EXOME SEQUENCING FOR UNDERSTANDING PHENOTYPIC VARIABILITY IN SUBJECTS WITH 16p11.2 CNV

by

Jila Dastan

M.D. Medicine, Shiraz University, 1993
M.Sc. Medical Genetics, Glasgow University, 1998

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Medical Genetics)

The University of British Columbia
(Vancouver)

April 2016

Abstract

Microduplication of 16p11.2 (dup16p11.2) is associated with a broad spectrum of neurodevelopmental disorders (NDD) confounded by variable expressivity. I hypothesized that while some unique features reported in individuals with dup16p11.2 may be explained by the over-expression of its integral genes, co-occurrence of other genetic alterations in the genome may account for the variability in their clinical phenotypes. This hypothesis was explored in two unrelated subjects with NDD who each inherited the dup16p11.2 from an apparently healthy carrier parent.

First, I performed a detailed phenotypic analysis of individuals with dup16p11.2 (published and current study). I did not find evidence of phenotypic commonality and consistent syndromic phenotype pattern among carriers of dup16p11.2. Next, I assessed the effect of dup16p11.2 on the expression of genes located within and nearby this region which showed that RNA expression of *KCTD13*, *MAPK3* and *MVP* from 16p11.2 region was inconsistent in carriers of dup16p11.2. *KTCD13* has been identified as a driver of a mirror brain phenotype of 16p11.2 CNV in zebrafish. However, the data presented here demonstrated that dup16p11.2 did not result in increased protein expression of KCTD13 in either probands or the healthy carrier parent, indicating that *KCTD13* is not the sole cause of microcephaly in cases with dup16p11.2. Finally, whole exome sequencing (WES) was used to investigate the presence of genomic sequence changes in dup16p11.2 carriers that could explain such clinical variability. Compound heterozygous variants of *VPS13B* in proband A and missense variants of *SYNE2* in proband B were identified. Mutations of *VPS13B* cause Cohen syndrome in keeping with proband A's phenotype (ID, microcephaly, facial gestalt, retinal dystrophy, joint hypermobility and episodic neutropenia) and low RNA expression. The protein encoded by *SYNE2*, Nesprin 2, plays critical

roles in neurogenesis in mice. Over-expression of Nesprin 2 identified in proband B may cause NDD phenotypes via a dominant-negative effect.

In conclusion, pathogenic variants were identified in genes outside of the 16p11.2 region which could contribute to the clinical variability between parent-offspring dup16p11.2 carriers in this study. This suggests discordance in phenotype of dup16p11.2 carriers warrants further study by WES and individualized genetic assessment and counselling.

Preface

For chapter 2, I reviewed the charts of two patients, extracted and categorized their clinical data. I also searched, extracted and categorized the phenotype of published carriers of dup16p11.2 (including affected and unaffected subjects), and subsequently performed detailed phenotypic analysis of dup16p11.2 carriers (published and current study). Chromosome microarray analysis results for all subjects except the brother of proband B were obtained previously in the research laboratory. Microarray for the brother of proband B was performed by research assistant, Sally Martell.

For chapter 3, lymphoblasts were transformed on a service basis at the CFRI (Dr. John Priatel lab), and then were grown and maintained by me. RNA samples (probands and two controls) used for the whole genome expression (WGE) study were extracted by Sally Martell. The Illumina Expression BeadChip array (HumanRef-8 v3.0) was run by the CFRI core facility. Background-correction and normalization of data was done in collaboration with Dr. Paul Pavlidis (Department of Psychiatry, Centre for High-throughput Biology). I analyzed the expression of genes within and nearby the 16p11.2 region, and also performed the pathway enrichment analysis for over-and under-expressed genes shared between two probands. I also simultaneously extracted RNA and protein samples of two families and two controls, and subsequently performed qPCR and western blotting for candidate genes within the 16p11.2 region (*KCTD13*, *MAPK3* and *MVP*).

A version of Chapter 4 is submitted, entitled; "*Whole exome sequencing in familial study of 16p11.2 duplication carriers*" (Dastan J, Chijiwa C, Tang F, Martell S, Qiao Y, Rajcan-Separovic E and Lewis MES). Whole exome sequencing was done by PerkinElmer Inc. A VCF file was generated and imported into Golden Helix software by co-op student Flaming Tang. I

analyzed all variants and identified candidate genes for both families, also designed primers, performed PCR for each candidate variant and prepared the PCR products for Sanger sequencing. Sanger sequencing was performed at the CFRI/CMMT core facility. Further, I performed qPCR and western blotting for *VPS13B* and *SYNE2* genes. I performed *in-silico* prediction analysis (Alamut software) of two *VPS13B* variants. To validate this prediction, I designed two different sets of primers covering exons 9-12 and 26-29 of *VPS13B*, and then performed PCR using the cDNA samples from Proband A and one control to send for Sanger sequencing after PCR clean-up. Lastly, I extracted all the reported phenotypes of patients with Cohen syndrome (CS), and joined Dr. Suzanne Lewis in the clinic for re-evaluation of proband A for the presence of any of CS features, as well as examination of Mother A for presence of dup16p11.2 features.

The collection of the samples for these studies was approved by the University of British Columbia Clinical Research Ethics Board, approval number C01-0509.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABD	Actin binding domain
AD	Autosomal dominant
ADHD	Attention-deficit/hyperactivity disorder
ALDOA	Aldolase A, fructose-bisphosphate
AR	Autosomal recessive
AS	Angelman syndrome
ASD	Autism spectrum disorder
B.C.	British Columbia
BAM	Binary alignment/map
BBS7	Bardet-Biedl syndrome 7
BMI	Body mass index
BOLA2/B	Bo1A family member 2/B
BWA	Burrows-wheeler aligner
C/S	Caesarean section
C16orf54/92	Chromosome 16 open reading frame 54/92
CA1	Cornu ammonis 1
CAs	Congenital anomalies
CDIPT	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
CEDNIK	Cerebral dysgenesis–neuropathy–ichthyosis–keratoderma
CEP290	Centrosomal protein 290kDa
CH	Calponin homology
CHD1L	Chromodomain helicase DNA binding protein 1-like
CLN3	Ceroid-lipofuscinosis, neuronal 3
CMA	Chromosomal microarray
CNS	Central nervous system
CNV	Copy number variation
CO2	Carbon dioxide
CORO1A	Coronin, actin binding protein, 1A
CPK	Creatine phosphokinase

CS	Cohen syndrome
CSF	Cerebrospinal fluid
CT	Computerized tomography
CVID	Common variable immunodeficiency
DCD	Developmental coordination disorder
DD	Developmental delay
DE	Differentially expressed
del	Deletion
Deletion	Microdeletion
DMD	Duchenne muscular dystrophy
DMD	Dystrophin
DOC2A	Double C2-Like domains, alpha
DS	DiGeorge syndrome
DSM	Diagnostic statistical manual
dup	Duplication
Duplication	Microduplication
EBV	Epstein-Barr virus suspension
EDMD	Emery-Dreifuss muscular dystrophy
EEG	Electroencephalogram
EHMT1	Euchromatic histone-lysine N-methyltransferase 1
ERK	Extracellular-signal-regulated kinases
FATHMM	Functional analysis through hidden markov models
FBS	Fetal bovine serum
FISH	Fluorescent in situ hybridization
FMR1	Fragile X mental retardation 1
FUS	FUS RNA binding protein
FX	Fragile X
GATK	Genome analysis toolkit
GDPD3	Glycerophosphodiester phosphodiesterase domain containing 3
GERP	Genomic evolutionary rate profiling

GH	Golden helix
GIYD1/2	GIY-YIG domain-containing protein 1/2
GO	Gene ontology
GOF	Gain of function
GWAS	Genome wide association study
HBT	Human brain transcriptome
HC	Head circumference
HIRIP3	HIRA interacting protein 3
HPO	Human phenotype ontology
ID	Intellectual disability
INM	Inner nuclear membrane
iPSCs	Induced pluripotent stem cells
IQ	Intelligence quotient
ITGAX	Integrin subunit alpha X
KASH	Klarsicht–ANC–Syne-homology
kb	Kilobase
KCTD13	Potassium channel tetramerization domain containing 13
kDa	Kilodalton
LCL	Lymphoblasts cell line
LCRs	Low copy repeats
LINC	Linkers of nucleoskeleton and cytoskeleton
LOF	Loss of function
MAF	Minor allele frequency
MAPK3	Mitogen-activated protein kinase 3
mat	Maternal
MAZ	MYC-associated zinc finger protein (purine-binding transcription factor)
Mb	Megabase
MECP2	Methyl CpG binding protein 2
mGluR5	Metabotropic glutamate receptor 5
MKS	Meckel-Gruber syndrome

MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MVP	Major vault protein
NAHR	Non-allelic homologous recombination
NDD	Neuro-developmental disorders
NF1	Neurofibromin 1
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NHLBI	National heart, lung and blood institute
NMD	Nonsense-mediated decay
NSD1	Nuclear receptor binding SET domain protein 1
ODD	Oppositional defiant disorders
OFC	Occipito-frontal circumference
ONM	Outer nuclear membrane
PAH	Phenylalanine hydroxylase
pat	Paternal
PCR	Polymerase chain reaction
PDD-NOS	Pervasive developmental disorder-not otherwise specified
PhyloP	Phylogenetic p-values
PLS	Potocki-Lupski syndrome
PolyPhen2	Polymorphism phenotyping v2
PPP4C	Protein phosphatase 4, catalytic subunit
PRKAB2	Protein kinase, AMP-activated, beta 2 non-catalytic subunit
PRRT2	Proline-rich transmembrane protein 2
PWS	Prader-Willi syndrome
qPCR	Real-time quantitative PCR
QPRT	Quinolinate phosphoribosyltransferase
RAB6	RAB6A, member RAS oncogene family
RAI1	Retinoic acid induced 1
rRNA	Risbosomal RNA

RT-PCR	Reverse-transcriptase PCR
SCID	Severe combined immune deficiency
SD	Segmental duplication
SD	Standard deviation
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SH2B1	SH2B adaptor protein 1
SHANK3	SH3 and multiple ankyrin repeat domains 3
SIFT	Sorting intolerant from tolerance
SMS	Smith-Magenis syndrome
SNAP29	Synaptosomal-associated protein, 29 kDa
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SPN	Sialophorin
SR	Spectrin repeat
SULT1A3/4	Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3/4
SYNE1/2	Spectrin repeat containing, nuclear envelope 1/2
SYNE3/4	Spectrin repeat containing, nuclear envelope family member 3/4
TAOK2	TAO kinase 2
TBX1	T-Box 1
TMEM219	Transmembrane protein 219
TS	Tuberous sclerosis
TSC1	Tuberous sclerosis 1
TSC2	Tuberous sclerosis 2
UBE3A	Ubiquitin protein ligase E3A
VCF	Variant call format
VOUS	Variants of uncertain significance
VPS13B	Vacuolar protein sorting 13 homolog B (Yeast)
WB	Whole blood
WBC	White blood cell
WBDD	Winter-Baraitser Dysmorphology Database

WBS	Williams Beuren syndrome
WES	Whole exome sequencing
WGE	Whole genome expression
WGS	Whole genome sequencing
XL	X-linked
y/o	Year old
YPEL3	Yippee-like 3 (Drosophila)
%ile	Percentile

Acknowledgements

I would like to offer my enduring gratitude to my supervisors, Professor Suzanne Lewis and Dr. Evica Rajcan-Separovic for their constant guidance and support. I am grateful for the opportunity to conduct research under their supervision.

I would also like to thank members of my advisory committee; Dr. Wendy Robinson, Dr. Anna Lehman and Professor David Huntsman for their expert advice; and the Medical Genetics graduate program assistant, Cheryl Bishop for her kindness and friendship.

I am truly grateful for friendship and support of members of the Lewis and Rajcan-Separovic Labs: Dr. Ying Qiao, Sally Martell, Kristina Calli, Hani Bagheri, Chansonette Badduke, and Flamingo Tang.

I also would like to thank the funding agencies, and the families involved in this project for their valued cooperation and participation.

Lastly, special thanks are owed to my parents and husband, who have supported me throughout my years of education. I missed my two beloved sons every single day that I lived in Canada by myself. I could not have made it through this challenging experience without their admirable understanding and encouragement.

Chapter 1: Introduction

1.1 Neurodevelopmental disorders

Neurodevelopmental disorders (NDD) are a heterogeneous group of clinical disorders characterized by impairment in growth and development of the brain often causing cognitive, neurological, or psychiatric dysfunction [1]. The estimated frequency of NDD in children of industrial countries is about 15% [2]. According to the Diagnostic Statistical Manual of mental disorders (DSM-5), NDD is an umbrella term that can extend to diverse disorder classifications including intellectual disabilities (ID), autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), communication disorders, specific learning disorder, motor disorders, and other specified and unspecified NDD [3]. There is debate that current classified neuropsychiatric disorders are mostly NDD as well. It assumes that etiological elements of NDD occur during fetal development or/until early childhood, thus, the broad spectrum of these elements poses a specific challenge in nosology and classification and consequently diagnosis and treatment of NDD [4]. Having a diagnosis of one disorder within the NDD continuum, greatly increases the risk of meeting full or sub-threshold criteria for another related disorder, suggesting that they share common, underlying susceptibilities [5]. The best examples of this commonality are ID and ASD.

"ID" is distinguished by significant limitations both in cognitive (IQ < 70) and in adaptive functions, with the age of onset before 18 years [6]. During infancy or early childhood, ID may appear with focal or global developmental delay (DD). The estimated prevalence of DD/ID in the general population is 1-3% [7], disproportionately affecting more males than females (1.6-1.7 times more) [8]. Unlike severe ID, the prevalence of mild ID is variable and depends on environmental factors including access to education and healthcare [8-10].

"ASD" is characterized by two categories of behavioral problems: a) unusual reciprocal social interactions and communication deficits; and b) stereotyped, repetitive behaviors with restricted interests [3]. The estimated prevalence of ASD is 1 in 68 [11] with male to female ratio of 4:1 and is typically diagnosed using objective psychometric measures by age 3-4 years [12].

Teasing ID and ASD apart one from the other is very difficult. Up to 70% of individuals with ASD also have ID. The remaining 30% are technically not intellectually disabled, although a subset of individuals may suffer from common, co-occurring disabilities such as speech and language deficits, as well as behavioral problems. On the other hand, approximately 10% of children with ID demonstrate autistic symptoms [13-15].

1.2 Genetics of NDD

While most cases of NDD have unknown etiology, genetic causes of NDD are mainly chromosomal abnormalities and monogenic disorders, and less frequently imprinting/epigenetic disorders [16]. A summary of common genetic etiologies of NDD is presented below.

1.2.1 Monogenic disorders

Monogenic disorders are caused by pathogenic variants of a certain single gene harbouring a direct genotype-phenotype relationship, of which some are responsible for a subset of NDD.

Fragile X (FX) and Tuberous sclerosis (TS) are well-known genetic disorders associated with ASD and ID, resulting from unstable expansion of a CGG repeat (>200 repeats) in the *FMR1* gene (Xq27.3), and mutations of *TSC1* (9q34) or *TSC2* (16p13.3), respectively [17-19]. According to Willemsen and Kleefstra (2013), FX is the most common cause of ID after trisomy 21, with a frequency of 0.5-3% [16, 20-23]. About 30% cases of FX are also diagnosed with autism and an additional 30% categorized with pervasive developmental disorder-not otherwise

specified (PDD-NOS) [24]. TS is associated with a range of NDD including ID, learning disability, epilepsy and ASD. About 90% of patients with TS present with epilepsy [25].The frequency of autism among patients with TS is 40-50% [26-28], in which 70% suffer from coexisting cognitive impairments [29].

Many other known monogenic disorders also present with ID and ASD such as neurofibromatosis 1 (*NF1*) [30], untreated phenylketonuria (*PAH*) [31, 32] and Cohen syndrome (*VPS13B*) [33, 34].

1.2.2 Chromosome abnormalities

Microscopically visible chromosomal anomalies, and submicroscopic gains and losses (also known as copy number variations or CNVs) are identified using karyotype and chromosomal microarray analyses, respectively.

1.2.2.1 Microscopically visible cytogenetic aberrations

Microscopically visible chromosomal anomalies (>5 MB) are responsible for ∼15% and 5-7% of cases with ID and ASD, respectively [10, 35]. Autosomal trisomies and X- chromosome aneuploidies typically cause some degree of ID of which trisomy 21 is the most frequent form, accounting for 8% of ID cases [16]. Many microscopically visible chromosomal rearrangements are also associated with NDD such as deletion of chromosome 4p16 (Wolf–Hirschhorn syndrome), duplication of chromosome 4p16 (Partial trisomy 4p syndrome) [36-38], and detectable chromosomal rearrangements, mostly deletions, involving 15q11-13 (Prader-Willi syndrome and Angelman syndrome) [39, 40]. Maternally derived duplication of 15q11-13 with a frequency of 1% is the most frequently observed microscopically visible chromosomal abnormality in ASD patients [41].

1.2.2.2 Submicroscopic gains and losses (Copy number variations)

Submicroscopic microdeletions (deletions) and microduplications (duplications) not detectable from standard karyotype were initially discovered using region-specific DNA probes and Fluorescent *In Situ* Hybridization (FISH). This technique is useful in the diagnosis of several deletions accountable for clinically distinct ID syndromes such as Williams Beuren syndrome (WBS), Prader-Willi /Angelman syndromes (PWS/AS), and DiGeorge syndrome (DS) [42-45]. Subtelomeric unbalanced rearrangements are also causative for 0.5-7.4% of patients with ID [46].

