted from de Vries et al. 2001 [1]

Items	Score
Family history of ASDs or ID	
Compatible with Mendelian Inheritance	1
Incompatible with Mendelian Inheritance	2
Prenatal onset of growth retardation	2
Postnatal growth abnormalities	
For each of the following score 1 point (maximum=2)	2
Macrocephaly or Microcephaly (orbito-frontal circumference (OFC)$\geq$98%ile or $\leq$2%ile)	
Tall stature or Small stature(Height, Weight and OFC $\geq$97%ile or $\leq$3%ile)	
$\geq$2 Craniofacial dysmorphic features	2
(hypertelorism, nose anomalies, ear anomalies, etc.)	
$\geq$1 Non-Craniofacial dysmorphisms and Congenital anomalies	2
For each anomaly score 1 point (maximum=2)	
(hand and feet anomalies, heart anomaly, etc.)	

2.2 DNA Isolation, Quantitation and Quality Control

Blood was collected from affected proband(s) and non-affected family members for the purpose of DNA isolation. Blood was also collected from probands for cytogenetic studies when there was no previous cell pellet available from the cytogenetic laboratories at Children's and Women's hospital in Vancouver or Royal Columbian hospital in Burnaby, and the remainder from clinical tests done at those centers kept for confirmatory FISH studies.

1. Each sample was first assigned a unique laboratory identification number. Blood was transferred from EDTA tubes to properly labeled 50 ml conical tubes and DNA extracted using the Puregene (Gaithersburg, MD, USA) DNA Isolation Kit using the manufacturer's protocol.

2. To determine the concentration of DNA in solution, each sample was analyzed using a spectrophotometer (Pharmacia Biotech Ultraspec 3000). 2.5µl of DNA was diluted in 97.5 µl of ddH2O. One at a time, samples were loaded into the holding well of the spectrophotometer and three readings were obtained; absorbance at 260nm, absorbance at 280nm and ratio of the two readings.

3. To obtain the concentration of the sample, the following formula was used:

   Formula: Abs260*40*50

   (Where Abs260 is the absorbance by the sample at 260nm, 40 is the dilution factor and 50 is the DNA constant)

4. To ensure that it was not degraded, DNA was analyzed by electrophoresis on 1% agarose gel. All gels were observed under UV light and an image was captured and printed.

2.3 Array CGH

2.3.1 Definition of Pathogenic and Benign CNVs

One database cataloguing putatively benign CNVs (Database of Genomic Variants (DGV), http://projects.tcag.ca/variation/) was used to assess whether array-detected CNVs had been reported in normal populations. CNVs reported in at least two independent studies of healthy control subjects are typically referred to as BCNVs and are not investigated further, while other CNVs are referred to as disease-specific CNVs. Those CNVs that fulfill criteria strongly suggesting their pathogenicity in individuals affected by a specific disorder, either 1) through association with a known clinical syndrome; 2) having a *de novo* origin (or a maternally inherited X-linked origin in males); 3) involving multiple genes not known to show copy number variation according to current databases; 4) involving a gene for which a dosage effect is known; or 5) were $>$1 Mb and encompass multiple well-defined genes, even when the inheritance of the CNV cannot be determined, are referred to as PCNVs [108, 109]. Databases cataloguing PCNVs like DECIPHER (https://decipher.sanger.ac.uk/) and the ACRD (http://projects.tcag.ca/autism/), and available literature were also used to interpret the clinical significance of changes.

2.3.2 Spectral Genomics 1Mb Array Protocol

The array CGH platforms and reagents were purchased from a commercial company (Spectral Genomics, Houston, TX, USA) and the manufacturer's protocol was followed. A pool of normal male or female control DNA (Promega, Madison, WI, USA) was used as reference DNA to match the sex of the samples studied.

1. **DNA Sonication**

   To ensure efficient random prime labeling, each DNA sample was sonicated to produce smaller fragments:

   (1) $X\mu$l for 2μg of test DNA was added to $(200-X)\mu$l of ddH2O; same was done for sex matched reference DNA. (where X= the volume of DNA)

   (2) Samples were then sonicated for 9 seconds.

2. **DNA Purification**

   Following sonication, samples were cleaned in order to eliminate traces of protein, RNA and other impurities using DNA Clean & Concentrator TM from Zymo Research (Orange, CA, USA). DNA cleaning involved the following procedure using the manufacturer's protocol:

   (1) 2 volumes of DNA binding buffer was added to each volume of DNA sample (400μl for 200μl of sample).

   (2) Samples were loaded into a spin column and columns were placed into a 2ml collection tube.

   (3) Columns were centrifuged at 13 000 rpm for 10 seconds to draw the excess fluid into the collection tube.

   (4) 200μl of DNA wash buffer was added to the columns. Columns were centrifuged at 13 000 rpm for 10 seconds. This process was repeated twice.

   (5) 27.5μl of water was added to the columns. Columns were placed in 1.5ml tubes and centrifuged for 30 seconds to elute the DNA. This process was also repeated twice, to increase the recovery of DNA from the column matrix. The final volume after elution was 55μl.

(6) Following purification, 5µl of each DNA sample was analyzed by agarose gel electrophoresis to ensure adequate sonication. Optimal sonication produced homogeneous smears extending 600bp to approximately 2kb.

3. **Labeling**

(1) Each sample was divided into two tubes of 25µl each. Tubes were labeled Cy3 and Cy5 respectively (at this stage, for each patient there were 4 tubes: Cy3 Reference, Cy3 Test, Cy5 Reference and Cy5 Test).

(2) Master Mixes for each dye were prepared and kept on ice according to the following composition for each patient:

- 12.5µl Labeling buffer (Spectral Genomics, Houston, TX, USA)
- 7.5µl Cyanine 3-dCTP / Cyanine 5-dCTP fluorescent dye (Perkin -Elmer, Turku, Finland)
- 5.0µl Klenow enzyme (Invitrogen BioPrime labeling kit) (Invitrogen Life Technologies, Carlsbad, CA, USA)

(3) 20µl of 2.5x random prime mix (Invitrogen's Bio Prime labeling kit) was added to each tube. Tubes were vortexed and centrifuged briefly. DNA was denatured into single strands by heating each sample for 5 minutes at 100°C. Following denaturation, samples were cooled in an ice water slurry for 5 minutes.

(4) 5µl of Cy5/Cy3 Master Mix was added to each of the tubes. Samples were incubated in a 37°C water bath for 1 hour. Following incubation, each sample was again heated at 100°C and cooled on ice for 5 minutes each. A second dose (5µl) of the Cy3/Cy5 Master Mix was added to each of their designated samples and incubated at 37°C for 1 hour.

(5) Following the second hour of incubation, the reaction was stopped by adding 5µl of 0.5M EDTA (pH 8) to each tube. The enzyme was denatured using heat (72°C for 10 minutes). To ensure that the reaction had taken place, a 5µl aliquot from each sample was removed and analyzed by gel electrophoresis on 1% agarose gel.

4. **DNA Precipitation**

At this junction there were 4 tubes for each sample: Cy3 and Cy5 Test samples and Cy3 and Cy5 Reference samples. Differentially labeled DNA samples were combined together in test/reference pairs.

(1) 45µl of hybridization buffer (Spectral Genomics, Houston, TX, USA) was added to each of the two tubes.

(2) 12.9µl of 5M NaCl and 130µl of isopropyl alcohol were added to each sample to initiate the precipitation reaction. Samples were incubated in the dark at room temperature for 20 minutes.

(3) Samples were centrifuged at 13 000 rpm for 20 minutes to obtain DNA pellets. Purple pellets were desired, indicating equal amounts of Cy3 and Cy5 labeled DNA (in experiments where the pellet(s) did not turn out to be purple, the pellets were discarded and a new experiment was initiated).

(4) Supernatant was discarded and pellets were washed with 500µl of 70% ethanol. Samples were centrifuged for 3 minutes at 13 000 rpm and supernatant was removed. Pellets were air dried (in the dark) for 10 minutes.

5. **Hybridization**

(1) DNA was resuspended in 10µl of sterile water.

(2) Following resuspension, 30µl of hybridization buffer II (Spectral Genomics) was added to each sample.

(3) Samples were incubated at 72°C for 10 minutes to denature DNA to single strands, and then put in the water bath for 30 minutes at 37°C.

(4) Each sample was applied to a separate array platform. The sample was pipetted as a line of liquid down the centre of the platform. A 22X60 glass cover slip was immediately placed over the slide to spread the solution over the entire face of the slide.

(5) Each slide was placed in a hybridization chamber. 10µl of sterile water was added to the wells on both sides of each chamber to ensure that the inside of the chamber remains moist throughout the incubation period.

(6) Chambers were placed in a Kapak pouch lined with wet paper towel (to prevent evaporation). The pouch was heat sealed and placed in an incubator at 37°C for 16 hours.

6. **Post Hybridization Washing**

Each slide was washed in an individual Petri dish containing 25ml of wash solution in a rocking platform incubator.

Washes:

(1) At room temperature: 2X SSC, 0.5% SDS. Washed for 5 seconds by gently rocking the Petri dish and slid the cover slip off the array.

(2) At 50°C: 2X SSC/50% Formamide*. Incubated in rocking incubator for 20 minutes.

(3) At 50°C: 2XSSC/0.1%Igepal*. Incubated in rocking incubator for 20 minutes.

(4) At 50°C: 0.2X SSC*. Incubated in rocking incubator for 10 minutes.

(5) At room temperature: double distilled water. Each slide was washed twice (for 5 seconds each time).

*wash solutions were heated and maintained at 50°C prior to use.

Following the second water wash, the slides were IMMEDIATELY blown dry with a stream of nitrogen gas. Slides were placed in a dark slide box to protect from light.

7. **Scanning and Data Analysis**

Scanning of the slides took place on the same day as the washing to ensure minimum loss of signal. Slides were scanned using a GenePix 4000B scanner (Axon Instruments, Union City, CA, USA) with the help of GenePix Pro 6.0 software.

Data analysis was facilitated by Spectralware software Version 2 (Spectral Genomics, Houston, TX, USA). Clones with an average ratio value of less than 0.80 and greater than 1.20 were deemed of possible interest and were further investigated. Values from each duplicate clone on the array (each clone is spotted twice)

needed to be above cut-offs and both dye-swap experiments needed to show deviating ratios for the clones to be further investigated as described previously [121].

2.3.3 Agilent Technologies 105K Array Protocol

The array CGH platforms and reagents were purchased from a commercial company (Agilent Technologies, Santa Clara, CA, USA) and the manufacturer's protocol was followed. A pool of normal male or female control DNA (Promega, Madison, WI, USA) was used as reference DNA to match the sex of the samples studied.

1. **DNA Digestion**

   (1) 1.5µg of genomic DNA was added to a 0.5ml microfuge tube and nuclease-free water added to bring the final volume to 20.2µl.

   (2) The digestion Master Mix was prepared by mixing the following components (for each reaction) in the order indicated, and kept on ice:

   - 2.0µl of nuclease-free water
   - 2.6µl of 10X buffer C (Agilent Technologies, Santa Clara, CA, USA)
   - 0.2µl of Acetylated BSA (10µg/µl) (Agilent Technologies, Santa Clara, CA, USA)
   - 0.5µl each of Alu I and Rsa I (10U/µl) (Agilent Technologies, Santa Clara, CA, USA)

   (3) 5.8µl of digestion Master Mix was added to each reaction tube to make a total volume of 26µl.

   (4) The sample tubes were transferred to a circulating water bath at 37°C for 2 hours. Samples were then transferred to a heating block at 65°C for 20 minutes and then kept on ice for 5 minutes. At this point, DNA can be stored overnight at -20°C.

   (5) 2µl per sample were taken at this point and run on a 1% agarose gel at 90V for 20 minutes, to assess the completeness of the digestion. The majority of digested products should be between 200bp and 500bp in length.

2. **DNA Labeling**

(1) 5µl of Random Primers (Agilent Genomic DNA Labeling Kit PLUS, Agilent Technologies) were added to each sample and the reaction tubes were transferred to a Thermocycler (Eppendorf Canada, Mississauga, ON) preheated to 95°C and incubated for 3 minutes. Tubes were then put immediately on ice for 5 minutes.

(2) The labeling Master Mix was prepared by mixing the following components (for each reaction) in the order indicated, and kept on ice: (All reagents supplied with the Agilent Genomic DNA Labeling Kit PLUS).

- 2.0µl of nuclease-free water
- 10µl of 5X buffer
- 5µl of 10X dNTPs
- 3µl of Cy3/Cy5 (Reference samples Cy5, Test sample Cy3) *flick tubes before next step
- 1.0µl of Exo-Klenow fragment

(3) 21µl of labeling Master Mix was added to each reaction tube and gently mixed. Tubes were transferred to a circulating water bath at 37°C for 2 hours and then to a heating block at 65°C for 10 minutes to inactivate the enzyme, and finally moved on ice. At this point, DNA can be stored overnight at -20°C.

3. **Clean-up of Labeled Genomic DNA**

(1) 430µl of 1X TE (pH8.0) was added to each reaction tube.

(2) A Microcon YM-30 filter was placed into a 1.5ml microfuge tube (supplied) and the samples loaded into the filter. Tubes were spun at 10 000 rpm for 10 minutes at room temperature and the flow-through discarded.

(3) 480µl of 1X TE (ph 8.0) was added to each filter. Tubes were spun at 10 000 rpm for 10 minutes at room temperature and the flow-through discarded.

(4) The filter was inverted into a fresh 1.5ml tube (supplied) and spun for 1 minute at 9000 rpm at room temperature, to collect the purified sample.

(5) The volume of each eluate was measured and recorded. If sample exceeded 41µl, it was returned to its filter and spun again for 1 minute at 9000 rpm. This step was repeated until the volume was under 41µl and the flow-through was discarded.

(6) The sample volume was brought to 41µl with 1X TE (ph 8.0)

(7) 1.5µl of each tube was taken to determine yield and specific activity by using the NanoDrop ND-1000 UV-VIS Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). 1X TE (ph 8.0) was used as a blank.

(8) DNA was quality-controlled according to the following criteria:

- Specific activity of Cy3 (pmol/µg) was to be between 25 and 55
- Specific activity of Cy5 (pmol/µg) was to be between 20 and 40
- The yield in µg had to be between 5 and 10

(9) The appropriate Cy5 labeled reference was combined to the appropriate Cy3 labeled test sample for a total mixture volume of 79µl in a new 0.5ml heat-resistant microfuge tube. At this point, DNA can be stored overnight at -20°C.

4. **Preparation of Labeled Genomic DNA for Hybridization**

(1) The Blocking agent was prepared in advance in the following manner and stored at -20°C: 1350µl of nuclease-free water was added to the vial containing lyophilized 10X Blocking Agent (Agilent Oligo aCGH Hybridization Kit, Agilent Technologies). It was then left at room temperature for 60 minutes to reconstitute the sample before it was stored.

(2) The following component were added to the Labeled genomic DNA mixture in the following order:

- 25µl of Cot-1 DNA (1.0mg/µl)
- 26µl of 10X Blocking Agent
- 130µl of 2X Hybridization buffer (Agilent Oligo aCGH Hybridization Kit, Agilent Technologies)

(3) The samples were mixed by pipetting up and down, then quickly spun to drive contents to the bottom of the tube.

(4) The samples were transferred to a Thermocycler (Eppendorf Canada, Mississauga, ON) pre-heated to 95°C for 3 minutes, and immediately transferred to a circulating water bath at 37°C for 30 minutes.

(5) Samples were then quickly spun to collect the sample at the bottom of the tube.

5. **Array Assembly**

The Gasket was placed in the Hybridization chamber Agilent side up and the hybridization solution added to the Gasket side. The array slide was then placed on the Gasket slide Agilent side down and the Hybridization chamber placed in the oven at 65°C for 40 hours with a rotation speed of 16 rpm.

6. **Microarray Washing**

(1) Slide on top and Gasket on the bottom were put into a jar containing Oligo aCGH Wash Buffer 1 and disassembled using a plastic holder.

(2) Slides were then transferred to a second jar containing the same buffer and put on a stirring plate for 5 minutes.

(3) Slides were transferred to another jar containing Oligo aCGH Wash Buffer 2 that was pre-heated to 37°C and put on a stirring plate for 1 minute. *time sensitive

(4) Slides were then transferred to another jar containing Acetonitrile at room temperature for 1 minute on a stirring plate. *time sensitive

(5) Slides were taken out very slowly of the Acetonitrile, at which point they should be dry. Slides were put in a dehydrator until scanned.

7. **Scanning and Analysis**

Scanning of the slides took place on the same day as the washing to ensure minimum loss of signal. Slides were scanned using a GenePix 4000B scanner (Axon Instruments, Union City, CA, USA) with the help of GenePix Pro 6.0 software.

Data analysis was done using the Agilent Feature Extraction Software version 9.5.1 and the Agilent CGH Analytics software 3.5.14 (Agilent Technologies,

Santa Clara, CA, USA). The aberration algorithm selected was The Aberration Detection Method 2 (ADM-2) which is more robust in identifying small aberrant regions. A minimum number of 3 aberrant probes was needed with a minimal average absolute log ratio of 0.25 for a CNV to be called [108].

2.3.4 Affymetrix Human Mapping 250K Nsp GeneChip

Note: The following experiments were performed by collaborators at Queen's University. These additional data were available for 4 cases and were used to supplement copy number data acquired with the above-mentioned arrays. I was involved only at the data analysis step of this process and the first part of the following section was provided by Dr. Xudong Liu.

1. **DNA Digestion, Ligation, PCR and Hybridization**

   DNA samples were processed using the GeneChip Human Mapping 250K Nsp Assay Kit (Affymetrix, Santa Clara, CA, USA) and according to the GeneChip Mapping 500K Assay Manual (https://www.affymetrix.com/support/downloads/manuals/500k_assay_manual.pdf).

   (1) 250ng of genomic DNA were digested with Nsp restriction enzyme. Adapter sequences were ligated to the digested DNA and then subjected to PCR conditions that are optimized to amplify 250 - 1000 bp fragments.

   (2) The amplified sample was fragmented, end labeled and 20μg of DNA hybridized to the Human Mapping 250K Nsp GeneChip for 18 hours, 49°C at 60 rpm.

   (3) All GeneChips were stained and washed by the Affymetrix GeneChip Fluidics Station 450 and scanned with the Affymetrix GeneChip Scanner 3000 (Affymetrix, Santa Clara, CA, USA).

2. **Data Analysis**

   (1) Scanned images were analyzed with GTYPE 4.1 and genotype calls produced using the BRLMM algorithm (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf)

(2) Probe level data (data in the .cel file) was imported into Partek Genomics Suite 6.3 beta, 6.07.0629 (Partek, St. Louis, MO, USA).

(3) Data analysis steps were as follows:

- Data from all samples in a plate were imported using Partek defaults and adjusting for GC content
- Batch effect on scan date was removed
- Baseline was created from all samples
- Created copy number from baseline
- Hidden Markov model or segmentation algorithm was used to identify amplifications and deletions
- Gene annotations were added using Mapping_250k_Nsp.na23.annot.csv file from Affymetrix.

(4) Data was also manually viewed using the view chromosome option.

2.4 Secondary Confirmation Methods

In this study, FISH was used as a secondary, independent method to confirm array CGH findings for two reasons. First, it is used to ensure that the deleted or duplicated clone of interest found by the array is not an experimental artefact and that each clone maps to the chromosomal location outlined by the array analysis software and bioinformatics websites like NCBI (http://www.ncbi.nlm.nih.gov/) and UCSC (http://genome.ucsc.edu/). Second, FISH is used to provide any information regarding possible structural rearrangements that the deleted/duplicated clone(s) may be involved in. FISH was used to confirm deletions and duplications whenever sufficient cell pellet was available. In one case (06-43A), pellet was not available and a custom array was designed to confirm changes in this family. Two small changes found only by the Affymetrix array in two cases (07-102A and 06-107A) are being followed up by real-time quantitative PCR, to provide the most complete information possible for those two cases with an already defined PCNV.

2.4.1 Fluorescent *in situ* Hybridization

FISH analyses were performed using the BAC DNA clones purchased from The Centre for Applied Genomics (TCAG: http://www.tcag.ca/).

1. **Precipitation of Labeled BAC**

   (1) Per reaction: 2µl of labeled probe, 2µl of Cot-1 DNA and 9.5µl of sdH20 were added to a 2ml microfuge tube and mixed.

   (2) Per reaction: 1.4µl of 5M NaCl and 30µl of 100% isopropanol was added to the mixture to precipitate.

   (3) The mixture was then left at room temperature for 20 minutes in the dark, and then spun at 13 000 rpm for 20 minutes.

   (4) The supernatant was removed and the pellet rinsed with 50µl of 70% ethanol.

   (5) The pellet was then spun at 13 000 rpm for 3 minutes, the supernatant removed again, at which point the pellet was air dried for 5 minutes.

   (6) The pellet was then resuspended in 11µl (per reaction) of Hybrisol VII (Q-BIOgene, Irvine CA, USA), and incubated at 37°C for 10-15 minutes.

   (7) The probe was then denatured at 75°C in a heating block for 5 minutes, and then transferred to a water bath at 37°C for a pre-annealing step for 30 minutes.

