

**THE PATHOGENICITY OF COPY NUMBER VARIANTS IN CHILDREN WITH
INTELLECTUAL DISABILITY**

by

Farah R Zahir

B.Sc., Genetics, Biochemistry, Microbiology, Bangalore University, 2003

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2011

© Farah R Zahir 2011

Abstract

Intellectual disability affects 1-3% of individuals globally, and, for half the cases, the cause is unknown. Recent studies using whole genome microarray genomic hybridization have shown that submicroscopic genomic imbalance causes intellectual disability in at least 10% of idiopathic cases with normal conventional cytogenetic analysis. I established genotype-phenotype correlations for *de novo* copy number variants detected by previous whole genome array genome hybridization studies performed by our group in children with intellectual disability. These genotype-phenotype correlations show that genomic imbalance of genes belonging to the epigenetic regulatory category, among others, are causative of intellectual disability.

I hypothesized that dosage changes in the broad functional category of genes encoding epigenetic regulatory proteins are more likely to be pathogenic for intellectual disability than dosage changes in other kinds of genes. Epigenetic regulatory proteins include those with DNA methylation, histone modification or chromatin remodeling activity. I have selected all known genes encoding epigenetic regulatory proteins and defined probes to interrogate these candidate genes for copy number alteration as part of a custom targeted microarray design that selectively investigates all candidate genes associated with intellectual disability. We have conducted comparative genome hybridization on 177 patients with idiopathic intellectual disability using this array and on both normal parents of each affected child. We identified and independently validated 16 cases with *de novo* CNVs involving the epigenetic regulatory candidates. 7 of the 16 CNVs involve the same exon of the *JARID2* gene, while the other 9 CNVs affect different genes. I discuss genotype-phenotype correlations for these cases and show that epigenetic perturbation by way of disruption of genes that encode epigenetic regulators is an important cause for intellectual disability.

Preface

This dissertation is comprised of both published and unpublished material as follows;

Chapter 2, 3 and 4 are verbatim reproductions of published manuscripts. The papers detail genotype-phenotype correlations for four patients identified to have previously unknown genetic defects by studies undertaken at the Medical Genetics Research Unit in collaboration with the British Columbia Genome Sciences Center and headed by Dr. Jan M. Friedman. This study and the subsequent genotype-phenotype correlations studies arising from this work that are detailed below were approved by the University of British Columbia Clinical Research Ethics Board (certificate number C04-0537).

For chapter 2, **'Novel deletions of 14q11.2 associated with intellectual disability and similar minor anomalies in three children'** (Journal of Medical Genetics. 2007 Sep;44(9):556-61), chapter 3, **'A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1 α '** (Journal of Medical Genetics. 2008 Apr;45(4):239-43) and chapter 4, **'A novel de novo 1.1 Mb duplication of 17q21.33 associated with cognitive impairment and other anomalies'** (American Journal of Medical Genetics Part A. 2009 Jun;149A(6):1257-62.), I was solely in charge of researching the genotype-phenotype correlation and preparing the manuscripts with input from all contributing authors.

Chapters 5 and 6 are verbatim reproductions of published manuscripts for which I was a contributing author. Chapter 5, **'Duplications of the critical Rubinstein-Taybi deletion region on chromosome 16p13.3 cause a novel recognizable syndrome'** (Journal of Medical Genetics. 2010 Mar;47(3):155-61), contains a clinical comparison and discussion of candidate genes for a newly defined Intellectual Disability syndrome. I contributed clinical data on the patient from our center included in this study and also assisted with researching candidate genes for the syndrome. Chapter 6 **'The duplication 8q12 Case: A characteristic syndrome associated with microduplication of 8q12, inclusive of CHD7'** (European Journal of Medical Genetics. 2009 Nov-Dec;52(6):436-9.) contains a detailed clinical case report of a patient with a novel genetic defect detected in the same study

headed by Dr. Jan M Friedman that led to the works appearing in chapters 2, 3 and 4. For this case I contributed an extensive genotype-phenotype correlation analysis, delineating candidate genes. The manuscript was prepared by Dr. Anna Lehman, who provided the clinical description of the patient.

Chapters 1 and 8 contain large sections that are taken from two published review papers that I prepared. The discussion of microarrays appearing in chapter 1, the main technology underlying experiments detailed in this thesis, is taken from **'The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility'** (Clinical Genetics. 2007 Oct;72(4):271-87) a review article published in 2007 written with Dr. Jan M. Friedman as my co-author. I have updated the text as required to reflect current knowledge for sections reproduced from this work. The sections introducing epigenetics and discussing the contribution of epigenetic regulation to neurodevelopment contained in chapter 1 and chapter 8 are taken from the review article, **'Epigenetic impacts on neurodevelopment: pathophysiological mechanisms and genetic modes of action'** (Pediatric Research. 2011 May;69(5 Pt 2):92R-100R) that I wrote in collaboration with Dr. Carolyn Brown.

Chapter 7 is previously unpublished. The chapter details the main project of my doctoral research that I conducted in collaboration with Dr. Tracy Tucker who was in charge of a study to design and test a microarray probing selected genes. I was in charge of the selection of candidate genes to test my research hypothesis, which I contributed to the design (36% of the design). Dr. Jan Friedman, Dr. Sylvie Langlois and Dr. Patrice Eydoux provided the rest of the gene selection. Dr. Jacques Michaud (CHU Sainte-Justine, Montreal) provided patient samples and clinical reports. I was solely in charge of the entire validation component of the study that used an independent technology to assess the findings of the microarray. I have contributed this chapter as a joint primary author to a manuscript in preparation by Dr. Tracy Tucker detailing the results of this project.

This study was approved by the University of British Columbia Clinical Research Ethics Board (certificate number H07-00392).

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	v
List of Tables	xi
List of Figures	xii
List of Abbreviations.....	xiv
Acknowledgements.....	xv
Dedication.....	xvi
 Chapter 1: Introduction	 1
1.1 Definition of Intellectual Disability.....	1
1.2 Etiology of Intellectual Disability	1
1.2.1 The genetic etiology of Intellectual Disability	2
1.2.1.1 Cytogenetically detectable genetic aberrations	2
1.2.1.2 Genetic aberrations not detectable cytogenetically.....	2
1.2.1.3 Single gene mutations.....	3
1.3 Epigenetics - definition and overview of epigenetic processes	4
1.4 Epigenetic perturbation in neurogenetic disorders	7
1.4.1 Dosage change of genes encoding epigenetic factors as the mechanism of pathogenicity	11
1.5 Microarray technology	12
1.5.1 Types of arrays and factors that influence their effective resolution	14
1.5.2 The bioinformatics of AGH	17
1.5.3 Comparison to reference genome(s).....	18
1.5.4 Conclusions regarding use of arrays to identify CNVs.....	20
1.6 Independent confirmation of CNVs detected by AGH	20

1.7 Characterizing the pathogenicity of a CNV: Genotype-Phenotype correlations	21
1.7.1 Differentiating between benign CNVs and those that are pathogenic for Intellectual Disability: <i>Inheritance status</i>	22
1.7.2 Differentiating between benign CNVs and those that are pathogenic for Intellectual Disability: <i>connection to known pathogenic genetic lesions</i>	23
1.7.3 Differentiating between benign CNVs and those that are pathogenic for Intellectual Disability: <i>other considerations</i>	23
1.7.4 Summary: novel pathogenic CNVs, new Intellectual Disability syndromes and candidate genes	24
1.8 Genotype-Phenotype correlation studies undertaken in a cohort of patients with <i>de novo</i> CNVs identified by AGH.....	28
1.8.1 AGH study results.....	28
1.8.2 Summary of Genotype-Phenotype correlation study methodology	29
1.8.3 Summary of Genotype-Phenotype correlation findings	30
1.8.4 Genotype-Phenotype correlation studies published as first author	31
1.8.5 Genotype-Phenotype correlation studies published as contributing author included in this thesis.....	32
1.9 Custom aCGH to probe all genes encoding epigenetic regulators	33
1.9.1 Experimental plan summary.....	34
1.10 Doctoral research objectives.....	34
1.11 Thesis outline	34
Chapter 2: Novel deletions of 14q11.2 associated with mental retardation and similar minor anomalies in three children	36
2.1 Introduction.....	36
2.2 Results	37
2.2.1 Vancouver case 5566	37
2.2.2 Vancouver case 8326	38
2.2.3 Case DECIPHER#126	39

2.3 Discussion	40
2.3.1 Previously-reported cases with chromosome 14q11.2 deletion	40
2.3.2 Candidate genes.....	42
2.3.3 Low copy repeat (LCR) sequences in the region.....	44
2.3.4 Common deletion polymorphisms in the region.....	45
2.4 Conclusion.....	46
Chapter 3: A patient with vertebral, cognitive and behavioural abnormalities and a <i>de novo</i> deletion of <i>NRXN1α</i>	47
3.1 Introduction.....	47
3.2 Case report.....	47
3.3 Discussion	50
3.4 Conclusion.....	57
Chapter 4: A novel <i>de novo</i> 1.1 Mb duplication of 17q21.33 associated with cognitive impairment and other anomalies	58
4.1 Introduction.....	58
4.2 Case report.....	58
4.3 Discussion	61
4.4 Conclusion.....	64
Chapter 5: Duplications of the critical Rubinstein Taybi deletion region cause a novel recognizable syndrome characterized by mild arthrogryposis.	66
5.1 Introduction.....	66
5.2 Patients, materials and methods.....	66
5.3 Results	67
5.4 Discussion	74
Chapter 6: A characteristic syndrome associated with microduplication of 8q12, inclusive of <i>CHD7</i>	78
6.1 Methods of detection	78

6.1.1 Cytogenetics	78
6.1.2 Array comparative genomic hybridization	78
6.1.3 Chromosomal anomaly	78
6.2 Method of confirmation	78
6.3 Clinical description	79
6.4 Discussion	80
Chapter 7: Custom aCGH to investigate incidence of CNVs involving genes encoding epigenetic regulators pathogenic for ID	84
7.1 Study Design	84
7.2 Material and methods	85
7.2.1 Patients and samples	85
7.2.2 Selection of genes involved in epigenetic regulation	85
7.2.3 Custom microarray design	86
7.2.4 aCGH	89
7.2.5 Defining <i>de novo</i> CNVs	89
7.2.6 Validations of <i>de novo</i> CNVs	90
7.2.7 Genotype-Phenotype correlations analyses	90
7.3 Results	92
7.4 Genotype Phenotype correlations	102
7.4.1 CNVs affecting the whole gene	102
7.4.1.1 <i>SMARCA2</i> loss	102
7.4.1.2 <i>MEF2C</i> loss	103
7.4.2 CNVs affecting exon1 (and the promoter) but not the whole gene	105
7.4.2.1 Copy number losses	105
7.4.2.1.1 <i>CHD6</i> exon 1 loss	105
7.4.2.2 Copy number gains	108

7.4.2.2.1 <i>CHD7</i> exon 1 gain and <i>ARID1B</i> exon 1 gain	108
7.4.2.2.2 <i>ARID2</i> exons 1, 2 and 3 gain.....	112
7.4.2.2.3 <i>ARID4B</i> exons 1 and 2 gain	114
7.4.3 CNVs affecting single intra-genic exons other than exon 1	115
7.4.3.1 <i>JMJD1A</i> exon 6 loss	116
7.4.3.2 <i>JMJD1C</i> exon 4 gain.....	117
7.4.3.3 <i>JARID2</i> exon 5 gains and losses.....	119
7.4.3.3.1 <i>JARID2</i> exon 5 copy number losses.....	121
7.4.3.3.2 <i>JARID2</i> exon 5 copy number gains	123
7.4.3.3.3 Functional analysis of <i>JARID2</i> exon 5 CNVs.....	124
7.5 Discussion	128
7.5.1 Interpretation of study findings in relationship to the hypothesis.....	128
7.5.2 Limitations of study	129
7.5.3 Importance of study	130
Chapter 8: General discussion	132
8.1 Genotype-Phenotype studies for whole genome array genome hybridization data.	132
8.2 Epigenetic impacts on neurodevelopment: pathophysiological mechanisms and genetic modes of action.....	135
8.2.1 Pleiotropy and functional complexity	135
8.2.2 Endophenotypes and epigenetic modes of action in the brain.....	138
8.3 Conclusion.....	140
References.....	144
Appendices.....	170
Appendix A: Summary of clinical features in children with de novo CNVs detected by WGS with mapping 100K arrays and confirmed by FISH.....	170

Appendix B: Genomic imbalance detected by 500K GeneChip® AGH in a cohort of 100 trios with idiopathic ID.....	171
Appendix C: Methodology for investigating genotype-phenotype correlations for <i>de novo</i> CNVs identified in children with idiopathic intellectual disability.	176
Appendix D: Primer sequences for qPCR validation (SYBR green $\Delta\Delta CT$ method) of custom aCGH results of CNVS involving epigenetic regulatory genes.	177

List of Tables

Table 1.1	Genes encoding epigenetic regulators that have been implicated in neurodevelopmental pathologies.....	9
Table 1.2	Genomic effects that can influence the expression of genes that lie outside of a CNV.....	26
Table 2.1	Other genes within 500K of our minimal critical deletion region: 20.4 Mb to 21.5 Mb on chromosome 14.....	44
Table 3.1	Details of other reported cases with genomic aberrations expected to cause NRXN1 loss of function.....	54
Table 4.1	Important RefSeq genes in the duplicated segment.....	65
Table 5.1	Genotypic and phenotypic characteristics of patients with an interstitial 16p13.3 duplication.....	70
Table 6.1	Comparison of physiological systems in two children with microduplication of 8q12 versus CHARGE syndrome.....	81
Table 7.1	Summary of validated <i>de novo</i> CNVs involving epigenetic regulatory genes.....	94
Table 7.2	Genotype-Phenotype correlations summary for 16 validated CNVs involving epigenetic regulatory genes.....	95
Table 7.3	Phenotypic comparison for patients with <i>de novo</i> CNVs of only <i>JARID2</i> exon 5.....	120

List of Figures

Figure 1.1	Illustrating interactions between DNA methylation, histone modification and chromatin remodelling.....	6
Figure 1.2	Comparison of cytogenetic CGH, aCGH and AGH.....	13
Figure 1.3	Factors to be considered when choosing an aCGH Platform.....	15
Figure 1.4	Overview of number of trios analyzed by 100K and 500K AGHs, and results obtained.....	30
Figure 2.1	Photographs of the faces and ears of our three cases.....	41
Figure 2.2	SNP genotyping data on case 5566 and both her parents by AGH using Affymetrix Genechip®	42
Figure 2.3	The genomic region of chromosome14 that contains the deleted region of our three cases.....	45
Figure 3.1	(A) chest X-ray image of patient taken at 4 years of age, showing the left extra rib, (B) Facial photograph of patient at 11 years of age.....	50
Figure 3.2	Schematic structure of Neurexin β and Neurexin α proteins.....	53
Figure 4.1	Mildly dysmorphic face of patient (A) with prominent cupped ears (B) at 10 years of age.....	60
Figure 4.2	A and B, spine radiographs at 10 and 14 years of age, respectively.....	61
Figure 4.3	Showing the ~1.2 Mb genomic sequence for the maximally affected region on the UCSC genome browser	62
Figure 5.1	The extent of the duplications identified in the present study.....	69
Figure 5.2	Phenotypic features of patients 1, 2, 4 and 6–12. Please note that patient 2 has had a surgical correction for ptosis.....	73
Figure 6.1	Genomic intervals of the 8q12 microduplications found in our patient and the previous patient displayed, overlaid on an adapted	

	segment from the UCSC genome browser.....	81
Figure 7.1	Showing location of probes against a chromosome ideogram.....	88
Figure 7.2	Overview of AGH results.....	93
Figure 7.2	Nexus images of validated CNVs involving epigenetic regulatory genes.....	99
Figure 7.3	Comparison of facial photographs of patient from Monfort et al (A and B) with that of our patient 33739 (C and D).....	111
Figure 7.4	I-TASSER modeling for normal <i>JARID2</i> and aberrant <i>JARID2</i>	125

List of Abbreviations

Basepair (bp)

Chromosome (Chr)

Copy Number Variant (CNV)

DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER)

Database of Genomic Variants (DGV)

Deoxyribonucleic acid (DNA)

Intellectual Disability (ID)

Kilobase (kb)

Loss of Function (LOF)

Megabase (Mb)

Mental Retardation (MR)

Messenger ribonucleic acid (mRNA)

Microarray Comparative Genome Hybridization or Array Genome Hybridization (aCGH)

Polymerase Chain Reaction (PCR)

Quantitative Polymerase Chain Reaction (qPCR)

Acknowledgements

I would like to express my deep and abiding thanks to my research supervisor Dr. Jan M. Friedman, for his visionary guidance, steadfast support and for believing in my abilities. His help and assistance over the years has not only transformed my life, it has benefited the lives of many others dependent on me. He has taught me the value of a high standard in research and shown me that I can contribute to the field as an independent researcher. I have been very blessed to be supervised by Dr. Friedman. I would also like to acknowledge my heartfelt gratitude to my dear colleague and friend Ms. Patricia Birch, whose 'motherly' help over the years has carried me through many a difficult time and who has always been a source of strength and happiness.

My sincere thanks also go out to my thesis advisory committee, Dr. Sylvie Langlois, Dr. Marco Marra and Dr. Carolyn Brown. They have guided my work, always being there with their advice and encouragement. I will always be very grateful to them.

I would like to thank Ms. Cheryl Bishop, for constantly looking out for my affairs and assisting me patiently and cheerfully through the years. My sincere thanks to Dr. Tracy Tucker, Shelin Adam and 'Ravi-mama' for their encouragement and ready support.

A very special thank you to my parents, my sister and the rest of my family for supporting me as I 'went against the grain' and for their immense struggle on my behalf so that I may immigrate to Canada to take up graduate study. And my deep thanks to my wonderful friends who have become 'family' as I've made a life in this beautiful Canadian city.

To my grandmother,
Umm Razeena Rasheed

And to the patients we are privileged to serve

Chapter 1: Introduction

1.1 Definition of Intellectual Disability

As defined by the American Association on Intellectual and Developmental Disabilities (AAIDD) guidelines, Intellectual Disability (ID) is a disorder characterized by:

- 1) *Below average intellectual functioning*, usually quantified as an IQ below 70.
- 2) *Impairment in one or more adaptive behaviour skills*. Adaptive behaviour refers to skills necessary for an individual to conduct his necessary normal daily functions. They can be sub-classified as social, conceptual or practical skills.
- 3) *A diagnosis made before adulthood*. A diagnosis made within the individual's formative years (up to age 18) is indicative that the disorder represents a condition of impaired neuronal functioning during the prenatal or postnatal development stage.

ID is separate from disorders involving impaired mental functioning that result from degeneration of once normal brain activity, as occurs in Alzheimer disease, Parkinson disease or brain trauma. ID is manifested not only by poor intellectual functioning but also by inability to conduct functions required to live a normal life. An ID patient presents developmental delay in various areas of adaptive behaviour: conceptual skills (e.g., speech, reading and writing), social skills (e.g., interpersonal relations, responsibility, appreciation, gullibility, etc.), and practical skills (e.g., eating, drinking, toileting, etc.). Therefore, ID is a condition of arrested or incomplete development of the mind and not specific to a particular type of brain function.

1.2 Etiology of Intellectual Disability

The incidence of ID of any degree (mild, moderate or severe) is about 3% [1, 2]. The etiology of ID is diverse and has undergone various classifications [2, 3]. The cause of ID can be primarily categorized as genetic or non-genetic, i.e., external insults during development (prenatal or teratogenic, perinatal or postnatal causes). In over half of cases, no etiology is recognized [3], and these cases are termed 'idiopathic'. When an etiology is identified, over half of cases are found to be due to genetic causes [4].

1.2.1 The genetic etiology of Intellectual Disability

The genetic etiologies of ID can be categorized in several ways. They can be considered in terms of whether they involve an entire genome (e.g., polyploidy, which is not viable for humans except in a mosaic state), an entire chromosome (e.g., chromosomal aneusomies, such as Down syndrome) or parts of a chromosome such as translocations, inversions, deletions and duplications. While aberrations that involve an entire chromosome are recognizable cytogenetically, those that affect only sections of a chromosome may or may not be large enough to be seen under a microscope. Submicroscopic chromosomal aberrations refer to deletions, duplications and inversions of too small a region to be detected cytogenetically. Another class of genetic defect that is causative for ID is where a single gene is affected. These single gene disorders, which show Mendelian inheritance patterns, can be due to gene sequence mutations or other genic mutations such as deletions affecting only that gene or a part of the gene.

1.2.1.1 Cytogenetically detectable genetic aberrations

Cytogenetically detectable chromosomal abnormalities account for about 10-15% of cases in most ID series [3, 5, 6]. The most commonly recognized cause of ID is Down syndrome [2], which is usually produced by trisomy 21. Other trisomies, such as trisomy 18 and trisomy 13, also account for a significant proportion of occurrence of ID syndromes recognizable at birth. Segmental aneusomies (the presence of chromosomal segments in more or less than the normal two copies, i.e., deletions and duplications) also account for many well-known ID syndromes. Structural and numerical aberrations that can be detected by microscopy have been reported in up to 15% of ID patients tested [2, 6].

1.2.1.2 Genetic aberrations not detectable cytogenetically

High resolution screens that are able to detect submicroscopic chromosomal imbalances, also called copy number variants (CNVs), have found that submicroscopic dosage imbalance due to CNVs causes ID at least as frequently as the larger cytogenetically detectable aberrations [7-9]. CNVs were first defined as DNA segments >1 kb in length that are present in fewer or more than the expected number of copies in the genome and that did not arise from the insertion or deletion of transposable elements [10, 11]. However the current recommendation is that the original size consideration of >1kb not be retained as it was a reflection of the findings of early

technologies [12]. More recent technologies have led to the frequent discovery of genomic copy number variation smaller than 1kb. Lee and Scherer discuss these definitions in depth [12]. In this thesis I will, therefore, use the term 'CNV' for genomic copy number changes even if they are <1 kb in size.

Several common clinically delineated ID syndromes are caused by microdeletions or microduplications [2], i.e., by pathogenic CNVs in critical regions. The genetic anomaly in such cases is often suspected because of characteristic clinical features and may be diagnosed by locus-specific Fluorescence In Situ Hybridization (FISH). For example, the DiGeorge/Velo-Cardio-Facial syndrome is caused by deletion of an approximately 3 Mb segment of chromosome 22q11, while duplication of the same segment causes the microduplication 22q11 syndrome [13, 14]. Smith-Magenis syndrome is caused by a microdeletion at 17p11.2, which is commonly 3.5 Mb in size [15], while the corresponding duplication of this region is causative of Potocki-Lupski syndrome [16].

1.2.1.3 Single gene mutations

A significant proportion of ID syndromes is due to single gene perturbations, either dosage change or mutation [17]. A few years ago the number of genes recognized to contribute to ID was reported as approximately 300 [17, 18]. Since then many more causative genes have been identified, as indicated by reported candidate genes from the plethora of microarray studies focused on elucidating novel genetic causes for ID that have been published (e.g.[7, 19, 20],). However these candidates have not been collated as far as we are aware. Dr. Hans Roper, in his recent overview of the genetics of ID, calculates a total number of genes implicated in autosomal ID to be at least between 800-850, based on the evidence that the 91 X-linked genes that are known to be causative of ID account for 10-12% of ID in males [2]. Genes pathogenic for ID can be classified according to their molecular function into many categories, of which two are particularly over-represented. These are: 1) genes involved in synapse formation and function, and 2) genes controlling epigenetic regulation and related transcriptional activity [18].

My thesis is focussed on dosage imbalance, i.e., copy number variations, involving genes controlling epigenetic regulation. The next section discusses epigenetics in the context of neurodevelopmental disease.

1.3 Epigenetics - definition and overview of epigenetic processes

Epigenetics is broadly defined as those heritable changes not dependent on the genomic sequence. It is therefore a method to regulate gene function without involving alteration of the genomic sequence itself. The epigenetic machinery includes factors that can ‘write’ (covalently attach), ‘read’ (differentially bind) and ‘erase’ (remove) chemical moieties to chromatin, thereby moderating genomic expression. These modifications are often dynamic and may be amenable to control. Broadly we can discuss epigenetic processes as DNA methylation, histone modifications and chromatin remodeling. In this thesis I will collectively refer to the effector proteins for these functions as epigenetic regulators.

In vertebrates, DNA methylation results in the covalent addition of methyl groups to the 5' position of cytosines and occurs predominantly at cytosines that are situated next to a guanine (written as CpG, with p reflecting the phospho-diester bond) in the DNA strand. CpG dinucleotides are found concentrated around gene promoter regions in what are termed “CpG islands” [21]. Therefore methylation of CpG islands serves as an ‘epigenetically modifiable’ mark. While most cytosines in CpG islands remain unmethylated (thus the gene is functional), methylation of CpG islands occurs in gene silencing events such as X-inactivation and silencing of imprinted genes [21, 22]. The majority of CpGs outside of islands are methylated, and while variation in such methylation may impact local chromatin structure, many current strategies [23] measuring genome-wide methylation have focused on gene promoter methylation [24].

Eukaryotic nuclear DNA is present as chromatin, which is made up of repeating units of nucleosomes. A nucleosome consists of a length of approximately 147 bp of DNA wrapped around a core of histone proteins. How tightly the DNA is wrapped around the histones impacts the amenability of the DNA to transcription. Modification of the histone protein tails can significantly alter the binding properties of DNA to the histones and the compactness of the nucleosomes to each other. Therefore such modifications serve as a key transcriptional

regulatory mechanism. For example, acetylation of lysine residues is associated with transcriptional activation while silencing is associated with certain lysine methylation signatures (e.g. K9me3 or K27me3) and other lysine methylation signatures are considered activating (e.g. K4me3)[25].

In addition there are large multiunit chromatin remodeling complexes that are necessary for the assembly or displacement of nucleosomes, and also serve to insert variants of the histones, which alter chromatin compactness. The net outcome of these processes is to render chromatin as transcriptionally active or euchromatic, or conversely transcriptionally inactive or heterochromatic. The effects of chromatin remodeling on neurodevelopment are reviewed in depth by Yoo and Crabtree [26].

DNA methylation, histone modifications and chromatin remodeling complexes work together and there are significant interactions in their recruitment. This 'cross-talk' is multi-directional and multi-modal [27, 28], for instance, a DNA methylase can recruit a histone modifier which in turn can recruit chromatin remodeling complexes. Not only the effector molecules (e.g, DNA methyl transferase), but also the regulatory marks themselves (e.g., DNA methylation) can be involved in recruiting other regulators. Figure 1.1 illustrates the interplay between these factors.

In summary we see that epigenetic regulation is a complex and intricate process. It is a finely orchestrated system involving the synchronized working together of many diverse proteins, often in large multi-component complexes that act upon vast portions of the genome. Therefore even small changes in the balance of factors comprising the machinery may be pathogenic.

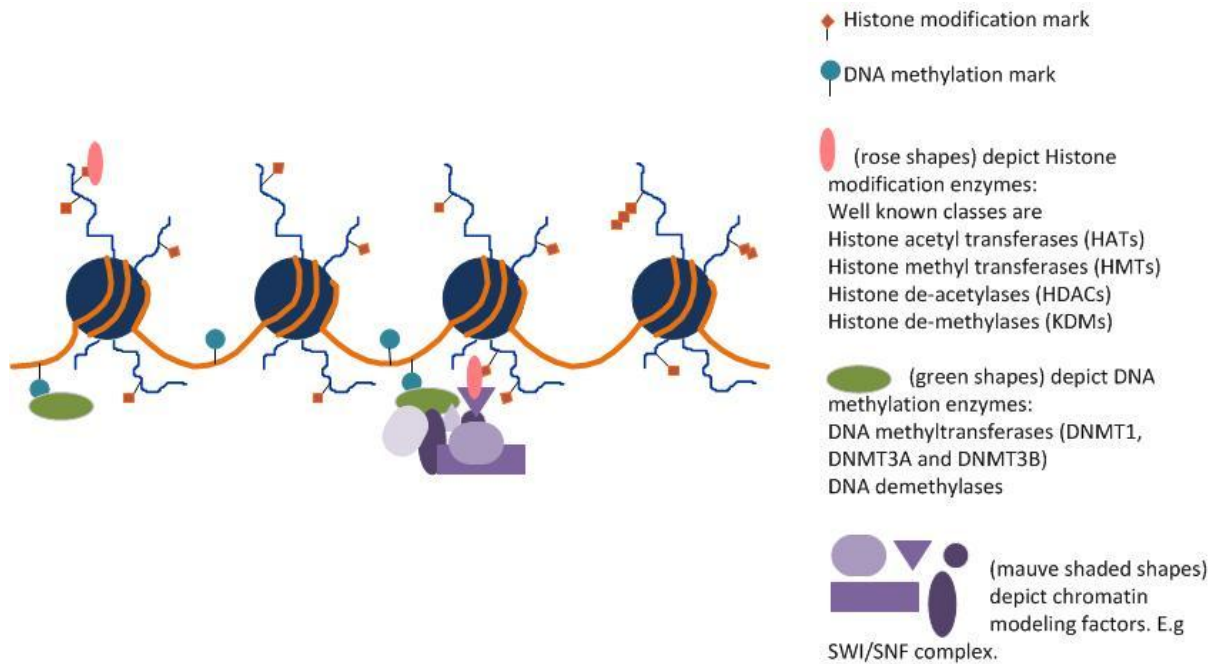


Figure 1.1: Illustrating interactions between DNA methylation, histone modification and chromatin remodeling. The DNA strand (orange) is wrapped around histone protein cores (large blue circles), to form repeating nucleosomes that make up chromatin. Histone tail modifications are depicted as orange diamonds attached to histone tails, and DNA methylation marks are depicted as teal circles attached to the DNA strand. Epigenetic regulators that have DNA methylation (green shapes), histone modification (rose shapes) or chromatin remodeling (mauve shaded shapes) interact with each other and display cross-recruitment.

1.4 Epigenetic perturbation in neurogenetic disorders

Over a decade following the availability of the human genome sequence [29], we are still far from understanding the genetic basis of many neurodevelopmental disorders [2, 30].

Epigenetics is growing in prominence as a significant contributor to the etiology of diseases [31, 32]. A growing body of work is highlighting the extent of epigenetic involvement in neurological disease [30-34].

Several well-characterized ID syndromes are caused by perturbation of genes involved in epigenetic regulation [18]. For example, Rett syndrome, one of the most common causes of ID in women [35], is due to mutations in the *MECP2* gene [36]. The protein encoded by this gene is a member of the methyl CpG-binding domain (MBP) protein family and is an epigenetic regulator of transcription [37]. Rubinstein-Taybi syndrome is caused by mutations in the *CREBBP* gene [38]. CBP, the protein produced by this gene, is known to have intrinsic histone acetyltransferase activity and is also a transcriptional co-activator. Coffin-Lowry syndrome is caused by mutations of *RSK2* [39], which codes for a modulator of the CBP protein. Loss of function mutations of *DNMT3B* are causal for another ID syndrome, ICF (Immunodeficiency-Centromeric instability-Facial anomalies) syndrome [40]. DNMT3B is a DNA methyltransferase necessary for methylation of cytosine residues. ATRX (Alpha Thalassemia- X linked MR) is caused by mutations of the *ATRX* gene, which codes for a protein that is a chromatin remodeler [41]. *CHD7*, which codes for an ATP-dependant chromatin remodeling enzyme, is the causative gene for CHARGE syndrome, characterized by a non-random pattern of congenital anomalies including heart, ear and eye phenotypes in addition to ID [42]. In addition to these well-known pathologies, some of the genes above are also being discovered to be causative of rarer phenotypes. Table 1.1 gives a list of genes known to encode epigenetic regulators that have been implicated as causative for neurodevelopmental pathologies when perturbed.

The examples mentioned above are monogenic disorders that may be defined as those caused by a disruption of epigenetic regulation due to dysfunction of genes encoding epigenetic regulators. Not unexpectedly, genes that are under epigenetic regulation can also be candidates for causing ID due to disruption of the elements involved in the recruitment of the epigenetic mark. An important example of this effect is the deregulation of imprinted genes.

Imprinting refers to when a gene, though present in two copies, is only expressed by one chromosome, dependent on parent of origin, i.e., a preferential transcription for either the maternal or the paternal allele [43]. Imprinting is currently known for a handful of human genes[43]. The parent-of-origin dependent expression is regulated by which parental allele is differentially methylated, thus imprinting is an effect brought about by epigenetic control. Prader-Willi and Angelman syndrome both include ID in their phenotypic spectrum, and are caused by parent-of-origin specific defects of 15q11-q13 [44]. Buiting [45] reviews the fascinating causative mechanisms for these disorders.

Intriguingly, copy number variation of the 15q11-q13 region is also associated with autism [46-48], and there is emerging evidence for autism spectrum disorder symptomatology presenting with the imprinting disorder Prader-Willi Syndrome [47]. Other ID syndromes are also now being understood to include autism phenotypes (e.g., 16q11.2 microdeletion and microduplication syndrome [49], 22q13 microdeletion [50]). Genes implicated in ID syndromes are also being found to be pathogenic for autism (e.g. the *MECP2* gene was thought to be lethal if deleted in males; however, there is now evidence of milder phenotypes including autism behavioural phenotypes manifesting in males with deficient *MECP2* (Table 1.1 and [51])).

Therefore we see that defects of genes encoding epigenetic regulatory factors are a recognized cause of neurodevelopmental disease. The causative gene defects can be of many types, with dosage imbalance, i.e., the loss or gain of genetic material such that the gene is disrupted from its normal two-copy state, also implicated in several instances. We discuss dosage imbalance for genes encoding epigenetic regulators in more detail in the following section.

Table 1.1: Genes encoding epigenetic regulators that have been implicated in neurodevelopmental pathologies. DMT= DNA methyltransferase. HMT= Histone methyltransferase. HAT= Histone acetyltransferase. HP=Histone phosphorylation. HD= Histone demethylase. HDAC = Histone deacetylase. DMD-CR= DNA methylation dependant chromatin remodelling. CR=chromatin remodeling protein.

Gene	Protein	Epigenetic Class	Pathogenicity	OMIM #	Defect	Selected References
<i>DNMT3B</i>	DNMT3B	DMT	Immunodeficiency, Centromeric instability and Facial Dysmorphisms (ICF) Syndrome	242860	homozygous or compound heterozygous mutations	Okano et al. 1999[52], Xu et al. 1999[40]
<i>NSD1</i>	NSD1	HMT	Sotos Syndrome	117550	heterozygous deletions and truncating mutations	Kurotaki et al. 2002[53]
<i>EHMT1</i>	EHMT1	HMT	9q Subtelomeric Deletion syndrome	610253	heterozygous deletions and truncating mutations	Kleefstra et al. 2009[54]
<i>CREBBP</i>	CBP	HAT	Rubinstein-Taybi Syndrome	180849	heterozygous microdeletions and truncating mutations	Petrij et al. 2000[55]
<i>CREBBP</i>	CBP	HAT	Rubinstein-Taybi Syndrome	180849	heterozygous deletions	Stef et al. 2007[56], Gervasini et al. 2007[57]
<i>CREBBP</i>	CBP	HAT	Incomplete Rubinstein-Taybi Syndrome	180849	heterozygous missense mutation	Bartsh et al. 2002[58]
<i>CREBBP</i>	CBP	HAT	16p13.3 Duplication syndrome	613458	heterozygous duplications	Thienpont et al. 2010[59]
<i>EP300</i>	P300	HAT	Rubinstein-Taybi Syndrome	180849	heterozygous mutations	Roelfsema et al. 2005[60]
<i>RPS6KA3</i> (X linked)	RSK2	HP	Coffin-Lowry Syndrome	303600	heterozygous deletions, nonsense and missense mutations	Trivier et al. 1996[39], Delaunoy et al. 2006[61]
<i>RPS6KA6</i> (X linked)	RSK4	HP	Non Syndromic MR	300303	Deletion	Yntema et al. 1999[62]
<i>PHF8</i> (X linked)	PHF8	HD	Siderius X-linked MR Syndrome	300263	deletions and mutations	Laumonnier et al. 2005[63]
<i>PHF8</i> (X linked)	PHF8	HD	ASD and ID	300263	deletion encompassing other genes	Qiao et al. 2008[64]
<i>HDAC4</i>	HDAC4	HDAC	Brachydactyly-MR Syndrome	600430	heterozygous deletions	Williams et al. 2010[65]
<i>HDAC4</i>	HDAC4	HDAC	SCZ	300055	Association	Kim et al. 2010[66]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	Rett Syndrome	312750	deletions and severe loss of function mutations (females)	Amir et al. 1999[67], Smeets et al. 2005[68]

Gene	Protein	Epigenetic Class	Pathogenicity	OMIM #	Defect	Selected References
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	Severe Neonatal Encephalopathy	300673	severe loss of function mutations (males)	Schanen et al. 1998[69], Hardwick et al. 2007[70]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	ASD	300496	severe loss of function mutations (females)	Carney et al. 2003[71]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	X linked MR	300055	mild loss of function mutations (males)	Couvert et al. 2001[72]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	X linked MR and MECP2 Duplication Syndrome	300260	duplications (males)	Lugtenberg et al. 2009[73]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	ASD	300496	over and under expression	Nagarajan et al. 2006[74], Ramocki et al. 2009[75]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	SCZ	300055	non-synonymous mutations	Cohen et al. 2002[76]
<i>MECP2</i> (X linked)	MeCP2	DMD-CR	Angelman Syndrome	105830	mutations	Watson et al. 2001[77]
<i>ATRX</i> (X linked)	ATRX	CR	Alpha-Thalasemia X linked MR	301040	mutations and intragenic duplications leading to loss of function	Gibbons et al. 1995[78], Thienpont et al. 2007[79]
<i>ATRX</i> (X linked)	ATRX	CR	MR-Hypotonic Facies Syndrome, X-linked	309580	mutations	Villard et al. 1996[80], Abidi et al. 1999[81], Villard et al. 2000[82]
<i>CHD7</i>	CHD7	CR	CHARGE Syndrome	214800	heterozygous deletions and truncating mutations	Vissers et al. 2004[42], Lalani et al. 2006[83]
<i>JARID1C</i> (X linked)	JARID1/ SMCX	CR	X linked MR	300534	mutations (males)	Jensen et al. 2005[84], Abidi et al. 2008[85]
<i>PHF6</i> (X linked)	PHF6	CR	Borjeson-Forssman-Lehmann syndrome	301900	truncating and missense mutations (males and one report of female)	Crawford et al. 2006[86]
<i>ZEB2</i>	ZEB2	CR	Mowat Wilson syndrome	235730	heterozygous deletions	Cacheux et al. 2001[87]
<i>ZEB2</i>	ZEB2	CR	Hirschprung disease-MR Syndrome (Mowat-Wilson syndrome variants)	235730	heterozygous mutations (often truncating)	Zweier et al. 2006[88], Heinritz et al. 2006[89]
<i>REST</i>	REST	CR	Down Syndrome	190685	reduced expression	Bahn et al. 2002[90]
<i>CDKL5</i> (X linked)	CDKL5/ STK9	CR	Atypical Rett Syndrome, Infantile Spasms, and Severe MR	300672	mutation (female)	Weaving et al. 2004[91]

1.4.1 Dosage change of genes encoding epigenetic factors as the mechanism of pathogenicity

Many ID syndromes are due to deletion or duplication of specific regions of the genome and therefore are caused by abnormal gene dosage. Autosomal microdeletion syndromes, in which one or both copies of a small genomic region or single gene are deleted, result in disease due to haploinsufficiency in the hemizygous state or complete loss of function in the homozygous null state. Microduplication syndromes are the converse, in which disease results from the presence of an extra copy of small genomic region or gene.

Several ID syndromes are microdeletion syndromes [2], some of which include dosage change of genes controlling epigenetic regulation. Over 15% of Rubinstein-Taybi cases are caused by deletions [92] of the epigenetic regulatory gene *CREBBP*. Causative genes for CHARGE syndrome and the recently characterized 9q34 syndrome were first identified in microdeletion cases for these syndromes and validated upon identification of sequence mutations in the same genes in other cases [42]. *CHD7* and *EHMT1* are the causative genes for CHARGE and 9q34 syndromes respectively. They are both active in epigenetic regulation, the former is a member of the chromodomain family of proteins which are important in chromatin remodelling, and the latter is a histone modifier.

Genes which have epigenetic function can also be pathogenic for ID when present in dosage excess to the normal. Rett-like syndrome is an ID syndrome caused by a duplication or triplication of a < 1 Mb segment of chromosome Xq28 involving the *MECP2* gene which has DNA methylation-specific transcription factor activity. Numerous other ID syndromes result from microdeletions and microduplications and have been tabulated by Dr. Hans Ropers in his review of genetic causes of ID[2]. Therefore there is strong evidence supporting dosage change as the mechanism of pathogenicity for ID.

Microarrays have been at the vanguard of the drive to detect submicroscopic dosage changes, thereby also defining novel microdeletion and microduplication syndromes. There are many techniques by which microarrays are utilized for this purpose, two of the most common are termed array genomic hybridization (AGH) and array comparative genome hybridization

(aCGH). The technical limits and strengths of these methods and their utilization toward detecting CNVs are discussed next.

1.5 Microarray technology

Microarray technology is useful for a range of studies [93], among which detection of potentially pathogenic submicroscopic copy number variants (CNVs) has become a main focus. aCGH originated as an improvement of cytogenetic comparative genomic hybridization (CGH). CGH in turn developed from FISH (Fluorescence In Situ Hybridization), as a way of identifying deletions or duplications genome-wide in a single experiment [94, 95]. Metaphase CGH involves the isolation and differential fluorescent labeling of total genomic DNA from test (patient) and reference (normal) samples and hybridization in equal amounts to a normal metaphase chromosome spread. The differential labeling reveals genomic copy number gains or losses in the test DNA, as regions of inequality in the corresponding fluorescent signal, in comparison to that of the reference DNA. A major limitation of this method is that its resolution is no better than that of conventional cytogenetic analysis, typically about 10 Mb for copy number gains and over 10Mb for losses [96].

aCGH overcomes the limited resolution of cytogenetic CGH by using an array of DNA segments representing or covering a reference genome. Since the first arrays for CGH were developed [97, 98], many variations on this common theme have been used to detect CNVs.

aCGH technologies can be grouped according to the nature of the DNA fragments (called “targets” or “probes”) on the array and whether the genomic comparison is made by hybridization of differentially-labeled test and reference DNA samples on the array (aCGH), or by *in silico* comparison of the signals produced by array hybridization of a single labeled test DNA to similar data from reference samples (AGH). Metaphase CGH, aCGH and AGH are compared in Figure 1.2


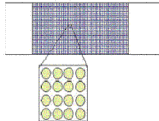






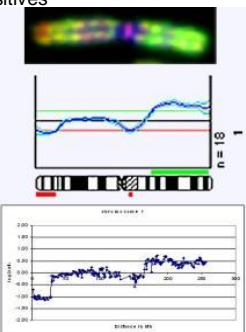
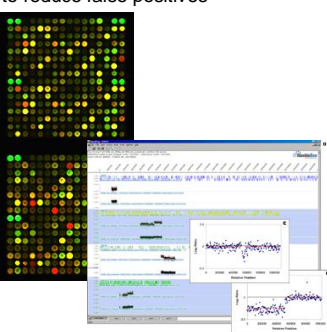
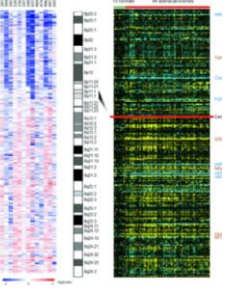
Technology	Cytogenetic CGH	AGH with Sanger 1Mb BAC array	AGH with Affymetrix GeneChip®
Target	Normal metaphase spread 	3500 BACs at ~1Mb intervals throughout the genome, spotted onto a glass slide. 	100,000 sets of oligonucleotides, synthesized <i>in situ</i> on a chip 
DNA samples, differentially labeled	Test sample and reference sample labeled in 2 different fluoro-chromes Test  Control 	Test sample and reference sample labeled in 2 different fluoro-chromes Test  Control 	Test sample labeled with single fluoro-chrome. Test 
Hybridization	Test and reference samples hybridized in equal amounts to targets in presence of Cot-1 DNA	Test and reference samples hybridized in equal amounts to targets in presence of Cot-1 DNA	Test sample hybridized to targets using standard conditions
Data analysis	Relative intensity of test and reference sample fluorescence along chromosome used to detect gains and losses; second hybridization with dye flip may be used to reduce false positives 	Location of each BAC mapped back to genome; relative intensity of test and reference sample fluorescence for each BAC used to detect gains and losses; second hybridization with dye flip used to reduce false positives 	Location of each target mapped back to genome; data normalized and compared <i>in silico</i> to reference standard; statistical evaluation and congruency among adjacent targets used to reduce false positives 

Figure 1.2: Comparison of metaphase CGH, aCGH and AGH. Note: Cot-1 DNA is human DNA enriched for repetitive sequences. It functions to reduce non-specific hybridization by competitively binding to repetitive sequences present in the hybridization mixture.

As previously discussed, we refer to submicroscopic deletions and duplications detected by aCGH and AGH studies as “copy number variants” (CNVs). An important advantage of calling these changes “CNVs” as opposed to “microdeletions” or “microduplications” is that the term “CNV” is neutral with respect to both pathogenicity and frequency of occurrence. The pathogenicity of a CNV detected in a patient with ID may not always be clear and in some cases may be contingent on other factors such as parent of origin, the state of the normal alleles, epigenetic factors, or genetic background. Resolution of this issue is discussed in detail below. The frequencies with which many CNVs occur are not yet known, but in instances in which the population frequency has been shown to be >1% and the change *per se* does not cause disease, the term “copy number polymorphism” may be used instead of “copy number variant”.

1.5.1 Types of arrays and factors that influence their effective resolution

There are a number of microarray platforms that can be used for AGH or aCGH. They can be categorized by the content, type and number of arrayed elements and by how the array is manufactured. Figure 1.3 summarizes these issues, which will be discussed below.

Greater resolution is the *raison d'être* for using AGH or aCGH to identify CNVs in children with ID. Arrays with more elements providing more comprehensive and higher resolution whole genome coverage may detect more pathogenic CNVs but also detect many more normal or benign variants that must be distinguished from those that cause disease. In addition, false positive calls occur more often when there are more elements on an array. Therefore, greater resolution does not necessarily translate into more meaningful data.

Use of a so-called “targeted” array for AGH or aCGH provides an alternative that avoids many of these issues because such arrays include fewer probes, most of which detect CNVs that are known to be associated with ID or other important clinical abnormalities. However, like locus-specific FISH and multiplex telomere tests, targeted assays cannot detect CNVs of genomic regions that are not represented in the assay.

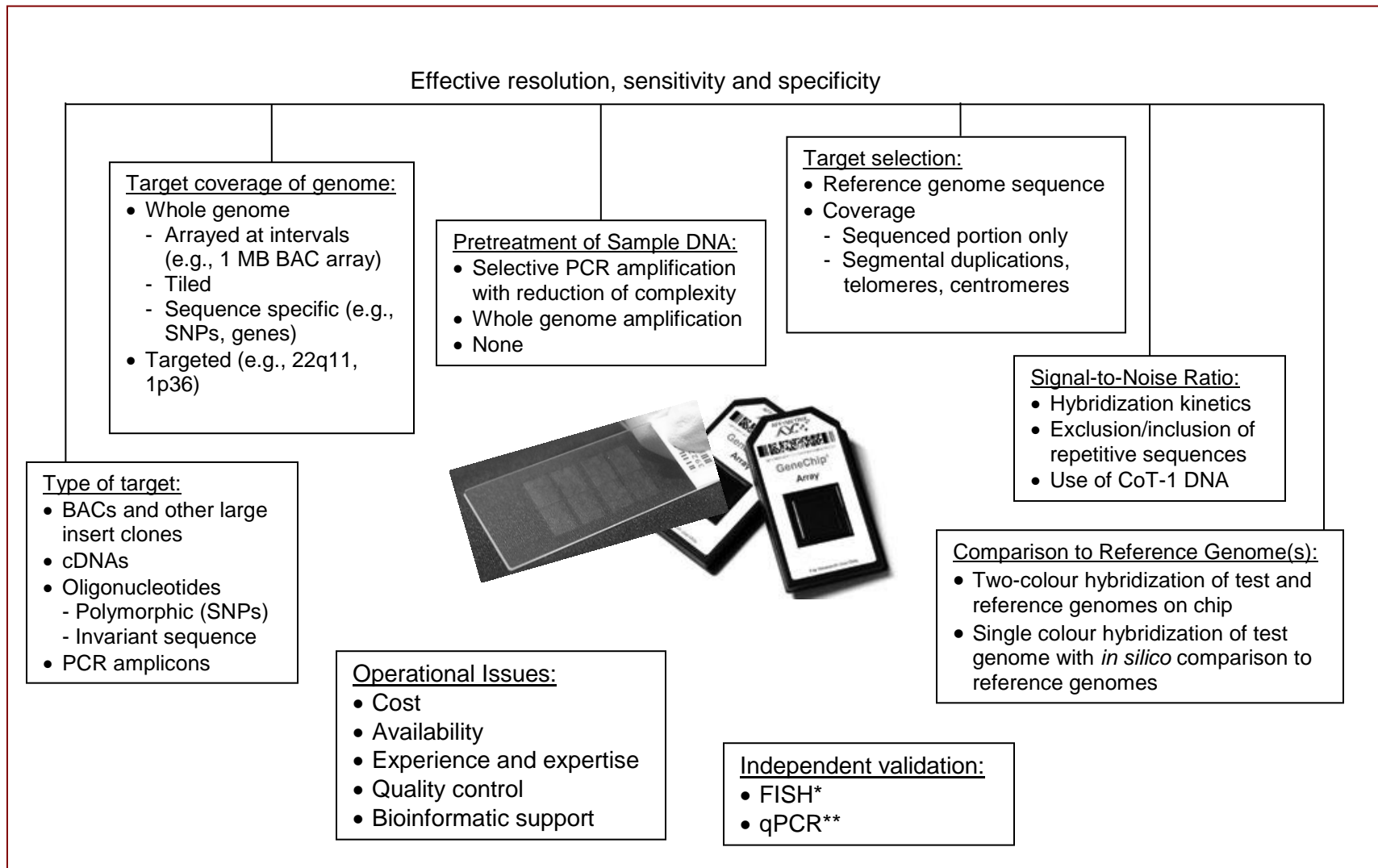


Figure 1.3: Factors to be considered when choosing an AGH or aCGH platform. * FISH – Fluorescence In Situ Hybridization, **qPCR – Quantitative Polymerase Chain Reaction[99]

Achieving higher resolution depends on a number of factors: type and size of the DNA used as array targets, genetic content of the targets, distribution of target sequences across the genome, and specificity of the hybridization reaction. Several different types of DNA fragments have been used as array targets, including BAC, PAC and YAC clones [97, 100, 101], single- [102] and double-stranded oligonucleotides[103], cDNAs[104] and PCR products[105]. Of these, BAC and oligonucleotide arrays have been the most widely used for whole genome studies in idiopathic ID.

Oligonucleotide targets range in size from 20-60 bp, but, in general, the shorter the target the less specific the hybridization. This would be expected to reduce the signal-to-noise ratio, although other factors in array (or chip) design may compensate, such as placing a large number of identical oligonucleotides in each spot, using a set of overlapping oligonucleotides to measure a particular sequence, or optimizing the hybridization conditions.

The resolution of an array depends on the way the targets are distributed across the genome as well as on the type of target used. The earliest arrays achieved a whole genome resolution of ~1 Mb by using approximately 3000 BAC clones spaced across the genome at 1 Mb intervals [106, 107]. Later improvements resulted in a resolution less than the size of an individual BAC obtained by tiling approximately 30,000 BAC clones with overlap across the genome [100, 108]. Oligonucleotide arrays are also able to achieve genome-wide resolutions of at least 50-100 Kb [101, 109-111] and they have proved to be the array of choice with the introduction of their availability commercially. Currently available commercial arrays can have up to 2 million oligonucleotide probes yielding possible resolutions of <10 Kb [112].