Detection of copy number variants (CNVs) in a genome wide manner was first described in 1999 using chromosomal microarray (CMA) analysis [47-49]. CNVs are defined as deletions or duplications of DNA segments greater than 1 kb size which do not arise from insertion or deletion of transposable elements [50]. CNVs are a major class of human genomic variation and play a significant role in human diversity, evolution and disease [51]. Almost 35% of the human genome includes CNVs [52, 53]. The frequency of large CNVs (>500 kb) in the general population is 5-10% with rare CNVs comprising only <1% [54].

Several mechanisms contribute to CNV formation. One of them is non-allelic homologous recombination (NAHR) that occurs between low copy repeats (LCRs) or segmental duplications (SDs) [55-57]. LCRs are repeated DNA segments of >1-5 Kb in size and >95% sequence identity occurring twice or more in the genome [58]. LCRs are found in about 9% of the human genome (130 hotspots) predisposing the genome to recurrent CNV formation during meiosis. As a result about 3000 genes located in these hotspots are prone to dosage alteration [59, 60]. However, most CNVs are non-recurrent and individually rare, created by mechanisms other than NAHR within chromosomal regions with little or no homology but complex genomic

architecture. Mechanisms such as non-homologous end joining (NHEJ), and micro-homology mediated repair mechanisms contribute to non-recurrent CNVs formation [61].

1.2.2.2.1 CNV classification

Based on their clinical relevance, CNVs are classified into three categories including pathogenic, benign, and variants of uncertain significance (VOUS) [62, 63]. Different factors contribute to the pathogenicity of CNVs including gene content, size, type of CNV (gain or loss), inheritance pattern and its frequency in the general population [64]. Classification of CNVs is aided by catalogues of CNVs reported in controls (DGV: http://dgv.tcag.ca/dgv/app/home) and patients (DECIPHER: https://decipher.sanger.ac.uk/; and ECARUCA: www.ECARUCA.net).

Pathogenic CNVs are considered disease-causative and include CNVs that overlap with known genomic syndromes such as deletion of 7q11.2 (WBS) and deletion of 22q11.2 (DS), or CNVs that are commonly found in individuals of similar phenotype. *De novo*, large (>400 kb), gene-rich CNVs which contain either dosage sensitive genes, genes that are members of established family/pathway, or genes involved in synapse formation, transcription and embryonic development are more likely to be pathogenic [50, 65, 66]. Deletions clinically are more likely to be pathogenic than duplications [64, 67, 68]. Duplications are better tolerated than deletions as the duplicated genes have redundant functions [69]. In addition, CNVs smaller than 400 Kb with relevant gene content or close to regions with known reported phenotypes are also treated as pathogenic. Familial CNVs may be considered as pathogenic if the CNV contains imprinted genes [44], is located on the chromosome X in affected males showing X-linked familial segregation (e.g., deletion of Xp21) [70], or when associated with biallelic deletion/mutation of a candidate disease gene unmasking an autosomal recessive condition [71, 72]. Of note, rare CNVs

with higher frequency in affected subjects compared to controls are referred to as putatively pathogenic or susceptibility factors [73-75].

Benign CNVs are present in >1% of the general population (copy number polymorphism) or detected repeatedly in phenotypically normal individuals. Benign CNVs are more likely located outside of genic and ultra-conserved areas [76]. Genes located in benign CNVs are mostly involved in sensory perception, cell adhesion and immunity [77, 78]. CNVs transmitted from a normal parent are usually considered as benign, other than those aforementioned [50].

Variants of Unknown Significance (VOUS) CNVs cannot be classified as either benign or pathogenic [62, 63]. They are more frequent among inherited CNVs than *de novo*, with higher rate of duplications than deletions [64, 79]. Kaminsky *et al.* reported that VOUS CNVs account for 9.3% of individuals with DD/ID, ASD, and/or multiple congenital anomalies; and also that 9% of *de novo* CNVs are considered as VOUS [64].

1.2.2.2.2 CNVs involved in NDD

Individually, rare CNVs contribute to 15–25% of NDD [62, 79]. *De novo* or rare inherited deletions and duplications occur in 10-15% of ID subjects [50]. *De novo* CNVs account for 5–8% of ASD cases in simplex families [80, 81]. Genomic disorders (known recurrent deletion (del) or duplication (dup) syndromes) are frequently associated with cognitive and developmental abnormalities. These genomic disorders include del(7)(q11.23) (WBS), del(15)(q11-q13)pat (PWS), del(15)(q11-q13)mat (AS), del(17)(p11.2) (Smith-Magenis syndrome (SMS)), dup(17)(p11.2) (Potocki-Lupski syndrome (PLS)), del(22)(q11.2) (DS) and dup(X)(q28) (MECP2 duplication); all impacting ID and/or ASD phenotype [82].

Many recurrent rare CNVs are emerging and assumed to be important risk factors for NDD. These include CNVs at 1q21.1 and 16p11.2, dup15q11-q13, dup22q11.2, and delXq28

which all have been found in individuals with ID and/or ASD and other NDD including, schizophrenia and epilepsy, but also in cognitively unaffected controls [82]. Generally, the frequency of deletions is higher in affected individuals compared to duplications [67, 83]. Incomplete penetrance and variable expressivity are common findings among patients with such recurrent CNVs, therefore are considered predisposing to NDD based on large case-control studies showing higher incidence in affected versus control populations [73-75].

1.3 Approaches for culprit-gene discovery within CNVs and beyond

CNVs involved in NDD mostly contain several unrelated genes and thus identification of functionally relevant genes for a certain phenotype is very challenging. So far, several approaches have been employed to identify culprit gene(s) within CNVs.

1.3.1 Phenotype-genotype correlation

The purpose of this approach is to find the minimal critical region among patients with similar phenotypes and with overlapping CNVs yet different breakpoints. Investigation of such CNVs helped to narrow down the 17q21.31 region and led to identification of *KANSL1* associated with 17q21.31 microdeletion syndrome [67, 84]. Identification of a 100kb region at the distal end of del22q13 in a patient with autism and mild ID helped in the discovery of *SHANK3* [85]. Thereafter, study of overlapping del22q13 in 45 ID patients led to further characterization of *SHANK3* [86]. Moreover, atypical and small deletions and duplications of 17p11.2 pointed towards *RAI1* as the major disease-causing gene due to haploinsufficiency or over-expression in SMS and PLS, respectively [87-89]. In Chapter 2, I use this approach to look for phenotypic features of subjects with a dup16p11.2.

1.3.2 Gene expression analysis

Study of the gene products, namely RNA and protein, helps to better understand the relation between the genome and cellular function. A growing number of studies indicate that CNVs affect the expression of genes within the CNV and its nearby genes sometimes megabases away from the CNV's breakpoints [90-96]. Phenotypic effects of genetic variations like CNVs appear to reflect altered expression levels, either by directly affecting the genes within CNVs, or indirectly through position effects of downstream or regulatory pathways [97, 98]. Therefore, the integration of gene expression study and CNV data may help to prioritize CNV candidate gene regions.

Expression microarrays are widely used to determine the genome wide effect of CNVs on gene expression, The first study to explore the impact of CNVs on gene expression in lymphoblast cell lines (LCLs) reported that CNVs are contributing to ~20% of variation in gene expression [91]. Harvard *et al.* performed whole genome expression (WGE) using LCL samples of carriers of 1q21.1 CNV and demonstrated a positive association between copy number and the expression levels of 50% of integral genes, and identified that the protein expression of two genes within the CNV region, *CHD1L* and *PRKAB2,* were reciprocally altered in deletion and duplication carriers [99]. Soon after, Ye *et al.* performed WGE in cerebellar tissue samples of cases with psychiatric disorders and showed a positive correlation between copy number and expression level of genes within 1q21.1 and 22q11.2 CNV regions. Interestingly, they found that the effect of duplications was smaller than that of deletions [100]. In Chapter 3, I used this approach to look for candidate genes in the 16p11.2 duplication region.

1.3.3 Phenotype consequences of CNV gene knock-down

Traditionally, the causality of candidate genes is confirmed when it is found in several unrelated but phenotypically similar patients or when it is validated by *in vivo* or *in vitro* modelling. Animal models are important tools to demonstrate pathogenicity and to establish a link between the specific phenotype and a certain gene or CNV. Mouse, zebrafish and fruit fly are common animal models to evaluate the functional consequence of rare CNVs.

Mouse is the standard animal model for functional study, partly because 99% of mouse genes have human orthologs and exhibit a large synteny with the human genome [101]. Mouse models have been generated for many rare CNVs such as CNVs at 7q11.23 [102-104], 15q13.3 [105], 16p11.2 [106, 107], 17p11.2 [108, 109] and 22q11.2 [110-112].

Zebrafish is an interesting animal model due to the ability to study large populations in a short period of time and to easily inspect the organ involvement through their transparent embryos. About 70% of the zebrafish genome has human orthologs [113, 114]. To date, zebrafish has been used to evaluate the effect of CNVs and to prioritize candidate genes within CNVs at 8q24.3 [115], 16p11.2 [116, 117] and 22q11.2 [118, 119].

Drosophila has also been used as an animal model for many years in many different research fields including behavioural and neuronal studies due to short generation time and low cost. About 75% of human- disease genes have orthologs in drosophila [120, 121]. Drosophila has been used in the study of many genes such as *UBE3A* and *EHMT1* associated with AS and Kleefstra syndromes, respectively [122, 123].

1.3.4 Whole exome sequencing (WES)

Next generation sequencing (NGS) is an exciting tool for screening sequence changes within protein coding genes integral to pathogenic CNVs, or genome-wide (WES) [124-126]. NGS

represents a significant breakthrough in para-clinical (relating to techniques or findings that are not entirely clinical such as those of laboratory or radiology) and emerges to remodel medical research and clinical practice by increasing diagnostic yield (i.e. identification of a causative variant in a known or candidate gene), reducing time and cost, and thus avert and alleviate prolonged uncertainties and disease burden shared by patients and their families [127, 128]. The first clinical application of WES in 2009, as a proof of principle, revealed that candidate genes for monogenic disorders can be identified by exome sequencing of unrelated patients [129]. In 2011, WES was adopted for clinical testing in molecular diagnostic laboratories. Currently, individuals with severe ID, ASD, epilepsy, multiple congenital anomalies generally are good candidates for WES. Diagnostic yield from WES varies from 13% in individuals with unexplained ID [130], 20% in sporadic ASD [131], 25% in heterogeneous patient populations and 45% in NDD patients [127].

Notably, *de novo* variations are the most common type of pathogenic mutations in children with NDD with a diagnostic yield of 44% [128]. The influence of advanced paternal age on the incidence of *de novo* mutations in affected offspring underscores the importance of genetic counselling in individuals at-risk [132]. Rauch *et al.* revealed that *de novo* mutations could explain 45–55% of patients with severe ID [133]. WES studies of cases with ASD showed that the vast majority of *de novo* mutations are only potential risk factors, not exclusively causative, and occur in several hundred different genes, indicative of an oligogenic/polygenic model and extreme genetic heterogeneity of ASD [134-137]. Furthermore, WES uncovered recessive-acting mutations that are directly causative of ASD/ID, in both consanguineous [138-141] and non-consanguineous families [142]. Lim *et al.* reported that about 3% of ASD cases

may be affected by rare autosomal recessive variants, and that rare hemizygous mutations on the X chromosome account for 1.7% of male ASD [143].

Although genomic microarrays are recommended as first-tier tests for the postnatal evaluation of individuals with NDD and/or multiple congenital anomalies [62, 144], some pathogenic and putatively pathogenic CNVs detected in patients cannot completely explain complex patient phenotypes, particularly when an unaffected parent carries the same submicroscopic imbalance. Multiple studies demonstrated the power of WES to find the genetic etiology of clinical variability among such patients [124-126]. WES helped to discover that the presence of variants on the non-CNV homolog chromosome may unmask biallelic mutations in an autosomal recessive condition [124, 125], or that damaging variants in other parts of the genome may contribute to such variable expressivity [126]. The results of these studies suggest that uncommon or variable phenotypes in patients with known pathogenic CNVs or in patients with CNVs inherited from an unaffected parent may indicate co-occurrence of a secondary contributory genomic event on the alternate homolog or elsewhere in the genome. In Chapter 4, I provide further details of the studies and results encompassing WES for cases with 16p11.2 CNV.

1.4 Genomic and phenotypic characteristics of 16p11.2 CNV- overview and challenges

Reciprocal 16p11.2 deletion/duplication represents one of the best examples of recurrent CNV associated with incomplete penetrance and variable expressivity. The reported frequency of 16p11.2 CNV varies across different studies, but what is mostly known is that this CNV accounts for ~1% of autism cases [145-148] and 1.5% of children diagnosed with significant developmental or language delays [147] compared to 0.04-0.07% of control populations [147, 149].

Human chromosome 16, with a total size of 88.7 Mb, contains 880 protein-coding genes, 19 transfer RNA genes and 341 pseudogenes. Chromosome 16 has one of the highest levels of SDs amongst the human autosomes. Notably, 91 genes are located in SD regions which are prone to instability. In particular, SDs that cluster along the "p" arm of the chromosome and within the pericentromic region (16p11) represent the largest inter-chromosomal duplication zone which mostly map to pericentromeric regions of other chromosomes [150].

A chromosomal region of 593 kb at 16p11.2 (genomic coordinates 29.5 Mb to 30.1 Mb in hg 19) is flanked by two major 147 kb segments of LCRs which share 99.6% identity (Figure 1.1). Adjacent and distal to the telomeric 147 kb repeat, there is a ~72 kb region sharing 98.6% similarity with two counterparts within both 147 kb LCRs and in direct orientation [151] (Figure 1.1). The regions of LCRs confer genomic instability and predispose to NAHR during meiosis causing *de novo* copy number alteration [73]. This 16p11.2 CNV region contains 26 protein-coding genes, and the flanking LCRs include three additional duplicated genes (*BOLA2/B, GIYD1/2, SULT1A3/4*) [147] (Table 1.1), of which 22 genes are expressed in the developing human brain [152].

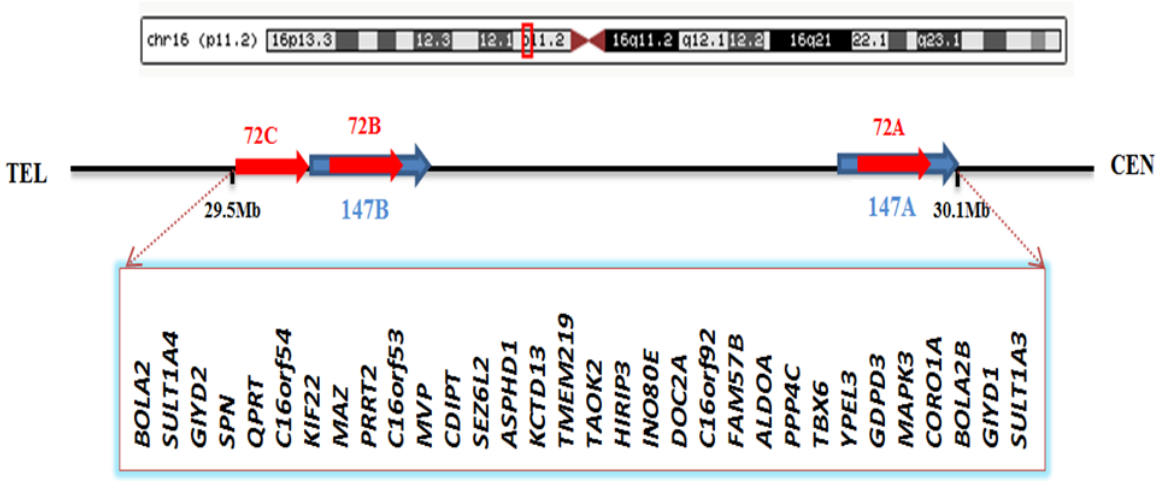

Figure 1.1: 16p11.2 genomic region.

Red and blue bars represent low copy repeats (LCRs) flanking the recurrent rearrangement region (Modified from Shinawi et al. 2010[151]).

Table 1.1: The list and function of genes located within the recurrent 16p11.2 CNV region.