2. **Metaphase Slide Preparation**

   Note: Patient slides were prepared by Celina Fawcett (Technician) or Chansonette Harvard (Technician). Protocol for this method was provided by Dr. Rajcan-Separovic.

   (1) Metaphase slides using cell pellets were prepared one day prior to hybridization. To prepare the slide, 2 drops of fixative solution (3:1 ratio of methanol and acetic acid) containing cells were placed on a pre-cooled blank slide. The slide was then covered with fresh fixative and placed in a humid chamber. Once dry, the slides were evaluated for quality.

(2) While the probe was precipitating, the slides were aged in 2X SSC at 37°C for 30 minutes.

(3) While the probe was in the pre-annealing step, the slides were put in 70% Formamide/2X SSC at 75°C in a heating block for 5 minutes.

(4) Slides were then dehydrated in an alcool series (2 minutes in 100% ethanol, 2 minutes in 80% ethanol and 2 minutes in 70% ethanol). Slides were then air dried.

3. **Hybridization**

   The probe was applied to the slide and the area marked with a pencil. A 22X22 glass cover slip was applied to the area and sealed with rubber cement. The slides were then placed in a humid chamber at 37°C for approximately 16 hours.

4. **Post-Hybridization Washes**

   Slides were removed from the incubator and the rubber cement and slip cover were taken off. Slides were placed in a solution of 0.4X SSC/0.3%NP-40 at 74°C for 2 minutes followed by placement in a solution of 2X SSC/0.1%NP-40 at room temperature for 1 minute. Following the second wash, slides were drained and air dried. 14µl of DAPI was applied to each portion of the slide containing a FISH experiment and covered with a cover-slip. The cover-slip was sealed using nail polish. Slides were placed in a -20°C freezer for 30 minutes in order for the DAPI to be absorbed.

5. **Slide Evaluation and Image Capturing**

   Note: To ensure accuracy, a blind evaluation and count was performed on each patient, parent and control slide by Dr. Evica Rajan-Separovic. Signals were then counted again by myself.

   Slides were evaluated using Zeiss Axioplan 2 fluorescence microscope and images were captured using MacProbe software (Applied imaging software, Santa Clara, CA, USA). For each FISH probe, at least 10 metaphase cells were analyzed. In case of a duplication detected with the array, up to 100 nuclei were scored for the number of FISH signals/nucleus.

2.4.2 Real-time Quantitative PCR

Note: All confirmations and parent of origin studies completed by qPCR experiments were performed by collaborators at Queen's University. I was not involved in the methodological aspects of this work, but only in the prioritization of additional changes detected by Affymetrix arrays requiring Real-time qPCR confirmation. The content of the following section was provided by Dr. Xudong Liu.

Two additional potentially pathogenic changes were detected by Affymetrix arrays in two different patients with confirmed CNVs detected by Spectral (06-107A) and Agilent (07-102A) arrays. These two changes were not detected by Spectral or Agilent arrays and were too small to be confirmed by FISH, therefore they are currently being confirmed by Real-time quantitative PCR to provide a more complete picture of CNVs for these patients. Briefly, changes were studied by using 3-5 non-polymorphic markers within the region using SYBR Green I detection. Bioinformatics tools including the publicly available human genome databases like NCBI (http://www.ncbi.nlm.nih.gov/) and UCSC (http://genome.ucsc.edu/) were used to design the primers. If the rearrangement was not detected by any of these markers, a further group of markers (2-4) were tested. If the second group of markers were negative, no additional testing was performed. Real time detection of PCR products was performed using an ABI Prism 7900HT system.

The system allows one to see the threshold cycle (CT) during the experimental phase of amplification (when none of the PCR reagents are limiting) and quantify each allele such that a single allele at a test locus in a person with a deletion would show 50% less amplification than in a person with 2 copies of that allele. Non-contiguous markers were used as controls in each experiment. The ratio of amplification is approximately 1 (range 0.8-1.2:1) if both loci have the equivalent of 2 alleles. If the test locus is deleted on one chromosome, the ratio of test to control markers is expected to be 1:2 (range of 1:1.7-2.5). If the test locus is duplicated, the ratio of the test to control markers is expected to be 3:2 (range of 2.6-3.5:2).

2.4.3 Agilent 8X15K Custom Design Arrays

In the case of one patient (06-43A), cell pellet was not available and thus FISH could not be performed for secondary confirmation. A custom array was therefore designed and DNA from the patient and both parents was used to confirm the CNVs of interest.

1. **Custom Array Design**

   Note: Arrays were designed by Chansonette Harvard (Technician). The following design protocol was provided by her.

   Custom Arrays were designed using eArray (Agilent technologies, Santa Clara, CA, USA). Agilent-optimized probes were selected from the *H. Sapiens* (UCSC hg18) probe set and searches were done using the standard HD probe search for user defined genomic intervals.

   To query aberrations larger than 100kb, the aberration area was decreased on either boundary by 25 kb and was filtered at an average probe spacing of 200 bp. A breakpoint of 50 kb, 25 kb on either side of the reported aberration boundary, was filtered at an average probe spacing of 100 bp. A further extension of 100 kb on either side of the breakpoints was filtered at an average probe spacing of 1000 bp.

   The remaining area was filled with a selection of CNVs and any unused spots were filled from Agilent's Human Genome 44K probe set. CNVs were filtered at an average probe spacing 1000 for changes less than 250 kb and an average probe spacing of 1500 for changes greater than 250 kb.

2. **DNA Digestion and Labeling**

   Please refer to Section 2.3.3 as these two steps are identical for both arrays.

3. **Clean-up of Labeled Genomic DNA**

   (1) 430µl of 1X TE (pH8.0) was added to each reaction tube.

   (2) A Microcon YM-30 filter was placed into a 1.5ml microfuge tube (supplied) and the samples loaded into the filter. Tubes were spun at 11 000 rpm for 10 minutes at room temperature and the flow-through discarded.

(3) 480µl of 1X TE (ph 8.0) was added to each filter. Tubes were spun at 11 000 rpm for 13 minutes at room temperature and the flow-through discarded.

(4) The filter was inverted into a fresh 1.5ml tube (supplied) and spun for 1 minute at 9000 rpm at room temperature, to collect the purified sample.

(5) The volume of each eluate was measured and recorded. If sample exceeded 10µl, it was returned to its filter and spun again for 1 minute at 9000 rpm. This step was repeated until the volume was under 10µl and the flow-through was discarded.

(6) The sample volume was brought to 10µl with 1X TE (ph 8.0)

(7) 1.5µl of each tube was taken to determine yield and specific activity by using the NanoDrop ND-1000 UV-VIS Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). 1X TE (ph 8.0) was used as a blank.

(8) DNA was quality-controlled according to the following criteria:

- Specific activity of Cy3 (pmol/µg) was to be between 25 and 55
- Specific activity of Cy5 (pmol/µg) was to be between 20 and 40
- The yield in µg had to be between 5 and 10

(9) 8 µl from the appropriate Cy5 labeled reference was combined to 8µl from the appropriate Cy3 labeled test sample for a total mixture volume of 16µl in a new 0.5ml heat-resistant microfuge tube. At this point, DNA can be stored overnight at -20°C.

4. **Preparation of Labeled Genomic DNA for Hybridization**

Please refer to Section 2.3.3 as the steps are exactly the same except for an adjustment of volume (below). The following component were added to the Labeled genomic DNA mixture in the following order:

- 2µl of Cot-1 DNA (1.0mg/µl)
- 4.5µl of 10X Blocking Agent
- 22.5µl of 2X Hybridization buffer (Agilent Oligo aCGH Hybridization Kit, Agilent Technologies)

5. **Array Assembly**

   The Gasket was placed in the Hybridization chamber Agilent side up and the hybridization solution added to the Gasket side. The array slide was then placed on the Gasket slide Agilent side down and the Hybridization chamber placed in the oven at 65°C for 24 hours with a rotation speed of 20 rpm.

6. **Microarray Washing and Scanning**

   Please refer to Section 2.3.3 as these two steps are exactly the same.

7. **Data Analysis**

   This was very similar. Data analysis was done using the Agilent Feature Extraction Software version 9.5.1 but this time the new Agilent DNA Analytics software 4.0.76 was used (Agilent Technologies, Santa Clara, CA, USA). The aberration algorithm selected was The Aberration Detection Method 2 (ADM-2) which is more robust in identifying small aberrant regions. A minimum number of 3 aberrant probes was needed with a minimal average absolute log ratio of 0.25 for a CNV to be called [108].

2.5 Gene Prioritization

I used a computational candidate gene prioritization software, Endeavor [122] to help pinpoint potential ASD-related genes within the deleted/duplicated genomic regions. The training (reference) gene set includes 19 putative and known autism-related genes summarized from recent publications: *FMR1*, *NLGN3*, *NLGN4X*, neurexin 1 (*NRXN1*), *SHANK3*, *CNTNAP2*, protocadherin 10 (*PCDH10*), deleted-in-autism 1 (*DIA1*), solute carrier family 9 (sodium/hydrogen exchanger), member 9 (*SLC9A9*), solute carrier family 9 (sodium/hydrogen exchanger), member 6 (*SLC9A6*), contactin 3 (plasmacytoma associated) (*CNTN3*), ataxin 2-binding protein 1 (*A2BP1*), *MECP2*, *UBE3A*, *EN2*, *SLC6A4*, *MET*, sodium channel, voltage-gated, type VII, alpha (*SCN7A*), ring finger protein 8 (*RNF8*) [123]. Based on the reference set, the Endeavor software prioritizes genes in each altered genomic region, which have similar functions or biological processes/pathways with the reference genes.

In contrast to other gene prioritization tools, the publicly available Endeavor software can access many more data sources (currently up to 20 for *H. Sapiens*) including categories of functional annotations, protein interactions, expression profiles, regulatory information, sequence-based and text-mining data [124] with all referenced information validated by *in vivo* experiments [122]. However, the rank of prioritized genes in a specific locus is mainly dependent on the training set of genes and the selection of models in the database.

Because the reference list of genes contains loci for which association with autism has not been definitively proven, and lacks many other possibly interesting genes that have not been studied yet, I felt it important to complement this computed analysis with manual investigations using the UCSC (www.http://genome.ucsc.edu/cgibin/hgGene), NCBI (www.http://www.ncbi.nlm.nih.gov/sites/entrez) and OMIM (www.ncbi.nlm.nih.gov/Omim/) databases and available literature. Such a manual investigation was undertaken when less than 5 genes were present in a region, or when Endeavor did not pinpoint any interesting candidates with p-values below 0.05.

2.6 Statistical Analysis

Exact Fisher tests were used for all categorical phenotype comparisons, while the non-parametric Mann-Whitney test was used to compare numerical data across categories. All statistical tests were performed using the Statistical Computation Web Site from Vassar College (http://faculty.vassar.edu/lowry/VassarStats.html).

Chapter 3

Case Reports

A summary of the genomic findings and clinical features of the 6 subjects described in this Chapter can be found in Table 3.1 and Table 3.2, respectively.

3.1 Case 1: Patient 06-107A

3.1.1 Clinical Presentation

Family and Prenatal History This patient is a 6-year-old male, the second child of two unrelated and healthy parents. His older brother is healthy and developmentally normal, despite previous diagnosed sensory integration disorder and mild gross motor delays, now resolved. Family history is positive for speech delay in one cousin on the maternal side, and suspected ASD in another maternal cousin, however no formal testing was performed in this individual. The remainder of the family history is non-contributory. Maternal and paternal ages were both 35 at delivery. Pregnancy was complicated by one episode of food poisoning at 7 weeks of gestation, followed by erratic low grade fevers as well as two episodes of mild spotting at 26 and 32 weeks. In addition, the mother described fetal movements as being of decreased intensity.

Medical and Developmental History Delivery was by repeat Cesarian section (C-Section) at 38 weeks with a birth weight of 4115g (just below 97%ile). The neonatal period was complicated by respiratory distress and the baby was discharged from hospital 5 days after birth. Significant health concerns arose during infancy including several episodes of vomiting with bile up to 11 months of age, asthma, gastro-esophageal reflux (GERD) and several upper respiratory tract infections before he was 15 months old. He was also diagnosed with anemia at 1 year of age. The patient is now on medication for GERD and asthma and is doing well.

Systems review is positive for nasal speech, dysphagia, drooling, a history of nasal regurgitation, hyperactive gag reflex, late eruption of teeth and bilateral plugged lacrimal ducts. An asymptomatic grade 1/6 systolic ejection murmur has been assessed in cardiology and no structural anomalies were found. An electromyogram (EMG) and nerve conduction studies were performed on the basis of longstanding hypotonia and were negative. There is no history of seizures or unusual staring spells and his hearing has been tested twice and is normal. A computed tomography (CT) scan in 2004 showed subtle hypodensity in the posterior cerebral white matter and two cranial MRIs performed in 2005 and 2006 revealed dilated Virchow-Robins spaces.

Gross motor development was delayed, with sitting unsupported at 14 months, sitting up on his own and rolling over at 18 months and walking at 30 months of age. No regression in skills is reported.

ASD Diagnosis and Cognitive Testing The proband was diagnosed with PDD-NOS at 2.11 years using standard objective measures (ADOS-G and ADI-R) [119, 120]. A diagnosis of PDD-NOS was given even though all ADI-R criteria were above cut-off for AD, because he did not consistently demonstrate the characteristic behaviours or exhibited them only mildly. He demonstrates hyperacusis, diminished pain sense, tactile defensiveness and self-injurious behaviours.

At the time of diagnosis, further psychometric indices were evaluated using the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and Vineland Adaptive Behaviour Scale (VABS). Using the WPPSI he was found to perform within the bor-

derline intellectual range (PQ=87, VIQ=79, FS=81). Adaptive testing with VABS was overall in the mild handicap range (composite score of 56), markedly below what would be expected from his level of intellectual functioning.

Clinical Genetics Examination Physical and dysmorphology examinations revealed several mild craniofacial dysmorphisms and minor physical anomalies and a height, weight and orbito-frontal circumference(OFC) all measuring around the 98%ile for his age. Both parents are also tall statured.

Craniofacial exam reveals a square-ish forehead and prominent chin, such that his overall craniofacial profile is triangular in shape. A low anterior hairline with frontal cowlick, wide-spaced eyes with down-slanting palpebral fissures and a short, upturned nose with wide and high nasal root were also present. Dentition is widely spaced on the upper and lower margins and he has a high-arched palate. In addition his philtrum is short and flat, there is thickening of the superior helices of both ears and he presents with a flat occiput with mild right-sided plagiocephaly.

Other minor physical anomalies include distal pectus excavatum and widened internipple distance, broad thumbs and halluces with strikingly prominent fingertip and toe pads. Musculoskeletal exam revealed generalized ligamentous laxity, and overall generalized hypotonia. He has bilateral tight heel cords, calcaneovalgus and pes planus, a hair-like patch at the distal spine and quite a deep sacral dimple. His phenotype score is 5.

3.1.2 Genomic Findings

Array Findings and Confirmation One database cataloguing putatively benign CNVs (DGV, http://projects.tcag.ca/variation/) was used to assess whether array-detected CNVs had been reported in normal populations. CNVs reported in at least two independent studies of healthy control subjects are typically referred to as BCNVs and are not investigated further, while other CNVs are referred to as disease-specific CNVs. In total, 3 CNVs were detected in this patient through array CGH (Spectral Genomics), of which

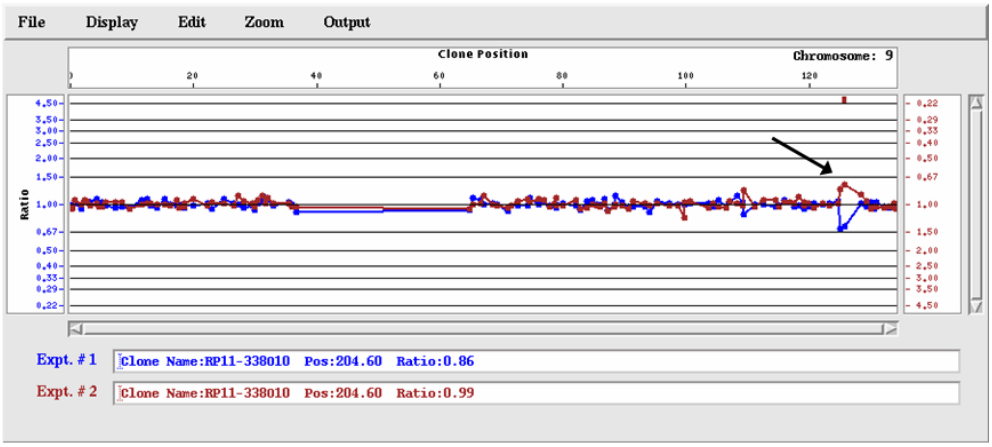

Figure 3.1: Example of Spectral Array Profile: del(9)(q34.11q34.13) in 06-107A

two were BCNVs and only one was a potential PCNV.

The only disease-specific CNV in this patient is a *de novo* deletion of about 2.5Mb (Figure 3.1), confirmed by FISH (Figure 3.2) at 9q34.11q34.13 , which was found to extend from position 130,525,315bp to 133,060,517bp by Affymetrix arrays. The deletion is not flanked by segmental duplications on either side, and is very gene-rich (Figure 3.3). Several very small deletions and duplications within the 2.5Mb region are reported in the Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation/), and all are reported only in single individuals in single studies except for a small duplication that was reported in 14 individuals in one study. In combination, deletions only cover about 5% of the region, with one additional large deletion (overlapping 300kb) that was detected by BAC-array, for which the precise size is unknown. Duplications in the region cover about 17% (Figure 3.3).

Overlap with PCNV Databases and Literature Databases cataloguing PCNVs like DECIPHER (https://decipher.sanger.ac.uk/) and the ACRD (http://projects.tcag.ca/autism/), and available literature were used to interpret the clinical significance of disease-specific CNVs detected in probands. For the PCNV in this patient, there is no overlap of cases in the Autism Chromosome Rearrangement Database (ACRD) (http://projects.tcag.

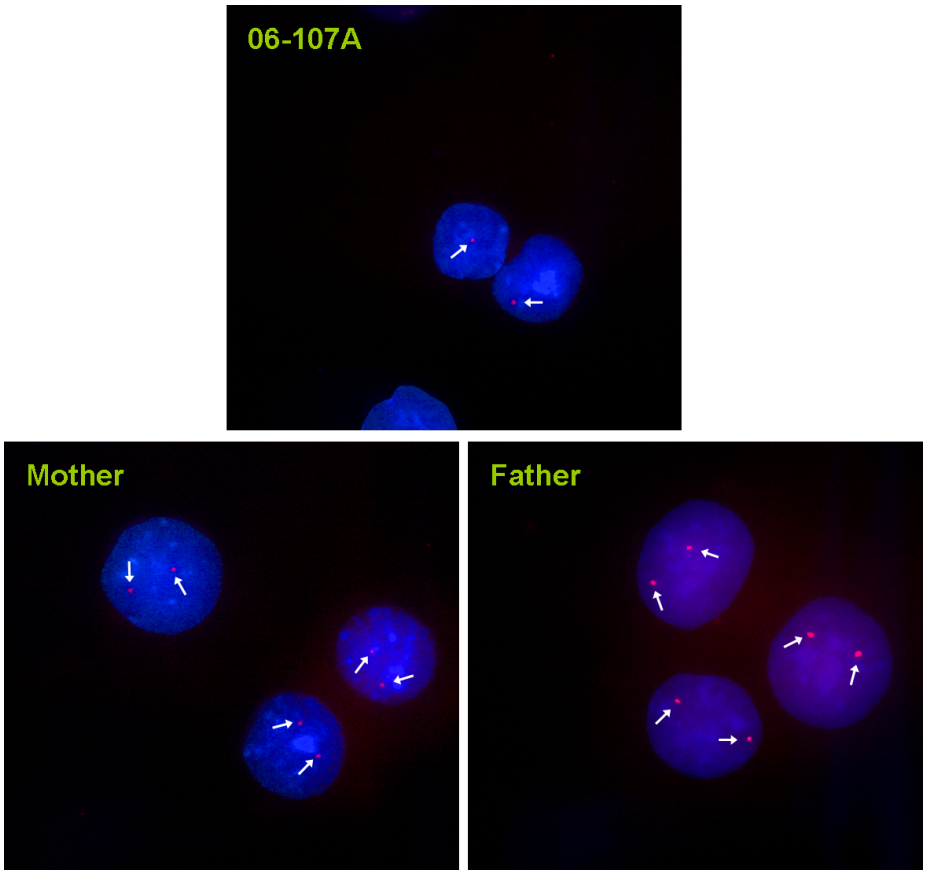

Figure 3.2: FISH Example of *de novo* CNV: del(9)(q34.11q34.13) in 06-107A

ca/autism/) or DECIPHER (https://decipher.sanger.ac.uk/) databases and only limited data about possible overlapping deletions in the literature (please see below). This novel microdeletion likely represents a new syndrome and a definitive biological explanation for the unique constellation of somatic and behavioral phenotypic features seen in this boy.