The information generated from a target depends not only on its size but also on its composition. DNA targets of a given size that contain only unique sequence would be expected to have a higher signal-to-noise ratio than those containing substantial amounts of repetitive sequence. Large insert clones such as BACs and non-polymorphic oligonucleotide targets can only provide information on copy number. SNP arrays, on the other hand, simultaneously provide information on genotypes, which can be used to create a complementary screen for loss of heterozygosity, parent of origin and Mendelian errors if family data are available [113] [114].

Pretreatment of sample DNA prior to hybridization can also influence the resolution and signal-to-noise ratio. Sample DNA may be pretreated to reduce the concentration of repetitive sequences, reduce genomic complexity, increase the concentration of DNA in the hybridization mixture, or increase the concentration of fragments that correspond to target sequences on the array. “Genomic representation” is used to improve hybridization kinetics in AGH studies done with platforms such as the Affymetrix 100K and 500K GeneChips®[115]. The usual approach is to digest sample DNA with a restriction endonuclease and use PCR to amplify sequences of a particular size selectively [101, 116]. In order for a genomic representation produced by this method to be useful in assessing copy number, every DNA fragment that is selected must be amplified to the same extent. The process by which elements of a particular type were selected for inclusion in an array must also be considered. The targets are usually chosen from a human reference genome sequence, which means that the array will not include targets for gaps in the assembly. Almost half of sequence assembly gaps occur in areas of sequence duplication [117], and CNVs are more frequent in regions that are enriched for segmental duplications [118-120]. Array platforms with elements chosen exclusively from the human reference sequence, therefore, may not be able to detect all CNVs that are present in a DNA sample. In addition, arrays based entirely on unique SNPs may offer poor coverage of telomeric and centromeric regions, which contain many repeat elements and few SNPs within sufficiently long DNA sequences to be unique within the genome. Similarly, arrays that focus on coding sequence may not detect variants in non-coding regions that could potentially involve important regulatory elements.

1.5.2 The bioinformatics of AGH

Because of the large volumes of data generated by high-resolution genome scans, bioinformatic analysis is essential for visualization and interpretation of the results. Three different kinds of bioinformatic processing are employed in microarray experiments. The first involves obtaining the hybridization image, identifying the array target that corresponds to every portion of the scanned image, and reading the hybridization signal associated with that target properly.

The second step is converting the raw image data into information regarding copy number at particular points in the genome. This usually involves both data normalization and smoothing. Normalization refers to the process of making data generated by an aCGH or AGH experiment independent of the variation caused by the particular experimental conditions [121]. With chips

that scan the genome at very high resolution, a high level of “noise” may be generated by chance fluctuations over tens or hundreds of thousands of data points, and “smoothing” is carried out to increase the signal-to-noise ratio. A number of statistical approaches that can be used for normalizing and smoothing AGH data are succinctly discussed by Pinkel et al. [122].

Most commercially-available aCGH or AGH systems include user-friendly software tools, and software for AGH has also been developed by third parties [121, 123-131]. The performance of these programs varies substantially, and choice and optimization of the software can have a profound effect on the sensitivity and specificity of AGH analysis.

The third task that must be accomplished by bioinformatics is data visualization. The distribution of targets on an array is usually randomized with respect to genomic location to minimize the effects of local hybridization artifacts on the analysis. However, the clinician needs to understand where in the genome a CNV is and how large a genomic region is involved. Various bioinformatic tools have been developed to convey this information, and they may also provide additional useful data like the predicted copy number at each site, the gene content of a region involved in a CNV, the location of known polymorphic CNVs, or the likelihood that a particular CNV represents a chance fluctuation in hybridization intensity.

A key issue in the analysis is how many consecutive targets (i.e., target sequences on the array that lie consecutively within the genome) are used to “call” a CNV. For example, calls can usually be made on the basis of the information from a single target on a 3000 element (1 MB) BAC array, but several consecutive targets showing the same hybridization pattern (i.e., an increased or decreased signal) are needed to call a CNV with confidence on a high-density oligonucleotide array [132, 133]. Requiring a greater number of consecutive targets showing the same hybridization pattern usually increases the specificity and decreases the sensitivity of a call to different degrees depending on the characteristics of the array, the hybridization reaction and the software.

1.5.3 Comparison to reference genome(s)

Interpretation of AGH involves comparison of the patient hybridization data to a reference standard. This can be done on the chip, using the approach that was originally developed for

cytogenetic CGH, by hybridizing equal amounts of differentially-labeled test (patient) and reference (control) DNA simultaneously [134-137]. Alternatively, the comparison can be done *in silico* by comparing the intensities obtained by hybridization of the sample to the results obtained from reference individuals using the same kind of chip and the same hybridization conditions [132].

Whichever method is used, the question of whose DNA to use as the reference genome arises. There are several possibilities:

- All test hybridizations can be compared to a ***single normal reference individual***. This is a practical approach that has the advantage of making all analyses comparable. Normal (polymorphic) CNVs that happen to be present in the reference sample can confound interpretation of similar regions in the test sample, but these changes can be discounted once the reference DNA has been used enough to recognize all of the normal variants that are present. One limitation is that DNA obtained from a fresh blood sample of any individual will eventually run out, and the individual may not always be available to provide another sample. In addition, there are reports of the occurrence of somatic CNVs that may produce batch-to-batch variability in the results obtained from a single individual [120]. This problem can be ameliorated by preparing a cell culture from the reference subject, although cell lines can accumulate culture-induced rearrangements over time [114].
- A variation on the same theme is to use a ***single male reference DNA for all female test samples and a single female reference DNA for all male test samples***. This provides the advantage of having an X-chromosome copy number difference that can serve as an internal control in all normal samples [134, 138-140]. Otherwise, using the same two reference samples for all experiments offers most of the same advantages and limitations as using just one, although the male and female reference samples will have different benign CNVs.
- Using ***normal family members*** is a particularly appealing approach in intellectual disability studies. If an affected child's sample is compared to samples from both of his or her normal parents, inherited CNVs in the child can be distinguished from those that arose *de novo*. Adding a comparison of the parents to each other permits CNVs that are present in one parent to be distinguished from those that are present in the other [132]. The observation that a CNV in a child with intellectual disability has occurred *de novo* may be an important clue to the pathogenicity of the lesion but is often insufficient *per se* to either prove or exclude a causal relationship. This issue is discussed in detail below.

- Using a “pooled” reference standard that contains DNA from **many different normal donors** offers the advantage of “averaging out” all benign CNVs that are not extremely common in the population and of providing a large amount of DNA that can be used for many experiments. However, even pooled samples may run out, and using fresh pools allows for batch-to-batch variability due to culture induced variation [114] or somatic variation [120] if the same individuals are used as donors, or due to the pool containing a different set of benign CNVs if the donor individuals change.

This list is not exhaustive, and some investigators have used somewhat different reference comparisons. The best reference sample may depend on the nature of the analysis that is being performed, and there is no consensus on the best approach. An important advantage of *in silico* comparisons as opposed to comparisons based on co-hybridization on the chip is the ability to compare the hybridization data from a single test sample to several different reference samples. This is useful because different comparisons may reveal different things about the genomic content of the test sample [132]. Using *in silico* comparisons also means that the results are not heavily dependent on the sample chosen as the reference and that an unlimited number of test samples can be compared to the same reference without running out of DNA.

1.5.4 Conclusions regarding use of arrays to identify CNVs

The most effective platform, optimal resolution and best distribution of elements for whole genome aCGH and AGH in the diagnosis of ID have not yet been established. An alternative approach that is being used clinically is AGH with a targeted microarray designed to detect CNVs that are known to cause intellectual disability while avoiding many of the complications inherent in whole genome AGH [139-141]. The advantage of this approach is its practicality – minimizing the likelihood of picking up benign variants and CNVs of unknown clinical significance. The major disadvantage is that most previously uncharacterized pathogenic CNVs will not be identified. In our studies, we have used both strategies to identify potentially pathogenic CNVs in children with idiopathic ID. Findings from these studies and subsequent genotype-phenotype correlations for detected aberrations will be presented in this thesis.

1.6 Independent confirmation of CNVs detected by AGH

Given the limitations of AGH, validation of submicroscopic CNVs by FISH or another independent method like qPCR (quantitative PCR) is recommended, especially in instances

where the detected CNV is not known to be recurrent and pathogenic for a clinically identified condition [132, 142]. FISH is used in instances where the lesion detected by AGH is large enough for a FISH probe to hybridize, producing sufficient fluorescence to be detected microscopically, while qPCR is the method of choice when the lesion is too small to be “FISH-ed”. qPCR is also useful for the confirmation of duplications for which FISH confirmation may be difficult, especially in metaphase preparations. FISH and qPCR can be applied to any unique locus within a region of interest and are, therefore, ideal for confirmation of novel CNVs for which no standard assay exists. Other PCR-based techniques, such as multiplex ligation-dependent probe amplification (MLPA) [143] or multiple amplifiable probe hybridization (MAPH)[144] can also be used when the region of interest includes one of the loci available in the probe set. These techniques, which are designed to assess copy number at multiple loci in a single reaction by using sequence-specific probe hybridization and amplification by universal primers [145-147], are less easily adaptable to a novel CNV in an individual family. AGH on a different platform can also be used to confirm the findings of a CNV call. This method has been widely used by groups undertaking large-scale validation studies [112, 114].

1.7 Characterizing the pathogenicity of a CNV: Genotype-Phenotype correlations

It is now recognized that normal individuals also carry many CNVs, which are apparently benign, or at least do not by themselves cause ID or major malformations. Every normal person carries a surprisingly large number of apparently benign CNVs. Estimates range from upwards of 1000 CNVs per individual involving from ~24Mb of sequence (considering CNVs >1kb) to up to 48Mb of sequence (considering CNVs of all sizes) [112, 148].

CNVs in normal individuals have been reported to include up to ~3% of RefSeq genes [112, 148]. Although found in normal people, these CNVs might, nevertheless, be associated with disease. For example, a polymorphic CNV of the *CCL3L1* locus is associated with susceptibility to HIV infection[149]. Many disease-associated genes are found within benign CNVs [114, 142]. Redon *et al.* calculated that a staggering 14.5% of genes in the OMIM morbid map overlap with CNVs detected in normal individuals [114]. Although benign CNVs appear to be preferentially located outside genic and ultra-conserved regions [114], many functional genes have been identified within CNVs. Genes related to sensory perception, including members of the highly redundant olfactory receptor family [150, 151], have frequently been found within CNVs [114, 142]. Other types of genes reported to be over-represented in CNVs include those involved in cell adhesion [114] and immunity [142] which are also gene families

with many members, leading us to believe loss or gain of these genes are well tolerated in the population. Therefore it is necessary to distinguish between benign and pathogenic structural variation once a CNV is identified. Strategies to do so are discussed below.

1.7.1 Differentiating between benign CNVs and those that are pathogenic for Intellectual Disability: *Inheritance status*

One factor that can often (but not always) help to distinguish pathogenic and benign CNVs is the inheritance status [132, 134, 136, 152]. If a CNV is inherited from a normal parent, it is likely to be benign, while if it arose *de novo* in a child with intellectual disability, it is more likely to be pathogenic. Therefore, testing both biological parents as well as the child with intellectual disability is often very important for interpretation of AGH studies. However, some pathogenic CNVs are inherited rather than occurring *de novo*. For example, a pathogenic microdeletion of Xp21 has been reported to have been transmitted to a boy with intellectual disability from his normal mother, in this case the mutation which is recessive would not manifest in the mother but would be causative in the male [153].

Inherited CNVs that contain an imprinted locus may also be associated with intellectual disability in a child but not in his or her parent. For example, a child with Angelman syndrome was found to have a 570 Kb microdeletion of chromosome 15(q11.2q12) involving the imprinted *UBE3A* gene [154]. The child inherited this CNV from his normal mother, who, in turn, inherited it from her father. *UBE3A* is expressed on the maternal but not the paternal chromosome, so the mother had a normally active copy on her maternal chromosome. The gene is repressed on the paternal chromosome, so the child's only normal allele, which was on his paternal chromosome, was inactive.

Inheritance of a submicroscopic deletion from one parent and an allelic sequence mutation from the other can cause an autosomal recessive form of intellectual disability. A good example of this was found in a family in which two boys with the autosomal recessive Peters Plus syndrome were found by AGH to have a ~1.5 Mb deletion of chromosome 13(q12.3q13.1) [155]. Both affected boys also had allelic splice-site mutations of *B3GALT*, which lies within the CNV and is the locus that is mutated in Peters Plus syndrome. The children inherited the microdeletion of chromosome 13 from their normal mother and the *B3GALT* splice-site mutation from their normal father. Rare cases have also been reported in which a CNV that

causes intellectual disability in a child has been inherited from a normal parent who is a gonadal mosaic for the CNV [156].

On the other hand, benign CNVs are usually inherited, although they may occur *de novo* [150, 157]. Redon *et al.* estimate that less than (and probably substantially less than) 0.2% of benign CNVs occur as *de novo* events [114].

1.7.2 Differentiating between benign CNVs and those that are pathogenic for Intellectual Disability: *connection to known pathogenic genetic lesions*

Compelling evidence that a CNV in an individual is pathogenic is present if the submicroscopic deletion or duplication is known to cause the patient's phenotype, e.g., if a child with features of Williams-Beuren syndrome has a 7q11.2 deletion or a child with a 22q11.2 duplication has features of the syndrome associated with that pathogenic CNV [13, 140]. Similarly, pathogenicity is supported when a CNV includes a gene that is known to cause the patient's phenotype when inactivated (if the CNV is a deletion) or over-expressed (if the CNV is a duplication). For example, a deletion encompassing the *NSD1* gene, haploinsufficiency of which is known to cause Sotos syndrome, was found by AGH in a patient with features of this phenotype [158].

1.7.3 Differentiating between benign CNVs and those that are pathogenic for Intellectual Disability: *other considerations*

Known functions and expression patterns of involved genes can also indicate whether or not a CNV is likely to be pathogenic. Genes that control synapse formation and function, transcription, or embryonic development have been suggested as good candidates for involvement in intellectual disability syndromes [18]. Whether the genes are dosage sensitive or members of the same family or pathway as previously-established intellectual disability genes may also be considered. For example, cardio-facio-cutaneous (CFC) syndrome has phenotypic overlap with Noonan syndrome and Costello syndrome, which are caused by mutations of genes encoding members of the RAS signaling pathway. Mutations that cause CFC were identified by screening genes that encode members of this pathway in CFC patients [159].

Comparing a CNV found in a child with intellectual disability against databases such as the Toronto Database of Genomic Variants (<http://projects.tcag.ca/variation>) or the Human Structural Variation Database (<http://humanparalogy.gs.washington.edu/structuralvariation/>) may be informative because CNVs that occur as benign polymorphic variants are unlikely to be pathogenic in a child with intellectual disability. Comparing the extent of patient CNVs that overlap but are larger than those seen frequently in normal individuals can also be used to hone in on candidate genes, which are likely to lie outside of the polymorphic region.

1.7.4 Summary: novel pathogenic CNVs, new Intellectual Disability syndromes and candidate genes

Traditionally, genotype-phenotype correlations for patients with intellectual disability have been identified by recognizing a clinical syndrome among children who share a similar pattern of congenital anomalies and then demonstrating a recurrent genetic abnormality among patients who have this clinical syndrome. This was the sequence of events in Down syndrome [160, 161], Allagille syndrome [162-164] and dozens of other dysmorphic syndromes. However, as more and more patients are studied by AGH, it has become apparent that subtle patterns of anomalies that have not previously been recognized clinically may be shared by individuals who have similar CNVs. The deletion 17q21.31 syndrome [4, 165, 166] and deletion 12q14 syndrome [167] provide excellent examples of how such “reverse dysmorphology” can establish a genotype-phenotype correlation.

Most apparently-pathogenic CNVs reported in whole genome aCGH or AGH studies of patients with intellectual disability have been unique, so there are no patients known to have similar submicroscopic lesions whose phenotypes can be compared. In this circumstance, or when only very few similar patients are known, it may prove useful to screen large numbers of additional intellectual disability patients for similar CNVs by an inexpensive PCR-based method that assesses only the locus of interest. It may also be useful to test patients with similar phenotypes for sequence mutations of candidate genes. The first approach proved successful with the 17q21.31 deletion syndrome, in which, after recognizing the initial patients by AGH, one group found two additional cases in a cohort of 840 mentally retarded individuals who were screened by MLPA [4]. The second approach proved successful in establishing loss of function mutations of *JAG1* as causal for Alagille syndrome [168] and loss-of-function mutations of *CHD7* as causal for the CHARGE syndrome [42] after submicroscopic deletions had been found to account for a small fraction of cases.

The recognition that submicroscopic CNVs are an important cause of intellectual disability is facilitating the identification of many new candidate intellectual disability genes. Extreme pathogenetic heterogeneity has made identifying autosomal intellectual disability genes difficult, especially when they behave as dominants and usually result from new mutations. These are exactly the conditions in which most apparently pathogenic CNVs are being recognized in people with intellectual disability, and the small size of some of these genomic lesions may implicate only a few loci as candidate intellectual disability genes. For example, *MAPT* was identified as a strong candidate for the gene that causes intellectual disability in the 17q21.31 deletion syndrome by simple scrutiny of the 500-600 kb region of overlap among several affected patients [4, 165, 166].

While genes that are completely or partially duplicated or deleted by a CNV are always considered first as probable causal factors for the disease phenotype, CNVs may alter genetic function in other ways. These include position effects, epigenetic effects and epistatic effects (Table 1.2).

Table 1.2: Genomic effects that can influence the expression of genes that lies outside of a CNV.

EFFECT	DESCRIPTION	EXAMPLES	DIAGRAM
Position effect	A deletion or duplication that involves the regulatory sequence of a gene can deregulate the gene's expression even though the coding sequence of the gene is not involved in the CNV event.	<p>The function of <i>ALX4</i> is reduced by a 1.3 Mb deletion >15 Kb downstream from the gene in Potocki-Shaffer syndrome [169].</p> <p>Cleidocranial dysplasia can be caused by a translocation 800 kb upstream of the <i>CBFA1</i> gene [170]</p>	<p>Normal transcription</p> <p>Transcription down regulated</p> <p>Transcription up regulated</p> <p>Normal gene regulation</p> <p>Vs. Deletion of gene regulator causes aberrant gene regulation</p> <p>Key: Coding exons. Gene regulatory element.</p>
Epigenetic effect	A deletion or duplication that alters the function of an imprinting control region interferes with gene expression.	A deletion that removed the paternal imprint control locus (PWS-SRO) but maintained the maternal imprint control locus (AS-SRO) at the 15q11 imprinting center (IC) mimicked maternal UPD and caused a Prader-Willi phenotype [171].	<p>AS-SRO PWS-SRO</p> <p>IC</p> <p>Normal imprint maintained by Imprinting center with intact AS-SRO and PWS-SRO.</p> <p>Vs. Microdeletion removing PWS-SRO mimics maternal UPD due to loss of paternal imprinting control. Removal of the AS-SRO would cause the opposite and mimic paternal UPD.</p>

EFFECT	DESCRIPTION	EXAMPLES	DIAGRAM
Epistatic effect	A CNV of a gene involved in the regulation of a second gene that lies elsewhere in the genome can alter the expression of the second gene.	A deletion of <i>MECP2</i> , an X-chromosome locus that encodes a DNA binding protein involved in histone modification and chromatin remodeling, affects the function of many other genes throughout the genome [172, 173]	<p>Normal epistasis</p> <p>Vs.</p> <p>MECP2 deleted</p> <p>Deletion of epistatic gene causes deregulated function of other genes, leading to disease.</p>

1.8 Genotype-Phenotype correlation studies undertaken in a cohort of patients with *de novo* CNVs identified by AGH

We used high-resolution whole genome AGH on two large patient cohorts in order to detect causative genetic lesions for idiopathic ID. The first cohort comprised 100 children with idiopathic ID and both normal parents of each child (100 trios). The AGH platform used was the Affymetrix 100K GeneChip® assay[132]. This platform provides whole genome coverage using more than 100,000 approximately evenly spaced oligonucleotide probes and provided an effective resolution of ~100 Kb. The second cohort comprised another 100 children and both normal parents of each who were tested on a higher resolution array: the Affymetrix 500K GeneChip® assay [7]. In this case the platform comprised more than 500,000 oligonucleotide probes evenly spaced across the genome and provided an effective resolution of ~25 Kb. We also re-tested an additional 54 trios from the initial series on this higher resolution microarray. These projects were completed in 2006 and 2009, respectively.

In both cases the child, mother and father were individually compared to a reference genome bioinformatically to detect CNVs in each, and then *de novo* CNVs were identified as those present only in the child and not in either parent. My contribution to both projects was the characterization of genotype-phenotype correlations for detected and validated *de novo* CNVs in the patients. I will discuss selected patients from these studies in this thesis. Detailed study design for each project is available in the respective publications [7, 132].

1.8.1 AGH study results

In the first series studied using 100K Affymetrix AGH (hereafter referred to as “the 100K study”), we detected and independently validated 11 *de novo* genomic aberrations in 11 patients (Appendix A). Eight of the CNVs were deletions and ranged in size from 178 kb (family 5566) to >11 Mb (family 3476). Two were duplications of 1.1 Mb (family 6168) and 2.9 Mb (family 4794). One was a mosaic trisomy for chromosome 9 (family 5994).

In the second series using the 500K Affymetrix AGH (hereafter referred to as “the 500K study”) all 10 *de novo* CNVs found by the 100K study detailed above were validated, and in some cases the increased resolution led to a better determination of the CNV boundary or breakpoint. In addition, the higher resolution assay was able to detect a *de novo* CNV missed

by the previous 100K AGH (family 4840). Nineteen of 100 children with ID in the new cohort were found by the 500K study to have *de novo* genomic imbalance that was independently confirmed (Appendix B). One child (Patient 8056) had mosaic trisomy 9, and two were found to have *de novo* unbalanced reciprocal translocations - a der(10)t(2;10)(q37;q26.13) in Patient 873 and a der(4)t(4;8)(p16.1;p23.1) in Patient 5814 - each producing both a terminal duplication and a terminal deletion. Two children had uniparental disomies (UPDs): a partial UPD for chromosome 11 (11p11.2-pter) in Patient 6904 and a UPD for the complete chromosome 16 in Patient 1658. We found 14 *de novo* submicroscopic deletions in 13 other patients and 3 *de novo* submicroscopic duplications in 3 other patients. The deletions ranged in size from 89 kb (family 216) to 11.0 Mb (family 9346); six were less than 1 Mb. The duplications ranged in size from 362 kb (family 9979) to 11.1 Mb (family 873). Figure 1.4 diagrams the results from both tests.

I considered only the *de novo* deletions, duplications and translocations for detailed assessment of genotype-phenotype correlation. In total, this included 31 *de novo* structural variants (10 *de novo* CNVs from the 100K AGH and 21 *de novo* CNVs from the 500K AGH study) in 28 patients.

1.8.2 Summary of Genotype-Phenotype correlation study methodology

I scrutinized *de novo* CNVs and the clinical details of the patients who bear them using a variety of web-based tools as well as both online and published catalogues. For each CNV studied, I analyzed the genomic content of the region involved. This included thorough assessment of all publically available data on gene function and expression in order to define causative ID genes and their possible mechanism of pathogenicity. I also searched the medical literature for previously reported cases of the same CNV regions and looked for distinctive phenotypic similarities between the affected patient and other reported patients with similar genomic alterations to identify possible syndromic associations. All *de novo* CNVs were also checked against databases of known benign CNVs to assess CNV pathogenicity as well as genomic tolerance of copy number variation. The methodology for this research is described in detail in Appendix C.

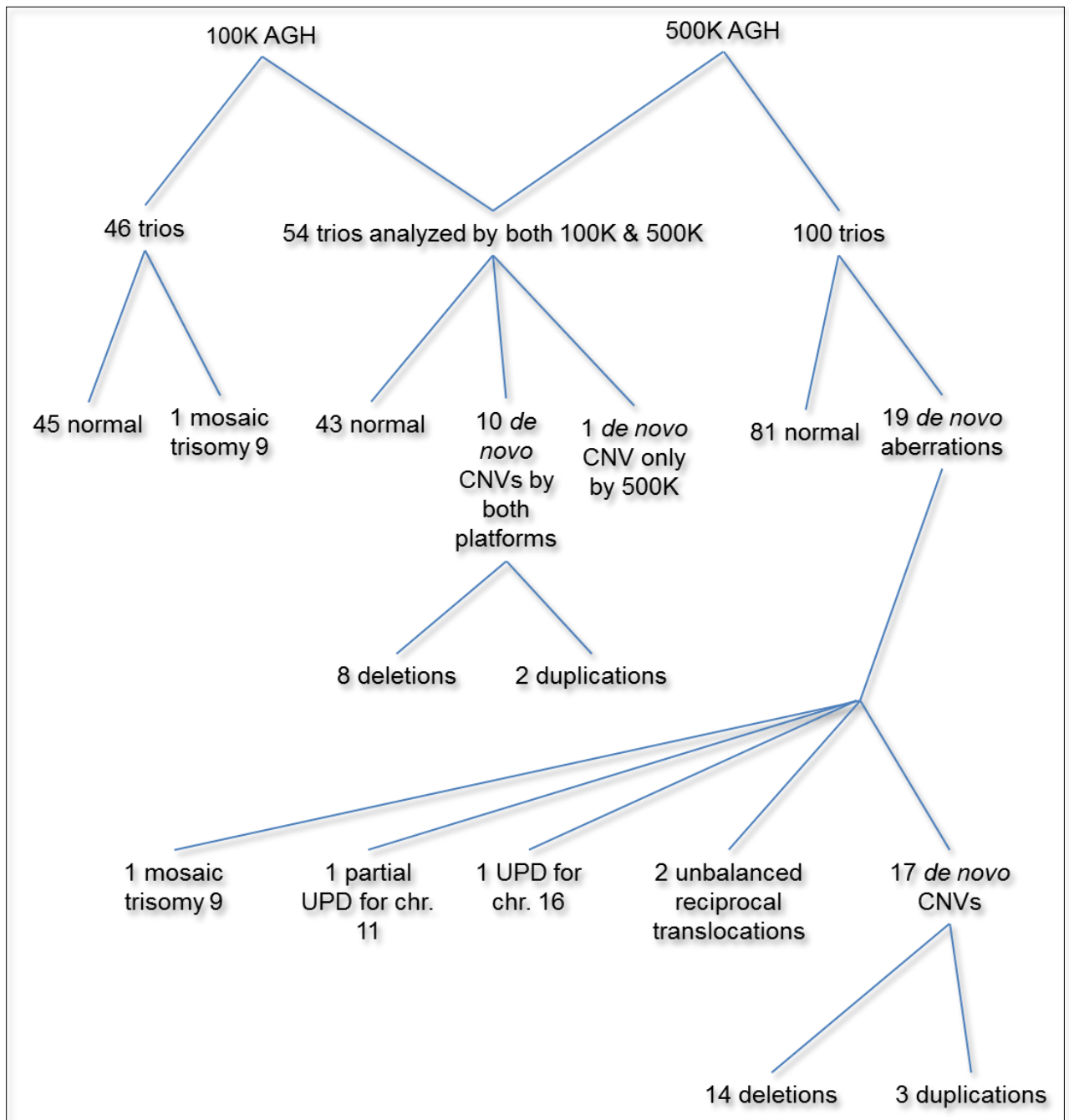


Figure 1.4: Overview of number of trios analyzed by 100K and 500K AGHs, and results obtained.

1.8.3 Summary of Genotype-Phenotype correlation findings

In several instances the detailed genotype-phenotype correlation resulted in unique findings that we published as research papers. I led three of these published studies involving four patients, and these are included in this thesis (chapters 2, 3 and 4). For another three patients, we contributed to collaborative efforts with other local and international researchers resulting in three published papers. I have included two of these collaborative studies as chapters in this thesis (chapter 5 and 6) as their findings contributed to the development of the main hypothesis of my thesis, as explained in chapter 7.

We were able to make a clinical diagnosis in six other patients, as we identified CNVs affecting regions known to be causative for described ID syndromes and showed that the patient's phenotype was characteristic of these rare conditions. Of the remaining 15 patients, nine had CNVs that do not affect critical regions for known microdeletion or microduplication syndromes, however as the CNV in each case was larger than 5 Mb and involved many genes, a meaningful correlation with the phenotype at the level of individual genes was precluded. For two other patients, the *de novo* CNVs involved only pseudogenes and redundant members of very large gene families such as the immunoglobulin gene family, and we therefore deemed them likely not pathogenic [7].

For those patients for whom a detailed genotype-phenotype correlation study has not been published, I contributed comments on the clinical outcome for the *de novo* CNVs in the main publications of the AGH study [7]. These comments are incorporated in the table in Appendix A and B.

1.8.4 Genotype-Phenotype correlation studies published as first author

A) The 14q11.2 Microdeletion Syndrome: **Novel deletions of 14q11.2 associated with intellectual disability and similar minor anomalies in three children [174]**

In this work we show a phenotypic and genotypic overlap between three children: two from our center and one from the UK. All three children had overlapping *de novo* deletions of chromosome 14q11.2 and a characteristic phenotype including mild ID and a recognizable facies with mild dysmorphism. By comparing the deleted region identified in each child, we were able to recognize two candidate genes for the phenotype, *SUPT16H* and *CHD8*. Both

genes encode factors that have epigenetic regulatory functions. This published manuscript comprises chapter 2 of this thesis. In addition, the results informed a larger study of genes encoding epigenetic regulatory factors, which is presented in chapter 7.

B) The *NRXN1* α Microdeletion Case: A patient with vertebral, cognitive and behavioural abnormalities and a *de novo* deletion of *NRXN1* α [175]

We identified a *de novo* CNV involving only the α transcript of the *NRXN1* gene in a child with ID and features of ASD. *NRXN1* encodes Neurexin, a protein that has an important function in synaptogenesis. Our study was one of the earliest to implicate this gene in pathology for ID and the first showing perturbation of only the α -isoform of this protein. Since publication of this paper, others have identified mutations of this gene to be causative for autism and schizophrenia as well as ID. The published manuscript of our study is given in chapter 3 of this thesis.

C) The microduplication 17q21.33 Case: A novel *de novo* 1.1 Mb duplication of 17q21.33 associated with cognitive impairment and other anomalies [176]

In this paper we presented a child who had a novel *de novo* duplication of chromosome 17 in band q21.33. A detailed assessment of the genes involved in the lesion resulted in identification of two candidate genes, *PPP1R9B*, a good candidate for the ID and microcephaly, and *COL1A1*, a good candidate for the hearing loss and skeletal deformities seen in our patient. This published manuscript is presented in chapter 4 of this thesis.

1.8.5 Genotype-Phenotype correlation studies published as contributing author included in this thesis

A) The Duplication 16p13.3 Case: Duplications of the critical Rubinstein-Taybi deletion region on chromosome 16p13.3 cause a novel recognizable syndrome [59]

We collaborated with other researchers to compare genotype-phenotype correlations for patients who were found to have microduplications of 16p13.3, which includes the Rubinstein-Taybi Syndrome (RBS) critical region. RBS is caused by deletions of *CREBBP*, which encodes an epigenetic regulatory protein. This work identified occurrence of a distinctive severe ID syndrome and suggested *CREBBP* overexpression as the causative factor. This study is

included as chapter 5 of this thesis. The results also informed the study of genes encoding epigenetic regulatory factors that is presented in chapter 7.

B) The duplication 8q12 Case: **A characteristic syndrome associated with microduplication of 8q12, inclusive of *CHD7* [177]**

In this report we described genotype-phenotype correlations for a patient who was found to have a microduplication that included *CHD7*. Loss of function mutations of this gene, which encodes an epigenetic regulator, are causative for CHARGE syndrome. In this work we show genotypic and phenotypic similarity between our patient and previous reports suggesting that a dosage increase of *CHD7* may be pathogenic. This paper comprises chapter 6 of the thesis, and the results informed the study of genes encoding epigenetic regulatory factors that is presented in chapter 7.

1.9 Custom aCGH to probe all genes encoding epigenetic regulators

In our AGH studies we found a high incidence of involvement of genes that encode epigenetic regulators to be involved in *de novo* pathogenic CNVs. We found genes encoding epigenetic regulators to be good candidates for the phenotype in four patients with three newly-described ID syndromes. We have also found genes encoding epigenetic regulators to be involved in other *de novo* CNVs in our patients. These genes include: *MYST* (duplicated in patient 6168), *DYRK2* (deleted in patient 4818), *SAP18* (deleted in patient 1895), *HDAC4* (duplicated in patient 873), *HDAC2* (deleted in patient 3160), and *HDAC9* (deleted in patient 2200). In addition, mutations of genes encoding other epigenetic regulators are known to cause several established ID syndromes (*vide* section 1.4).

These data led me to ask the question: how significant is dosage imbalance of genes encoding epigenetic regulators as a class in the pathogenicity of ID? In order to investigate this, I formulated the following specific hypothesis: ***Genetic imbalance involving genes regulating epigenetic processes causes 5% of ID.*** I then designed a targeted custom aCGH study to test this hypothesis.

1.9.1 Experimental plan summary

I designed a custom oligonucleotide microarray that contains probes covering all genes with epigenetic regulatory function known at time of design (2007). The array was used to screen for CNVs of these genes in 177 previously-unstudied patients with idiopathic ID and both normal parents of each child. Results of this targeted screen were analyzed for copy number variation of the candidate genes. All *de novo* CNVs identified were validated by qPCR. Results were analyzed with respect to the phenotypes of the patients, and genotype-phenotype correlations were established.

The aCGH for this study was performed by Dr. Tracy Tucker. I was in charge of selecting the genes encoding epigenetic regulators as well as the qPCR validations for all *de novo* CNVs detected. I also assisted with the bioinformatic analyses of the aCGH output and identification of *de novo* CNVs. This project and its findings are detailed in chapter 7 of this thesis.

1.10 Doctoral research objectives

The first objective of my thesis was to establish genotype-phenotype correlations for *de novo* potentially pathogenic CNVs detected in children with idiopathic ID using whole genome array genome hybridization [7, 132].

The second objective of my research was to investigate the contribution of imbalance of genes that control epigenetic processes in the pathogenesis of ID, thereby assessing the importance of correct epigenetic control toward normal neurodevelopment and function.

1.11 Thesis outline

This thesis is composed of both published manuscripts and previously-unpublished chapters. Chapters 2 to 6 are texts of published manuscripts reproduced verbatim. They detail genotype-phenotype correlation studies completed for patients found to have *de novo* CNVs from the 100K and 500K whole genome aCGH studies as previously elaborated. Chapters 1, 7 and 8 have not previously been published. Chapter 7 details the major project of my doctoral research: the custom aCGH to probe pathogenicity for dosage imbalance of genes encoding epigenetic regulators outlined in section 1.9.

The introduction (chapter 1) and the conclusion (chapter 8) include major sections from two published review articles of which I was first author: 'The impact of array genomic hybridization on intellectual disability research: a review of current technologies and their clinical utility' [178] and 'Epigenetic impacts on neurodevelopment: pathophysiological mechanisms and genetic modes of action' [179]. I have incorporated relevant sections from both papers where suitable in chapter 1 and chapter 8 but modified the text and added additional material as necessary.

Chapter 2: Novel deletions of 14q11.2 associated with mental retardation and similar minor anomalies in three children¹

2.1 Introduction

The incidence of mental retardation (MR) is about 3% globally [1]. Chromosomal abnormalities are the most frequently diagnosed cause of MR [5], and recent studies show that submicroscopic deletions or duplications cause MR at least as frequently as cytogenetically detectable chromosomal abnormalities [132, 134, 180].

We performed array genome hybridization (AGH) on 100 children with idiopathic MR and both their clinically normal parents using Affymetrix Genechip[®] Mapping 100K Assays [132]. We detected and independently validated apparently pathogenic *de novo* submicroscopic copy number variants (CNVs) in 10 children. These cases were recorded in DECIPHER (Database of Chromosomal Imbalance and Phenotype in Human using Ensembl Resources, <http://www.decipher.sanger.ac.uk>), a database of clinical and genomic findings on individuals with submicroscopic chromosomal imbalance. Here we discuss genotype-phenotype correlations for three patients with mild MR and minor physical abnormalities who have similar submicroscopic deletions of chromosome 14q11.2, two from our center and one case from the United Kingdom that we located on DECIPHER. They share similar features of minor dysmorphism which also resemble those reported for some children with MR and cytogenetically-visible deletions of proximal chromosome 14q. The deletions reported in these three patients overlap at a common genomic region of approximately 35 kb which we believe is critical as this genomic segment contains portions of two genes that are strong candidates for involvement in the phenotypes.

These studies were approved by appropriate ethics boards in Canada and the United Kingdom.

¹ **Zahir F**, Firth H.V., Baross A. et al. J Med Genet. 2007 Sep;44(9):556-61

2.2 Results

2.2.1 Vancouver case 5566

Case 5566, a female, was the second child born after an uncomplicated 38-week gestation to healthy non-related parents of English-Canadian descent; their first child is a healthy boy. The family history is unremarkable except for bilateral 2-3 cutaneous syndactyly of the toes in the maternal grandfather and maternal great aunts.

The infant's birth weight was 3900 g (91st percentile), birth length, 54 cm (98th percentile), and birth head circumference, 38cm (99.6th percentile). She had prominent strawberry nevi over her left eye-lid, philtrum and occiput at birth. The occipital nevus has persisted while the others subsequently faded. She had hypotonia as an infant that resolved as she grew older.

On evaluation at age 28 months, her height was 89 cm (75th percentile), weight 12 Kg (25th percentile), and head circumference 51.5 cm (97th percentile). She had mildly dysmorphic facies (Figure 2.1a) with short palpebral fissures (5th percentile). Her eyebrows were rounded with a triangular medial aspect and distal tapering. She had normally positioned ears with unusual auricles in which the helical root swept into a horizontal bar of cartilage dividing the concha into a lower portion and an upper portion limited superiorly by a horizontal inferior crux of the antihelix (Figure 2.1a). The posterior portion of the helix and antihelix were fused bilaterally. She had a short nose with small nasal tip and flat nasal bridge, long philtrum and small mouth with prominent Cupid's bow of the upper lip. She also had micrognathia. Her heels were prominent and her feet, flat with 2-3 cutaneous syndactyly of the toes. Formal developmental assessment at age 4 years, 6 months found her gross motor skills at the 2-3 year range and fine motor skills at the 3-4 year range.

Cytogenetic analysis at >550 band resolution showed a normal female karyotype. AGH using the Affymetrix GeneChip[®] Mapping 100K Assay on Case 5566 and both her parents showed a *de novo* deletion of approximately 200 kb of chromosome 14q11.2 [132]. The presence and *de novo* status of the CNV were confirmed by FISH [132]. Due to the small size of the lesion, we reanalyzed the child and both parents by AGH using the Affymetrix GeneChip[®] Mapping 500K Assay and found the deletion to be only 101 kb, with breakpoints at 20,896,740 bp and 20,998,178 bp based on the locations of the most proximal and distal deleted SNPs

respectively (UCSC build hg18, March 2006) (Figure 2.2). The chromosome with the deletion is likely maternal, based on presence of one informative paternal SNP in the interval.

2.2.2 Vancouver case 8326

Case 8326, a male, was the first child of healthy unrelated parents of Japanese descent. He was born at 37 weeks gestation after an uncomplicated pregnancy and normal vaginal delivery. The family history is noncontributory.

The infant's birth weight was 2730g (~ 50th percentile), length, 51 cm (75th percentile) and head circumference, 32 cm (10th percentile). The neonatal course was complicated by respiratory distress, and a ventricular septal defect and large patent ductus arteriosus were diagnosed by echocardiogram. The ductus was ligated at 10 days of age, and the septal defect closed spontaneously by age 6 months. There were no subsequent cardiac problems.

The patient has a history of hypotonia, acquiring good head control only at age 6 months. Walking was achieved at age 2 years. He also showed delays in fine motor and language skills. A Gesell Developmental test at 26 months of age estimated his gross and fine motor skills at the 13 month level, his adaptive skills between the 11 and 13 month level and his language skills at the 15 month level. Ophthalmologic and auditory evaluations were both normal. On physical examination at age 3 years 8 months, his height was 97.7 cm (25th percentile), weight, 13.8 Kg (10th percentile), and head circumference, 49 cm (25th percentile). He had mildly dysmorphic facies (Figure 2.1b) with a broad forehead, epicanthic folds, very short palpebral fissures (2 standard deviations below the mean), and rounded eyebrows with a triangular medial aspect and distal tapering. His nose was small with a very flat broad nasal bridge and small nasal tip. His ears were normally placed with malformed auricles that were remarkably similar to those in Case 5566 (Figure 2.1b). He also had a small mouth, a very narrow high-arched palate, a long philtrum, and a prominent Cupid's bow. He had a bridged palmar crease on the right hand and a single palmar crease on the left hand. He had significant pronation of his left foot, mild 2-3 syndactyly of his toes and clinodactyly of his fourth toe bilaterally.

Cytogenetic analysis at >550 band resolution showed a normal male karyotype. AGH on the child and both his parents using Affymetrix GeneChip® Mapping 100K Assays showed a *de*

*nov*o ~1.6 Mb deletion of chromosome 14q11.2, with proximal and distal breakpoints at 19,584,863 bp and 21,207,935 bp (UCSC build hg18, March 2006), respectively [132]. The presence and *de novo* status of this deletion were confirmed by FISH [132]. SNP genotype data determined the deletion-bearing chromosome to be paternal.

2.2.3 DECIPHER case 126

Case DECIPHER #126, a female, was the first child of healthy unrelated parents of English descent with no relevant family history. She was born at 40 weeks gestation by emergency caesarian section for fetal distress, weighing 3657g (50th-75th percentile). She did not require resuscitation but was admitted to the neonatal unit with respiratory problems and required nasogastric feeding for the first few days of life. At 6-8 weeks of age she still had marked head lag and was generally hypotonic. Her first year of life was characterized by lack of social interaction. She had a brief febrile convulsion at 10 months and sat at 11 months of age.

On review at age 14 months she did not give eye contact, and she demonstrated hand regard and bruxism. She walked at 26 months. At 29 months, her developmental level was assessed at the 12-17 month range, with speech and language skills and spontaneous play both at approximately the 10 month level. A moderate left alternate convergent squint and hypermetropia were noted. She had rather deep-set eyes, a prominent philtrum and prominent antihelix of both ears (Figure 2.1c).

On review at 4 years 10 months of age, she had some single words, showed an increased desire to communicate and some imitative play. On physical examination, her height was 102 cm (9th percentile), weight 15.9 kg (9th-25th percentile) and head circumference 50.7cm (9th-25th percentile).

Cytogenetic analysis at >550 band resolution and cranial MRI scan were normal. An MLPA-based telomere assay showed deletion of one of the 'control' regions on chromosome 14q11. FISH analysis demonstrated a *de novo* deletion of the region defined by clones RP11-203M5 and RP11-524O1 on chromosome 14q11.2, identifying a ~1.079 Mb minimal deletion from 19,853,310 to 20,932,827 bp (UCSC build hg18, March 2006). The deletion is flanked by the clones RP11-597A11 (19.2 Mb) and RP11-124D2 (22.8 Mb), both of which are present in two

copies. Studies of her parents confirmed that the deletion was *de novo*. The parental origin of the deletion-bearing chromosome was not determined.

2.3 Discussion

The three patients described present similar degrees of MR and a similar mild dysmorphic appearance, despite having different ethnic origins. All three have widely-spaced eyes, a broad flat nasal bridge and short nose, a long philtrum, prominent Cupid's bow of the upper lip and full lower lip, similar eyebrows and an usual abnormality of helical root formation of the ear. Hypotonia in infancy was also reported in all three children.

2.3.1 Previously-reported cases with chromosome 14q11.2 deletion

We found published reports on ten patients with deletions of cytogenetic band 14q11.2 [181-188]. FISH mapping was performed for five of these cases [182], but none was found to have a deletion involving the region deleted in our three cases. No molecular characterization is available for four other published cases, precluding molecular genotype-phenotype correlations. However, some phenotypic features found in our patients were also noted among these cases, including psychomotor delay [188], hypotonia [186, 188], alternating esotropia [188], micrognathia [186] and other facial dysmorphisms [184, 186]. Two unpublished cases with cytogenetic deletions of 14q11.2 were found in the online Chromosome Abnormality Database (<http://www.ukcad.org.uk>, accessed Oct 2006). The unavailability of molecular breakpoint mapping for these cases excluded them from molecular genotype-phenotype correlations.

The only previously-reported case with a deletion shown molecularly to overlap the region involved in our cases is that described by Zanolli *et al.* [181]. In this patient, a der(5)t(5;14)(5qter→5pter::14q11.2→14qter) chromosome replaced both a normal chromosome 5 and a chromosome 14, deleting the entire short arm and the centromere of chromosome 14 (chromosome 5 bore a very small deletion distal to the subtelomeric region, involving a region composed of repetitive sequence). This was a *de novo* rearrangement involving the paternal chromosomes. The distal breakpoint on chromosome 14 was mapped by FISH between D14S72 (20.4 Mb, deleted) and D14S64 (23.6 Mb, retained) [181]. Therefore, the deletion in this case overlaps those in our Case 8326 and DECIPHER Case 126 up to 20.4 Mb distally, and may also overlap our minimal critical deleted region. The patient described by

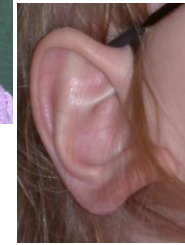
Figure 2.1: The first row shows photographs of the faces and ears of our three cases; Vancouver case 5566 (at 2 yrs. Age), Vancouver case 8326 (at 2 yrs. Age) and DECIPHER case 126 (at 3 yrs. Age). Consent was obtained to publish these photographs. The second row shows the face of previously published case A of Zanolli *et al.* [181], which has been reprinted with permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc.



a) Case 5566: Face and ear
Deletion: 20,896,740 to 20,998,178
Size: ~ 101 kb



b) Case 8326: Face and ear
Deletion: 19,584,863 to 21,207,935
Size: ~ 1.62 Mb



c) Case 126: Face and ear
Deletion: 19,853,310 -20,931,827
Size: ~1.08 Mb



d) Previously published case
'case A' from Zannolli *et al.*

Zanolli et al. was a 2 year-old girl whose facial appearance is quite similar to that of our patients (Figure 2.1d). She had delayed psychomotor development and hypotonia, which are common to all our cases. She also had an alternating squint, a palmar crease on both hands and skin pigmentation abnormalities, each of which was found in at least one of our cases.

2.3.2 Candidate genes

The minimal region of overlap for our three cases lies between the proximal breakpoint of Case 5566 and the distal breakpoint of Case 126, a region of 35 kb between 20,896,740 bp and 20,931,827 bp on chromosome 14. Portions of only two RefSeq [189] genes are included in this region: *SUPT16H* (20,889,478-20,922,265, minus strand) and *CHD8* (20,923,470-20,975,242, minus strand) (Figure 2.3). Both *CHD8* and *SUPT16H* show high expression levels in normal human adult and fetal brain (<http://symatlas.gnf.org/SymAtlas,Version 1.2.4>).

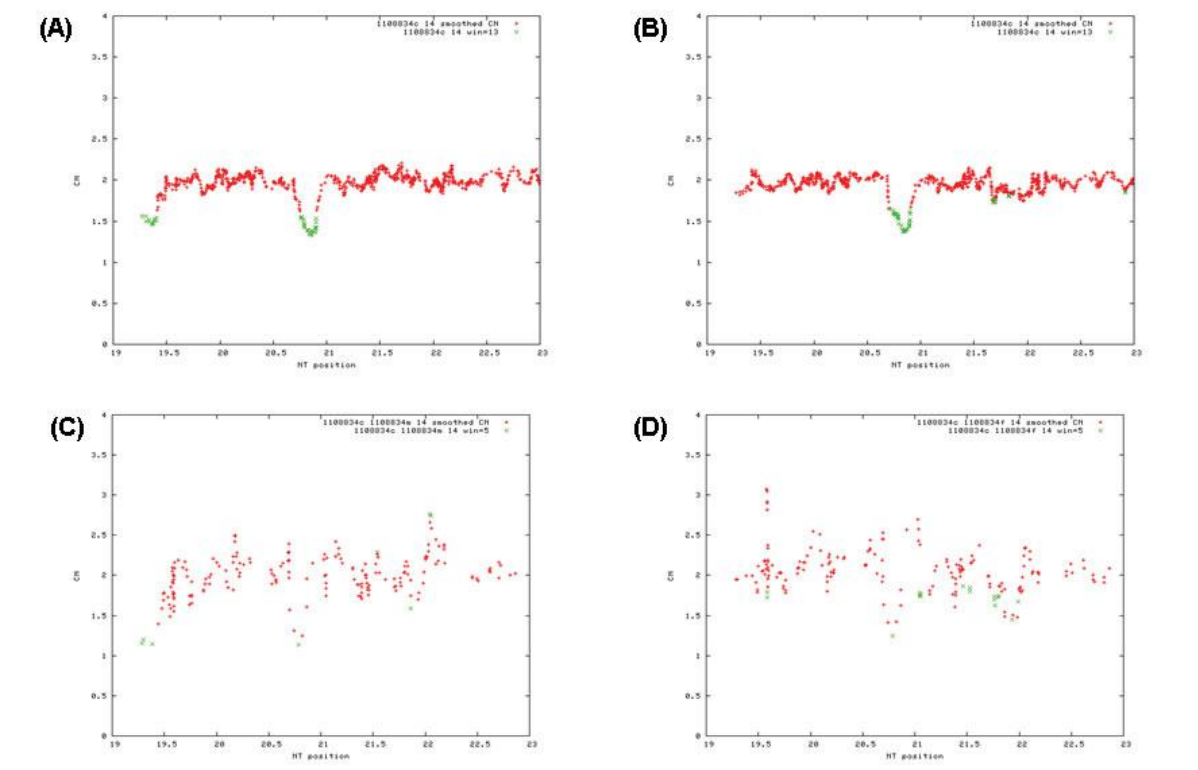


Figure 2.2 : SNP genotyping data on case 5566 and both her parents by AGH using Affymetrix 500K Genechip® (top panel, (A), - child vs. father comparison, and (B)- child vs. mother comparison) were able to refine the data generated by AGH using Affymetrix 100K Genechip® assays (bottom panel, (C) – child vs. father comparison and (D) – child vs. mother comparison). All graphs show enlarged views of chromosome 14 in band q11.2 (19Mb to 23Mb), generated using proprietary and locally developed software.

SUPT16H (Entrez GeneID: 11198) encodes the larger subunit of the conserved FACT (Facilitates Chromatin Transcription) complex [190]. The gene lies almost entirely within our minimal critical region, and the deleted segment includes all known transcription start sites. (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=35g&c=Gene&l=SUPT16H>). Therefore, the deletion probably produces a null mutation of *SUPT16H*.

The FACT complex is an essential facilitator of transcription in yeast [191]. FACT has also been directly implicated in maintenance of chromatin structure in yeast [192, 193] and is known to associate with CHD1, a chromatin remodeling factor in yeast and higher organisms [194]. A role in histone methylation has also been suggested because FACT interacts with the PAF complex [195, 196], which is known to regulate transcription-related histone modification in yeast [197]. In addition, FACT may be involved in DNA replication and DNA repair [194].

The possibility that haploinsufficiency of *SUPT16H* may be the cause of the congenital anomalies and MR in our patients is supported by the fact that mutations in other genes that encode transcription factors or chromatin remodeling complexes are recognized causes of MR [18].

CHD8 (Entrez GeneID: 57680) encodes at least 7 different transcripts, of which only 3 form complete protein products. Approximately 10kb of the gene's 3' end, which includes the 3' ends of all of the complete transcripts is found in our minimal critical deletion region. Therefore the deletions in all of our patients would be expected to result in altered and possibly non-functional protein products (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&l=CHD8>).