Genes	**Protein Function** (http://www.genecards.org/)
BOLA2/ BOLA2B	Involved in cell proliferation /cell-cycle regulation
GIYD1/ GIYD2	Regulators of genome stability
SULT1A3/ SULT1A4	Catalyze the sulfate conjugation of phenolic monoamines (such as dopamine, norepinephrine and serotonin) and catechol and phenolic drugs
SPN	Involved in the physicochemical properties of the T-cell surface and lectin binding
QPRT	Plays a role in catabolism of quinolinic acid. Quinolinic acid has a potent toxic effect on neurons
C16orf54	Unknown
KIF22	Plays a role in spindle formation and the movement of chromosomes during mitosis and meiosis
MAZ	May act as a transcription factor with dual roles in transcription initiation and termination
PRRT2	A trans-membrane protein containing a proline-rich domain in its N-terminal half. Studies in mice suggest that it is mainly expressed in brain and spinal cord in embryonic and postnatal stages
C16orf53	A part of a Set1-like multiprotein histone methyltransferase complex that displays histone H3 lysine-4 (H3K4) methyltransferase activity
MVP	An essential component of the vault complex. Vaults may be involved in nucleo-cytoplasmic transport. The encoded protein may involve in multiple cellular processes by regulating the MAP kinase, phosphoinositide 3-kinase/Akt and JAK/STAT signaling pathways. This protein also plays a role in drug resistance, and may also be a prognostic marker for several types of cancer
CDIPT	Involved in catalysing the biosynthesis of phosphatidylinositol (PtdIns) as well as PtdIns:inositol exchange reaction. PtdIns is an important signal molecule within the nervous system
SEZ6L2	May contribute to specialized endoplasmic reticulum functions in neurons. Increased expression of this gene has been found in lung cancers
ASPHD1	Unknown
KCTD13	Involved in the regulation of actin cytoskeleton structure and cell migration
TMEM219	Cell death receptor specific for insulin-like growth factor-binding protein (IGFBP3), may mediate caspase-8-dependent apoptosis upon ligand binding
TAOK2	Plays a role in many various cellular processes including cell signaling, microtubule organization and stability, and apoptosis
HIRIP3	Maybe involved in chromatin function and histone metabolism

Genes	**Protein Function** (http://www.genecards.org/)
INO80E	May act as regulatory component of the chromatin remodeling INO80 complex which is involved in transcriptional regulation, DNA replication and probably DNA repair
DOC2A	Involved in Ca^{2+} -dependent neurotransmitter release. It is mainly expressed in brain
C16orf92	Unknown
FAM57B	Involved in ceramide synthesis. Ceramides are a family of waxy lipid molecules
ALDOA	Involved in glycolysis and gluconeogenesis. In addition, may also act as scaffolding protein
PPP4C	Plays a role in many processes such as microtubule organization at centrosomes, maturation of spliceosomal snRNPs, DNA repair, DNA damage checkpoint signaling, apoptosis, tumor necrosis factor (TNF)-alpha signaling, activation of c-Jun N-terminal kinase MAPK8, regulation of histone acetylation, NF-kappa-B activation and cell migration
TBX6	Act as a regulator of developmental processes. It may play an essential role in left/right axis determination
YPEL3	Plays a role in proliferation and apoptosis in myeloid precursor cells
GDPD3	Unknown
MAPK3	A member of the MAP kinase family. MAP kinases act in a signaling cascade that regulates different cellular processes such as proliferation, differentiation, and also cell cycle progression in response to a variety of extracellular signals
CORO1A	May be an essential component of the cytoskeleton of highly motile cells

Deletion of 16p11.2 was first reported in monozygotic twins presenting with ID, seizure and aortic valve anomalies [153]. Thereafter, a *de novo* del16p11.2 was observed among simplex autism families [73]. Currently, a growing number of studies are pointing to the 16p11.2 CNV region as being associated with a broad spectrum of NDD. Several studies have proposed that recurrent deletion and duplication of the 16p11.2 region have mirror phenotypes, with autism [145, 147, 154], developmental delay [147, 151, 155, 156], and macrocephaly more often observed with del16p11.2 [151]; whereas dup16p11.2 presents with psychiatric disorders (ADHD and schizophrenia) [157] and microcephaly [151]. Additionally, studies show that del16p11.2 confers risk of obesity often associated with hyperphagia [95, 156], whereas the corresponding reciprocal duplication is more common to an underweight body mass index (BMI) with increased risk of anorexia [94]. These observations may support a diametric model of autism versus psychotic behavior and respective mirror somatic phenotypes in reciprocal deletion and duplication involving 16p11.2 [151]. However, debate still continues on the specific associated risk of autism, ID and psychiatric disability among deletion versus duplication carriers. A summary of phenotype features seen in individuals with 16p11.2 CNV is described in more detail in the introduction to chapter 2.

Growing evidence indicates that recurrent reciprocal deletion and duplication of 16p11.2 are delineated by a broad spectrum of neurocognitive phenotypes associated with variable expressivity. The phenotype of some individuals with 16p11.2 CNV, particularly duplication, can also fall within a normal range [146, 151, 155]. The estimated penetrance of deletion and duplication of the 16p11.2 CNV region are 46.8% and 27.2%, respectively [149]. There is familial coincidence of both phenotypically affected and unaffected carriers in some

families [146, 147, 158, 159]. Furthermore, dupl16p11.2 has also been reported in control cohorts without ID or autism, at 0.03-0.04% frequency [67, 149].

Finding the explanation for inter- and intra-familial clinical variability among cases with dup16p11.2 is challenging. Although genomic gains of 16p11.2 may directly affect clinical phenotype by affecting dosage-sensitive genes within the CNV, this cannot explain the normal phenotypes of parents who carry the same CNV as their affected children. Common and/or rare functional variants in one or more genes within the CNV region may confer susceptibility to NDD; however, there has been no firm evidence to date supporting this notion [148]. Additionally, there are no imprinted genes recognized within the 16p11.2 region [151]. There is also a possibility that carriers described as having no particular phenotype might still exhibit an anomalous phenotype if examined by a clinical geneticist or psychiatrist. However, concrete evidence of specific phenotype-genotype correlation in a large cohort is needed. Finally, there is the possibility that dup16p11.2 is not pathogenic, and its co-occurrence with one or more different cryptic imbalances in another part of the genome is responsible for the large phenotype variability observed.

1.5 Research objectives

The overall goal of my thesis is to apply some of the above described gene discovery approaches to look for genetic causes of phenotypic variability in individuals with familial dup16p11.2.

I hypothesize that while some unique features found in individuals with dup16p11.2 may be due to over-expression of its integral genes, co-occurrence of other genetic alterations in the genome may account for the variability in their clinical phenotypes. Objectives of this study are:

1) To investigate the clinical phenotypes of carriers of dup16p11.2.

2) To investigate the expression of integral genes within and flanking the 16p11.2 region in dup16p11.2 carriers.
3) To investigate the genome of patients with dup16p11.2 for presence of sequence changes that could explain their clinical variability.

This study aims to improve our understanding of the bases for significant phenotypic variability observed among individuals with dup16p11.2, the pathogenic mechanisms involved, and relevance to NDD etiology. Further, this study aims to assist clinicians in providing more meaningful, personalized genetic counselling for individuals and families living with these disorders.

Chapter 2: Phenotype investigation of subjects with dup16p11.2

In the medical setting, "phenotype" refers to deviations from normal morphology, physiology and behavior. Detailed phenotyping is the precise and comprehensive analysis of phenotypic differences observed in the individual [160], and has inherent powers to enable the diagnosis, especially in rare diseases, and to distinguish similar disorders. Such phenotyping along with genotyping helps to sub-classify genetic diseases more precisely, which subsequently aids informed genetic counselling and guides best-fit prevention, intervention and treatment options. Therefore, detailed phenotyping is beneficial not only in clinical practice, but also in medical genetic research, particularly for the meaningful interpretation of CMA and NGS, hampered by genotype-phenotype variation. [161]. Detailed phenotype analysis is the essential translational bridge that connects genomic biology to human pathology.

A growing number of CMA and NGS studies have shown an increased chance of detecting VOUS, in which extensive filtering strategies and particularly phenotyping are required to make a decision about the pathogenicity of each variant. The study of 1000 exomes revealed that each individual exome contains almost 20,000 single nucleotide variations (SNVs), of which 239 are disease-causing variants [162]. Although reverse phenotyping (phenotype clarification based on genotype) is an emerging new approach [163], detailed and standardized phenotyping of the studied subjects is needed to narrow down the list of candidate disease-causing variants. However, imprecise clinical terms, lack of accurate description of patient features, and utilizing different technical or measuring standards diminish the integrative power of phenotype-genotyping. To overcome this challenge, the usage of standardized terminologies and measurements is crucial. Consistent with stringent quality standards applied for genotype analysis, phenotyping also mandates precise and reproducible methods [163].

The usage of human phenotype ontology (HPO) terms allows researchers a better standard to communicate and correlate phenotype-genotype findings and offers an opportunity to optimize scientific insight, data integration and derive clearer clinical meaning from automated genome analyses. Detailed phenotyping facilitates new gene/disease discovery and helps to find the pathophysiology of pleiotropic diseases as well as genotype-phenotype relationships.

Recently, we published the results of detailed phenotyping of 78 patients with ID using standardized HPO terms derived from the Winter-Baraitser Dysmorphology Database (WBDD) integrated with microarray data [164]. The analysis showed an increased frequency of cranial and forehead abnormalities in cases with rare, and more likely pathogenic CNVs (*de novo* and familial) compared to ID cases with common CNVs [164]. In addition, systematic categorization of patients' phenotypes according to 34 organ systems and 169 sub-classified phenotypes showed that abnormalities of head, hands and feet are most frequently detected in subjects submitted for CMA (> 50% of ID cases in this cohort).

WES and whole genome sequencing (WGS) have recently been applied broadly, not only in patients with high suspicion of genetic disease but also in asymptomatic people who are interested in NGS for their own curiosity or health screening purposes. According to Frebourge *et al.*, it is important that the medical genetics community inform the patient, the public and policy makers that genomic analysis has limited value without precise and detailed phenotyping [165]. Detailed phenotyping is particularly important for assessment of the significance of familial CNVs, necessitating close clinical examination of both the affected children and their apparently unaffected carrier parent. Recently, integration of phenotype and genotype data from exome sequencing of a large cohort of ID patients led to the discovery of four new autosomal disorders caused by mutations of *KIAA0586*, *HACE1*, *PRMT7* or *MMP21* [166].

It is expected that loss or gain of genes or chromosome regions produce different clinical features. However, patients with deletion and duplication of 16p11.2 region have very broad and often overlapping neurocognitive phenotypes, and thus it is difficult to differentiate cases with deletion from duplication based only on their clinical phenotypes. A summary of reported features among patients with deletion and duplication of this region is described in below.

2.1 Phenotypes associated with 16p11.2 CNV (literature review)

2.1.1 Morphometric abnormalities

The first study focusing on head circumference (HC) and height of patients with 16p11.2 CNV was conducted by McCarthy *et al.* in 2009. In this study, the authors showed that HC was greater in individuals with deletion than duplication, but they did not find a significant difference in the height of duplication and deletion groups [158]. The association of macrocephaly and del16p11.2 was further confirmed by Shinawi *et al.*, who also noticed that a subset of duplication carriers were microcephalic, yet did not reach statistical significance when compared to control populations [151]. Subsequently, it has become accepted that 16p11.2 deletion confers risk of obesity [95], whereas the corresponding reciprocal duplication is more common in patients with low BMI [94]. However, recent evidence is suggestive of significantly higher BMI in children with del16p11.2 compared to controls, while no significant difference was seen in adult deletion carriers [167].

2.1.2 NDD, psychiatric disorders and epilepsy

Individuals with 16p11.2 CNV have been found among very diverse cohorts of NDD and psychiatric disorders. The estimated frequencies of 16p11.2 CNV among cases with ID and ASD are slightly different, depending on the type of family (simplex or multiplex) or specific population involved (e.g. Icelandic) [147, 154]. However, the cumulative frequency of deletion

and duplication of 16p11.2 region is 1% among individuals with ASD [145-148] and 1.5% among children diagnosed with significant developmental or language delays without autism [147], compared to 0.04-0.7% in the control population [147, 149]. Moreover, the frequency of dup16p11.2 in schizophrenia patients is higher (0.3-0.6%) compared to controls (0.03-0.04%) [158]. McCarthy *et al*. performed a meta-analysis on phenotype data derived from large cohorts of subjects manifesting developmental delay and psychiatric disorders and found a significant reciprocal association between subjects with dup16p11.2 having autism and schizophrenia, whereas deletion subjects showed developmental delay and autism but not schizophrenia or bipolar disorder [158]. Hanson *et al.* recently reported that more than 90% of patients identified with the del16p11.2 presented with developmental and psychiatric disorders; and 24% with ASD. Further, these individuals demonstrated varying degrees of ID with the average IQ 1 SD below the population mean and 1.8 SD below family controls [168]. Thus, notable controversy exists on the correlation of deletion or duplication of 16p11.2 with specific types of NDD.

Other psychiatric problems, especially ADHD, are also common among individuals with 16p11.2 CNV. The notion of higher frequency of ADHD among patients with duplication was first introduced by Shinawi *et al.* [151]; but multiple studies have since emerged showing that some deletion cases are also affected by ADHD [151, 155, 169, 170]. Of note, dup16p11.2 has also been found in patients with Alzheimer's disease who suffer from psychosis (0.46%), but not among those without psychosis [171].

Speech and language impairment, another subgroup of NDD, is one of the most common observed findings among patients with both deletion and duplication of 16p11.2 [145, 146, 151, 153, 156, 172, 173]. Approximately 71% of individuals with the deletion have a speech and

language-based developmental disorder including articulation problems, and expressive or mixed receptive-expressive language deficits [174].

Motor delay is another common feature of patients with 16p11.2 CNV [146, 155, 156, 175]. Recent reports suggest that Developmental Coordination Disorder (DCD) including impaired motor skills, clumsiness and failure to achieve developmental milestones presents in 53% patients with del16p11.2 [174].

The estimated frequency of dup16p11.2 among cases with rolandic/atypical rolandic epilepsies is 1.3-1.5%; 25-fold higher than the general population [176]. Although epilepsy has been considered as an important feature of 16p11.2 duplication, it also presents among cases with deletion [146, 151, 156, 173, 177].

2.1.3 Dysmorphic features and congenital anomalies

Dysmorphic features are common among individuals with 16p11.2 CNV [146, 151, 178]. Carriers of del16p11.2 and particularly duplication also suffer from a broad range of congenital anomalies (CAs). CAs involve brain [151, 153, 172, 177, 179-181], heart [153, 156], urinary/ reproductive tract [146, 151, 152, 155, 182], eye [156, 183] and vertebra [146, 155, 159, 172, 177, 184]. Umbilical and diaphragmatic hernias have also been reported among both deletion and duplication carriers [145, 146, 151, 177]. Observed structural brain abnormalities include cerebral cortical atrophy [153], partial temporal lobe agenesis [155], large cerebral ventricles [178, 179] and hyperplasia of corpus callosum [152]. 48% of deletion carriers are either left-hand or mixed-hand dominant compared to 14% of non-carrier family members [174], which suggests potential differences in brain development. Correlation of molecular and neuroimaging techniques indicate that the number of genomic copies of 16p11.2 negatively correlate with gray matter volume and white matter tissue properties in cortico-subcortical regions involved in

reward, social and language cognition [181]. Further, carriers of del16p11.2, compared to controls, present with increased axial diffusivity in most areas of central white matter including the anterior corpus callosum, bilateral internal and external capsule [180]. Apart from the structural brain abnormalities, presence of spinal anomalies has frequently been reported among patients with both deletion and duplication of 16p11.2. These anomalies include hemi-vertebrae, syringomyelia, osteoarticular malformation and scoliosis [146, 155, 159, 167, 172, 177, 184].

2.1.4 Susceptibility to infection

Increased susceptibility to infection is frequently reported in patients with del16p11.2. This susceptibility ranges from recurrent urinary tract infection [146], otitis media [151, 156], upper respiratory infection [153, 185] to fulminant hepatitis [153] and severe combined immune deficiency (SCID) [170]. A recent genome wide association study (GWAS) conducted by Maggadottir *et al.* identified an association between common variable immunodeficiency (CVID) with variants at 16p11.2 region. Nonetheless, they did not find any corresponding rare variants within candidate genes (*FUS*, *ITGAM*, and *ITGAX*) located in this region [186].

2.2 Patients and methods

Two dup16p11.2 carrier families involving an apparently healthy carrier parent and child with NDD were randomly selected from consented patients recruited for ID and ASD research through the B.C. Provincial Medical Genetics Program and Child & Family Research Institute of BC Children's and Women's Health Center. Ethics approval for clinical research involving human subjects was obtained through the joint Clinical Research Ethics Board of the University of British Columbia and BC Children's and Women's Health Center (approval number: C01-0507; Vancouver, B.C.).

The clinical data from these two study subjects were collected principally through chart review as both families had previously undergone clinical genetic assessment before inception of this project. I thoroughly reviewed their charts and summarized the positive findings and then systematically classified their clinical features according to the WBDD, also known as the "London Dysmorphology Database" (http://www.lmdatabases.com/about_lmd.html) downloaded in June 2012.

WBDD contains a very detailed classification of human clinical phenotypes comprising 34 body systems, 196 sub-classifications and > 1000 dysmorphic features, congenital anomalies and clinical abnormalities. In order to simplify these categories, I adopted the body systems category, and reduced them to 12 groups by merging some categories. The list of merged categories of WBDD ontology is shown in Table 2.1. The detailed phenotypes of Proband A and B are categorized and classified according to the revised WBDD categories in Table 2.2. Next, I extracted the clinical phenotypes of published patients with dup16p11.2. To do this, I searched the "PubMed" using the keyword "16p11.2" and then reviewed each publication abstract. Among 269 hits (18th September 2015), five publications presented the clinical features of 16 patients with dup16p11.2 [146, 151, 152, 172, 178]. I systematically reviewed and categorized their clinical features using the WBDD classification.

Table 2.1: List of merged categories of the Winter-Baraitser Dysmorphology Database

Merged categories	New category
Abdomen, thorax	Abdomen/thorax
Back and spine, build, neck, stature	Build/stature
Eye (associated structures), Eye (globes)	Eye
Face, forehead, nose	Face
Hair, skin, blood vessels	Hair/skin
Hand, feet, joint, lower extremities, nails, skeletal, upper extremities	Extremities
Haematology/Immunology, endocrine	Haematology/Immunology
Mouth, oral region, teeth	Mouth/oral region
Neurology, speech (voice)	Neurology
Genitalia, pelvis, urinary tract	Genitalia/urinary tract

Table 2.2: Detailed phenotyping of Proband A and B.

Body system	Proband A-male (4 and 8 y/o)	Proband B-male (14y/o)
Abdomen/thorax	Diastasis recti, hypoplastic nipples	-
Build/stature	Underweight, short stature	-
Cranium	Microcephaly, flat occiput	Microcephaly
Ears	Large, low set, posteriorly rotated ears, auricular pits, unilateral sensorineural hearing loss	Thickened helixes
Eye	Hypertelorism, bilateral ptosis, blepharophimosis, thick eyebrow, long straight eyelashes, high myopia, diffuse retinal dystrophy	Synophrys, widened palpebral fissure, myopia, nystagmus
Extremities	Bilateral club foot, tibial torsion, prominent heels, hyper-extensible joint	Low muscle tone, bilateral prominent toe pads
Face	Shallow slanting forehead, low anterior hairline, micrognathia, mild retrognathia, malar hypoplasia, depressed nasal root, short triangular nose	Short forehead , low anterior hairline, hemi-facial asymmetry, high nasal bridge and root
Genitalia/urinary tract	Bilateral cryptorchidic testes, hypospadias	Right cryptorchidism
Haematol/Immunology	Chronic anemia, recurrent UTI	Occasional URI
Hair/skin	Double hair whorls, small cafe au lait spots	Double hair whorls with widow pick, linear hypo-pigmentation streak, small cafe au lait spots, dry skin
Mouth/oral region	Small mouth, thick upper lip, thickened alveolar ridges, high palate, anterior tongue tie	High palate, thickened alveolar ridges
Neurology	Generalized hypotonia, ID, ASD, verbal apraxia, motor delay, broad-based gait	Hypotonia, ID, ADHD, ODD, self-injurious behavior, onset of hallucinations, infantile seizure, tremor, motor delay, speech delay, vocal tics

Abbreviation: UTI: urinary tract infection; URI: upper respiratory infection; ODD: oppositional defiant disorders.