In total, only 22 cases of interstitial 9q deletions have been reported [125], with only 3 of those (from 9q32q34) possibly overlapping the region of interest in our patient [126–128]. Some of the craniofacial features reported in those patients overlap with our proband's, including low frontal hairline, hypertelorism, epicanthus, downslanting

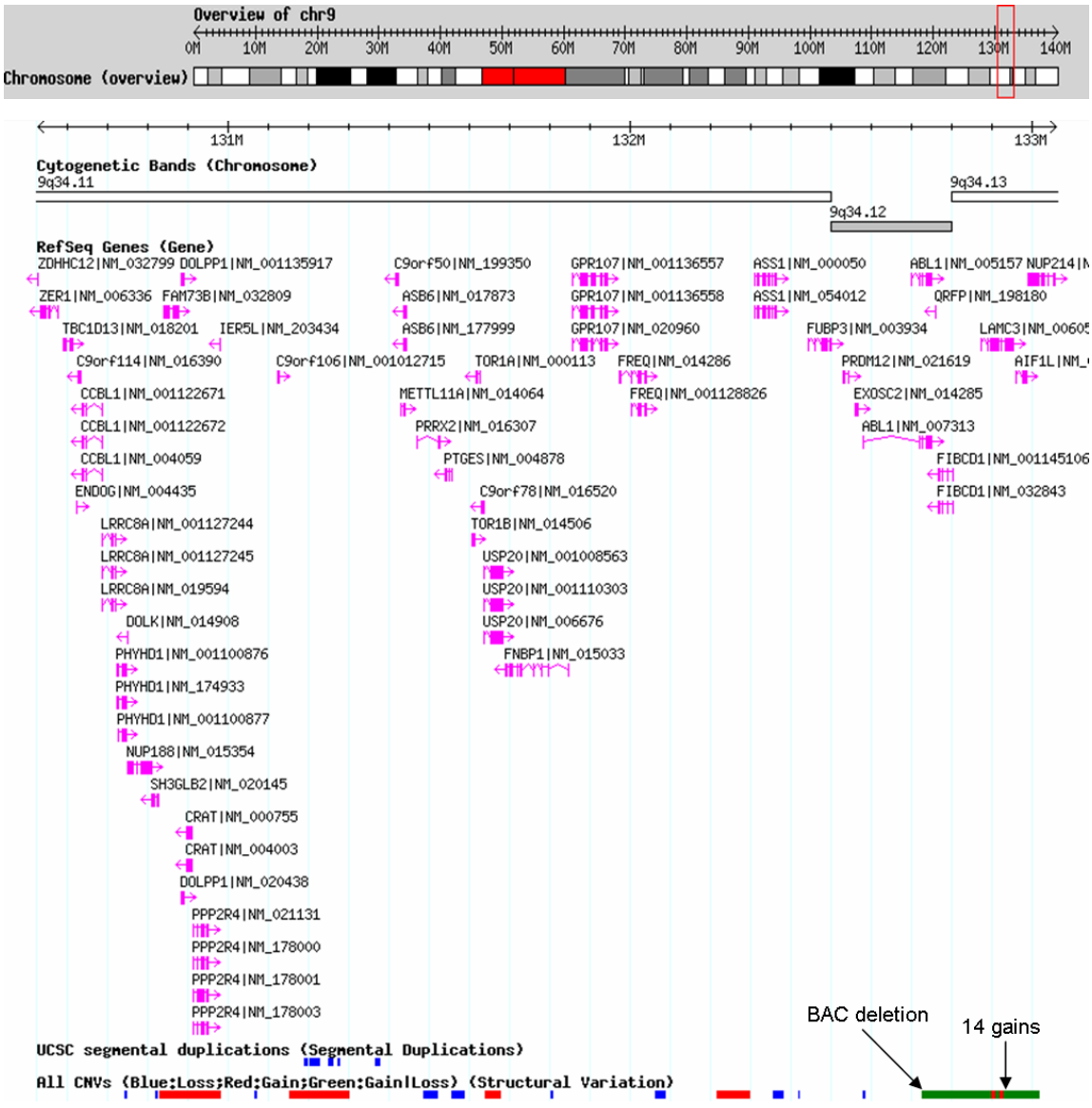

Figure 3.3: DGV Window for 06-107A

palpebral fissures and a broad nasal root. Other phenotypes reported include psychomotor and developmental delay, hypotonia, and short stature and microcephaly, the last two being in sharp contrast to the symmetric tall stature reported in our proband. No autism has been reported in any of these cases.

All three deletions were detected karyotypically, and thus a more precise determination of breakpoints was not reported. This precludes us from determining the extent of overlap, if any, with our case. Furthermore, the similarity of banding patterns between 9q22q32 and 9q23q34.1 can make it difficult to accurately assign breakpoints based on chromosome analysis alone [125].

It is interesting that there is phenotypic somatic and neurodevelopmental overlap between our patient's phenotype and the one typically seen in combination with the more common 9q34 terminal deletion syndrome, which is caused either by terminal deletions or mutations that lead to haploinsufficiency of euchromatic histone-lysine N-methyltransferase 1 (*EHMT1*), some 6Mb away. Phenotypic overlap includes hypotonia, hypertelorism, a short upturned nose and frequently reported cardiac murmurs and behavioural problems [129]. However there are also significant differences, including moderate to severe ID and microcephaly, which are hallmarks of the terminal syndrome and more proximal 9q21q33 deletions [125]. Whether or not there is gene-gene interactions between dosage-sensitive genes in the 9q region remains to be seen. The family has agreed to participate in RNA studies and expression array analyses, which will be conducted in the future.

Gene Content and Clinical Significance Using the gene prioritization software Endeavor (see Section 2.5) 2 genes were identified in the interval that are good candidates for autism: c-abl oncogene 1, receptor tyrosine kinase (*ABL1*) and torsinA (*TOR1A*), both with $p<0.05$. Additional manual investigations of candidates with $p<0.1$ did not reveal any other candidates except for torsinB (*TOR1B*), a gene very similar to *TOR1A*.

ABL1 is a tyrosine kinase better known for its role in cancers of the blood, but recent studies suggest that it is also involved in axon guidance, neurite outgrowth and synaptic structure and function [130, 131]. In effect, the protein has been shown to provide

an essential link between neuronal guidance receptors and the cytoskeletal machinery in the central nervous system (CNS) and peripheral nervous system (PNS) [131]. In Drosophila, mutations in the homolog gene cause certain axons to terminate prior to reaching peripheral muscle targets, and mutations that cause upregulation of the gene cause CNS axons to cross the midline inappropriately and peripheral motor axons to grow past their targets [132, 133].

In addition to their role in neuronal development, this family of kinases also appears to regulate synaptic structure and function [131]. The protein localizes to both presynaptic and dendritic spines of synapses in the hippocampal CA1 region [131] and electrophysiological studies reveal that the Abl protein modulates the efficiency of neurotransmitter release from the presynaptic terminal [134].

As for *TOR1A* and *TOR1B*, both are expressed in several regions of the brain, and both have a similar spatial and temporal expression pathway, with expression of both genes being undetectable during human fetal life, but becoming abundant between 4-8 weeks postnatally, with a stable expression in neurons throughout adult life [135, 136]. Inhibition of TorsinA protein expression by antisense treatment in human neuronal cell lines leads to longer neurites and an increase in growth cone area [135] while overexpression of the mutant protein (which causes autosomal dominant Early-onset torsion dystonia (OMIM: 128100)) interferes with neurite extension [137].

Additional Changes Detected by Affymetrix Another gene of interest is found to be disrupted in this patient, this time by a deletion of 20kb on 3p26.1 that was detected by Affymetrix array, from position 7,653,574bp to 7,673,488bp. This deletion is in an intron of glutamate receptor, metabotropic 7 (*GRM7*), a receptor for L-glutamate, the major excitatory neurotransmitter. Metabotropic glutamate receptor proteins have been linked to the Fmr protein encoded by *FMR1* in Drosophila and shown to modulate synaptic excitability [138]. Also, glutamatergic transmission has been implicated in self-injurious behaviours in rats [139]. Confirmation of this change by real-time qPCR is underway.

3.2 Case 2: Patient 07-93A

3.2.1 Clinical Presentation

Family and Prenatal History This patient is a 3-year-old male, the only child of two unrelated and healthy parents. Family history is negative for ASD, but is positive for severe ID in a paternal brother and sister. Maternal and paternal ages at delivery were 26 and 31, respectively. The patient was born at 39 weeks of gestation by spontaneous vaginal delivery, with a birth weight of 3200g (10-50%ile). Pregnancy was unremarkable except for a cold-like upper respiratory tract illness at 12-14 weeks of gestation, with a mild fever, yet no rash or other symptoms. The patient was discharged from hospital after two days and has remained very healthy since.

Medical and Developmental History The patient's vision has been tested and is normal and a hearing test could not be performed, yet there are no concerns with his hearing. There is no history of regression or suspected seizures, and a cranial MRI was performed and normal. The remainder of the systemic review is negative.

ASD Diagnosis and Cognitive Testing The proband was diagnosed with Autistic Disorder at 2.6 years of age using DSM-IV criteria [2], at which time the Bailey Scales of Infant Development (BSID), third edition and VABS were administered. The BSID placed him in the mild cognitive delay range at the 1st centile. Adaptive testing found him to be functioning in the borderline range overall, with an uneven profile of skills. Communication scores were at the 2nd centile (mild delay), daily living skills were scored at the 16th centile and socialization skills at the 7th centile (both in the borderline range), while motor scores were at the 25th centile, in the adequate range.

Clinical Genetics Examination Physical and dysmorphology examinations revealed a height between the 70th and 90th centile, a weight above the 98%ile and OFC at

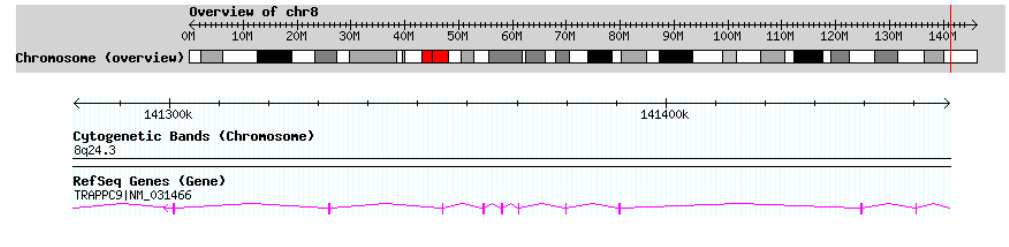

Figure 3.4: DGV Window for 07-93A

the 50%ile. Minor craniofacial dysmorphisms include downslanting palpebral fissures, medial eyebrow flare, a broad and high nasal root and a high-arched palate. Minor systemic anomalies include a small (1-2cm) hypopigmented lesion mid-thorac region over spine and prominent fingertip pads. His phenotype score is 6, including 2 points for non-mendelian family history of ID.

3.2.2 Genomic Findings

Array Findings and Confirmation By array CGH (Agilent 105K), 7 CNVs were detected, of which 6 are considered BCNVs. Affymetrix data were not available for this patient. The disease-specific CNV is a deletion of approximately 175kb at 8q24.3 detected by Agilent array, from position 141,280,797bp to 141,457,076bp. The deletion was confirmed by FISH and shown to be of maternal origin. There are no variants for this interval in the DGV (Figure 3.4), and the deletion involves several exons of one large gene known as trafficking protein particle complex 9 (*TRAPPC9*) or NIK and IKK(beta) binding protein (*NIBP*), and is located 500kb distal to potassium channel, subfamily K, member 9 (*KCNK9*).

Overlap with PCNV Databases and Literature There is no overlap with other cases in the literature, the ACRD or the DECIPHER databases.

Gene Content and Clinical Significance The change in this patient is considered to be putatively pathogenic, since both *NIBP* and *KCNK9* are good candidates for ASD. Investigation of *NIBP* reveals that it is expressed in neurons, where it functions as an activator of the NF-kappaB (*NFKB*) signalling pathway through increased phosphorylation of IKK(beta) (*IKKB*) [140]. *NFKB* is a transcription factor that regulates the expression of numerous genes of diverse function, including synaptic plasticity and neuronal survival and growth during development [141]. Knockdown of *NIBP* expression by small interfering RNA was shown *in vitro* to reduce tumor necrosis factor alpha (*TNFA*)-induced activation of *NFKB*, and prevents nerve growth factor (*NGF*)-induced neuronal differentiation [140].

NIBP is located only 27kb distal to *KCNK9*, an imprinted gene that has been shown in two independent studies to be expressed exclusively from the maternal allele in human fetal brains [142, 143]. The homologous region in mouse is imprinted as well and contains paternally expressed gene 13 (*PEG13*) and a germline maternally methylated differentially methylated region (DMR) (Catalogue of Parent of Origin effects, www.otago.ac.nz/IGC). A homolog of *PEG13* has not been found in humans, and a DMR has not been found in this region [142]. It is quite possible that *NIBP* is also imprinted, given its close proximity to *KCNK9*. The deletion of *NIBP* in this patient is maternally inherited and so if *NIBP* is shown to be imprinted and expressed exclusively from the maternal allele like its neighbour *KCNK9*, the deletion of the maternal copy of the gene in this patient would completely abolish the expression of the gene in this proband and possibly contribute to his ASD. Studies to elucidate the imprinting status of *NIBP* and its possible monoallelic expression in the brain, or any other tissues, have not been performed to date and will be crucial to determine its possible involvement in causing ASD behaviours in this patient.

A mutation in *KCNK9* has been found to cause the Birk Barel Mental Retardation and Dysmorphism syndrome (OMIM: 612292) in an extended pedigree presenting with maternal transmission of the disorder [144]. The phenotype in this family includes moderate to severe ID, severe hypotonia and a recognizable pattern of craniofacial dysmorphisms, a phenotype that does not resemble our patient's. The mutation causing this syndrome in this family is a missense mutation that has a dominant-negative effect. It abolishes all current in the channel, and results in a 4-fold decrease in current

capacity when co-expressed with wild-type potassium channel, subfamily K, member 3 (*KCNK3*), which forms a heterodimer with *KCNK9 in vivo* [144]. Although the phenotype in this pedigree is quite different from our patient's, we cannot rule out that the deletion on the maternal chromosome close to *KCNK9* could lead to dysregulation of the gene and somehow participate in the etiology of autism in this individual.

3.3 Case 3: Patient 06-43A

3.3.1 Clinical Presentation

Family and Prenatal History Patient 06-43A is a 13-year-old female, the only child of two unrelated healthy parents. Family history is negative for ASD or ID. Maternal and paternal ages at delivery were 35 and 32, respectively. Pregnancy was complicated by a fall on the stairs at 36 weeks of gestation, and gestational diabetes diagnosed at 28 weeks of gestation. Delivery at 40 weeks of gestation was complicated first by a trial of induction for failure to progress, followed by emergency C-Section. The baby was healthy with a birth weight of 4004g (just below the 90%ile) and the neonatal period was uneventful.

Medical and Developmental History The patient had her first seizure at 10 years of age, with a presentation consistent with Lennox-Gastaut syndrome. EEG testing was clearly abnormal on 3 different occasions between 2003 and 2005. Vision and hearing have been tested and are both normal. She achieved developmental milestones within normal limits and was described as a social, happy baby. She took her first steps at 9 months and walked well by 13 months of age. First words were spoken at 12 months of age and she was speaking in short sentences by 2 and a half years, at which point a significant regression in language skills was noted, as well as socialization difficulties and anxiety. Regression of bowel movements control followed at 3 years of age. She is also known to have fine motor difficulties, laboured and nasal speech, as well as difficulties with balance and coordination. She has tactile defensiveness, sensitivities to

sounds and touch, diminished pain sense and self-abusive behaviours (hand-biting). A cranial MRI done in 2005 was normal.

ASD Diagnosis and Cognitive Testing The proband was diagnosed with PDD-NOS at 7 years of age (using the Asperger's checklist) with semantic/pragmatic language difficulties and high anxiety interfering with functioning in stressful situations. A cognitive assessment at the age of five using the Stanford-Binet and Wechsler Individual Achievement Test (WIAT) reported variable cognitive skills with scores in the average range to well below average. She demonstrated word finding problems, difficulties with comprehension and non-command reactions and below average visual-motor integration and fine motor skills difficulties. Follow-up psychoeducational and speech and language assessments continued to document her difficulties with language processing, as well as difficulties with fine motor skills and coordination. At 7 years of age she had not yet developed hand dominance. She has also consistently been seen to have difficulties focusing and sustaining attention, yet not at a clinically significant level typical of ADHD.

Clinical Genetics Examination Physical and dysmorphology examination revealed a height and weight both at the 10%ile, with OFC at the 50%ile. Minor craniofacial dysmorphisms include a shallow forehead with flat supra-orbital ridges, strabismus, downslanting palpebral fissures, a broad and high nasal root and right facial asymmetry with left malar flattening. Other minor physical anomalies observed were left cubitus valgus, prominent fingertip pads, bilateral pes planus and ligamentous laxicity. Her phenotype score is 4.

3.3.2 Genomic Findings

Array Findings and Confirmation By array CGH (Agilent 105K), a total of 9 CNVs were detected in this patient, of which only two are disease-specific and potentially pathogenic, specifically two duplications on chromosome 16, at 16p11.2 and 16p13.2.

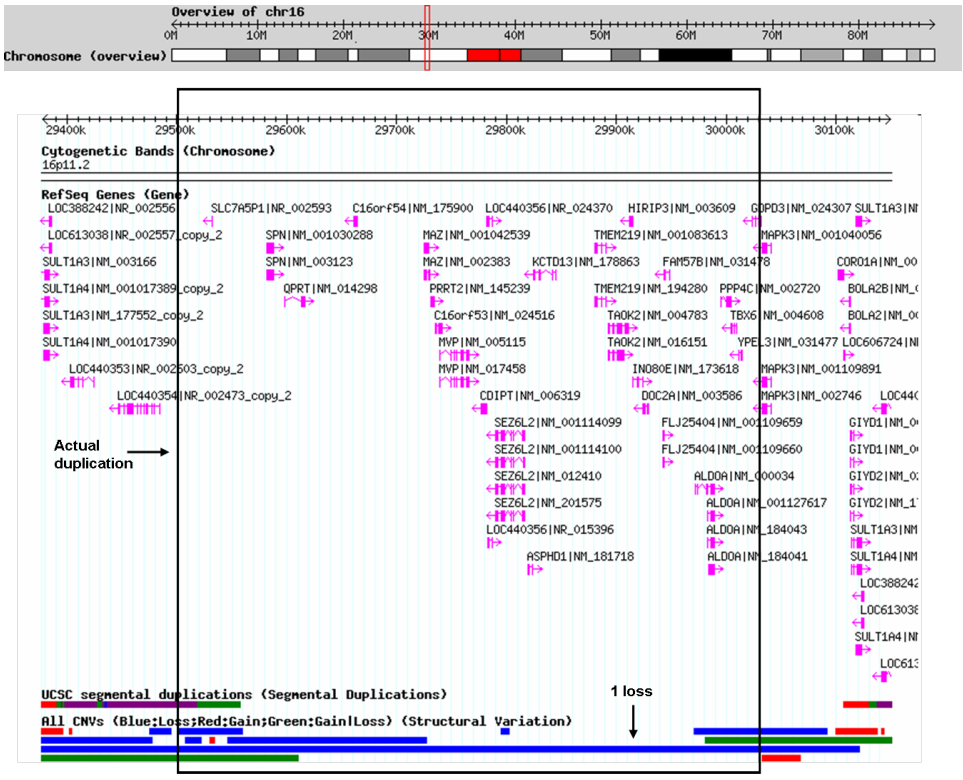

Figure 3.5: DGV Window for dup(16)(p11.2) in 06-43A. Zoom out shows Segmental Duplications

The duplication at 16p11.2 is around 530kb in size, from positions 29,500,084bp to 30,027,413bp, and is flanked by segmental duplications on both sides (Figure 3.5). The duplication at band p13.2 measures 220kb, from positions 8,709,048bp to 8,926,411bp (Figure 3.6). The first duplication at 16p11.2 is *de novo* while the 16p13.2 duplication is of maternal origin, both confirmed by a custom array design (Agilent 8X15K) (Figure 3.7). Overlaps in the DGV for dup(16)(p11.2) are partial and/or limited to areas close to flanking segmental duplications, except for one completely overlapping deletion reported in a single individual in a single study (Figure 3.5). As for dup(16)(p13.2), there are no overlap of normal variants in the DGV (Figure 3.6). Affymetrix array results were not available for this proband.