The CHD8 protein directly interacts with the chromatin insulator-binding protein CTCF and is involved in epigenetic regulation of transcription [198]. Ishihara *et al.* showed that RNAi-mediated knockdown of *CHD8* in HeLa cells alters CpG methylation and histone acetylation around CTCF binding sites [198]. These authors concluded that CTCF-CHD8 has a role in insulation and epigenetic regulation at active insulator sites [198], suggesting the importance of adequate *CHD8* function in development. The potential pathogenicity of *CHD8* haploinsufficiency is also suggested by analogy to the known pathogenicity of *CHD7* loss-of-function mutations, which are causative of CHARGE syndrome [199]. *CHD7* is a member of the conserved chromodomain helicase DNA binding protein family to which CHD8 belongs.

To examine possible position effects and regional genomic effects, we expanded our candidate gene search to include regions ~500 kb proximal and distal to our minimal critical region. This region contains thirteen additional RefSeq genes, of which six belong to the highly redundant olfactory receptor and RNase gene families, undermining the possibility that a heterozygous deletion of any of these genes would contribute to pathogenicity. The remaining seven genes are listed in Table 2.1.

Table 2.1: Other genes within 500 Kb of our minimal critical deletion region: 20.4 Mb to 21.5 Mb on chromosome 14. The best candidates of these based on function, *NDRG2* and *SALL2*, are highlighted.

	Gene Name	Genomic Co-ordinates (UCSC genome browser, build hg18, March 2006)	Function
1.	<i>NDRG2</i>	Chr14:20554763-20563775	Member of the N-myc downregulated gene family. Role in neurite outgrowth[200].
2.	<i>SALL2</i>	Chr14:21059074-21075177	Transcription regulator [201]
3.	<i>METT11D1</i>	Chr14:20527805-20535032	Methyl transferase activity-related
4.	<i>KIAA0737</i>	Chr14:21015175-21037157	Epidermal Langerhans cell protein
5.	<i>METTL3</i>	Chr14:21036129-21049291	Methyl transferase activity-related
6.	<i>FLJ10357</i>	Chr14:20608367-20627875	Hypothetical protein
7.	<i>HNRPC</i>	Chr14:20748939-20807424	Heterogeneous nuclear ribonucleoprotein C

2.3.3 Low copy repeat (LCR) sequences in the region

We scanned the genomic sequence of chromosome 14 from 19.0 to 21.50 Mb (covering sequence ~500 kb upstream and downstream of our largest deletion) for known LCR sequences (duplications of >1 kb of non-repeat masked sequence with over 90% similarity). There were no LCR pairs flanking any of our deletion breakpoints. However, there were a total of nine LCR pairs located within this region (Figure 2.3). The duplication-enrichment-index (DEI), defined as the ratio of the observed percentage of duplications in a region to the percentage in the whole genome, is a measure of how enriched a specific region of the genome is for repeat elements [202]. We calculated a DEI of 4.0 for the region between 19.0 and 21.5 Mb on chromosome 14 (supplementary table 1 available online at <http://jmg.bmj.com/content/early/2007/07/14/jmg.2007.050823/suppl/DC1>), indicating moderate LCR enrichment for this region.

The lack of recurrent breakpoints and the absence of LCR pairs that flank any of the *de novo* deletions in our cases argue against non-allelic homologous recombination as the mutation

mechanism. However, the relative abundance of LCR pairs within the genomic region is consistent with a suggested stimulatory effect on deletion formation by other mechanisms [203], such as non-homologous end joining [203] and replication fork shifting [204]. Sequence information for the deletion breakpoints and surrounding genomic segments of our cases is required to assess these possibilities.

2.3.4 Common deletion polymorphisms in the region

We searched the Toronto Database of Genomic Variants (<http://projects.tcag.ca/variation/>) for reported benign CNVs within the region encompassed by the three *de novo* CNVs reported here. In total, 2 loss, 2 loss/gain and 4 gain polymorphisms mapped to this region (accessed March 17th 2007, supplementary table 2 available online at <http://img.bmj.com/content/early/2007/07/14/img.2007.050823/suppl/DC1> and Figure 2.3). Of these, only gain variant #0174 partially overlaps *CHD8*, while none of the other variants overlap either *SUPT16H* or *CHD8*. We also searched the Human Structural Variation Database (<http://humanparalogy.gs.washington.edu/structuralvariation/>) and found a somatic variant mapping to our deletion region (supplementary table 2 available online at <http://img.bmj.com/content/early/2007/07/14/img.2007.050823/suppl/DC1>), providing further evidence of the region's genomic instability (Figure 2.3).

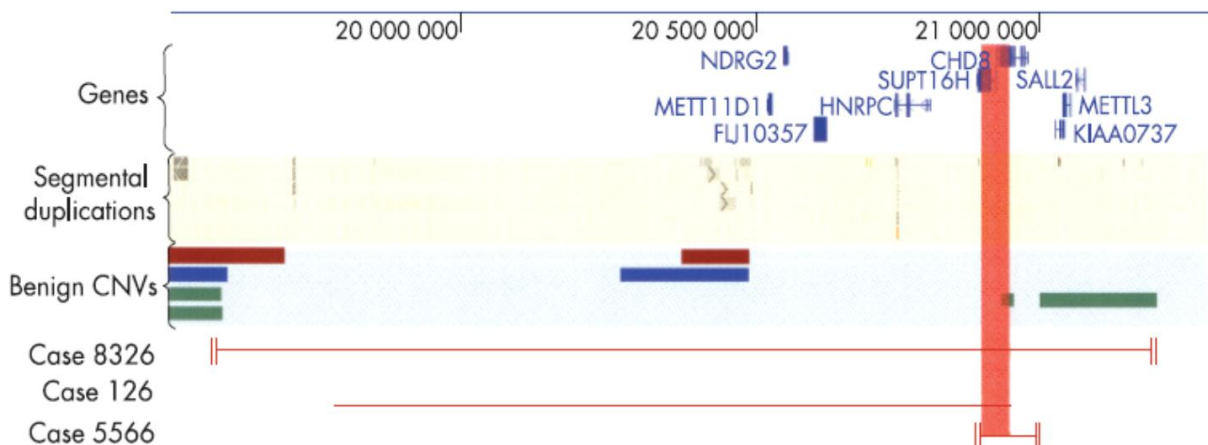


Figure 2.3: The genomic region of chromosome 14 that contains the deleted region of our three cases (red horizontal lines; those ending in short vertical bars represent most probable deletion regions detected by SNP genotyping), genes in the critical region as well as candidate genes within 500 Kb (only those in Table 2.1) of our minimal critical deletion region, benign copy number variants, and segmental duplications according to UCSC genome browse (build hg 18). On the benign copy number variant track, the brown bars represent common loss variants, blue, common loss/gain variants, and green, common gain variants. The minimal critical deletion region is highlighted by the vertical pink bar and the deletes almost all of *SUPT16H* and disrupts the coding transcripts of *CHD8*. The genomic coordinates are given by the ruler at the top of the figure.

2.4 Conclusion

We found a characteristic pattern of MR and minor congenital anomalies associated with submicroscopic deletions of a minimal critical region of only 35 kb on chromosome 14 on band q11.2. The only two RefSeq genes included in this minimal deletion, *SUPT16H*, and *CHD8*, are both good candidate MR genes. Further studies are needed to determine whether this pattern of similar minor dysmorphism constitutes a novel MR syndrome. By using AGH and a subsequent genotype to phenotype correlation approach we have been able to identify a similar pattern of dysmorphism associated with MR, and candidate genes for the phenotype.

Chapter 3: A patient with vertebral, cognitive and behavioural abnormalities and a *de novo* deletion of *NRXN1*²

3.1 Introduction

In mammals, there are three neurexin genes that encode a family of highly polymorphic proteins with high levels of expression in the brain. Each neurexin gene produces a longer α -neurexin transcript and a shorter β -neurexin transcript from different promoters [205]. The structure of neurexins and their localization at the synapse suggest a role as cell-surface receptors or cell adhesion molecules [206]. Cell adhesion is important in coordinating synaptic activity in the brain [207], and the formation, maintenance and modification of synapses are fundamental to learning, memory, and cognition [208]. The genetic locus for neurexin 1 (*NRXN1*) has recently been implicated as one of several involved in susceptibility to autism [209, 210].

We describe a child with developmental delay, unusual autistic-like behaviours, multiple vertebral anomalies and an unusual facial appearance, who was found to have a *de novo* 321 kb deletion of chromosome 2p16.3 by array genomic hybridization (AGH). This deletion involves the most 5' portion of *NRXN1*. We believe that this case supports an essential role of *NRXN1* α in normal neurodevelopment and function. These studies were approved by the University of British Columbia Research Ethics Board.

3.2 Case report

The male patient was born to healthy non-consanguineous parents. He has one sibling, an older maternal half-brother, who was diagnosed with Langerhans cell histiocytosis of a cervical lymph node several years ago. Our patient was the product of an uncomplicated term pregnancy and normal vaginal delivery. His birth weight was 2820 g (25th centile), birth length was 50 cm (50th centile), and birth head circumference was 34.5 cm (50th centile). His mother reported smoking less than half a pack of cigarettes per day during the pregnancy.

² Zahir F.R., Baross A, Delaney A.D., et al. J Med Genet. 2008 Apr;45(4):239-43

The patient's growth has been normal, but his development has been somewhat slow. He began walking at 17 months, said his first words between 2 and 3 years of age, and was only able to communicate in 3-4 word sentences by 3 years 6 months of age. He has been receiving speech therapy since the age of 2 years 6 months for delayed language development.

The patient has had numerous developmental assessments. Psychoeducational evaluation at 5 years 7 months of age found him to be in the borderline range for verbal comprehension, non-verbal information processing and perceptual organization skills. He was also noted to be delayed in fine-motor and visual-motor integration skills, and continued to demonstrate speech and language difficulties.

At age 7 years 6 months, an educational consultant reported him to have severe language impairment in addition to motor impairment. However, it was noted that the patient displayed an unusual ability to speak at length about subjects that interested him at a level above that of his classmates.

A developmental pediatrician assessed the patient at age 7 years 6 months and noted difficulty with visual, spatial and perceptual integration, difficulty with conceptual tasks like problem solving, difficulty with organizing and task sequencing, mental fatigue and emotional difficulty with anxiety. The pattern was thought to be consistent with a non-verbal learning disability.

A neuropsychiatry assessment at age 8 years found inappropriate social behaviours, poor control of bowel movements, worries, inability to accept criticism and non-compliance. In addition, the child was noted to have difficulty with temper, argumentativeness, touchiness, fears, odd ways of relating and odd pre-occupations on a parental checklist.

The patient's most recent assessment was by a developmental psychologist at age 9 years 7 months. At this time, the child was noted to have a tendency to live in his own world, to be fixated on certain topics well past the point of interest for his playmates or family, to over-react if his routine was disturbed, to display a lack of accommodation with different behaviours, to miss social cues, and to line things up and make patterns or rituals in certain activities. Two parent report measures, the Autism Spectrum Screening Questionnaire [211] and the

Australian scale for Asperger's syndrome [212], showed more than the average number of autistic features but did not meet the threshold for diagnosis of autism spectrum disorder. The Wechsler Intelligence Scale for Children showed a verbal IQ of 90 and a performance IQ of 70. The full scale IQ was not calculated because of the 20 point discrepancy between the verbal and performance scores.

The patient has had intermittent asthma that has required hospitalization in association with upper respiratory infections about once per winter despite chronic treatment with albuterol and fluticasone. Spine radiographs performed for mild dorsal scoliosis noted clinically at 4 years of age showed vertebral fusion of T2, T3, T4, L4 and L5, 13 ribs on the left with possible fusion of the second and third ribs posteriorly, and a bifid right second rib (Figure 3.1, A). Renal ultrasound examination at age 4 years and CT scan of the head at age 6 were both normal. Evaluation at age 5 years by a cardiologist revealed borderline biventricular hypertrophy with some septal dyskinesia. His overall left ventricular function was considered to be borderline but qualitatively normal. Assessment at 6 years of age by an ophthalmologist was normal. The patient underwent circumcision at 11 years of age for recurrent bladder infections. His medical history is otherwise unremarkable, and no other surgeries were reported.

On physical examination at almost 12 years of age, his height was 170.4 cm (75th to 90th centile), his weight was 72.6 kg (95th centile), and his head circumference was 57.8 cm (just above the 98th centile). He had a somewhat unusual facial appearance with prominent frontal bossing and bi-temporal recession of the hairline (Figure 3.1, B). In addition, the right antero-temporal hairline was irregular with a small tongue of hair spared in the region where the hair is normally absent. His posterior hairline was low. There was mild prominence of the nipples, which was considered to be normal for a boy of this age. There was mild dorsal scoliosis. He had mild hyperextensibility of the distal interphalangeal joints of his fingers, but he was not generally hyperextensible. There was mild 2-3 syndactyly of the toes on both feet.

Cytogenetic analysis of peripheral blood lymphocytes revealed a normal male (46,XY) karyotype at 600-650 band resolution. Array genomic hybridization using the Affymetrix GeneChip® Mapping 100K Assay on the child and both parents revealed a *de novo* heterozygous deletion of ~320 kb on chromosome 2 in band p16.3 between genomic coordinates 50,799,281 and 51,120,644 bp (NCBI build 36.1) based on location of the most

proximal and distal SNPs included in the deletion, SNP_A-1659073 and SNP_A-1696422 respectively[132]. The hemizygous deleted SNPs are flanked by SNP_A-1658967 at 50,799,203 and SNP_A-1694106 at 51,160,893 both present in two copies. The deletion was confirmed to be *de novo* by FISH on blood lymphocytes from the child and both parents using BAC probe RP11-1151G3.

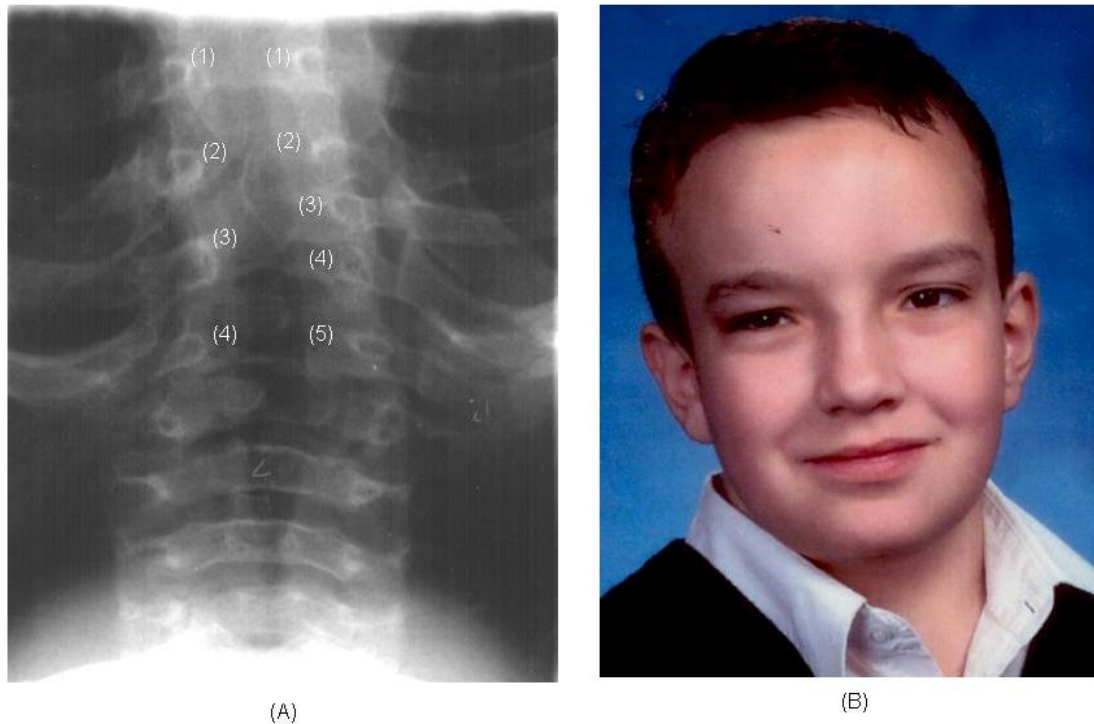


Figure 3.1: (A) chest X-ray image of patient taken at 4 years of age, showing the left extra rib, (B) Facial photograph of patient at 11 years of age.

We also sequenced all of the NRXN1 α and β exons in the patient and both his parents. No inactivating mutations were found in the remaining copy of *NRXN1* of the patient, while both copies in both parents were non-mutated.

3.3 Discussion

The only coding regions annotated in the genomic segment deleted in this patient are exons 1-5 of *NRXN1* (Figure 3.2). The deletion also includes the *NRXN1* α promoter.

We searched the published literature as well as the Chromosome Abnormality Database, CAD (<http://www.ukcad.org.uk>), European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations, ECARUCA (<http://agserver01.azn.nl:8080/ecaruca/ecaruca.jsp>), Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources, DECIPHER (<http://www.sanger.ac.uk/decipher/>), and the Developmental Genome Anatomy Project database, DGAP (<http://www.bwhpathology.org/dgap/>) for other patients who carry a deletion involving *NRXN1*. We found seven other patients with cognitive and/or behavioural abnormalities who are thought to have genomic alterations producing heterozygous *NRXN1* loss of function (Table 3.1). However, only limited clinical information is available on each of these cases, and we were not able to conduct a detailed phenotypic comparison with our patient. Notably, the patient described here is the only individual reported to date with haploinsufficiency of *NRXN1α* without involvement of *NRXN1β*.

Neurexins are a family of proteins that function as cell adhesion molecules and receptors in the vertebrate nervous system [213]. Neurexin proteins are encoded by three unlinked genes – *NRXN1* (Entrez gene ID 9378, chromosome 2), *NRXN2* (Entrez gene ID 9379, chromosome 11) and *NRXN3* (Entrez gene ID 9369 chromosome 14). *NRXN1* and *NRXN3* each span more than 1 Mb of genomic sequence, but the genomic extent of *NRXN2* is much smaller at ~110 kb [214]. Nevertheless, the exon size and number are almost identical in all three genes. Each neurexin gene encodes two major isoforms, designated α and β . The promoter for α -neurexin transcripts lies upstream of exon 1, while the promoter for β -neurexin transcripts is located in the intron downstream of exon 17 [214]. All three α -neurexin transcripts begin with a short non-coding exon that is spliced to a second exon containing the start codon. The complete *NRXN1α* and *NRXN2α* transcripts each contain 23 exons, while the *NRXN3α* transcript contains 24 exons. The first transcribed exon of the β -neurexins (alternative exon 17) contains the start codon, is spliced to exon 18, and is not included in any of the α -neurexin transcripts [214].

Extensive alternative splicing is characteristic of all three neurexins [206] and appears to be of functional importance [215-217]. There are five sites of alternative splicing in the α -transcripts, of which two are also part of the β -transcripts [206]. The independent utilization of different splice sites in various combinations suggests that hundreds of different neurexin gene products may exist [205, 206, 216, 218].

Both α - and β -neurexins are single transmembrane proteins with distinct extracellular N-terminal receptor-like sequences. In α -neurexins, the N-terminal signal peptide is followed by three repeat sequences; each repeat consists of two LNS (laminin A, neurexin and sex-hormone binding) domains that flank an EGF (epidermal growth factor)-like domain. The repeats are followed by a carbohydrate attachment sequence and a single transmembrane region, then by a conserved short intracellular cytoplasmic tail [206]. In β -neurexin, the LNS-EGF-LNS repeats are replaced by a single LNS domain [206], while the rest of the protein is very similar to α -neurexins (Figure 3.2).

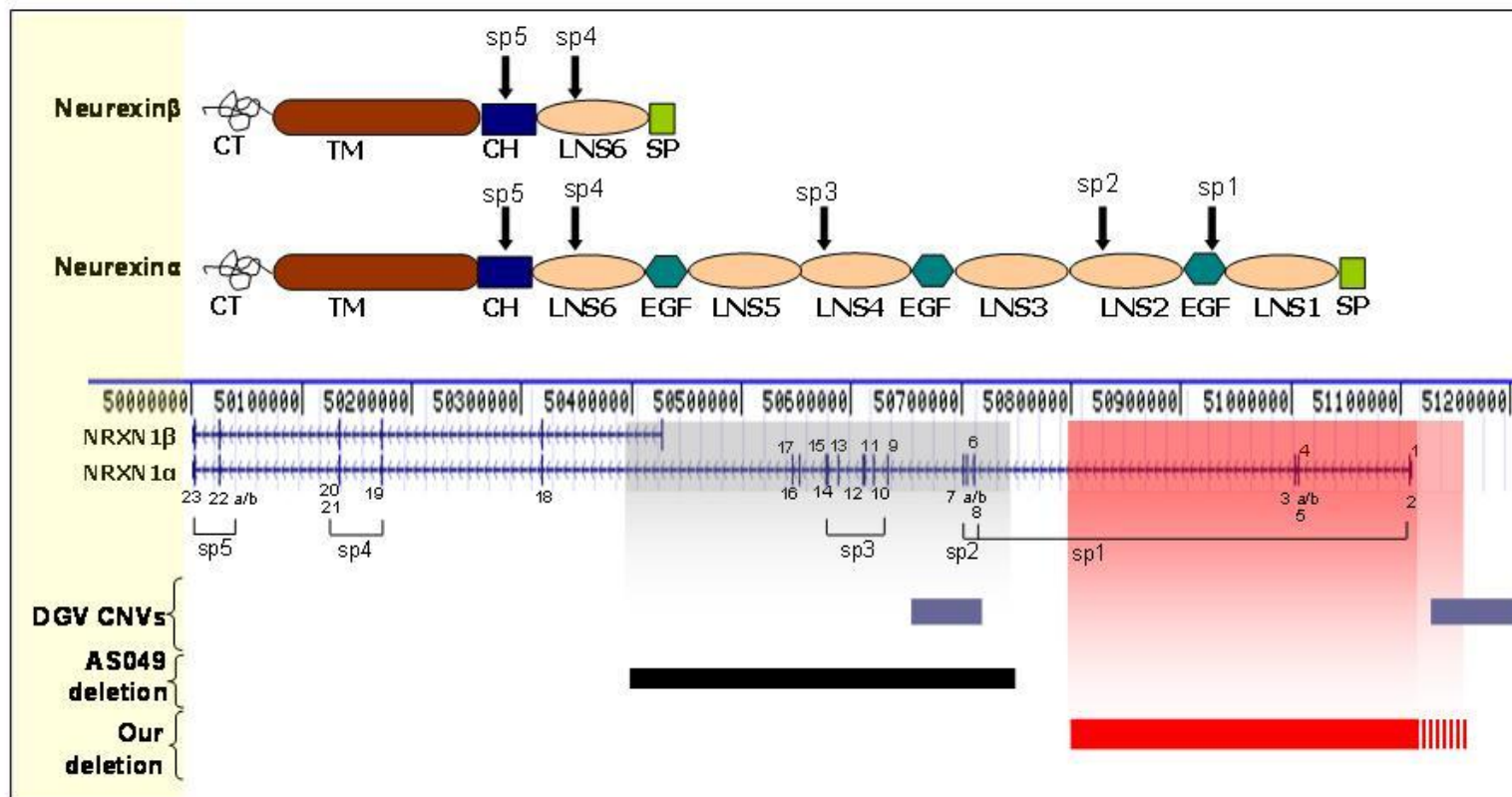


Figure 3.2: Top panel, schematic structure of Neurexinβ and Neurexinα proteins (SP- signal peptide, LNS 1-6 –laminin, neurexin, sex-hormone binding domains 1-6, EGF – Epidermal growth factor-like domains, CH – carbohydrate binding region, TM –transmembrane region, CT – cytoplasmic tail. Arrows sp1-sp5 indicate sites of alternative splicing. Bottom panel, mapped gene and exon position of NRXN1β and NRXN1α transcripts (UCSC hg 18, March 2006). Exons are numbered and splice sites indicated [214]. Benign CNVs in the region (mauve bars), the deletion reported by the Autism Genome Consortium study [210] (black bar), and our patient's deletion (red solid bar represents region of absent SNP probes, striped region represents flanking region to non-deleted adjacent SNP probe) are shown. Genomic co-ordinates are marked by the centre ruler. DGV is the Database of Genomic Variants (<http://projects.tcag.ca/variation/>).

Table 3.1: Details of other reported cases with genomic aberrations expected to cause NRXN1 loss of function.

Case	Source	Genomic feature	Method	Reported phenotype
1, Male patient with abnormal karyotype (both parents were normal)	Published report, Fryns et al. 1979[219]	47, XY,del (2)(p11p21), +ace	Karyotyping, subsequent AGH with Affymetrix GeneChip® Mapping 500K Assay was normal	Mental retardation, marked speech delay, low hair line, marked hyperkyphosis
2&3, Two female siblings, family number AS049	International Autism Consortium[210]	Both carried identical 300 Kb deletions on 2p16.3 (Figure2)	AGH with Affymetrix GeneChip® Mapping 10K Assay. <i>De novo</i> and validated by qPCR	Both girls presented with typical autism and language regression
4, BRI 92/2626	Chromosome Abnormality Database (www.ukcad.org.uk)	46,XY,del(2)(p16.2p21)	Not known	Mental retardation and behavioural abnormalities
5, DGAP124	Developmental Genome Anatomy Project (http://www.bwhpathology.org/dgap/)	46,XY,ins(16;2)(q13;p16.2p21)	Not known	Significant speech delay as a child and subsequent speech impediment. Obsessive-compulsive tendencies.
6, DGAP123		46,XX,ins(16;2)(q13;p16.2p21)pat.ish ins(16;2)(wcp2+;wcp2+)	Whole chromosome paint, chromosome 2.	Developmental delay, obsessive-compulsive disorder, Tourette syndrome, expressive apraxia and dysarthria, central auditory processing disorder, seizures, and left temporal lobe arachnoid cyst
7, DECIPHER case #613	DECIPHER database (www.sanger.ac.uk/PostGenomics/decipher/)	del(2)(p16.3p16.3) (1.4 Mb), del(2)(p11.1;q11.1) (5.8 Mb)	Not known	Joint laxity and developmental delay

Three neurexin ligands have been recognized in the mammalian brain – neuroligins, neurexophilins and dystroglycan. Neurexophilin binding is specific to α -neurexins, while dystroglycan shows preferential binding to α -neurexins but also binds β -neurexins. Neuroligins are well-known binding partners of β -neurexins but also interact with α -neurexins [220]. Binding of neuroligins to β -neurexins can induce the formation of glutamatergic and GABAergic synapses *in vitro* [221]. Both dystroglycan and neurexophilin show highly selective expression in the brain [222, 223], but the function of the interaction of neurexins with dystroglycan or neurexophilin is not known [224].

The hemizygous deletion in our patient involves the segment from 50,799,281 to 51,120,644 on chromosome 2p16.3. The *NRXN1* gene is transcribed from the minus strand: the α -precursor from 51,109,107 to 50,000,992, and the β -precursor from 50,428,398 to 50,000,992 (NCBI genome build 36.1, UCSC hg18, March 2006) (Figure 3.2). Therefore, the portion of *NRXN1* between 51,109,107 and 50,799,281, affecting the promoter and part of the coding sequence of the α -neurexin transcript is hemizygously deleted in our patient (Figure 3.2). The deleted segment also extends ~11.5 kb (from 51,109,107 to 51,120,644) upstream of the *NRXN1* α -transcription start site, which likely includes α -transcription enhancer elements. The deletion does not affect the β -neurexin promoter or coding sequence. About 371 kb of sequence upstream of the β -promoter also remains intact, and any enhancer elements that lie in this region are unlikely to be involved in the deletion.

Alignment of the complete RefSeq α -transcript (AB_035356.1) to the *NRXN1* genomic sequence of the deletion shows that the α -promoter region and first five exons, corresponding to extra-cellular signal peptide, outermost ligand binding LNS domain and part of the adjacent EGF domain, are involved in our patient's deletion. The next exon lies approximately ~700 kb downstream and is not included in the deletion. The deletion would be expected to produce complete absence of the α -neurexin transcript from the affected homolog, but β -neurexin expression is probably not affected. The remaining copy of *NRXN1* is non-mutated and expected to be fully functional in our patient.

The Toronto Database of Genomic Variants (<http://projects.tcag.ca/variation/>) does not include any copy number alterations in the region deleted in our patient and only one variant that maps

elsewhere within the large *NRXN1* gene. Variant # 2383, reported by Redon et al. [114] maps to chr2:50,651,965-50,714,553 (NCBI Build 36.1). This variant includes exons 6, 7 and 8 of *NRXN1α*. It was identified in only one subject of 276 studied, by only one of the two platforms utilized, and was not independently validated.

The 2 Mb genomic segment containing the *NRXN1* locus, as well as 500 kb upstream and 500 kb downstream of the gene, does not include any segmental duplications annotated on the UCSC genome browser (UCSC hg18, March 2006). This observation and the paucity of other pathogenic or benign copy number variants in the region suggest that the region is unlikely to be genomically unstable.

Ulrich *et al.* [216] showed that expression patterns of the α - and β -isoforms of all three neurexin genes differed in different regions of the rat brain, although most neurons expressed multiple neurexins. *NRXN1α* expression was widely distributed in all brain regions, while *NRXN1β* showed selective expression. Moreover, different splice variants of each neurexin isoform were expressed independently in different regions of the brain, and the five variable splice sites within *NRXN1α* were used in all possible combinations [216]. Ullrich and associates hypothesized that alternative splicing provides a very large number of distinct neurexin products with different ligand interactions. Thus, haploinsufficiency for α -neurexins in our patient could affect a variety of brain functions.

Geppert *et al.* [213] created mice that were homozygous null for the *NRXN1α*, but not the *NRXN1β*, transcripts. These mice were viable, fertile and indistinguishable in appearance from wild type mice. However, the females were less able than control mice to care for their litters, with the result that more pups died when attended by neurexin 1 α -deficient mothers. Poor parenting ability in mice is considered to be a model of human autism [225].

Missler *et al.* [226] used triple neurexin 1 α , neurexin 2 α , and neurexin 3 α knock-out mice to show that α -neurexins are required for normal neurotransmitter release through the action of pre-synaptic calcium channels. Others have used α -neurexin knock outs in various combinations to show that α -neurexins are important in post-synaptic as well as pre-synaptic

function [227], are required for efficient neurotransmitter release at neuromuscular junctions [228], act as organizer molecules at synapses [229], and are essential for Ca^{2+} -triggered exocytosis from neurons and endocrine cells [230]. These data support the idea that alteration of α -neurexin expression may impair normal function and development of the mouse brain.

The relationship of *NRXN1 α* haploinsufficiency to our patient's unusual vertebral anomalies is unknown. To our knowledge, no *NRXN* expression studies have been reported in mammalian embryos, but α -neurexins are expressed in a segmental pattern along the developing spinal cord in *Xenopus* embryos [231]. Expression of the *Drosophila* homolog of the vertebrate α -neurexins has also been demonstrated during early embryogenesis, with enrichment in the nerve cord [232].

3.4 Conclusion

This report is unique in that our patient exhibits haploinsufficiency of *NRXN1 α* but not *NRXN1 β* . His clinical presentation includes cognitive impairment, autistic features, vertebral anomalies, and mild facial dysmorphism. These findings support the idea that α -1-neurexin is necessary for normal neurological development and suggest that correct neuronal development and functioning is *NRXN1 α* dosage sensitive.

Chapter 4: A novel *de novo* 1.1 Mb duplication of 17q21.33 associated with cognitive impairment and other anomalies³

4.1 Introduction

Whole genome array genome hybridization (AGH) is now recognized to provide at least twice the ability of conventional cytogenetic analysis to identify genetic imbalances that cause mental retardation (MR). We previously reported detecting potentially pathogenic *de novo* copy number variants (CNVs) in 11 of 100 trios, each comprised of a child with idiopathic MR and both normal parents, using Affymetrix GeneChip® Human Mapping 100K assays [233]. Here we present an in-depth genotype-phenotype analysis of one case from this cohort, who was found to have a ~1.1 Mb duplication of 17q21.33. Two genes within the duplicated segment are strong candidates for causing the abnormalities observed in this patient. The affected region is distinct from that involved in the recently described 17q21.31 microdeletion/microduplication syndrome [9, 234]. These studies were approved by the University of British Columbia Research Ethics Board.

4.2 Case report

The female patient was the first child born to healthy non-consanguineous parents. She was delivered by emergency cesarean section at 37-38 weeks gestation because of placenta praevia. Her birth weight was 2665 g (10th centile). The pregnancy was otherwise uncomplicated. The patient was noted at birth to have small, abnormally shaped ears.

Her growth has remained significantly below average, and her development has always occurred more slowly than expected. At age 7 years her overall cognitive level was assessed to be in the low range (full scale IQ =4th centile) using the Weschler Intelligence Scale for Children, with significant difficulties in both the verbal (5th centile) and performance (5th centile) areas. Additional testing at age 8 years and 3 months using two non-verbal tests, a test of visual and verbal memory, and two parent report questionnaires, showed her to be at the upper

³ **Zahir F.R.**, Langlois S, Gall K., et al. Am J Med Genet A. 2009 Jun;149A(6):1257-62.

end of intellectual disability. She was found to have gross motor impairment (<1% on the Movement Assessment Battery for Children), fine motor skills below average on the Bruininks-Oseretsky Test of Motor Proficiency, and a complex learning profile. She received individualized education with an aide throughout elementary school.

The patient was found to have moderate-severe conductive hearing loss in both ears at 18 months of age. She has been wearing hearing aids since the age of 27 months, and her speech has improved with the aids. She had recurrent otitis media in early childhood and was fitted with three sets of pressure equalization tubes between the ages of 2 and 5 years. The ear infections became less frequent as she grew older. She was noted to have slight astigmatism (more severe on the left than the right) at 8 years and 6 months of age. Her astigmatism subsequently worsened, and she was prescribed glasses. She has never had seizures, but her parents report she has had episodes of migraine headache since around 9 years of age. At 9 years 9 months of age she was noted to have mild scoliosis. At 14 years of age she had extensive orthodontic work to advance her lower jaw to correct a severe overbite.

On physical examination at age 14 years, her head circumference was 50 cm (<3rd centile), her height was 157 cm (25th centile) and her weight was 41.4 kg (10th centile). She had bilateral epicanthic folds. Her inner canthal distance was 2.4 cm (<3rd centile), which was proportional to the head. Her palate was high, narrow and arched. Both her ears were small (<3rd centile), with the right being slightly smaller than the left. Both ears were cupped and prominent, with a small nodule on the superior aspect of the helix and very small lobes (Figure 4.1). Pre-auricular tags were present bilaterally. She had a normal nose, philtrum and mouth. She had a narrow, high-arched palate. There was mild scoliosis with the right scapula being more prominent than the left. There was limitation of extension of the 4th and 5th digits of both hands, and limitation of ankle flexion bilaterally. She had mild 2-3 syndactyly of the toes of both feet. Her pubertal development was delayed, with Tanner 3 pubic hair development and Tanner 2 breast development.

A head CT scan at about 10 years of age was normal, and a radiograph of the thoracolumbar spine at the same age demonstrated a 5-6° S-shaped curve, with the thoracic component convex to the right and the lumbar component convex to the left (Figure 4.2,A). Subsequent

radiographs at age 14 years showed an S-shaped scoliosis with a 27° thoracic convexity to the right and a compensatory 16° lumbar convexity to the left. There was also rotation of the middle to lower thoracic curvature (Figure 4.2,B). Hand radiographs at 14 years of age showed a mild contracture of the left 4th proximal interphalangeal joint and a minor degree of lateral abduction of both 5th digits at the metacarpophalangeal joints (Figure 4.2 C). Ankle radiographs at the same time showed an os trigonum in the left ankle (Figure 4. 2 D).

Cytogenetic analysis of peripheral blood lymphocytes revealed a normal female (46,XX) karyotype at 600–650 band resolution. AGH using the Affymetrix GeneChip® Mapping 100K Assay revealed a heterozygous duplication of ~1.1 Mb on chromosome 17 in band q21.33 between genomic coordinates 45,093,544 and 46,196,038 bp (NCBI build 36.1), based on location of the most proximal and distal probes included in the duplication, SNP_A-1722156 and SNP_A-1744474, respectively. The duplication was not present in either parent by AGH using the same method and was confirmed to be *de novo* by fluorescence in situ hybridisation (FISH) on blood lymphocytes from the child and both parents using bacterial artificial chromosome probes RP11-2A16 and RP11-121F10.



(A)



(B)

Figure 4.1: Mildly dysmorphic face of patient (A) with prominent cupped ears (B) at 10 years of age.

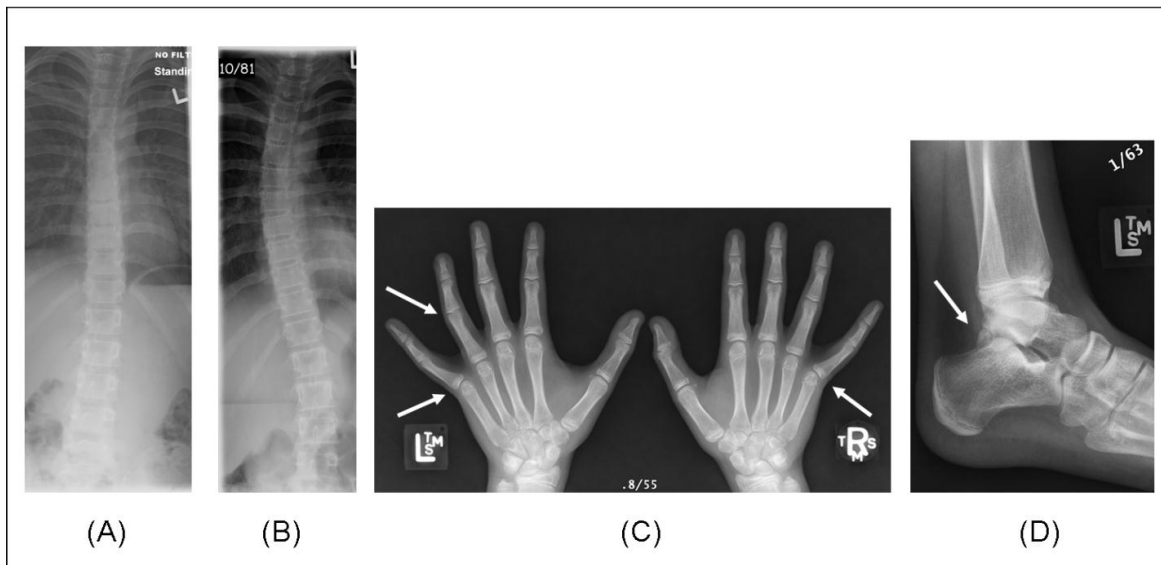


Figure 4.2: A and B, spine radiographs at 10 and 14 years of age, respectively, showing scoliosis that increased with age. C, hand radiographs (arrows indicate flexion of the left 4th proximal interphalangeal joint and lateral abduction of both 5th digits at the metacarpophalangeal joints) and D, left ankle radiograph (arrow indicates os trigonum) at 14 years of age.

4.3 Discussion

Our patient has an unusual phenotype and a 1.1 Mb duplication of 17q21.33. We found no other reports of overlapping submicroscopic duplications (Publications or online repositories DECIPHER (www.decipher.sanger.ac.uk), CAD (www.ukcad.org.uk), or ECARUCA (<http://agserver01.azn.nl:8080/ecaruca/ecaruca.jsp>). A patient with a cytogenetically apparent *de novo* duplication of 17q21.33 [46,XY,inv dup(17)(q21.3q22)] characterized by FISH has been reported [235]. Leana-Cox et al. confirmed that the duplicated material was from chromosome 17 but did not determine the precise genomic region affected. Their patient, a 3 day-old male, presented with up-slanting palpebral fissures, large beaked nose with prominent nasal tip, micrognathia, wide-spaced nipples, syndactyly, hypoplastic nails and coronal hypospadias. However, the lack of exact molecular breakpoint data precludes further genotype-phenotype correlations with our case.

Four variants within the 1.1 Mb region of 17q21.33 duplicated in our patient are reported in the Database of Genomic Variants (<http://projects.tcag.ca/variation>, as of July 2nd 2008) (Figure 4.3). Three copy number losses are described: variant #5017 [45,949,241-46,110,829] – 12 cases, variant #5868 [45,955,070-45,962,857] – 1 case and variant #5867 – [45,181,829-

45,182,888]1 case. The only possible gain reported is an indel, variant #23523 [~2.5 kb at 45,473,490], described in a single individual's genome [236]. In total, less than 163 kb of our critical region is involved in the reported benign variation (the indel does not affect coding sequence). This 163 kb includes a total of six genes, one of which (*CACNAG1*) we have included in our list of important genes below (Table 4.1). Five of these six genes, including *CACNAG1*, are located within a single BAC clone that was found to be deleted (Variant #5017) in a cohort composed of normal and diseased individuals [142]. We do not believe the normal variation reported in this region undermines the pathogenicity of our patient's CNV.

The maximal deleted region of 1,235,938 bp between the proximal and distal non-duplicated SNP probes in our analysis includes a total 17 reviewed RefSeq genes and 12 provisional RefSeq genes (Figure 4.3). Several of these genes have important functions in development (Table 4.1), but two genes, *PPP1R9B* and *COL1A1*, are especially interesting with respect to our patient's phenotype, indicating a possible contiguous gene duplication disorder [16, 237].

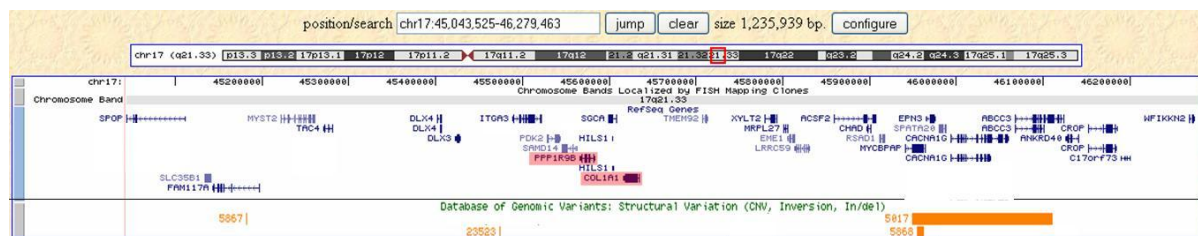


Figure 4.3: Showing the ~1.2 Mb genomic sequence (from 45,043,525 to 46,279,463) for the maximally affected region on the UCSC genome browser (NCBI build 36.1). Genes in the region are shown in the top portion with *COL1A1* and *PPP1R9B* highlighted. The bottom portion shows benign variants in the region reported in the Database of Genomic Variants (DGV CNVs).

PPP1R9B (protein phosphatase 1, regulatory (inhibitory) subunit 9B, Entrez gene ID 84687) encodes spinophilin, a protein located in the heads of neuron dendritic spines. Spinophilin functions as a targeting and regulatory subunit of protein phosphatase 1, an enzyme involved in postsynaptic signal integration [238], and has an essential modulatory function for synaptic transmission and dendritic spine morphology in mouse [239].

In a whole genome expression study comparing lymphoblastoid cell lines from patients with fragile X mental retardation to those of age-matched controls, Bittel et al. found *PPP1R9B* to be one of 90 genes showing at least a 1.5 fold change in expression level [240]. On subsequent

RT-PCR analysis, *PPP1R9B* expression was found to be increased in the fragile X samples by up to 3.5 fold. In another study, D'Agata et al. compared the brain expression patterns of 6789 genes in fragile X syndrome mouse models to normal mice using cDNA microarrays [241]. Their gene set did not include the mouse ortholog of *PPP1R9B* but did include another member of the protein phosphatase 1 family – the mouse ortholog of *PPP1CB* (protein phosphatase 1, catalytic subunit, beta isoform), which has a similar function. Interestingly, *PPP1CB* showed >2 fold increased expression in the brains of the fragile X mice in comparison to normal mice. These data are consistent with the hypothesis that *PPP1R9B* over-expression is pathogenic for intellectual disability in our patient.

COL1A1 (Collagen Type 1 α 1, Entrez gene ID 1277) encodes the pro- α 1 chain of type 1 collagen. Type 1 collagen is the major structural protein in bones, ligaments, skin, tendons, sclera, blood vessels, hollow organs and corneas. Type 1 collagen is a triple helix made up of two α 1 chains (encoded by *COL1A1*) and one α 2 chain twisted around each other [242]. A correct stoichiometrical balance between the chains is necessary to achieve the proper three-dimensional structure of collagen [243]. Duplication of *COL1A1* may lead to increased production of collagen type1 α 1 subunits, which may in turn cause production of aberrant type 1 collagen and contribute to the skeletal abnormalities in our patient. Collagen type 1 gene over expression has been reported in fibroblasts of patients with scleroderma, a connective tissue disorder characterized by the excessive deposition of extra cellular matrix macromolecules (including collagen) in skin and other organs [244, 245]. The spontaneous mutant tight skin 2 (*Tsk2/+*) mouse has been studied as a model of the condition. Studies of collagen gene expression in *Tsk2* mice dermal fibroblasts have shown that *COL1A1* gene transcription is increased relative to wild type [246, 247]. Barisic-Dujmovic et al also report that the *Tsk2/+* mice are considerably smaller than *+/+* mice, with significant decrease in body weight as well as body length [246]. These findings are of interest with respect to the reduced weight and height in our patient. However, we have not noticed dermal thickening in our patient. Mutations of *COL1A1* are causative of osteogenesis imperfecta (OI) and Ehlers Danlos syndrome (EDS) [248, 249]. Our patient does not have clinical features of either OI or EDS, but hearing loss and scoliosis, which are present in our patient, are commonly seen in OI, and scoliosis is frequent in EDS. OI type I and OI type IV, in which hearing loss is most prevalent, are usually caused by dominant negative sequence mutations of *COL1A1* which is also the

most common mechanism of action for *COL1A1* mutations causing EDS. Other important RefSeq genes in the interval are given in Table 4.1.

We screened the affected region and sequence flanking the breakpoints up to 1 Mb in either direction for duplicons that could have mediated the *de novo* CNV in our patient via non-allelic homologous recombination. We found no segmental duplications, although the region does contain numerous SINEs, LINEs and other low complexity repeats that may promote the formation of CNVs through other mechanisms [250].

CNVs that are *de novo*, gene-rich and large in size are more likely to be pathogenic than those that are inherited, gene-poor and small [251-253]. The ~1.1 Mb CNV in our patient is much larger than the median size of benign CNVs, which is <100 kb [252, 254]. Although duplications are generally less likely to be pathogenic than deletions, numerous examples of duplications involving genomic regions as small or smaller than that observed in our patient are known to be pathogenic [237, 255, 256]. The fact that the duplicated region in our patient contains many genes, including those for which over-expression has been associated with pathogenicity (*PPP1R9B*, *COL1A1*), strongly supports this duplication being causative of our patient's phenotype.

4.4 Conclusion

We report the first apparently pathogenic submicroscopic duplication of 17q21.33. This ~1.1 Mb lesion includes *PPP1R9B*, a good candidate for the developmental delay and microcephaly, and *COL1A1*, a good candidate for the hearing loss and skeletal deformities seen in our patient.

Table 4.1: Important RefSeq genes in the duplicated segment. # These genes are considered good candidates and discussed in detail in the text.

Gene	Known function
# PPP1R9B (<i>protein phosphatase 1, regulatory subunit 9B</i>)	Located in the heads of dendritic spines and involved in postsynaptic signal integration. Further discussion in text.
# COL1A1 (<i>collagen type 1 α1</i>)	Encodes pro- α 1 chain of type 1 collagen, a major structural protein in the body. Further discussion in text.
DLX3 (<i>distal less homeobox 3</i>)	Defects in <i>DLX3</i> cause autosomal dominant trichodontoosseous syndrome, as well as autosomal dominant amelogenesis imperfecta hypomaturational-hypoplastic type with taurodontism. Amelogenesis imperfecta (AI) represents a group of developmental conditions affecting the structure and clinical appearance of the enamel of all or nearly all the teeth in a more or less equal manner. In both disorders the indicated mechanism of pathogenicity is LOF or dominant negativity. <i>DLX3</i> is one of a 6 member family of proteins implicated in human hair, tooth, placental and craniofacial morphogenesis [257, 258].
DLX4 (<i>distal less homeobox 4</i>)	A member of the DLX family, may regulate placental development. Murthi et al show increased <i>DLX4</i> expression in placenta's of fetuses with IUGR [259].
CACNA1G (<i>voltage-dependent calcium channel alpha 1G</i>)	Voltage-sensitive calcium channels (VSCC) mediate the entry of calcium ions into excitable cells and are also involved in a variety of calcium-dependent processes, including muscle contraction, hormone or neurotransmitter release, gene expression, cell motility, cell division and cell death. Highly expressed in fetal and adult brain, and moderate expression in fetal heart, kidney and lung tissue as well moderate expression in the same adult organs.
TAC4 (<i>Tachykinin 4 isoform delta</i>)	Member of the tachykinin family of neurotransmitter encoding genes. Tachykinin receptors play a role in endocrine and paracrine signaling. Exact role of this gene's product not known.
PDK2 (<i>Pyruvate dehydrogenase</i>)	Inhibits the mitochondrial pyruvate dehydrogenase complex. Expression level association with cervical cancer reported [260].
ACSF2 (<i>acyl-CoA synthetase family member 2</i>)	Not much information known about this gene. However another gene which likely has related function, <i>ACSL4</i> (<i>acyl-coA synthetase long chain family member 4</i>) LOF is reported to cause MR [261].
ITGA3 (<i>Integrin, isoform 3 alpha</i>)	Cell surface receptor, known role in neuronal migration (associated with Reelin).
SPOP (<i>speckle type POZ protein</i>)	Involved in proteasome degradation pathway and protein ubiquitinylation pathway. Epigenetic modulator.
SCGA (<i>alpha sarcoglycan a.k.a Adhelin</i>)	This gene encodes a protein that is a component of the sarcoglycan complex, a subcomplex of the dystrophin-glycoprotein complex which forms a link between the F-actin cytoskeleton and the extracellular matrix. LOF causes adhelin deficient limb girdle muscular dystrophy.
MYST2 (<i>MYST histone acetyltransferase</i>)	Histone acetyltransferase. Epigenetic modulator with global function.

Chapter 5: Duplications of the critical Rubinstein Taybi deletion region cause a novel recognizable syndrome characterized by mild arthrogryposis.⁴

5.1 Introduction

Chromosomal imbalances are an important cause of genetic disorders, especially when the patient presents with a syndromic phenotype. The identification of such imbalances on a genome-wide level was until recently only feasible by investigating metaphase chromosomes under a microscope. The latest developments in microarray technology however enable assessment of copy number of thousands to millions of loci across the human genome at a resolution far surpassing that of karyotyping metaphase chromosomes. The introduction of these technologies in to the diagnostic work-up of patients with congenital disorders represents a revolution in this field of which the importance cannot be underestimated. It led to a vast improvement in the etiological diagnosis of patients with idiopathic congenital abnormalities, mental retardation (MR) and/or psychiatric problems [135]. It moreover enabled the identification and delineation of novel microdeletion and microduplication syndromes, thus enabling a more detailed assessment of the phenotypical consequences associated with specific chromosomal imbalances and a more accurate and targeted counseling and therapy of patients and parents [262]. The present study describes the identification and delineation of a novel microduplication syndrome, microduplication 16p13.3, which is complementary to microdeletions of 16p13.3 which cause Rubinstein Taybi syndrome (RTS) [263].