2.3 Results

I present the results for this Chapter as detailed phenotypes for the two families (2.3.1) and comparison of their phenotypes with findings from literature (2.3.2).

2.3.1 Detailed phenotyping of the dup16p11.2 study subjects

2.3.1.1 Proband A: Subject 12-32A

Proband A is an 11 year old (y/o) boy introduced to our clinic with global developmental delay and verbal apraxia at the age of four. He is the third of four-children of healthy non-consanguineous parents of Chinese descent. His mother and his paternal grand-mother have a history of recurrent spontaneous pregnancy losses. His parents and three siblings are apparently healthy (Figure 2.1). Proband A was born after 39 weeks of uneventful pregnancy via caesarean section (C/S) for fetal distress with Apgar scores of 8 and 9 at one and five minutes after birth, respectively. His birth weight was 2175 gram (<3rd percentile (%ile)), length was 47 cm (10th %ile) and occipito-frontal circumference (OFC) was 34 cm (25th %ile). The patient exhibited feeding difficulty, low muscle tone, bilateral ptosis, club foot, bilateral undescended testes (cryptorchidism), and flexion contracture of hand and wrist.

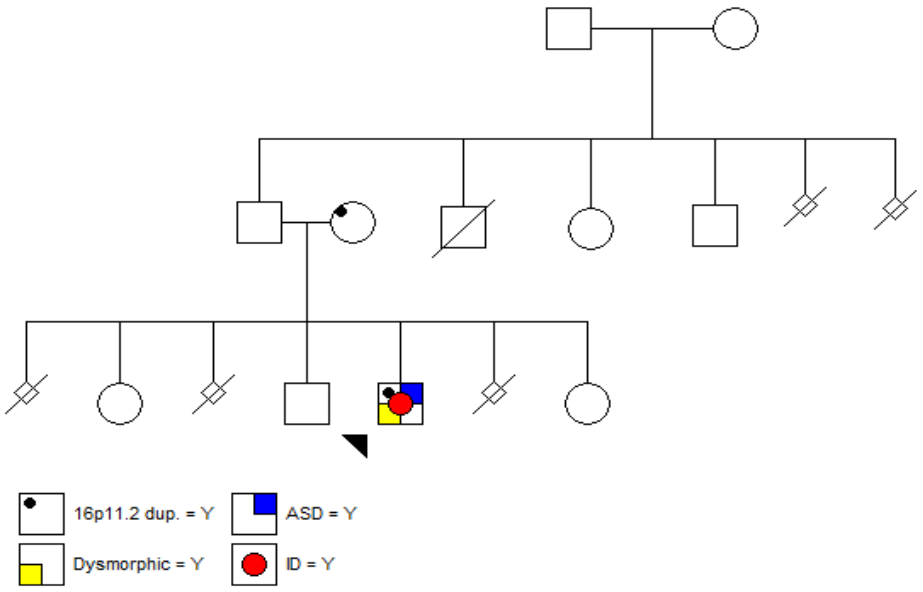

Figure 2.1: Family pedigree of Proband A.

Proband A was examined by clinical geneticist, Dr. Suzanne Lewis, at the ages of four and eight years (Table 2.2). Weight, length and head circumference respectively at age four and eight years consistently plotted at <3rd %ile, <3rd %ile, and >-3SD. The proband's laboratory diagnostic workup was normal and included routine karyotype, subtelomeric FISH, fragile X, biochemical assessment, cranial magnetic resonance imaging (MRI) and computerized tomography (CT) imaging. Affymetrix Genome-Wide Human SNP Array 6.0 revealed a 709.2 kb duplication of 16p11.2 (29,425,199- 30,134,432) in the proband, confirmed by FISH and parental studies indicating maternal inheritance. The proband's siblings were not tested for dup16p11.2 per the family's request.

2.3.1.2 Proband B: Subject 06-32

Proband B is a 15 y/o boy seen by Dr. Lewis with ID, ADHD, oppositional defiant disorder (ODD) and history of epilepsy. Proband B is the second child of non-consanguineous parents originally from Kenya. Family history revealed his older brother to have ADHD and mild ODD with learning disability (diagnosed at age 9 years). His mother has a past history of learning disability, especially difficulty with math, science, and language comprehension, as well as significant attention and memory problems. Proband B's father is apparently healthy. The mother's paternal and maternal cousins had epilepsy. Her maternal female cousin also was physically and mentally handicapped and died at age 23 of unknown cause (Figure 2.2).

Proband B was born at 35 weeks gestation with breech presentation and birth weight of 1873.3 gr (<3rd %ile). His mother had a history of vaginal bleeding and gestational diabetes during pregnancy.

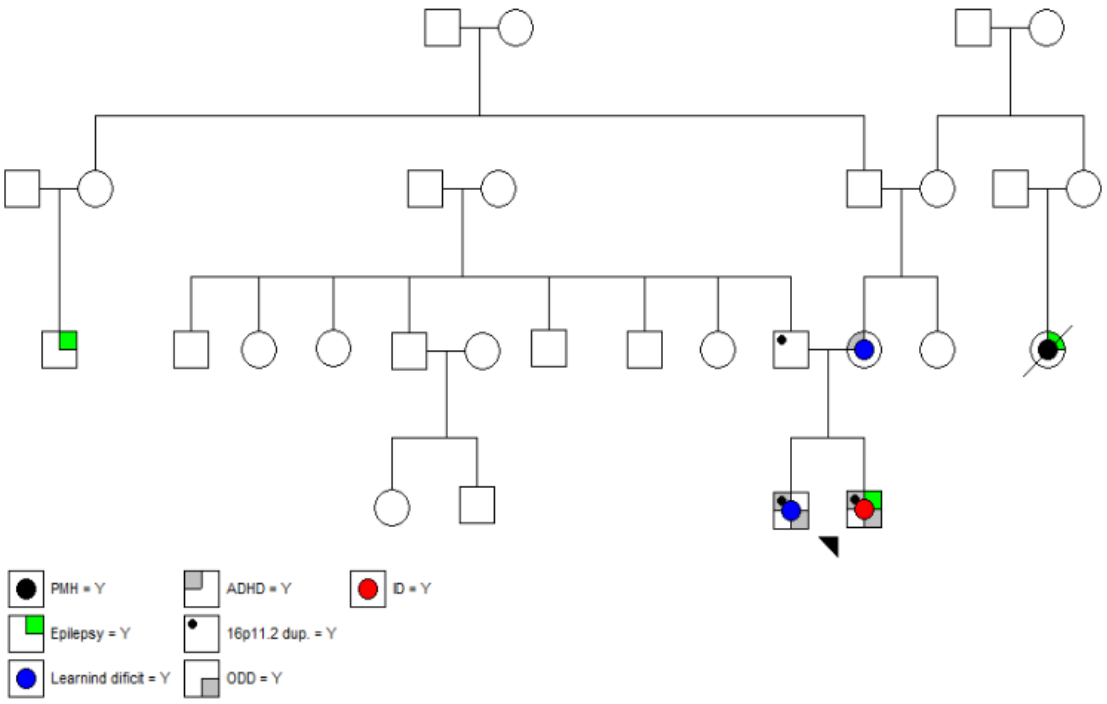

Figure 2.2: Family pedigree of Proband B.

Proband B had been evaluated by Dr. Lewis at the age of 11 and 14. His clinical summary is illustrated in Table 2.2. The MRI imaging performed at two years of age showed mild wavy configuration at the superior border of the corpus callosum as well as faint increase in T2 signal in the globus pallidus likely within normal limits. His electroencephalogram (EEG) report at the age of seven was indicative of benign rolandic epilepsy, and the cerebrospinal fluid (CSF) and neurotransmitter metabolites were normal. Moreover, due to presence of myopathy, a muscle biopsy was performed that suggested "type 1 fiber predominance". Other diagnostic workup including cytogenetic, fragile X, biochemical and hearing assessment were in normal range. The subject was evaluated by two different microarray platforms. Agilent Human Genome CGH Microarray 105K and Affymetrix Cytogenetics Whole-Genome 2.7M Array respectively found a 526.976 kb (29,500,284- 30,027,260) and 811.531 kb (29,303,502- 30,115,033) dup16p11.2, and the paternal origin was confirmed by FISH.

The last evaluation by Dr. Lewis, at the age of 14, revealed that his growth parameters had changed since age 11. For instance, his weight increased from the 5^{th} to 20^{th} %ile, and his height dropped from 10^{th} to 5^{th} %ile; OFC remained consistently below the 2^{nd} %ile.

Since the older brother also shared some of Proband B's clinical features, we initiated testing for dup16p11.2. Subsequent microarray analysis (Agilent105K) confirmed him also to be a carrier of dup16p11.2. The brother of Proband B was not included in the reference cohort as no detailed phenotype information was available for him, however his family physician reported that his height and weight were in the normal range at the age of 13. The 16p11.2 CNVs of Proband A and B are shown in Figure 2.3.

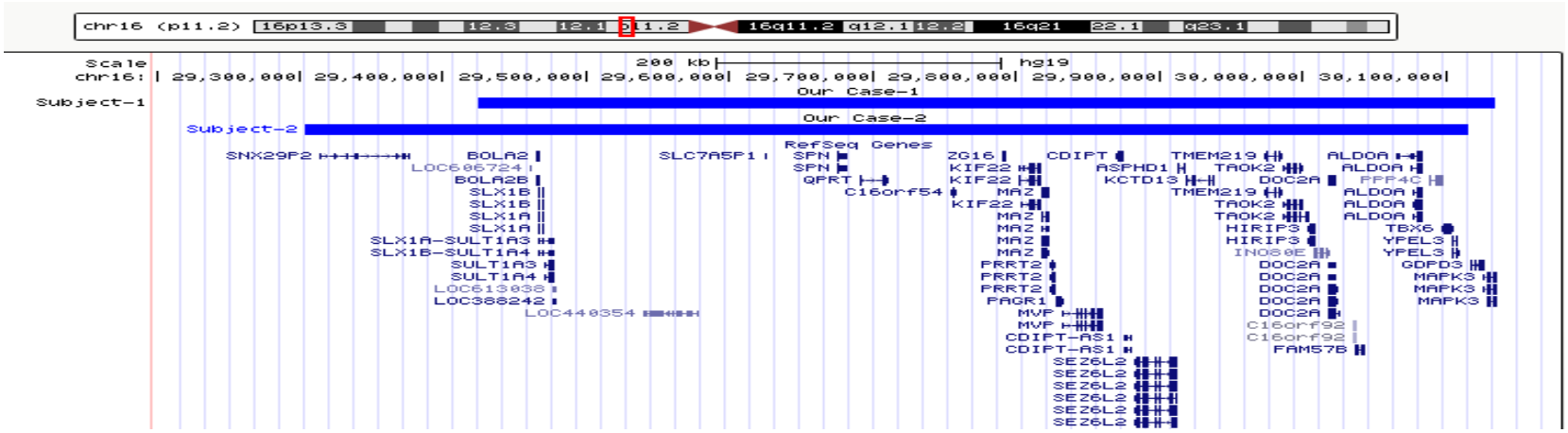

Figure 2.3: 16p11.2 duplications in Proband A and B.

UCSC Genome Browser (http://genome.ucsc.edu/; GRCH37/hg19) screenshot annotates the whole genome array profiles in proband A and B showing 16p11.2 duplications and genes located in this region.

2.3.2 Phenotypic comparison of Probands A/B to published dup16p11.2 cases (from literature review)

In order to compare and contrast our study subjects' phenotypes with those reported for dup16p11.2 cases in the published literature, I systematically reviewed and catalogued positive findings from 16 published cases [146, 151, 152, 172, 178]. Unfortunately, detailed phenotypes were not available for all published cases. Absence of any comment on some features was confounding as it cannot be determined whether the described individual was negative for a specific phenotype or simply was not assessed for that feature, a problem inherent to most retrospective phenotype analyses from published literature. All features were grouped in four categories as described in sections 2.4.2.1 to 2.4.2.3.

2.3.2.1 Morphometric abnormalities

It has been proposed that patients with dup16p11.2 are underweight and microcephalic. I tested this by evaluating the 18 dup16p11.2 subjects in my reference cohort (including current subjects and published cases). I used <3rd %ile as cut-off for OFC/head circumference, length/height and weight. The frequency of microcephaly, short stature and being underweight (postnatally) among the reference cohort are 44.4% (8/18), 11.1% (2/18) and 11.1% (2/18); respectively. The OFC, height and weight of Proband A (10y/o) were <2nd, <2nd and 5-10 %ile, and Proband B (14y/o) were <2nd, 5th and 20th %ile; respectively (Table 2.3).

Table 2.3: Birth and growth parameters of patients with dup16p11.2.

	Birth (%ile)			18mo-18y/o (%ile)		
Study (case)	**OCF**	**Lt**	**Wt**	**OFC/HC**	**Lt/Ht**	**Wt**
Shinawi (1)	-	-	50-90	50	22	65
Shinawi (2)	-	-	25	5	20	2
Shinawi (3)	-	-	80	12	80	50
Shinawi (4)	-	-	50	<3	1	8
Shinawi (5)	-	-	-	<3	20	<3
Shinawi (6)	-	-	85	<3	25	50
Shinawi (7)	-	-	-	88	28	60
Shinawi (8)	-	-	>97	97	67	89
Shinawi (9)	-	-	35	<3	70	90
Bedoyan	-	-	40	2	14	61
Fernandez (1)	-	-	-	50	<5	<5
Fernandez (2)	90	90	75-90	50	5	<5
Fernandez (3)	-	-	50-75	10	3	25
Schaaf	-	-	50	50	50-75	50-75
Filges (1)	25	50	5-10	75	90	40
Filges (2)	-	-0.77*	0.03*	-3.2*	-0.6*	-1*
Current study (Proband A)	25	10	<3	<2	<2	5-10
Current study (Proband B)	-	-	<3	<2	5	20

Abbreviation: Wt: weight; Lt: length; OCF: occipito-frontal circumference; Ht: height;
*: Standard deviations, cut-off of >2SD below the mean is considered as significant.

2.3.2.2 NDD, psychiatric disorders and epilepsy

Individuals with dup16p11.2 have been frequently identified amongst cohorts of NDD and psychiatric disorders. Epilepsy is also reported as a common feature of patients with dup16p11.2. Phenotypic analysis of the reference cohort (Table 2.4) indicates that 100% of patients with dup16p11.2 suffer from ID (18/18) and speech impairments separately (17/17). The frequency of ASD, psychiatric disorders, motor deficit and epilepsy are 56.2% (9/16), 73.3% (11/15), 76.4% (13/17) and 60% (9/15); respectively. Proband A presented with ID, verbal apraxia, ASD and motor delay; and Proband B suffers from ID with a history of speech and motor delay, psychiatric disorder and epilepsy.

2.3.2.3 Dysmorphic features and congenital anomalies

Dysmorphic features are common in patients with dup16p11.2. In order to elicit a clinically recognizable pattern, I characterized all reported facial features of the reference cohort (Table 2.5). Most reported patients did not have detailed facial descriptions, and although some phenotypic traits such as broad nasal bridge/root and smooth philtrum are common, they were not consistent in all patients. Short and smooth philtrum was described in Proband A (10y/o), and broad high nasal root in Proband B. Both probands have low anterior hairline and thick oral alveolar ridges, but their facial features are different from each other and from the rest of cohort.

Evidence suggests that CAs are very common among carriers of dup16p11.2. The analysis of the reference cohort shows that 77.7% (14/18) of patients have one or more CAs including abnormal brain MRI/CT scan, yet no consistent neuro-structural features identified. Multiple CAs were detected in Proband A including cryptorchidism which is the only CA found in Proband B (Table 2.6).

Although susceptibility to infection was not mentioned in any published 16p11.2 cases, Proband A and B both have a positive history of recurrent urinary tract infection (UTI), and upper respiratory infection (URI); respectively.

Table 2.4: NDD, psychiatric disorders and epilepsy in dup16p11.2 patients.

Study (case)	ID/DD	ASD/Autistic	Psychiatric dis.	Speech imp.	Motor imp.	Epilepsy
Shinawi (1)	+	+	+	+	+	-
Shinawi (2)	+	-	-	+	-	-
Shinawi (3)	+	-	+	+	+	-
Shinawi (4)	+	-	+	+	+	+
Shinawi (5)	+	+	-	+	-	-
Shinawi (6)	+	-	-	+	+	-
Shinawi (7)	+	-	+	+	+	+
Shinawi (8)	+	-	+	+	+	+
Shinawi (9)	+	+	+	+	-	+
Bedoyan	+	N/S	N/S	N/S	+	+
Fernandez (1)	+	+	+	+	-	+
Fernandez (2)	+	+	N/S	+	+	N/S
Fernandez (3)	+	+	N/S	+	+	N/S
Schaaf	+	+	+	+	N/S	N/S
Filges (1)	+	N/S	+	+	+	+
Filges (2)	+	+	+	+	+	+
Current study (Proband A)	+	+	-	+	+	-
Current study (Proband B)	+	-	+	+	+	+

Abbreviation: dis: disorders; imp: impairment, -: negative, +: positive, N/S: not specifically reported as positive or negative.