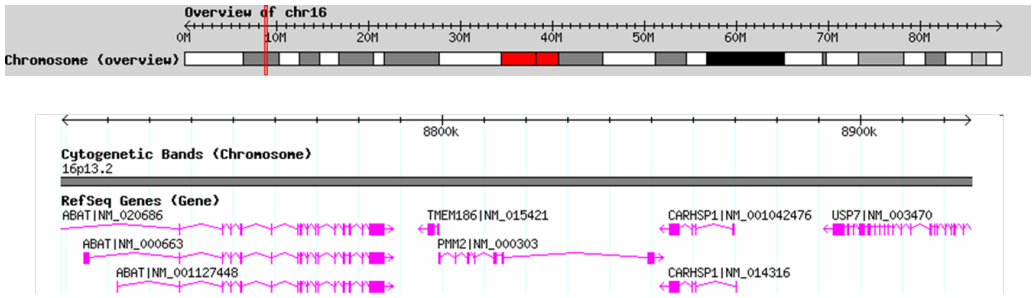

Figure 3.6: DGV Window for dup(16)(p13.2) in 06-43A

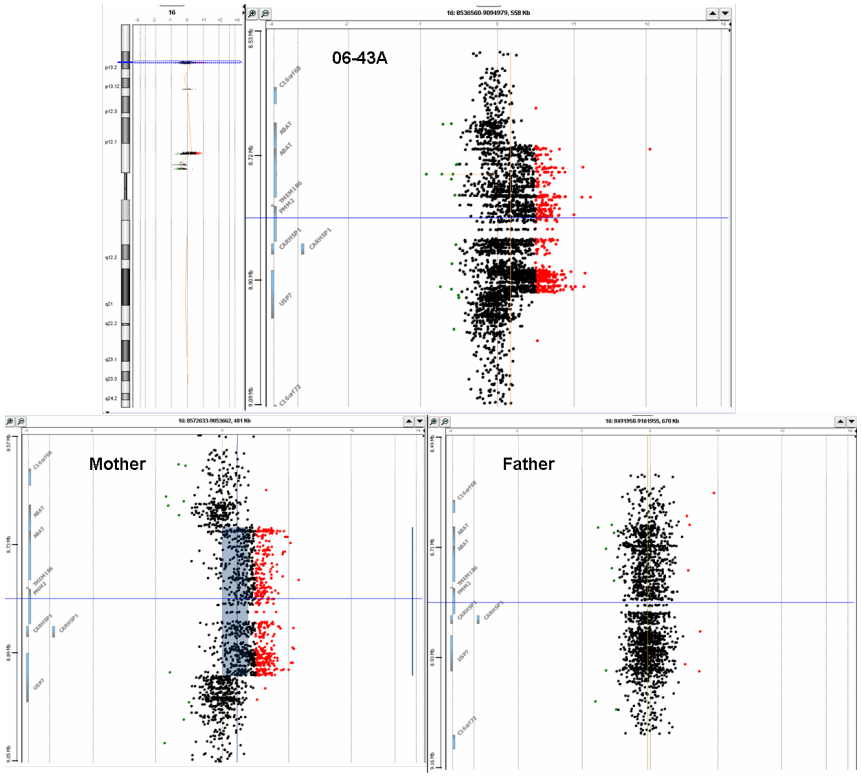

Figure 3.7: Example of Custom Array Profile: dup(16)(p13.2) in 06-43A

Overlap with PCNV Databases and Literature Del(16)(p11.2) and dup(16)(p11.2) with almost identical overlap are reported in DECIPHER with phenotypes including ID, speech delay and complex partial seizures. In the ACRD, there are numerous deletions and duplications of the same region as well. This region has been reported in many patients with autism and other psychiatric disorders or milder behavioural problems in the literature with the duplication sometimes inherited from apparently normal parents [114] (For more information, please refer to Section 1.4.4). In addition, a recent paper pinpointed the region as one of several potential PCNV regions for neurological disorders, including autism [145] (for more information, please refer to Section 3.4).

For the 16p13.2 duplication, there are two overlapping duplications in DECIPHER, one without a reported phenotype, the other with a phenotype including a craniofacial phenotype that does not overlap, percepto-motor delay without ID, no seizures and hypotonia. There is no overlap in the ACRD for this region and no overlap of cases in the literature.

Gene Content and Clinical Significance The gene prioritization software Endeavor pinpointed 3 possible candidate genes in the 16p11.2 region: sialophorin (*SPN*), seizure-related 6 homolog (mouse)-like 2 (*SEZ6L2*) and myc-associated zinc finger protein (*MAZ*). *SPN* is expressed on the surface of human T lymphocytes, monocytes, granulocytes and some B lymphocytes and after careful investigations, the reasons why this gene was prioritized by the software are unclear. One recent association study of genes in this region prioritized other genes in addition to *SEZ6L2* and *MAZ*, the most interesting of them being double C2-like domains, alpha (*DOC2A*) [146].

SEZ6L2 is a homolog of seizure-like 6 (*SEZ6*) in mouse, named as such because it is acutely upregulated in response to seizure-inducing agents [147]. Expression analysis of *SEZ6L2* in human fetal brains showed high expression in post-mitotic cortical layers, hippocampus, basal ganglia, amygdala and thalamus [146]. In mouse, postnatal expression of the *SEZ6* protein has been shown to be marked in regions with ongoing plasticity, such as the hippocampus, cerebellum and olfactory bulb [148]. The *SEZ6* protein acts at the cell surface to sculpt the dendritic arbor and influence synaptic connectivity [148]. The behaviour of null mice was normal on many tests, though reductions in general

activity, motor function, anxiety-related and depression-related behaviours and spatial memory were documented [148]. One rare variant associated with autism in *SEZ6L2* was identified in an initial study, which was not confirmed in follow-up [146].

The gene product of *MAZ* promotes the expression of glutamate receptor, ionotropic, N-methyl D-aspartate 1 (*NR1*) during neuronal differentiation [149]. *NR1* is the only indispensable NMDA receptor subunit for the formation of proper receptors, which are responsible for a major portion of excitatory synaptic transmission in the CNS [150]. *MAZ* has also been shown to regulate transcription of phospholipase C-like 1 (*PRIP1*), which is expressed predominantly in the brain and regulates brain-derived neurotrophic factor (*BNDF*)-dependent regulation of gamma-aminobutyric acid (GABA)-alpha receptors, which are responsible for fast synaptic inhibition [151].

The gene product of *DOC2A* is a brain-specific synaptic vesicle-associated protein thought to serve as a calcium sensor for neuronal activity and short-term priming of synapses [152]. Mice deleted for this gene show alterations in synaptic transmission, long-term potentiation and exhibit long-term memory deficits [153]. In their screen for rare variants of genes in the 16p11.2 region, Kumar et al. 2009 found one autism-specific amino acid substitution predicted to affect protein function, located in the domain which interacts with calcium and phospholipids. They also identified a promoter variant in another patient that is predicted to alter transcription factor binding sites for several brain-expressed genes [146].

Two other genes in 16p11.2 are of interest for a possible association with epilepsy, namely major vault protein (*MVP*) and quinolinate phosphoribosyltransferase (*QPRT*). *QPRT* encodes a key enzyme in the catabolism of quinolinate, which acts as a most potent endogenous excitotoxin to neurons through hyperstimulation of N-methyl-D-aspartic acid (NMDA) receptor [154]. Quinolinate is thought to be involved in the pathogenesis of epilepsy [155], Alzheimer's disease [156] and Huntington's disease [157]. *MVP* is the main component of vault particles [158], and is highly over-expressed in many multidrug resistant cancer cell lines (see [159] for review). *MVP* was postulated to contribute to multidrug resistance in epileptic pathologies, and upregulation has been shown in surgically resected human brain material from patients with common causes of refractory epilepsy [160]. The duplication of the *MVP* gene in this proband could

possibly contribute to her form of seizures which have been resistant to many different medications and are to date not controlled properly.

Besides *MVP* and *QPRT*, a third candidate for seizures in this patient is found in the 16p13.2 region and the duplication breakpoint is in between 2 alternate promoters that give rise to identical proteins. The gene is 4-aminobutyrate aminotransferase (*ABAT*), and its product is responsible for the catabolism of GABA. The enzyme activity has been studied in platelets of patients suffering from different forms of epilepsy, with some seizure types associated with decreased activity while others seem to be associated with increased activity [161, 162]. Furthermore, a study found that the drug phenylethylidenehydrazine, and inhibitor of *ABAT*, reduced epileptiform activity in rat hippocampal slices [163]. Finally, *ABAT* has been associated with the autism phenotype in one study to date [164].

In summary I consider both dup(16)(p11.2) and dup(16)(13.2) to be putatively pathogenic in this patient, mainly because of the abundance of interesting candidate genes in both regions. Furthermore, there is good evidence that deletions and duplications of this particular region of 16p11.2 are found more often in individuals with autism than controls (see Section 1.4.4). The pathogenicity of dup16p13.2 in this patient is less clear, because it was inherited from a normal mother, but we cannot exclude the possibility of gene interactions between the two duplicated loci in the proband that would influence her phenotype.

3.4 Case 4: Patient 07-102A

3.4.1 Clinical Presentation

Family and Prenatal History This patient is a 3-year-4-month-old male, the only child of two unrelated parents. Family history is positive for bipolar disorder in the mother, and for a possible ASD in a cousin on the maternal side, however no formal assessment has been performed in this individual. Maternal and paternal ages at delivery were 31 and 37, respectively. Pregnancy was complicated by daily use of lithum

medication at doses of 900mg daily during the first trimester and 1200mg during the second and third trimester, until delivery. Triple maternal screen was negative for Down Syndrome, Open Spina Bifida and Trisomy 18, but hCG levels were elevated, at 12.68 multiples of the median. An ultrasound at 18 weeks coupled with a fetal echocardiogram was normal, but a follow up ultrasound at 22 weeks revealed the presence of Ebstein's anomaly in the fetus and mild polyhydramnios. An ultrasound at 26 weeks confirmed the cardiac defect and moderate polyhydramnios. Gestational diabetes developed in the third trimester and was controlled by diet. The baby was delivered at 37 weeks of gestation by C-Section prompted by abnormal fetal heart rate monitoring. His birth weight was 3600g, measuring just below the 90th centile.

Medical and Developmental History The baby had a complicated neonatal course, requiring resuscitation and intubation. Hepatomegaly and an enlarged head with associated plagiocephaly and craniosynostosis were noted at birth. Apgar scores were 2,6 and 7 at 1, 2 and 5 minutes respectively and neonatal course was further complicated by hypercalcemia. Postnatal cardiology evaluation revealed a dysplastic tricuspid valve of a normal aortic arch and wide patent ductus arteriosus (PDA) with a small pericardial effusion causing cardiomegaly. He underwent a balloon septostomy of the atrium with ligation of the PDA on day two.

He had an early delay in gross motor acquisition crawling at 10 months and walking at 18 months, which was felt to be related to his cardiac surgery and recovery. At 3.4 years he is now quite mobile in terms of walking and running but is felt to be a little clumsier than peers. Development is marked by regression in play and social skills as well as communication, pronunciation and building vocabulary from the age of 28-30 months. CNS review is positive for frequent and unusual staring spells. An EEG was performed and slightly abnormal, and cranial MRI is normal. He has hypersensitivity to sounds and intermittent vocal tics. Vision and hearing have been tested and are both normal.

ASD Diagnosis and Cognitive Testing The proband was diagnosed with ASD at 3.1 years of age using standard objective measure (ADI-R, ADOS-G) [119, 120], at which

time the BSID was administered, which revealed cognitive scores around the 25 month level, with communication skills ranging from 15-18 months and a relative strength in motor skills, with fine and gross motor skills evaluated at the 27 and 19 month levels, respectively.

Clinical Genetics Examination Physical and dysmorphology examinations revealed a height between the 75th and 95th centiles, a weight at the 25%ile and OFC above the 98%ile, which is also documented in his father. Minor craniofacial dysmorphisms include right facial asymmetry with a flat left occiput, a broad nasal root and his two lower front teeth are joined. Other minor physical anomalies include long and tapered fingers with prominent fingertip pads as well as bilateral pes planus and calcaneus valgus. His phenotype score is 5.

3.4.2 Genomic Findings

Array Findings and Confirmation In total, 9 CNVs were detected in this individual by array CGH (Agilent 105K), of which only one is believed to have a possible pathogenic effect. This PCNV is a deletion on chromosome 15 in bands q25.2q25.3 of approximately 620kb, from positions 82,901,902bp to 83,522,315bp (Figure 3.8). The deletion was confirmed by FISH and found to be of maternal origin. The region is flanked by segmental duplications and rare partial duplication variants cover around 30% of the area between duplicons, with partial deletions reported in single control individuals covering around 25% of the region (Figure 3.9).

Overlap with PCNV Databases and Literature One deletion of unknown origin reported in DECIPHER overlaps with almost identical breakpoints to a patient reported to have proportionate tall stature and ulnar deviation of the fingers. Another deletion reported in DECIPHER overlaps the more proximal 140kb of the region and is of maternal origin. The phenotype reported in this patient includes macrocephaly, ID, hypertelorism and feeding problems in infancy.

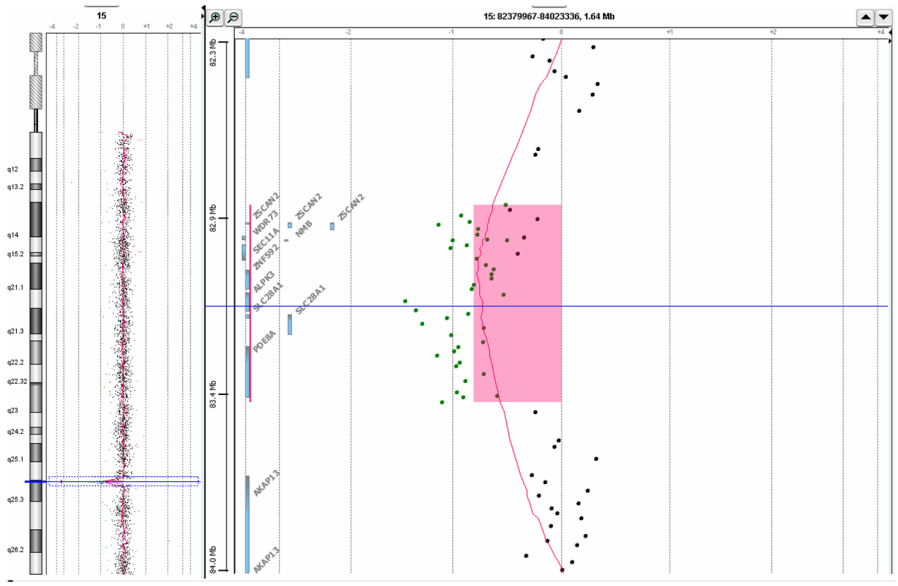

Figure 3.8: Example of Agilent Array Profile: del(15)(q25.2q25.3) in 07-102A

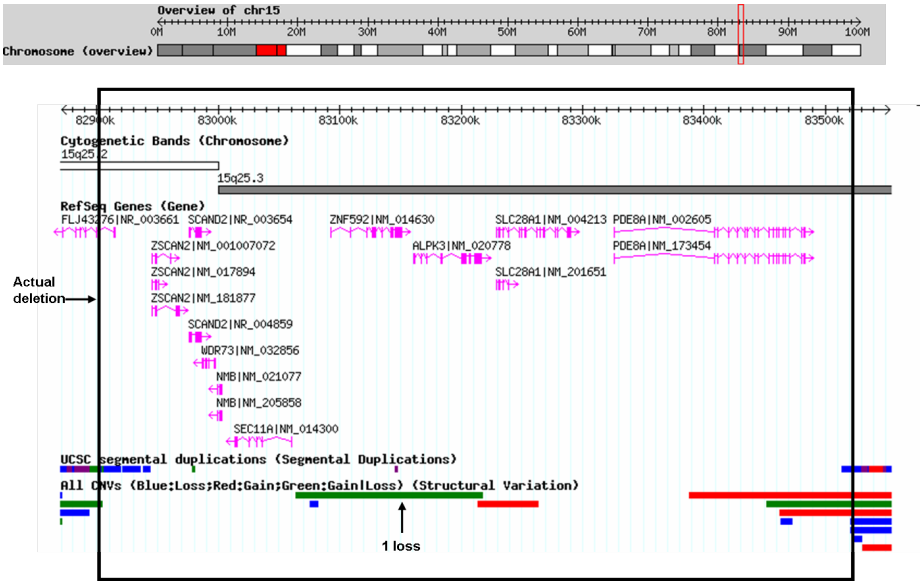

Figure 3.9: DGV Window for 07-102A. Zoom out shows Segmental Duplications

In addition, there are 2 reported deletions with practically identical breakpoints reported in the ACRD, 1 inherited and one of unknown origin [99, 100]. One additional *de novo* duplication is reported, detected microscopically (15q25.2qter) with an associated phenotype of severe language delay, postnatal overgrowth, strabismus and seizures in addition to autism [165]. Two additional deletions overlapping 15q25 are reported in the literature, with phenotypes very different from our proband, including microcephaly, hypotonia, growth retardation and renal anomalies [166]. Those deletions are much larger, spanning the region between q24q26.1.

Several cases with duplications of terminal 15q are reported in the literature [165, 167–169], and the phenotype reported includes postnatal overgrowth, learning difficulties, motor delays, renal anomalies and a long and thin face with a prominent chin and nose. Only one of those cases is reported to present with autism [165] and was reported in the ACRD. In many of those patients, only height and OFC are reported and when weight is reported, it is always reported to be disproportionately lower, which was also found in our proband. Two studies reporting microscopically detectable duplications of 15q25qter also report craniosynostosis as part of the presenting phenotype [167, 168].

Of interest, the region between segmental duplications on 15q25.2q25.3 was reported in a recent large scale study of CNVs in normal populations as a region for possible pathogenicity in ASD, ID and schizophrenia [145]. Copy number variable loci of 500kb or more (to minimize platform effects) were compared from different studies of normal, autistic, intellectually disabled and schizophrenic populations in a total of 6860 affected and 5674 controls. Loci were ranked for their potential pathogenic effects by calculating allele frequency differences of gains, losses and total CNVs in affected versus control individuals. Loci ranked highest included deletions and duplications of 22q11, 15q11q13 and 16p11.2, as well as deletions of 15q25.2q25.3, which were found twice in autistic individuals (see above) and twice in schizophrenic subjects, but never in controls.

Gene Content and Clinical Significance The gene prioritization software Endeavor did not highlight any interesting candidate genes in the region, and neither did a manual

investigation of each gene in the region using web databases (UCSC, NCBI, OMIM) and available literature.

Literature searches were also done to investigate the link between prenatal lithium exposure and Ebstein's anomaly as well as developmental disorders. In a review of 4 case-control and 2 cohort studies in 1994, it was found that the risk of Ebstein's anomaly following prenatal exposure to lithium was between 0.05 and 0.1%, 10-20 times higher than the population prevalence [170]. Long-term effects on infant neurodevelopment are undetermined, but limited data suggests normal behavioural patterns in childhood [171].

In conclusion, this del(15)(q25.2q25.3) is considered to be putatively pathogenic in this individual, even in the absence of good candidate genes, because of overlaps in the ACRD and DECIPHER databases, and data supporting that this region could be linked to behavioural and cognitive deficits [145]. Furthermore, several genes in the region do not have well established functions, and further studies are needed before we can definitively exclude them from being involved in ASD pathophysiology.

Additional Changes Detected by Affymetrix Affymetrix array data was available for this patient and revealed an additional change that warrants further study. A duplication of 99kb at 2q24.2 was detected from positions 163,333,466bp to 163,432,554bp which contains the first two exons as well as 25kb of sequence 5' to potassium voltage-gated channel, subfamily H (eag-related), member 7 (*KCNH7*). The protein is expressed exclusively in the brain, predominantly in the frontal lobe and cerebellum in humans and is known to be blocked by at least 2 antipsychotic drugs [172]. Confirmation of this change by real-time qPCR is underway.

3.5 Case 5: Patient 06-134A

3.5.1 Clinical Presentation

Family and Prenatal History This patient is a 15-year-old male born to unrelated and healthy parents. He has one older sister who is developmentally normal and extended family history is negative for ASD or ID. Maternal and paternal ages at delivery were 29 and 30, respectively. Pregnancy and neonatal periods were unremarkable, and he was delivered by repeat C-Section at 37 weeks of gestation with a birth weight of 3680g, which measures just above the 90th centile.

Medical and Developmental History The patient is a very healthy boy, with no hospitalizations or surgeries and no recurring infections. Vision and hearing have been tested and are normal and there is no history of seizures. An EEG and cranial MRI have both been performed and are reported to be normal.

Development was normal up until 1 year, after which little progress was made in language development. At 15 years of age, language is still very limited, and consists mostly of short phrases used for requests.

ASD Diagnosis and Cognitive Testing The proband was diagnosed at 3.1 years of age with moderate to severe AD using DSM-IV criteria [2]. Upon re-evaluation at 12 years of age, cooperation for formal cognitive testing was insufficient to get a good estimate of intellectual ability, but it was felt to be in the moderate ID range based on a composite VABS score of 46. Behavioural problems include hyperactivity, aggressive outbursts, mood swings and high anxiety, and he has a history of head-banging which has now resolved.