5.2 Patients, materials and methods

Patients described in this study were identified from different groups of patients referred for idiopathic MR and/or congenital anomalies. Patients were followed in the collaborating centers and clinical data and informed consent were obtained from all patients or their legal representatives. A genome-wide copy number profile of the patient DNA was obtained by subjecting it to microarray analysis using the Affymetrix 500K GeneChip platform (patient 2), the genome-wide 105K V7 OLIGO array (a custom designed array manufactured by Agilent Technologies, Inc. (Santa Clara, California, USA), containing oligonucleotides as probes

⁴ Thienpont B, Béna F, Breckpot J, et al. J Med Genet. 2010 Mar;47(3):155-61

chosen genome-wide with an average interval of 30 kb plus an enrichment of probes in most known regions associated with syndromes, pericentromeric and subtelomeric regions) (patients 1, 3, 6 and 8), the Agilent 244K platform (patients 7, 9 and 11) or the Agilent 44K platform (patient 12), all according to manufacturer's recommendations, or using a BAC array containing a tiling resolution of chromosome 16 (patient 4) as described [264]. All genome coordinates mentioned in this study are according to human genome build 18 (NCBI 36.1). The inheritance of each of these duplications was investigated by analysing parental samples using microarrays (patient 2) or targeted approaches such as fluorescence in situ hybridisation (FISH) (patients 1, 3, 4, 6, and 8), MLPA (patients 9 and 11) or quantitative polymerase chain reaction (PCR) (patients 7 and 12). Positions of segmental duplications were downloaded from the segmental duplication database website (<http://humanparalogy.gs.washington.edu/build36/>). The duplications in patients 2, 5 and 10 were previously reported [132, 237, 265].

5.3 Results

Nine patients with an interstitial duplication of 16p13.3 are described for the first time in the present study. Of the three previously reported patients with an interstitial duplication of 16p13.3 (patients 2, 5 and 10) [132, 237, 265], we have added more accurate genotypic and phenotypic data for patient 2. The results of microarray analyses for patients 1-4, 6-9 and 11-12 are depicted in Figure 5.1 and the molecular data are summarized in Table 5.1. The duplication occurred *de novo* in 10 cases and was inherited from a clinically normal parent in two cases (patient 2 (maternal) and patient 8 (paternal)).

Comparing the extent of all duplicated regions enabled the delineation of a critically duplicated region, which contained a single gene: *CREBBP* (Figure 5.1). We also reanalyzed the molecular karyotype of the patient reported by de Ravel [266] using a higher-resolution platform (Agilent 244K). This revealed that the duplication described in this patient was not interstitial but telomeric (extending from 0 kb to 8,633-8,648 kb), and is 2-10 times larger than the interstitial duplications in patients 1-12. This patient was, therefore, not included in the present analysis.

There was no family history of birth defects except for patient 2, whose father had one brother with a congenital heart defect who died at 8 months of age and one sister with a

history of cardiac arrhythmias, and patient 3, where both parents had heart murmurs at birth that resolved spontaneously after a few years.

Phenotypically, interstitial duplications of 16p13.3 are associated with variable mental development (ranging from normal to moderately delayed), and occasional behavioural problems such as attention deficit hyperactivity disorders, aggressive behaviour, and autism spectrum disorders. Pregnancy was uncomplicated in most; the mother of patient 3 had one abnormal non-stress test and mild oligohydramnios was noted, and patient 12 had intrauterine growth retardation. There was no exposure to alcohol or other known teratogens except in patient 3, where the mother reported occasional alcohol consumption in the first weeks of pregnancy. Birth was at term in all patients. Growth was normal in all individuals except patient 12 who had precocious puberty.

Many patients shared similar facial features. There was midfacial hypoplasia in young children and a longer face in older individuals. The nose was prominent and had a bulbous tip, and the eyes were often upslanting with narrow palpebral fissures, sometimes with ptosis. The upper lip typically was thin, and the ears low set and/or protruding (figure 5.2). We frequently observed mild abnormalities of the hands (the thumbs were often proximally implanted and short, fingers were long and tapering, the fifth finger was often short, there was also often camptodactyly or mild cutaneous syndactyly) and of the feet (club feet, camptodactyly, or syndactyly) (Figure 5.2). Less frequent were other anomalies of the skeleton (congenital hip dislocation, vertebral fusion), the eyes (blepharophimosis, epicanthus inversus, strabismus, astigmatism, or ptosis), and the heart (atrial septal defect, tetralogy of Fallot). Occasional findings include inguinal hernia (twice), precocious puberty, cryptorchidism (twice), submucosal cleft palate, and mild periventricular heterotopia on magnetic resonance imaging (MRI) examination of the brain. An overview of the phenotypic data of the individual patients is given in Table 5.1. The frequent occurrence of features such as club feet, camptodactyly of the toes and fingers, congenital hip dislocation, and incomplete extension of the elbows is indicative of mild arthrogryposis in these patients. A total of about 25,700 patients have been analyzed for duplication of *CREBBP* in our centers. In this patient population, 11 duplications of *CREBBP* were found, suggesting a frequency of this duplication of around 0.043% in these patients. As about 2-3% of the human live births fit the inclusion criteria for our patient cohort (MCA/MR) and about 80% of these are idiopathic before molecular karyotyping [6], we estimate the frequency of this duplication to be 1 in 97,000 to 146,000 live births.

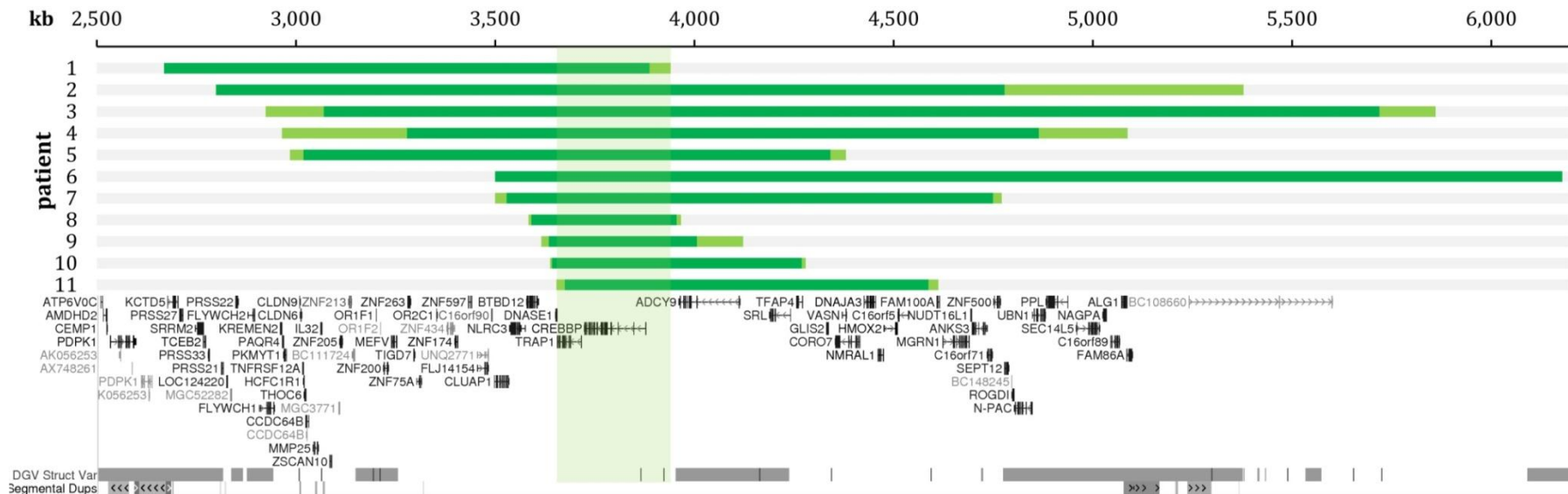


Figure 5.1: *Upper panel:* the extent of the duplications identified in the present study. Grey bars indicate normal copy number, green bars indicate duplicated regions, light green bars indicate the breakpoint containing regions. Patients are ordered according to the telomeric breakpoint. The transparent green box demarcates the smallest region of overlap. *Lower panel:* a view of the UCSC genome browser (NCBI build 36.1, March 2006), illustrating the genes mapped to the implicated region (first track), the regions that are copy number variable in the normal population (second track), and the regions that have paralogues in the human genome (low copy repeats, segmental duplications; third track). All breakpoints appear unique and do not map specifically in low copy repeats.

Table 5.1: Genotypic and phenotypic characteristics of patients with an interstitial 16p13.3 duplication. Patients 2 [132], 5 [265] and 10 [237] were reported previously. Abbreviations: ADHD = Attention-deficit hyperactivity disorder, ASD = Atrial septal defect, AuSD = Autism spectrum disorder, bcr = breakpoint containing region, OFC = occiputofrontal circumference, VCS = vena cava superior.

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Tel bcr (kb)	1863–1880	2680–2682	2812–2813	2937–3083	2978–3292	2998–3032	3506–3513	3513–3542	3597–3604	3629–3648	3651–3656	3667–3688
Cen bcr (kb)	4665–4693	3874–3927	4765–5365	5706–5847	4851–5074	4328–4367	6165–6177	4736–4758	3942–3953	3993–4109	4281–4290	4574–4599
Inheritance	De novo	De novo	Maternal	De novo	De novo	De novo	De novo	Paternal	De novo	De novo	De novo	De novo
Gender	Female	Male	Male	Male	Female	Male	Male	Female	Female	Male	Female	Male
Biometry W	p50	p50	p50	p5	p10	p50	p50	p10–25	p3	p25	p10	–
At birth L	p50	–	<p5	–	p75	–	p25	p5–10	<p3	p25–50	p10	p90
OFC	–	–	p50–75	–	p25	–	p75	p25	–	p50–75	p50	–
Age last consult	4½ months	13 y 6 months	3 y	11 y	16 y	8 y	18 y	3 y	8 y	15 y	2 y 9 months	10 y
Growth	Normal H p10; W p25; OFC p10	Normal H p90; W p75–90; OFC p90	Normal Wp25–50; H p10; OFC p3	Normal W p3, H p50, OFC p10	Normal	Normal	Normal W p25–50; H p75–97; OFC p25–50	Normal W p25; H p90	Normal W p10; H p10–25	Normal W, L and OFC at p75	Normal W and H p25–p50; OFC p50–p75	Precocious puberty (W >p97, H >p97, OFC p40)
Delay in development	–	Mod (WISC-III IQ 46); pronounced speech delay	Mild (LIPS-R: brief IQ 68—fluid reasoning 73)	Mod (WISC-R IQ 51—perf: 52; verb: 54)	Mod (WISC-R IQ 45 perf 63; verb 39) severe verbal dyspraxia	Mod (WISC IQ 50)	Mild-mod (IQ 56); speech articulation problems	Mild (PLS-4: expressive language 74; total 77)	Low normal intelligence; speech articulation defect	Mod	Mild	Low normal intelligence
Behaviour	–	No behavioural problems	No behavioural problems	Behavioural problems	No behavioural problems	AuSD; ADHD	AuSD; ADHD; aggressive	No behavioural problems	No behavioural problems		No behavioural problems	No behavioural problems

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Hands	Short, proximally implanted, adducted thumbs	Short fifth fingers	Hypoplastic thumbs and thenar eminences; long fingers	Short, proximally implanted thumbs; camptodactyly II-V	Tapering fingers; short fifth finger	Camptodactyly 2–5	Short fingers	Tapering long fingers	Proximally implanted thumbs; long fingers; hyperconvex nails	Mild cutaneous syndactyl camptodactyly (fingers 2–4)	Small, proximally implanted and adducted thumbs	Large, proximally implanted thumbs; mild syndactyly 2–3, camptodactyly (2, 3 and 4)
Feet	Vertical talus; dorsiflexion; proximally placed toes	Pes cavus; camptodactyly; 4–5 clinodactyly	Medially deviating, broad toes; abnormal nail beds	Camptodactyly; sandal gap	Bilateral equinovag; short toes; camptodactyly; sandal gap	Camptodactyly 2–4	Sandal gap; camptodactyly; sandal gap	Mild 2–3 syndactyly; club foot right	Dysplastic toe nails; camptodactyly; hallux valgus	–	Short halluces	Large halluces; sandal gap
Musculoskeletal findings	–	C5–C6 vertebral fusion; mild pectus excavatum	Plagiocephaly; leg length discrepancy; normal joint mobility	Mild pectus excavatum; flat, asymmetric thorax	–	–	Scoliosis	Hip dislocation; spinal lipoma	Severe congenital left hip luxation	Incomplete extension elbows	–	–
Eyes	–	Blepharophimosis; bilateral ptosis; epicanthus inversus	Bilateral ptosis, esotropia	Ptosis left eye, eye motility disturbance	–	–	Deep set	Normal	–	–	Strabismus	Normal
Heart	ASD-I; VSD; leaky AV valves	Normal	ToF	Normal	–	ASD	Normal (VSD spontaneously closed)	Normal	Normal	ASD type II	Normal	–

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Dysmorphism	Bi temporal narrowing ; short upslanting palpebral fissures; left preauricular tag; lacrimal duct obstruction	Low set, mildly dysplastic ears; overturned helices of the ears; upslanting palpebral fissures; bulbous nose	Sacral dimple; bifid uvula; short neck with excess skin; micrognathia; thin upper lip; double hair whorl	Webbed neck; anteverted nares; thin upper lip; mildly protruding ears; epicanthus; long face; pointed chin; narrow palpebral fissures; smooth philtrum	Deep set eyes; small, upslanting palpebral fissures; protruding , abnormal ears; broad nasal bridge; round nasal tip with prominent glabella, long philtrum, midfacial hypoplasia and prominent mandible	Protruding ears; bulbous nose; deep set eyes; tented upper lip	Sacral dimple; upslanting palpebral fissures; hypertelorism; synophrys ; low set ears; webbed neck; small chin	Midfacial hypoplasia; posteriorly rotated, protruding ears; upslanting palpebral fissures; short nose	Short hypoplastic nose; low set and posteriorly rotated small ears; long flat philtrum; small teeth; facial hypotonia	Mildly protruding ears, broad nose, hirsutism; synophrys; low frontal hair line	Dolichcephaly; upslanting palpebral fissures; large low set ears; low nasal bridge; micrognathia	Smooth philtrum; large mouth; posteriorly rotated, protruding and malformed ears; short nose; upslanting, narrow palpebral fissures
Other	–	–	Right cryptorchidism; small corpus callosum; EEG normal	Cryptorchidism	Normal MRI and EEG	Inguinal hernia; trachea-bronchomalacia; recurrent respiratory infections; submucous cleft	Inguinal hernia; pyloric stenosis; hypotonia ; high palate	Cerebellar angioma; periventricular leucomalacia; thin hair	Permanent drooling	–	Hypotonia; recurrent ear infections	Mild periventricular heterotopia precocious puberty

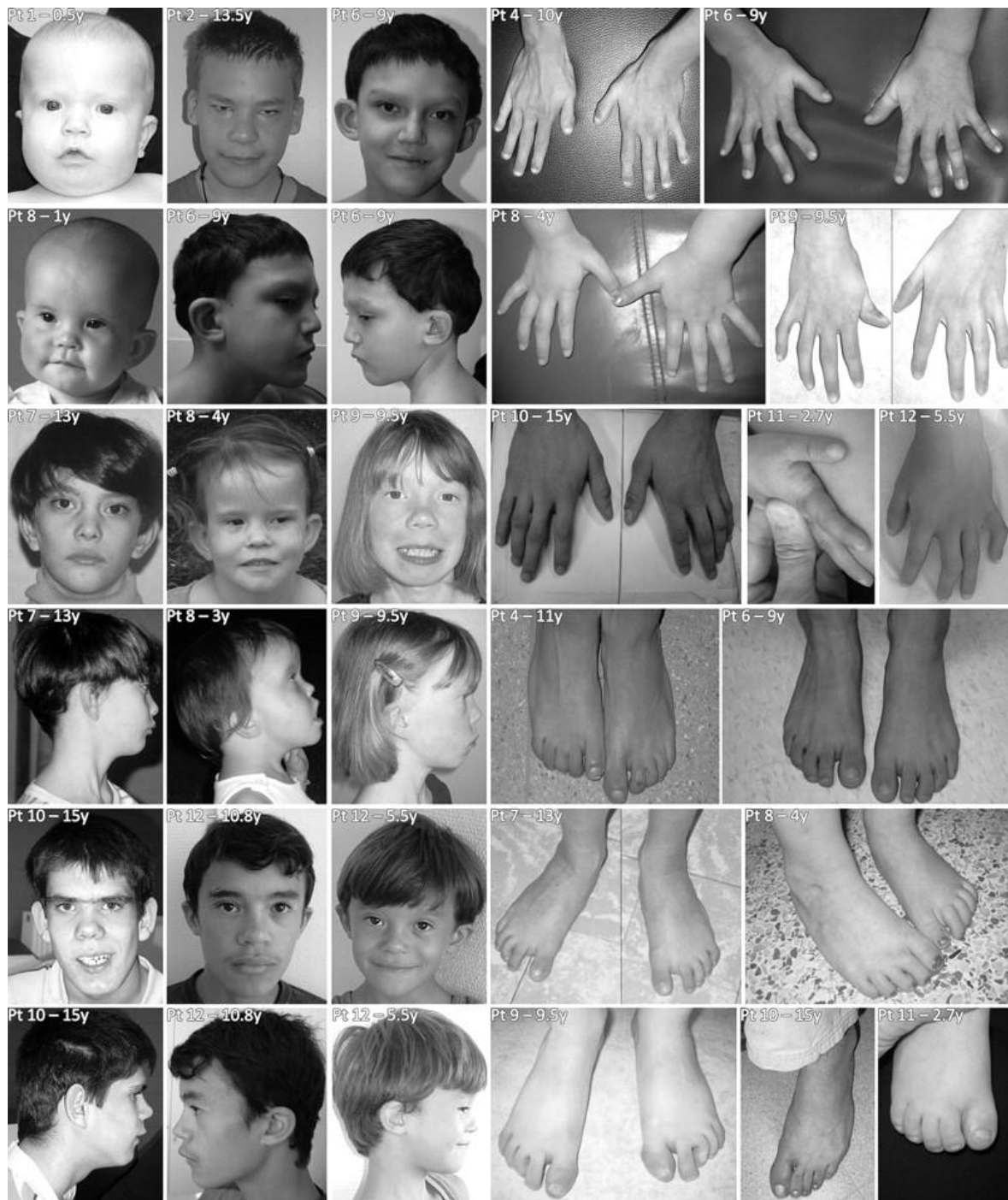


Figure 5.2: Phenotypic features of patients 1, 2, 4 and 6–12. Please note that patient 2 has had a surgical correction for ptosis.

Three different mechanisms have been proposed to generate chromosome imbalances : non-allelic homologous recombination (NAHR), non-homologous end-joining (NHEJ) and fork

stalling and template switching (FoSTeS) [267]. When a region is recurrently found to be deleted as well as duplicated, NAHR has been suggested as a potential underlying mechanism. However, similar to the breakpoints of deletions in this region, all breakpoints of duplications appear unique, arguing against a mechanism of non-allelic homologous recombination. To further exclude a local homology-based molecular mechanism, we assessed the presence of segmental duplications (SD; length > 10kb, over 90% homology) in the breakpoint containing regions. SDs encompass 7.7% of the region between 2.5 and 6.5 Mb, but none of the breakpoints fall clearly within an SD, and only two out of 22 breakpoint containing regions (9%) encompass a SD. We therefore exclude NAHR as the mechanism generating these imbalances. The lack of recurrent breakpoints for the CNVs in these cases indicate other mechanisms, such as NHEJ or FoSTeS as possibly causative. There were no complex imbalances identified, a hallmark of the recently proposed FoSTeS mechanism, although the resolution of the analysis was probably insufficient to assess this critically in most cases. None of the parents reported exposure to known mutagens.

5.4 Discussion

The present report details genotypes and phenotypes of nine patients with an interstitial 16p13.3 duplication. Three patients that were previously reported were also included [132, 237, 265], with more precise genotypic and phenotypic data for one of these (patient 2) [132]. Careful assessment of the phenotypic features of these individuals enabled the description of a characteristic phenotype, with normal to moderately retarded mental development, mild arthrogryposis-like anomalies of the musculoskeletal system (club feet, congenital hip dislocation, or camptodactyly of the fingers and toes), mild facial dysmorphism that changes with age and occasional anomalies of the heart (atrial septal defect, tetralogy of Fallot).

It is noteworthy that all patients were selected for study because of MR and/or congenital anomalies. There might thus be—as in most genetic disorders—a bias towards the more severe end of the spectrum in the described phenotype; normal or mildly affected individuals do not present at the genetics clinic. In two instances the duplication was indeed found to be inherited from an apparently normal parent (patients 3 and 8) who had both followed normal schooling, function normally in society, and do not present the typical face. However, the frequent de novo occurrence (in 10 of 12 patients) indicates that this duplication is associated

in most cases with a reduced reproductive fitness. In line with this is the finding that *CREBBP* duplications were never identified in genome-wide copy number profiling studies of more than 4000 individuals recruited from normal control populations [114, 268, 269]. In the patient populations we studied, we identified an interstitial 16p13.3 duplication in 11 out of 25 700 individuals, suggesting that it occurs in about 0.043% of patients with MR or birth defects.

Three typical patients (9,10 and 11) carry a very small duplication (maximal sizes 356, 480 and 639 kb) indicating this region is associated with the described phenotype of moderate MR, typical facial gestalt, mild hand and feet anomalies and heart malformations. This small region encompasses only 6 genes. Of these genes, *CREBBP* is the most attractive candidate. The smallest region of duplication overlap (186-260kb) in the 11 described individuals contains only *CREBBP* entirely. The distally flanking gene, *ADCY9*, is unaffected in patient 1. The proximally flanking gene, *TRAP1*, is only partially duplicated in patient 12 rendering the extra copy probably not functional. *CREBBP* encodes a histone acetyl transferase (HAT) and thus functions as a transcriptional co-activator by decondensing chromatin and activating gene transcription. Heterozygous loss-of-function mutations and deletions of this gene have been shown to cause RTS, demonstrating [270] that human development is sensitive to *CREBBP* dosage.

None of the recurrent phenotypic findings associate specifically with a region proximally or distally from *CREBBP*, except for ptosis, which was found only in three patients with a more telomerically extending duplication (patients 2-4). In contrast to duplications extending to the subtelomere [271], interstitial duplications are not associated with microcephaly or growth retardation. Other features reported for terminal duplications seem to associate with duplication of *CREBBP*. These include heart defects, mental retardation and the arthrogryposis-like features [271].

Reciprocal duplications are now being clinically delineated for many of the the previously described microdeletion syndromes. Examples include 7q11.23 duplications (reciprocal to Williams-Beuren syndrome deletions) [272], 15q11.2 duplications (reciprocal to Prader-Willi and Angelman syndrome deletions) [46], 17p11.2 duplications (known as Potocki-Lupski syndrome, reciprocal to Smith-Magenis syndrome deletions) [16], 22q11.2 duplications

(reciprocal to DiGeorge syndrome deletions) [273, 274] and Xq28 duplications (*MECP2* duplications, reciprocal to Rett syndrome deletions) [275] and 17p13.3 duplications (reciprocal to Miller–Dicker syndrome deletions) [276]. Some recurrent themes are emerging. In general, the phenotypic manifestations are milder and much more variable as compared to the deletions. Of interest, non-penetrance has been reported for several of the duplications (e.g. dup 22q11.2, dup 15q11.2) [46, 273].

The defects observed in reciprocal deletion and duplication syndromes often involve the same organs or functions, for example the heart in DiGeorge and duplication 22q11.2, the speech in Williams-Beuren and duplication 7q11.23, the teeth in Smith-Magenis and Potocki-Lupski syndromes or the nerves in Charcot-Marie tooth 1A and hereditary neuropathy with liability to pressure palsies [277, 278]. Similarly, defects of the hands, the feet and of the heart are seen in both deletion and duplication of 16p13.3.

In some instances, an intriguing reciprocal phenotype is seen in the deletion versus duplication. For instance, 7q11.2 deletions have a high verbal functioning, compared to the deficiencies in speech and language development observed in the duplication 7q11.2 [272]. The bulbous nasal tip observed in the present patients may perhaps be a reciprocal phenotype to the long columella seen in RTS, the short and proximally implanted thumbs may be reciprocal to the broad thumbs typical for RTS, and the arthrogryposis-like feature arguably compare to the joint hypermobility [279].

The present findings indicate that normal *CREBBP* dosage is restricted not only by a lower threshold (as demonstrated in RTS) but also by an upper threshold. The described duplications most likely cause only a slight increase in *CREBBP* expression, with the normal dose being only 33% lower than the dose upon duplication. This poses significant challenges on the recently proposed strategy to treat RTS patients with histone deacetylase inhibitors [280]: the dosage-dependency of *CREBBP* suggests that such pharmaceutical interventions will prove beneficial in a very limited range of concentrations only. On the other hand, administering a precise dosage of HAT inhibitors such as curcumin [281] present a valuable line of investigation to develop a therapy for this novel genomic disorder. Curcumin is a cell-permeable inhibitor of multiple cellular targets including *CREBBP* and p300 (a paralogue

mutated in an allelic form of RTS), which is currently being tested in clinical trials for neoplastic and immunological disorders [282]. Although errors in development will not be curable by such strategies, the observed problems in mental function might benefit from the proposed therapeutic interventions.

Chapter 6: A characteristic syndrome associated with microduplication of 8q12, inclusive of *CHD7*.⁵

6.1 Methods of detection

6.1.1 Cytogenetics

Cytogenic analysis was performed on the patient's peripheral blood lymphocytes according to standard procedures. The patient's karyotype was normal 46,XX (550 band resolution).

6.1.2 Array comparative genomic hybridization

As part of a larger study on the detection of copy number variations in children with developmental disabilities, array genomic hybridization analysis was performed on the child and both of her parents using Affymetrix GeneChip® 500K mapping assays [7]. Chip-to-chip normalization, standardization to a reference, genotype detection, and copy number estimation on a single SNP basis were performed using Affymetrix Power Tools (version 1.6.0) software (www.affymetrix.com) as previously described [283].

6.1.3 Chromosomal anomaly

This analysis demonstrated a *de novo* 6.9 Mb duplication at 8q12.1q12.3 between genomic coordinates 58,388,614 and 65,306,097 bp (NCBI build 36.1) based on the location of the most proximal and distal SNPs involved in the duplication, SNP_A-2293240 and SNP_A-4266991 respectively. These SNPs are flanked by SNP_A-4212967 (at 58386565 bp) and SNP_A-1900852 (at 65328199 bp), both present in two copies. This was the only *de novo* event identified.

6.2 Method of confirmation

The duplication was confirmed by a more intense fluorescence *in situ* hybridization (FISH) signal on one metaphase chromosome 8 and by 3 signals in 43/50 interphase nuclei. The

⁵ Lehman A.M., Friedman J.M., Chai D, et al. Eur J Med Genet. 2009 Nov-Dec;52(6):436-9.

finding of 86% of cells with three signals is consistent with a nonmosaic state according to ours and others' experience [284]. BACs were selected from the UCSC genome browser (University of California, Santa Cruz <http://genome.ucsc.edu/>) and included probes RP11-1149A23 and RP11-831F23. FISH experiments were performed on acetic acid-fixed chromosome preparations in metaphase cells and interphase nuclei from white blood cell samples from the child and her parents. FISH with the probe RP11-832J12, which overlaps a Duane anomaly locus on chromosome 8q13.1, confirmed that it was not included in or disrupted by the duplication. Additional BAC probes, RP11-585E8 (within the duplicated region in 8q12.1, labeled with FITC green) and RP11-915H16 (within the duplicated region in 8q12.3, labeled with Texas Red) showed that the duplication is of a direct tandem orientation.

6.3 Clinical description

The uncomplicated first pregnancy of unrelated parents of East Asian ethnicity ended in the term spontaneous delivery of a normocephalic, 3.6 kg female. Atrial and ventricular septal defects were diagnosed and surgically corrected at 3 months of age. A urinary tract infection at 4 months of age led to the identification of grade II genitourinary reflux and mild right pelviectasis. Because of poor feeding, gastroesophageal reflux, aspiration, and failure to thrive, gastric tube feedings were initiated at 11 months of age. During her first year the patient was diagnosed with a right-sided Duane anomaly of extraocular movement (impaired abduction and globe retraction upon adduction), and severe hearing loss on the right side with moderate loss on the left. A cranial CT scan showed a dilated right vestibule, severely hypoplastic cochlear nerve, and abnormal modiolus on the right side, and a hypoplastic distal turn of the cochlea on the left side. She walked at age 23 months, became independent of gastric tube nutritional support at age 30 months, achieved toilet training at age 30 months, and spoke in sentences by 40 months. Mild constipation and moderate eczema have been ongoing.

Examination at 43 months showed microbrachycephaly (2nd %-ile) and normal height (75th %-ile) and weight (60th %-ile). Telecanthus and epicanthal folds were present; the interpupillary distance was normal (50th %-ile). Arched eyebrows, bilateral preauricular ear pits, upturned ear lobules, mild overfolding of the helices, and mild microstomia were observed. Ear length was normal (5.5 cm, 50th %-ile). The total hand length was 10 cm and her 3rd digit length measured 4.2 cm (both <-2 SD). Her feet were also small and demonstrated pes planus. The visceral

examinations were within normal limits. Extraocular abduction was limited, and adduction of the right eye resulted in globe retraction. Her neurological examination was otherwise significant for mild hypotonia, impaired balance and dyscoordination.

6.4 Discussion

The advent of array-based comparative genomic hybridization has led to the rapid identification of many new genomic disorders, some of which are clinically recognizable syndromes, and some that are not, owing to nonspecific or highly variable features. We propose that duplication 8q12 is a new genomic syndrome on the basis of a consistent phenotype associated with overlapping copy number variants found in a boy previously described [285] and in a girl reported here. Both patients were affected by deafness, congenital heart defects, cognitive impairment, and a Duane anomaly of extraocular movement.

We describe a girl with a *de novo* 6.9 Mb duplication within 8q12.1q12.3. She had a difficult first year with hypotonia, feeding difficulties, developmental delay, and medical and surgical interventions for congenital anomalies. She has a Duane anomaly, deafness with a Mondini malformation, atrial and ventricular septal defects, and genitourinary reflux. In addition, she has minor malformations of the external ears and face. Her cognitive development continues to progress more slowly than normal, although the delayed progression of her speech development must be considered in the context of severe hearing impairment.

The duplicated region encompasses 15 protein-coding, RefSeq annotated genes (UCSC Browser, <http://genome.ucsc.edu>), including *CHD7*, which, when inactivated or deleted, is causative of CHARGE syndrome [42]. Within this duplicated interval, copy number variants have been detected in normal individuals for *UBXN2B*, *SDCBP*, *NSMAF*, *NKAIN*, and *YTHDF3* but not for *CHD7* (Database of Human Genomic Variants, <http://projects.tcag.ca/variation>, accessed Apr 15 2009). One other individual with submicroscopic duplication encompassing *CHD7* has been reported. He has moderate intellectual disability, hypotonia, sensorineural deafness, mild dysmorphisms, ventricular septal defect, pulmonary stenosis, and, in conspicuous similarity to our patient, a Duane anomaly (Table 6.1) [285]. His 3.0 Mb duplication includes *CHD7* among his eight duplicated genes. The duplication in this patient lies

entirely within the duplication in our patient, spanning the region from about 59.5 to 62.5 Mb (Figure 6.1).

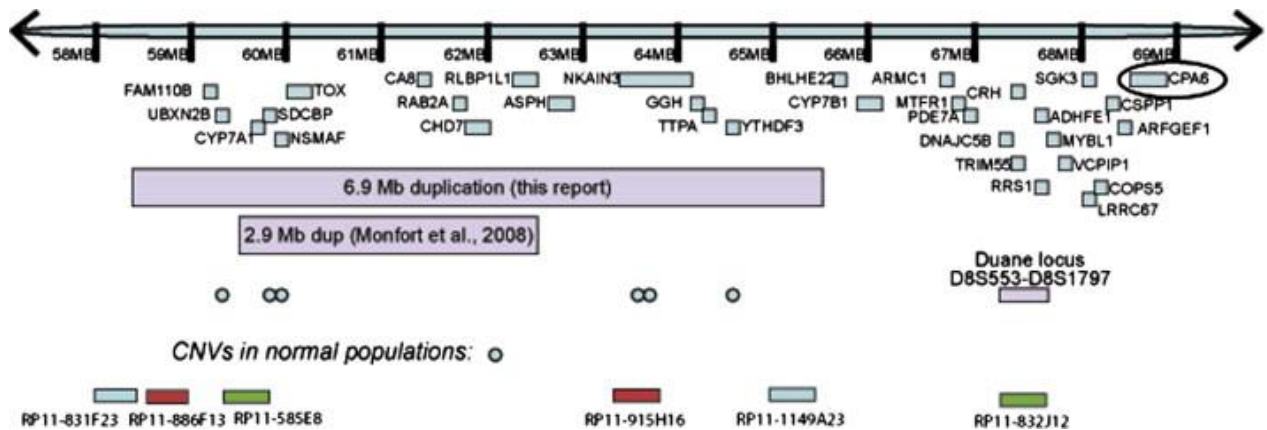


Figure 6.1: Genomic intervals of the 8q12 microduplications found in our patient and the previous patient displayed, overlaid on an adapted segment from the UCSC genome browser. In the telomeric direction, a Duane anomaly locus is shown and also the candidate Duane gene *CPA6/CPAH*. RefSeq protein coding validated genes are shown; BAC probes used in this study are represented. The figure is not precisely to scale.

System	This child	Child in Monfort et al. [285]	CHARGE syndrome
Growth	Microbrachycephaly	?	Short stature
Visual	Duane anomaly	Duane anomaly	Coloboma, strabismus, microphthalmia
Hearing	Deafness with Mondini	Deafness (inner ear undescribed)	Deafness +/- Mondini malformation
Craniofacial	Telecanthus, high eyebrows, preauricular pits	Wide nasal root, high eyebrows	Choanal atresia, external ear defects, palsies, abnormal semicircular canals
Developmental	Developmental delay	Development delay, hyperactivity	Developmental delay, hyperactivity
Cardiac	ASD, VSD	VSD, pulmonary stenosis	Conotruncal, ASD, VSD, other defects
Swallowing	GER, aspiration	?	TEF, GER, aspiration
Genitourinary	Mild vesicular reflux	?	Genital hypoplasia, renal anomalies
Limbs	Proportionate brachydactyly	?	Hand anomalies

Table 6.1: Comparison of physiological systems in two children with microduplication of 8q12 versus CHARGE syndrome. ASD = atrial septal defect; VSD = ventral septal defect; GER = gastroesophageal reflux; TEF = tracheoesophageal fistula.

The Duane anomaly is a form of complex strabismus characterized by limited ocular abduction and variably limited adduction accompanied by retraction of the globe, which narrows the palpebral fissure. It is typically caused by congenital hypoplasia of the abducens nerve or nucleus and aberrant innervation of the lateral rectus by fibers of the oculomotor nerve [286]. For a minority of affected individuals, the anomaly is familial. Various nearby genetic loci have been implicated in the Duane anomaly: deletion of 8q12.2-q21.2 [287], 8q13 deletion [288] and a translocation at 8q13 interrupting the carboxypeptidase gene *CPAH/CPA6* [289]. The relative proximity of the distal breakpoint of our patient's duplication with the 8q13 Duane anomaly locus (approximately 1.5 Mb) argues for positional epigenetic regulatory effects of the duplication (Figure 6.1). In mouse models, strain-specific CNVs occurring within noncoding regions of the genome contribute to trait variability, presumably through complex regulatory mechanisms rather than gene dosage effects [290]. It is plausible, therefore, that downstream effects of either the duplicated genes or duplicated nonprotein coding regulatory elements impact genes more proximally responsible for the Duane anomaly. Alternatively, genes in both 8q13 and 8q12 may be responsible for a Duane anomaly phenotype. Notably, CHARGE syndrome can occasionally feature strabismus, although the Duane anomaly specifically has not been reported.

Mondini malformation refers to development of only one and a half turns of the normal two and one half turns of the cochlea. There is incomplete partition with resulting confluency of the middle and apical turns. The associated hearing loss is probably sensory in origin. The vestibule and semicircular canals may or may not be normal [291]. CHARGE syndrome is one of a handful of genetic conditions featuring this rather uncommon malformation (OMIM#214800).

We support the proposal of Monfort and colleagues [285] that the phenotype associated with the 8q12 duplication may be attributable to abnormal *CHD7* dosage and perhaps also to an effect on a nearby Duane anomaly locus. *CHD7* encodes a chromodomain helicase DNA-binding protein that is thought to have pivotal roles in early embryonic development by affecting chromatin structure and gene expression [42]. The visceral structures involved in these two individuals with 8q12 duplications are similar to those involved in CHARGE syndrome, which has been attributed to *CHD7* haploinsufficiency in half of affected individuals [199]. As there are no phenotypic differences between CHARGE patients with sequence mutations of *CHD7* and

patients with microdeletions involving neighboring genes, dosage changes of the genes neighboring *CHD7* in individuals with CHARGE are not thought to be consequential [42]. Phenotypic overlap resulting from both deletion and duplication of dosage sensitive genes has been observed previously in 22q11 deletion and duplication syndromes, for example, which both can variably feature cardiac defects, velopharyngeal insufficiency, and urogenital abnormalities [273].

An attempt was made to identify genomic sequences around the breakpoints that could predispose toward non-allelic homologous recombination. Reference genomic sequence (NCBI Build 36.1) corresponding to the maximally duplicated region, as well as sequence up to 500 kb upstream and downstream, was scanned for known segmental duplications (duplications of over >1 kb of non-repeat masked sequence with over 90% similarity, also known as low copy repeats (LCRs)). The breakpoints as recorded by the AGH (either the minimally duplicated region given by most outward affected probes, or the maximally duplicated region given by most outward non-affected probes) did not map to any LCR pairs. We also examined the breakpoints reported by Monfort et al. [285] and could identify no flanking LCR pairs.

The lack of recurrent breakpoints in these two cases and the absence of any LCR pairs that flank these *de novo* events do not support non-allelic homologous recombination as the mutation mechanism. Other mechanisms that have been suggested to produce non-recurrent rearrangements include non-homologous end joining and fork stalling and template switching [203, 250].

In conclusion, duplication of 8q12 region that includes *CHD7* produces an unusual and consistently characteristic multi-organ phenotype that includes hearing loss, congenital heart defects, intellectual disability, hypotonia in infancy, and Duane anomaly.

Chapter 7: Custom aCGH to investigate incidence of CNVs involving genes encoding epigenetic regulators pathogenic for ID

In the Introductory chapter to this dissertation I discussed the contribution of disruption of epigenetic regulatory genes to the etiology of ID and identified the following hypothesis:

Genetic imbalance involving genes regulating epigenetic processes causes 5% of ID.

In this chapter I will present the main project of my thesis, which was undertaken to test this hypothesis. We designed a custom microarray capable of screening in ultra-high resolution for CNVs of every known gene that encodes a factor involved in epigenetic regulation, viz., DNA methylation, histone modification or chromatin remodeling. This custom microarray was part of a larger project that also included testing for other candidate loci involved in ID. The larger project, which was led by Dr. Tracy Tucker, began in 2007 and ended in 2010. I was solely responsible for the selection of the epigenetic regulatory genes that were included on the array and for the validation of all *de novo* copy number variants identified at any of the candidate loci. In this chapter, I will only discuss in detail the aspects of this project that were relevant to my thesis. Although I will briefly describe elements of the design, methods and results pertaining to other candidate loci on the microarray to provide context for my findings, I will only elaborate on data relevant to my focus on genes encoding epigenetic regulators.

7.1 Study Design

The study included five main phases: a) bioinformatic selection of genes and design of probes, b) pilot phase NimbleGen 385K aCGH with validation on control samples from patients with known CNVs, c) selection of best performing probes and design of the final NimbleGen 12x135K array, d) aCGH of DNA from 177 children with idiopathic ID and both parents of each child on the final NimbleGen 12x135K array and bioinformatic analysis of results, d) Validation of *de novo* CNVs identified by aCGH using quantitative PCR, and e) genotype-phenotype correlation of validated *de novo* CNVs.

7.2 Material and methods

7.2.1 Patients and samples

For the pilot phase microarray we selected ten patients with *de novo* CNVs that had previously been ascertained and validated [7]. For the final phase study, we selected 177 trios, each comprised of a child with idiopathic ID and both normal parents. Patients with ID and at least one of the following additional characteristics were selected for study: 1) growth retardation of pre- and/or post-natal onset, 2) microcephaly or macrocephaly, 3) one or more major malformations, and 4) more than two facial dysmorphic features. The cause of the ID in each child was unknown despite full evaluation by a clinical geneticist, a karyotype at ≥ 500 band resolution and subtelomeric FISH studies. Some of these patients have also had clinical aCGH conducted; details are included in the discussion. DNA isolated from blood was used for the aCGH and qPCR validation. In a few instances DNA isolated from lymphoblastoid cell lines were used for the qPCR validation in absence of blood DNA. This study was approved by the University of British Columbia Clinical Research Ethics Board and informed consent was obtained from each family.

7.2.2 Selection of genes involved in epigenetic regulation

As a first step, I used the search terms 'DNA methylation', 'chromatin modeling/remodeling' and 'histone modification' to identify genes from the following four databases; a) Entrez Gene, b) Gene Cards, c) GeneOntology and d) OMIM. We then selected genes included in at least two of these databases and found 62 genes that fulfilled this criterion.

The second step was to select all members of gene families represented by these 62 genes. I used NCBI RefSeq curations, GeneCards, EMBL GeneHarvester and manual literature searches to identify other gene family members. I added another 135 genes by this process for a total of 197 genes. The regions selected for dense coverage on the custom microarray also included 100 kb flanking stretches upstream and downstream of the gene boundaries in order to capture possible regulatory sequence. We defined an exon or a regulatory gene-flanking sequence as a 'region'. The selection resulted in a total of 4725 regions that included all protein-coding segments and upstream and downstream segments of the 197 selected genes involved in epigenetic processes. Overall this accounted for 36% of the 13087 regions included on the custom chip.

7.2.3 Custom microarray design

Our goal was to design a custom NimbleGen 12-plex array with 135,000 probes covering suspected or known genes involved in the development of ID, with a minimum of eight probes per exon in each of our selected genes. In addition to all genes known to be involved in epigenetic regulation, the design also included other genes of interest: 1) Genes previously shown to cause ID, 2) All genes within reported microdeletion/microduplications (<100 Kb) in which the phenotype included ID, and 3) Genes involved in synaptogenesis. In addition, previously reported CNVs involving >100 Kb of genomic DNA were also included if the phenotype included ID. If there were candidate genes within these CNVs, probes were placed at higher density within these genes. Based on the reported number of genes with epigenetic regulatory function already known to cause ID (Table 1.1), and the instances where we have found epigenetic regulatory genes within large CNVs, we estimate approximately 20 genes belonging to the category of epigenetic regulation would also be included under the selection criteria of (1) and as representative genes for large CNVs selected to be probed. No overlap of genes is expected between genes categorized as involved in synaptogenesis and those that have epigenetic regulatory function.

The probes were designed according to NimbleGen's specifications, and the array design protocol was similar to a previously published protocol [292]. The start and end of each exon within each gene of interest was identified using Human Genome Build March 2006 (Hg 18). For exons <1200 bp in length, we included flanking regions on either side of the exon to accommodate our minimum of eight probes per exon. 75% of the regions of interest were 1200 bp or greater.

Once the regions of interest were selected, the sequence was retrieved from the UCSC Genome Browser and RepeatMasker [293] was used to identify repeat elements. Probes were generated using a series of Perl scripts. Specifically, a probe was required to have: (a) a length of 54 bp +/- 10 bp (b) a T_m of 76 °C +/- 4.5 °C using the Nearest Neighbour approach [294, 295], (c) less than 10% of its length representing repeat-masked sequence, (d) a free-energy of hairpin folding greater than -10.0 kcal/mol determined using MultiRNAFold (www.rnasoft.ca) [296], (e) a free-energy of dimerization greater than -26.0 kcal/mol, (f) low complexity sequence (e.g., simple tandem repeats, polypurine and AT-rich regions) occupying 10% or less of the

probe length, (g) no non-specific blast hits of 80% of the probe length or greater (using BLAST 2.2.15), (h) no more than one non-specific blast hit of 75% of the probe length or greater, (i) no more than four non-specific blast hits of 50% of the probe length or greater, and (j) no more than 178 cycles required for synthesis according to NimbleGen's cycle calculator.

The first phase design included 385,000 probes. The probes were selected by cycling through all target regions and selecting probes for each region until 99% of 385,000 probes were identified. At every cycle for each region, the 'best' probe was determined by considering its T_m and length as well as its distance from probes already selected within the region. The selection algorithm attempted to maximize the distance between probes, promote even coverage of each region, and minimize probe overlap. The remaining 1% of the array was filled by selecting random probes as negative controls that do not bind to any sequence in the human genome. The negative control probes, which should not show any specific hybridization on aCGH, were chosen to represent uniformly the range of T_m and length of all other probes selected. Manufacture of the arrays was conducted by NimbleGen using their proprietary technology for both the first phase and final design.

DNA samples from ten patients with known pathogenic CNVs were hybridized to the pilot phase array to determine sensitivity of the platform. An analysis of the variability in probe \log_2 ratios for each of the ten patients (excluding those regions with known pathogenic variants for that patient only) was performed. We then selected the 'best' 135,000 probes for inclusion on the final array by cycling through all target regions and selecting probes with the lowest \log_2 variability until 99% of the 135,000 probes were selected. The selection algorithm attempted to maximize the distance between probes, promote even coverage of each region, and minimize probe overlap. Figure 7.1 shows the location of all probes and coverage obtained by our design.

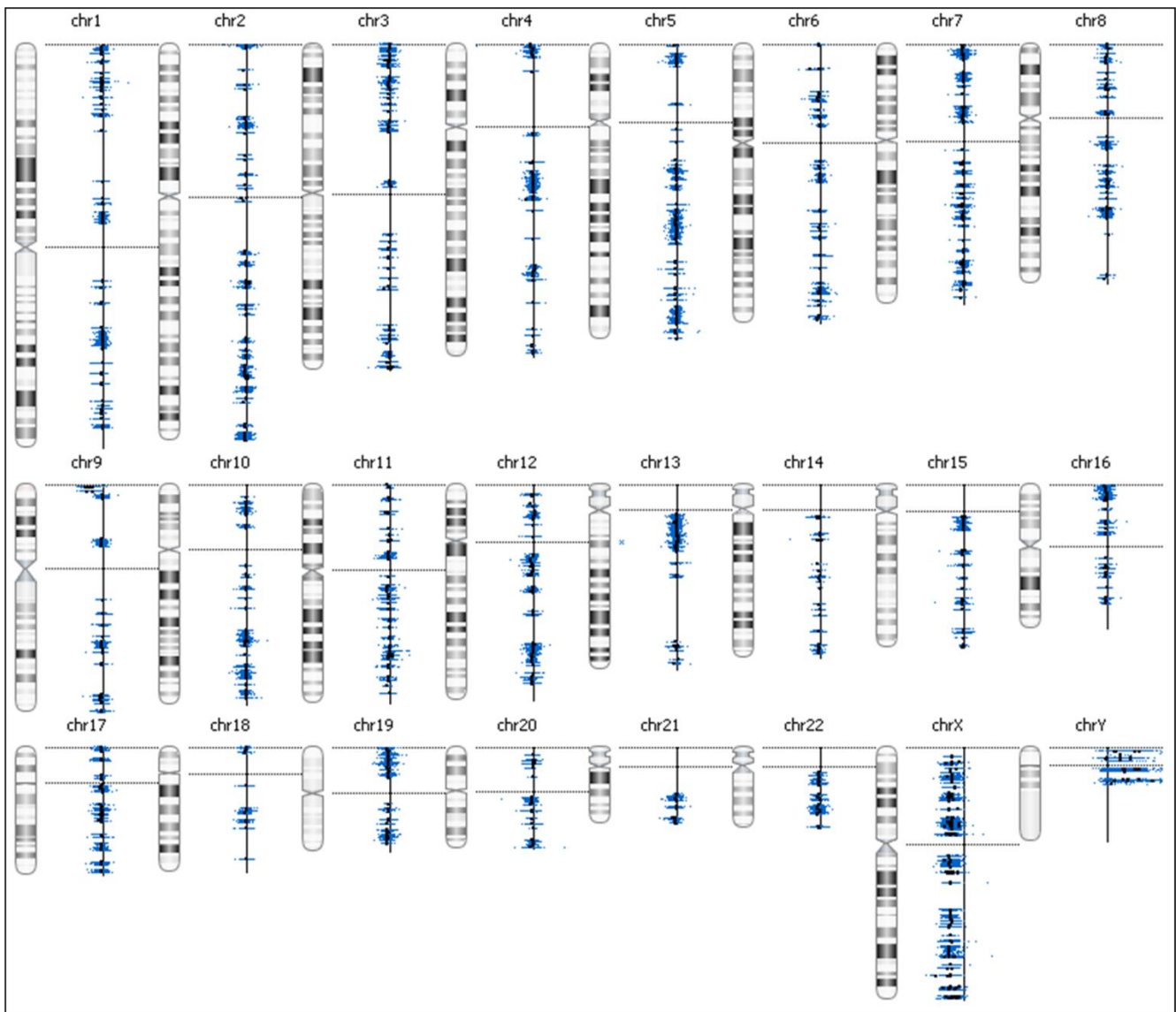


Figure 7.1: Showing location of probes against a chromosome ideogram. The blue dots clustered around the black vertical line represent individual probes with the vertical line representing the normal \log_2 ratio signal of 0. N.B. This patient has a deletion of chromosome 9pter as seen by a shift to the left from the normal (vertical line) of probes at that region. In addition, the hybridization in this case is of a male child versus the mother, as can be seen by a shift toward the left (signifying a loss) of all chromosome X probes and shift to the right (signifying a gain) of all chromosome Y probes.

7.2.4 aCGH

The labeling, hybridization and washing of the array was performed according to the manufacturer's specifications

(http://www.nimblegen.com/products/lit/NG_CGHCVN_Guide_v8p0.pdf). Briefly, 1 µg of high-quality genomic DNA was labeled with Cyanine-3 or Cyanine-5 using a random priming method. Two arrays were used for each trio, one in which the child's DNA was hybridized against the father's DNA, and another in which the child's DNA was hybridized against the mother's DNA. 20 µg each of the child's and one parent's DNA were labeled with opposite fluorochromes and hybridized to the array. After the hybridization was completed, the arrays were scanned at 5 µm resolution on a GenePix 4000AL Scanner. Unaveraged SegMNT files were generated using NimbleScan version 2.4.1 segMNT algorithm [297] with the default parameters, as follows: 1) the minimum segment difference, which represents the minimum difference in the \log_2 ratio that two segments must exhibit before they are identified as separate segments, was set at 0.0; 2) the minimum segment length, which represents the minimum number of consecutive probes that must exhibit a change in \log_2 ratio in order to call a segment, was set at 5; and 3) the acceptance percentile, which represents the stringency with which initial segment boundaries are selected, was set at 0.999.

CNVs containing five or more probes with the recommended default setting of a mean absolute \log_2 ratio of 0.2 or more were identified using BioDiscovery Nexus Software. The gender of each individual was input into the software, which corrects for the gender mismatch that occurs in the hybridization against one of the parents, allowing CNVs to be identified on the X chromosome.

7.2.5 Defining *de novo* CNVs

The analyses measured the strength of the hybridization signal obtained with the child's DNA in comparison to that of each of his or her parent's DNA. Each CNV call in which the child's signal was less than that of the parent (called a "deletion" in the tally) could actually represent either a loss of copy number in the child or a gain of copy number in the parent. Similarly, each CNV call in which the child's signal was greater than that of the parent (called a "duplication" in the tally) could actually represent either a gain of copy number in the child or a loss of copy number in the parent.

7.2.6 Validations of *de novo* CNVs

Primers were designed within ID candidate genes within the identified CNVs, avoiding benign polymorphisms listed in the DGV (version- variation.hg18.v10.nov.2010)(Table S8). Primers were designed using Primer Express (Applied Biosystems). The patient's DNA was diluted in PCR-grade water, and the quality and concentration was assessed using Nanodrop technology. Primers were optimized for qPCR by standard PCR amplification (50ng/μl sample DNA concentration) on a positive and negative control. PCR product was visualized on a 2% agarose gel stained with ethidium bromide. The presence of only a single band of the expected size in the control DNA, the absence of primer-dimers, and the absence of any amplification on the blank was considered indicative of a primer set that could be used for qPCR.