Table 2.5: Dysmorphic features of patients with dup16p11.2.

Study (case)	Face	Forehead	Nose	Mouth	Oral region	Ear	Eye
Shinawi (1)	Prominent maxilla	-	Broad nasal bridge	Flat philtrum, thin lips	-	-	-
Shinawi (2)	-	-	Broad nasal bridge and tips	-	Cleft palate	Prominent, mildly posterior rotated ears	-
Shinawi (3)	Facial asymmetry, micrognathia	-	-	-	-	Large low set ears	Hypertelorism, myopia
Shinawi (4)	Maxillary overbite	-	-	-	-	Prominent ears	Epicanthal folds
Shinawi (5)	-	-	Broad nasal tip, hypoplastic alae nasi	Thin upper lip	-	-	-
Shinawi (6)	-	-	bulbous nose	Wide mouth	-	-	Epicanthal folds
Shinawi (7)	-	-	Broad nasal root and tip	Cleft lip	Cleft palate	-	Hypertelorism, mild synophrys
Shinawi (8)	-	Frontal bossing	-	-	-	Prominent ears	Deep set eyes, telecanthus, myopia
Shinawi (9)	-	Bitemporal narrowing	Tubular nose	-	-	-	-
Bedoyan	-	-	-	-	-	-	Esotropia
Fernandez (1)	-	-	-	Smooth philtrum	-	Abnormal ears	Hypertelorism
Fernandez (2)	-	-	-	Smooth philtrum	-	-	Synophrys
Fernandez (3)	Flat face	Frontal bossing, flat supraorbital ridges/ face	-	Smooth philtrum, thin upper lip	-	-	Sparse eyebrow and eyelashes, deep set eyes
Schaaf	-	-	-	-	Wide spaced teeth	-	-
Filges (1)	Square face	-	Prominent alae nasi	Short philtrum	-	Small ears and lobules	Myopia
Filges (2)	Overbite	-	Broad nasal bridge, bulbous nose	Smooth philtrum, thin lips	-	prominent ears	Deep set eyes, myopia, nystagmus, progressive vision deterioration
Current study (Proband A)	Micrognathia, retrognathia	Shallow slanting forehead, low anterior hairline	Depressed nasal root, very short triangular nose	Small mouth, thick upper lip, short smooth philtrum	Thick alveolar ridge, high palate, tongue tie	Large ear, low set posterior rotator, preauricular sinus opening, hearing loss	Hypertelorism, diffuse retinal dystrophy, high myopia, bilateral ptosis
Current study (Proband B)	Facial asymmetry	Short forehead, low anterior hair line	Broad, high nasal root	-	Thick alveolar ridge	Thick helixes	Synophrys, Myopia, astigmatism

Table 2.6: List of congenital anomalies among patients with dup16p11.2.

Study	Congenital anomalies
Shinawi (1)	-
Shinawi (2)	Hypospadias, cleft palate, pectus excavatum, long third toes
Shinawi (3)	Left torticollis, pectus carinatum, abnormal brain MRI
Shinawi (4)	-
Shinawi (5)	-
Shinawi (6)	Abnormal brain MRI
Shinawi (7)	Cleft lip and palate, pes planus, abnormal brain MRI
Shinawi (8)	Pes planus, phimosis
Shinawi (9)	Tethered cord
Bedoyan	Spastic quadriparesis, cryptorchidism, abnormal brain MRI
Fernandez (1)	Congenital diaphragmatic hernia, scoliosis
Fernandez (2)	-
Fernandez (3)	Oligohydramnios sequence
Schaaf	Thoracolumbar syringomyelia, low set nipple, shawl scrotum
Filges (1)	Scoliosis and lordosis, short thumb, leg discrepancy, planovalgus, Abnormal MRI
Filges (2)	Abnormal MRI
Current study (Proband A)	Club foot, bilateral ptosis, cryptorchidism, hypospadias, hypoplastic nipple, abdominal rectus diastasis, scoliosis
Current study (Proband B)	Cryptorchidism

2.3.3 Clinical phenotypes of dup16p11.2 carrier parents

The bases for phenotypic variability among carriers of dup16p11.2 are difficult to identify when the CNV is inherited from an unaffected parent, as not all features are evaluated consistently or in a standardized manner. There remains a possibility that carrier parents described in the literature as having no particular phenotype may evidence an abnormal phenotype upon detailed phenotypic evaluation.

The mother of proband A, who is a carrier of dup16p11.2 showed no sign of ID, ASD, underweight or microcephaly. She was negative for history of other known phenotypes of dup16p11.2 including epilepsy, speech and motor delay, and CAs.

The father of proband B who is also a carrier of dup16p11.2 was unfortunately not available; nonetheless, we were able to determine the past medical history and examination of pictures revealing no sign of microcephaly, short stature or reduced body mass. By report, he is an educated man with no manifestations of learning disability, autism or psychiatric problems.

Inheritance pattern was reviewed for the reference dup16p11.2 cohort. The inheritance pattern for 13/18 patients (including my subjects) was available, in which 5/13 cases (38.4%) were *de novo* and 6/8 familial dup16p11.2 were maternally inherited (75%). Very little phenotypic information was available for four out of six carrier parents (excluding my subjects). One case reported by Shinawi *et al.* was an affected mother which I included in my reference cohort. A second carrier mother reported by Shinawi *et al.* was phenotypically normal. Fernadez *et al.* also provided limited phenotypic description for the two carrier-parents; one was an apparently healthy mother who also had a healthy dup16p11.2 carrier-daughter with no sign of ASD or learning difficulties, while the second individual was a dup16p11.2 carrier father diagnosed with bipolar disorder.

2.4 Discussion

In this chapter, I systematically reviewed and classified phenotypes of the two study subjects for comparison to similar correlates mined from 16 published dup16p11.2 cases. This was done first as a screening for any consistent, recognizable phenotypic pattern among the reference dup16p11.2 cases, and secondly to define the most relevant phenotype data for my probands to use in the final filtering steps of WES analysis of the study trios (chapter 4).

Of note, while the size of the dup16p11.2 in each of two study probands was different, the 16p11.2 CNV is a recurrent CNV with consistent breakpoints and gene content. In addition, use of two different microarray platforms in Proband B identified different sizes of the dup16p11.2 region, suggesting that the size discrepancy of dup16p11.2 is likely due to the difference in resolution and coverage of the different microarray platforms used within and between Proband A and B.

Proband A and B share some features similar to some cases of the reference cohort including ID, speech and motor delay, and being under-weight (at birth). Relative to each other, they are discordant for ASD, major psychiatric disorder, seizures, verbal apraxia, and facial gestalt (Table 2.7).

Table 2.7: Reported features of patients with 16p11.2 CNV.

Clinical features of Proband A and B in comparison to reported clinical findings of cases with 16p11.2 CNV.

Observed phenotypes in patients with16p11.2 CNV	Proband A	Proband B
DD/ID	+	+
ASD	+	-
Psychiatric disorders	-	+
Speech /language delay	+	+
Childhood apraxia of speech	+	-
Motor delay	+	+
Epilepsy	-	+
Macrocephaly/microcephaly	Microcephaly (postnatal)	Microcephaly (postnatal)
Over-weight/under-weight	Under-weight (at birth)	Under-weight (at birth)
Eating disorders	-	-
Hypotonia	+	+
Facial dysmorphism	+	+
Visual impairment	+	+
Hearing impairment	+	-
Brain anomalies	-	-
Eye anomalies	+	-
Heart anomalies	-	-
Renal anomalies	-	N/A
Vertebra anomalies	+	N/A
Hernia	-	-
Pyloric stenosis	-	-
Abnormal sexual development	+	+
Recurrent infection	+	+
Severe combined immunodeficiency	-	-

Abbreviation: N/A: not available.

More importantly, some of their features (OFC, height, weight and even facial features) changed with age, and were less consistent with features of dup 16p11.2. For instance, Proband A and B both were underweight (<3rd %ile) at birth, however, their weights later changed to 5-10th %ile (10y/o) and 20th %ile (14y/o), respectively. Microcephaly has been proposed as a well-recognized phenotype of dup16p11.2 patients, yet no information is available to clarify if it is congenital or acquired. In this study, Proband A showed postnatal microcephaly as his birth-OFC decreased from the 25th %ile (birth) to <2nd %ile at the age of 10. The birth-OFC of Proband B is unknown, however he had microcephaly at the age 14 years (<2th %ile). Both probands are positive for dysmorphic features; however, they don't share any specific facial features except low anterior hairline and thick alveolar ridge, which are not reported in any of the reference cohort. Both probands and four cases in the reference cohort are myopic. Among all studied cases, only Proband A has ptosis and retinal dystrophy and hearing impairment. Cryptorchidism was detected in both probands and one case introduced by Bedoyan *et al.* (Table 2.5, Table 2.6).

The brother of Proband B was later identified in the course of this study to be a carrier of dup16p11.2. Earlier reports showed that he has ADHD/ODD with some degree of learning deficit without being underweight or short statured. Of note, the neurocognitive phenotypes in family B segregated with the maternal side. In addition, the dup16p11.2 carrier mother of Proband A, and carrier father of Proband B do not manifest any recognized phenotype of dup16p11.2 including microcephaly, reduced BMI and stature or other abnormality.

In brief, this study involves five individuals with dup16p11.2, yet none with a characteristic phenotype, or sharing any major similarity to the data collected from 16 previously reported cases. This confirms significant variable expressivity among dup16p11.2 CNV carriers.

The fact is that lack of full phenotype description of patients in the literatures, use of vague and imprecise clinical terms, clinical bias towards the examination or report of particular features, make the phenotypic comparison of patients from different studies very challenging. In addition, repetitive reports of the same patients in different studies [159, 178] lead to over-estimation of the frequency of dup16p11.2 and its comorbidities. However, use of well-known phenotype ontology terms from the WBDD database helped me to overcome this obstacle.

Phenotypic analysis of the reference dup16p11.2 cohort revealed that only 44% and 11% of patients with dup16p11.2 have post-natal microcephaly and low BMI (the age between 18 months-18 y/o); respectively. Of note, one patient from the Shinawi *et al.* study has macrocephaly (97^{th} %ile). Neurologic disorders and congenital anomalies are common features in this cohort as well as 100% respective frequency of ID and speech impairment. No recognizable dysmorphic pattern was detected in dup16p11.2 carriers including carrier-parents discordant to their own offspring phenotype or more common phenotypes of dup16p11.2.

It is important to note that microarray testing is biased towards individuals with ID, learning disabilities, ASD and congenital anomalies [62, 144], and very few studies attempt to find the prevalence of a particular CNV in the general population [54, 187]. Therefore, high frequency of NDD among dup16p11.2 carriers may be due to ascertainment bias of affected individuals referred for medical genetics investigation. Further, clinical findings like speech impairment, psychiatric disorders, ASD and congenital anomalies are known comorbidities of individuals with ID. In fact, 66% of ID cases show one or more comorbid features [188]. Speech impairment is the most common childhood disability [189]. The prevalence of ASD and psychiatric disorders among ID cases are 5-30% and 32-40%, respectively [190, 191]. Hence,

it’s not surprising to find high frequency of such phenotypes in cohorts of ID patients who also carry dup16p11.2.

In this chapter, I have shown that significant clinical variability with no recognizable pattern of clinical features exists among carriers of dup16p11.2. The absence of known features of dup16p11.2 in a subset of patients with this CNV, as well as carrier-parents and -siblings suggest that this CNV by itself cannot explain the variable spectrum of phenotypes reported in carriers of this CNV. Further support for this hypothesis will benefit from expression study of genes within and nearby the 16p11.2 region in probands and their healthy carrier parents. The results of such expression analyses are discussed in the next chapter.

Chapter 3: Whole genome expression, and analyses of genes within 16p11.2 region

3.1 Introduction

Phenotypic effects of genetic variations like CNV appear to reflect altered expression levels, either by directly affecting the genes within CNVs, or indirectly through position effects of downstream or regulatory pathways [97, 98]. Therefore, the integration of gene expression studies and CNV data may help to prioritize CNV-candidate regions.

It is well known that the optimal tissue for expression study of patients with NDD is central nervous system (CNS) tissue, particularly brain, preferably studied during early development; with knowledge of which cell types are studied. However, brain biopsy is not a possible option. Further, post-mortem brain samples are not frequently available and would likely not represent a viable sample [192]. Hence, several studies have used LCLs to investigate the expression of genes within CNVs as they are often the only available tissues for high-throughput studies [94, 96]. Initial transcriptome-wide study of CNVs in LCLs showed that CNVs found in the general population affect the gene expression of ~20 % of genes within or nearby (1 Mb) the CNV region [91].

Expression studies of LCLs from patients with deletion and duplication of 16p11.2 showed a positive correlation of transcript levels and copy numbers of genes located in this region; however the effect of copy number on the number of differentially expressed (DE) genes (i.e., over- and under-expressed genes) within this region is controversial. Jacquemont *et al.* reported that the expression levels of all 27 studied-genes was positively correlated with dup16p11.2 in patients, but not in the flanking region [94], whereas a study conducted by Luo *et al.* displayed a significant positive correlation between expression level of 12/19 and 8/19 genes in deletion and duplication cases, respectively. Luo *et al.* also acknowledged that the expression

results of healthy carrier parents were similar to controls. Gene ontology (GO) enrichment analysis of 16p11.2 deletion cases performed by the Luo group demonstrated neural-related pathway enrichment, but no such enrichment was detected in duplications. The most consistent expression change in both deletions and duplications was reported for *KCTD13*, *ALDOA* and *MAZ* and the strongest correlation with head circumference was seen for *TAOK2* [96].

Recently, post-mortem brain samples from patients with 16p11.2 CNV were used for transcriptome studies. Transcriptome analysis of prefrontal cortex of post-mortem brain tissues of patients with psychiatric disorders illustrated the effect of 16p11.2 imbalance on the expression of several genes including *CORO1A*, *TAOK2*, *DOC2A* [192]. Blumenthal *et al.* [193] performed RNA sequencing using two different tissues, LCLs from ASD patients with 16p11.2 CNV (dels and dups), and cerebral cortex from mice with 16p11.2 CNV (dels and dups), and proposed that both deletion and duplication resulted in expression change of genes located within 16p11.2 region [193]. The most consistent results between two tissues were *cis* positional effect of the CNV on transcriptional level of genes outside of the CNV, especially in the distal region 1-5 Mb away. The DE genes from these regions were enriched in chromatin modification, synapticity, and known ASD/ID genes. However; the strongest correlation between the copy number and the expression of genes, in samples from both 16p11.2 CNV mice and ASD patients, was within the CNV region itself, especially for deletions [193].

Migliavacca *et al.* recently performed a transcriptome analysis on LCLs of carriers of 16p11.2 CNV and showed that genes which showed expression alterations were enriched for developmental pathways including ASD and ciliopathy genes. The authors identified a significant enrichment of DE genes (genome-wide) in 16p11.2 CNV carriers with genes implicated in two ciliary disorders: Bardet-Biedl and Joubert syndromes. They validated the

effect of ciliary gene dysregulation in 16p11.2 CNV transgenic mice models, and demonstrated a significant shortening of neuronal cilia in the CA1 region of the hippocampus in the 16p11.2 duplicated mouse model [194], associated with spatial memory [195]. Further, they employed zebrafish models and showed that over-expression of *BBS7* (4q27), associated with Bardet-Biedl syndrome, rescued macrocephaly in *kctd13*-morpholinos, and that up- or down-regulation of *CEP290* (12q21.32), associated with Joubert syndrome, rescued the brain phenotype of under- or over-expression of *kctd13*-zebrafish, respectively. In brief, the authors proposed that ciliary gene alterations may be responsible for the clinical variability of 16p11.2 CNV [194].

In this chapter, I first reviewed the information on expression of 16p11.2 genes in different tissues. Next, I explored the effect of dup16p11.2 on the RNA expression of 16p11.2 integral genes and genome-wide. To do this, I used WGE data to screen DE genes within and nearby the 16p11.2 region and across the whole genome. Then, I applied real-time quantitative PCR (qPCR) to validate the results of WGE in candidate genes within 16p11.2, and to compare their expression in each proband with their healthy carrier parent (mother of A) and non-carrier parent (mother of B). I also performed western blotting to evaluate the protein expression of one gene from dup16p11.2 (*KCTD13*). This study is the first to evaluate the protein expression of genes integral to 16p11.2 CNV. Finally, I performed enrichment analysis for genome wide-DE genes that were shared in both probands.

3.2 Materials and methods

3.2.1 Transformed lymphoblasts

LCLs were obtained from the two dup16p11.2 probands (Proband A: 12-32A, Proband B: 06-32), their mothers (Mother A: 12-56A, Mother B: 09-21), and two unaffected adult controls (Control 1: Nr-101M, Control 2: Nr-104F). Mononuclear cell fractions from sodium heparinized

whole blood were collected and processed using Ficoll-Paque Plus (GE Healthcare Bio-Sciences, Cat. No. 17-1440-02) and transformed using an Epstein-Barr virus (EBV) suspension to obtain LCLs. After transformation, cells were cultured and maintained at 37^{o}C under 5% carbon dioxide (CO_2) in a T25 flask using RPMI 1640 (Invitrogen, Cat. No.11875-093), supplemented with 15% Fetal Bovine Serum (FBS) (GIBCO, Cat. No. 12483-020) and 10% conditioned medium from an EBV producing cell line, and Cyclosporin A, until stable and exponentially growing cells were established. Transformed cells were stored at -80^{o}C for future study.