Clinical Genetics Examination Physical and dysmorphology examinations reveal a height and weight both measuring at the 50%ile and OFC greater than the 98%ile, a find-

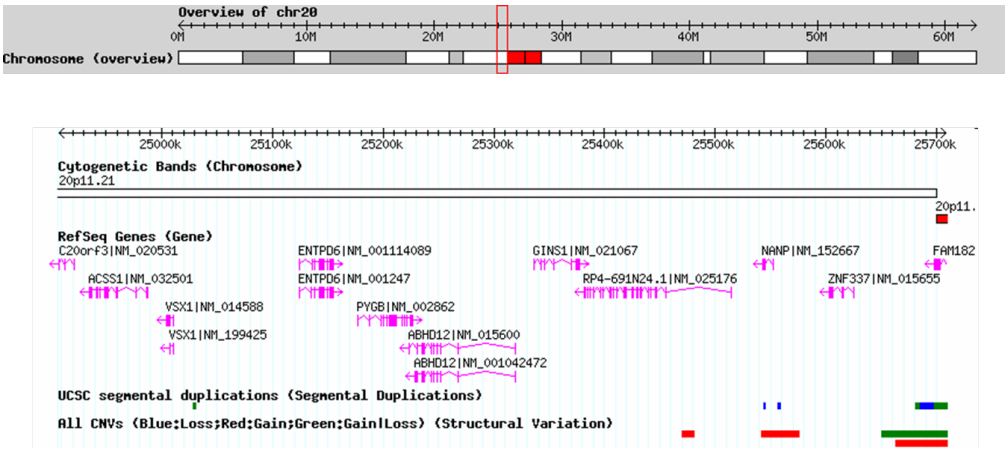

Figure 3.10: DGV Window for 06-134A

ing also documented in his father. Minor craniofacial dysmorphisms include a prominent forehead, upslanting palpebral fissures with a medial eyebrow flare, right facial asymmetry with malar flattening and prognathia, prominent lips, a high-arched palate and thick alveolar ridges and he is missing both lower lateral bicuspids. Prominent fingertip pads is the only other minor physical anomaly noted on examination. His phenotype score is 4.

3.5.2 Genomic Findings

Array Findings and Confirmation By performing array CGH (Spectral Genomics) a total 2 CNVs were detected in this patient, of which only one is considered to have potential clinical significance at the moment, specifically a deletion of approximately 800kb at 20p11.21. The deletion was confirmed by FISH and shown to be of paternal origin and Affymetrix array data was used to more precisely define breakpoints, from positions 24,907,412bp to 25,708,297bp. The deletion is flanked by a segmental duplication on one side only, and only two small gains have been reported in DGV (Figure 3.10). Investigations of Affymetrix array results did not reveal additional CNV regions of interest.

Overlap with PCNV Databases and Literature One maternally inherited duplication of 400kb overlaps with this patient's deletion in DECIPHER in a patient reported to have ID and a short attention span, but no ASD. Upon investigation for a possible overlap in the ACRD, I came across 3 translocation breakpoints mapping to 20p11.2, but more information about the precise physical locations of those breakpoints is lacking to ascertain where they fall in band 20p11.2. In addition, one overlapping microscopically visible deletion of 12Mb on 20p11.2p12.2 is reported, with another deletion of 8.7Mb at 20p11.22p11.23 which maps about 2Mb away from our patient's. Overlap for these deletions can be precisely estimated, as both have been further refined with arrays [173]. Both deletions are much larger, and it is thus not surprising that a much more complex phenotype is reported in addition to autism and ID.

Gene Content and Clinical Significance The change in this patient is considered to be of unknown clinical significance, as there is very minimal overlap with other cases in published and unpublished reports and investigation of the genes in the region did not reveal any interesting candidates, either by the Endeavor software of through manual investigations. Moreover, the deletion is inherited from a normal father. As for many other regions in the human genome that contain many genes, several have not been functionally investigated and some have not even been proven to be expressed at the protein level yet, so we cannot rule out that a gene of unknown function might be involved in this patient's phenotype, in which case the deletion could be unmasking a recessive deleterious allele on the other intact chromosome. Genetic interaction with modifier genes or imprinting of certain genes in the region are other possibilities.

3.6 Case 6: Patient 06-119A

3.6.1 Clinical Presentation

Family and Prenatal History This patient is a 3-year-old male, the only child of two unrelated and healthy parents. Family history is negative for ASD or ID. Maternal and

paternal ages at delivery were 31 and 34, respectively. He was born by spontaneous vaginal delivery at term, with a birth weight of 3923g (50th-90th centile), following an unremarkable pregnancy. He was discharged from the hospital after two days and has remained healthy since, except for recurrent ear infections (X3).

Medical and Developmental History His hearing has been tested and is normal, and there are no concerns regarding any vision problems. There are no suspected seizures nor staring spells and an EEG has been performed and is normal. The remainder of the systemic review is negative.

The patient achieved motor milestones within average limits, but has still yet to develop language. At 22 months of age, his receptive language was judged to be at the 9 month-old level, while his expressive communication skills were at the 6-9 month-old level, both significantly delayed. His oral motor function was regarded to be adequate for the production of speech. Vocalizations were observed and used to gain attention, but communication gestures were quite limited and eye contact was inconsistent.

ASD Diagnosis and Cognitive Testing The proband was diagnosed with AD at 2.7 years of age using standard objective measures (ADI-R and ADOS-G) [119, 120], at which time the BSID, second edition, and the VABS were administered. Using the BSID, he was found to present with overall moderate developmental delay, but inconsistent with global delay, since he successfully completed tasks for his age. His overall developmental age equivalent was around 18 months, but with a wide range of 5-27 months. Individual domains on the VABS tests were consistently in the mildly handicapped range and again a broad scatter of ability was noted (from 8-21 month-old levels).

Clinical Genetics Examination Physical and dysmorphology examinations revealed a height between the 50th and 70th centile, a weight at the 90%ile and OFC at the 50%ile. Minor dysmorphisms include facial asymmetry with right prominence, frontal bossing, epicanthal folds, a broad nasal root and prominent fingertip pads. His phenotype score is 3.

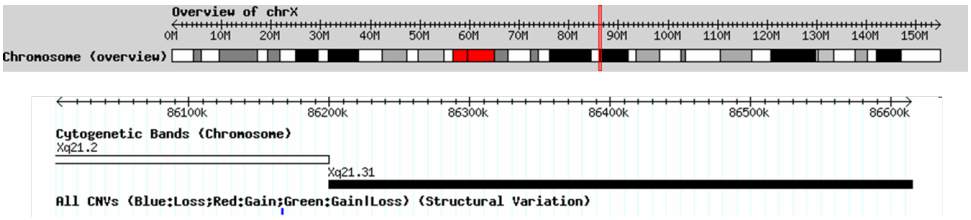

Figure 3.11: DGV Window for 06-119A

3.6.2 Genomic Findings

Array Findings and Confirmation By array CGH (Spectral Genomics), a total of 4 CNVs were detected in this individual, 3 of which have been reported in normal populations, including a maternal 2q terminal polymorphism that was also detected by FISH with subtelomeric probes in a clinical setting. The fourth CNV detected was a duplication in Xq21.2q21.31 involving two clones (RP11-192B18 and RP11-145I17). The duplication was confirmed by FISH and shown to be of maternal origin (Figure 3.12). X-inactivation studies were pursued as a clinical service through the Molecular Diagnostic Laboratory at BC Children's and Women's Health Center and were negative for preferential inactivation of one X chromosome in the mother. Affymetrix array results were available for this patient and were used to establish more precise breakpoints. The duplication was found to be approximately 600kb in size, from positions 86,005,960bp to 86,614,379bp. There is only one very small variant (<2kb) for this interval in the DGV (Figure 3.11). Investigations of Affymetrix array results did not reveal additional CNV regions of interest.

Overlap with PCNV Databases and Literature There is no overlap with other cases in the literature, the ACRD or the DECIPHER database.

Gene Content and Clinical Significance There are no genes in the interval, but two genes closely flank the duplication: dachshund homolog 2 (*DACH2*), 31kb proximally and kelch-like 4 (*KLHL4*), 45kb distally. The dup(X)(q21.2q21.31) in this patient is

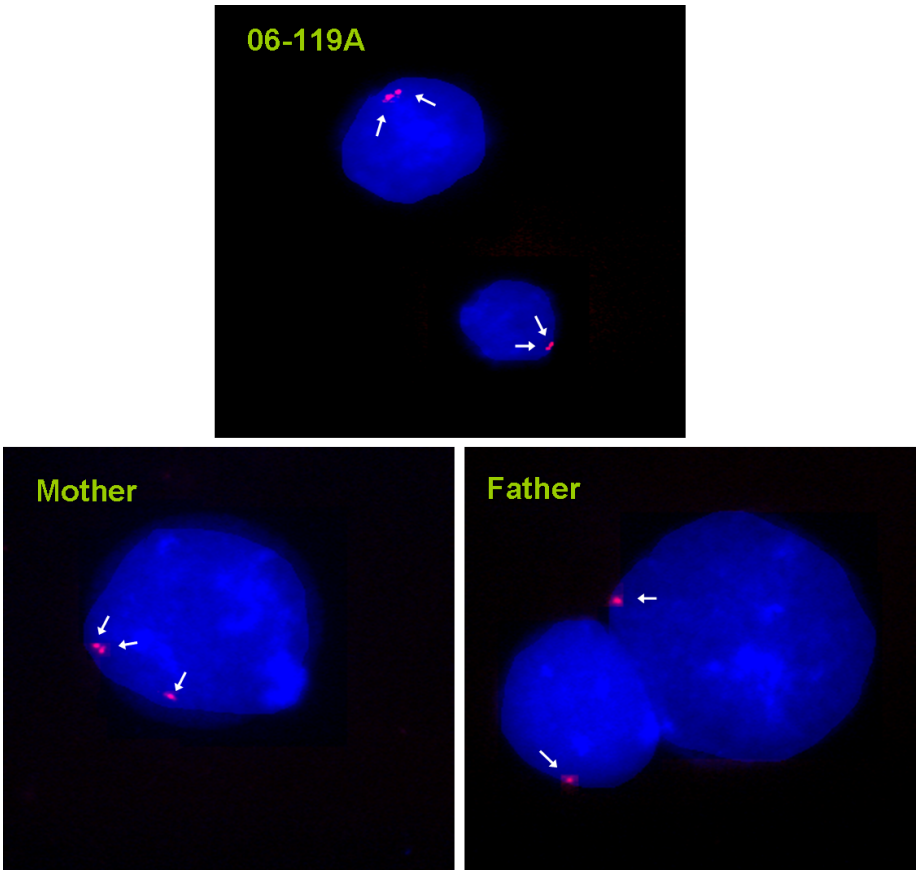

Figure 3.12: FISH Example of Inherited CNV: dup(X)(q21.2q21.31) in 06-119A

considered as being of "unknown significance", since it is maternally inherited in the absence of skewed X-inactivation, lacks any overlap in relevant rearrangement databases and does not contain any genes. Investigations of *DACH2* and *KLHL4* do not suggest that they would be good candidates for autism.

The gene *DACH2* encodes a protein similar to the Drosophila transcription factor Dachshund and in vertebrates, the gene is expressed in embryonic nervous tissue, sensory organs and limbs, a pattern similar to dachshund homolog 1 (*DACH1*), suggesting partially redundant roles in development [174]. Both *DACH1* and *DACH2* knock-out mice models do not show eye, limb or brain malformations [175, 176]. No behavioural

differences between mouse *DACH2* null mutants and their wild-type counterparts was noticed [176].

Much less is known about *KLHL4*, and the function of the protein is unknown. The gene was identified through a search for novel genes in the X-linked cleft palate and Ankyloglossia (CPX) critical region, and was found to be expressed in a range of fetal tissues including brain, palate, mandible and tongue [177]. It is not possible at this time to totally exclude the possibility that the duplication in this patient somehow leads to dysregulation of one or both of these genes, but more research into their function is needed before we can consider them good candidates for autism susceptibility.

Table 3.1: Summary of Disease-specific CNVs

Type of CNV	Pathogenic					Unknown significance	
Subject (Platform)	06-107A (Spectral)	07-93A (Agilent)	06-43A (Agilent)		07-102A (Agilent)	06-134A (Spectral)	06-119A (Spectral)
CNV	Del(9) (q34.11q34.13)	Del(8) (q24.3)	Dup(16) (p11.2)	Dup(16) (p13.2)	Del(15) (q25.2q25.3)	Del(20) (p11.21)	Dup(X) (q21.2q21.31)
Position (kb)	130,525-133,061	141,281-141,457	29,500-30,027	8,709-8,926	82,902-83,522	24,907-25,708	86,006-86,614
Size (kb)	2,535	176	527	217	620	801	608
Confirmation method	FISH, Affymetrix	FISH	Custom Agilent	Custom Agilent	FISH, Affymetrix	FISH, Affymetrix	FISH, Affymetrix
Origin	*de novo*	Maternal	*de novo*	Maternal	Maternal	Paternal	Maternal
No. of genes	39	1	25	5	10	11	0
Candidate genes	*ABL1*, *TOR1A* *TOR1B*	*NIBP* *KCNK9*[a]	*MAZ*, *DOC2A* *SEZ6L2*, *MVP*, *QPRT* [b]	*ABAT*	-	-	-
DECIPHER cases	-	-	5 Dels 3 Dups	2 Dups	2 Dels	1 Del	-
ACRD cases	-	-	17 Dels 6 Dups	-	2 Dels 1 Dup	1 Del	-

[a] gene is 500kb distal and potentially imprinted
[b] *MVP* and *QPRT* are candidates for epilepsy

Table 3.2: Summary of Clinical Features of Subjects with Disease-specific CNVs

	Pathogenic CNVs				Unknown significance	
Subject	06-107A	07-93A	06-43A	07-102A	06-134A	06-119A
CNVs Detected	Del(9) (q34.11q34.13)	Del(8) (q24.3)	Dup(16)(p11.2) Dup(16)(p13.2)	Del(15) (q25.2q25.3)	Del(20) (p11.21)	Dup(X) (q21.2q21.31)
Age (sex)	3 (M)	3 (M)	13 (F)	3 (M)	15 (M)	3 (M)
Family type	SPX	SPX	SPX	SPX	SPX	SPX
Diagnosis	PDD-NOS	AD	PDD-NOS	ASD	AD	AD
Modified de Vries score[a]	5	4	4	5	4	3
Growth parameters (H, W, OFC)	97,97,98	70-90,98,50	10,10,50	75-95,25,98	50,50,98	50-75,90,50
MPAs	21	7	12	8	11	5
MCAs	2	-	-	2	-	-
Craniofacial Dysmorphisms	13	5	8	4	10	4
Systemic Anomalies [b]	10	2	4	6	1	1
Health issues	4	-	-	-	-	-
Seizures	-	-	+	+	-	-
ID Severity [c]	Borderline to Mild	Mild	No ID	Mild	Moderate	Mild
Obstetric Complications	4	-	2	9	-	-
Birth weight %ile	$\geq$97	10-50	$\sim$90	$\sim$90	$\sim$90	50-90

[a] Modified=original, except 07-93A: unmodified=6 (positive family history of ID)

[b] Includes MCAs

[c] Severe: IQ=20-35; Moderate: IQ=35-50; Mild: IQ=50-70

Chapter 4

Whole-body Phenomic Analyses of ASDs and their Genomic Correlates

4.1 Cohort and Clinical Demographics

In the course of this study, several clinical variables, growth parameters and physical features were recorded for each of the 40 patients included in this study. The cohort was composed of 34 males and 6 females, for an overall male-to-female ratio (M:F) of 5.7:1. This ratio may be artificially biased towards a higher M:F ratio as individuals to be included in the study were non-randomly chosen to make sure that a significant proportion (at least 10 individuals) would present with more essential forms of ASDs (devoid of significant dysmorphology and/or ID). Populations with essential forms of autism, including normal intelligence, have higher M:F ratios that deviate from the usual 4:1 ratios [4, 9, 97]. For purposes of patient selection, probands with phenotype scores $\leq$3 excluding points for family history (see Table 2.1) were considered as having essential autism while those with scores $\geq$4 were considered complex cases. The 40 subjects studied included 14 individuals with essential autism (3 females and 11 males) and 26 individuals with complex autism (3 females and 23 males).

Noted co-morbidities included epilepsy (17.5%), ID (IQ <70) (72.5%) and significant health issues (35%) as well as MCAs in a minority of individuals (22.5%) (Table 4.1). The most common health issues reported were chronic constipation, hypotonia and food allergies, while major anomalies were very varied, with a third consisting of various brain anomalies detected with MRI or CT scans. Minor craniofacial anomalies were observed in 90% of cases while other minor systemic anomalies were present in 83%, with 75% of subjects having both (Table 4.1). The mean number of MPAs (craniofacial + systemic) was 8.60 for the whole cohort, with a standard deviation (SD) of 5.29. Phenotype scores were computed (see Section 2.1) and the mean score was 4.25 with a SD of 1.29 (Table 4.2). Most patients had scores between 3 and 5, with only 6 individuals above and 2 below.

The study of growth parameters revealed that macrocephaly (OFC $\geq$98%ile) is quite common, with 25% of subjects presenting with the phenotype, while microcephaly (OFC $\leq$2%ile) was only observed in one individual (2.5%). Interestingly, and in line with our previous findings [102], an enrichment for individuals with large heights ($\geq$97%ile) (15%, two-tailed p=0.02) and large weights ($\geq$97%ile) (20%, two-tailed p=0.002) was seen in this cohort. Subjects delivered prematurely ($\leq$36 weeks) composed 12.5% of cases and 5% and 8% of the group were found at birth to be small-for-gestational age ($\leq$5%ile) and large-for-gestational age ($\geq$95%ile), respectively (Table 4.1).

4.2 Characterization of CNVs in ASD Subjects

Amongst 40 subjects with ASD, 7 unique, disease-specific CNVs were identified (see Table 3.1) that are categorized as potentially pathogenic (PCNVs) in 6 probands (15%); 5 males and 1 female (M:F of 5:1). Of those 7 PCNVs, 2 might be better categorized as of "unknown significance" (see Chapter 3), but they are nevertheless considered as PCNVs for analysis in this chapter. I based this decision on the fact that the numbers are small, that the changes are indeed unique and disease-specific and on the observation that preliminary analyses excluding both cases did not significantly alter the demographics of the group carrying PCNVs on any measure of co-morbidity, physical dysmorphology or growth parameters. All PCNVs were identified in SPX cases, for an overall frequency of 6/24 (25%). Four changes were familial, one was a maternally inherited X-linked

Table 4.1: Cohort and Clinical Demographics: Discrete Variables

Phenotypes		No. (%) of cases with phenotype	No. (%) of cases with PCNVs with phenotype	No. (%) of cases without PCNVs with phenotype	two-tailed p-value (Fisher's test)
Total cases		40	6 (15)	34 (85)	
Autism Subtype	Essential	14 (35)	1 (17)	13 (38)	0.40
	Complex	26 (65)	5 (83)	21 (62)	
Gender	Male	34 (85)	5 (83)	29 (85)	1.00
	Female	6 (15)	1 (17)	5 (15)	
ID	ID (IQ<70)	29 (72.5)	5 (83)	24 (71)	0.46 (one-tail)
	No ID (IQ>70)	11 (27.5)	1 (17)	10 (29)	
Epilepsy		7 (17.5)	2 (33)	5 (15)	0.57
OFC	macrocephaly (≥98%ile)	10 (25)	2 (33)	8 (24)	1.00
	microcephaly (≤2%ile)	1 (2.5)	0	1 (3)	0.85 (one-tail)
Stature	Small height (≤3%ile)	2 (5)	0	2 (6)	1.00
	Large height (≥97%ile)	6 (15)[a]	1 (17)	5 (15)	1.00
	Small weight (≤3%ile)	2 (5)	0	2 (6)	1.00
	Large weight (≥97%ile)	8 (20)[b]	2 (33)	6 (18)	0.58
≥2 craniofacial anomalies		36 (90)	6 (100)	30 (88)	1.00
≥1 systemic anomalies		33 (83)	6 (100)	27 (79)	0.51
≥2 craniofacial + ≥1 systemic anomalies		30 (75)	6 (100)	24 (60)	0.31
only one category present		10 (25)	0	10 (29)	
premature delivery (≤36 weeks)		5 (12.5)	0	5 (15)	0.58
small-for-gestational age (≤5%ile)		2 (5)	0	2 (6)	1.00
large-for-gestational age (≥95%ile)		3 (8)	1 (17)	2 (6)	0.39
MCAs present		9 (22.5)	2 (33)	7 (21)	0.60
Significant health issues present		14 (35)	1 (17)	13 (38)	0.40

[a] significantly different from normal curves (p=0.02)
[b] significantly different from normal curves (p=0.002)

CNV in a male and 2 were *de novo* (2/24, 8.3%) (see Chapter 3 and Table 3.1).