CNVs were validated by qPCR ($\Delta\Delta C_t$ method) using SYBR Green (Applied Biosystems). Testing was performed on an ABI7500 fast DNA sequencer using both parents and a pooled reference sample from Promega (Catalogue#: G3041 – male and female and G1521-female only) and the hexose-6-phosphate dehydrogenase (*H6PD*) gene as a locus control. CNVs on the X chromosome were validated against a pooled reference sample composed of female only DNA, while those on autosomes were validated against a pooled reference sample composed of both male and female DNA. All primer sequences can be found in appendix D.

7.2.7 Genotype-Phenotype correlations analyses

Factors taken into consideration when deducing genotype-phenotype correlations are discussed in section 1.7 of this thesis and the study methodology is included in appendix C. The following resources were utilized to investigate the effect on genotype-phenotype for those CNVs involving intragenic regions. Gene functional information was obtained from the Universal Protein Database, UniProt (<http://www.uniprot.org/uniprot/>). Entrez Gene (<http://www.ncbi.nlm.nih.gov/gene/>), the OMIM database (<http://www.ncbi.nlm.nih.gov/omim/>) and the Jackson lab mouse informatics database (<http://www.informatics.jax.org/>) were accessed to mine gene functional, disease association and mouse model information. The Human Gene Nomenclature Committee (<http://www.genenames.org/index.html>) database was used to obtain official gene names. Involvement of CNVs that are known to be normal polymorphisms of disease-causing variants was checked by using the UCSC genome browser (<http://genome.ucsc.edu/> build NCBI36/hg18) and accessing the Database of Genome

Variants (<http://projects.tcag.ca/variation/>) and DECIPHER database (<http://decipher.sanger.ac.uk/>) tracks, as well as tracks reporting structural variations.

Effects of CNVs involving a single or few exons were analyzed in the following manner:

A. Assessing the probable effect on protein products

We determined if single exon deletions/duplications caused frame-shifts using the Graphics view of the GenBank entry corresponding to the affected gene, accessed via the Nucleotide database of the NCBI suite (<http://www.ncbi.nlm.nih.gov/nuccore/>). The analysis included consideration of alternative splice forms where these were documented by GenBank. For instances in which a frame-shift was noted to occur, the normal protein and the mutant protein predicted by deletion and/or duplication of the exons involved in the observed CNV were compared after *in silico* translation of the complete coding sequence in the correct reading frame using ExPASy (<http://www.expasy.ch/tools/dna.html>). As a further check, we used SMART (<http://smart.embl-heidelberg.de/>) to compare the mutant translation product to the original by submitting the translated coding sequences and visualizing the domain structure of the proteins produced. If the mutant coding sequence included open reading frames of over 10 amino acids that could be translated from either of the other two 5' to 3' reading frames, we included these in the SMART analysis as well.

For CNVs that affected one or more exons within a gene, we also used the commercial software Alamut (Interactive Biosoftware. Version 1.5) to check for frame-shift mutations in addition to the manual analysis described above. We used I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>), an automated protein structure and function prediction algorithm, to calculate 3D protein folding models [298, 299] when possible. The I-TASSER program accepts sequences between 10 and 1500 amino acids in length. We input the amino acid sequence modeled to include the CNV effect (e.g., removal of exon 5 in the case of *JARID2* exon 5 deletion CNV) and submitted these sequences without any restraints, i.e., with default settings, for analysis.

B. Assessing whether transcription is likely to occur

In cases where exon 1 was involved in a CNV, we checked to see if the known promoter for the gene was also included. We entered the sequence of the maximal affected genomic region into FirstEF (<http://rulai.cshl.org/tools/FirstEF/>), which calculates the likelihood of a promoter sequence being present and gives the most probable locations of the promoter and first exon.

7.3 Results

We performed aCGH on 177 trios, each comprised of a child with idiopathic ID and both of his or her unaffected parents, using a custom NimbleGen microarray containing 135,000 probes selected for possible involvement in ID. CNVs were called by analysing the CGH data using Nexus® software. A CNV was considered to be *de novo* if it was identified in both the hybridization of the child versus the mother and the child versus the father. Inherited variants (CNVs that were seen on a hybridization of the child versus one parent but not the other) were analyzed manually and where there was possibility for the variant to be *de novo*, i.e. an apparent shift of the probe log₂ ratio in the same direction observed by eye, though not called by the software, in comparison to the other parent, these were included as well.

We identified 176 *de novo* CNVs in total, of which 35 CNVs were validated to be *de novo* by qPCR. The confirmed CNVs are shown in Table 7.1. The ~25% validation rate we achieved reflects our use of software settings intended to maximize sensitivity and is in keeping with other reported studies [300]. Sixty seven of the *de novo* CNVs called by the software involved epigenetic regulatory genes, and 16 of these (in 15 patients) were confirmed by qPCR. Therefore, a total of 8.5% of the 177 ID trios studied had a validated CNV involving an epigenetic regulatory gene. We confirmed nine loss and seven gain CNVs in these 15 trios (Table 7.1). Two of these individuals, patient 32858 and patient 33739, each exhibited two confirmed CNVs, but only patient 32858 had both CNVs that included epigenetic regulatory candidates. In two instances (patient 32861 and patient 32858), a whole gene was involved and the affected region may also include adjacent genes. Figure 7.2 over views these results.

Of the 12 CNVs that involved only a single exon, seven found in seven different individuals involved only exon 5 of *JARID2*. Two of these CNVs were copy number gains, and five were copy number losses. Five other CNVs involved either only exon 1 or exon 1 and the adjacent few exons; these cases were analysed for involvement of the promoter, as discussed below. The smallest CNV involving an epigenetic regulatory gene was only 100 to ~1600 bp in length (patient 31823). The largest CNV involving an epigenetic regulatory gene was 178kb to 6600 kb long (patient 32861). The imprecision of these estimates reflects the uneven distribution of probes on the custom array.

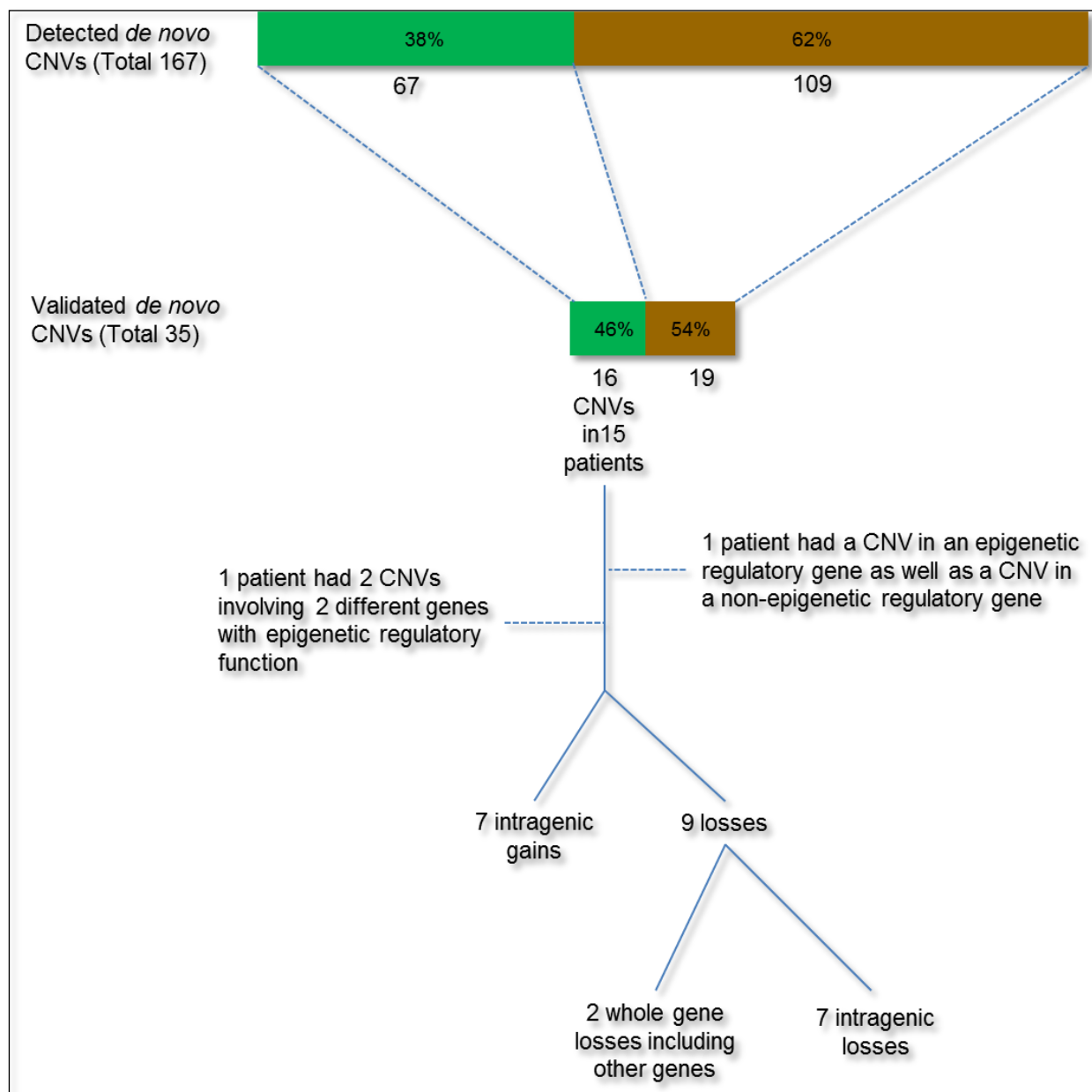


Figure 7.2: Overview of AGH results; The colored bar graphs total detected (top most bar) and validated (second bar) *de novo* CNVs with CNVs involving genes with epigenetic regulatory function depicted in green color CNVs involving other candidate genes included in the design depicted in brown. The number of calls is given under the relevant section of the bar graph. *De novo* validated CNVs involving genes with epigenetic function are further categorized.

Table 7.1: Summary of validated *de novo* CNVs involving epigenetic regulatory genes. The minimal size of the CNV is defined as the region between the locations of the first and last probes with an abnormal copy number. Because the custom microarray used in this study provides discontinuous coverage, the breakpoints of the CNV cannot be precisely determined but are likely to lie between the locations of the last normal and first abnormal probes proximally and between the last abnormal and first normal probes distally. 'Chr' = chromosome.

Trio ID	Gene	Loss/ Gain	Chr	Exon	Last Normal Probe Location (proximal)	First Abnormal Probe Location (proximal)	Last Abnormal Probe Location (distal)	First Normal Probe Location (distal)	Minimum Size of CNV	Maximum Size of CNV
32861	SMARCA2	1 copy loss	9	whole gene	0	2,005,341	2,183,623	6,649,166	178,282	6,649,166
32858	MEF2C	1 copy loss	5	whole gene	87,100,982	88,162,900	88,163,700	88,257,129	800	1,156,147
33459	CHD6	1 copy loss	20	1	39,613,550	39,680,100	39,680,750	39,688,521	650	74,971
33739	CHD7	6 copy gain	8	1	61,748,366	61,753,950	61,754,150	61,816,132	200	67,766
33739	ARID1B	1 copy gain	6	1	157,133,828	157,141,000	157,142,500	157,192,105	1,500	58,277
30848	ARID2	6 copy gain	12	1,2,3	44,403,154	44,409,000	44,412,000	44,491,454	3,000	88,300
33665	ARID4B	1 copy gain	1	1 & 2	233,556,621	233,556,700	233,557,700	233,559,503	1,000	69,551
31823	JMJD1A	1 copy loss	2	6	86,535,860	86,537,100	86,537,200	86,537,443	100	1,583
33806	JMJD1C	1 copy gain	10	4	64,647,390	64,694,400	64,694,700	64,721,535	300	74,145
31916	JARID2	1 copy loss	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836
32094	JARID2	1 copy loss	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836
32146	JARID2	1 copy loss	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836
32858	JARID2	1 copy loss	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836
33650	JARID2	1 copy gain	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836
33723	JARID2	1 copy loss	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836
34335	JARID2	1 copy gain	6	5	15,568,665	15,576,500	15,577,250	15,595,501	750	26,836

Table 7.2: Genotype-Phenotype correlations summary for 16 validated CNVs involving epigenetic regulatory genes

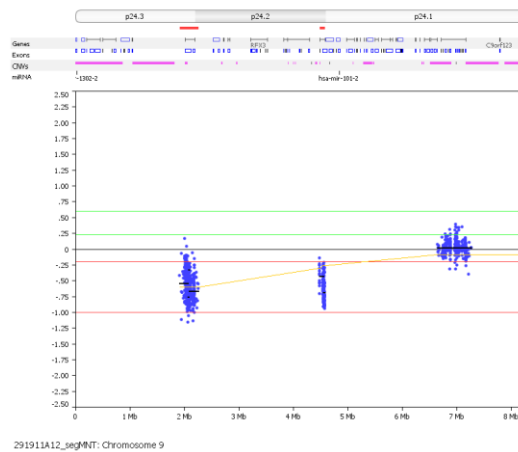
Proband ID and CNV	Gene (Uniprot ID)	Protein Function	Expression	Expected effect of CNV on protein	Known Disorders/reports /comment	Patient Phenotype	DGV entries mapping to the named gene	Interpretation
32861 loss	<i>SMARCA2</i> (P51531). Probable global transcription activator SNF2L2	Belongs to the npBAF and the nBAF complex. Involved in vitamin D-coupled transcription regulation via its association with the WINAC complex, a chromatin-remodeling complex recruited by vitamin D receptor.	Ubiquitous	Whole gene is deleted.	14 DECIPHER reports that include the gene (4 gains and 10 losses); all are very large and span many other genes.	ID	DGV #53551, 95804, 82867 involve a few exons. Two other small variants do not involve exons.	Good candidate for pathogenicity
32858 loss	<i>MEF2C</i> (Q06413) Myocyte-specific enhancer factor 2C	Forms a complex with class II HDACs in undifferentiating cells. May interact with HDAC7 and CARM1 (By similarity). Interacts with HDAC4, HDAC7 AND HDAC9; the interaction with HDACs represses transcriptional activity.	Expression is highest during the early stages of postnatal development and moderate later on.	Whole gene is deleted.	Candidate gene for recently described 5q14.3 microdeletion ID syndrome (severe ID, epilepsy, stereotypic movements)	ID, spasticity, epilepsy, acquired microcephaly	10 small variants, all located within intronic sequence, except for variant # 93139, which involves the final exon.	Good candidate for pathogenicity
33459 loss	<i>CHD6</i> (Q8TD26) Chromodomain-helicase-DNA-binding protein 6	Probable transcription regulator.	Ubiquitous Highest in brain	Maximal deleted region includes promoter. No transcription expected.	Homozygous null mice display impaired coordination that is not due to muscle weakness or bradykinesia (MGI)	ID, autism, epilepsy	Two small variants mapped to introns.	Good candidate for pathogenicity

Proband ID and CNV	Gene (Uniprot ID)	Protein Function	Expression	Expected effect of CNV on protein	Known Disorders/reports /comment	Patient Phenotype	DGV entries mapping to the named gene	Interpretation
33739 gain	<i>CHD7</i> (Q9P2D1) Chromodomain-helicase-DNA-binding protein 7	Probable transcription regulator. May interact with CTCF	Widely expressed in fetal and adult tissues	Effect of 6 copy gain unpredictable	Loss of function causes CHARGE syndrome. Candidate gene for recently described 8q12 micro-duplication syndrome (ID, Duane anomaly, CHD)	ID, cataract	One small variant mapping to intron of <i>CHD7</i>	Good candidate for pathogenicity
33739 gain	<i>ARID1B</i> (Q8NFD5) AT rich interactive domain 1B (SWI1-like)	Belongs to the npBAF and the nBAF complex. The npBAF complex is essential for self-renewal/proliferative capacity of multipotent neural stem cells. The nBAF complex along plays a role regulating the activity of genes essential for dendrite growth	Widely expressed with high levels in brain, heart, skeletal muscle and kidney	No frame-shift if the duplication is tandem; an extra coiled-coil motif is predicted.	10 DECIPHER reports that include the gene (3 gains and 8 losses), all except #250455 (deletion) are very large and span many other genes.	ID, cataract	DGV#111727 gain variant includes only exon1 of <i>ARID1B</i>	Benign or possibly contributory
30848 gain	<i>ARID2</i> (Q68CP9) AT-rich interactive domain-containing protein 2	Required for the stability of the chromatin remodeling complex SWI/SNF-B (PBAF). May be involved in targeting the complex to different genes.	Highly expressed in heart and testis. Some expression in brain	Effect of 6 copy gain unpredictable. Single tandem duplication causes frame-shift and premature STOP	One DECIPHER case (#254078), ~10 Mb duplication including many more genes	ID, facial dysmorphism, short stature, unilateral deafness	DGV # 3887 gain variant begins before exon 3 and extends beyond gene.	Good candidate for pathogenicity

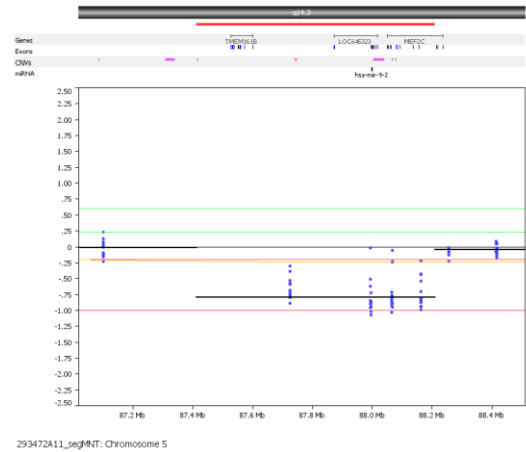
Proband ID and CNV	Gene (Uniprot ID)	Protein Function	Expression	Expected effect of CNV on protein	Known Disorders/reports /comment	Patient Phenotype	DGV entries mapping to the named gene	Interpretation
33665 gain	<i>ARID4B</i> (Q4LE39) AT-rich interactive domain-containing protein 4B	A subunit of the histone deacetylase-dependant SIN3A transcriptional corepressor complex, which functions in diverse cellular processes such as proliferation, differentiation and apoptosis	Ubiquitous	No frame-shift with tandem duplication of exons 1 and 2; an extra coiled-coil motif is predicted.	Hypomorph alters PW/AGS imprinting center regulation	ID, autism	DGV #48278 gain variant, begins upstream of gene and includes up to exon 7 of gene.	Possibly pathogenic
31823 loss	<i>JMJD1A/KDM3A</i> (Q9Y4C1) Lysine-specific demethylase 3A	Histone demethylase that specifically demethylates 'Lys-9' of histone H3. Preferentially demethylates mono- and dimethylated H3 'Lys-9' residue	Ubiquitous	Manual curation and Alamut result: Frame-shift causes premature STOP results in loss of all functional domains.	Male mice hypomorphic and nullomorph display impaired spermatogenesis. Nullomorph mice also display obesity	ID	None.	Good candidate for pathogenicity
33806 gain	<i>JMJD1C</i> (Q15652) jumonji domain containing 1C	Probable histone demethylase that specifically demethylates 'Lys-9' of histone H3. May be involved in hormone-dependent transcriptional activation by participating in recruitment to androgen-receptor target genes	Ubiquitous	Manual curation: no frame-shift, no additional motifs added with tandem duplication of exon 4. Alamut curation: frame-shift and premature STOP. Discrepancy of exon size from source curation.	DECIPHER entries #242960 (~7 Mb loss, ID, hypotonia)	ID	None	Possibly contributory

Proband ID and CNV	Gene (Uniprot ID)	Protein Function	Expression	Expected effect of CNV on protein	Known Disorders/reports/comment	Patient Phenotype	DGV entries mapping to the named gene	Interpretation
31916 Loss	JARID2 (Q92833) Jumonji/ARID domain-containing protein 2	Regulator of histone methyltransferase complexes that play an essential role in embryonic development, including heart and liver development, neural tube fusion process and hematopoiesis	During embryogenesis, predominantly expressed in neurons and particularly in dorsal root ganglion cells.	Deletion of exon 5 or addition of a tandemly duplicated exon 5 does not cause a frame-shift. Exon 5 codes a low complexity motif without known function	Mouse model: Jmj-deficient mice show several morphologic abnormalities, including neural tube and cardiac defects, and die <i>in utero</i> around embryonic day 11.5. Overexpression of Jmj may inhibit cell growth	ID, short stature, ASD	None	Possibly contributory
32094 loss						ID, CL/P, deafness, facial dysmorphisms		Possibly contributory
32146 loss						ID, overgrowth		Possibly contributory
32858 loss						ID, spasticity, epilepsy, acquired microcephaly		Possibly contributory
33650 gain						ID, autism		Possibly contributory
33723 loss						ID with ASD		Possibly contributory
34335 gain						ID with minor facial dysmorphisms		Possibly contributory

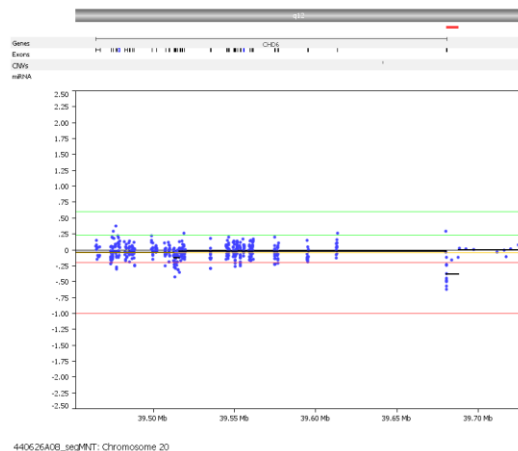
Figure 7.3: Nexus images of validated CNVs involving epigenetic regulatory genes. Each box includes the following panels: topmost – schematic of chromosome (ideogram) with red bars depicting area called as a deletion and green bars depicting area called as a duplication. This is followed by tracks showing genes ('Gene'), exons ('Exon'), normal variants as reported in the DGV ('CNV') and micro RNA transcripts ('miRNA'). The genes, exons and IDNAs are derived from the NCBI reference databases as accessed by the Nexus software. The boxed area is a view of the aCGH probe log₂ ratios (Y axis) versus genomic co-ordinates in Mb (X axis) of the region represented in the ideogram. The black horizontal line within the box indicates a log₂ ratio of 0, i.e., the normal signal. The red and green horizontal lines indicate default cut-off values for a one-copy loss (-0.2), two-copy loss (-0.1), one-copy gain (0.2) and two-copy gain (0.56). The yellow line indicates the moving average for the probes, with each probe's signal depicted by a blue dot.



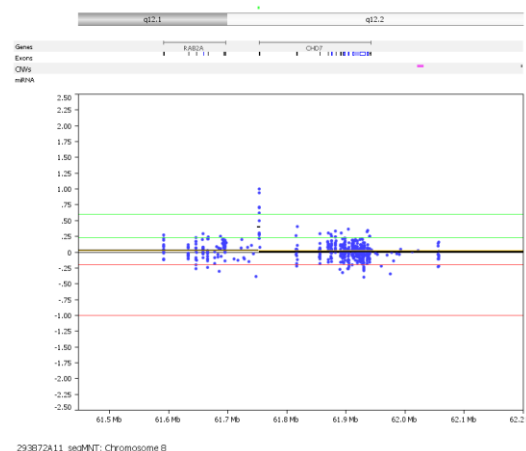
A) *SMARCA2* deletion in patient 32861



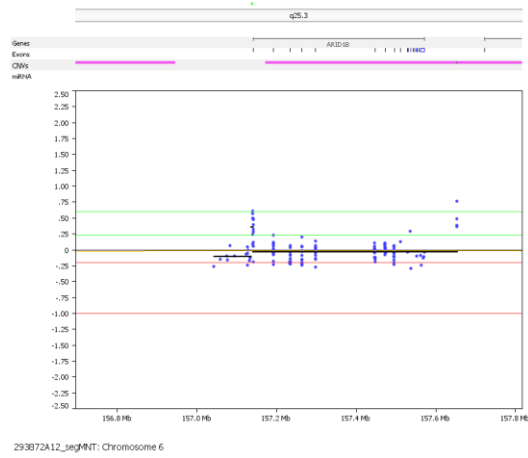
B) *MEF2C* deletion in patient 32858



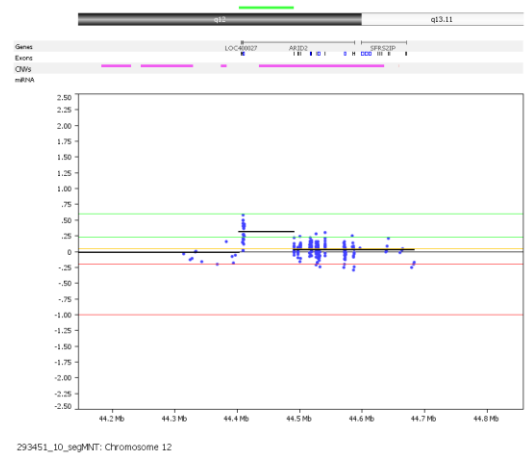
C) *CHD6* exon 1 deletion in patient 33459



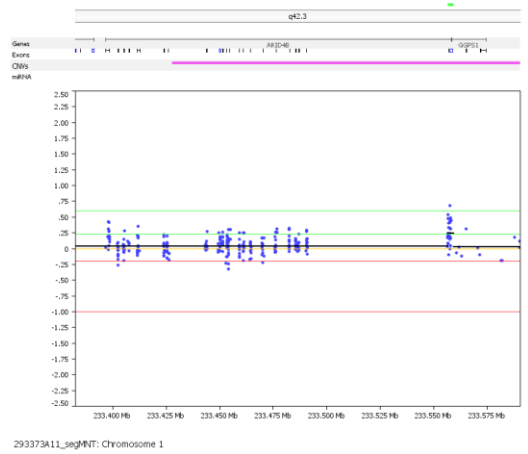
D) *CHD7* exon 1 gain in patient 33739



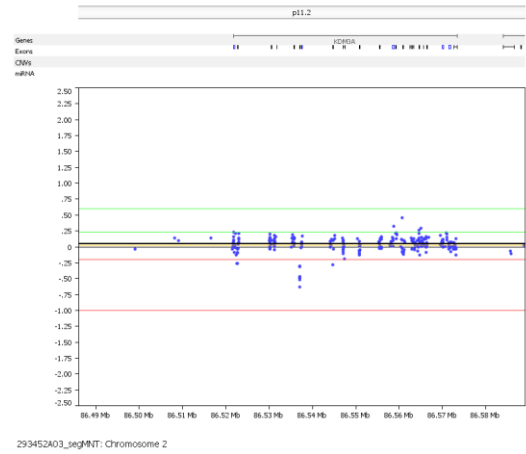
E) *ARID1B* exon 1 gain in patient 33739



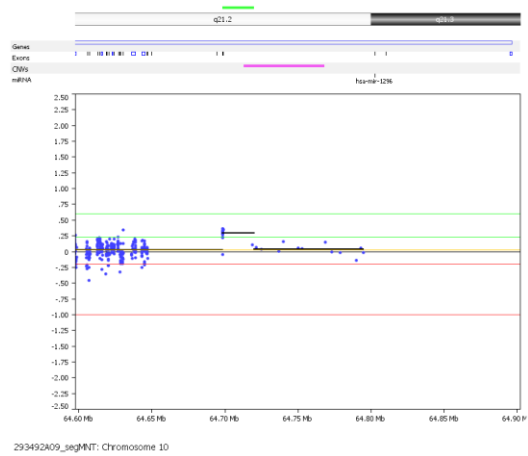
F) *ARID2* exons 1 to 3 gain in patient 30848



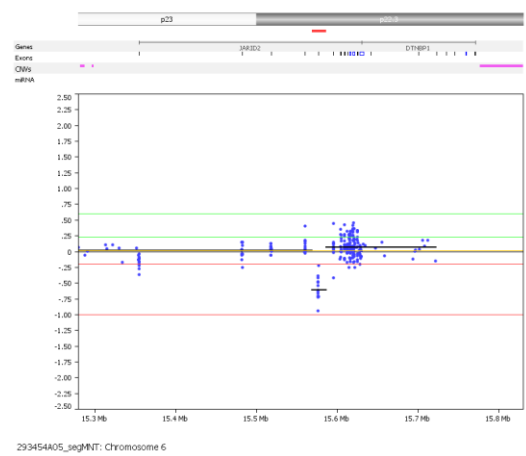
G) *ARID4B* exons 1 and 2 gain in patient 33665



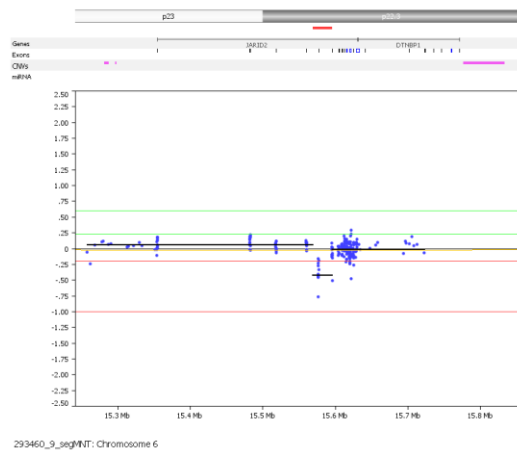
H) *JMJD1A/KDM3A* exon 6 loss in patient 31823



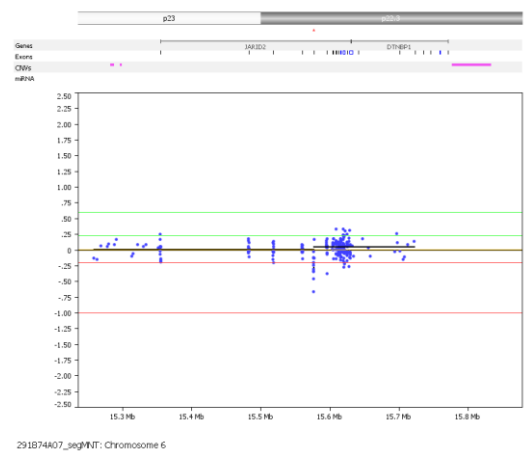
I) *JMJD1C* exon 4 gain in patient 33806



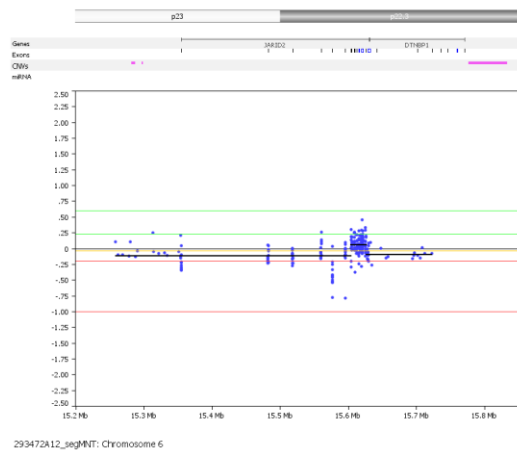
J) *JARID2* exon 5 loss in patient 31916



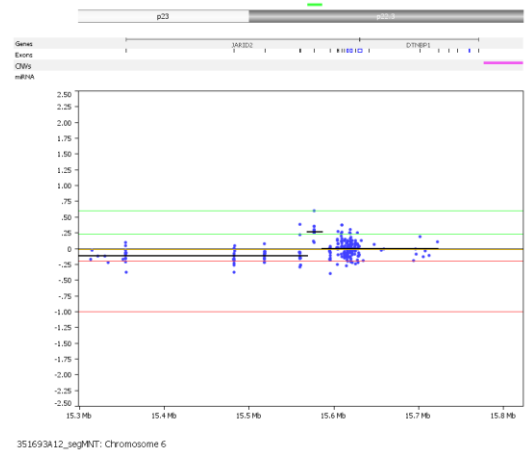
K) *JARID2* exon 5 loss in patient 32094



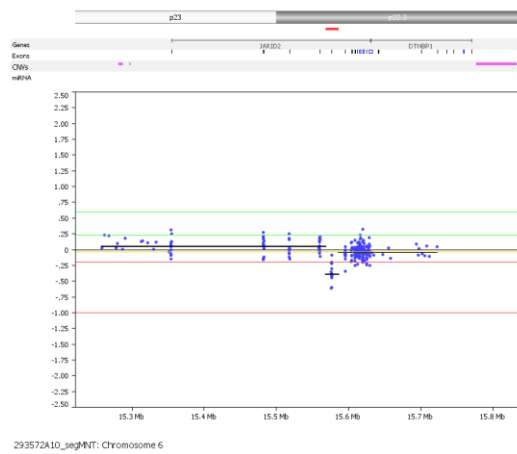
L) *JARID2* exon 5 loss in patient 32146



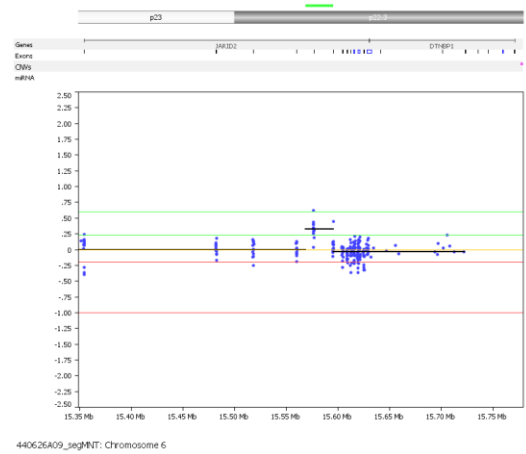
M) *JARID2* exon 5 loss in patient 32858



N) *JARID2* exon 5 gain in patient 33650



O) *JARID2* exon 5 loss in patient 33723



P) *JARID2* exon 5 gain in patient 34335

7.4 Genotype Phenotype correlations

Table 7.2 provides a summary of the expected effect of the confirmed CNVs listed in Table 7.1 on the genes involved. We discuss each case in more depth below.

7.4.1 CNVs affecting the whole gene

7.4.1.1 *SMARCA2* loss

SMARCA2 is involved in a large (0.2-6.6 Mb) telomeric deletion of the short arm of chromosome 9 in patient 32861 (Figure 7.3A). Clinical whole genome aCGH using a 135,000 oligonucleotide array with probes evenly spaced across the genome confirmed this deletion as a 5Mb deletion of 9p24 which includes our minimally affected region. In addition, the patient was also shown to carry a *de novo* 3.5 Mb duplication of chromosome 16q24.1 on the clinical aCGH. Both events were confirmed by FISH. It is unclear if both CNVs are independent events or due to a translocation as karyotyping is unavailable. Nonetheless the large size of both CNVs precludes an accurate genotype-phenotype correlation as the maximally affected regions contain many genes as well as many records for pathogenic CNVs from the DECIPHER database.

SMARCA2, also known as *BRM*, is a key member of the highly conserved ATP-dependent chromatin remodelling SWI/SNF complex. The SWI/SNF complex is the best studied of three recognized ATP-dependant chromatin remodelling complexes present in mammals and, among many functions, is important in embryonic development [301]. Furthermore, *SMARCA2* is also a key member of the neuronal progenitor-specific chromatin remodelling sub-complex of the SWI/SNF complex, nBAF [302]. *SMARCA2* has been reported to be involved in several neurofunctional disorders. Notably, it has been reported to be associated with schizophrenia by both genome wide association analysis [303] and as a result of rare structural variants that disrupt the gene [304]. Others have demonstrated a key interaction between *SMARCA2* and *MECP2* (causative gene for Rett syndrome) in transcriptional silencing [305, 306], implicating the *SMARCA2* protein product in the same disease pathway. In another study, Ehrlich et al. show that *SMARCA2* is up-regulated in patients with ICF syndrome in comparison to controls[307]. ICF syndrome, a rare immunodeficiency disease, is caused by mutations to

DNMT3B, a DNA methyltransferase. The mouse ortholog of *SMARCA2* has been shown to be a modifier for epilepsy in animal studies [308].

In summary, the important role of the gene product supports the notion that *SMARCA2* disruption could contribute to the pathogenicity in this case. The large size of both events indicates a strong likelihood for the CNVs to be pathogenic, though a more definitive genotype-phenotype correlation is precluded.

7.4.1.2 *MEF2C* loss

MEF2C was completely involved in a deletion that also includes only one other gene, a pseudogene (hypothetical protein encoding gene LOC645323) in patient 32858 (Figure 7.3B). Karyotype at the 500 band level was normal, but since enrolling in this research study the patient was tested using clinical aCGH (135,000 oligonucleotide whole genome array) and was found to have a 1 Mb deletion of chromosome 5q14.3 mapping to genomic coordinates 87134779bp-88171786bp (NCBI build 36). These co-ordinates map to the maximally affected region identified by our targeted array. This patient also has a deletion of exon 5 of *JARID2*, which is discussed below (section 7.4.3).

Our patient, a girl, is of French Canadian ethnicity. Her family history is unremarkable. She had a birth weight of 2.86kg (25th centile) with unremarkable perinatal history. At 5 months of age she presented with poor visual contact and axial hypotonia. An ophthalmologic evaluation suggested delayed visual maturity, and an EEG showed multifocal epileptic activity. She had a few generalized seizures between ages 1 and 2 years but has since been well controlled with valproic acid.

At age 10 years and 5 months old, she was able to walk with help and stand without support. She was unable to feed herself. She did not respond to verbal orders and could not designate body parts. No dysmorphic features were noted on physical examination. Quadriparesis with increased reflexes was noted, and she had decreased mobility of the left upper arm.

At her most recent assessment at age 13 years and 11 months old, she was able to walk short distances without assistance, although her gait was unstable. She did not respond to verbal orders, and did not speak or use sign language, although she did utter syllables. She did feed herself. Her weight was 41.2 kg (10th-25th centile), height was 152 cm (10th centile) and head circumference was 50.5 cm (<3rd centile). No dysmorphic features were noted. There was mild spastic quadriparesia that was more severe on the right side. Head CT scan and audiogram were both normal. She had strabismus that was corrected surgically.

MEF2C has been reported to be causative for a recently defined 5q14.3 microdeletion syndrome [309] characterized by severe ID, stereotypic movements and epilepsy and/or cerebral malformation. Le Meur *et al* [309] described five unrelated patients: four with complete *de novo* deletions of *MEF2C* and one with a partial *de novo* deletion of the gene. These authors also reported an additional patient with a truncating mutation in *MEF2C*. Poor visual contact, absent speech, inability to walk independently and severe ID were common to all of these patients and also occurred in our patient. Microcephaly, acquired in our patient, was reported for 3 of the previously reported patients. Importantly, generalized seizures were reported in 4 of the patients in the Le Meur study and also occurred in our patient. Le Meur *et al.* did not describe a characteristic facial phenotype, although mild dysmorphisms were noted in some of their patients.

Thus, severe developmental delay, hypotonia, absent speech, and epileptic seizures characterize this syndrome, which is caused by *MEF2C* loss of function. The fact that our patient shares the same phenotype and the same genetic defect provides strong evidence for pathogenicity of the 5q14.3 microdeletion in patient 32858.

MEF2C encodes a protein that belongs to the MADS (MCM1, agamous, deficiens, SRF) family of transcription factors, an evolutionarily ancient family. In vertebrates four genes, *Mef2a*, *b*, *c* and *d* are known. The function of *MEF2* involves the recruitment of and cooperation with a multitude of other factors, in particular with class II histone deacetylases [310]. A clear connection has been established between *MEF2* and *HDAC* activity during development: in many cell types, differentiation appears to be dictated by a balance between the *MEF2* proteins as transcription activating factors and *HDACs* as transcription repressors of a number

of downstream targets [310]. In mice, expression of the homolog of human *MEF2C*, *Mef2c*, is highly enriched in skeletal muscle, cardiac muscle and neuronal tissues [311, 312]. Moreover, an intricate protein-protein interaction has been shown for MEF2 and various HDACs [313-315]. For example, in transgenic mice, it has been shown that *MEF2C* activates expression of *Hdac9*, which in turn modulates *MEF2C* activity in a negative feedback loop [316].

In humans *MEF2C* has been shown to be highly expressed in fetal and adult brain, and a critical role for the protein in postmitotic neuronal differentiation has been suggested [317]. *MEF2C* has been shown to have a key function in synaptogenesis in mouse models [312], and the protein has been implicated in causing seizures in humans [309]. Le Meur *et al.* [309] postulate a shared biological pathway for *MECP2* (an epigenetic regulator that is causative of Rett syndrome) and *MEF2C*, based on presence of putative binding sites between the proteins. Taken together these data suggest a key role in epigenetic regulation for *MEF2C*. Our finding corroborates the existence of a characteristic ID syndrome caused by disruptions of *MEF2C*, as first reported by Le Meur *et al.* [309].

7.4.2 CNVs affecting exon1 (and the promoter) but not the whole gene

7.4.2.1 Copy number losses

7.4.2.1.1 *CHD6* exon 1 loss

CHD6 was involved in a single exon copy number loss in patient 33459 (Figure 7.3C). Testing by four other whole genome aCGH platforms returned a normal result [318]. This exon is completely contained within the minimal affected region of the CNV, and, although the maximally affected region is ~75 kb, exon 1 of *CHD6* is the only protein-coding segment involved. qPCR validation confirmed that exon 1 was indeed deleted. The upstream breakpoint appears to be close to the start of exon 1, but the downstream breakpoint is uncertain and could lie anywhere in the large first intron. Exon 2 is not deleted (Figure 7.3C). Schuster *et al.*, who first identified this gene (originally named *CHD5*), predicted the promoter sequence to be 120 bp upstream of the start of exon 1 [319]. This predicted promoter region is also included in the minimally affected region of this deletion. We entered the maximally affected sequence into FirstEF, a promoter prediction algorithm, and also found that the promoter is likely to be involved in the deletion. The FirstEF program predicted the promoter to lie in a sequence that

included the complete first exon and extended 500 bp upstream. These data strongly support complete loss of transcription of the affected allele of *CHD6*.

Our patient is a 6-year-old girl of North African origin. Her parents are consanguineous. She had an unremarkable perinatal history. She walked at 18 months of age and at 2 years of age was able to say a few words. However, she regressed thereafter. At 2 years of age she had a few generalized seizures and was diagnosed with epilepsy. An EEG showed epileptic activity in the left hemisphere of the brain. Her seizures were well controlled with valproic acid.

At 3 years 6 months old, she did not utter any words or syllables and was unable to respond to verbal orders. She showed no visual contact or interactive play and was not able to eat with a spoon. Assessment on the Griffiths Developmental Scale showed performance at the 8-month level. She was diagnosed with autism by a developmental paediatrician, although formal assessment was not conducted. At this age, her weight was 14.7 kg (25th – 50th centile), her height was 96.9 cm (25th centile) and her head circumference was 49.5 cm (50th centile). No craniofacial dysmorphic features were noted. Her extremities were normal except for clinodactyly of the right 5th finger and ligamentous laxity.

At 6 years of age she still had no words, although she could understand a few signs. She could not draw a circle but could scribble and could eat with a spoon. Her weight was 20.4 kg (50th centile), height 111.4 cm (>25th centile) and head circumference was 52 cm (50th centile). She had an unremarkable neurological examination, a normal audiogram, normal brain MRI and normal brain spectroscopy. Testing for fragile X syndrome, amino acids, organic acids, lactic acid, ammonia, and transferrin glycosylation were all normal. aCGH by Signature Genomics using a 4685 BAC targeted clinical array was normal, but this platform does not include a probe for *CHD6*.

Yamada *et al.* recently reported a patient with a *de novo* translocation that disrupted only *CHD6* [320]. Their patient, a female, presented with severe ID and brachydactyly of the toes. A published photograph and X-ray image of the hand show clinodactyly of the 5th digit (Figure 1.D

from Yamada *et al.* [320]). They report a normal brain MRI and EEG and note minor facial dysmorphisms.

Yamada and associates demonstrated 50% expression of *CHD6* mRNA in lymphoblastoid cell lines from their patient and concluded that the translocation causes reduced expression of *CHD6* as a result of disruption of the coding sequence. These authors posit that this disruption of *CHD6* is causative for the abnormal phenotype in their patient.

We found one other report for a *CHD6* mutation, again caused by a translocation [321]. This patient, also a girl, was mildly to moderately delayed, with normal height, weight and head circumference. She crawled at 16 months and stood at 17 months of age. At 15 years old, she was able to speak in sentences, read and write simple text, and understand simple tasks. She did not have epilepsy. On physical examination, she had a broad square face, hypertelorism, flat nasal bridge, prominent ears and short neck. Her left leg was longer than the right, and she had bilateral single palmar creases.

The translocation breakpoint in this patient interrupts both *CHD6* and *TCF4*. Mutations for *TCF4* have been shown to cause Pitt-Hopkins syndrome, defined by severe ID, epilepsy, growth retardation, and distinctive facial features [322, 323]. This patient does not have classical Pitt-Hopkins syndrome, but phenotypic comparison with our patient is complicated by the involvement of *TCF4* with the transcription of a *TCF4-CHD6* fusion product [321].

We did not find any patients in the DECIPHER database with *CHD6* affected, and polymorphisms of this gene are not reported in DGV, a repository of benign copy number variation.

CHD6 is a member of the CHD family of chromatin remodelling proteins, which are known to play an important role in gene regulation and have an essential role in normal neural development and brain function [324, 325]. The nine member family is divided into three subfamilies. Of these *CHD6-CHD9* make up subfamily III, which encode the largest CHD proteins. Other members of this subfamily are known or strongly suspected to be disease causing: deletions and inactivating mutations of *CHD7* cause CHARGE syndrome [42], and we also

have found that duplications of *CHD7* lead to a distinctive ID syndrome (*vide* chapter 6). *CHD8* is one of two proposed candidate genes for the novel 14q11.2 microdeletion ID syndrome we identified (*vide* chapter 2).

CHD6 is expressed ubiquitously, with highest levels in the brain [325]. Yamada et al. show that reduced expression of *CHD6* in HeLa cells causes extensive misalignment of chromosomes, and, importantly, *CHD7* depletion causes a similar effect [320]. Furthermore, the same authors show an increased frequency of aneuploidy in both *CHD6*- and *CHD7*-mutation bearing lymphoblastoid cell lines [320]. Mice homozygous null for exon 12 of *CHD6* have been shown to have impaired motor co-ordination [326], but we could find no phenotypic characterization for heterozygous loss of the gene in this animal model.

In summary, the functional importance of *CHD6* and other reports for pathogenicity caused by loss of function of this gene strongly support the interpretation that heterozygous loss of *CHD6* in our patient is causative for her ID.

7.4.2.2 Copy number gains

We found 4 instances of copy number gains involving only exon 1 of four different genes. I will discuss each case below.

7.4.2.2.1 *CHD7* exon 1 gain and *ARID1B* exon 1 gain

Patient 33739 was found to carry two CNVs, both copy number gains involving only exon 1 (the only exon involved in both the maximal and minimal affected region) of the affected gene. One of the copy number gains involves *CHD7*, and the other involves *ARID1B* (Figure 7.3D and 7.3E). qPCR validation shows the CNV of the *CHD7* exon1 to be a gain of 6 copies above the normal 2 copy state while the gain in *ARID1B* exon 1 was validated to be a gain of only a single extra copy. With respect to the clearly validated gain of *CHD7* exon 1, it is not certain that the number of copies present is actually eight, as qPCR is unreliable for precise quantitation of extreme copy number events[300].

When considering the maximally affected region for the CHD7 exon 1 CNV, a first EF analysis shows the predicted promoter sequence is involved in the CNV. Exon 1 of this gene does not contain sequence that is part of the open reading frame of the protein. If we consider a tandem duplication in the same orientation, that would result in a longer 5' UTR. However the effect of gaining 6 copies of CHD7 exon 1 cannot be deduced with any certainty as we are unsure whether the duplications are tandem to each other, inverted, or even located at different regions in the genome. Therefore, a gain of function, loss of function or non-effect is possible depending on the orientation and position of the duplication and how that affects the gene promoter.

In the case of the ARID1B single copy gain for exon 1, a first EF analysis of the maximal affected region shows the predicted promoter sequence is involved in the CNV. A duplication of the promoter might cause increased expression levels or, alternatively, might adversely affect expression by causing aberrant binding of the transcription machinery. Assuming a tandem duplication of one allele, a frame-shift would not be expected because the duplicated exon ends after the third nucleotide of a translated codon. SMART analysis of the domain structure of the transcript with duplication of exon 1 predicted an extra motif but no effect on the only well-defined major domain of the protein. Although we are unable to predict the outcome of the *ARID1B* copy number gain on the ARID1B protein with certainty, it is unlikely to be a loss of function.

Our patient, a girl, is of French Canadian descent. Her parents are first cousins who both report learning difficulties at school. Her perinatal history is unremarkable, and her birth weight was 3.4kg (50th centile). She spoke her first words by 15 months but did not walk until 20 months of age. A psychomotor evaluation performed by her school at 6 years of age indicated that she had mild ID. At 7 years of age, bilateral posterior lamellar cataracts were discovered and removed. She had a congenital V pattern esotropia, but globe retraction characteristic of Duane anomaly was not noted. Assessment when she was 7 years and 6 months old showed a weight of 27.5 kg (90th centile), height of 127.8 cm (90th centile) and head circumference of 53 cm (60th centile). She had synophrys, epicanthal folds, narrow nasal root, broad nasal base and dental malocclusion. She was mildly hypotonic. She has had the following tests, all with normal results: EEG, brain CT scan, brain MRI, fragile X testing, testing for galactose-1-phosphate and

galactose uridyl transferase activity in red blood cells, plasma amino acid analysis and urine organic acid analysis.

Duplications of *CHD7* have recently been reported by our group [177] and another group [285] in association with a characteristic syndrome that is distinct from the well-recognized CHARGE syndrome, which is caused by *CHD7* loss of function. CHARGE syndrome is a multiple congenital anomaly ID syndrome characterized by choanal atresia and retinal, inner ear and heart abnormalities [42]. Our patient's phenotypes do not fulfill diagnostic criteria for CHARGE syndrome [327], though her eye abnormalities and ID are noted among the minor signs of the syndrome.

The first reported case of duplication of *CHD7* was described by Monfert et al. [285] in a patient who had mild to moderate ID, hypotonia, pulmonary stenosis and dysmorphic facial features including high eyebrows, right palpebral ptosis, short nose, long philtrum, and carp-shaped mouth with full lower lip (Figure 7.4a). In addition, deafness and Duane anomaly were noted. This child has a 3 Mb duplication that includes *CHD7* and also an apparently balanced translocation [t(4;7)(q11;q12)pat] inherited from his healthy father. The second published report of a patient with a duplication of *CHD7* appears in this thesis as chapter 6, where her phenotype is described in detailed. To summarize, she too, was developmentally delayed and hypotonic and had Duane anomaly among other dysmorphic features. She had a *de novo* 6.9 Mb duplication at 8q12.1q12.3 that included *CHD7* as well as 14 other genes.

Patient 33739 shares mild ID, hypotonia and gross motor delay with the two published cases with *CHD7* duplication. Figure 7.4 shows a photographic comparison of patient 33739 and the patient reported by Monfort *et al.* Although our patient does not have Duane anomaly, she did have strabismus and bilateral cataracts, which have been reported in patients with CHARGE syndrome [328, 329]. The presence of some phenotypes in common with the two other patients with *CHD7* duplications as well as those seen in patients with CHARGE syndrome supports the notion that the CNV affecting *CHD7* in this child is pathogenic. However further study is necessary to delineate the effect of the mutation on gene function and the clinical phenotype.

The second CNV found in patient 33739 is a *de novo* copy number gain of exon 1 of *ARID1B*. *ARID1B* encodes a factor that is part of a neuronal specific chromatin remodeling complex. We found a recently published report demonstrating reduced *ARID1B* expression as possibly involved in pathogenicity for autism [330], a phenotype not assessed in our patient. Nord *et al.* [330] found an intragenic deletion of *ARID1B* in a child with autism using high resolution aCGH and showed that the deletion produced a mutant truncated product. This deletion did not include exon1 of *ARID1B*, which is the region involved in the duplication in our patient. A search of the DECIPHER database shows three copy number gains and eight copy number losses that include this gene. However, the CNVs in all of the DECIPHER cases except # 250455 are very large, involving several dozen genes and therefore precluding meaningful comparison. DECIPHER #250455 is a copy number loss event that only affects *ARID1B*. Autism, coarse facial features with an asymmetrical face, low hairline, and choanal atresia, as well as ID and self-mutilation, have been recorded for DECIPHER #250455.