3.2.2 RNA and protein extraction

Frozen LCL samples were categorized in two groups. One group consisted of Proband A, Mother A and two controls (control 1 and 2) and the other included Proband B, Mother B and the same controls. LCLs of each group were simultaneously thawed and re-cultured at 37^{o}C under 5% CO_2 in upright T25 flasks using RPMI 1640 (Invitrogen, Cat. No.11875-093), supplemented with 10% FBS (GIBCO, Cat. No. 12483-020) and 1% Penicillin/Streptomycin (GIBCO, Cat. No. 15140). Next, LCLs were split into multiple flasks and allowed to grow. RNA and protein was extracted simultaneously from three to four different cultures, harvested ~4 days apart (different replicates).

RNAs were isolated from harvested cells using an RNeasy Plus Mini Kit (QIAGEN, Cat. No.74134), according to the manufacturer's protocol. The optional on-column DNase digestion was performed using the RNase-free DNase set (QIAGEN, Cat. No. 79254). Nanodrop spectrophotometry (ND-1000, software v.3.8.1) was used to obtain RNA concentration and purity (260/280 $\geq$ 2.0, 260/230 $\geq$ 1.80). RNA samples were diluted to the concentration of 300-400 ng/µl and then aliquoted and stored at -80^{o}C. To determine the quality and degradation, three µl of each RNA sample was sent to the CMMT core facility and analyzed using the

"Eukaryote Total RNA Nano Assay" on a Bioanalyzer 2100 (Agilent). RNA samples with RNA index numbers (RIN) of 10, and 28S/18S ribosomal RNA ratio of >1.9 were used for WGE arrays and qPCR experiments.

To isolate protein, the aforementioned cultured LCL samples were also centrifuged (same day of RNA extraction), washed twice with cold PBS and then lysed with cold RIPA Buffer (Thermo Scientific, Cat. No. 89900) containing Halt Protease Inhibitor Cocktail (Thermo Scientific, Cat. No. 87785), according to the manufacturer's instructions. Cold supernatants were aliquoted to small tubes and stored at -80 ^{o} C for future use. Prior to use, protein concentration was determined using the Bio-Rad™ *DC* Protein Assay kit (Biorad, Cat. No. 500-0116). To measure the protein concentration, the standard curve was run using dilution series of Bovin Serum Albumin (BSA), and then absorbance was read using the EnSpine 2300 Mulitlabel Reader (Perkin Elmer, Enspire Manager Software v1.00 Rev2).

3.2.3 Whole genome expression (WGE)

RNA from Proband A and B and two controls were used for WGE study. Samples were hybridized on Illumina Whole Genome Expression Array using HumanHT-12 v4 Expression BeadChip Kit, according to standard protocols. The BeadChip was scanned using the BeadArray Reader (Illumina), and GenomeStudio Software (Illumina) converted the image data to numerical data. This was done by the CMMT Genotyping Core Facility (CFRI). Next, subsequent analyses using "R" package (http://www.R-project.org/) was performed. The data normalization (quantile normalization) was performed for all the genes across all the samples. After normalization, Student's t-test was applied to the normalized gene expression levels.This was done by Dr. Paul Pavlidis' group (Department of Psychiatry, Centre for High-throughput Biology). Then, the gene content within 16p11.2 and flanking regions (genomic coordinates of

25.2-31.4 Mb) was determined to include 50 genes. The expression data were not available for two genes including *C16orf92* and *GDPD3*. The expression data of 44% of genes (21/48) were available only from one probe, corresponding to a well-known transcript, including genes of interest such as *KCTD13*. I included all 48 genes in my evaluation and calculated the mean expression of each gene for each proband and control separately, when more probes were available (Table 3.1). Fold change in comparison to control of <0.8 was considered as an indication that the gene is under-expressed, while fold change of >1.2 in comparison to controls was considered to represent over-expression.

Table 3.1: RNA expression of genes within and nearby 16p11.2 region using WGE data.

Change in expression is given as the mean transcript level (all probes) in the two probands and controls.

	Genes	Start- End (bp)	Strand	Probe	Proband A	Proband B	Control 1	Control 2
Telomeric	*AQP8*	25,227,052- 25,240,261	+	1	1.04	1.01	1.00	1.00
	ZKSCAN2	25,247,322- 25,269,252	-	1	0.92	1.01	1.00	1.00
	HS3ST4	25,703,347- 26,149,009	+	1	1.01	1.04	1.00	0.99
	IL4R	27,324,989- 27,376,099	+	2	1.08	1.02	1.01	0.99
	IL21R	27,413,483- 27,463,363	+	4	1.10	0.99	1.01	0.99
	KIAA0556	27,561,454- 27,791,692	+	1	1.01	1.08	1.02	0.98
	CLN3	28,477,983- 28,506,896	-	2	1.07	1.16	1.01	0.99
	EIF3C	28,699,879- 28,747,051	+	1	0.90	0.98	0.91	1.09
	ATXN2L	28,834,356- 28,848,558	+	5	1.04	1.06	1.03	0.97
	SH2B1	28,857,921- 28,885,534	+	1	1.02	1.12	1.04	0.96
	LAT	28,996,147- 29,002,104	+	4	0.92	0.97	0.98	1.02
	RRN3P2	29,086,163- 29,128,039	+	1	1.09	1.09	1.02	0.98
	RUNDC2C	29,262,829- 29,519,817	+	2	1.08	1.01	1.07	0.93
16p11.2	*BOLA2*	29,454,570- 29,466,285	-	6	1.14	1.16	0.96	1.03
	GIYD2	29,465,822- 29,469,540	+	2	1.02	1.02	0.95	1.04
	SULT1A4	29,466,401- 29,476,300	+	3	1.03	0.98	1.00	0.99
	SPN	29,674,300- 29,682,187	+	3	1.12	**1.25**	0.99	1.01
	QPRT	29,690,329- 29,710,020	+	1	**2.42**	**2.25**	1.15	0.85
	C16orf54	29,753,784- 29,757,340	-	1	1.05	1.05	0.98	1.02
	KIF22	29,802,040- 29,816,706	+	1	1.05	1.14	0.91	1.09
	MAZ	29,817,427- 29,823,649	+	3	1.03	1.02	1.00	1.00
	PRRT2	29,823,177- 29,827,202	+	1	1.09	1.02	1.00	1.00
	C16orf53	29,827,528- 29,833,816	+	2	0.92	0.99	0.90	1.10
	MVP	29,831,715- 29,859,355	+	2	**1.86**	**2.03**	1.09	0.91
	CDIPT	29,869,678- 29,875,057	-	2	1.11	1.09	1.05	0.95
	SEZ6L2	29,882,480- 29,910,868	-	4	1.04	1.06	0.97	1.03
	ASPHD1	29,911,696- 29,931,185	+	3	1.04	1.03	0.97	1.02
	KCTD13	29,916,333- 29,938,356	-	1	**1.34**	**1.76**	1.05	0.95
	TMEM219	29,952,206- 29,984,373	+	1	**1.52**	**1.84**	1.11	0.89

	Genes	Start- End (bp)	Strand	Probe	Proband A	Proband B	Control 1	Control 2
	TAOK2	29,984,962- 30,003,582	+	3	1.05	1.03	0.97	1.03
	HIRIP3	30,003,645- 30,007,757	-	1	**1.20**	0.91	1.05	0.95
	INO80E	30,006,615- 30,017,114	+	1	1.12	**1.27**	1.00	1.00
	DOC2A	30,016,830- 30,034,591	-	2	1.06	1.09	0.99	1.01
	C16orf92	30,034,655- 30,039,057	+	-	-	-	-	-
	FAM57B	30,035,748- 30,064,299	-	1	1.06	1.04	0.98	1.02
	ALDOA	30,064,411- 30,081,778	+	4	1.16	**1.24**	1.11	0.89
	PPP4C	30,087,299- 30,096,698	+	1	**1.56**	**1.56**	1.01	0.99
	TBX6	30,097,114- 30,103,208	-	3	0.99	0.94	1.00	1.00
	YPEL3	30,103,635- 30,108,236	-	2	**1.48**	1.17	1.06	0.93
	GDPD3	30,116,131- 30,125,177	-	-	-	-	-	-
	MAPK3	30,125,426- 30,134,827	-	2	**1.34**	**1.53**	0.97	1.03
	CORO1A	30,194,148- 30,200,397	+	1	0.95	**1.23**	1.00	1.00
	BOLA2B	30,204,255- 30,205,627	-	1	1.09	1.04	1.00	1.00
	GIYD1	30,205,208- 30,208,882	+	2	0.98	1.01	0.97	1.03
	SULT1A3	30,205,743- 30,215,650	+	6	1.01	1.02	1.02	0.98
Centromeric	*SEPT1*	30,389,454- 30,407,312	-	2	0.94	0.93	1.07	0.93
	ZNF688	30,580,667- 30,584,055	-	3	0.94	0.95	0.97	1.03
	BCL7C	30,844,947- 30,906,281	-	3	0.99	0.97	0.99	1.01
	POL3S	31,094,745- 31,100,949	-	1	0.96	1.02	1.04	0.96
	ITGAD	31,404,633- 31,437,826	+	1	1.00	1.02	1.01	0.99

To evaluate the significance of the putative functional association between genome-wide-DE genes in two dup16p11.2 probands with known cellular process or pathways, I performed pathway enrichment analysis. I calculated the mean expression of each specific gene available in the WGE data, and used a more stringent cut-off to define DE genes for this part of experiment (<0.7, >1.4 fold). From total of >31,000 annotated genes with >47000 probes, the sum of under-expressed genes (<0.7 fold) and over-expressed genes (>1.4 fold) shared in both probands was 127 and 171, respectively. Next, I analyzed the pathway and phenotype enrichment in over-expressed and under-expressed genes separately using "WEB-based Gene SeT Analysis toolkit" (WebGestalt). After applying Bonferroni multiple test adjustment, I selected "Pathway common" and "Phenotypic analysis" tools to find the top 10 enriched pathways as well as phenotypic enrichments shared in both probands for the under and over-expressed genes.

3.2.4 Quantitative real time PCR (qPCR)

qPCR was used to validate the results of microarray expression for three selected genes (*KCTD13*, *MAPK3* and *MVP*) within the 16p11.2 region. cDNA from three different replicates of RNA samples from two probands, two mothers (carrier and non-carrier) and two adult controls (i.e. three individuals with dup16p11.2 and three controls) were obtained by reverse-transcription using the EasyScript™ cDNA Synthesis Kit (Applied Biological Materials Inc., Cat. No. G234) with provided Oligo (dT), according to the manufacturer's protocol. cDNAs were diluted using IDTE (1x TE solution), and then aliquoted and stored at -20 ^{o} C for future use.

"TaqMan Gene Expression assay" (Applied Biosystems Inc.) overlapping exon-exon boundaries were used for three candidate genes; *KCTD13* (Hs00923251), *MAPK3* (Hs00946872) and *MVP* (Hs00245438); and a housekeeping gene (*Beta-actin* (Hs99999903)). Primer test runs were performed and melt curves were analyzed for each reaction to confirm the presence of a

specific product and to ensure that efficiencies between reactions were approximately equal (within 5%). Next, qPCR was performed on Applied Biosystems StepOne Plus™ Real-time PCR system using cDNA samples, specific TaqMan assay and TaqMan universal Master Mix II (Applied Biosystems Inc.), according to manufacture protocols. Each reaction was run on a 96 well plate in triplicate with three separate runs done per gene per sample with the endogenous control gene included in reach run. The level of RNA was quantified using ΔΔCt method [196]. The mean of RNA expression was obtained from three independent replicates for each gene per sample.

3.2.5 Western blotting

Western blotting was used to determine protein expression of KCTD13 in two probands, two mothers (carrier and non-carrier) and two controls. After determining the cell lysate protein concentrations using a DC™ Protein Assay as described above, samples were separated according to size by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The fractionated proteins were blotted to a sheet of nitrocellulose and then exposed to anti-POLDIP1 (anti-KCTD13) antibody (ab173202). To standardize the amount of protein loaded into each lane, blots were re-probed with a polyclonal antibody against HSP60 (ab6530). The Amersham ECL kit (GE Healthcare Life Sciences, Product No. RPN2232) was used to detect the amount of each antibody bound to antigen and resultant photographic films analyzed using UV densitometry (GeneSnap and Gene Tools Software). Next, normalized absorbance values were obtained for KCTD13 relative to the corresponding loading control (HSP60). The mean of the protein expression was obtained from three independent replicates for each sample.

3.3 Results

3.3.1 Tissue expression of 16p11.2 genes (GeneCards)

I searched the literature in order to determine the expression profile of 16p11.2 genes, with a particular interest in their expression in brain (the affected tissue) and white blood cells (the origin of studied tissue). Using GeneCards (http://www.genecards.org/), I searched for the tissue expression of 32 genes from 16p11.2 in whole blood (WB), brain, retina, spinal cord, and placenta (Table 3.2). Since no data was available for the expression of genes in LCLs in "GeneCards" (Microarray-BioGPS), I also included white blood cells (WBC) as the origin of LCLs in my assessment. The expression data for 3/32 genes (*C16orf54, PRRT2 and C16orf92*) was not available.

All genes within 16p11.2 exhibit a ubiquitous pattern of expression yet with different intensity. Very few genes, such as *CORO1A*, act in a very high tissue specific manner with the expression intensity (EI) of 175 in blood, and 3-10 in other tissues. Among 29/32 genes, the EI of nine genes located within 16p11.2 region in different tissues is ≥ 10 (*CDIPT, SEZ6L2, KCTD13, FAM57B, ALDOA, PPP4C, YPEL3, MAPK3* and *CORO1A*). Three genes including *ALDOA* (EI: 380), *CDIPT* (EI: 104) and *YPEL3* (EI: 63) showed the highest expression in normal human brain, with similar expression in both brain and retina. Surprisingly, there is greater similarity between their expression in brain and WB than brain and spinal cord. A similarity between their expressions in brain with placenta was noted.

The expression level of *KCTD13*, known as a major driver of micro and macrocephaly and brain abnormalities in zebrafish, was relatively low in brain (EI:10) compared to the three above genes (*ALDOA, CDIPT* and *YPEL3*), suggesting *KCTD13* is not the most expressed gene within 16p11.2 region in the human brain. The EI of *MAPK3*, the modifier of *KCTD13*, was 29 in brain

and retina. Interestingly, the expression of *MAPK3* in WB (EI: 23) is similar to brain, suggesting that WB would be good alternative to brain for expression study of this gene. *MVP,* another modifier of *KCTD13,* shows high expression in WB (EI: 13), but similarly low expression in other tissues including brain (EI: 5).

Table 3.2: RNA expression intensity of genes located within 16p11.2 region in normal human tissue (GeneCards- Microarray-BioGPS).

Gene	Whole Blood	White blood cells	Brain	Retina	Spinal cord	Placenta
BOLA2	9	-	5	7	5	4
GIYD2	3	-	2	2	2	3
SULT1A4	8	-	3	5	3	4
SPN	3	-	3	4	3	3
QPRT	3	-	2	4	3	11
C16orf54	-	-	-	-	-	-
KIF22	5	-	4	5	5	4
MAZ	10	-	8	10	9	9
PRRT2	-	-	-	-	-	-
C16orf53	4	-	3	5	4	4
MVP	13	-	5	7	6	7
CDIPT	89	-	104	105	52	75
SEZ6L2	3	3	16	7	3	2
ASPHD1	6	-	7	6	12	5
KCTD13	5	-	10	5	5	5
TMEM219	6	-	3	5	5	3
TAOK2	4	-	4	4	4	4
HIRIP3	6	-	7	6	7	6
INO80E	4	-	4	4	4	4
DOC2A	3	-	5	3	3	3
C16orf92	-	-	-	-	-	-
FAM57B	25	25	36	34	36	30
ALDOA	234	-	380	288	195	144
PPP4C	53	53	13	10	7	36
TBX6	2	2	2	2	2	2
YPEL3	176	176	63	50	22	37
GDPD3	3	3	2	3	3	4
MAPK3	23	23	29	29	27	19
CORO1A	175	175	10	3	4	3
BOLA2B	5	5	4	3	4	3
GIYD1	3	3	2	2	2	3
SULT1A3	8	8	3	5	3	4

3.3.2 Expression study of genes within and nearby 16p11.2 region

3.3.2.1 WGE analysis

I used WGE analysis as a screening method to find DE genes within and nearby the 16p11.2 region. I compared the mean RNA expression of 48/50 genes (genomic coordinates of 25.2 Mb to 31.4 MB) in two probands with two controls (Table 3.1). 30 genes were from 16p11.2 CNV and 18 genes were from the flanking regions. The expression of 12 genes located within 16p11.2 region were above 1.2 fold, of which six genes (*QPRT*, *MVP*, *KCTD13*, *TMEM219*, *PPP4C*, and *MAPK3)* were over-expressed (>1.3 fold) in both probands. The remaining six genes (*SPN*, *HIRIP3*, *INO80E*, *ALDOA*, *YPEL3*, and *CORO1A*) had a change in one of the probands, but not both. I did not find the effect of dup16p11.2 on the expression of genes located proximal (centromeric) or distal (telomeric) to the CNV interval (Figure 3.1).