Detection rates for unique CNVs were not different between platforms, with 3/22 cases detected by Spectral (13.6%) and 3/18 cases detected by Agilent arrays (16.7%) showing disease-specific changes. Six out of seven regions contained genes, and one locus is a single-gene containing region (see Chapter 3 and Table 3.1). In addition to disease-specific PCNVs, we identified 52 putative BCNVs by Spectral and 112 BCNVs

Table 4.2: Cohort and Clinical Demographics: Continuous Variables

Phenotype	Cohort Mean (SD)	Cases with PCNVs Mean (SD)	Cases without PCNVs Mean (SD)	two-tailed p-value (Mann-whitney test)
No. of MPAs	8.60 (5.29)	10.67 (5.68)	8.23 (5.22)	0.33
No. of distinguishing MPAs	7.75 (4.69)	9.33 (4.32)	7.47 (4.75)	0.37
No. of OCs	2.28 (1.95)	2.5 (3.56)	2.24 (1.60)	0.57
phenotype score	4.25 (1.29)	4.5 (1.04)	4.21 (1.34)	0.58
phenotype score excluding family history	3.66 (1.25)	4.17 (0.75)	3.56 (1.31)	0.25
No. of craniofacial anomalies	6.48 (4.11)[a]	7.33 (3.67)	6.32 (4.22)	0.52
No. of systemic anomalies	2.55 (2.34)	4.00 (3.52)	2.29 (2.04)	0.23

[a] excluding macrocephaly

by Agilent arrays, for a total of 164 copy number variable regions. 95 were deletions and 69 were duplications for an overall del:dup ratio of 1.38:1. Sizes of BCNVs cannot be accurately ascertained when detected by Spectral arrays, but data from Agilent showed that deletions are significantly smaller than duplications (Mann-Whitney two-tailed p=0.003).

4.3 Genome-phenome Comparisons

4.3.1 Correlation of PCNVs with Specific Phenotypes

Dysmorphology scores were computed by adding all MPAs reported for each proband and then MPA scores were compared for patients with PCNVs to those without (see Table 4.2). Phenotype scores were also computed (see Section 2.1) and the same comparison was performed. Both scores showed an increase in patients with PCNVs (29.6% for MPAs and 6.9% for phenotype score) compared to patients without PCNVs, but both associations were not statistically significant (p=0.33 and p=0.58, respectively)(Table 4.2). Furthermore, adjusting MPA scores to reflect only those MPAs that had been shown by Miles et al. 2008 [26] to distinguish dysmorphic from non-dysmorphic ASD individu-

als did not decrease the p-value. The list of features was kindly provided to us by the authors of this study.

Because array CGH studies of subjects with ASDs have consistently shown a higher discovery rate of PCNVs in individuals without family history of autism (Table 1.1), I reasoned that a phenotype score without points for family history would be a more sensitive indicator of PCNV risk in autistic individuals. Indeed, when family history was taken out of the equation (modified phenotype score), a more significant increase in phenotype score in the group of patients with PCNVs (17.1%) and a more significant association between the presence of PCNVs and higher phenotype score were discovered (p=0.25 compared to p=0.58)(Table 4.2).

Macrocephaly evaluated on its own did not show a significant association with PCNVs (p=1.00), nor the items $\geq$2 craniofacial anomalies (p=1.00) or $\geq$1 systemic anomalies (p=0.51) (Table 4.1). When patients with both craniofacial and systemic anomalies were separated from those with only one or the other, those with "whole-body" involvement had a slightly increased risk for PCNVs (RR=1.42, 95%CI (1.14-1.76)), even though the p-value was not significant (p=0.31) (Table 4.1). Probands with PCNVs also had a 74.7% increase in the number of systemic anomalies compared to individuals without PCNVs (mean numbers of 4.00 and 2.29 respectively), but again the p-value was not significant (Table 4.2). Microcephaly had been shown to be correlated to PCNVs risk in our previous group of patients [102] but it was only observed in one patient and did not seem to correlate with increase PCNV risk in this small cohort (one-tailed p=0.85).

No other significant associations between PCNV detection and specific phenotypes were found in this group of patients (Table 4.1), including the presence of ID (one-tailed p=0.46) or specifically moderate ID (one-tailed p=0.59), which had shown associations in a previous study [102]. Seizures and the number of obstetric complications did not differ between those individuals with PCNVs and those without (both p=0.57)(Table 4.1 and Table 4.2) and pathogenic CNVs did not associate with one gender specifically (p=1.00).

4.3.2 Correlation of Dysmorphology Measures with other Phenotypic Features

Individuals with a higher number of MPAs are etiologically distinct [9, 10, 30]. Individuals with significant dysmorphologies are reported to have lower IQs, more seizures, a higher proportion of cases without family history, a higher proportion of females, more brain anomalies and a higher number of hospitalizations. I was thus interested in seeing whether the group with the highest number of anomalies would be different from the group with the lowest.

Probands were separated into approximate MPA score quartiles (ties in scores did not allow more precision) and both groups were compared on a number of items (Table 4.3). Both MPA quartiles were made of 12 subjects (30%). The same comparisons were performed in subgroups based on modified phenotype scores (Table 4.4). The complex group comprised individuals with modified scores of $\geq$4 and the essential group included subjects with scores of 0,1,2 or 3. Groups were separated in this manner because most individuals with 3 points or less (10/14, 71%) have a very simple phenotypic presentation that excludes either craniofacial or systemic anomalies, or present with only 1 systemic anomaly (4/14, 29%) while individuals with $\geq$4 points all had both craniofacial and systemic anomalies. In addition, individuals with 4 points clearly included more complex cases, with only 1/26 individual with less than 6 MPA, which was the cutoff in Miles et al. in 2000 and 2005 [9, 10] for the complex subgroup.

The presence of seizures or MCAs was not higher in the highest MPA group, but the presence of significant health issues, ID or moderate ID all showed trends towards association with high MPA scores (one-tailed p=0.20, p=0.19 and p=0.16, respectively) (Table 4.3). When moderate ID and/or health issues were looked at in combination, the association became significant (one-tailed p=0.05) (Table 4.3). Four individuals in the lowest MPA group showed one or both of these features (33%) versus 9/12 (75%) individuals in the highest MPA group.

In this small sample, females were not significantly more represented in the highest group (18%) or under represented in the lower group (17%) but individuals without family history of autism seemed to be over represented when the highest and lowest

Table 4.3: Clinical Demographics of Complex Individuals: MPA Scores

Phenotype	No. of cases (%) with phenotype in lowest MPA quartile (12 cases)	No. of cases (%) with phenotype in highest MPA quartile (12 cases)	one-tailed p-value (Fisher's test)
ID (IQ<70)	7 (58)	10 (83)	0.19
Moderate ID (IQ<55)	1 (8)	4 (33)	0.16
Health issues present	3 (25)	6 (50)	0.20
MCAs present	2 (17)	3 (25)	1.00 (two-tailed)
Epilepsy	1 (8)	2 (17)	0.47
Mod. ID or Health issues	4 (33)	9 (75)	**0.05**
Female sex (15%)	2 (17)	2 (17)	0.67
SPX (60%)	5 (42)	9 (75)	0.11

quartiles of MPA scores were compared (p=0.11). Finally, the number of obstetric complications was not correlated with higher MPA score (two-tailed p=0.38) and neither was the absolute number of health issues (two-tailed p=0.24) (data not shown). It was not possible to statistically look at brain anomalies in relation to MPA scores because only 3 individuals were shown to have brain anomalies through imaging studies, but interestingly 2/3 of these subjects were the two patients with the highest MPA scores.

When modified phenotype scores were looked at, this complex group also showed a trend for the presence of health issues (p=0.17). On the other hand, when moderate ID was looked at it was not significant (p=0.31) although the presence of any degree of ID was in fact associated with the complex subgroup, with only 4/26 (15%) of complex individual having normal intelligence while 7/14 (50%) of subjects with essential autism did not show cognitive deficits (p=0.03) (Table 4.4). Again major congenital anomalies and seizures were not significantly more prevalent in either subgroup, and females were not over represented in the complex subgroup. However, individuals without family history of autism were again over represented in the complex subgroup (p=0.10) (Table 4.4). Finally, the number of obstetric complications was not correlated with either subgroup (two-tailed p=0.52) but the number of health issues suggested a positive trend (two-tailed p=0.14) (data not shown).

Table 4.4: Clinical Demographics of Complex Individuals: Modified Phenotype Scores

Phenotype	No. of cases (%) with phenotype in essential subgroup (14 subjects)	No. of cases (%) with phenotype in complex subgroup (26 subjects)	one-tailed p-value (Fisher's test)
ID (IQ<70)	7 (50)	22 (85)	**0.03**
Moderate ID (IQ<55)	2 (14)	7 (27)	0.31
Health issues present	3 (21)	11 (42)	0.17
MCAs present	1 (7)	8 (31)	0.66 (two-tailed)
Epilepsy	2 (14)	5 (19)	0.53
Female sex (15%)	3 (21)	3 (12)	0.35
SPX (60%)	6 (43)	18 (69)	0.10

4.3.3 Correlation of Total CNV Load with Clinical Features

The average number of CNVs per person was 2.4 for Spectral arrays and 6.2 for Agilent arrays. Small numbers precluded the use of a non-parametric statistical test, but these numbers did not seem to differ from those found in individuals with PCNVs (which are expected to have 1 extra CNV). The 3 cases with PCNVs detected by Spectral array had an average of 3.0 CNVs, while the 3 cases with PCNVs detected by Agilent arrays had on average 8.3 CNVs.

Since individuals with PCNVs had a 29.6% increase in the number of MPAs and 17.1% higher modified phenotype scores compared to probands without PCNVs, these two measures were looked at in relation to total CNV load. Again, small numbers restricted the ability to perform certain comparisons as there was only 4 probands investigated with Spectral arrays with modified phenotype scores of 0-3 and only 4 probands in the highest MPA score quartile investigated with Agilent arrays. It was thus only possible to investigate MPA score quartiles in the Spectral array group and modified phenotype scores in the Agilent group (0-3 versus 4-7). While the association was not present between higher MPA score and higher CNV load in the Spectral group (Mann-whitney two-tailed p=0.38), there was a trend for a higher number of CNVs detected by Agilent in persons with modified phenotype scores of 4-7 versus those with scores of 0-

3 (Mann-whitney two-tailed p=0.08). Finally, absence of family history and presence of moderate ID and/or health issues were not found to be associated with a higher number of CNVs in either platform.

4.4 Gene Prioritization

The computational candidate gene prioritization software, Endeavor, was used to assist with further pinpointing potential ASD-related genes that share similarity to selected autism-related reference genes [122]. The rank of prioritized genes in a specific locus is mainly dependent on the training set of genes and the selection of sources in the databases. A list of 19 autism-related genes from a recent review [123] was used as the training set and all models were selected in the analysis. Because this list of gene contains loci for which association with autism has not been definitively proven, and lacks many other possibly interesting genes that have not been studied yet, I felt it important to complement this computed analysis with manual investigations using web databases (UCSC, NCBI, and OMIM) and available literature. Such a manual investigation was undertaken when less than 5 genes were present in a region, or when Endeavor did not pinpoint any interesting candidates. In total, 9 good candidate genes were identified in three patients, involved in axon guidance, neurite outgrowth, synaptic structure, function and plasticity, neurotransmitter release, glutamatergic and GABAergic transmission and neuronal differentiation. For a full discussion of each candidate gene, please refer to individual cases in Chapter 3.

Chapter 5

Discussion

5.1 Characterization of CNVs in ASD Subjects

In this work 40 individuals diagnosed with ASD were studied using array CGH technology and putative PCNVs were detected in 6 subjects (15%), for which I provide detailed phenotypic and genomic information. All seven changes detected in those 6 individuals were identified in cases without family history of autism (6/24, for a detection rate of 25% in SPX families) and 2 were *de novo* (2/24 for a detection rate of 8.3% in SPX families). This *de novo* detection rate of 8.3% in simplex families and the fact that more changes are detected in SPX than MPX families is in general agreement with previous genome-wide array CGH studies of autism (see Table 1.1). The overall detection rates of 15% and 25% (in SPX families) are higher than most studies, but if both changes of "unknown significance" are taken out of the equation, these rates decrease to 10% and 17%, highlighting the reduced power of this small sample to establish accurate PCNV risk estimates in the more general ASD population.

Inability to detect disease-specific CNVs in MPX-I and MPX-E families in this study is probably due again to the small number of individuals investigated. We have previously reported on a cohort of 100 individuals diagnosed with ASD in which we reported rates of PCNVs in MPX-E families that were more similar to SPX detection rates, at

12.5% [102]. Only six families with extended family history of autism were included in this study, and larger cohorts with detailed recording of extended family histories will be needed to resolve this discrepancy. Similarly, only 6 females were included in this study, and this limited number probably explains why, in contrast to every other study, a significantly reduced M:F ratio in PCNV cases was not found (Table 1.1).

In this work, investigations using either the Spectral Genomics 1Mb array platform or the Agilent Technologies 105K platform revealed a total of 7 disease-specific and potentially pathogenic CNVs. One of those changes is a large and *de novo* deletion at 9q34.11q34.13, and I believe it represents a novel microdeletion syndrome associated with autism (Section 3.1). One patient carries two duplications, at 16p13.2 and 16p11.2 (Section 3.3), the latter of which has been reported in numerous individuals, with a change being present in approximately 1% of individuals at a rate significantly higher than control populations [114]. Other changes include a maternally inherited deletion at 15q25.2q25.3 (Section 3.4), a region that has been recently highlighted as being one of a number of copy number variable loci possibly involved in susceptibility for autism, ID and schizophrenia [145] and for which there is overlap with previously reported cases in the literature and web databases (Section 3.4). The fourth patient carries a small maternal deletion at 8q24.3, a region close to at least one very interesting gene exclusively expressed from the maternal allele and which disrupts another interesting gene with unknown imprinting status (Section 3.2). Lastly two other cases (Section 3.6 and Section 3.5) were found to have inherited changes with less obvious clinical significance. For a full discussion of each genomic region and possible phenotype-genotype relationships, please refer to the individual case discussions in Chapter 3.

Of the seven disease-specific and potentially pathogenic CNVs discovered in this study, 5 are inherited from normal parents, including one duplication on the X chromosome in a male that was inherited from the mother, who does not show preferential inactivation of one X chromosome. In addition, one small deletion on 8q24.3 in patient 07-93A is inherited from his normal mother and includes a gene that is potentially imprinted and could be expressed exclusively from the maternal allele. CNVs inherited from unaffected parents could show incomplete penetrance because of such imprinting influences, or in the case of an inherited deletion, could uncover deleterious recessive alleles on the intact chromosome. Gene-gene and gene-environment interactions could

also explain why some chromosome imbalances do not lead to a clinically recognizable phenotype in all individuals that carry them.

All the above-mentioned disease-specific CNVs were investigated for the presence of potentially good candidate genes for autism using a software called Endeavor and manual investigations using the NCBI, UCSC and OMIM online databases (see Section 2.5). In total, 9 good candidates were identified (Table 3.1), either for their possible relationship to ASD susceptibility, or in one proband for their putative role in epilepsy. Candidate genes pinpointed in this study encode molecules involved in axon guidance, neurite outgrowth, synaptic structure, function and plasticity, neurotransmitter release, glutamatergic and GABAergic transmission and neuronal differentiation. For a full discussion of each candidate gene, please refer to individual cases in Chapter 3.

5.2 Correlation of PCNVs with Specific Phenotypes

As part of this study I also undertook an investigation into possible medical, physical and obstetric indicators of PCNV risk in autism spectrum disorders. The results do not indicate that pregnancy, labor or neonatal complications are good etiologic factors in autism, in agreement with previous studies suggesting that obstetric complications are an epiphenomenon of autism or genetic factors predisposing to the disorder, and not on the causal pathway to developing the condition [42, 43].

Dysmorphology scores were computed by adding all reported MPAs for each proband and compared in patients with PCNVs and those harbouring only BCNVs. Phenotypic scores were then computed based on de Vries criteria (see Section 2.1) and the same comparison was performed. Based on those measures, increases of 29.6% in MPA scores and 6.9% in phenotype scores were found in individuals with PCNVs when compared to probands with BCNVs only. Adjusting MPA scores to reflect only those reported by Miles et al. 2008 as distinguishing dysmorphic from non-dysmorphic ASD subjects did not strengthen the statistical significance of this trend (p=0.33 and p=0.37, respectively), but adjusting phenotype scores to exclude points for positive family history did

(17.1% compared to 6.9% increase and p=0.25 compared to p=0.58), in accordance with the fact that an absence of family history is a good indicator of PCNV risk (see Table 1.1 and Section 1.4.4).

Specific items from the phenotype scoring checklist, as well as several growth parameters and medical co-morbidities were investigated separately for any association with increased PCNV risk and none of them were shown to be significantly linked to the presence of PCNVs in this small cohort of patients. In contrast to our previous cohort study of a 100 patients [102], I did not find an association between the presence of ID and the detection of PCNVs (one-tailed p=0.46). Whereas in the previous study none of the subjects without ID showed detectable PCNVs, 1 out of 6 patients with PCNVs in this cohort did not have a secondary diagnosis of ID. Furthermore, I did not find an association of moderate to severe ID with the presence of PCNVs in this subject group (one-tailed p=0.59), which also contrasts with our previous results and the results of a previous genome-wide array CGH study of patients with ID [178], in which subjects with more severe ID were found to have a higher risk of PCNVs compared to those with mild or borderline ID. Further studies are thus needed with larger samples of patients with known ID status in order to better define the role of cognitive functioning on the risk of PCNVs, and to determine relative risk and positive predictive value of ID on the risk of carrying PCNVs in ASDs.

The presence of seizures, MCAs and significant health issues was not associated with PCNV risk, and neither was prematurity or being small-for-gestational or large-for-gestational age at birth (see Table 4.1). In addition, macrocephaly, microcephaly, large or small heights and large or small weights were not associated with the presence of PCNVs. Results for seizures, macrocephaly, height and weight are consistent with those of our previous cohort, while major congenital anomalies and significant health issues were not studied previously [102]. Microcephaly was shown to be associated with PCNVs in our previous study, and has been shown by Miles et al. 2005 [10] to be a good etiologic indicator for ASD , but I did not repeat these findings in this work, most probably because only one of the 40 individuals studied presented with the phenotype.

When specific items from the phenotype scoring checklist were looked at, the items $\geq$2 craniofacial anomalies and $\geq$1 systemic anomalies did not significantly associate

with the presence of PCNVs in isolation, but when patients were separated based on whether they exhibited both types of anomalies or only one, those with "whole-body" involvement had a slightly elevated relative risk of harbouring PCNVs (RR=1.42, 95%CI (1.14-1.76)) and individuals with PCNVs also had a 74.7% higher number of systemic anomalies (4.00 compared to 2.29) (Table 4.1 and Table 4.2). This is consistent with the idea that contiguous gene deletions and duplications are expected to impact more body areas than point mutations or smaller single-gene containing CNVs, which possibly explain the etiology in those cases without PCNVs detectable with both platforms used.

5.3 Correlation of Dysmorphology Measures with other Phenotypic Features

Next I was interested in whether those individuals exhibiting more complex dysmorphology based on MPA scores and modified phenotype scores would be different from those exhibiting low scores, as previous research by Miles et al. in 2000 and 2005 [9, 10], and Links et al. 1980 [30] had shown that ASD subjects with significant dysmorphology are etiologically different and have lower outcome measures. More specifically, those individuals were shown to have lower IQs, more seizures, more brain anomalies, more frequent hospitalizations, and a higher proportion of females and individuals without family history.

I thus proceeded to separate my group of 40 patients into approximate MPA score quartiles and compared the highest quartile with the lowest quartile with respect to ID, moderate to severe ID, seizures, MCAs, significant health issues that could lead to hospital visits and finally female gender and absence of family history. I performed the same comparisons in subgroups based on modified phenotype scores. The complex group comprised individuals with modified scores of $\geq$4 and the essential group included subjects with scores of 0,1,2 or 3.

Even in such a small sample, those individuals with more minor dysmorphic features, either based on higher MPA score or a modified phenotype score $\geq$4 are indeed different from those with a less complex presentation. Specifically, results show trends of association for ID, moderate to severe ID, and the presence of significant health is-

sues with the highest quartile of MPA scores (p=0.19, p=0.16 and p=0.20, respectively). When both moderate to severe ID and significant health issues are combined to define individuals positive for one or both items, the association becomes significant (p=0.05) (Table 4.3).

When individuals are subgrouped based on modified phenotype scores, the presence of health issues is still showing a trend for an association with the complex autism subtype (p=0.17), but moderate ID does not (p=0.31). In this complex subgroup however, individuals without ID are significantly under represented (p=0.03) (Table 4.4). Furthermore there is a trend for an absolute increase in health issues in individuals in the complex subgroup based on modified phenotype score (p=0.14). On the other hand, female gender, seizures and major anomalies did not show an association on either subgrouping measures possibly because of small numbers in each category. Absence of family history did show a trend of association with the highest MPA quartile (p=0.11) and the complex subgroup (p=0.10) and interestingly, we report that both individuals with the highest absolute MPA scores composed the majority of our very small number of patients with brain anomalies (2/3). Finally, there was no association between higher MPA scores or modified phenotype scores $\geq$4 and the number of obstetric complications (p=0.38 and p=0.52, respectively), a finding consistent with a study by Links et al. 1980 [30].

5.4 Correlation of Total CNV Load with Clinical Features

Lastly, because results had shown an increased number of MPAs and higher modified phenotype scores in individuals with PCNVs, these two measures were looked at for a possible association with total CNV load, which is the total number of pathogenic and benign CNVs found in one individual. Since the average number of CNVs detected per person was much higher in those cases investigated with the Agilent arrays, I had to separate the cohort into two groups to perform this analysis. The Agilent group was composed of 18 individuals while the Spectral group was composed of the remaining 22, with 3 people in each group found to have PCNVs. These small numbers restricted the ability to perform certain statistical analyses, but I was able to look at MPA score

quartiles in the Spectral group and modified deVries scores in the Agilent group (0-3 versus 4-7 points).