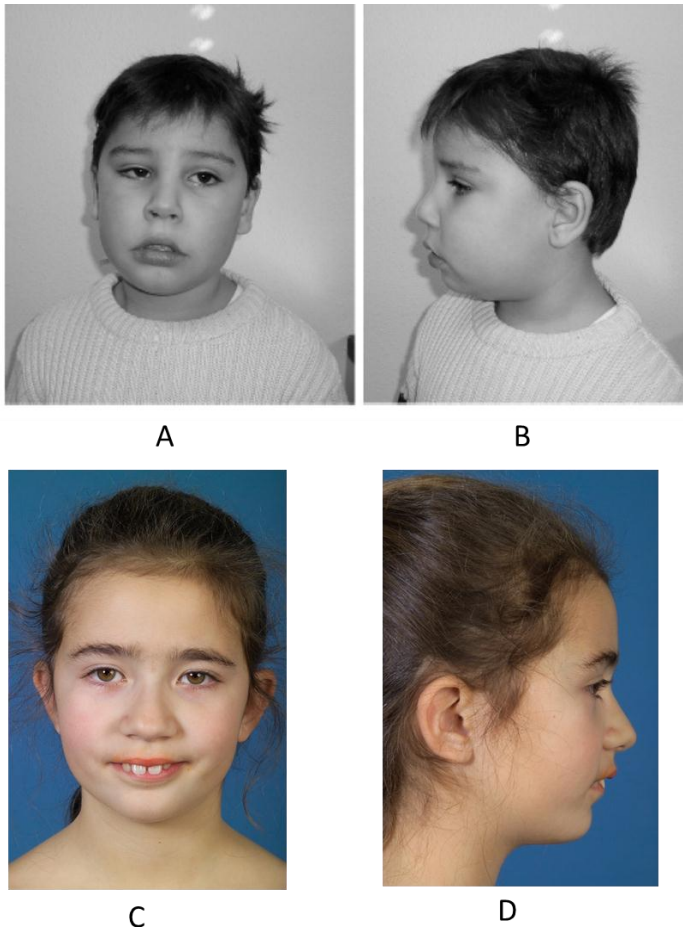


Figure 7.4: Comparison of facial photographs of patient from Monfort *et al.* [285] (A and B) with that of our patient 33739 (C and D). The images from Monfort *et al.* images are reproduced with permission from the publishers, and the images of our patient are included with the family's informed consent.

The above two reports support loss of function of *ARID1B* to cause autism. A recent report by Park et al. investigating Asian specific benign CNVs identified exon 1 of *ARID1B* in two independently validated gain CNVs in two clinically normal individuals(DGV# 111727)[331]. Taken together, these data indicate that the duplication of *ARID1B* exon1 is likely to be benign, although it is premature to rule out any involvement of this CNV in pathogenicity.

We believe that the copy number gain of *CHD7* exon 1 of *CHD7* is more likely to be causative for ID in patient 33739 than the copy number gain of *ARID1B*. However, this does not mean that duplication of exon 1 of *ARID1B* is necessarily benign.

7.4.2.2.2 *ARID2* exons 1, 2 and 3 gain

We found a gain of genomic sequence involving only the first three exons of *ARID2* in the minimally (and the maximally) affected region of patient 30848 (Figure 7.3F). Testing on four other whole genome aCGH platforms returned a normal result [318]. Results from the qPCR validation indicate a gain of an additional 6 copies versus the normal 2. It is not certain that the number of copies present is actually eight, as qPCR is unreliable for precise quantitation of extreme copy number events [300]. In this case qPCR primers were located to a sequence in intron 2 as representative of the entire CNV called by the array and we did not test each exon separately. When considering the maximally affected region, a first EF analysis shows the predicted promoter sequence is involved in the CNV. The coding sequence for *ARID2* begins at the start of exon 1, and if we consider a tandem duplication of exon 1, 2 and 3 in the same orientation, it would cause a premature STOP by altering the open reading frame. While it is not possible to infer the exact outcome of gaining 6 copies of *ARID2* exon 1, 2 and 3, based on the effect of a single duplication event it is reasonable to suppose the overall outcome will be a loss of function regardless of the number of copies gained as even if additional copies re-establish the correct reading frame the coding sequence itself will be altered by several exons.

Patient 30848 is a boy of non-consanguineous French-Canadian descent. His father reported he had learning difficulties at school. The perinatal history is normal. The patient's birth weight was 3.18kg (25th centile). He walked at 2 years of age. At 3 years and 9 months old he was seen by a developmental paediatrician who reported language delay (50 single words and ability to make short sentences of 3 words), and a global developmental quotient of 70 with

homogenous social, language, visuo-spatial, performance and practical reasoning scores on the Griffiths Development Scale. At this time the child's weight was 15.2 kg (50th centile), his height was 94.5 cm (5th centile) and his head circumference was 50 cm (50th centile). Both parents had normal heights (father 178 cm and mother 168 cm). Patient 30848 had no facial dysmorphism. His neurological examination was normal except for unilateral sensorineural deafness.

At 15 years 6 months old, the patient's weight was 43.4 kg (3rd centile), his height was 146.9 cm (<3rd centile) and his head circumference was 55 cm (50th centile). He was noted to have a narrow face, bulbous nasal tip and small mouth. There was bilateral clinodactyly of the fifth fingers. Patient 30848 had a normal brain CT-scan at 3 years of age, a negative test for fragile X syndrome, and a normal karyotype at the 550 band level.

One report in DECIPHER (#254078) includes *ARID2* in an apparently pathogenic copy number gain, but the affected region is up to 10 Mb in size and includes many other genes as well. A single DGV variant (#3887) is reported in the region; however, it does not include the first three exons of *ARID2* and was reported as a loss event. This variant has only been observed in one normal HapMap individual and did not meet the authors' criteria of independent confirmation [114].

ARID2 is an essential member of the SWI/SNF-B (PBAF) chromatin remodelling complex [332]. The SWI/SNF family is one of three major subfamilies of DNA-dependent ATPases that make up the chromatin remodelling machinery, which is structurally and functionally conserved across the animal and plant kingdoms [333]. The SWI/SNF complexes are a category of chromatin remodelling complexes that have been implicated in diverse biological processes, including transcriptional regulation, development, differentiation and tumorigenesis [334].

In humans, the SWI/SNF complex can be divided into two main components, BAF and PBAF, which have similar subunit compositions but are thought to have distinct functions [335]. *ARID2*, which encodes a factor that is also known as BAF200, is one of two molecules that is specific to the PBAF complex, i.e., not found in BAF, and is considered to be an unique

accessory component that regulates differential targeting of the holocomplex [333]. Furthermore BAF200 is required for PBAF to mediate expression of the interferon-responsive *IFITM1* gene, and *IFITM1* gene expression specifically depends on PBAF but not BAF regulation [332]. IFITM is a small conserved gene family that encodes factors with key roles in early development and control of cell growth [336]. IFITM1 in particular has been shown to have a negative effect on proliferation in both human cell cultures and mouse models [337-339]. These results point to the possibility that deregulated *ARID2* could in turn cause altered *IFITM1* activity, which could impact on cell proliferation and growth. This is interesting in light of the growth retardation seen in our patient, but this hypothesis requires a better understanding of the molecular effect of the *ARID2* copy number gain event in our patient. Nevertheless, the important functional role of *ARID2* and the extreme copy number gain in patient 30848 suggests that involvement of *ARID2* may be pathogenic, but confirmation will require the recognition of additional cases and genotype-phenotype correlation.

7.4.2.2.3 *ARID4B* exons 1 and 2 gain

This is a gain of an extra copy of genomic sequence involving minimally all of exons 1 and 2 of *ARID4B* in patient 33665 (Figure 7.3G), who presented with severe non-syndromic ID with autism. The maximally affected region includes flanking sequence in intron 2 downstream as well as sequence upstream of exon 1 that is predicted to include the promoter. The first exon of GGPS, an upstream flanking gene is also included in the maximal affected region. GGPS was not included in our design but does locate within the sequence we selected to be possibly regulatory for *ARID4B*. GGPS encodes the metabolic enzyme, geranylgeranyl-pyrophosphatase which catalyzes an important factor in the isoprenoid biosynthesis pathway[340], whose members are active in various biological pathways [341, 342]. We will focus our discussion on *ARID4B* as its involvement in the CNV has been independently verified.

When considering the maximally affected region, a first EF analysis shows the predicted promoter sequence is involved in the CNV. A duplication of the promoter might cause increased expression levels or, alternatively, might adversely affect expression by causing aberrant binding of the transcription machinery. Considering a tandem duplication of both exons 1 and 2, results in a larger 5' UTR region but does not affect the coding sequence. Thus

although it is not possible to predict with certainty the effect of this CNV on *ARID4B* expression, abnormal gene function would be expected.

ARID4B functions in histone deacetylation as part of the SIN3A co-repressor chromatin remodelling complex [343]. Wu *et al.* have shown that deregulation of *Arid4b* can lead to imprinting defects in the Prader-Willi syndrome and Angelman syndrome imprinting control loci in the mouse model [344]. It is noteworthy that autism, a phenotype present in our patient, is known to occur in some patients with imprinting defects caused by mutation of the Prader-Willi syndrome and Angelman syndrome critical region [345].

We found three large pathogenic CNVs that include this gene in the DECIPHER database: DECIPHER case #253950 (~15 Mb duplication inherited from a normal parent), #252439 (~20 Mb *de novo* duplication) and 251817 (~23Mb *de novo* duplication; patient also has another *de novo* large CNV) all include this gene in the affected region. However, the large size of these CNVs precludes a meaningful comparison with our case.

A single benign CNV is reported in the DGV that includes *ARID4B*. DGV# 48278 is a ~400 kb duplication reported by Sheikh *et al.* in two normal individuals of a cohort of over 2000, however this variant has not been independently validated [346].

Although the functional importance of the *ARID4B* product argues in favour of the pathogenicity of this CNV in patient 33665, we are unsure of the molecular impact of the CNV. More definitive conclusions require further studies.

7.4.3 CNVs affecting single intra-genic exons other than exon 1

We found a total of nine copy number changes that affected a single exon other than exon 1 of a gene involved in epigenetic regulation. Seven of these events involve *JARID2*. The other two cases are a copy number loss of exon 6 in *JMJD1A* and a copy number gain of exon 4 in *JMJD1C*.

7.4.3.1 *JMJD1A* exon 6 loss

A single copy deletion involving only exon 6 of *JMJD1A* within the minimal (and maximal) affected region was found in patient 31823 (Figure 7.3H). She was found to be normal on testing using a whole genome Affymetrix 6.0 platform. I modeled the transcription product and found that removal of exon 6 results in an altered reading frame, leading to a premature STOP codon. The same result was predicted by the Alamut analysis. Entering the truncated transcript into SMART showed that most of the protein would be lost, including the major JmjC functional domain.

Patient 31823 is a girl who was born to a French-Canadian mother and Middle Eastern father. The parents are non-consanguineous. The family history and perinatal history are unremarkable. At 4 years of age she suffered a *de novo* status epilepticus, and an EEG at that time showed epileptic activity in the right temporal lobe. This was not treated, and no seizures have been reported since. On physical examination at 6 years and 4 months of age, she exhibited global developmental delay. She could walk with help and eat with a spoon. She had no pincer grasp. She had no words or sign language and did not respond to simple verbal commands. Her visual contact was poor. She had a flapping hand mannerism. Her weight was 17.7 kg (10th centile), height was 107 cm (< 3rd centile) and head circumference was 51.6 cm (50-75th centile). No specific facial dysmorphisms were noted, and her tone was normal.

ADOS (Autism Diagnostic Observation Schedule) was positive, and she was diagnosed as having autism, but the specificity of this diagnosis was low because of the severity of her ID. A karyotype at the 500 band level was normal, and multiprobe subtelomeric FISH was negative. Methylation studies of the Angelman syndrome region were normal. Sequencing and MLPA of *MECP2* were negative. Analyses of amino acids, organic acids, acylcarnitine, ammonia, lactic acid and transferrin glycosylation were all normal. Brain MRI performed at 1 year of age showed mild increase in the size of the subarachnoid space and ventricles.

JMJD1A, also known as KDM3A, encodes a histone demethylase with specific H3K9 mono- and di- demethylation activity [347]. We found no other reports of either pathogenic or benign copy number losses involving this gene in humans. In mice, the orthologous *JMJD1A* protein has an essential role in embryonic stem cell differentiation [348]. In a study investigating the

downstream effects of knock down of the *JMJD1A* product, Loh *et al.* showed that a double depletion of *Jmjd1a* and another closely related gene, *Jmjd2c*, reduced the number of embryonic stem cell colony forming units and led to more rapid differentiation. *Jmjd1a* knock-down alone was shown to affect a large number of downstream targets; analysis of global mRNA expression showed that 125 genes were up-regulated and 100 genes were down-regulated [348]. Another study by Ko *et al.* found that *Jmjd1a* is a key target of STATs, an important class of upstream signaling molecules in mouse embryonic stem cells as well as in a human embryonic stem cell system[349]. Complete knock out of *Jmjd1a* results in adult onset obesity, hyperlipidemia and other signs of metabolic syndrome in mice [350][351].

In summary, the proven importance of the *JMJD1A* gene function in development and lack of benign variation involving this gene support the notion that this CNV is pathogenic, but further studies are required to establish a pathogenic relationship.

7.4.3.2 *JMJD1C* exon 4 gain

A copy number gain involving only exon 4 of *JMJD1C* in the minimally (and maximally) affected segment was observed in patient 33806 (Figure 7.3I). The patient also had a confirmed deletion of most of *UBE2A*. This gene does not belong to the epigenetic regulatory class and was not part of our gene selection, but it was included on the array among the other gene categories studied.

JMJD1C has two major isoforms: RefSeq accession numbers NM32776.1 and NM4241.2. NM32776.1 corresponds to the consensus coding sequence (CCDS) 41532.1, and, according to this annotation, the affected region in this patient's CNV only affects exon 4. No CCDS exists for NM4241.2, and according to this annotation the affected region affects only exon 3, which is completely absorbed in the 5' UTR sequence for the gene. Therefore, it is possible that this duplication results in an extra exon 4 or just in the incorporation of additional sequence into the 5' UTR, which is less likely to cause a defective gene translation product. If NM32776.1 is the relevant isoform and exon 4 is tandemly duplicated, no frame-shift was apparent on manual analysis of the *in silico* translated product. However, Alamut called a frame-shift mutation in NM32776.1 with a tandem duplication of exon 4. We considered the NCBI reference sequence record for our analysis, where exon 4 appears as a 114 bp sequence. Alamut calls exon 4 a 106 bp sequence. This discrepancy may be due to the fact that Alamut uses Ensemble

annotations. These differences in annotation and the occurrence of multiple isoforms for this gene cloud the functional analysis of this mutation.

Patient 33806 is a boy of French Canadian descent. The family history and perinatal history are unremarkable. He was noted to be hypotonic and developmentally delayed during his first year of life and had generalized tonic-clonic seizures from the age of one year. At 21 months of age he demonstrated limited interest in his surroundings, could not sit, and could not stand. He had no language capability, and he was hypotonic. His weight was 12.8 kg (75th centile), his height was 85 cm (50th centile), and his head circumference was 48.6 cm (50th centile). No facial dysmorphisms were noted. At 9 years of age he underwent surgery for unilateral cryptorchidism.

He had a normal EEG, but a CT scan showed cerebral atrophy. Cytogenetic assessment at the 500 band level was normal, and fragile X testing was negative. FISH testing for Prader-Willi syndrome and multiprobe subtelomeric FISH were normal.

Two CNVs were confirmed in patient 33806, a duplication of exon 4 of *JMJD1C* and a deletion of *UBE2A*. *UBE2A* is not an epigenetic regulator but the CNV of *UBE2A* is likely to be pathogenic for this patient's ID. The CNV we found deletes exons 3-6 of *UBE2A* and causes a frame shift that results in a prematurely truncated product with no functional domains. Mutations of *UBE2A* are known to cause X-linked ID in some boys [352, 353]. De Leeuw *et al.* described a *UBE2A* deficiency syndrome characterized by ID, absent speech, seizures and urogenital anomalies [352]. All of these features were also found in patient 33806. Some patients with *UBE2A* loss of function mutations have been reported to have various cerebral defects including white matter lesions and cysts [352] and white matter hypodensity [353]. Our patient had cerebral atrophy.

JMJD1C codes for a histone demethylase that specifically demethylates 'Lys-9' of histone H3, an important regulatory feature of chromatin [354]. *JMJD1C* may also be involved in hormone-dependent transcriptional activation by participating in recruitment to androgen-receptor target genes [355].

There are no reported benign copy number variants involving *JMJD1C* in DGV. The only DECIPHER CNV that involves this gene is a ~25 Mb pathogenic duplication event. Castermans *et al.* reported a patient with autism who has a translocation that disrupts *JMJD1C* as well as *REEP3* (Receptor Expression-Enhancing Proteins 3, a probable cytoplasmic vesicle trafficking regulator) and suggested that disruption of both of these genes had caused their patient's phenotype [356].

Thus, the *UBE2A* deletion in our patient appears to be sufficient to explain his ID. The effect, if any, of the copy number gain involving *JMJD1C* exon 4 cannot be determined from the available information.

7.4.3.3 *JARID2* exon 5 gains and losses

We found seven confirmed *de novo* CNVs that involved only exon 5 of *JARID2* in both maximal and minimal regions in all cases. Copy number losses were identified in Patients 33723, 32858, 31916, 32146 and 32094, and copy number gains were seen in Patients 33650 and 34335 (Figure 7.3 J-P). These CNVs were the only *de novo* copy number changes found in six of the patients. Patient 32858 also carried a deletion involving *MEF2C* which was identified on clinical array testing(*vide supra*), however the *JARID2* exon 5 CNV was missed on that platform. Patients 32146 and 32094 were tested on four whole genome aCGH platforms with normal results [318].

The phenotypes of these patients are not strikingly similar. Table 7.3 compares the phenotypes, and further details of six of the cases are given in the text following. Patient 32858 is described above in the section on his *MEF2C* deletion, which is considered likely to have caused his phenotype (section 7.4.1).

Table 7.3: Phenotypic comparison for patients with *de novo* CNVs of only *JARID2* exon 5. (W= weight, H=height, H.C = head circumference. '%' refers to centile score. Mo = months, PDD = pervasive developmental disorder, ASD= arterial septal defect, PF = palpebral fissures

	31916 (loss)	33723 (loss)	32146 (loss)	32094 (loss)	33650 (gain)	34335 (gain)
Family history	French-Canadian, mother has a half-brother with Asperger	French-Canadian. Unremarkable	South American. Unremarkable	French-Canadian. Unremarkable	African (Haiti). Unremarkable	North European. Unremarkable
Prenatal history	Unremarkable	Unremarkable	Gestational diabetes	Delivery at 35 weeks gestation	Delivery at 39 weeks gestation	Delivery at 40 weeks gestation
Birth	W 2.9kg. (<25 th %) H.C 34cm (25 th %)	APGAR 9-9-9. W 4.1kg (95 th %)	APGAR 9-10-10. W 4.9 kg (>97 th %)HC 39cm (>97 th %)	APGAR 8-9-9 W 2.5kg(<25 th %)	APGAR 9-9-9. W 2.7 kg (<25 th %). HC 32.5cm (5 th %)	W 3.2 kg (50 th %)
Growth	At 3yr, 9 mo.: W 12.8kg (3 rd %). H 92.3 cm (3 rd %). H.C 50.5cm (25-50 th %)	At 2yrs 11mo. W 17.2kg (90-97 th %). H 110.6 cm (90-97 th %). HC 52cm (97 th %)	At 7 yrs,4 mo: W 61.5kg (>97 th %) H 147.2 (>97 th %) BMI 28.4 HC 57.5 cm (>97 th %)	At 7rs: W 18.2kg(3 rd %), H 113 cm (3-10 %) HC 52.2 cm (50 th %)		At 6 yrs; normal W, H and HC
Developmental	Walked at 19 mo. At 3 yrs 9 mo., assessed PDD	Walked at 16 mo. Autism	Walked at 11 mo. Moderate ID	ID with attention deficit. Walked at 2 yrs.	Walked at 12 mo. Diagnosed with Autism at 5yrs.	-
Neurological	Normal	Normal	Normal	Bilateral neuro-sensorial deafness	Normal	-
Cardiac		-	-	ASD and valvular pulmonary stenosis	-	-
Craniofacial	None	None	None	Bilateral cleft lip palate. Absence of lacrimal glands. Upslanting PF. Prominent nasal root and tip. Telecanthus. Low set small ears	None	-
Ophthalmology	Exotropia treated surgically	-	Normal	-		-
Skeletal	Duplication of anterior arc of 5 th rib	Bilateral clinodactyly of the 5 th fingers	Bone age of 8-9 yrs at 6.5 yrs old	-	-	-
Genitourinary		-	-	Hypogonadotropic hypogonadism	-	-

7.4.3.3.1 *JARID2* exon 5 copy number losses

Patient 31916:

Patient 31916 is a boy whose parents are both French-Canadian. The family history is unremarkable except that the mother has a maternal half-brother with Asperger syndrome. The prenatal history was unremarkable. Birth weight was 2.9 kg (<25th centile). Patient 31916 first walked at 19 months of age. On evaluation at 3 years 9 months of age, he did not speak or respond to simple commands. He could write his name, recognize a few written words and draw a circle. He had no dysmorphic features, and his neurological exam was normal.

Psychometric evaluation with the Wechsler Preschool and Primary Scale of Intelligence - Third Edition (WPPSI-III), Leiter-R and Griffiths Developmental Scales showed non-verbal cognitive abilities within normal limits. A developmental pediatrician diagnosed pervasive developmental disorder with specific language impairment; ADOS-ADI (Autism Diagnostic Observation Schedule-Autism Diagnostic Interview) was not used for this assessment. Karyotype at 500 band resolution was normal. Fragile X testing, multiprobe subtelomeric FISH studies and brain MRI were all normal. Chest radiograph showed a duplication of the anterior portion of the 5th rib.

Patient 33723:

Patient 33723 is a girl whose parents are both French Canadian. The family history is unremarkable. Prenatal history is unremarkable. Birth weight was 4.1 kg (95th centile). She walked at 16 months of age and had her first words at 10 years of age. At 2 years and 11 months of age the child was assessed by a developmental paediatrician who diagnosed autism and global developmental delay. At the same age a physical exam presented weight 17.2 kg (90-97th centile), height 110.6 cm (90-97th centile), and head circumference at 52 cm (97th centile). She had no dysmorphisms except bilateral clinodactyly of the 5th fingers noted at the time. She had a normal neurological examination.

Her karyotype showed an apparently balanced inherited translocation: 46,XX,t(6;7)(q21;q11.23)mat. Fragile X testing was normal. Her brain CT scan was also normal.

Patient 32146:

Patient 32146 is a boy whose parents are both South American. His family history is unremarkable. The mother had gestational diabetes. Delivery was normal at 40.5 weeks gestation. Birth weight was 4.9 kg (>97th centile) and head circumference was 39 cm (>97th centile). No neonatal hypoglycemia was noted. He first walked at 11 months of age. At 7 years and 4 months of age, his weight was 61.5 kg (>97th centile), his height was 147.2 cm (>97th centile), and his head circumference was 39 cm (>97th centile). He could say a few words, understand a few simple commands, and draw a face. He did not know the letters and could not count. A psychological evaluation at that time showed moderate ID. He did not exhibit any dysmorphic features, and his neurological examination was normal. He was diagnosed with moderate intellectual disability and overgrowth.

Karyotyping at 550 band resolution, multiprobe subtelomeric FISH studies, fragile X testing, and *NSD1* sequencing were all normal. Ophthalmological examination at 8 years old was normal. Radiographic bone age was between 8 and 9 years at a chronological age of 6 ½ years. Brain CT scan was normal.

Patient 32094:

Patient 32094 is a boy whose parents are both French Canadian. His family history was unremarkable. He was delivered at 35 weeks of gestational age. Birth weight was 2.5 kg (<25th centile). At birth he was noted to have bilateral cleft lip and palate, atrial septal defect, and valvular pulmonary stenosis. He was subsequently found to have hypogonadotropic hypogonadism, bilateral sensorineural deafness, and absence of the lacrimal glands. He walked at 2 years of age. At 4 years and 1 month a psychometric assessment using the WPPSI-III gave a result of 'non-verbal intelligence: mild deficiency'.

On physical examination at 7 years and 1 month of age, he was noted to have upslanting palpebral fissures, prominent nasal root and tip, telecanthus and low-set, small ears. CT scan showed normal internal and middle ears. Abdominal ultrasound examination was normal. Karyotype, sequencing and MLPA of the *MID1* gene, sequencing and MLPA of *CHD7*, and whole genome aCGH using a135k oligonucleotide microarray were all normal.

7.4.3.3.2 *JARID2* exon 5 copy number gains

Patient 33650:

Patient 33650 is a boy who was born to parents who are both Haïtian and of African ancestry . The family history is unremarkable. He had an unremarkable prenatal history and was delivered at 39 weeks of gestation. Birth weight was 2.7 kg (<25th centile) and head circumference was 32.5 cm (5th centile). He walked at 12 months of age and had his first words at 15 months. However, regression of language and social interactions was noted around 20 months of age. At age 3 years he was found to be delayed in all spheres except gross motor function. He was non-verbal except for a few words that were not spoken in context; he was able to respond to a few commands. ADI-R and ADOS-G assessment were diagnostic of autism. A Griffiths Developmental Scale showed gross motor skills at a 39 month level, autonomy at 26 months, graphism at 24 months, language at 16 months, and knowledge at 19 months. A Vineland adaptive behaviour assessment showed mild to moderate deficiency. No dysmorphic features were noted on physical examination, and his neurological examination was normal. Karyotyping and electroencephalogram were normal.

At 5 years of age, a Mullen Scales of Early Learning showed visual perception at a 25 month level, fine motor skills at 30 months, receptive language at 10 months, expressive language at 10. The overall global assessment at this time was severe intellectual disability.

Patient 34335:

Patient 34335, a girl, is of North European descent. Her family history is unremarkable.

She was delivered at 40 weeks of gestation. Birth weight was 3.2 kg (50th centile). At 6 years of age, she could say fewer than 20 words, and she was diagnosed with ID. Her growth measurements were normal at this time.

Karyotyping was normal. MLPA for 15q11-q13, methylation study of the Angelman region, sequencing of the *MECP2* gene, MLPA of the subtelomeric regions, and brain MRI were all normal.

7.4.3.3.3 Functional analysis of *JARID2* exon 5 CNVs

No patients are reported in DECIPHER with CNVs involving only gains of exon 5 of *JARID2*, but DECIPHER does include one patient with a deletion involving only *JARID2* exon 5 (DECIPHER #255155). This patient shows language retardation, moderate cognitive impairment aggression and hyperactivity. Motor development and clinical examination were normal. The deletion was inherited from the patient's father, who had school difficulties. [Personal communication from Dr. Bertrand Isidor].

In order to assess the effect on JARID2 protein of exon 5 CNVs, we carried out extensive *in silico* analyses. Neither loss nor gain (assuming a tandem duplication) of exon 5 causes a frame-shift in the transcript (Note- Alamut analysis was not available for this mutation). SMART analysis predicted that removal or duplication of exon 5 only alters a low complexity element within the protein and not its major functional domain. I then used I-TASSER, a programme that models protein folding *in silico* [298, 299], to investigate whether deletion or duplication of *JARID2* exon 5 would alter the three-dimensional folding properties of the resultant proteins. I submitted the normal *JARID2* sequence, the *JARID2* sequence minus exon 5 and the *JARID2* sequence with a tandem duplication of exon 5 to the I-TASSER programme. The output shows the best five model predictions for each sequence configuration (Figure 7.5). This analysis predicts that the protein products resulting from exon 5 copy number gain or loss are dissimilar to the normal protein and different from each other.

Given the high number of CNVs involving this particular exon of *JARID2*, we re-assessed our CGH data and found that exon 5 was involved in a number of apparently inherited CNVs (that is, those that appear on the aCGH in comparison to one parent but not the other). We found 39 such events (16 deletions and 13 duplications) and given our confirmation rate of ~ 25% for aCGH calls, we estimate that at least 10 of these are true positives, suggesting this is a region involved in frequent copy number variation in the unaffected parents of our ID patients.

Although neither gain nor loss of *JARID2* exon 5 is listed in DGV as a CNV that has been observed in normal populations, it seems unlikely that such small changes would have been

recognized in most of the studies listed in DGV. The phenotypes of our patients with copy number loss or gain of *JARID2* exon 5 do not show any marked similarity. All of our probands had ID, but this varied greatly in severity (Table 7.3). In addition, one of the patients with a *JARID2* exon 5 CNV (patient 32858) also has another CNV that is more likely to be pathogenic. Moreover, the fact that we found many inherited CNVs of *JARID2* exon 5, indicating that one of the parents also carried the CNV, strongly suggests that many people who carry deletions or duplications of *JARID2* exon 5 are intellectually normal.

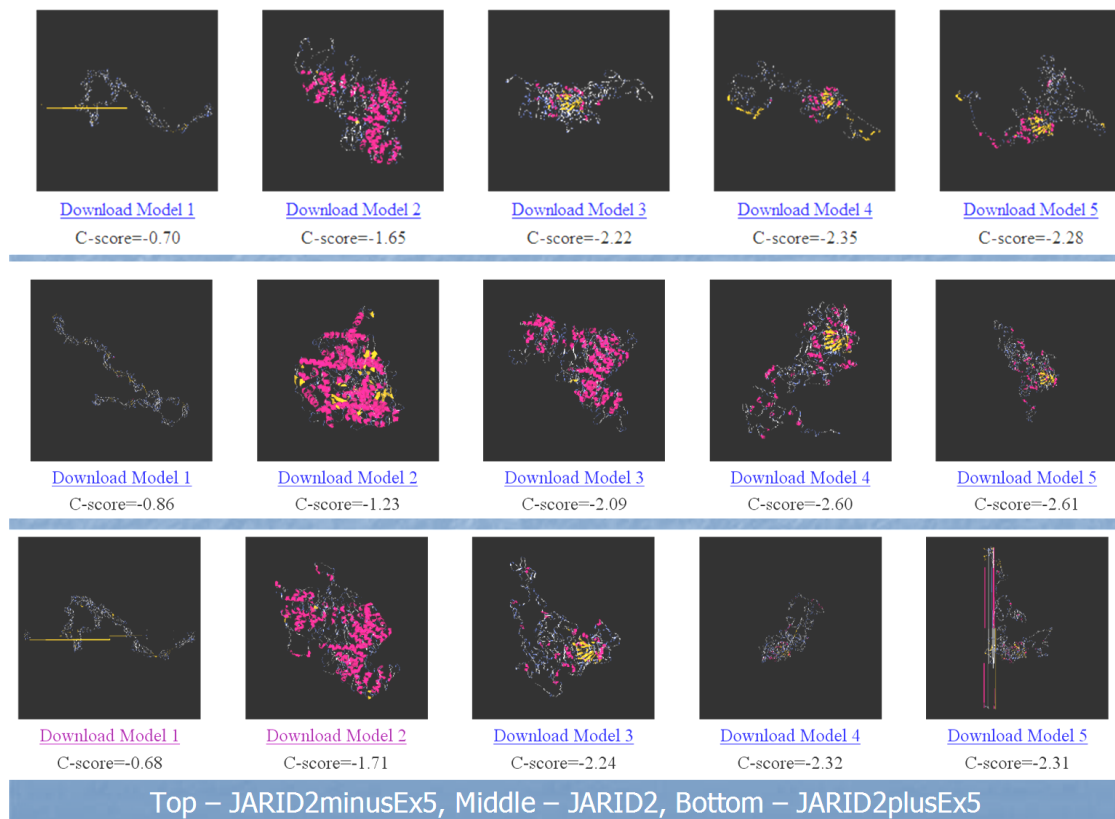


Figure 7.5: I-TASSER modeling for normal JARID2 protein (middle panel), mutant JARID2 with exon 5 deletion (top panel) and mutant JARID2 with tandem duplication of exon 5 (bottom panel). Each panel shows five best predictions in sequence from right to left with the higher C score signifying a greater confidence for the prediction.

Autistic features are described in two of the five deletions (31916 and 33723) and in one of two duplications (33650). Autism was also diagnosed in DECIPHER patient #255155, who has a deletion of *JARID2* exon 4. Although various dysmorphic features were described among these patients, there is no apparent overlap among the children with either copy

number losses or gains of *JARID2* exon 5 (Table 7.3). In addition, one of the children (patient 32094) with a deletion of *JARID2* exon 5 had multiple major and minor malformations, but other patients with similar deletions had no dysmorphic features at all. No two patients had the same dysmorphic features.

JARID2 encodes a histone methyltransferase with essential activity in many mammalian systems. It is the founding member of the largest family of histone demethylases identified – the Jumonji family [357]. The mouse ortholog, *Jarid2*, has been shown to have pleiotropic effects and an essential role in development [358]. For example, it is necessary for normal formation of the neural tube [359, 360] and the heart [361] and is involved in differentiation of embryonic stem cells [362]. Over and under expression of the protein have been reported to have different phenotypic outcomes, with over-expression inhibiting cell growth and under-expression enhancing growth in mice [363]. Landeira and Fisher review more than 15 published studies investigating *Jarid2* mutant mouse models; phenotypes have been reported in the brain, heart, spinal cord, liver, spleen and thymus. Interestingly, the phenotypic outcome for a given mutation may vary depending on the genetic background [358], indicating a complex activity for this protein.

In humans, *JARID2* is expressed by cells in the adult brain and heart [364, 365]. Weiss *et al.* conducted a genome-wide linkage and association analysis of more than 1000 independent families that included over 1500 individuals with autism [366]. The most strongly associated SNP found in this study was rs7766973, which is located in intron 1 of *JARID2* ($P = 6.8 \times 10^{-7}$). However, the authors were unable to replicate this association in an independent set of more than 2000 additional autism trios of diverse ethnic origin. A recent reanalysis of these data and those of another study of that obtained cases from the same source [367] using a latent class method that permits consideration of disease subtypes within families found associations of autism spectrum disorder with other SNPs within *JARID2* (rs6459404, also and rs6921502, both located in intron 1) although the effect did not reach a genome-wide level of statistical significance [368]. The association was found specifically with disease sub-type class 7, characterized by high levels of symptoms pertaining to social interaction, communication and language and low levels of stereotyped and repetitive behavior symptoms.

The idea that *JARID2* has an effect on neurofunctioning is further strengthened by reports from two independent groups of its association in schizophrenia [369, 370]. Pedrosa *et al.* genotyped nine markers located in the 3' ends of both *JARID2* and *DTNBP1* (an adjacent gene oriented in the opposite direction that is also suspected to be involved in the pathogenesis of schizophrenia) in 172 Caucasian and 152 African American patients and ethnically matched controls. A significant association was found with one marker located in intron 1 of *JARID2* [369]. Liu *et al.* genotyped 119 unrelated schizophrenia patients and 119 unrelated controls from a very well documented population isolate in a remote region of China [370]. One of three markers that achieved statistical significance in this analysis was located in *JARID2* (D6S289, located in intron 1). This association was not seen in two other populations from more heterogeneous regions of China. They then typed four SNPs located in the distal portion of *JARID2* and 10 SNPs located in the proximal portion of the adjacent candidate gene *DTNBP1* in an extended sample, and found a significant association with only rs2235258 ($P = 0.0087$, OR = 1.88) and rs9654600 ($P = 0.46$, OR = 1.72), both located in *JARID2* (both map close to exon 15 of the gene). Others have implicated *JARID2* in other phenotypes: Scapoli *et al.* report *JARID2* as involved in nonsyndromic cleft lip with or without cleft palate based on an initial linkage based identification and subsequent functional assays [371, 372] and Volcik *et al.* suggest *JARID2* defects pose a risk for spina bifida (SB) and congenital heart defects (CHD) upon identifying eight different variants in the gene (located in exon 4, 5, 6 and introns 3, 7, 9, 15 and 16) in a cohort of 74 infants with SB and 41 infants with CHD [373]. Of the three variants located in exons of *JARID2*, the variant in exon 5 was non-synonymous while the other two were synonymous. However, except for the variant in exon 6 (OR = 2.0 for SB and 3.5 for SB and CHD) and in intron 9 (OR= 2.1 for SB and 2.0 for SB and CHD), the others were detected at equal frequencies in a control population. These data do not offer compelling evidence for pathogenicity but indicate a possible small gene-effect size contribution of *JARID2* in these diverse developmental conditions.

In summary, the association of *JARID2* with multiple human diseases and its established importance in embryonic development raise the possibility that *JARID2* exon 5 CNVs might be involved in the pathogenesis of some cases of ID. However, it is unclear whether either a single copy deletion or tandem duplication of exon 5 has any functional effect on the *JARID2* protein. The high frequency of *JARID2* exon 5 gain and loss CNVs among the unaffected

parents of our probands and the lack of a consistent phenotype among our patients who carry *de novo* deletions or duplications of *JARID2* exon 5 argues against a direct pathogenic relationship to ID.

7.5 Discussion

7.5.1 Interpretation of study findings in relationship to the hypothesis

We have found 16 *de novo* CNVs that involve epigenetic regulators in 15 patients (8.5%) in our cohort of 177 patients with ID. Of these, the CNVs most likely to be pathogenic are the five that produce *SMARCA2* loss, *MEF2C* loss, *CHD6* loss, *JMJD1A* loss and *ARID2* gain. We also consider the high copy number gain for *CHD7* likely to be pathogenic, while the copy number gains for *ARID1B*, *ARID4B*, and *JMJD1C* may possibly be pathogenic. The recurrent copy number losses and copy number gains of *JARID2* that we observed are unlikely to be pathogenic for ID. In contrast, only 19 *de novo* CNVs (in 19 patients, 10.7%) were found that included other candidate genes included in the design, viz., genes involved in synaptogenesis, genes known to be causative for ID syndromes or representative of larger potentially pathogenic CNVs [7, 132]. Therefore, the rate of confirmed CNVs identified for genes in the epigenetic regulatory class (46% of the total from 36% of the probes on the array) is higher than that of all of the other classes put together (54% of the total CNVs from 64% of the probes on the array). This value does reach statistical significance (chi square test = 4.30, degrees of freedom = 1, two tailed P = 0.0372). When considering the number of genes involved as opposed to percentage of sequence, we are able to compare the set of epigenetic regulatory genes with the only other distinct functional gene class included on the array, that of genes involved in synaptogenesis (192 genes totally). Only 7 of the 19 *de novo* CNVs (20% of the total) we identified belong to the class of genes encoding factors involved in synaptogenesis. This means that as a functional class epigenetic regulatory genes are significantly more likely to be involved in *de novo* CNV events than genes involved in synaptogenesis (Z test for two proportions, $z=1.654$, $p=0.049$). If we consider only CNVs that are potentially pathogenic, i.e., disregard the *JARID2* exon 5 CNVs, the comparison is not significant (Z test for two proportions, $z=0.285$, $p=0.38$). However this analysis is limited by the fact that only these two functional classes may be compared based on our design and the assumption that all CNVs involving genes with function in synaptogenesis are pathogenic.

We began this study with the hypothesis that genomic imbalance for genes involved in epigenetic regulation would be pathogenic for 5% of cases. The arbitrary number of 5% was chosen as representative of a significant contribution to ID pathogenicity. Disregarding the *JARID2* exon 5 CNVs, we have identified CNVs that involve epigenetic regulatory genes in 5% of cases (9 of 177 patients). While it is not possible to determine with certainty whether all of the confirmed CNVs we found are pathogenic or not, our genotype-phenotype correlations suggest that genomic imbalance of our candidate epigenetic regulatory genes are an important cause of ID.

7.5.2 Limitations of study

There are several limitations inherent in the design of our custom microarray: (1) The sparse and irregular coverage afforded by our design prevents us from knowing whether pathogenic CNVs of untested regions are present in these patients. Seven of the patients (patients 30848, 31823, 32094, 32146, 32858, 32861 and 33459) we discussed above have had whole genome aCGH testing in addition to being tested on our custom aCGH. Of these, only patient 32861 in whom we found a deletion of *SMARCA2*, was found to have another possibly causative CNV on whole genome aCGH. (2) Our determination of where the CNV breakpoints are may not be correct because of the limitations of our aCGH bioinformatic analysis. In our experience, as is also generally observed [12, 374], we have found that breakpoints as defined by aCGH can be incorrect as they are based on a statistical inference for how the probe signal would be considered (as per the segmentation algorithm used by the analysis software) rather than an actual probe hybridization signal. Nevertheless our estimates *within* exons should be more precise because of the density of our coverage per exon. The fact that breakpoints within an intronic sequence cannot be localized precisely should not alter our interpretation of a CNV as long as the canonical splice donor and acceptor sequences are intact. However, the interpretation may be incorrect if the splice donor/acceptor sequence, an unrecognized part of an exon, or promoter or cis regulatory sequences are involved in the CNV but are not called accurately by the software. (3) Our quantitative estimates of how many copies are present in either a deletion or duplication are only approximate. In aCGH we compare the child's DNA separately to that of each parent. Our qPCR confirmations take a different approach – separate tests are performed on the child and each parent, and each is compared separately to a pooled reference sample. As a result, it is possible that we have

misinterpreted the direction (i.e., loss or gain) and parent of origin of some pathogenic CNVs.

Finally it must be borne in mind that in those cases where a clear syndromic association has not been established, it is possible the defect we have detected is not necessary and sufficient for the ID in the affected child. There may be other contributory factors involved in the pathogenesis that we are not able to detect by this method. This is especially relevant to the many patients harbouring CNVs involving *JARID2* exon 5. In chapter 8 of this thesis we discuss in more detail the relationship between epigenetic perturbation and disease manifestation.

7.5.3 Importance of study

This is the first study that specifically assesses genomic imbalance of epigenetic regulatory genes as a cause of ID. Our findings demonstrate the importance of this class of proteins in normal neurodevelopment and support the hypothesis that genomic imbalance of genes encoding epigenetic regulators is a major cause of ID.

The study design provided the ability to identify single exon copy number gains or losses and highlights the potential for even exonic imbalance events to be pathogenic. There has been an increase in focus on intragenic CNVs recently, and a number of investigators have attempted exon specific aCGH for a small set of carefully chosen genes that are important for certain diseases [375-378]. Efforts have also been undertaken to detect intragenic variation in normal individuals [112, 379]. However, there have been far fewer studies assessing a large set of candidate genes thought to be causative for a complex condition. The Wellcome Trust Case Control Consortium tested over 10,000 loci for CNVs in a large sample of cases and controls for common complex disease such as bipolar disorder, hypertension, etc. [380].

Boone *et al.* recently reported results from their test of over 3700 cases referred to the Medical Genetics Laboratory at Baylor College of Medicine with their custom targeted aCGH probing ~1700 candidate genes at an average coverage of four probes/exon [381]. The

patients they studied have a varied clinical presentation, but many, no doubt have ID. Genes with epigenetic function that are already established to be disease causing make up three of the 31 intragenic CNVs these authors report as potentially pathogenic (*MECP2* deletions as causative for Rett syndrome, *EP300* deletions as causative for Rubinstein-Taybi syndrome 2, and *CREBBP* deletions as causative for Rubinstein-Taybi syndrome), further corroborating our findings that intragenic CNVs of genes of this class can be pathogenic.

Of the 16 *de novo* CNVs involving genes encoding epigenetic factors we identified, 14 affect only a part of the gene. Half of these, that is all except the seven CNVs involving exon 5 of *JARID2*, are good or likely candidates for pathogenicity. In four cases the first exon and possibly the promoter are affected; *CHD6* exon1, *CHD7* exon 1, *ARID2* exons 1,2 and 3, as well as *ARID4B* exon 1, and in two cases other exons are affected (*JMJD1A* exon 6 and *JMJD1C* exon 4). Our data have broad implications for the clinical assessment of CNVs, as current diagnostic standards exclude most single exon genomic imbalance from consideration as being pathogenic. The most recent consensus statement on clinical use of aCGH (both for oligonucleotide and SNP probe arrays) offers as a guideline that only CNVs $\geq 400\text{kb}$ throughout the genome be called by clinical testing for further consideration of pathogenicity [374]. This resolution is too low to detect most single gene defects and certainly would not detect single exon CNVs because exons range in size from 100 bp-1000 bp on average. Our findings support the need for re-assessment of the diagnostic approach taken with aCGH results in regards to CNVs that include only parts of genes. In the clinical context, a study of intragenic CNVs may not be feasible due to the extensive number of results needing further in-depth validation such an approach would generate. However a filter based on function of the genes/parts of a gene is suggested to be useful in order not to disregard potentially pathogenic CNVs that fall below the current diagnostic size threshold.

Chapter 8: General discussion

Genotype-phenotype correlations for CNVs identified by aCGH have led to identification of novel candidate genes, clinical description of previously unrecognized syndromes and greater insight regarding the pathogenesis of these conditions. In addition, findings from these studies have informed subsequent research, as exemplified by our custom aCGH to detect CNVs involving genes that encode epigenetic regulatory function. The contributions to our current understanding of the genetic basis of ID from the studies presented in this thesis are discussed below, with particular consideration of the role of epigenetic regulation in neurodevelopment and function.

8.1 Genotype-Phenotype studies for whole genome array genome hybridization data.

In this thesis I have presented five detailed genotype-phenotype studies (chapters 2-6), described the genotype-phenotype relationships more briefly for other patients with idiopathic ID who were found to have *de novo* CNVs in our 100K and 500K AGH cohorts (appendices A and B), and used genotype-phenotype correlations to assess the findings from our custom aCGH study with respect to epigenetic regulatory genes (chapter 7).

In chapter 2, we showed that AGH findings are able to establish new ID syndromes by a 'reverse dysmorphology' approach, i.e., initial identification of a similar genotype enables the recognition of a characteristic clinical presentation. This paper, '**Novel deletions of 14q11.2 associated with mental retardation and similar minor anomalies in three children**', published in 2007, reported the identification of a characteristic pattern of minor anomalies and ID produced by submicroscopic deletions of chromosome 14q11 that disrupt the genes *CHD8* and *SUPT16H*. Since publication of this work, we have received reports of three new patients who appear to have the same condition; one family from Palestine and two families from the USA with children who have similar characteristic facial features and deletions of chromosome 14q11. These reports support the notion that we have indeed identified a novel ID syndrome; however a more thorough analysis of these reports as well as other possible cases is required in order to characterize this condition better and establish its frequency in the population.

In our description of the three original cases, we identified two possible causative genes by overlapping the deleted regions in the patients. Both candidate genes play important roles in epigenetic regulatory processes. This finding led us to explore further the importance of correct epigenetic regulation for normal brain development and functioning.

In chapter 3, we presented another genotype-phenotype analysis that showed a phenotypic effect as the result of the disruption of only one transcript of *NRXN1*. *NRXN1* has two different promoters that code for the major isoforms of neurexin proteins, which play a prominent role in synaptogenesis. In this work entitled, '**A patient with vertebral, cognitive and behavioural abnormalities and a *de novo* deletion of *NRXN1α***', we correlated the loss of only one isoform with specific phenotypic conditions. The work was important as it was the first to identify the *NRXN1* gene, which had already been reported as a candidate for autism [209], as also causative of ID. Since publication of our paper, others have shown the involvement of the *NRXN1* gene in ID [382]. In addition, *NRXN1* has now also been implicated in schizophrenia [383], suggesting that disruption or deregulation of a single biochemical or neural pathway can lead to ID, autism and schizophrenia phenotypes. Although a detailed discussion of this idea is beyond the scope of this thesis, I do consider the pathogenic relationships between ID, autism and schizophrenia as a result of epigenetic deregulation in a review article [179] and in section 8.2, below. These data point to the need for further studies into the biochemical and pathophysiological basis of complex neurodevelopmental disease.

Chapter 4 details a genotype-phenotype correlation for another case that is interesting because it delineates a clinical outcome resulting from the possible over-expression of the causative genes. This is in contrast to most CNV studies, which implicate reduced gene expression or haploinsufficiency as the mode of pathogenicity [2]. In this publication, '**A novel *de novo* 1.1 Mb duplication of 17q21.33 associated with cognitive impairment and other anomalies**', we reported a 14 year-old girl with mild cognitive impairment, deafness, and an unusual pattern of congenital anomalies associated with a previously-unreported *de novo* duplication of chromosome 17q21.33. The patient has microcephaly, unusual cup-shaped ears, scoliosis and other skeletal defects. We identified two genes, *PPP1R9B* and *COL1A1*, as strong candidates for producing her phenotype. Further work is needed to establish the pathogenicity of these genes and their mode of action. However, a

recent report also highlights this region as being causative for ID in multiple patients who also report skeletal defects including scoliosis [384].

My thesis includes published manuscripts of two additional genotype-phenotype correlation studies on which I collaborated with other researchers. Both of these contributed to the development of my main research hypothesis as they implicate genes with epigenetic regulatory function as causative for characteristic ID syndromes. The study '**Duplications of the critical Rubinstein-Taybi deletion region cause a novel recognizable syndrome characterized by mild arthrogryposis**' is presented in chapter 5, and the study '**A characteristic syndrome associated with microduplication of 8q12, inclusive of *CHD7***' is given as chapter 6. In both instances we see that genes encoding epigenetic regulatory factors (viz., *CREBBP* and *CHD7*) that cause recognized ID syndromes when deleted (Rubenstein Taybi syndrome and CHARGE syndrome, respectively) produce quite different patterns of anomalies when duplicated. These findings support the notion that correct dosage for genes encoding epigenetic regulators is necessary for normal brain functioning.

In chapter 7, I present my main research project: a custom aCGH study that was designed to assess the contribution of genomic imbalance for the class of genes encoding epigenetic regulatory proteins to ID pathogenicity. We studied 177 trios, each composed of a child with idiopathic ID and both normal parents, on a custom array that included coverage at exonic resolution for all known genes encoding epigenetic regulators. The results support my hypothesis that genomic imbalance for epigenetic regulatory factors is an important cause of ID. Furthermore, I have found that CNVs at the intragenic level, which is far below the resolution of most clinical aCGH testing, can contribute to disease. Most of the genes I found to be involved in children with ID encode factors that occur in complex multi-unit chromatin remodeling complexes.

In addition, many of the genes I found that appear to cause ID when their copy number is abnormal have also been found by others to be involved in the pathogenesis of related neuro-functional conditions such as autism or schizophrenia. This observation indicates the complex effects of abnormal epigenetic control during development and neurocognitive function. Epigenetic regulation provides fine scale orchestration of control, impacting diverse

down-stream targets, thus it is an attractive target for study to help us unravel how the brain develops and functions. These ideas are elaborated in the following section, in which I discuss epigenetic impacts on neurodevelopment in more detail.

8.2 Epigenetic impacts on neurodevelopment: pathophysiological mechanisms and genetic modes of action

Epigenetics is attractive in the context of complex disease, as it is able to define a molecular mechanism that links environmental effects to gene function. That is, epigenetic modulation is able to act as an interface between the environment and the genome. This is especially relevant when discussing neurofunctional disorders as they often involve a large environmental component in their etiology. For this reason there is an increasing attention on epigenetics in pathophysiological studies in not only ID but also in schizophrenia and autism.

8.2.1 Pleiotropy and functional complexity

As epigenetic modulators are often involved in multi-protein complexes [28] and given that epigenetic change is impacted by multiple chromatin modifying pathways [26] I suggest that mutations in epigenetic modifiers may be particularly prone to exhibiting pleiotropy. This certainly seems to be the case for *MeCP2* as mutations of the gene are known to cause a number of different phenotypes (Table 1.1).

Pleiotropy can arise when the dysfunctional gene's product affects a number of downstream targets [385]. In the epigenetic context, the mutated gene could encode an epigenetic regulator, and the anomalous product (or absence/over-expression of the product) therefore would cause deregulated expression of a number of other genes. A good example of this is provided by *MECP2*, which specifically binds to methylated cytosine residues of CpG islands and recruits other factors that contribute to establishing an inactive chromatin state [37]. There have been thorough reviews of the phenotypic outcome of the large number of mutations found in *MECP2* [37, 386]. Focusing on the hypothalamus in mice, Chahrour et al. show that the MECP2 protein can act as both an activator and a repressor, and that it serves as a direct transcriptional regulator for the majority of genes that it affects [387].