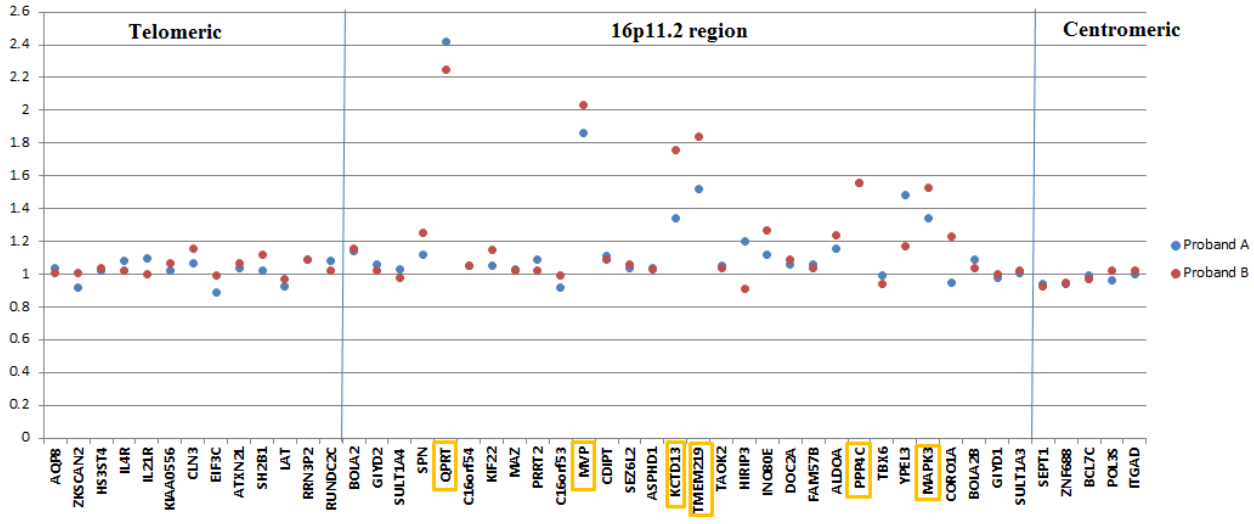

Figure 3.1: Scattered diagram of RNA expression of genes within and nearby 16p11.2.

The expression of *QPRT*, *MVP*, *KCTD13*, *TMEM219*, *PPP4C* and *MAPK3* is >1.3 fold in both probands.

3.3.2.2 RNA expression validation of *KCTD13, MAPK3, MVP*; and patients-parents comparison

The aim of this part of the project was firstly, to use qPCR to validate the WGE analysis results, and secondly to compare the 16p11.2 gene expression in my patients with their dup16p11.2 carrier and non-carrier parents. Among six over-expressed genes detected in both probands, I selected *KCTD13*, *MAPK3* and *MVP* for further evaluation, with rationale as follows: *KCTD13* has been reported as a major driver of micro- or macrocephaly in zebrafish due to respective duplication or deletion of the 16p11.2 region, and *MAPK3* and *MVP* were identified as modifiers of *KCTD13* [117]. Thus, I performed qPCR for these three genes using cDNA samples made from three different times-series of extracted RNAs, and then calculated the mean expression of each gene for each family member including two controls, separately.

In family A, the expression of three genes in Proband A did not differ from two controls and a non-carrier parent (Mother B). However, Mother of A (dup16p11.2 carrier) showed a slight increase in the expression of three genes (>1.27 fold).

In family B, results show that *KCTD13*, *MAPK3* and *MVP* genes are over-expressed (>1.51 fold) in Patient B compared to controls. The expression of three genes in his non-carrier mother was similar to controls (Figure 3.2). In brief, a discrepancy between WGE and qPCR results in Proband A was detected.

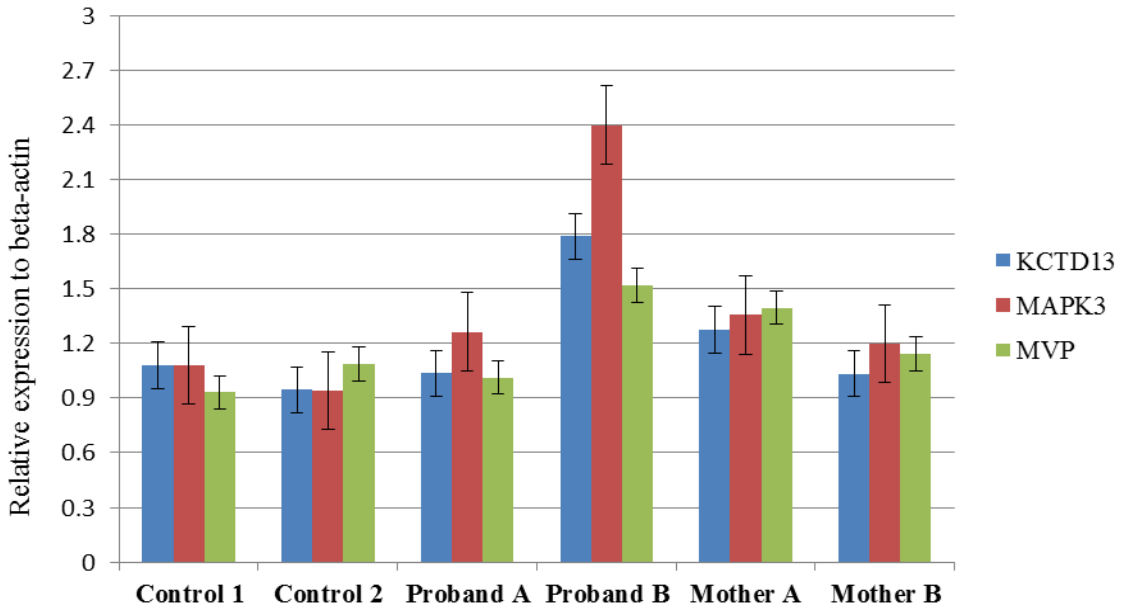

Figure 3.2: qPCR results of *KCTD13*, *MAPK3* and *MVP* in family A and B.

Diagram shows the mean expression of three genes in Proband A and B, Mother A (carrier parent), Mother B (non-carrier parent), and two controls. Error bars indicate standard errors from three replicates.

3.3.2.3 Protein expression of KCTD13

The level of RNA expression of *KCTD13* was not consistent in two probands, with the relative expression of 1.03 fold in proband A and 1.79 fold in proband B. To test if the protein expression of KCTD13 follows its RNA expression, I performed Western Blotting using cell lysate samples of two probands, their two mothers (one carrier and one non-carrier) and two controls. Surprisingly, the intensity of the band corresponding to KCTD13 was equally normal for all individuals even in patient B who previously showed significantly increased RNA expression (Figure 3.3).

The normal protein expression of KCTD13 in both probands, raises the possibility that *KCTD13* expression might be time-specific, up-regulated in the prenatal stage of brain development and then down-regulated postnatally. To answer this question, I used "Human Brain Transcriptome" (HBT) software (http://hbatlas.org/) to investigate the expression of *KCTD13* in a normal brain at different stages of life. Interestingly, the result showed that its expression level remains similar (EI: 8-10) from prenatal stages to adulthood (Figure 3.4), suggesting that the expression of *KCTD13* is not time-specific.

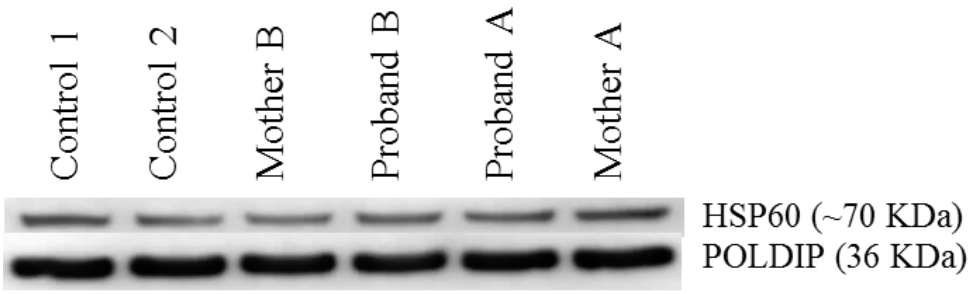

Figure 3.3: Western blotting of KCTD13.

Anti-POLIDIP1 antibody (corresponding to KCTD13) and anti-HSP60 antibody (loading control) were used in this experiment. Bands of approximate size of 36 KDa (KCTD13) and 70 KDa (HSP60) were detected.

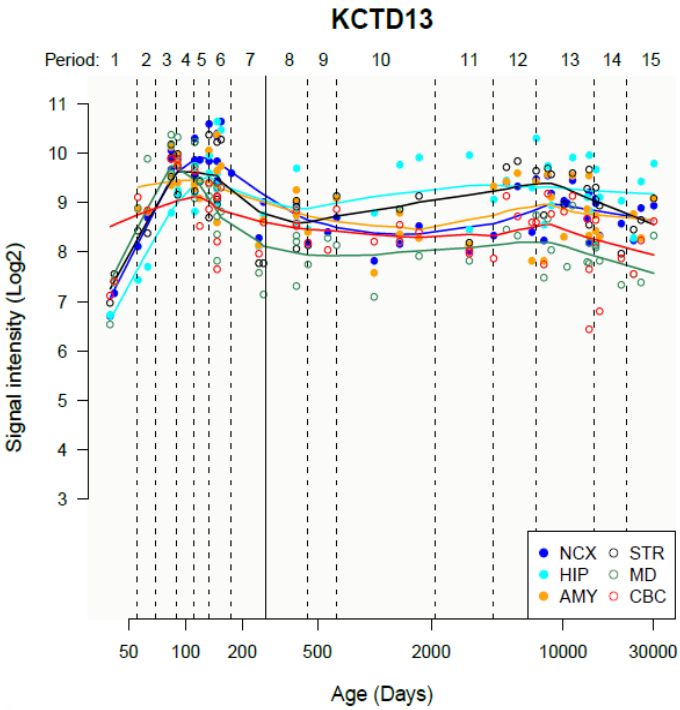

Figure 3.4: Brain expression intensity of *KCTD13* in different stage of human life.

The image is captured from "Human Brain Transcriptome" software. The Y-axis represents the signal intensity of *KCTD13* (log2) and the X-axis indicates different period of life, periods 1-7 (prenatal) and 7-15 (post-natal).

Abbreviation: NCX: neocortex, HIP: hippocampus, AMY: amygdala, STR: striatum, MD: mediodorsal nucleus of the thalamus, CBC: cerebellar cortex.

3.3.3 Pathway enrichment analysis (Pathway common and Phenotypic analysis)

Recently, Migliavacca et al. identified significant enrichment of developmental signaling (glucocorticoid receptor, TGF-b receptor, BMP and FGFR signaling) and ciliary pathways among the genome-wide DE genes in cases with 16p11.2 CNV [194].

Although my sample size was very small, I was interested to test if there is an enrichment in these or other pathways in my probands. The top enriched pathways using the "Pathway common" analysis for under-expressed genes in both probands (127 genes) included "DNA replication and cell cycle pathways". The same approach for over-expressed genes in both probands (171 genes) demonstrated that "signaling pathways" are significantly enriched in this group (Table 3.3), of which Class I PI3K [197], Arf6 [198], EGF receptor (ErbB1) [199] and mTOR [200] signaling pathways; and Glypican 1 network [201] are associated with brain function.

I also performed "Phenotype enrichment" analysis for the genome-wide up-regulated and down-regulated genes. None of the phenotypes were statistically significant. However, a positive correlation was detected between under-expressed genes and "infantile age of onset", short stature, abnormality of oral cavity, CNS and heart (Figure 3.5). The phenotype analysis of over-expressed genes displayed their involvement in abnormalities of multiple organs including the genitourinary system, eyes, musculoskeletal system, nervous system and vasculature (Figure 3.6).

Table 3.3: Top 10 enriched pathways among under-expressed (127) and over-expressed (171) genes detected genome wide, and shared in proband A and B.

Adjusted *p*-value (Bonferroni multiple test adjustment) of <0.05 is considered as significant.

Under-expressed genes		Over-expressed genes	
Pathway commons	**Adjusted p-value**	**Pathway commons**	**Adjusted p-value**
Cell cycle, Mitotic	3.44E-05	Glypican 1 network	8.94E-12
Mitotic M-M/G1 phases	1.10E-03	LKB1 signaling events	1.05E-11
DNA replication	1.70E-03	Glypican pathway	1.80E-11
E2F mediated regulation of DNA replication	2.60E-03	IL12-mediated signaling events	4.34E-11
Glycine cleavage complex	3.50E-03	Class I PI3K signaling events	5.48E-11
Aura B signaling	7.50E-03	Arf6 signaling events	5.48E-11
Mitotic G1-G1/S phases	1.58E-02	Insulin pathway	5.48E-11
M phase	3.16E-02	Arf6 trafficking events	5.48E-11
Pyrimidine deoxyribonucleotides *de novo* biosynthesis	3.16E-02	EGF receptor (ErbB1) signaling pathway	5.48E-11
E2F transcription factor network	3.95E-02	mTOR signaling pathway	5.48E-11

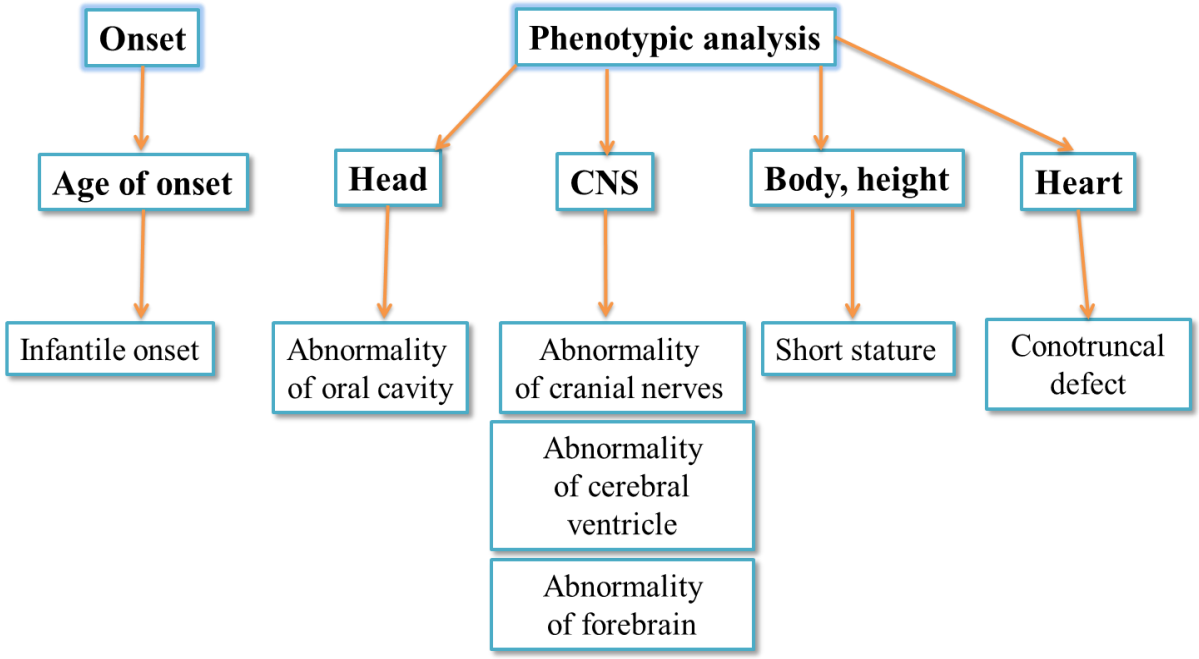

Figure 3.5: Phenotypic analysis of under-expressed genes (genome-wide) shared in proband A and B.

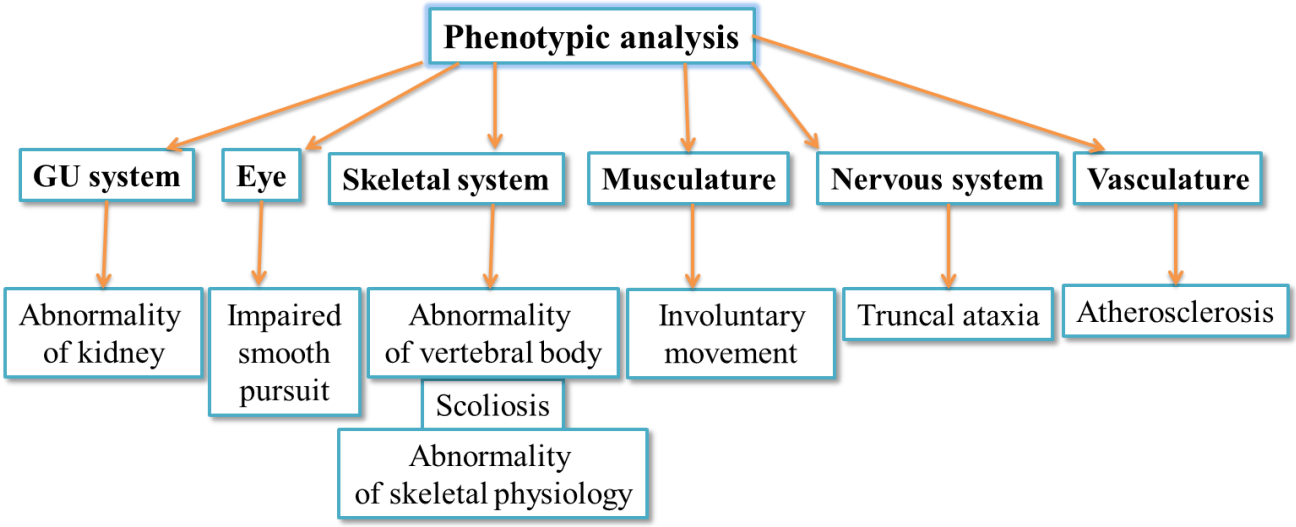

Figure 3.6: Phenotypic analysis of over-expressed genes (genome-wide) shared in proband A and B.

3.4 Discussion

Transcriptomic and proteomic study of genes in complex diseases requires accessible and sufficient amount of samples. Brain is considered to be an organ affected in patients with 16p11.2 CNV, but is an inaccessible tissue for expression studies in living patients. To gain insight into the expression of genes from the 16p11.2 CNV in brain and other tissues, I first used the data from cell lines of different tissues and determined that all the genes located within 16p11.2 were ubiquitously expressed. The gene with highest expression in CNS is *ALDOA* with relatively similar expression in WB. Among my three candidate genes (*KCTD13, MAPK3, MVP*), the brain expression of *MAPK3* (EI: 29) was higher compared to both *KCTD13* (EI: 10) and *MVP* (EI: 5).