When CNV load had been investigated in relation to phenotype scores in our previous study of a 100 ASD subjects [102], we had found that those individuals with scores of 3 had significantly more chances of having no detectable CNVs (by Spectral) than those with scores of $\geq$4 (all 100 subjects in this study had scores of at least 3). In the current study, an association was found between higher CNV load and higher (modified) phenotype scores only in the Agilent group (p=0.04). In the Spectral group, the number of patients with scores of 0-3 was insufficient to perform a statistical analysis, but higher MPA scores did not associate with a higher CNV load.

Since modified phenotype scores were looked at in the current study and the group of subjects was biased to include individuals with more simple phenotypic presentation (compared to our previous group of patients), it is hard to explain the discrepancy of results regarding a possible association between phenotype and CNV load. However, the association between modified phenotype scores and higher CNV load in the Agilent group might suggest that when a higher resolution platform is used and more CNVs are uncovered, an additive effect of small benign CNVs on minor abnormal phenotypic variants can be seen.

On the other hand, absence of family history, presence of moderate ID and/or health issues were not found to be associated with a higher number of CNVs in either platform, consistent with results in our previous larger group of ASD subjects [102]. I also looked at whether those individuals with PCNVs would have a higher number of detectable BCNVs, again separating the cohort in two groups. Consistent with results from our previous cohort, I did not find a consistent increase in either group (Spectral or Agilent), apart from the small increase of 1 extra CNV that is expected in patients carrying one (and in one Agilent case 2) PCNVs.

5.5 Limitations and Future Directions

There are several limitations to this study that need to be taken into account when interpreting the results. First and foremost, the small size of the sample is limiting the statis-

tical power to detect phenomic predictors of PCNV risk. With only 6 cases harbouring pathogenic copy number variants, it is difficult to conclude anything about phenotype-genotype correlations with any mathematical certainty, and the next step will be to combine the 40 patients analyzed here with the previous cohort of 100 patients that our laboratory has studied in the past [102]. This larger combined group of patients with detailed phenotypic and genomic information will provide better statistical power to identify potential phenomic predictors for a CNV-based etiology of ASDs. However, the long term goal is to analyze a much larger group of patients, with hundreds of cases and dozens of individuals with potentially clinically relevant changes.

Another very important limitation comes from the uncertainty when it comes to interpreting the clinical significance of any array-detected copy number variant. We are not currently able to distinguish between benign and pathogenic CNVs for all cases, nor do we fully understand the relationship between many presumably "benign" CNVs and human disease. Variants reported in the Database of Genomic Variants were not followed up in this study, but some of them might be linked to autism susceptibility in the future, and such CNV-genome wide association studies will be needed to understand the role of so-called benign CNVs in disease.

Future investigations of genes within disease-specific CNVs detected in the present cohort are also needed, and should include investigations at the RNA and protein expression levels, the creation of mouse models, and bioinformatics analyses of pathways involved, which could lead to the identification of other candidate genes that could be involved in ASD pathogenesis. Furthermore, each PCNV should be investigated for its possible recurrence in larger autistic samples.

Chapter 6

Conclusions and Future Directions

Over the last few years, whole-genome array CGH analyses have greatly contributed to our understanding of the etiology of ASD and participated in the development of a revised view of the genetics of the disorder that acknowledges the role of rare submicroscopic copy number variants in a more significant proportion of idiopathic cases than what was previously hypothesized. But to grasp the full spectrum of genomic changes in a disorder as complex as ASD will require a continuous unmasking and characterization of disease-specific CNVs in probands and their families and a better understanding of what constitutes normal variation.

Our understanding of specific biological and molecular pathways involved in the pathophysiology of autism is still limited, but the identification of a number of genes through the use of various methodologies is starting to show convergence, pointing to dysregulated protein translation [76] and aberrant activity-dependent changes in neuronal function [94]. Our work adds to this discussion by pinpointing nine additional candidate genes that could be involved in the susceptibility of autism.

This work has also demonstrated the importance of comprehensive phenotypic evaluations of subjects undergoing array CGH investigation. Performing and reporting detailed developmental, physical, cognitive and behavioural profiles of probands is critical for defining novel microdeletion and microduplication syndromes, and for the compar-

ison of cases presenting with overlapping changes in different studies. Current results indicate that individuals with more minor dysmorphic features are at an increased risk of carrying pathogenic changes and perhaps that a clinical presentation including minor physical variants from different body areas is suggestive of changes involving multiple contiguous genes. Investigations into etiologic indicators for pathogenic DNA dosage changes is also likely to improve our understanding of natural history and causal pathways leading to autism whilst enriching the evidence base for genetic counseling that awaits refinement from the growing spectrum of ASD-associated microdeletion and microduplication syndromes.

Interpreting the effect on human phenotypes of smaller or inherited CNVs, or those reported very rarely in ill-defined control populations is a challenge, and one that we are increasingly facing as investigators are turning to higher resolution arrays. Investigations into the functional significance of such DNA changes will be key to our understanding of genome-phenome links and to the identification of the specific molecules involved in the diseased molecular pathways. Research using whole-genome expression arrays, the generation of mouse models and research into imprinting defects in autism are critically needed in the future so that we can more fully understand what autism is at the biological level and better diagnose and treat the individuals living with this spectrum of disorders.

Bibliography

[1] de Vries, BBA, White, SM, et al. (2001) Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *J Med Genet* **38(3)**:145–150 → pages vii, 28, 29

[2] American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders DSM-IV Fourth Edition (American Psychiatric Publishing) → pages 1, 56, 71

[3] Fombonne, E (2005) Epidemiology of Autistic Disorder and Other Pervasive Developmental Disorders. *J Clin Psychiatry* **66(10)**:3–8 → pages 2

[4] Chakrabarti, S and Fombonne, E (2005) Pervasive Developmental Disorders in Preschool Children: Confirmation of High Prevalence. *Am J Psychiatry* **162(6)**:1133–1141 → pages 2, 80

[5] Szatmari, P, Jones, MB, et al. (1998) Genetics of Autism: Overview and New Directions. *J Autism Dev Disord* **28(5)**:351–368 → pages 2, 3

[6] Simonoff, E (1998) Genetic Counseling in Autism and Pervasive Developmental Disorders. *J Autism Dev Disord* **28(5)**:447–456 → pages 2, 3

[7] Persico, AM and Bourgeron, T (2006) Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends Neurosci* **29(7)**:349–358 → pages 2, 10

[8] Geschwind, DH (2008) Autism: Many Genes, Common Pathways? *Cell* **135(3)**:391–395 → pages 2, 3, 10, 20

[9] Miles, H and Hillman, E (2000) Value of a clinical morphology examination in autism. *Am J Med Genet* **91(4)**:245–253 → pages 2, 6, 7, 80, 85, 93

[10] Miles, JH, Takahashi, TN, et al. (2005) Essential versus complex autism: Definition of fundamental prognostic subtypes. *Am J Med Genet* **135A(2)**:171–180 → pages 2, 7, 85, 92, 93

[11] Schaefer, GB and Mendelsohn, MJ (2008) Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. *Genet Med* **10(1)**:4–12 → pages 2, 4

[12] Schaefer, GB, Mendelsohn, MJ, et al. (2008) Clinical genetics evaluation in identifying the etiology of autism spectrum disorders. *Genet Med* **10(4)**:301–305 → pages 2

[13] Abrahams, BS and Geschwind, DH (2008) Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* **9(5)**:341–355 → pages 2, 4, 10, 11, 13, 16, 18, 21

[14] Folstein, S and Rutter, M (1977) Infantile Autism: A study of 21 twin pairs. *J Child Psychol Psychiatry* **18(4)**:297–321 → pages 3

[15] Steffenburg, S, Gillberg, C, et al. (1989) A Twin Study of Autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry* **30(3)**:405–416 → pages 3, 8

[16] Bailey, A, Le Couteur, A, et al. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25(1)**:63–77 → pages 3

[17] Folstein, SE and Rosen-Sheidley, B (2001) Genetics of austim: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* **2(12)**:943–955 → pages 3, 4

[18] Bolton, P, Macdonald, H, et al. (1994) A Case-Control Family History Study of Autism. *J Child Psychol Psychiatry* **35(5)**:877–900 → pages 4

[19] Pickles, A, Starr, E, et al. (2000) Variable Expression of the Autism Broader Phenotype: Findings from Extended Pedigrees. *J Child Psychol Psychiatry* **41(4)**:491–502 → pages 4

[20] Schaefer, GB and Lutz, RE (2006) Diagnostic yield in the clinical genetic evaluation of autism spectrum disorders. *Genet Med* **8(9)**:549–556 → pages 4

[21] Freitag, CM (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry* **12(1)**:2–22 → pages 4, 10, 13

[22] Xu, J, Zwaigenbaum, L, et al. (2004) Molecular Cytogenetics of Autism. *Curr Genomics* **5(4)**:347 – 364 → pages 12, 13, 15, 16, 18

[23] Veenstra-VanderWeele, J and Cook, E H Jr (2004) Molecular genetics of autism spectrum disorder. *Mol Psychiatry* **9(9)**:819–832 → pages 4, 13, 15, 16

[24] London, E (2007) The Role of the Neurobiologist in Redefining the Diagnosis of Autism. *Brain Pathol* **17(4)**:408–411 → pages 4

[25] Burmeister, M, McInnis, MG, et al. (2008) Psychiatric genetics: progress amid controversy. *Nat Rev Genet* **9(7)**:527–540 → pages 4, 5

[26] Miles, H, Takahashi, N, et al. (2008) Development and validation of a measure of dysmorphology: Useful for autism subgroup classification. *Am J Med Genet A* **146A(9)**:1101–1116 → pages 5, 7, 83

[27] Gottesman, II and Gould, TD (2003) The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *Am J Psychiatry* **160(4)**:636–645 → pages 5

[28] Kanner, L (1943) Autistic disturbances of affective contact. *Nerv Child* **2**:217–250 → pages 5

[29] Walker, HA (1977) Incidence of minor physical anomaly in autism. *J Autism Dev Disord* **7(2)**:165–176 → pages 5, 6

[30] Links, P, Stockwell, M, et al. (1980) Minor physical anomalies in childhood autism. Part I. Their relationship to pre- and perinatal complications. *J Autism Dev Disord* **10(3)**:273–285 → pages 5, 85, 93, 94

[31] Gualtieri, CT, Adams, A, et al. (1982) Minor physical anomalies in alcoholic and schizophrenic adults and hyperactive and autistic children. *Am J Psychiatry* **139(5)**:640–643 → pages 5

[32] Ozgen, HM, Hop, JW, et al. (2008) Minor physical anomalies in autism: a meta-analysis. *Mol Psychiatry* **-(advance online publication 15 July 2008)**:– → pages 5, 6, 7

[33] Aase, JM (1990) Diagnostic Dysmorphology (Plenum Medical Book Co., New York) → pages 5

[34] Rodier, PM, Bryson, SE, et al. (1997) Minor Malformations and Physical Measurements in Autism: Data From Nova Scotia. *Teratology* **55(5)**:319–325 → pages 6

[35] Tripi, G, Roux, S, et al. (2008) Minor physical anomalies in children with autism spectrum disorder. *Early Hum Dev* **84(4)**:217–223 → pages 6, 7

[36] Hultman, CM, Sparen, P, et al. (2002) Perinatal Risk Factors for Infantile Autism. *Epidemiology* **13(4)**:417–423 → pages 6, 7

[37] Lauritsen, MB, Mors, O, et al. (2002) Medical Disorders Among Inpatients with Autism in Denmark According to ICD-8: A Nationwide Register-Based Study. *J Autism Dev Disord* **32(2)**:115–119 → pages

[38] Wier, ML, Yoshida, CK, et al. (2006) Congenital anomalies associated with autism spectrum disorders. *Dev Med Child Neurol* **48(6)**:500–507 → pages 6

[39] Juul-Dam, N, Townsend, J, et al. (2001) Prenatal, Perinatal, and Neonatal Factors in Autism, Pervasive Developmental Disorder-Not Otherwise Specified, and the General Population. *Pediatrics* **107(4)**:e63 → pages 6, 7

[40] Gillberg, C and Gillberg, IC (1983) Infantile autism: A total population study of reduced optimality in the pre-, peri-, and neonatal period. *J Autism Dev Disord* **13(2)**:153–166 → pages

[41] Piven, J, Simon, J, et al. (1993) The etiology of autism: pre-, peri-, and neonatal factors. *J Am Acad Child Adolesc Psychiatry* **32(6)**:1256–1263 → pages 8

[42] Bolton, PF, Murphy, M, et al. (1997) Obstetric complications in autism: consequences or causes of the condition. *J Am Acad Child Adolesc Psychiatry* **36(2)**:272–281 → pages 8, 9, 91

[43] Zwaigenbaum, L, Szatmari, P, et al. (2002) Pregnancy and Birth Complications in Autism and Liability to the Broader Autism Phenotype. *J Am Acad Child Adolesc Psychiatry* **41(5)**:572–579 → pages 8, 9, 91

[44] Glasson, EJ, Bower, C, et al. (2004) Perinatal Factors and the Development of Autism: A Population Study. *Arch Gen Psychiatry* **61(6)**:618–627 → pages 7

[45] Kolevzon, A, Gross, R, et al. (2007) Prenatal and Perinatal Risk Factors for Autism: A Review and Integration of Findings. *Arch Pediatr Adolesc Med* **161(4)**:326–333 → pages 7, 8

[46] Thomson, AM and Barron, SL (1983) Perinatal Mortality in Obstetrical Epidemiology (London: Academic Press) → pages 8

[47] Jones, M and Szatmari, P (1988) Stoppage rules and genetic studies of autism. *J Autism Dev Disord* **18(1)**:31–40 → pages 9

[48] Bolton, P and Holland, A (1994) Chromosomal abnormalities In: Child and Adolescent Psychiatry: Modern approaches (Blackwell Scientific Oxford: England) → pages 9

[49] Nelson, KB (1991) Prenatal and perinatal factors in the etiology of autism. *Pediatrics* **87**:S761–766 → pages 9

[50] The Autism Genome Project Consortium (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* **39(3)**:319–328 → pages 11, 18, 19

[51] Hoekstra, RA, Bartels, M, et al. (2007) Heritability of Autistic Traits in the General Population. *Arch Pediatr Adolesc Med* **161(4)**:372–377 → pages 11

[52] Constantino, JN and Todd, RD (2003) Autistic Traits in the General Population: A Twin Study. *Arch Gen Psychiatry* **60(5)**:524–530 → pages 11

[53] Alarcón, M, Abrahams, BS, et al. (2008) Linkage, Association, and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. *Am J Hum Genet* **82(1)**:150 – 159 → pages 12

[54] Bakkaloglu, B, O'Roak, BJ, et al. (2008) Molecular Cytogenetic Analysis and Resequencing of Contactin Associated Protein-Like 2 in Autism Spectrum Disorders. *Am J Hum Genet* **82(1)**:165 – 173 → pages 12

[55] Arking, DE, Cutler, DJ, et al. (2008) A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. *Am J Hum Genet* **82(1)**:160 – 164 → pages 12

[56] Strauss, KA, Puffenberger, EG, et al. (2006) Recessive Symptomatic Focal Epilepsy and Mutant Contactin-Associated Protein-like 2. *N Engl J Med* **354(13)**:1370–1377 → pages 12

[57] Susser, ES, Schwartz, S, et al. (2006) Psychiatric Epidemiology (Oxford University Press USA) → pages 12, 13

[58] Wang, K, Zhang, H, et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459(7246)**:528–533 → pages 13, 21

[59] Coutinho, A, Sousa, I, et al. (2007) Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. *Hum Genet* **121(2)**:243–256 → pages 13

[60] Cantor, RM, Kono, N, et al. (2005) Replication of Autism Linkage: Fine-Mapping Peak at 17q21. *Am J Hum Genet* **76(6)**:1050 – 1056 → pages 13

[61] Badner, JA and Gershon, ES (2002) Regional meta-analysis of published data supports linkage of autism with markers on chromosone 7. *Mol Psychiatry* **7(1)**:56 – 66 → pages 13

[62] Trikalinos, TA, Karvouni, A, et al. (2006) A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Mol Psychiatry* **11(1)**:29 – 36 → pages 13

[63] Lesch, KP, Bengel, D, et al. (1996) Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region. *Science* **274(5292)**:1527–1531 → pages 14

[64] Wassink, TH, Hazlett, HC, et al. (2007) Cerebral Cortical Gray Matter Overgrowth and Functional Variation of the Serotonin Transporter Gene in Autism. *Arch Gen Psychiatry* **64(6)**:709–717 → pages 14

[65] Campbell, DB, Sutcliffe, JS, et al. (2006) A genetic variant that disrupts MET transcription is associated with autism. *PNAS* **103(45)**:16834–16839 → pages 14

[66] Campbell, B, D'Oronzio, R, et al. (2007) Disruption of cerebral cortex MET signaling in autism spectrum disorder. *Ann Neurol* **62(3)**:243–250 → pages 14

[67] Devlin, B, Cook, EH, Jr, et al. (2005) Autism and the serotonin transporter: the long and short of it. *Mol Psychiatry* **10(12)**:1110–1116 → pages 14

[68] Klauck, SM, Poustka, F, et al. (1997) Serotonin transporter (5-HTT) gene variants associated with autism? *Hum Mol Genet* **6(13)**:2233–2238 → pages 14

[69] Betancur, C, Corbex, M, et al. (2002) Serotonin transporter gene polymorphisms and hyperserotonemia in autistic disorder. *Mol Psychiatry* **7(1)**:57 – 71 → pages 14

[70] Campbell, DB, Buie, TM, et al. (2009) Distinct Genetic Risk Based on Association of MET in Families With Co-occurring Autism and Gastrointestinal Conditions. *Pediatrics* **123(3)**:1018–1024 → pages 14

[71] Lepagnol-Bestel, AM, Maussion, G, et al. (2008) SLC25A12 expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects. *Mol Psychiatry* **13(4)**:385–397 → pages 14

[72] Modahl, C, Green, LA, et al. (1998) Plasma oxytocin levels in autistic children. *Biol Psychiatry* **43(4)**:270 – 277 → pages 14

[73] Bagni, C and Greenough, WT (2005) From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nat Rev Neurosci* **6(5)**:376–387 → pages 15

[74] Kwiatkowski, DJ and Manning, BD (2005) Tuberous sclerosis: a GAP at the crossroads of multiple signaling pathways. *Hum Mol Genet* **14(suppl_2)**:R251–258 → pages 15

[75] Wiznitzer, M (2004) Autism and Tuberous Sclerosis. *J Child Neurol* **19(9)**:675 – 679 → pages 15

[76] Kelleher, R J III and Bear, M F (2008) The Autistic Neuron: Troubled Translation? *Cell* **135(3)**:401 – 406 → pages 15, 17, 97

[77] Vorstman, JAS, Staal, WG, et al. (2006) Overview of cytogenetic regions of interest (CROIs) associated with the autism phenotype across the human genome. *Mol Psychiatry* **11(1)**:18–28 → pages 15, 16, 18

[78] Samaco, RC, Hogart, A, et al. (2005) Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum Mol Genet* **14(4)**:483–492 → pages 16

[79] Fine, SE, Weissman, A, et al. (2005) Autism Spectrum Disorders and Symptoms in Children with Molecularly Confirmed 22q11.2 Deletion Syndrome. *J Autism Dev Disord* **35(4)**:461–470 → pages 16

[80] Shprintzen, J (2008) Velo-cardio-facial syndrome: 30 Years of study. *Dev Disabil Res Rev* **14(1)**:3–10 → pages 16

[81] der Aa, NV, Rooms, L, et al. (2009) Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *Eur J Med Genet* **52(2-3)**:94 – 100 → pages 16

[82] Potocki, L, Bi, W, et al. (2007) Characterization of Potocki-Lupski Syndrome (dup(17)(p11.2p11.2)) and Delineation of a Dosage-Sensitive Critical Interval That Can Convey an Autism Phenotype. *Am J Hum Genet* **80(4)**:633 – 649 → pages 16

[83] Falk, E and Casas, A (2007) Chromosome 2q37 deletion: Clinical and molecular aspects. *Am J Med Genet C Semin Med Genet* **145C(4)**:357–371 → pages 16

[84] Galasso, C, Lo-Castro, A, et al. (2008) Deletion 2q37: An Identifiable Clinical Syndrome With Mental Retardation and Autism. *J Child Neurol* **23(7)**:802–806 → pages 16

[85] Jamain, S, Quach, H, et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* **34(1)**:p27 – 29 → pages 16, 17

[86] Durand, CM, Betancur, C, et al. (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* **39(1)**:25–27 → pages 16, 17

[87] Song, JY, Ichtchenko, K, et al. (1999) Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. *PNAS* **96(3)**:1100–1105 → pages 17

[88] Laumonnier, F, Bonnet-Brilhault, F, et al. (2004) X-Linked Mental Retardation and Autism Are Associated with a Mutation in the NLGN4 Gene, a Member of the Neuroligin Family. *Am J Hum Genet* **74(3)**:552 – 557 → pages 17