Notably, it is clear that the MECP2 activity is central to further epigenetic control of the genes targeted. Of the 2582 genes tested, they found that abnormally elevated or abnormally decreased MECP2 levels (engineered by using a gene construct with a hyperactive promoter and a null allele respectively, in transgenic mice of two different strains), affected the expression patterns of a staggering ~85% of genes. *A2bp1*, *Gamt* and *Gprn1* are among the target genes. Interestingly, disruption of *A2BP1* in humans has been implicated in autism susceptibility [388] as well as ID and epilepsy [389], *GAMT* deficiency has been shown to cause severe ID [390], and the human homolog of *Gprn1*, *GRIN* is well documented as a causative gene for schizophrenia [391-393]. Therefore pleiotropy in this example could be a manifestation of the multiplicity of binding targets for *MECP2*.

Pleiotropy can also be brought about by mutations altering the protein functionality in a domain-specific manner. Thus, the phenotypic outcome could vary according to which functional domain of the protein was altered by the mutational event. *MeCP2* also provides a good example of this, being a protein with multiple well-characterized functional domains. Three distinct domains are known for *MeCP2*: a methyl binding domain (MBD) that binds to methylated cytosine residues in the DNA strand, a transcriptional repression domain (TRD), which binds to other chromatin remodeling factors as a protein-protein interaction domain, and a C-terminal domain which can bind naked DNA and RNA splicing factors [387]. In this case, where one protein has many binding partners, it can be hypothesized that genetic changes which alter specific binding properties of the protein [37] can affect the phenotypic outcome in different ways.

In this context our findings of exon level genomic imbalance is important. As the deletion or addition of an exon (as is possible due to an intragenic CNV not altering the reading frame or affecting regulatory sequence for example) that codes for different functional domains may cause a different downstream phenotype. While a CNV affecting an exon coding a key functional domain may be highly penetrant, another impacting an exon encoding a less important functional sequence may be tolerated. This possibility is illustrated by the *JARID2* exon 5 CNVs that we found. While clearly the gene has an important function, CNVs involving this particular exon may not have as great an effect as those involving exons coding for more important domains in the protein [357, 358].

Another source of phenotypic variability possible due to dysfunction of epigenetic regulators depends on their extent of involvement in coincident pathways. This is illustrated by the case of *DNMT3B*. The enzyme does have a primary epigenetic programming function, being a DNA methyl transferase; however, it is one of three major DNA methyltransferases, the other two being *DNMT3A* and *DNMT1* [394]. Of these, *DNMT1* is considered to be the maintenance methyltransferase and *DNMT3A* and *DNMT3B* are termed *de novo* methyltransferases [395]. *DNMT1* is the most abundant methyltransferase in somatic cells [395]. Aberrant expression of *DNMT1* has been shown to result in extreme global DNA methylation defects, and embryonic lethality in mammals [396-398]. Members of the *DNMT3* family on the other hand, display more specific tissue expressivity [394, 395]. *DNMT3B* in particular appears to have a smaller effect size. Rhee *et al.*, reported that disruption of *DNMT3B* only reduced global methylation by <3%, however when both *DNMT3B* and *DNMT1* were disrupted, global methylation was changed by >95% [399]. The take-home message is that the overlap/redundancy with other family members can influence the pathogenicity of defects in epigenetic regulators.

In a similar vein, for epigenetic regulators that function in more than one multiunit complex, the alteration of its function can have a nuanced impact dependent on which multiunit complex is affected, and how. In this context genetic background would play an important role. A mutation in a member of a multiunit chromatin remodeling complex may not be phenotypically evident in one individual, however in another individual, who may have a variant in a different member of the same multiunit complex, or perhaps in a related or partially redundant complex, the combined effects of the mutated epigenetic regulators may manifest in a disease outcome. This would be analogous to the situation for CNV pathogenicity, where it has been shown that a CNV which is benign in one individual could be pathogenic in another individual who carries a second benign CNV affecting different genes [400]. In this case the combined effect would be pathogenic whereas each CNV on its own is non-disease causing [12].

Epigenetic regulatory protein expression levels can also contribute to pleiotropy. *CBP* deletions are considered a common cause of Rubenstein-Taybi syndrome [38, 56, 401] and duplications are causative for the recently described characteristic 16p13.3 syndrome (chapter 5 of this thesis) [59, 265] (Table1.1), indicating that the over and underexpression

of the gene product affects different molecular pathways, or the same pathways differently. CBP is a histone acetyltransferase and functions as a transcriptional co-activator [402]. The distinct phenotypes observed due to its under versus over expression highlight the sensitivity to incorrect copy number or dosage imbalance of epigenetic regulators for correct neurodevelopment.

CNVs are frequent in the population and are thought to be causative for about 15% of ID [2]. In keeping with our understanding of neurogenetic pathogenicities caused by CNVs in general, we see that the loss of an epigenetically functional gene is more frequently implicated than the gain of the same gene (Table 1.1 and [12]). The indication is not that losses occur more frequently than gains, but that losses are less well tolerated than gains [12].

It can be theorized that the over-expression of epigenetic regulators should not impact the overall functional outcome as having more product would not alter the normal sequestering of these factors. However the lack of sufficient epigenetic regulators would result in an impairment of the sum functional outcome. But given the context that many epigenetic regulators, especially those involved in chromatin remodeling complexes, do act as part of large multi-unit complexes, the situation may be much more complex.

8.2.2 Endophenotypes and epigenetic modes of action in the brain

The task of correctly correlating genotypes to phenotypes is particularly challenging for neurodevelopmental disorders. There is an overlap of features among different ID syndromes [2]. In addition an overlap is also observed between the broader clinical neurodevelopmental disorder categories. For instance, autism is often part of the presentation for ID cases [2], and patients with schizophrenia can display behaviors which are part of the spectra of other disease categories such as Bipolar Disorder (BD), Attention Deficit Hyperactivity Disorder (ADHD), ID or autism [403]. The finding of a phenotypic overlap as well as a common genotype (e.g. Table 1, *MECP2* defects causing Rett Syndrome, autism and schizophrenia) lends weight to the approach to study the genetic basis of these disorders by breaking them down into endophenotypes [403, 404]. An endophenotype can be defined as a largely heritable quantitative trait that is part of the

pathophysiology of a given disorder, but not necessarily sufficient to manifest the disorder itself (a comprehensive history and definition of the term is found here [405]). For example, there has been focused study of specific brain morphologies as an endophenotype of autism [406-409], ID [397] and schizophrenia [410]. The hope is that an endophenotype approach will help demystify the genotype to phenotype connection and articulate a more straightforward cause and effect model [404]. This is especially relevant in the context of neurodevelopmental disorders, where the syndromic presentation and behavioural phenotypes are complex and to say a given gene may directly control a given characteristic (e.g., IQ) is at best an over-simplification [411].

Another important perspective for understanding the role of epigenetic regulation in neurodevelopment, is looking at where and when epigenetic regulatory factors play a role in the developing central nervous system (CNS). MacDonald and Roskams, upon reviewing a number of studies researching the spatio-temporal expression patterns of epigenetic regulators in mouse brain, argue for the occurrence of ‘discrete epigenetic checkpoints’ depending on the time and place of expression of Hdacs, Dnmts and Mbd proteins [34]. This opinion is interesting as it allows us to think in terms of not only gene targets for the pathogenicity of epigenetic deregulation but enables us to view epigenetic deregulation as affecting spatial and temporally bound aspects of neurodevelopment. The work by Chahrour *et al.* [387] (*vide* section 8.2.1) lends support to this model as their groundbreaking findings came about due to approaching the investigation by looking not at the brain as a whole but at a specific brain region, in this case the hypothalamus.

Others have taken the study of spatio-temporal patterning a step further by looking at neuron subtype specific expression of epigenetic regulators. For example, studies have shown that aberrant epigenetic events deregulate important targets in GABAergic neurons of schizophrenia postmortem brains [412, 413]. Costa *et al.* detail the effects of *DNMT* expression levels in GABAergic neurons, showing a *DNMT* dosage dependent activity of key target genes [414]. These observations point to links between specific neuronal activity outcomes or endophenotypes and the way epigenetic factors play out in different brain regions and at different time-points in cognitive development and function, though we are far from a thorough understanding of these complex processes.

Genetic background offers another explanation for the variability possible due to epigenetic deregulation. For instance, in the case of CNVs, inherited variants are being shown to be benign or pathogenic in different individuals in an increasing number of ID cases [12]. While we can explain the different effect of the same mutation in different individuals in terms of variable expressivity and penetrance purely situated upon the genotypic background of the concerned individual [12], Van Winkel *et al.* point to the phenomenon that certain genotypes may be correlated with contradistinctive epigenetic signatures. They posit that the 'genetic background' should be discussed in terms of the epigenetic landscape as there could be individual specific genotype dependent differential DNA methylation states [415]. This moves epigenetic regulation into the paradigm of the inherent individual's hereditary (or genomic and epigenomic) variability and not simply limiting epigenetic mechanisms to a mode of action. Therefore while epigenetics is particularly attractive as a link between gene and environment, exactly how it functions in this context may not be straightforward.

8.3 Conclusion

As discussed in section 8.2, we understand that the role of epigenetic regulation in neurodevelopment is multi-faceted and complicated. Modeling the link between aberrant epigenetic control and neurocognitive disorders may be attempted from a number of angles including highlighting the relationships between molecular genetic functional pathways, the spatial and temporal aspects of regulation in the brain, and how the environment plays a part in the brain's function.

The relatively small number of genes encoding epigenetic regulators in comparison to the large number of genes involved in neurodevelopment which these regulators are able to control, emphasizes the extent of pleiotropy possible for defects involving genes encoding epigenetic regulators. The ability of mutations to affect particular functional domains within an epigenetic regulatory protein, the existence of multiple isoforms of some of these proteins, the partial functional redundancy of the genes, and the involvement of some epigenetic regulators in integrative pathways result in more layers of functional and consequent phenotypic variation when genomic imbalance of epigenetic regulatory genes occurs. Further, that many epigenetic regulators operate as part of multi-unit complexes allows for a higher sensitivity to dosage dependency, as the functional outcome may be fine-

tuned stoichiometry among components of the complex. This is evidenced by the frequent occurrence of genes encoding epigenetic regulators as part of pathogenic CNV events.

In chapter 7, I presented the main research project of my thesis where we used a microarray with resolution sufficient to identify CNVs involving only a single exon, and targeted to all genes with epigenetic function, and my finding of 9% of 177 idiopathic ID trios with *de novo* CNVs that include an epigenetic regulator. I have discussed genotype-phenotype correlations for each of the genes with epigenetic function identified as copy number variant in our study in section 7.4. Those results show that seven *de novo* CNVs, each involving a different gene and a different family, are definitely or probably pathogenic for ID. Two of these cases involve genes that were previously known to cause ID when present in abnormal copy number; the others are novel observations that implicate additional epigenetic regulatory genes in the pathogenesis of ID (*SMARCA2*, *CHD6*, *ARID2*, *ARID4B*, *JMJD1A*).

A striking result of the study was the identification of frequent CNVs involving only exon 5 of *JARID2*, a gene with proven important function, yet this particular CNV appears to be a common polymorphism in the population. These results should be considered in the context of the complexity inherently present in how imbalance of factors with epigenetic function can cause disease (*vide* section 8.2) and in how we assess normal neurological functioning itself. Particularly, while the *JARID2* exon 5 CNV is most likely benign it cannot be dismissed as completely non-contributory without further study, especially given the intricate nature of epigenetic regulation and the balance of many factors involved. Therefore the defect can be considered a factor contributing to the individual's genetic or epigenetic landscape. Any subtle effect this background may produce is expected to only shift the neurofunctional spectrum away from the normal but is expected to be far from what is required to produce a pathological state. The situation is similar to that of *NRXN1* (*vide* section 3.3); in this case also the gene is implicated in many neurodevelopmental disorders and benign intragenic CNVs are known within the gene. Extensive alternative splicing is considered to be one explanation of why different *NRXN1* defects can cause different phenotypes. The complex pleiotropy manifested by epigenetic regulatory systems offers the same mechanistic fluidity as extensive alternate splice effects, allowing for variable expressivity and penetrance not

only in the case of the *JARID2* exon 5 CNV but for all defects involving genes with epigenetic regulatory function.

The results generated by the focused study of genomic imbalance for genes encoding epigenetic regulators does raise the importance of this class as a substantial contributor to the pathogenicity of ID and upholds my hypothesis that imbalance of genes encoding epigenetic regulators is an important cause of ID. The fraction of genes with epigenetic regulatory function that are implicated in ID can be compared to the fraction of total genes that are pathogenic for ID;

$$\frac{ID_{er}}{G_{er}} : \frac{ID_t - ID_{er}}{G_t - G_{er}}$$

[where ID_t = estimate of the current number of ID genes; ID_{er} = the number of ID genes of the epigenetic regulatory class; G_t = the current estimate of the total number of genes; and G_{er} = the total number of genes of the epigenetic regulatory class.]

Adding the five newly implicated genes with epigenetic regulatory function (*vide supra*) to those already known to be causative for ID (Table 1.1) results in 21 genes with epigenetic function considered pathogenic for ID (ID_{er}). Since I selected genes with known epigenetic regulatory function for inclusion in our custom microarray, several more have been discovered. A more recent collection of genes with epigenetic function is by SigmaAldrich in their catalogue (<http://www.sigmaaldrich.com/life-science/epigenetics/epigenetics-genes.html> accessed may 2011), where they list 226 genes totally (G_{er}). The total number of genes implicated in ID on the autosomes and sex chromosomes can be estimated to be approximately 950 (ID_t) (*vide* section 1.2.1.3) of the approximately 20,000 genes recognized in the human genome (G_t). Therefore;

$$\frac{21}{226} : \frac{950 - 21}{20,000 - 226}$$

equals;

$$0.09 : 0.047$$

These calculations show that the fraction of genes with epigenetic regulatory function implicated in ID is almost twice as that of the fraction of total genes estimated to be pathogenic for ID.

The brain is a highly sophisticated organ, and it is required to have an extreme level of plasticity because it must continually react and adapt to diverse external stimuli [416]. Given that epigenetic regulation can act as a versatile 'switching mechanism' to fine-tune gene function, epigenetic regulation may be of critical importance in the central nervous system development and function. In the context of neurodevelopmental disease, delineating the role of epigenetics is particularly attractive because it offers the promise of novel therapeutic inventions that provide hope for people with these debilitating disorders.

References

1. Roeleveld, N., G.A. Zielhuis, and F. Gabreels, *The prevalence of mental retardation: a critical review of recent literature*. Dev Med Child Neurol, 1997. **39**(2): p. 125-32.
2. Ropers, H.H., *Genetics of early onset cognitive impairment*. Annu Rev Genomics Hum Genet, 2010. **11**: p. 161-87.
3. Leonard, H. and X. Wen, *The epidemiology of mental retardation: challenges and opportunities in the new millennium*. Ment Retard Dev Disabil Res Rev, 2002. **8**(3): p. 117-34.
4. Koolen, D.A., et al., *A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism*. Nat Genet, 2006. **38**(9): p. 999-1001.
5. van Karnebeek, C.D., et al., *Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness*. Eur J Hum Genet, 2005. **13**(1): p. 6-25.
6. Rauch, A., et al., *Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation*. Am J Med Genet A, 2006. **140**(19): p. 2063-74.
7. Friedman, J., et al., *Detection of pathogenic copy number variants in children with idiopathic intellectual disability using 500 K SNP array genomic hybridization*. BMC Genomics, 2009. **10**: p. 526.
8. Jaillard, S., et al., *Identification of gene copy number variations in patients with mental retardation using array-CGH: Novel syndromes in a large French series*. Eur J Med Genet, 2010. **53**(2): p. 66-75.
9. Koolen, D.A., et al., *Genomic microarrays in mental retardation: a practical workflow for diagnostic applications*. Hum Mutat, 2009. **30**(3): p. 283-92.
10. Freeman, J.L., et al., *Copy number variation: new insights in genome diversity*. Genome Res, 2006. **16**(8): p. 949-61.
11. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome*. Nat Rev Genet, 2006. **7**(2): p. 85-97.
12. Lee, C. and S.W. Scherer, *The clinical context of copy number variation in the human genome*. Expert Rev Mol Med, 2010. **12**: p. e8.
13. Ensenaer, R.E., et al., *Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients*. Am J Hum Genet, 2003. **73**(5): p. 1027-40.
14. Shaikh, T.H., et al., *Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis*. Hum Mol Genet, 2000. **9**(4): p. 489-501.
15. Vlangos, C.N., D.K. Yim, and S.H. Elsea, *Refinement of the Smith-Magenis syndrome critical region to approximately 950kb and assessment of 17p11.2*

- deletions. *Are all deletions created equally?* Mol Genet Metab, 2003. **79**(2): p. 134-41.
16. Potocki, L., et al., *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, 2007. **80**(4): p. 633-49.
 17. Inlow, J.K. and L.L. Restifo, *Molecular and comparative genetics of mental retardation*. Genetics, 2004. **166**(2): p. 835-81.
 18. Chelly, J., et al., *Genetics and pathophysiology of mental retardation*. Eur J Hum Genet, 2006. **14**(6): p. 701-13.
 19. Pickering, D.L., et al., *Array-based comparative genomic hybridization analysis of 1176 consecutive clinical genetics investigations*. Genet Med, 2008. **10**(4): p. 262-6.
 20. Manolakos, E., et al., *The use of array-CGH in a cohort of Greek children with developmental delay*. Mol Cytogenet, 2010. **3**: p. 22.
 21. Illingworth, R.S. and A.P. Bird, *CpG islands--'a rough guide'*. FEBS Lett, 2009. **583**(11): p. 1713-20.
 22. Bird, A., *DNA methylation patterns and epigenetic memory*. Genes Dev, 2002. **16**(1): p. 6-21.
 23. Laird, P.W., *Principles and challenges of genome-wide DNA methylation analysis*. Nat Rev Genet, 2010. **11**(3): p. 191-203.
 24. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics*. Nat Rev Genet, 2008. **9**(6): p. 465-76.
 25. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
 26. Yoo, A.S. and G.R. Crabtree, *ATP-dependent chromatin remodeling in neural development*. Curr Opin Neurobiol, 2009. **19**(2): p. 120-6.
 27. Vaissiere, T., C. Sawan, and Z. Herceg, *Epigenetic interplay between histone modifications and DNA methylation in gene silencing*. Mutat Res, 2008. **659**(1-2): p. 40-8.
 28. Riccio, A., *Dynamic epigenetic regulation in neurons: enzymes, stimuli and signaling pathways*. Nat Neurosci, 2010. **13**(11): p. 1330-7.
 29. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
 30. Kramer, J.M. and H. van Bokhoven, *Genetic and epigenetic defects in mental retardation*. Int J Biochem Cell Biol, 2009. **41**(1): p. 96-107.
 31. Urdinguio, R.G., J.V. Sanchez-Mut, and M. Esteller, *Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies*. Lancet Neurol, 2009. **8**(11): p. 1056-72.
 32. Graff, J. and I.M. Mansuy, *Epigenetic dysregulation in cognitive disorders*. Eur J Neurosci, 2009. **30**(1): p. 1-8.

33. Gavin, D.P. and R.P. Sharma, *Histone modifications, DNA methylation, and schizophrenia*. *Neurosci Biobehav Rev*, 2010. **34**(6): p. 882-8.
34. MacDonald, J.L. and A.J. Roskams, *Epigenetic regulation of nervous system development by DNA methylation and histone deacetylation*. *Prog Neurobiol*, 2009. **88**(3): p. 170-83.
35. Esteller, M., *Rett syndrome: the first forty years: 1966-2006*. *Epigenetics*, 2007. **2**(1): p. 1.
36. Amir, R.E., et al., *Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes*. *Ann Neurol*, 2000. **47**(5): p. 670-9.
37. Gonzales, M.L. and J.M. LaSalle, *The role of MeCP2 in brain development and neurodevelopmental disorders*. *Curr Psychiatry Rep*, 2010. **12**(2): p. 127-34.
38. Taine, L., et al., *Submicroscopic deletion of chromosome 16p13.3 in patients with Rubinstein-Taybi syndrome*. *Am J Med Genet*, 1998. **78**(3): p. 267-70.
39. Trivier, E., et al., *Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome*. *Nature*, 1996. **384**(6609): p. 567-70.
40. Xu, G.L., et al., *Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene*. *Nature*, 1999. **402**(6758): p. 187-91.
41. Xue, Y., et al., *The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies*. *Proc Natl Acad Sci U S A*, 2003. **100**(19): p. 10635-40.
42. Vissers, L.E., et al., *Mutations in a new member of the chromodomain gene family cause CHARGE syndrome*. *Nat Genet*, 2004. **36**(9): p. 955-7.
43. Biliya, S. and L.A. Bulla, Jr., *Genomic imprinting: the influence of differential methylation in the two sexes*. *Exp Biol Med (Maywood)*, 2010. **235**(2): p. 139-47.
44. Knoll, J.H., et al., *Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion*. *Am J Med Genet*, 1989. **32**(2): p. 285-90.
45. Buiting, K., *Prader-Willi syndrome and Angelman syndrome*. *Am J Med Genet C Semin Med Genet*, 2010. **154C**(3): p. 365-76.
46. Bolton, P.F., et al., *The phenotypic manifestations of interstitial duplications of proximal 15q with special reference to the autistic spectrum disorders*. *Am J Med Genet*, 2001. **105**(8): p. 675-85.
47. Dimitropoulos, A. and R.T. Schultz, *Autistic-like symptomatology in Prader-Willi syndrome: a review of recent findings*. *Curr Psychiatry Rep*, 2007. **9**(2): p. 159-64.
48. Piard, J., et al., *Clinical and molecular characterization of a large family with an interstitial 15q11q13 duplication*. *Am J Med Genet A*, 2010. **152A**(8): p. 1933-41.
49. Shinawi, M., et al., *Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size*. *J Med Genet*, 2010. **47**(5): p. 332-41.

50. Phelan, M.C., *Deletion 22q13.3 syndrome*. Orphanet J Rare Dis, 2008. **3**: p. 14.
51. Psoni, S., et al., *Phenotypic and genotypic variability in four males with MECP2 gene sequence aberrations including a novel deletion*. Pediatr Res, 2010. **67**(5): p. 551-6.
52. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*. Cell, 1999. **99**(3): p. 247-57.
53. Kurotaki, N., et al., *Haploinsufficiency of NSD1 causes Sotos syndrome*. Nat Genet, 2002. **30**(4): p. 365-6.
54. Kleefstra, T., et al., *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, 2009. **46**(9): p. 598-606.
55. Petrij, F., et al., *Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmid should be used for microdeletion detection and low number of protein truncating mutations*. J Med Genet, 2000. **37**(3): p. 168-76.
56. Stef, M., et al., *Spectrum of CREBBP gene dosage anomalies in Rubinstein-Taybi syndrome patients*. Eur J Hum Genet, 2007. **15**(8): p. 843-7.
57. Gervasini, C., et al., *High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints*. Genomics, 2007. **90**(5): p. 567-73.
58. Bartsch, O., et al., *Molecular studies in 10 cases of Rubinstein-Taybi syndrome, including a mild variant showing a missense mutation in codon 1175 of CREBBP*. J Med Genet, 2002. **39**(7): p. 496-501.
59. Thienpont, B., et al., *Duplications of the critical Rubinstein-Taybi deletion region on chromosome 16p13.3 cause a novel recognisable syndrome*. J Med Genet, 2010. **47**(3): p. 155-61.
60. Roelfsema, J.H., et al., *Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease*. Am J Hum Genet, 2005. **76**(4): p. 572-80.
61. Delaunoy, J.P., et al., *Identification of novel mutations in the RSK2 gene (RPS6KA3) in patients with Coffin-Lowry syndrome*. Clin Genet, 2006. **70**(2): p. 161-6.
62. Yntema, H.G., et al., *A novel ribosomal S6-kinase (RSK4; RPS6KA6) is commonly deleted in patients with complex X-linked mental retardation*. Genomics, 1999. **62**(3): p. 332-43.
63. Laumonnier, F., et al., *Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate*. J Med Genet, 2005. **42**(10): p. 780-6.
64. Qiao, Y., et al., *Autism-associated familial microdeletion of Xp11.22*. Clin Genet, 2008. **74**(2): p. 134-44.
65. Williams, S.R., et al., *Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems*. Am J Hum Genet, 2010. **87**(2): p. 219-28.

66. Kim, T., et al., *Association of histone deacetylase genes with schizophrenia in Korean population*. Psychiatry Res, 2010. **178**(2): p. 266-9.
67. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2*. Nat Genet, 1999. **23**(2): p. 185-8.
68. Smeets, E., et al., *Rett syndrome in females with CTS hot spot deletions: a disorder profile*. Am J Med Genet A, 2005. **132A**(2): p. 117-20.
69. Schanen, N.C., et al., *Neonatal encephalopathy in two boys in families with recurrent Rett syndrome*. J Child Neurol, 1998. **13**(5): p. 229-31.
70. Hardwick, S.A., et al., *Delineation of large deletions of the MECP2 gene in Rett syndrome patients, including a familial case with a male proband*. Eur J Hum Genet, 2007. **15**(12): p. 1218-29.
71. Carney, R.M., et al., *Identification of MeCP2 mutations in a series of females with autistic disorder*. Pediatr Neurol, 2003. **28**(3): p. 205-11.
72. Couvert, P., et al., *MECP2 is highly mutated in X-linked mental retardation*. Hum Mol Genet, 2001. **10**(9): p. 941-6.
73. Lugtenberg, D., et al., *Structural variation in Xq28: MECP2 duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy*. Eur J Hum Genet, 2009. **17**(4): p. 444-53.
74. Nagarajan, R.P., et al., *Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation*. Epigenetics, 2006. **1**(4): p. e1-11.
75. Ramocki, M.B., et al., *Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome*. Ann Neurol, 2009. **66**(6): p. 771-82.
76. Cohen, D., et al., *MECP2 mutation in a boy with language disorder and schizophrenia*. Am J Psychiatry, 2002. **159**(1): p. 148-9.
77. Watson, P., et al., *Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein*. J Med Genet, 2001. **38**(4): p. 224-8.
78. Gibbons, R.J., et al., *Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome)*. Cell, 1995. **80**(6): p. 837-45.
79. Thienpont, B., et al., *Partial duplications of the ATRX gene cause the ATR-X syndrome*. Eur J Hum Genet, 2007. **15**(10): p. 1094-7.
80. Villard, L., et al., *XNP mutation in a large family with Juberg-Marsidi syndrome*. Nat Genet, 1996. **12**(4): p. 359-60.
81. Abidi, F., et al., *Carpenter-Waziri syndrome results from a mutation in XNP*. Am J Med Genet, 1999. **85**(3): p. 249-51.
82. Villard, L., et al., *Identification of a mutation in the XNP/ATR-X gene in a family reported as Smith-Fineman-Myers syndrome*. Am J Med Genet, 2000. **91**(1): p. 83-5.

83. Lalani, S.R., et al., *Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation*. Am J Hum Genet, 2006. **78**(2): p. 303-14.
84. Jensen, L.R., et al., *Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation*. Am J Hum Genet, 2005. **76**(2): p. 227-36.
85. Abidi, F.E., et al., *Mutations in JARID1C are associated with X-linked mental retardation, short stature and hyperreflexia*. J Med Genet, 2008. **45**(12): p. 787-93.
86. Crawford, J., et al., *Mutation screening in Borjeson-Forssman-Lehmann syndrome: identification of a novel de novo PHF6 mutation in a female patient*. J Med Genet, 2006. **43**(3): p. 238-43.
87. Cacheux, V., et al., *Loss-of-function mutations in SIP1 Smad interacting protein 1 result in a syndromic Hirschsprung disease*. Hum Mol Genet, 2001. **10**(14): p. 1503-10.
88. Zweier, C., et al., *Atypical ZFX1B mutation associated with a mild Mowat-Wilson syndrome phenotype*. Am J Med Genet A, 2006. **140**(8): p. 869-72.
89. Heinritz, W., et al., *A missense mutation in the ZFX1B gene associated with an atypical Mowat-Wilson syndrome phenotype*. Am J Med Genet A, 2006. **140**(11): p. 1223-7.
90. Bahn, S., et al., *Neuronal target genes of the neuron-restrictive silencer factor in neurospheres derived from fetuses with Down's syndrome: a gene expression study*. Lancet, 2002. **359**(9303): p. 310-5.
91. Weaving, L.S., et al., *Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation*. Am J Hum Genet, 2004. **75**(6): p. 1079-93.
92. Bentivegna, A., et al., *Rubinstein-Taybi Syndrome: spectrum of CREBBP mutations in Italian patients*. BMC Med Genet, 2006. **7**: p. 77.
93. Hoheisel, J.D., *Microarray technology: beyond transcript profiling and genotype analysis*. Nat Rev Genet, 2006. **7**(3): p. 200-10.
94. Kallioniemi, A., et al., *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors*. Science, 1992. **258**(5083): p. 818-21.
95. du Manoir, S., et al., *Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization*. Hum Genet, 1993. **90**(6): p. 590-610.
96. Snijders, A.M., D. Pinkel, and D.G. Albertson, *Current status and future prospects of array-based comparative genomic hybridisation*. Brief Funct Genomic Proteomic, 2003. **2**(1): p. 37-45.
97. Pinkel, D., et al., *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays*. Nat Genet, 1998. **20**(2): p. 207-11.

98. Solinas-Toldo, S., et al., *Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances*. Genes Chromosomes Cancer, 1997. **20**(4): p. 399-407.
99. Ginzinger, D.G., *Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream*. Exp Hematol, 2002. **30**(6): p. 503-12.
100. Ishkanian, A.S., et al., *A tiling resolution DNA microarray with complete coverage of the human genome*. Nat Genet, 2004. **36**(3): p. 299-303.
101. Lucito, R., et al., *Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation*. Genome Res, 2003. **13**(10): p. 2291-305.
102. Wang, D.G., et al., *Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome*. Science, 1998. **280**(5366): p. 1077-82.
103. Carvalho, B., et al., *High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides*. J Clin Pathol, 2004. **57**(6): p. 644-6.
104. Pollack, J.R., et al., *Genome-wide analysis of DNA copy-number changes using cDNA microarrays*. Nat Genet, 1999. **23**(1): p. 41-6.
105. Geschwind, D.H., et al., *Klinefelter's syndrome as a model of anomalous cerebral laterality: testing gene dosage in the X chromosome pseudoautosomal region using a DNA microarray*. Dev Genet, 1998. **23**(3): p. 215-29.
106. Snijders, A.M., et al., *Assembly of microarrays for genome-wide measurement of DNA copy number*. Nat Genet, 2001. **29**(3): p. 263-4.
107. Fiegler, H., et al., *DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones*. Genes Chromosomes Cancer, 2003. **36**(4): p. 361-74.
108. Krzywinski, M., et al., *A set of BAC clones spanning the human genome*. Nucleic Acids Res, 2004. **32**(12): p. 3651-60.
109. Peiffer, D.A., et al., *High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping*. Genome Res, 2006. **16**(9): p. 1136-48.
110. Matsuzaki, H., et al., *Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays*. Nat Methods, 2004. **1**(2): p. 109-11.
111. Brennan, C., et al., *High-resolution global profiling of genomic alterations with long oligonucleotide microarray*. Cancer Res, 2004. **64**(14): p. 4744-8.
112. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome*. Nature, 2010. **464**(7289): p. 704-12.
113. Zhao, X., et al., *An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays*. Cancer Res, 2004. **64**(9): p. 3060-71.
114. Redon, R., et al., *Global variation in copy number in the human genome*. Nature, 2006. **444**(7118): p. 444-54.

115. Kennedy, G.C., et al., *Large-scale genotyping of complex DNA*. Nat Biotechnol, 2003. **21**(10): p. 1233-7.
116. Lisitsyn, N., N. Lisitsyn, and M. Wigler, *Cloning the differences between two complex genomes*. Science, 1993. **259**(5097): p. 946-51.
117. Eichler, E.E., R.A. Clark, and X. She, *An assessment of the sequence gaps: unfinished business in a finished human genome*. Nat Rev Genet, 2004. **5**(5): p. 345-54.
118. Sharp, A.J., et al., *Segmental duplications and copy-number variation in the human genome*. Am J Hum Genet, 2005. **77**(1): p. 78-88.
119. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome*. Nat Genet, 2004. **36**(9): p. 949-51.
120. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome*. Science, 2004. **305**(5683): p. 525-8.
121. Bouton, C.M. and J. Pevsner, *DRAGON View: information visualization for annotated microarray data*. Bioinformatics, 2002. **18**(2): p. 323-4.
122. Pinkel, D. and D.G. Albertson, *Array comparative genomic hybridization and its applications in cancer*. Nat Genet, 2005. **37 Suppl**: p. S11-7.
123. Awad, I.A., et al., *Caryoscope: an Open Source Java application for viewing microarray data in a genomic context*. BMC Bioinformatics, 2004. **5**: p. 151.
124. Chen, W., et al., *CGHPRO -- a comprehensive data analysis tool for array CGH*. BMC Bioinformatics, 2005. **6**: p. 85.
125. Chi, B., et al., *SeeGH--a software tool for visualization of whole genome array comparative genomic hybridization data*. BMC Bioinformatics, 2004. **5**: p. 13.
126. Kim, S.Y., et al., *ArrayCyGHt: a web application for analysis and visualization of array-CGH data*. Bioinformatics, 2005. **21**(10): p. 2554-5.
127. Lingjaerde, O.C., et al., *CGH-Explorer: a program for analysis of array-CGH data*. Bioinformatics, 2005. **21**(6): p. 821-2.
128. Menten, B., et al., *arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays*. BMC Bioinformatics, 2005. **6**: p. 124.
129. Autio, R., et al., *CGH-Plotter: MATLAB toolbox for CGH-data analysis*. Bioinformatics, 2003. **19**(13): p. 1714-5.
130. Wang, J., et al., *M-CGH: analysing microarray-based CGH experiments*. BMC Bioinformatics, 2004. **5**: p. 74.
131. Myers, C.L., et al., *Accurate detection of aneuploidies in array CGH and gene expression microarray data*. Bioinformatics, 2004. **20**(18): p. 3533-43.
132. Friedman, J.M., et al., *Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation*. Am J Hum Genet, 2006. **79**(3): p. 500-13.
133. Rauch, A., et al., *Molecular karyotyping using an SNP array for genomewide genotyping*. J Med Genet, 2004. **41**(12): p. 916-22.

134. de Vries, B.B., et al., *Diagnostic genome profiling in mental retardation*. Am J Hum Genet, 2005. **77**(4): p. 606-16.
135. Menten, B., et al., *Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports*. J Med Genet, 2006. **43**(8): p. 625-33.
136. Shaw-Smith, C., et al., *Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features*. J Med Genet, 2004. **41**(4): p. 241-8.
137. Vissers, L.E., et al., *Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities*. Am J Hum Genet, 2003. **73**(6): p. 1261-70.
138. Yu, W., et al., *Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions*. Hum Mol Genet, 2003. **12**(17): p. 2145-52.
139. Bejjani, B.A., et al., *Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more?* Am J Med Genet A, 2005. **134**(3): p. 259-67.
140. Shaffer, L.G., et al., *Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases*. J Pediatr, 2006. **149**(1): p. 98-102.
141. Cheung, S.W., et al., *Development and validation of a CGH microarray for clinical cytogenetic diagnosis*. Genet Med, 2005. **7**(6): p. 422-32.
142. Wong, K.K., et al., *A comprehensive analysis of common copy-number variations in the human genome*. Am J Hum Genet, 2007. **80**(1): p. 91-104.
143. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification*. Nucleic Acids Res, 2002. **30**(12): p. e57.
144. Armour, J.A., et al., *Measurement of locus copy number by hybridisation with amplifiable probes*. Nucleic Acids Res, 2000. **28**(2): p. 605-9.
145. Rooms, L., E. Reyniers, and R.F. Kooy, *Subtelomeric rearrangements in the mentally retarded: a comparison of detection methods*. Hum Mutat, 2005. **25**(6): p. 513-24.
146. Rooms, L., et al., *Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics*. Clin Genet, 2006. **69**(1): p. 58-64.
147. Sellner, L.N. and G.R. Taylor, *MLPA and MAPH: new techniques for detection of gene deletions*. Hum Mutat, 2004. **23**(5): p. 413-9.
148. Pang, A.W., et al., *Towards a comprehensive structural variation map of an individual human genome*. Genome Biol, 2010. **11**(5): p. R52.
149. Gonzalez, E., et al., *The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility*. Science, 2005. **307**(5714): p. 1434-40.

150. McCarroll, S.A., et al., *Common deletion polymorphisms in the human genome*. Nat Genet, 2006. **38**(1): p. 86-92.
151. Conrad, D.F., et al., *A high-resolution survey of deletion polymorphism in the human genome*. Nat Genet, 2006. **38**(1): p. 75-81.
152. Tyson, C., et al., *Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH*. Am J Med Genet A, 2005. **139**(3): p. 173-85.
153. des Portes, V., et al., *Inherited microdeletion in Xp21.3-22.1 involved in non-specific mental retardation*. Clin Genet, 1998. **53**(2): p. 136-41.
154. Burger, J., et al., *Familial interstitial 570 kbp deletion of the UBE3A gene region causing Angelman syndrome but not Prader-Willi syndrome*. Am J Med Genet, 2002. **111**(3): p. 233-7.
155. Lesnik Oberstein, S.A., et al., *Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase*. Am J Hum Genet, 2006. **79**(3): p. 562-6.
156. Petek, E., et al., *Mitotic recombination mediated by the JJAZF1 (KIAA0160) gene causing somatic mosaicism and a new type of constitutional NF1 microdeletion in two children of a mosaic female with only few manifestations*. J Med Genet, 2003. **40**(7): p. 520-5.
157. Locke, D.P., et al., *Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome*. Am J Hum Genet, 2006. **79**(2): p. 275-90.
158. Krepischi-Santos, A.C., et al., *Whole-genome array-CGH screening in undiagnosed syndromic patients: old syndromes revisited and new alterations*. Cytogenet Genome Res, 2006. **115**(3-4): p. 254-61.
159. Niihori, T., et al., *Germline KRAS and BRAF mutations in cardio-facio-cutaneous syndrome*. Nat Genet, 2006. **38**(3): p. 294-6.
160. Down, J., *Observations on an ethnic classification of idiots*. London Hosp. Clin. Lect. Rep. 3:259 only, 1866.
161. Lejeune, J., R. Turpin, and M. Gautier, *[Mongolism; a chromosomal disease (trisomy).]*. Bull Acad Natl Med, 1959. **143**(11-12): p. 256-65.
162. Alagille, D., et al., *Hepatic ductular hypoplasia associated with characteristic facies, vertebral malformations, retarded physical, mental, and sexual development, and cardiac murmur*. J Pediatr, 1975. **86**(1): p. 63-71.
163. Li, L., et al., *Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1*. Nat Genet, 1997. **16**(3): p. 243-51.
164. Oda, T., et al., *Mutations in the human Jagged1 gene are responsible for Alagille syndrome*. Nat Genet, 1997. **16**(3): p. 235-42.
165. Shaw-Smith, C., et al., *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability*. Nat Genet, 2006. **38**(9): p. 1032-7.

166. Sharp, A.J., et al., *Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome*. Nat Genet, 2006. **38**(9): p. 1038-42.
167. Menten, B., et al., *Osteopoikilosis, short stature and mental retardation as key features of a new microdeletion syndrome on 12q14*. J Med Genet, 2007. **44**(4): p. 264-8.
168. Spinner, N.B., et al., *Jagged1 mutations in alagille syndrome*. Hum Mutat, 2001. **17**(1): p. 18-33.
169. Wakui, K., et al., *Construction of a natural panel of 11p11.2 deletions and further delineation of the critical region involved in Potocki-Shaffer syndrome*. Eur J Hum Genet, 2005. **13**(5): p. 528-40.
170. Fernandez, B.A., et al., *Holoprosencephaly and cleidocranial dysplasia in a patient due to two position-effect mutations: case report and review of the literature*. Clin Genet, 2005. **68**(4): p. 349-59.
171. Camprubi, C., et al., *Imprinting center analysis in Prader-Willi and Angelman syndrome patients with typical and atypical phenotypes*. Eur J Med Genet, 2007. **50**(1): p. 11-20.
172. Archer, H.L., et al., *Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients*. J Med Genet, 2006. **43**(5): p. 451-6.
173. Bienvenu, T. and J. Chelly, *Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized*. Nat Rev Genet, 2006. **7**(6): p. 415-26.
174. Zahir, F., et al., *Novel deletions of 14q11.2 associated with developmental delay, cognitive impairment and similar minor anomalies in three children*. J Med Genet, 2007. **44**(9): p. 556-61.
175. Zahir, F.R., et al., *A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha*. J Med Genet, 2008. **45**(4): p. 239-43.
176. Zahir, F.R., et al., *A novel de novo 1.1 Mb duplication of 17q21.33 associated with cognitive impairment and other anomalies*. Am J Med Genet A, 2009. **149A**(6): p. 1257-62.
177. Lehman, A.M., et al., *A characteristic syndrome associated with microduplication of 8q12, inclusive of CHD7*. Eur J Med Genet, 2009. **52**(6): p. 436-9.
178. Zahir, F. and J.M. Friedman, *The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility*. Clin Genet, 2007. **72**(4): p. 271-87.
179. Zahir, F.R. and C.J. Brown, *Epigenetic impacts on neurodevelopment: pathophysiological mechanisms and genetic modes of action*. Pediatr Res, 2011.
180. Rosenberg, C., et al., *Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents*. J Med Genet, 2006. **43**(2): p. 180-6.

181. Zannolli, R., et al., *Corpus callosum agenesis, multiple cysts, skin defects, and subtle ocular abnormalities with a de novo mutation [45,XX,der(5), t(5;;14) (pter;q11.2)]*. Am J Med Genet, 2001. **102**(1): p. 29-35.
182. Kamnasaran, D., et al., *Defining the breakpoints of proximal chromosome 14q rearrangements in nine patients using flow-sorted chromosomes*. Am J Med Genet, 2001. **102**(2): p. 173-82.
183. Shapira, S.K., et al., *De novo proximal interstitial deletions of 14q: cytogenetic and molecular investigations*. Am J Med Genet, 1994. **52**(1): p. 44-50.
184. Levin SW, S.R., *Holoprosencephaly associated with 46,XX,del(14)(q11.2q13)*. Am J Hum Genet Supplement, 1991. **49**(Supplement): p. 269.
185. Bruyere, H., et al., *De novo interstitial proximal deletion of 14q and prenatal diagnosis of holoprosencephaly*. Prenat Diagn, 1996. **16**(11): p. 1059-60.
186. Govaerts, L., et al., *Another patient with a deletion 14q11.2q13*. Ann Genet, 1996. **39**(4): p. 197-200.
187. Grammatico, P., et al., *First case of deletion 14q11.2q13: clinical phenotype*. Ann Genet, 1994. **37**(1): p. 30-2.
188. Ramelli, G.P., et al., *Abnormal myelination in a patient with deletion 14q11.2q13.1*. Pediatr Neurol, 2000. **23**(2): p. 170-2.
189. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. Nucleic Acids Res, 2005. **33**(Database issue): p. D501-4.
190. Belotserkovskaya, R., et al., *FACT facilitates transcription-dependent nucleosome alteration*. Science, 2003. **301**(5636): p. 1090-3.
191. Orphanides, G., et al., *FACT, a factor that facilitates transcript elongation through nucleosomes*. Cell, 1998. **92**(1): p. 105-16.
192. Mason, P.B. and K. Struhl, *The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo*. Mol Cell Biol, 2003. **23**(22): p. 8323-33.
193. Kaplan, C.D., L. Laprade, and F. Winston, *Transcription elongation factors repress transcription initiation from cryptic sites*. Science, 2003. **301**(5636): p. 1096-9.
194. Reinberg, D. and R.J. Sims, 3rd, *de FACTo nucleosome dynamics*. J Biol Chem, 2006. **281**(33): p. 23297-301.
195. Costa, P.J. and K.M. Arndt, *Synthetic lethal interactions suggest a role for the Saccharomyces cerevisiae Rtf1 protein in transcription elongation*. Genetics, 2000. **156**(2): p. 535-47.
196. Squazzo, S.L., et al., *The Paf1 complex physically and functionally associates with transcription elongation factors in vivo*. Embo J, 2002. **21**(7): p. 1764-74.
197. Rosonina, E. and J.L. Manley, *From transcription to mRNA: PAF provides a new path*. Mol Cell, 2005. **20**(2): p. 167-8.

198. Ishihara, K., M. Oshimura, and M. Nakao, *CTCF-dependent chromatin insulator is linked to epigenetic remodeling*. Mol Cell, 2006. **23**(5): p. 733-42.
199. Jongmans, M.C., et al., *CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene*. J Med Genet, 2006. **43**(4): p. 306-14.
200. Takahashi, K., et al., *Ndr2 promotes neurite outgrowth of NGF-differentiated PC12 cells*. Neurosci Lett, 2005. **388**(3): p. 157-62.
201. Nagase, T., et al., *Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro*. DNA Res, 1997. **4**(2): p. 141-50.
202. Zhang, L., et al., *Patterns of segmental duplication in the human genome*. Mol Biol Evol, 2005. **22**(1): p. 135-41.
203. Lupski, J.R. and P. Stankiewicz, *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes*. PLoS Genet, 2005. **1**(6): p. e49.
204. Cleary, J.D. and C.E. Pearson, *Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability*. Trends Genet, 2005. **21**(5): p. 272-80.
205. Tabuchi, K. and T.C. Sudhof, *Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing*. Genomics, 2002. **79**(6): p. 849-59.
206. Missler, M., R. Fernandez-Chacon, and T.C. Sudhof, *The making of neurexins*. J Neurochem, 1998. **71**(4): p. 1339-47.
207. Garner, C.C., et al., *Molecular mechanisms of CNS synaptogenesis*. Trends Neurosci, 2002. **25**(5): p. 243-51.
208. Dean, C. and T. Dresbach, *Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function*. Trends Neurosci, 2006. **29**(1): p. 21-9.
209. Feng, J., et al., *High frequency of neurexin 1beta signal peptide structural variants in patients with autism*. Neurosci Lett, 2006. **409**(1): p. 10-3.
210. Szatmari, P., et al., *Mapping autism risk loci using genetic linkage and chromosomal rearrangements*. Nat Genet, 2007. **39**(3): p. 319-28.
211. Berument, S.K., et al., *Autism screening questionnaire: diagnostic validity*. Br J Psychiatry, 1999. **175**: p. 444-51.
212. Ehlers, S., C. Gillberg, and L. Wing, *A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children*. J Autism Dev Disord, 1999. **29**(2): p. 129-41.
213. Geppert, M., et al., *Neurexin I alpha is a major alpha-latrotoxin receptor that cooperates in alpha-latrotoxin action*. J Biol Chem, 1998. **273**(3): p. 1705-10.
214. Rowen, L., et al., *Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity*. Genomics, 2002. **79**(4): p. 587-97.
215. Patzke, H. and U. Ernsberger, *Expression of neurexin Ialpha splice variants in sympathetic neurons: selective changes during differentiation and in response to neurotrophins*. Mol Cell Neurosci, 2000. **15**(6): p. 561-72.

216. Ullrich, B., Y.A. Ushkaryov, and T.C. Sudhof, *Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons*. Neuron, 1995. **14**(3): p. 497-507.
217. Sugita, S., et al., *A stoichiometric complex of neurexins and dystroglycan in brain*. J Cell Biol, 2001. **154**(2): p. 435-45.
218. Ushkaryov, Y.A., et al., *Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences in receptor-like neuronal cell-surface proteins*. J Biol Chem, 1994. **269**(16): p. 11987-92.
219. Fryns, J.P., P. De Waele, and H. Van Den Berghe, *Interstitial deletion of the short arm of chromosome 2 in a moderately mentally retarded boy without gross clinical stigmata*. Hum Genet, 1979. **51**(2): p. 123-5.
220. Boucard, A.A., et al., *A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins*. Neuron, 2005. **48**(2): p. 229-36.
221. Graf, E.R., et al., *Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins*. Cell, 2004. **119**(7): p. 1013-26.
222. Levi, S., et al., *Dystroglycan is selectively associated with inhibitory GABAergic synapses but is dispensable for their differentiation*. J Neurosci, 2002. **22**(11): p. 4274-85.
223. Beglopoulos, V., et al., *Neurexophilin 3 is highly localized in cortical and cerebellar regions and is functionally important for sensorimotor gating and motor coordination*. Mol Cell Biol, 2005. **25**(16): p. 7278-88.
224. Craig, A.M. and Y. Kang, *Neurexin-neuroligin signaling in synapse development*. Curr Opin Neurobiol, 2007. **17**(1): p. 43-52.
225. Brodtkin, E.S., *BALB/c mice: low sociability and other phenotypes that may be relevant to autism*. Behav Brain Res, 2007. **176**(1): p. 53-65.
226. Missler, M., et al., *Alpha-neurexins couple Ca²⁺ channels to synaptic vesicle exocytosis*. Nature, 2003. **423**(6943): p. 939-48.
227. Kattenstroth, G., et al., *Postsynaptic N-methyl-D-aspartate receptor function requires alpha-neurexins*. Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2607-12.
228. Sons, M.S., et al., *alpha-Neurexins are required for efficient transmitter release and synaptic homeostasis at the mouse neuromuscular junction*. Neuroscience, 2006. **138**(2): p. 433-46.
229. Zhang, W., et al., *Extracellular domains of alpha-neurexins participate in regulating synaptic transmission by selectively affecting N- and P/Q-type Ca²⁺ channels*. J Neurosci, 2005. **25**(17): p. 4330-42.
230. Dudanova, I., et al., *Important contribution of alpha-neurexins to Ca²⁺-triggered exocytosis of secretory granules*. J Neurosci, 2006. **26**(41): p. 10599-613.
231. Zeng, Z., et al., *The expression and alternative splicing of alpha-neurexins during Xenopus development*. Int J Dev Biol, 2006. **50**(1): p. 39-46.

232. Zeng, X., et al., *Neurexin-1 is required for synapse formation and larvae associative learning in Drosophila*. FEBS Lett, 2007. **581**(13): p. 2509-16.
233. Friedman, J.M., et al., *Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation*. Am J Hum Genet, 2006. **79**(3): p. 500-13.
234. Kirchhoff, M., et al., *A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features*. Eur J Med Genet, 2007. **50**(4): p. 256-63.
235. Leana-Cox, J., et al., *Characterization of de novo duplications in eight patients by using fluorescence in situ hybridization with chromosome-specific DNA libraries*. Am J Hum Genet, 1993. **52**(6): p. 1067-73.
236. Levy, S., et al., *The diploid genome sequence of an individual human*. PLoS Biol, 2007. **5**(10): p. e254.
237. Thienpont, B., et al., *A microduplication of CBP in a patient with mental retardation and a congenital heart defect*. Am J Med Genet A, 2007. **143A**(18): p. 2160-4.
238. Allen, P.B., C.C. Ouimet, and P. Greengard, *Spinophilin, a novel protein phosphatase 1 binding protein localized to dendritic spines*. Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9956-61.
239. Feng, J., et al., *Spinophilin regulates the formation and function of dendritic spines*. Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9287-92.
240. Bittel, D.C., N. Kibiryeve, and M.G. Butler, *Whole genome microarray analysis of gene expression in subjects with fragile X syndrome*. Genet Med, 2007. **9**(7): p. 464-72.
241. D'Agata, V., et al., *Gene expression profiles in a transgenic animal model of fragile X syndrome*. Neurobiol Dis, 2002. **10**(3): p. 211-8.
242. Kielty CM, G.M., *Collagen: The Collagen family; structure, assembly and organization in the extracellular matrix*, in *Connective tissue and its heritable diseases*. 1993, Wiley-Liss: New York. p. 103-148.
243. Karsenty, G. and R.W. Park, *Regulation of type I collagen genes expression*. Int Rev Immunol, 1995. **12**(2-4): p. 177-85.
244. Kikuchi, K., et al., *Direct demonstration of transcriptional activation of collagen gene expression in systemic sclerosis fibroblasts: insensitivity to TGF beta 1 stimulation*. Biochem Biophys Res Commun, 1992. **187**(1): p. 45-50.
245. LeRoy, E.C., *Increased collagen synthesis by scleroderma skin fibroblasts in vitro: a possible defect in the regulation or activation of the scleroderma fibroblast*. J Clin Invest, 1974. **54**(4): p. 880-9.
246. Barisic-Dujmovic, T., I. Boban, and S.H. Clark, *Regulation of collagen gene expression in the Tsk2 mouse*. J Cell Physiol, 2008. **215**(2): p. 464-71.
247. Shibusawa, Y., et al., *Mouse model of dermal fibrosis induced by one-time injection of bleomycin-poly(L-lactic acid) microspheres*. Rheumatology (Oxford), 2008. **47**(4): p. 454-7.