[89] Vincent, B, Kolozsvari, D, et al. (2004) Mutation screening of X-chromosomal neuroligin genes: No mutations in 196 autism probands. *Am J Med Genet B Neuropsychiatr Genet* **129B(1)**:82–84 → pages 17

[90] Chih, B, Afridi, SK, et al. (2004) Disorder-associated mutations lead to functional inactivation of neuroligins. *Hum Mol Genet* **13(14)**:1471–1477 → pages 17

[91] Tabuchi, K, Blundell, J, et al. (2007) A Neuroligin-3 Mutation Implicated in Autism Increases Inhibitory Synaptic Transmission in Mice. *Science* **318(5847)**:71–76 → pages 17

[92] Meyer, G, Varoqueaux, F, et al. (2004) The complexity of PDZ domain-mediated interactions at glutamatergic synapses: a case study on neuroligin. *Neuropharmacology* **47(5)**:724 – 733 → pages 17

[93] Moessner, R, Marshall, CR, et al. (2007) Contribution of SHANK3 Mutations to Autism Spectrum Disorder. *Am J Hum Genet* **81(6)**:1289 – 1297 → pages 17

[94] Walsh, CA, Morrow, EM, et al. (2008) Autism and Brain Development. *Cell* **135(3)**:396 – 400 → pages 17, 97

[95] Feuk, L, Carson, AR, et al. (2006) Structural variation in the human genome. *Nat Rev Genet* **7(2)**:85–97 → pages 18

[96] Burbach, JPH and van der Zwaag, B (2009) Contact in the genetics of autism and schizophrenia. *Trends Neurosci* **32(2)**:69 – 72 → pages 18

[97] Jacquemont, ML, Sanlaville, D, et al. (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet* **43(11)**:843–849 → pages 18, 20, 26, 80

[98] Sebat, J, Lakshmi, B, et al. (2007) Strong Association of De Novo Copy Number Mutations with Autism. *Science* **316(5823)**:445–449 → pages 19, 20, 22, 26

[99] Marshall, CR, Noor, A, et al. (2008) Structural Variation of Chromosomes in Autism Spectrum Disorder. *Am J Hum Genet* **82(2)**:477 – 488 → pages 20, 26, 69

[100] Christian, SL, Brune, CW, et al. (2008) Novel Submicroscopic Chromosomal Abnormalities Detected in Autism Spectrum Disorder. *Biol Psychiatry* **63(12)**:1111 – 1117 → pages 20, 21, 22, 23, 26, 69

[101] Cusco, I, Medrano, A, et al. (2009) Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum Mol Genet* **26 Feb 2009 epub ahead of print** → pages 18, 20, 26

[102] Qiao, Y, Riendeau, N, et al. (2009) Phenomic Determinants of Genomic Variation in Autism Spectrum Disorders. *in press J Med Genet* → pages 19, 25, 26, 81, 84, 90, 92, 95, 96

[103] Glessner, JT, Wang, K, et al. (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459(7246)**:569–573 → pages 20

[104] Sebat, J, Lakshmi, B, et al. (2004) Large-Scale Copy Number Polymorphism in the Human Genome. *Science* **305(5683)**:525–528 → pages 21

[105] Iafrate, AJ, Feuk, L, et al. (2004) Detection of large-scale variation in the human genome. *Nat Genet* **36(9)**:949–951 → pages

[106] Sharp, AJ, Cheng, Z, et al. (2006) Structural Variation of the Human Genome. *Annu Rev Genomics Hum Genet* **7(1)**:407–442 → pages

[107] Redon, R, Ishikawa, S, et al. (2006) Global variation in copy number in the human genome. *Nature* **444(7118)**:444–454 → pages 21

[108] Fan, YS, Jayakar, P, et al. (2007) Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization. *Hum Mutat* **28(11)**:1124–1132 → pages 21, 30, 39, 46

[109] Lee, C, Iafrate, AJ, et al. (2007) Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* **39(S7)**:S48–54 → pages 21, 30

[110] Rajcan-Separovic, E, Harvard, C, et al. (2007) Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *J Med Genet* **44(4)**:269–276 → pages 21

[111] de Leeuw, N, Pfundt, R, et al. (2008) A newly recognised microdeletion syndrome involving 2p15p16.1: narrowing down the critical region by adding another patient detected by genome wide tiling path array comparative genomic hybridisation analysis. *J Med Genet* **45(2)**:122–124 → pages 22

[112] Chabchoub, E, Vermeesch, JR, et al. (2008) The facial dysmorphy in the newly recognised microdeletion 2p15-p16.1 refined to a 570 kb region in 2p15. *J Med Genet* **45(3)**:189–192 → pages 22

[113] Kumar, RA, KaraMohamed, S, et al. (2008) Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* **17(4)**:628–638 → pages 22

[114] Weiss, LA, Shen, Y, et al. (2008) Association between Microdeletion and Microduplication at 16p11.2 and Autism. *N Engl J Med* **358(7)**:667–675 → pages 22, 23, 63, 90

[115] Lupski, JR (2007) Genomic rearrangements and sporadic disease. *Nat Genet* **39**:S43–S47 → pages 22

[116] Ghebranious, N, Giampietro, F, et al. (2007) A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am J Med Genet A* **143A(13)**:1462–1471 → pages 23

[117] Mefford, HC, Sharp, AJ, et al. (2008) Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes. *N Engl J Med* **359(16)**:1685–1699 → pages 23

[118] Dibbens, LM, Mullen, S, et al. (2009) Familial and sporadic 15q13.3 microdeletions in Idiopathic Generalized Epilepsy: Precedent for Disorders with Complex Inheritance. *Hum Mol Genet* **10 Jul 2009 epub ahead of print** → pages 23

[119] Lord, C, Rutter, M, et al. (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24(5)**:659–685 → pages 27, 49, 66, 74

[120] Lord, C, Risi, S, et al. (2000) The Autism Diagnostic Observation ScheduleGeneric: A Standard Measure of Social and Communication Deficits

Associated with the Spectrum of Autism. *J Autism Dev Disord* **30(3)**:205–223 → pages 27, 49, 66, 74

[121] Tyson, C, Harvard, C, et al. (2005) Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. *Am J Med Genet A* **139(3)**:173–85 → pages 35

[122] Aerts, S, Lambrechts, D, et al. (2006) Gene prioritization through genomic data fusion. *Nat Biotechnol* **24(5)**:537–544 → pages 46, 47, 88

[123] Sutcliffe, JS (2008) GENETICS: Insights into the Pathogenesis of Autism. *Science* **321(5886)**:208–209 → pages 46, 88

[124] Tranchevent, LC, Barriot, R, et al. (2008) ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucl Acids Res* **36(suppl_2)**:W377–384 → pages 47

[125] Kulharya, S, Flannery, B, et al. (2008) Fine mapping of breakpoints in two unrelated patients with rare overlapping interstitial deletions of 9q with mild dysmorphic features. *Am J Med Genet* **146A(17)**:2234–2241 → pages 52, 54

[126] Farrell, SA, Siegel-Bartelt, J, et al. (1991) Patients with deletions of 9q22q34 do not define a syndrome: three case reports and a literature review. *Clin Genet* **40(3)**:207–214 → pages 52

[127] Park, JP, Moeschler, JB, et al. (1991) Schizophrenia and mental retardation in an adult male with a de novo interstitial deletion 9(q32q34.1). *J Med Genet* **28(4)**:282–283 → pages

[128] Turleau, C, de Grouchy, J, et al. (1978) Deletions intercalaires del9q. *Ann Genet* **21**:234–236 → pages 52

[129] Kleefstra, T, van Zelst-Stams, WA, et al. (2009) Further clinical and molecular delineation of the 9q Subtelomeric Deletion Syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. *J Med Genet* **Epub 4 March 2009** → pages 54

[130] Song, JK, Giniger, E, et al. (2008) The receptor protein tyrosine phosphatase PTP69D antagonizes Abl tyrosine kinase to guide axons in Drosophila. *Mech Dev* **125(3-4)**:247 – 256 → pages 54

[131] Moresco, EMY and Koleske, AJ (2003) Regulation of neuronal morphogenesis and synaptic function by Abl family kinases. *Curr Opin Neurobiol* **13(5)**:535 – 544 → pages 54, 55

[132] Wills, Z, Bateman, J, et al. (1999) The Tyrosine Kinase Abl and Its Substrate Enabled Collaborate with the Receptor Phosphatase Dlar to Control Motor Axon Guidance. *Neuron* **22(2)**:301 – 312 → pages 55

[133] Bashaw, GJ, Kidd, T, et al. (2000) Repulsive Axon Guidance: Abelson and Enabled Play Opposing Roles Downstream of the Roundabout Receptor. *Cell* **101(7)**:703 – 715 → pages 55

[134] Moresco, EMY, Scheetz, AJ, et al. (2003) Abl Family Nonreceptor Tyrosine Kinases Modulate Short-Term Synaptic Plasticity. *J Neurophysiol* **89(3)**:1678–1687 → pages 55

[135] Ferrari-Toninelli, G, Paccioretti, S, et al. (2004) TorsinA negatively controls neurite outgrowth of SH-SY5Y human neuronal cell line. *Brain Res* **1012(1-2)**:75 – 81 → pages 55

[136] Bahn, E, Siegert, S, et al. (2006) TorsinB expression in the developing human brain. *Brain Res* **1116(1)**:112 – 119 → pages 55

[137] Hewett, JW, Zeng, J, et al. (2006) Dystonia-causing mutant torsinA inhibits cell adhesion and neurite extension through interference with cytoskeletal dynamics. *Neurobiol Dis* **22(1)**:98 – 111 → pages 55

[138] Repicky, S and Broadie, K (2009) Metabotropic Glutamate Receptor-Mediated Use-Dependent Down-Regulation of Synaptic Excitability Involves the Fragile X Mental Retardation Protein. *J Neurophysiol* **101(2)**:672–687 → pages 55

[139] Muehlmann, AM and Devine, DP (2008) Glutamate-mediated neuroplasticity in an animal model of self-injurious behaviour. *Behav Brain Res* **189(1)**:32 – 40 → pages 55

[140] Hu, WH, Pendergast, JS, et al. (2005) NIBP, a Novel NIK and IKKbeta-binding Protein That Enhances NF-kappaB Activation. *J Biol Chem* **280(32)**:29233–29241 → pages 58

[141] Mattson, MP and Camandola, S (2001) NF-kappaB in neuronal plasticity and neurodegenerative disorders. *J Clin Invest* **107(3)**:247–254 → pages 58

[142] Ruf, N, Bahring, S, et al. (2007) Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Hum Mol Genet* **16(21)**:2591–2599 → pages 58

[143] Luedi, PP, Dietrich, FS, et al. (2007) Computational and experimental identification of novel human imprinted genes. *Genome Research* **17(12)**:1723–1730 → pages 58

[144] Barel, O, Shalev, SA, et al. (2008) Maternally Inherited Birk Barel Mental Retardation Dysmorphism Syndrome Caused by a Mutation in the Genomically Imprinted Potassium Channel KCNK9. *Am J Hum Genet* **83(2)**:193–199 → pages 58, 59

[145] Itsara, A, Cooper, GM, et al. (2009) Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. *The American Journal of Human Genetics* **84(2)**:148 – 161 → pages 63, 69, 70, 90

[146] Kumar, RA, Marshall, CR, et al. (2009) Association and Mutation Analyses of 16p11.2 Autism Candidate Genes. *PLoS ONE* **4(2)**:e4582 → pages 63, 64

[147] Shimizu-Nishikawa, K, Kajiwara, K, et al. (1995) Cloning and expression of SEZ-6, a brain-specific and seizure-related cDNA. *Mol Brain Res* **28(2)**:201 – 210 → pages 63

[148] Gunnersen, JM, Kim, MH, et al. (2007) Sez-6 Proteins Affect Dendritic Arborization Patterns and Excitability of Cortical Pyramidal Neurons. *Neuron* **56(4)**:621 – 639 → pages 63, 64

[149] Okamoto, S, Sherman, K, et al. (2002) Effect of the ubiquitous transcription factors, SP1 and MAZ, on NMDA receptor subunit type 1 (NR1) expression during neuronal differentiation. *Molecular Brain Research* **107(2)**:89 – 96 → pages 64

[150] Sucher, NJ, Awobuluyi, M, et al. (1996) NMDA receptors: from genes to channels. *Trends Pharm Neurosci* **17(10)**:348–355 → pages 64

[151] Murakami, A, Matsuda, M, et al. (2006) Characterization of the human PRIP-1 gene structure and transcriptional regulation. *Gene* **382**:129 – 139 → pages 64

[152] Groffen, A, Friedrich, R, et al. (2006) DOC2A and DOC2B are sensors for neuronal activity with unique calcium-dependent and kinetic properties. *J Neurochem* **97(3)**:818–833 → pages 64

[153] Sakaguchi, G, Manabe, T, et al. (1999) Doc2A is an activity-dependent modulator of excitatory synaptic transmission. *Eur J Neurosci* **11(12)**:p4262 – 4268 → pages 64

[154] Stone, TW (2001) Endogenous neurotoxins from tryptophan. *Toxicon* **39(1)**:61 – 73 → pages 64

[155] Schwarcz, R, Speciale, C, et al. (1986) Excitatory Amino Acids and Epilepsy (Plenum Publishing Corp., New York) → pages 64

[156] Guillemin, GJ and Brew, BJ (2002) Implications of the kynurenine pathway and quinolinic acid in Alzheimers disease. *Redox Rep* **7(4)**:199206 → pages 64

[157] Finkbeiner, S and Cuero, AM (2006) Disease modifying pathways in neurodegeneration. *J Neurosci* **26(41)**:1034910357 → pages 64

[158] Kedersha, NL, Heuser, JE, et al. (1991) Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. *J Cell Biol* **112(2)**:225–235 → pages 64

[159] Steiner, E, Holzmann, K, et al. (2006) Cellular Functions of Vaults and their Involvement in Multidrug Resistance. *Curr Drug Targets* **7(8)**:923 – 934 → pages 64

[160] Sisodiya, SM, Martinian, L, et al. (2003) Major Vault Protein, a Marker of Drug Resistance, Is Upregulated in Refractory Epilepsy. *Epilepsia* **44(11)**:1388–1396 → pages 64

[161] Rainesalo, S, Saransaari, P, et al. (2003) Uptake of GABA and activity of GABA-transaminase in platelets from epileptic patients. *Epilepsy Research* **53(3)**:233 – 239 → pages 65

[162] Rainesalo, S, Eriksson, K, et al. (2004) Uptake of GABA and Activity of GABA Transaminase in Blood Platelets from Children with Absence Epilepsy. *Neurochem Res* **29(10)**:1873–1877 → pages 65

[163] Duffy, S, Nguyen, PV, et al. (2004) Phenylethylidenehydrazine, a novel GABA-transaminase inhibitor, reduces epileptiform activity in rat hippocampal slices. *Neuroscience* **126(2)**:423 – 432 → pages 65

[164] Barnby, G, Abbott, A, et al. (2005) Candidate-Gene Screening and Association Analysis at the Autism-Susceptibility Locus on Chromosome 16p: Evidence of Association at GRIN2A and ABAT. *Am J Hum Genet* **76(6)**:950 – 966 → pages 65

[165] Bonati, T, Finelli, P, et al. (2005) Trisomy 15q25.2-qter in an autistic child: Genotype-phenotype correlations. *Am J Med Genet* **133A(2)**:184–188 → pages 69

[166] Roggenbuck, JA, Mendelsohn, NJ, et al. (2004) Duplication of the Distal Long Arm of Chromosome 15: Report of Three New Patients and Review of the Literature. *Am J Med Genet* **126A(4)**:398–402 → pages 69

[167] Zollino, M, Tiziano, F, et al. (1999) Partial duplication of the long arm of chromosome 15: Confirmation of a causative role in craniosynostosis and definition of a 15q25-qter trisomy syndrome. *Am J Med Genet* **87(5)**:391–394 → pages 69

[168] Van Allen, MI, Siegel-Bartelt, J, et al. (1992) Craniosynostosis associated with partial duplication of 15q and deletion of 2q. *Am J Med Genet* **43(4)**:688–692 → pages 69

[169] Tatton-Brown, K, Pilz, T, et al. (2009) 15q overgrowth syndrome: A newly recognized phenotype associated with overgrowth, learning difficulties, characteristic facial appearance, renal anomalies and increased dosage of distal chromosome 15q. *Am J Med Genet* **149A(2)**:147–154 → pages 69

[170] Cohen, LS, Friedman, JM, et al. (1994) A reevaluation of risk of in utero exposure to lithium. *JAMA* **271(2)**:146–150 → pages 70

[171] Austin, MPV and Mitchell, PB (1998) Psychotropic medications in pregnant women: treatment dilemmas. *Med J Aust* **169(8)**:428–431 → pages 70

[172] Kang, J, Chen, XL, et al. (2001) The Antipsychotic Drugs Sertindole and Pimozide Block erg3, a Human Brain K+ Channel. *Biochem and Biophys Res Commun* **286(3)**:499 – 504 → pages 70

[173] Kamath, BM, Thiel, BD, et al. (2009) SNP Array Mapping of Chromosome 20p Deletions: Genotypes, Phenotypes, and Copy Number Variation. *Hum Mutat* **30(3)**:371–378 → pages 73

[174] Davis, RJ, Shen, W, et al. (2001) Characterization of mouse DACH2, a homologue of Drosophila dachshund. *Mech Dev* **102(1-2)**:169 – 179 → pages 76

[175] Davis, RJ, Shen, W, et al. (2001) DACH1 Mutant Mice Bear No Gross Abnormalities in Eye, Limb, and Brain Development and Exhibit Postnatal Lethality. *Mol Cell Biol* **21(5)**:1484–1490 → pages 76

[176] Davis, J, Pesah, I, et al. (2006) Mouse DACH2 mutants do not exhibit gross defects in eye development or brain function. *Genesis* **44(2)**:84–92 → pages 76, 77

[177] Braybrook, C, Warry, G, et al. (2001) Identification and Characterization of KLHL4, a Novel Human Homologue of the Drosophila Kelch Gene That Maps within the X-Linked Cleft Palate and Ankyloglossia (CPX) Critical Region. *Genomics* **72(2)**:128 – 136 → pages 77

[178] Engels, H, Brockschmidt, A, et al. (2007) DNA microarray analysis identifies candidate regions and genes in unexplained mental retardation. *Neurology* **68(10)**:743–50 → pages 92

Appendix A: UBC Research Ethics Board's Certificate of Approval

The University of British Columbia
Office of Research Services
Clinical Research Ethics Board – Room 210, 828 West 10th Avenue, Vancouver, BC V5Z 1L8

ETHICS CERTIFICATE OF EXPEDITED APPROVAL: RENEWAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC CREB NUMBER:
Suzanne M.E. Lewis		H01-70507

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
Children's and Women's Health Centre of BC (incl. Sunny Hill)	Children's and Women's Health Centre of BC (incl. Sunny Hill)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Evica Rajcan-Separovic
Elena Lopez-Rangel
Elizabeth C. Mickelson
Linda Eaves
Niamh Lynch
Mary Connolly
Pratibha Reebye
Helena H. Ho

SPONSORING AGENCIES:
- Autism Speaks - "Clinical and Gene Signatures of Autism Spectrum Disorders"
- Canadian Institutes of Health Research (CIHR) - "Redefining Autism Spectrum Disorder: Correlations from Genotype to Phenotype & Toward Early Identification, Management & Prevention ii) Identification of Culprit Genes Using Genomic Microarrays & Molecular Assessments of Duplicon-Mediated ****"
- Canadian Institutes of Health Research (CIHR) - "The Identification of Susceptibility Genes and Phenotypic Subgroups for Autism Spectrum Disorders"
- Michael Smith Foundation for Health Research - "The Identification of Susceptibility Genes and Phenotypic Subgroups for Autism Spectrum Disorders"
- National Alliance for Autism Research - "The Identification of Susceptibility Genes, Phenotypic Subgroups and Craniofacial Endophenotypes Associated with Autism Spectrum Disorders"
- Unfunded Research - "Redefining Autism Spectrum Disorder: Correlations From Genotype to Phenotype & Applications Toward Early Identification, Management & Prevention - (ii) Phenotyping Genetic Subgroups"

PROJECT TITLE:
Clinical and Gene Signatures of Autism Spectrum Disorders

EXPIRY DATE OF THIS APPROVAL: November 12, 2009

APPROVAL DATE: November 12, 2008

CERTIFICATION:
In respect of clinical trials:
1. The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards defined in Division 5 of the Food and Drug Regulations.
2. The Research Ethics Board carries out its functions in a manner consistent with Good Clinical Practices.
3. This Research Ethics Board has reviewed and approved the clinical trial protocol and informed consent form for the trial which is to be conducted by the qualified investigator named above at the specified clinical trial site. This approval and the views of this Research Ethics Board have been documented in writing.

The Chair of the UBC Clinical Research Ethics Board has reviewed the documentation for the above named project. The research study, as presented in the documentation, was found to be acceptable on ethical grounds for research involving human subjects and was approved for renewal by the UBC Clinical Research Ethics Board.

Approval of the Clinical Research Ethics Board by