248. Nuytinck, L., et al., *Classical Ehlers-Danlos syndrome caused by a mutation in type I collagen*. Am J Hum Genet, 2000. **66**(4): p. 1398-402.
249. Martin, E. and J.R. Shapiro, *Osteogenesis imperfecta: epidemiology and pathophysiology*. Curr Osteoporos Rep, 2007. **5**(3): p. 91-7.
250. Lee, J.A., C.M. Carvalho, and J.R. Lupski, *A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders*. Cell, 2007. **131**(7): p. 1235-47.
251. Lee, C., A.J. Iafrate, and A.R. Brothman, *Copy number variations and clinical cytogenetic diagnosis of constitutional disorders*. Nat Genet, 2007. **39**(7 Suppl): p. S48-54.
252. Perry, G.H., et al., *The fine-scale and complex architecture of human copy-number variation*. Am J Hum Genet, 2008. **82**(3): p. 685-95.
253. Rodriguez-Revena, L., et al., *Structural variation in the human genome: the impact of copy number variants on clinical diagnosis*. Genet Med, 2007. **9**(9): p. 600-6.
254. Conrad, D.F. and M.E. Hurles, *The population genetics of structural variation*. Nat Genet, 2007. **39**(7 Suppl): p. S30-6.
255. Lee, J.A., et al., *Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease*. Hum Mol Genet, 2006. **15**(14): p. 2250-65.
256. Smyk, M., et al., *Different-sized duplications of Xq28, including MECP2, in three males with mental retardation, absent or delayed speech, and recurrent infections*. Am J Med Genet B Neuropsychiatr Genet, 2008. **147B**(6): p. 799-806.
257. Beanan, M.J. and T.D. Sargent, *Regulation and function of Dlx3 in vertebrate development*. Dev Dyn, 2000. **218**(4): p. 545-53.
258. Depew, M.J., et al., *Reassessing the Dlx code: the genetic regulation of branchial arch skeletal pattern and development*. J Anat, 2005. **207**(5): p. 501-61.
259. Murthi, P., et al., *Homeobox gene DLX4 expression is increased in idiopathic human fetal growth restriction*. Mol Hum Reprod, 2006. **12**(12): p. 763-9.
260. Lyng, H., et al., *Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer*. BMC Genomics, 2006. **7**: p. 268.
261. Bhat, S.S., et al., *Disruption of DMD and deletion of ACSL4 causing developmental delay, hypotonia, and multiple congenital anomalies*. Cytogenet Genome Res, 2006. **112**(1-2): p. 170-5.
262. Slavotinek, A.M., *Novel microdeletion syndromes detected by chromosome microarrays*. Hum Genet, 2008. **124**(1): p. 1-17.
263. Hennekam, R.C., *Rubinstein-Taybi syndrome: a history in pictures*. Clin Dysmorphol, 1993. **2**(1): p. 87-92.
264. Hannes, F.D., et al., *Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant*. J Med Genet, 2009. **46**(4): p. 223-32.

265. Marangi, G., et al., *Duplication of the Rubinstein-Taybi region on 16p13.3 is associated with a distinctive phenotype*. Am J Med Genet A, 2008. **146A**(18): p. 2313-7.
266. de Ravel, T., et al., *Trisomy of chromosome 16p13.3 due to an unbalanced insertional translocation into chromosome 22p13*. Eur J Med Genet, 2005. **48**(3): p. 355-9.
267. Gu, W., F. Zhang, and J.R. Lupski, *Mechanisms for human genomic rearrangements*. Pathogenetics, 2008. **1**(1): p. 4.
268. Zogopoulos, G., et al., *Germ-line DNA copy number variation frequencies in a large North American population*. Hum Genet, 2007. **122**(3-4): p. 345-53.
269. Itsara, A., et al., *Population analysis of large copy number variants and hotspots of human genetic disease*. Am J Hum Genet, 2009. **84**(2): p. 148-61.
270. Petrij, F., et al., *Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP*. Nature, 1995. **376**(6538): p. 348-51.
271. Digilio, M.C., et al., *16p subtelomeric duplication: a clinically recognizable syndrome*. Eur J Hum Genet, 2009. **17**(9): p. 1135-40.
272. Somerville, M.J., et al., *Severe expressive-language delay related to duplication of the Williams-Beuren locus*. N Engl J Med, 2005. **353**(16): p. 1694-701.
273. Portnoi, M.F., *Microduplication 22q11.2: a new chromosomal syndrome*. Eur J Med Genet, 2009. **52**(2-3): p. 88-93.
274. Ou, Z., et al., *Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes*. Genet Med, 2008. **10**(4): p. 267-77.
275. Van Esch, H., et al., *Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males*. Am J Hum Genet, 2005. **77**(3): p. 442-53.
276. Bi, W., et al., *Increased LIS1 expression affects human and mouse brain development*. Nat Genet, 2009. **41**(2): p. 168-77.
277. Chance, P.F., et al., *DNA deletion associated with hereditary neuropathy with liability to pressure palsies*. Cell, 1993. **72**(1): p. 143-51.
278. Lupski, J.R., et al., *DNA duplication associated with Charcot-Marie-Tooth disease type 1A*. Cell, 1991. **66**(2): p. 219-32.
279. Hennekam, R.C., *Rubinstein-Taybi syndrome*. Eur J Hum Genet, 2006. **14**(9): p. 981-5.
280. Kazantsev, A.G. and L.M. Thompson, *Therapeutic application of histone deacetylase inhibitors for central nervous system disorders*. Nat Rev Drug Discov, 2008. **7**(10): p. 854-68.
281. Balasubramanyam, K., et al., *Curcumin, a novel p300/CREB-binding protein-specific inhibitor of acetyltransferase, represses the acetylation of histone/nonhistone proteins and histone acetyltransferase-dependent chromatin transcription*. J Biol Chem, 2004. **279**(49): p. 51163-71.

282. Strimpakos, A.S. and R.A. Sharma, *Curcumin: preventive and therapeutic properties in laboratory studies and clinical trials*. Antioxid Redox Signal, 2008. **10**(3): p. 511-45.
283. Delaney, A.D., et al., *Use of Affymetrix mapping arrays in the diagnosis of gene copy number variation*. Curr Protoc Hum Genet, 2008. **Chapter 8**: p. Unit 8 13.
284. Ruangvutilert, P., et al., *Relative efficiency of FISH on metaphase and interphase nuclei from non-mosaic trisomic or triploid fibroblast cultures*. Prenat Diagn, 2000. **20**(2): p. 159-62.
285. Monfort, S., et al., *Detection of known and novel genomic rearrangements by array based comparative genomic hybridisation: deletion of ZNF533 and duplication of CHARGE syndrome genes*. J Med Genet, 2008. **45**(7): p. 432-7.
286. Miller, N.R., *Walsh and Hoyt's Clinical Neuro-Ophthalmology 4th Edition*. 4 ed, ed. M. NR. Vol. 2. 1982, Baltimore: Williams & Wilkins.
287. Vincent, C., et al., *A proposed new contiguous gene syndrome on 8q consists of Branchio-Oto-Renal (BOR) syndrome, Duane syndrome, a dominant form of hydrocephalus and trapeze aplasia; implications for the mapping of the BOR gene*. Hum Mol Genet, 1994. **3**(10): p. 1859-66.
288. Calabrese, G., et al., *Narrowing the Duane syndrome critical region at chromosome 8q13 down to 40 kb*. Eur J Hum Genet, 2000. **8**(5): p. 319-24.
289. Pizzuti, A., et al., *A peptidase gene in chromosome 8q is disrupted by a balanced translocation in a duane syndrome patient*. Invest Ophthalmol Vis Sci, 2002. **43**(12): p. 3609-12.
290. Cahan, P., et al., *The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells*. Nat Genet, 2009. **41**(4): p. 430-7.
291. Zheng, Y., et al., *The shortened cochlea: its classification and histopathologic features*. Int J Pediatr Otorhinolaryngol, 2002. **63**(1): p. 29-39.
292. Griffith, M., et al., *ALEXA: a microarray design platform for alternative expression analysis*. Nat Methods, 2008. **5**(2): p. 118.
293. Jiang, Z., et al., *DupMasker: a tool for annotating primate segmental duplications*. Genome Res, 2008. **18**(8): p. 1362-8.
294. Breslauer, K.J., et al., *Predicting DNA duplex stability from the base sequence*. Proc Natl Acad Sci U S A, 1986. **83**(11): p. 3746-50.
295. Sugimoto, N., et al., *Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes*. Nucleic Acids Res, 1996. **24**(22): p. 4501-5.
296. Andronescu, M., et al., *A new algorithm for RNA secondary structure design*. J Mol Biol, 2004. **336**(3): p. 607-24.
297. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data*. Biostatistics, 2004. **5**(4): p. 557-72.
298. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction*. Nat Protoc, 2010. **5**(4): p. 725-38.

299. Zhang, Y., *I-TASSER server for protein 3D structure prediction*. BMC Bioinformatics, 2008. **9**: p. 40.
300. Yu, S., et al., *Quantitative real-time polymerase chain reaction for the verification of genomic imbalances detected by microarray-based comparative genomic hybridization*. Genet Test Mol Biomarkers, 2009. **13**(6): p. 751-60.
301. Ko, M., et al., *Chromatin remodeling, development and disease*. Mutat Res, 2008. **647**(1-2): p. 59-67.
302. Wu, J.I., et al., *Regulation of dendritic development by neuron-specific chromatin remodeling complexes*. Neuron, 2007. **56**(1): p. 94-108.
303. Koga, M., et al., *Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia*. Hum Mol Genet, 2009. **18**(13): p. 2483-94.
304. Walsh, T., et al., *Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia*. Science, 2008. **320**(5875): p. 539-43.
305. Harikrishnan, K.N., et al., *Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing*. Nat Genet, 2005. **37**(3): p. 254-64.
306. Vecsler, M., et al., *MeCP2 deficiency downregulates specific nuclear proteins that could be partially recovered by valproic acid in vitro*. Epigenetics, 2010. **5**(1): p. 61-7.
307. Ehrlich, M., et al., *ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation*. Autoimmunity, 2008. **41**(4): p. 253-71.
308. Bergren, S.K., E.D. Rutter, and J.A. Kearney, *Fine mapping of an epilepsy modifier gene on mouse Chromosome 19*. Mamm Genome, 2009. **20**(6): p. 359-66.
309. Le Meur, N., et al., *MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations*. J Med Genet, 2010. **47**(1): p. 22-9.
310. Potthoff, M.J. and E.N. Olson, *MEF2: a central regulator of diverse developmental programs*. Development, 2007. **134**(23): p. 4131-40.
311. Black, B.L. and E.N. Olson, *Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins*. Annu Rev Cell Dev Biol, 1998. **14**: p. 167-96.
312. Lyons, G.E., et al., *Expression of mef2 genes in the mouse central nervous system suggests a role in neuronal maturation*. J Neurosci, 1995. **15**(8): p. 5727-38.
313. McKinsey, T.A., C.L. Zhang, and E.N. Olson, *Activation of the myocyte enhancer factor-2 transcription factor by calcium/calmodulin-dependent protein kinase-stimulated binding of 14-3-3 to histone deacetylase 5*. Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14400-5.
314. McKinsey, T.A., et al., *Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation*. Nature, 2000. **408**(6808): p. 106-11.

315. McKinsey, T.A., C.L. Zhang, and E.N. Olson, *Control of muscle development by dueling HATs and HDACs*. Curr Opin Genet Dev, 2001. **11**(5): p. 497-504.
316. Haberland, M., et al., *Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation*. Mol Cell Biol, 2007. **27**(2): p. 518-25.
317. Leifer, D., J. Golden, and N.W. Kowall, *Myocyte-specific enhancer binding factor 2C expression in human brain development*. Neuroscience, 1994. **63**(4): p. 1067-79.
318. Tucker, T., et al., *Comparison of genome-wide array genomic hybridization platforms for the detection of copy number variants in idiopathic mental retardation*. BMC Med Genomics, 2011. **4**: p. 25.
319. Schuster, E.F. and R. Stoger, *CHD5 defines a new subfamily of chromodomain-SWI2/SNF2-like helicases*. Mamm Genome, 2002. **13**(2): p. 117-9.
320. Yamada, K., et al., *Characterization of a de novo balanced t(4;20)(q33;q12) translocation in a patient with mental retardation*. Am J Med Genet A, 2010. **152A**(12): p. 3057-67.
321. Kalscheuer, V.M., et al., *Disruption of the TCF4 gene in a girl with mental retardation but without the classical Pitt-Hopkins syndrome*. Am J Med Genet A, 2008. **146A**(16): p. 2053-9.
322. Amiel, J., et al., *Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction*. Am J Hum Genet, 2007. **80**(5): p. 988-93.
323. de Pontual, L., et al., *Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome*. Hum Mutat, 2009. **30**(4): p. 669-76.
324. Brown, E., S. Malakar, and J.E. Krebs, *How many remodelers does it take to make a brain? Diverse and cooperative roles of ATP-dependent chromatin-remodeling complexes in development*. Biochem Cell Biol, 2007. **85**(4): p. 444-62.
325. Lutz, T., R. Stoger, and A. Nieto, *CHD6 is a DNA-dependent ATPase and localizes at nuclear sites of mRNA synthesis*. FEBS Lett, 2006. **580**(25): p. 5851-7.
326. Lathrop, M.J., et al., *Deletion of the Chd6 exon 12 affects motor coordination*. Mamm Genome, 2010. **21**(3-4): p. 130-42.
327. Pampal, A., *CHARGE: an association or a syndrome?* Int J Pediatr Otorhinolaryngol, 2010. **74**(7): p. 719-22.
328. Feigl, B., et al., *Cataract as an additional sign in CHARGE syndrome*. J Pediatr Ophthalmol Strabismus, 2000. **37**(2): p. 111-3.
329. McMain, K., et al., *Ocular features of CHARGE syndrome*. J AAPOS, 2008. **12**(5): p. 460-5.
330. Nord, A.S., et al., *Reduced transcript expression of genes affected by inherited and de novo CNVs in autism*. Eur J Hum Genet, 2011.

331. Park, H., et al., *Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing*. Nat Genet, 2010. **42**(5): p. 400-5.
332. Yan, Z., et al., *PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes*. Genes Dev, 2005. **19**(14): p. 1662-7.
333. Kwon, C.S. and D. Wagner, *Unwinding chromatin for development and growth: a few genes at a time*. Trends Genet, 2007. **23**(8): p. 403-12.
334. Vignali, M., et al., *ATP-dependent chromatin-remodeling complexes*. Mol Cell Biol, 2000. **20**(6): p. 1899-910.
335. Mohrmann, L. and C.P. Verrijzer, *Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes*. Biochim Biophys Acta, 2005. **1681**(2-3): p. 59-73.
336. Siegrist, F., M. Ebeling, and U. Certa, *The small interferon-induced transmembrane genes and proteins*. J Interferon Cytokine Res, 2011. **31**(1): p. 183-97.
337. Yang, G., et al., *IFITM1 plays an essential role in the antiproliferative action of interferon-gamma*. Oncogene, 2007. **26**(4): p. 594-603.
338. Xu, Y., G. Yang, and G. Hu, *Binding of IFITM1 enhances the inhibiting effect of caveolin-1 on ERK activation*. Acta Biochim Biophys Sin (Shanghai), 2009. **41**(6): p. 488-94.
339. Tanaka, S.S., et al., *IFITM/Mil/fragilis family proteins IFITM1 and IFITM3 play distinct roles in mouse primordial germ cell homing and repulsion*. Dev Cell, 2005. **9**(6): p. 745-56.
340. Kainou, T., et al., *Identification of the GGPS1 genes encoding geranylgeranyl diphosphate synthases from mouse and human*. Biochim Biophys Acta, 1999. **1437**(3): p. 333-40.
341. Hooff, G.P., et al., *Isoprenoids, small GTPases and Alzheimer's disease*. Biochim Biophys Acta, 2010. **1801**(8): p. 896-905.
342. Hall, A., *Rho GTPases and the control of cell behaviour*. Biochem Soc Trans, 2005. **33**(Pt 5): p. 891-5.
343. Fleischer, T.C., U.J. Yun, and D.E. Ayer, *Identification and characterization of three new components of the mSin3A corepressor complex*. Mol Cell Biol, 2003. **23**(10): p. 3456-67.
344. Wu, M.Y., T.F. Tsai, and A.L. Beaudet, *Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain*. Genes Dev, 2006. **20**(20): p. 2859-70.
345. Hogart, A., et al., *The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13*. Neurobiol Dis, 2010. **38**(2): p. 181-91.

346. Shaikh, T.H., et al., *High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications*. Genome Res, 2009. **19**(9): p. 1682-90.
347. Yamane, K., et al., *JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor*. Cell, 2006. **125**(3): p. 483-95.
348. Loh, Y.H., et al., *Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells*. Genes Dev, 2007. **21**(20): p. 2545-57.
349. Ko, S.Y., et al., *Identification of Jmjd1a as a STAT3 downstream gene in mES cells*. Cell Struct Funct, 2006. **31**(2): p. 53-62.
350. Tateishi, K., et al., *Role of Jhdm2a in regulating metabolic gene expression and obesity resistance*. Nature, 2009. **458**(7239): p. 757-61.
351. Inagaki, T., et al., *Obesity and metabolic syndrome in histone demethylase JHDM2a-deficient mice*. Genes Cells, 2009. **14**(8): p. 991-1001.
352. de Leeuw, N., et al., *UBE2A deficiency syndrome: Mild to severe intellectual disability accompanied by seizures, absent speech, urogenital, and skin anomalies in male patients*. Am J Med Genet A, 2010. **152A**(12): p. 3084-90.
353. Nascimento, R.M., et al., *UBE2A, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome*. Am J Hum Genet, 2006. **79**(3): p. 549-55.
354. Kim, S.M., et al., *Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance*. Nucleic Acids Res, 2010. **38**(19): p. 6389-403.
355. Wolf, S.S., V.K. Patchev, and M. Obendorf, *A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR*. Arch Biochem Biophys, 2007. **460**(1): p. 56-66.
356. Castermans, D., et al., *Identification and characterization of the TRIP8 and REEP3 genes on chromosome 10q21.3 as novel candidate genes for autism*. Eur J Hum Genet, 2007. **15**(4): p. 422-31.
357. Klose, R.J., E.M. Kallin, and Y. Zhang, *JmjC-domain-containing proteins and histone demethylation*. Nat Rev Genet, 2006. **7**(9): p. 715-27.
358. Landeira, D. and A.G. Fisher, *Inactive yet indispensable: the tale of Jarid2*. Trends Cell Biol, 2011. **21**(2): p. 74-80.
359. Takeuchi, T., et al., *Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation*. Genes Dev, 1995. **9**(10): p. 1211-22.
360. Jung, J., M.R. Mysliwiec, and Y. Lee, *Roles of JUMONJI in mouse embryonic development*. Dev Dyn, 2005. **232**(1): p. 21-32.
361. Takahashi, M., et al., *Cardiac abnormalities cause early lethality of jumonji mutant mice*. Biochem Biophys Res Commun, 2004. **324**(4): p. 1319-23.
362. Pasini, D., et al., *JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells*. Nature, 2010. **464**(7286): p. 306-10.

363. Toyoda, M., M. Kojima, and T. Takeuchi, *Jumonji is a nuclear protein that participates in the negative regulation of cell growth*. Biochem Biophys Res Commun, 2000. **274**(2): p. 332-6.
364. Berge-Lefranc, J.L., et al., *Characterization of the human jumonji gene*. Hum Mol Genet, 1996. **5**(10): p. 1637-41.
365. Bovill, E., et al., *Induction by left ventricular overload and left ventricular failure of the human Jumonji gene (JARID2) encoding a protein that regulates transcription and reexpression of a protective fetal program*. J Thorac Cardiovasc Surg, 2008. **136**(3): p. 709-16.
366. Weiss, L.A., et al., *A genome-wide linkage and association scan reveals novel loci for autism*. Nature, 2009. **461**(7265): p. 802-8.
367. Wang, K., et al., *Common genetic variants on 5p14.1 associate with autism spectrum disorders*. Nature, 2009. **459**(7246): p. 528-33.
368. Bureau, A., et al., *Latent class model with familial dependence to address heterogeneity in complex diseases: adapting the approach to family-based association studies*. Genet Epidemiol, 2011. **35**(3): p. 182-9.
369. Pedrosa, E., et al., *Positive association of schizophrenia to JARID2 gene*. Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(1): p. 45-51.
370. Liu, Y., et al., *Whole genome association study in a homogenous population in Shandong peninsula of China reveals JARID2 as a susceptibility gene for schizophrenia*. J Biomed Biotechnol, 2009. **2009**: p. 536918.
371. Scapoli, L., et al., *Expression and association data strongly support JARID2 involvement in nonsyndromic cleft lip with or without cleft palate*. Hum Mutat, 2010. **31**(7): p. 794-800.
372. Scapoli, L., et al., *Evidence of linkage to 6p23 and genetic heterogeneity in nonsyndromic cleft lip with or without cleft palate*. Genomics, 1997. **43**(2): p. 216-20.
373. Volcik, K.A., et al., *Evaluation of the jumonji gene and risk for spina bifida and congenital heart defects*. Am J Med Genet A, 2004. **126A**(2): p. 215-7.
374. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies*. Am J Hum Genet, 2010. **86**(5): p. 749-64.
375. Tayeh, M.K., et al., *Targeted comparative genomic hybridization array for the detection of single- and multiexon gene deletions and duplications*. Genet Med, 2009. **11**(4): p. 232-40.
376. Dhami, P., et al., *Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome*. Am J Hum Genet, 2005. **76**(5): p. 750-62.
377. Wong, L.J., et al., *Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions*. Clin Chem, 2008. **54**(7): p. 1141-8.

378. Saillour, Y., et al., *Detection of exonic copy-number changes using a highly efficient oligonucleotide-based comparative genomic hybridization-array method*. Hum Mutat, 2008. **29**(9): p. 1083-90.
379. Bailey, J.A., J.M. Kidd, and E.E. Eichler, *Human copy number polymorphic genes*. Cytogenet Genome Res, 2008. **123**(1-4): p. 234-43.
380. Craddock, N., et al., *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls*. Nature, 2010. **464**(7289): p. 713-20.
381. Boone, P.M., et al., *Detection of clinically relevant exonic copy-number changes by array CGH*. Hum Mutat, 2010. **31**(12): p. 1326-42.
382. Ching, M.S., et al., *Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders*. Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(4): p. 937-47.
383. Shah, A.K., et al., *Rare NRXN1 promoter variants in patients with schizophrenia*. Neurosci Lett, 2010. **475**(2): p. 80-4.
384. Al-Owain, M., A. Alazami, and F. Alkuraya, *An autosomal recessive syndrome of severe cognitive impairment, dysmorphic facies and skeletal abnormalities maps to the long arm of chromosome 17*. Clin Genet, 2010.
385. Knight, C.G., et al., *Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network*. Nat Genet, 2006. **38**(9): p. 1015-22.
386. Moretti, P. and H.Y. Zoghbi, *MeCP2 dysfunction in Rett syndrome and related disorders*. Curr Opin Genet Dev, 2006. **16**(3): p. 276-81.
387. Chahrour, M. and H.Y. Zoghbi, *The story of Rett syndrome: from clinic to neurobiology*. Neuron, 2007. **56**(3): p. 422-37.
388. Martin, C.L., et al., *Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism*. Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(7): p. 869-76.
389. Bhalla, K., et al., *The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene*. J Hum Genet, 2004. **49**(6): p. 308-11.
390. Verbruggen, K.T., et al., *Global developmental delay in guanidionacetate methyltransferase deficiency: differences in formal testing and clinical observation*. Eur J Pediatr, 2007. **166**(9): p. 921-5.
391. Begni, S., et al., *Association between the G1001C polymorphism in the GRIN1 gene promoter region and schizophrenia*. Biol Psychiatry, 2003. **53**(7): p. 617-9.
392. Rice, S.R., et al., *Identification of single nucleotide polymorphisms (SNPs) and other sequence changes and estimation of nucleotide diversity in coding and flanking regions of the NMDAR1 receptor gene in schizophrenic patients*. Mol Psychiatry, 2001. **6**(3): p. 274-84.

393. Zhao, X., et al., *Significant association between the genetic variations in the 5' end of the N-methyl-D-aspartate receptor subunit gene GRIN1 and schizophrenia*. Biol Psychiatry, 2006. **59**(8): p. 747-53.
394. Rottach, A., H. Leonhardt, and F. Spada, *DNA methylation-mediated epigenetic control*. J Cell Biochem, 2009. **108**(1): p. 43-51.
395. Portela, A. and M. Esteller, *Epigenetic modifications and human disease*. Nat Biotechnol, 2010. **28**(10): p. 1057-68.
396. Easwaran, H.P., et al., *Replication-independent chromatin loading of Dnmt1 during G2 and M phases*. EMBO Rep, 2004. **5**(12): p. 1181-6.
397. Biniszkiwicz, D., et al., *Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality*. Mol Cell Biol, 2002. **22**(7): p. 2124-35.
398. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality*. Cell, 1992. **69**(6): p. 915-26.
399. Rhee, I., et al., *DNMT1 and DNMT3b cooperate to silence genes in human cancer cells*. Nature, 2002. **416**(6880): p. 552-6.
400. Girirajan, S., et al., *A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay*. Nat Genet, 2010. **42**(3): p. 203-9.
401. Coupry, I., et al., *Analysis of CBP (CREBBP) gene deletions in Rubinstein-Taybi syndrome patients using real-time quantitative PCR*. Hum Mutat, 2004. **23**(3): p. 278-84.
402. Bannister, A.J. and T. Kouzarides, *The CBP co-activator is a histone acetyltransferase*. Nature, 1996. **384**(6610): p. 641-3.
403. Burmeister, M., M.G. McInnis, and S. Zollner, *Psychiatric genetics: progress amid controversy*. Nat Rev Genet, 2008. **9**(7): p. 527-40.
404. Inoue, K. and J.R. Lupski, *Genetics and genomics of behavioral and psychiatric disorders*. Curr Opin Genet Dev, 2003. **13**(3): p. 303-9.
405. Singh, S.M. and D. Basu, *The P300 event-related potential and its possible role as an endophenotype for studying substance use disorders: a review*. Addict Biol, 2009. **14**(3): p. 298-309.
406. Sacco, R., et al., *Clinical, morphological, and biochemical correlates of head circumference in autism*. Biol Psychiatry, 2007. **62**(9): p. 1038-47.
407. Aylward, E.H., et al., *Effects of age on brain volume and head circumference in autism*. Neurology, 2002. **59**(2): p. 175-83.
408. Fidler, D.J., J.N. Bailey, and S.L. Smalley, *Macrocephaly in autism and other pervasive developmental disorders*. Dev Med Child Neurol, 2000. **42**(11): p. 737-40.
409. Hazlett, H.C., et al., *Magnetic resonance imaging and head circumference study of brain size in autism: birth through age 2 years*. Arch Gen Psychiatry, 2005. **62**(12): p. 1366-76.

- 410. Gur, R.E. and R.C. Gur, *Functional magnetic resonance imaging in schizophrenia*. Dialogues Clin Neurosci, 2010. **12**(3): p. 333-43.
- 411. Hamer, D., *Genetics. Rethinking behavior genetics*. Science, 2002. **298**(5591): p. 71-2.
- 412. Huang, H.S., et al., *Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters*. J Neurosci, 2007. **27**(42): p. 11254-62.
- 413. Veldic, M., et al., *Epigenetic mechanisms expressed in basal ganglia GABAergic neurons differentiate schizophrenia from bipolar disorder*. Schizophr Res, 2007. **91**(1-3): p. 51-61.
- 414. Costa, E., et al., *GABAergic promoter hypermethylation as a model to study the neurochemistry of schizophrenia vulnerability*. Expert Rev Neurother, 2009. **9**(1): p. 87-98.
- 415. Van Winkel, R., et al., *REVIEW: Genome-wide findings in schizophrenia and the role of gene-environment interplay*. CNS Neurosci Ther, 2010. **16**(5): p. e185-92.
- 416. McEwen, B.S., *Stress, sex, and neural adaptation to a changing environment: mechanisms of neuronal remodeling*. Ann N Y Acad Sci, 2010. **1204 Suppl**: p. E38-59.

Appendices

Appendix A: Summary of clinical features in children with de novo CNVs detected by WGS with mapping 100K arrays and confirmed by FISH

Patient	Abnormality	Clinical Features	Evidence of Pathogenicity
1895	5.7 Mb, del(13)(q12.11q12.13)	6-year-old girl; microcephaly; mild growth retardation; several café-au-lait macules; mild anemia, thrombocytopenia, and neutropenia; moderate developmental delay	De novo CNV; large region of genetic imbalance
3476	11.1 Mb, del(4)(q21.21q22.1)	18-year-old girl; extremely short stature; deep-set eyes with narrow palpebral fissures; low-set, straight eyebrows; narrow nasal root and bridge; prominent columella with receding alae nasae, short philtrum, and thin, downturned lips; small hands and feet; severe hypotonia; marked pes planus; mild scoliosis; severe cognitive impairment	De novo CNV; very large region of genetic imbalance
4794	2.9 Mb, dup(16)(p13.3)	12-year-old boy; blepharophimosis and ptosis; low-set, mildly dysplastic auricles; mild pectus excavatum; C5-C6 vertebral fusion; pes cavus and clawed toes; normal growth; full-scale IQ of 46	De novo CNV; gene-rich region; striking similarity to other patients described with cytogenetically apparent 16p13 duplication
4818	3.4 Mb, del(12)(q14.2q15)	12-year-old boy; prenatal-onset growth retardation; partial anodontia; mild limitation of extension at elbows; mildly short and narrow fingers; tremor; mild developmental delay; osteopoikilosis on radiographic examination	De novo CNV; large region of genetic imbalance
5003	.32 Mb, del(2)(p16.3p16.3)	7-year-old boy; full-scale IQ of 74; learning problems in both parents; attention deficit disorder; Asperger syndrome; frontal bossing, high anterior hairline, frontal hair whorl, and low posterior hairline; mild dorsal scoliosis with 13 ribs on left, bifid right second rib, hemivertebrae and fusions of T2, T3, T4, and fusion of L4 and L5; asthma	De novo CNV; no known polymorphic CNVs in region
5566	.18 Mb, del(14)(q11.2q11.2)	2-year-old girl; large for gestational age at birth, head growth at 97th percentile subsequently, and normal height and weight; generalized hypotonia and joint hypermobility; short palpebral fissures; mildly dysplastic auricles; long toes with 2-3 cutaneous syndactyly; developmental delay, especially gross motor	De novo CNV; no known polymorphic CNVs in region; deletion lies within that of child 8326 and DECIPHER patient CAM126
5994	Mosaic trisomy 9 (~20%)	17-mo-old boy; plagiocephaly, torticollis, and broad and prominent forehead with bifrontal narrowing; low-set ears, small mouth, and short philtrum; puffy hands and feet; undescended testes; hypotonia; tracheomalacia; enlarged cerebral ventricles; bicuspid aortic valve; moderate-to-severe developmental delay	De novo CNV; cytogenetic confirmation; phenotype characteristic of trisomy 9 mosaicism
6168	1.1 Mb, dup(17)(q21.33)	9-year-old girl; microcephaly; moderate-to-severe conductive hearing loss; bilateral preauricular skin tags, small ears with abnormal pinnae, and small ear canals; mild retrognathia; mild MR	De novo CNV; large region of genetic imbalance
6545	3.6 Mb, del(7)(p22.2p22.1)	14-mo-old girl; prenatal-onset growth retardation and microcephaly; patent ductus arteriosus and perimembranous ventricular septal defect; failure to thrive; severe developmental delay; brachycephaly, epicanthic folds, midface hypoplasia, and lateral flare of eyebrows; 2-3 syndactyly of toes	De novo CNV; large region of genetic imbalance; DECIPHER case UPP969 has 0.2-Mb deletion within deleted region of this patient
7807	1.4 Mb, del(22)(q12.1q12.1)	5-year-old boy; microcephaly; frontal upsweep, hypertrichosis of forehead, and prominent brows; bifid uvula; undescended testes; developmental delay, especially of speech	De novo CNV; large region of genetic imbalance; 1.8-Mb deletion in DECIPHER patient CHG758 overlaps the CNV in this patient
8326	1.6 Mb, del(14)(q11.2q11.2)	2-year-old boy; small ventricular septal defect that closed spontaneously; large patent ductus arteriosus that required surgical closure; plagiocephaly; pseudostrabismus with very broad nasal root and short nose; preauricular pit; undescended testes and hypoplastic scrotum; moderate-to-severe developmental delay, especially of speech	De novo CNV; DECIPHER patient CAM126 has similar deletion

Appendix B: Genomic imbalance detected by 500K GeneChip® AGH in a cohort of 100 trios with idiopathic ID

Patient	CNV	Location	Start → End	Size (bp)	Validation	Phenotype	RefSeq Genes Involved*	Comments	Interpretation
9133	Loss	1p36.32-p36.33	769,185 → 3,581,308	2,812,123	FISH, MLPA	9 year-old female with obesity, moderate cognitive impairment, myoclonus, polyphagia, hypotonia, narrow frontal area, deep-set eyes, prominent orbital rims, short nose with low nasal bridge and upturned nasal tip, midface retrusion, short philtrum, tented upper lip, thoracic kyphosis	~70 genes including <i>AGRN</i> , <i>GNB1</i> , <i>PEX10</i> , <i>PRKCZ</i> , <i>SKI</i> , and <i>TP73</i>	This CNV is included in the 1p36 deletion syndrome critical region, and the patient's clinical features are compatible with that syndrome	Pathogenic
873	Gain [†]	2q37	231,577,285 → 242,663,303	11,086,018	FISH, MLPA	15 year-old male with severe cognitive impairment, birth weight < 1 st centile, birth length < 1 st centile, head circumference at birth < 10 th centile, hypotonia, microcephaly, contractures of hips and knees, hypoplastic scrotum, undescended testes, prominent, cup-shaped ears, narrow bifrontal diameter, broad nasal root, prominent epicanthal folds, bilateral clinodactyly V, ataxic gait, progressive joint contractures and muscle wasting of lower extremities, mixed hearing loss and hypoplastic inferior cerebellar vermis.	~100 genes including <i>AGXT</i> , <i>ATG16L1</i> , <i>CAPN10</i> , <i>CHRND</i> , <i>CHRNA3</i> , <i>COL6A3</i> , <i>D2HGDH</i> , <i>GBX2</i> , <i>HDAC4</i> , <i>MLPH</i> , <i>PDCD1</i> , <i>PER2</i> , <i>SAG</i> and <i>UGT1A1</i>		Pathogenic
	Loss [†]	10q26.13	126,415,527 → 134,032,911	7,617,384			~75 genes including <i>ADAM12</i> , <i>CTBP2</i> , <i>DOCK1</i> , <i>DPYSL4</i> , <i>FGFR2</i> , <i>OAT</i> , and <i>UROS</i>		
6473	Loss	4p16.3	190,631 → 3,277,436	3,086,805	Cytogenetic re-analysis, FISH	3 year-old male with fetal growth retardation, length 2 standard deviations below the mean for age, weight 3-4 standard deviations below the mean, head circumference 3-4 standard deviations below the mean, global developmental delay, seizures, triangular face, small jaw, posteriorly rotated ears, 2° hypospadias, ataxia	~45 genes including <i>ADD1</i> , <i>FGFR3</i> , <i>HTT</i> , <i>IDUA</i> , <i>LETM1</i> , <i>PDE6B</i> , <i>SH3BP2</i> , and <i>WHSC1</i>	Includes the Wolf-Hirschhorn syndrome critical region, and the clinical features are compatible with that syndrome	Pathogenic

Patient	CNV	Location	Start → End	Size (bp)	Validation	Phenotype	RefSeq Genes Involved*	Comments	Interpretation
5814	Loss [†]	4p16.1-p16.3	13,255 → 8,472,657	8,459,402	Cytogenetic re-analysis, MLPA	13 month-old female with length below the 3 rd centile, microcephaly, developmental delay, bilateral preauricular pits, and submucous cleft palate	~90 genes including <i>ADD1</i> , <i>ADRA2C</i> , <i>CRMP1</i> , <i>EVC</i>	The deletion includes the Wolf-Hirschorn syndrome critical region.	Pathogenic
	Gain [†]	8p23.1-p23.3	180,568 → 6,898,076	6,717,508			20 genes including <i>ARHGEF10</i> , <i>CLN8</i> , <i>DLGAP2</i> , and <i>MCPH1</i>		
216	Loss	6p21.33	29,937,087 → 30,026,517	89,430	MLPA	18 year-old male with postnatal onset growth retardation, unilateral sensorineural deafness, narrow face, bulbous nasal tip and mild intellectual disability.	HCG9, MICD and 5 pseudogenes	Deletion is homozygous in child, heterozygous in both parents (not <i>de novo</i> mutation)	Not pathogenic for ID
3160	Loss	6 q21-q22.31	111,979,175 → 121,506,916	9,527,741	FISH	8 year-old female with moderate developmental delay, hypotonia, microcephaly, brachycephaly, epicanthic folds, small ears with hypoplastic lobes, and micrognathia	~40 genes including <i>ASF1A</i> , <i>COL10A1</i> , <i>FRK</i> , <i>FYN</i> , <i>GOPC</i> , <i>HDAC2</i> , <i>LAMA4</i> , <i>MCM9</i> , <i>PLN</i> , <i>TSPYL1</i> , and <i>WISP3</i>		Pathogenic
2200	Loss	7p15.3	14,141,506 → 24,950,414	10,808,908	FISH	11 1/2 year-old female with head circumference at the 2 nd centile, mild cognitive impairment, sensorineural hearing loss, cleft palate, craniosynostosis, unilateral ptosis and esotropia, orbital rim hypoplasia, malar and midface hypoplasia, low-set ears with incomplete out-folding of superior helix, brachydactyly and syndactyly of digits, broad thumbs, decreased range of motion in elbows, and leg length discrepancy	~40 genes including <i>DFNA5</i> , <i>DGKB</i> , <i>DNAH11</i> , <i>FAM126A</i> , <i>HDAC9</i> , <i>IL6</i> , and <i>RAPGEF5</i>		Pathogenic

Patient	CNV	Location	Start → End	Size (bp)	Validation	Phenotype	RefSeq Genes Involved*	Comments	Interpretation
9938	Loss	7q22.1	98,211,585 → 100,553,755	2,342,170	FISH	14 year-old female with height < 5 th centile, weight < 5 th centile, head circumference < 5 th centile, severe cognitive impairment, left sensorineural hearing loss, close-set eyes, broad nasal root, marked retrognathia, high-arched palate, small and narrow feet, short 2-5th toes with hypoplastic nails, atrio-ventricular septal defect, and polyarticular arthritis	~70 genes including <i>ACHE</i> , <i>ACTL6B</i> , <i>CYP3A5</i> , <i>EPO</i> , <i>MUC3A</i> , <i>SERPINE1</i> , <i>SMURF1</i> , and <i>TFR2</i>		Pathogenic
1594	Gain	8q12	58,388,614 → 65,306,097	6,917,483	FISH	1 1/2 year-old female with height > 97 th centile, head circumference at 2 nd 5 th centile, developmental delay, Duane anomaly, broad glabella, epicanthic folds with telecanthus, upslanting palpebral fissures	15 genes including <i>ASPH</i> , <i>CHD7</i> , <i>RAB2A</i> , <i>RLBP1L1</i> , <i>TOX</i> , and <i>TTPA</i>		Pathogenic
663	Loss	9p13.3	34,144,847 → 38,736,451	4,591,604	FISH	5 1/2 year-old female with height at the 5 th centile, developmental delay, tremor, ocular hypertelorism, epicanthal folds, double hair whorl, bilateral ptosis, short upturned nose, flattened philtrum, underdeveloped genitalia, and pigmentary retinal changes	~75 genes including <i>CNTFR</i> , <i>DNAI1</i> , <i>DNAJB5</i> , <i>FANCG</i> , <i>GALT</i> , <i>GBA2</i> , <i>GNE</i> , <i>GRHPR</i> , <i>NPR2</i> , <i>PAX5</i> , <i>RECK</i> , <i>SHB</i> , <i>TPM2</i> , <i>UNC13B</i> , <i>VCP</i>	The deleted region in this patient is completely included in the region deleted in patient 9346.	Pathogenic
9346	Loss	9p11.2-p13.3	33,702,471 → 44,744,675	11,042,204	FISH	9 1/2 year-old female with moderate cognitive impairment, seizures, tremor, cataract, broad frontal area with bossing, arched eyebrows, low nasal bridge, and short, upturned nose	~85 genes including <i>CNTFR</i> , <i>DNAI1</i> , <i>DNAJB5</i> , <i>FANCG</i> , <i>GALT</i> , <i>GBA2</i> , <i>GNE</i> , <i>GRHPR</i> , <i>NPR2</i> , <i>PAX5</i> , <i>RECK</i> , <i>SHB</i> , <i>TPM2</i> , <i>UNC13B</i> , <i>VCP</i>	The deleted region in this patient includes the entire segment deleted in Patient 663.	Pathogenic

Patient	CNV	Location	Start → End	Size (bp)	Validation	Phenotype	RefSeq Genes Involved*	Comments	Interpretation
523	Loss	9q34.3	139,516,033 → 139,814,485	298,452	FISH	4 year-old female with moderate developmental delay, hypotonia, microcephaly, flat face with upslanting palpebral fissures, ocular hypotelorism, synophrys, and anteverted nares	7 genes including <i>EHMT1</i>	Within the critical region for the 9q subtelomeric deletion syndrome clinical features are compatible	Pathogenic
8056	Mosaic Trisomy	9	whole chromosome		FISH	2 1/2 year-old male with weight < 5 th centile, developmental delay, preauricular skin tags, hypospadias, and cryptorchidism	Numerous	Clinical features compatible with mosaic trisomy 9 syndrome	Pathogenic
6904	Uni parental Disomy	11p11.2-pter	196,767 → 44,589,530	44,392,763	Micro-satellite markers	11 year-old female with height < 5 th centile, gross and fine motor delay, hypotonia, and moderate mental handicap	Numerous	Mosaic paternal isodisomy; phenotype not compatible with Beckwith-Wiedemann syndrome	Uncertain
9897	Loss	13q33.3-q34	107,271,189 → 109,368,996	2,097,807	FISH	10 year-old female with fetal growth retardation, moderate cognitive impairment, upslanting palpebral fissures, and retrognathia	6 genes including <i>IRS2</i> , <i>LIG4</i> , and <i>MYO16</i>		Pathogenic
818	Loss	14q11.2	21,929,710 → >22,036,502	106,792	Fosmid FISH (variable) §	6 1/2 year-old male with weight < 5 th centile, height at 5 th centile, mild cognitive impairment with particular delay in language	Multiple T-cell receptor alpha-chain V, J, and region genes	Highly polymorphic region	Not pathogenic for ID
1658	Uni parental Disomy	16	Whole chromosome	Whole chromosome	Micro-satellite markers	5 year-old female with normal growth, severe mental handicap, seizures, self-abusive behavior.	Numerous		Uncertain
2106	Loss	17q21.31	41,049,321 → 41,564,451	515,130	FISH	15 year-old male with fetal growth retardation, mild cognitive impairment, attention deficit disorder, sagittal craniosynostosis, long face with malar hypoplasia and mild retrognathia, short and upslanting palpebral fissures, low-set ears.	5 genes including <i>MAPT</i>	Includes the critical region for the 17q21.31 deletion syndrome, clinical features are compatible.	Pathogenic

Patient	CNV	Location	Start → End	Size (bp)	Validation	Phenotype	RefSeq Genes Involved*	Comments	Interpretation
8619	Loss	21q22.11	33,902,218 → 34,087,893	185,675	Agilent 244 K AGH	24 year-old male with prenatal and postnatal growth retardation, moderate to severe intellectual disability, severe hypotonia, microcephaly, metopic craniosynostosis, cleft palate, down-slanting palpebral fissures, low-set ears, wide nasal base, retrognathia, tetralogy of Fallot, cryptorchidism, and joint hyperextensibility	5 genes including <i>SYNJ1</i>	No polymorphisms of region reported	Uncertain
	Loss	22q11.2	19,062,809 → 19,785,125	722,316	FISH		14 genes including <i>BCR</i> , <i>DGCR8</i> , <i>HIRA</i> , <i>MAPK1</i> , <i>PRODH</i> , <i>SNAP29</i> , <i>SEPT5</i> , <i>SERPIND1</i> , and <i>TBX1</i>	Included in the 22q11 deletion syndrome region. compatible with other reported cases of distal 22q11.2 microdeletion	Pathogenic
9979	Gain	22q11.21	19,429,297 → 19,791,607	362,310	FISH	20 year-old female with short stature, mild mental deficiency, cleft palate, and micrognathia	9 genes including <i>PI4KA</i> , <i>SERPIND1</i> , <i>LZTR1</i> , <i>SNAP29</i>	Polymorphic region	Uncertain
1128	Gain	Xq12-q21.1	67,088,023 → 76,204,344	9,116,321	FISH	11 year-old male with normal growth, severe developmental delay, hypotonia, brachycephaly, bilateral epicanthic folds, and posteriorly rotated ears with hypoplastic helices	~60 genes including <i>ABCB7</i> , <i>DLG3</i> , <i>EDA</i> , <i>EFNB1</i> , <i>GJB1</i> , <i>IGBP1</i> , <i>IL2RG</i> , <i>OPHN1</i> , <i>NAP1L2</i> , <i>NLGN3</i> , <i>PHKA1</i> , <i>SLC16A2</i> , <i>TAF1</i> , <i>ZDHHC15</i>		Pathogenic

The table includes all de novo CNVs, mosaic trisomy and UPD detected by 500 K AGH and confirmed by an independent method in 100 children with idiopathic ID. Breakpoints are shown on Human Genome Assembly Build 36.1. * The approximate number of RefSeq genes for each CNV is given, but only the most likely genes for the phenotype are named.† Unbalanced reciprocal translocation. § Interphase FISH in patient 818 showed some cells with no signals, some with one signal and some with two signals for a probe in the region of the deletion detected by AGH. This was interpreted as evidence of somatic mosaicism for this deletion. (Friedman et al. *BMC Genomics* 2009 10:526)

Appendix C: Methodology for investigating genotype-phenotype correlations for *de novo* CNVs identified in children with idiopathic intellectual disability.

(1) Search for other patients with MR who have similar genetic lesions to our cases

Database search –

- ✓ DECIPHER database(<http://www.sanger.ac.uk/PostGenomics/decipher/>). Input detailed genotype and phenotype data on to DECIPHER and use database tools to locate patients with similar genetic lesions
- ✓ The UK Association of Clinical Cytogeneticists (ACC) Chromosome Abnormality Database (CAD) [www.ukcad.org.uk]
- ✓ European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations [www.ecaruca.net]

Catalogue search – Search directories of published chromosomal aberrations in man;

- ✓ Catalogue of unbalanced chromosome aberration in man. Schinzel A (2001) 2nd Ed, de Gruyter, ISBN 3-11-011607-3
- ✓ Chromosomal variation in man; a catalogue of chromosomal variants and anomalies. Borgaonkar DS (1997), 8th Ed, Wiley, ISBN 0-471-24332-9

For cases with genetic lesions similar to those of our patients, obtain detailed molecular data on the patient's genomic lesion, the inheritance status of the lesion, as well as detailed phenotypic and clinical information for the patient, and compare to our cases in order to detect recurrent patterns of dysmorphology or other clinical presentations and identify syndromic associations.

(2) Catalogue benign CNVs that map to the genomic loci involved in our case CNVs

Database search – identify all benign CNVs recorded in online databases of normal variation in man such as;

- ✓ The Database of Genomic Variants [<http://projects.tcag.ca/variation/>] hosted by the Centre for Applied Genomics, Toronto,
- ✓ The Structural Variation Database maintained at the University of Washington [<http://humanparalogy.gs.washington.edu/structuralvariation>]

Benign CNVs reported in other reports

- ✓ Search published reports not repositied in the above databases for any possible benign variants mapping to our CNV loci. Overlay benign CNVs co-ordinates upon the potentially pathogenic CNV genomic co-ordinates in order to refine the region whose perturbation is more likely causal of the phenotypes (as it is less likely that any regions that map within benign CNVs would contain genes which are pathogenic in aberrant copy number).

(3) Analysis of gene content of regions involved in the patient's CNV.

Analysis of aberration size and gene content involves data mining the available genome and proteome browsers such as the NCBI web based resources, Ensembl, the UCSC genome browser for known genes involved in the critical deletion or duplication interval and how perturbation of the gene can be pathogenic for the observed phenotypes.

(4) Amalgamate gathered clinical and genetic information and identify and establish genotype-phenotype correlations.

Appendix D: Primer sequences for qPCR validation (SYBR green $\Delta\Delta$ CT method) of custom aCGH results of CNVS involving epigenetic regulatory genes.

	Forward primer				Reverse primer				Amplicon	
Target	Size	Tm	% GC	Primer	Size	Tm	% GC	Primer	Size	Pos on Genome Build 36 (source UCSC PCR)
H6PD	24	60	46	GGTGGATAGATGCAGAAACAA GGA	27	60	41	TATGAATGTGTAAGTCTGGA GGTCTT	110	chr1:9247701+9247810
JMJD1C	20	59	60	CATCCTCCGATCTGGGTCAG	21	58	52	ACAGTGGGAAAGGCTTTCTCC	113	chr10:64698812+64698924
JMJD1A/ KDM3A	26	58	38	AGGTGACAAAACTTAGTTGGT TCAG	21	58	38	TTTGATGCTGGGTTTCCATTT	102	chr2:86537074+86537176
JARID2	26	60	42	GTGTATTTTGGGAAGCTCTCAGG ATGA	23	59	43	ACTGGCATGAAGATGAAGCAT TG	103	chr6:15576778+15576880
ARID1B	20	59	50	GATCAACATGGCGGACAACA	20	60	55	CGGATAATGGTTGCACTGGC	141	chr6:157141512+157141652
ARID2	20	59	55	CGACAGAGGGACTTGCGATT	21	59	52	TGTGTTTTCCACCAGAGAGCCC	121	chr12:44410981+44411101
ARID4B	21	58	48	AACCCAGAGATTCGTTTTCC	20	58	50	TTTTGTTTCCCCTCCAGGTG	107	chr1:233556817+233556923
CHD6	20	59	60	AGACACCCTGCAAGGACACC	22	60	50	GCTTGAAATGGGCTTCTGTCT G	130	chr20:39680453+39680582
CHD7	19	58	58	TCGTGCTCGGGAAGTATCG	20	59	60	GGTGAGTTTCAGGGTGTCCG	102	chr8:61753957+61754058
SMARCA2	19	59	58	ATCGGTGGGTGGACATGGT	20	60	50	TGCTGGCAAAGACAAAGGG	102	chr9:2077201+2077302
MEF2C	30	58	27	GACCTTGAAATGGGAAGAAAT ACC	24	58	42	AAATGACATTCCACTTAAATAA CAGATGAT	107	chr5:88163076+88163182