The analysis of genes with available expression data for both WB and WBC demonstrated similar level of gene expression, suggesting that the expression level of genes in WB is a good predictor for gene expression in WBC, as most nucleated cells in WB are white blood cells. For one of the genes *MAPK3,* the EI was similar in brain (EI: 29) and WB/WBC (EI: 23/23). For the remaining two genes, *KCTD13* and *MVP*, the brain/WB expression intensity was 10/5 and 5/13; respectively. No expression data in WBC was available for these two genes.

LCLs have been typically used in expression studies of human disease as they provide accessible and unlimited samples, in contrast to WB which would require repeated sampling from children and families with NDD. Previous studies that used LCLs to investigate the expression of genes within 16p11.2 showed a positive correlation between copy number and expression of genes located in this region [94, 96]. Further, Hu *et al.* used LCLs from twins with ASD and showed that there is a quantitative correlation between the expression level of autism-candidate genes and severity of symptoms [202], suggesting that the expression study of some

genes in LCL may be a good predictor of their dysregulation in brain and phenotypic consequences. However, it must kept in mind that some of disease-causing genes may not be expressed in LCLs [203].

Using LCLs, I performed WGE analysis of 16p11.2 genes in my two probands. The expression level was in keeping with copy number for some of the genes located in this region. Among 48 studied genes within and nearby the 16p11.2 region, six genes (*QPRT*, *MVP*, *KCTD13*, *TMEM219*, *PPP4C*, and *MAPK3)* were over-expressed in both probands (>1.3 fold). Although some studies have found that all genes within the 16p11.2 CNV show altered expression [94, 193], there were others that identified only a subset of genes within the 16p11.2 region demonstrates a positive correlation between copy number changes and gene expression [95, 96]. Despite evidence suggesting CNVs affect the RNA expression of nearby genes [91, 93], I observed no such differences for the genes lying proximal (centromeric) or distal (telomeric) to the rearrangement. The controversy in the number of DE genes within the 16p11.2 CNV region may be due to the fact that techniques used for the validation of the expression results are not similar (qPCR, MLPA, another microarray platform or validation in animal models) [94, 96, 193], and that all genes are not validated in all samples [94, 96].

Luo *et al.* reported that the genes that had the strongest positive correlation of expression change to 16p11.2 gene dosage included *MAPK3, YPEL3, CORO1A and KCTD13* [96]. Golzio *et al.* showed that *KCTD13* acts as a major driver of micro- or macro-cephaly in zebrafish, and *MAPK3* and *MVP* serve as modifiers of *KCTD13*. Co-injection of either of these two modifier RNAs amplified *KCTD13* effect [117]. My WGE analysis showed *KCTD13, MAPK3* and *MVP* over-expression in both probands, however, the qPCR confirmed their over-expression only in proband B, but not in proband A. Interestingly, the expression of these three genes was slightly

increased in mother A who is an unaffected carrier of dup16p11.2, suggesting that there is no correlation between the expressions of these genes with the phenotypes of dup16p11.2 carriers.

The reasons for discrepancy between the two methods in Proband A may be due to unknown technical variation in the WGE, although replicates were not run to assess the extent of this. Normalization is a necessary step to make sure data from each sample run on array are comparable, and its purpose is to reduce artificial expression changes due to technical reasons. However, different types of normalization techniques can introduce bias in the data [204], and thus this may affect the WGE results. Overall, the results of microarray are only partially consistent with qPCR [205]. It is well-known that qPCR is more sensitive and accurate; and able to detect smaller differences in gene expression than microarrays. The advantage of qPCR is that measurement of the reaction product is quantitative and can be known at the end of the cycling process in comparison to expression of one endogenic control gene, while the expression of many genes on the array requires normalization across the genome. On the other hand, qPCR requires normalization to a housekeeping gene, which can itself vary between individuals. The advantage of WGE is that it does not require an internal control as the normalization is performed over all gene expression. This eliminates any potential biases due to the housekeeping gene chosen.

To date, all published expression data for 16p11.2 CNVs exclusively refer to RNA levels, with no protein expression data reported. Considering that the protein change is key to the disease-phenotype, it is conceivable to expect that copy number change of *KCTD13*, the proposed driver of the micro and macrocephaly, would affect its RNA and protein expression in affected carriers. However, the data presented here unexpectedly revealed that duplication of this region did not result in detectable increased protein expression of KCTD13 in any dup16p11.2

carriers, despite increased RNA expression in proband B, suggesting that *KCTD13* is not the sole cause of microcephaly in these two cases with dup16p11.2. However, it is important to note that the correlation between the mRNA and protein is weak [206-208]. The cellular turnover of proteins is the most important factor influencing RNA-protein correlation, and is highly variable ranging from seconds to several days. Post-transcriptional and translational modifications also may explain this poor correlation [209]. mRNA folding (secondary and tertiary) is sensitive to temperature, and temperature-dependent structural changes of mRNA may affect translational activity [208]. Additionally, quantification of proteins with extreme pH, very large or small size, and low abundance is challenging in Western blotting [210, 211].

It is also conceivable that the consequences of dup16p11.2 on KCTD13 protein expression alteration occurred during prenatal stages, and is not noticeable in childhood. However, available data from the HBT software suggests that the expression level of *KCTD13* in normal human brain remains almost always similar (EI: 8-10), and does not diminish post-natally. Nevertheless, it is worthy to note that my study is in LCLs and not in brain, and that the HBT database shows RNA and not protein expression of genes in brain during different stage of life. While the results of protein assessment for MAPK3 and MVP may differ from KCTD13, I did not pursue follow-up for these two genes given they are known as modifiers of the dup16p11.2 phenotype and not the drivers [117], and also their RNA expression changes are inconsistent in the studied dup16p11.2 carriers.

On a global scale, the DE genes in LCL from 81 carriers of 16p11.2 CNV (50 deletions and 31 duplications) showed significant enrichment in development signaling and ciliary pathways in samples [194]. Of note, the pathway enrichment analysis of over-expressed genes in Probands A/B also showed enrichment of “signaling pathways”, mainly associated in brain, but

no evidence of enrichment in ciliary pathways was found. The same approach for under-expressed genes showed “DNA replication and cell cycle” pathways enrichment. The result of the phenotype enrichment analysis of DE genes was not statistically significant, although several observations are worth noting. For instance, impaired smooth pursuit and involuntary movement were enriched among over-expressed genes. Smooth pursuit movement requires a precise coordination between many brain structures allowing eyes to follow the moving object. The smooth pursuit impairment is detected in a variety of disorders including schizophrenia and autism [212, 213]. Proband A was diagnosed with autism and Proband B suffers from psychiatric disorders (ADHD, ODD) and involuntary movement (tremor). Under-expressed genes were enriched for the infantile onset and abnormality of oral cavity seen in both probands (high arch palate and thick alveolar ridge).

In summary, the data presented in this chapter showed that expression changes of *KCTD13* neither segregate with dup16p11.2 nor with the microcephaly observed in my probands, and thus *KCTD13* is not likely to be the sole cause of microcephaly in these dup16p11.2 cases. This suggests that Proband A and B may harbour other disease-causing variants and highlights the need for further genomic analysis including WES. The results of such investigation are illustrated in the next chapter.

Chapter 4: Whole exome sequencing and variants in family A and B

4.1 Introduction

WES shows promise as an effective approach for identification of mutations in known and new candidate genes for NDD, uncovering a large number of single nucleotide variants (SNVs). Selection of those SNVs that are most likely pathogenic, as briefly described below, should be carefully performed using a step-wise approach and correlation of variants to detailed patient phenotypes to facilitate analyses [214]. Typically, the study patient is severely affected and it is highly unlikely that the causative variant will be synonymous or have a high frequency in the general population. Thus, excluding synonymous and common variants using public database such as 1000 Genome Project (http://www.1000genomes.org/), and National Heart and Lung and Blood Institute (NHLBI) Exome Sequencing Project (http://evs.gs.washington.edu/EVS/), or in-house database helps to filter out the variants of less likely effect. Furthermore, damaging variants are more likely to disrupt the expression, affinity and function of their corresponding gene. Multiple bioinformatics tools are designed to predict the severity of damage caused by non-synonymous variants such as SIFT, PolyPhen2, Mutation Taster. Finally, variants located within the conserved sites of the genome are expected to be deleterious, and hence applying tools such as GERP and PhyloP provide more insight to the conservation levels of the location of each variant [215].

WES has been largely used to identify SNVs in subjects with unknown cause of NDD, or to explore the causes of phenotypic variability detected in cases with the same genetic change. This includes, in some instances, cases with NDD that have a known pathogenic CNV but atypical phenotype or cases that share the same rare CNV with their unaffected parents. For example, McDonald-McGinn *et al.* used whole and targeted exome sequencing to explore the

cause of uncommon clinical features found in <10% of affected individuals with 22q11.2 microdeletion syndrome. Authors discovered that the presence of *SNAP29* mutations on the non-deleted chromosome unmasked an autosomal recessive CEDNIK (cerebral dysgenesis–neuropathy–ichthyosis–keratoderma) syndrome [125], thus leading to variable presentation of DiGeorge/velocardiofacial syndrome. Similarly, Paciorkowski *et al*. identified mutations in *NDE1* on the non-deleted chromosome 16 homolog in persons with inherited deletion of 16p13.11 and atypical phenotypes [124]. Of particular interest, Classen *et al*. evaluated three children with syndromic ID harboring putatively pathogenic CNVs; two inherited from normal parents, and discovered that rare inherited variations in other regions of the genome contributed to phenotypes atypical to those expected from the CNV alone in these CNV carriers [126]. These data suggest that co-occurrence of secondary genomic events in the other homolog or elsewhere in the genome may be responsible for the variable expressivity and incomplete penetrance in a subset of NDD patients harboring a CNV.

The purpose of this chapter is to describe the efforts to find an appropriate explanation for variable expressivity of individuals with dup16p11.2 using exome sequencing. Although zebrafish study suggests that dysregulation of *KCTD13* expression influences macro- or micro-cephaly, the results of investigations detailed in chapter 3 showed that *KCTD13* had similar protein expression in the probands who are microcephalic, their unaffected carrier parent, and controls. I have therefore expanded my search to look for the presence of damaging sequence variants that could explain the variable expressivity of dup16p11.2 subjects in this study. In my interpretation of SNVs, in addition to the tools described above, I used the RNA and protein expression data of the affected genes to determine more closely the effect of the variant on the gene function and to help interpreting their causality.

4.2 Materials and methods

4.2.1 DNA extraction

Genomic DNAs from Proband A and B and their family members were extracted from peripheral blood collected in EDTA tubes using the QIAamp DNA blood Mini Kit (QIAGEN). NanoDrop spectrophotometry (ND-1000, software v.3.8.1) was used to measure the concentration and purity of DNA samples, and the quality of samples was confirmed by agarose gel electrophoresis.

4.2.2 WES

DNA samples of both family trios were sent to PerkinElmer Company for exome enrichment using the TruSeq Exome Enrichment Kit (Agilent v5+UTR), followed by paired-end sequencing (Illumina HiSeq 2000, read length of 100 bp). Reads were mapped using BWA against the genomic reference sequence for Homo sapiens (Build 37). To generate a final genomic alignment, each BAM file underwent an additional post-alignment process using the Picard algorithm. Nucleotide variants were called using the GATK variant caller. These steps were done by PerkinElmer. WES data quality in two family trios is summarized in Table 4.1.

Table 4.1: Summary of sequencing and alignment data of family A and B.

	Family A			Family B		
	Proband	Mother	Father	Proband	Mother	Father
Total data yield (GB)	7.09	7.83	7.68	7.68	8.34	7.71
Number of reads (Million)	70.93	78.34	76.8	76.83	83.38	77.14
% of reads with quality ≥ 30	88.8	88.97	88.8	88.76	88.77	88.93
Number of reads mapped (Million)	69.87	77.21	75.75	75.54	81.73	75.78
Number of reads properly paired (Million)	67.21	74.22	73.39	71.76	75.74	71.06
Mean coverage	53	57	56	55	60	55
% of bases covered > 1x	99.1	99.1	99.2	99.2	99.1	99.2
% of bases covered > 10x	95	96	95.8	95.6	95.8	95.4
% of bases covered > 20x	84.2	87.2	86.7	85.8	87.4	85.4
% of bases covered > 30x	68.7	73.4	72.8	71.3	74.4	70.6

Using Golden Helix (GH) software (SNP & Variation Suite 7.7.8), the WES data from a single VCF file for all sequenced family members was analyzed. In order to reduce the number of false positive calls, genotype quality metrics were applied for all SNVs and Indels. Regions with at least tenfold depth of coverage, genotype quality of ≥40 and alternate allele ratio of ≥25% were used for calling variants. Available public databases including 1000 Genome Project 2012, and NHLBI ESP6500 were used to remove common variants with a minor allele frequency (MAF) of 1% for autosomal dominant (AD) and X-linked (XL) inheritance, and MAF of 5% for autosomal recessive (AR) categories. The inheritance pattern was determined and all modes were considered (AR including homozygous and compound heterozygous, *de novo*, and XL). Then, non-exonic (except variants in canonical splice sites) and synonymous variants were excluded. To predict the impact of SNVs on protein function and to determine how conserved the positions of variants are in the genome, several computational tools including SIFT, PolyPhen2, Mutation Taster, MutationAssessors, FATHMM, GERP and Phylop were used. In each inheritance category, candidate variants were selected if assigned as "damaging" or "disease causing" by two of five functional prediction tools, and if their locations in the genome were conserved (GERP >3 and PhyloP >1). Variants in the compound heterozygous inheritance category were excluded if one of two variants was not meeting the filtering criteria or if both variants were from one parent. Next, each candidate variant was further evaluated using data on gene function, literature (PubMed), and clinical findings in the probands. The filtering algorithm is shown in Figure 4.1. Finally, candidate SNVs were checked in the "GH Genome Browser" to confirm the read depth coverage and quality of sequencing data before confirmation by Sanger sequencing (Figure 4.2).

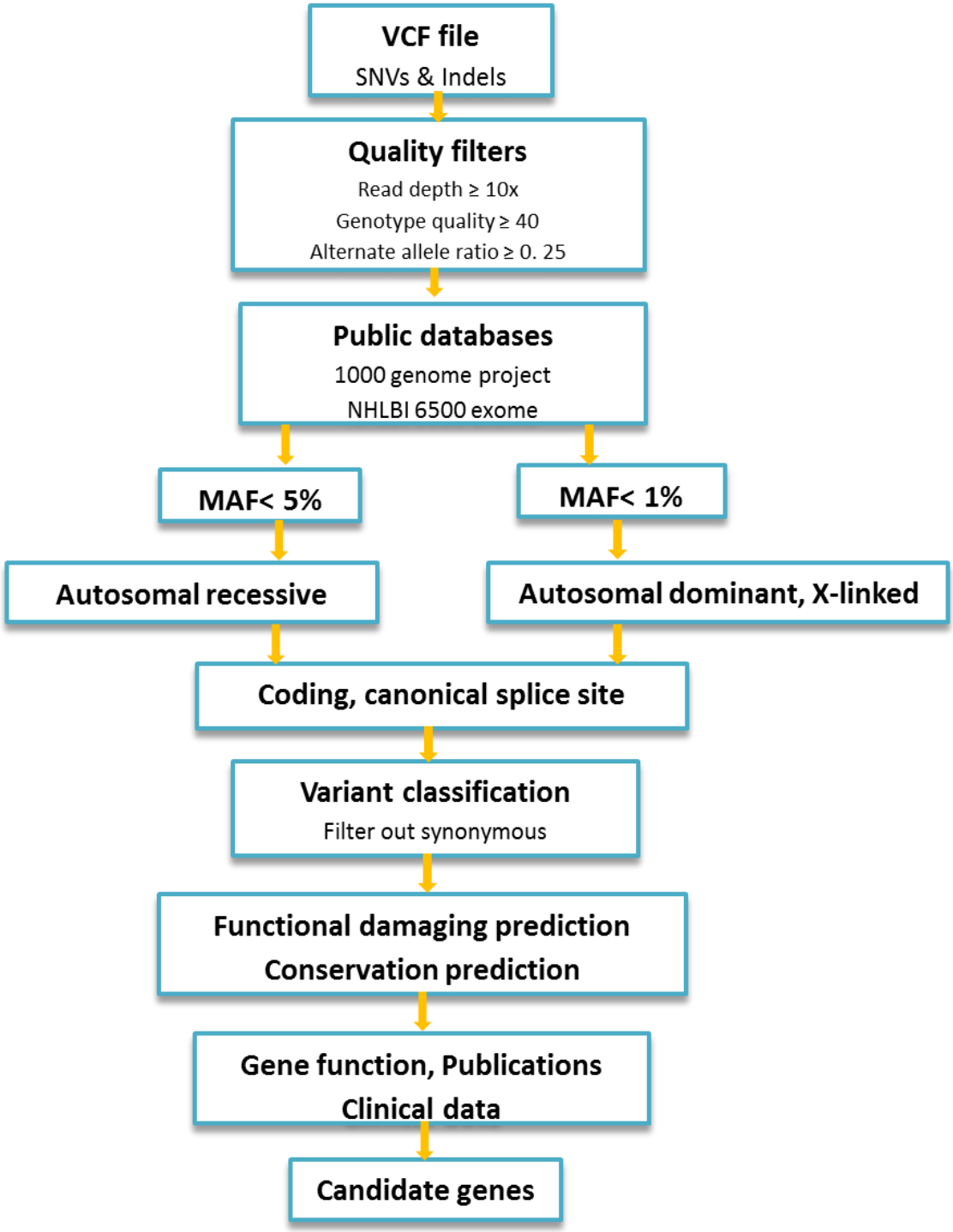

Figure 4.1: Filtering strategies used for analysis of WES data in family A and B.

Abbreviation: SNV: single nucleotide variation, NHLBI: National Heart, Lung and Blood Institute, MAF: minor allele frequency.

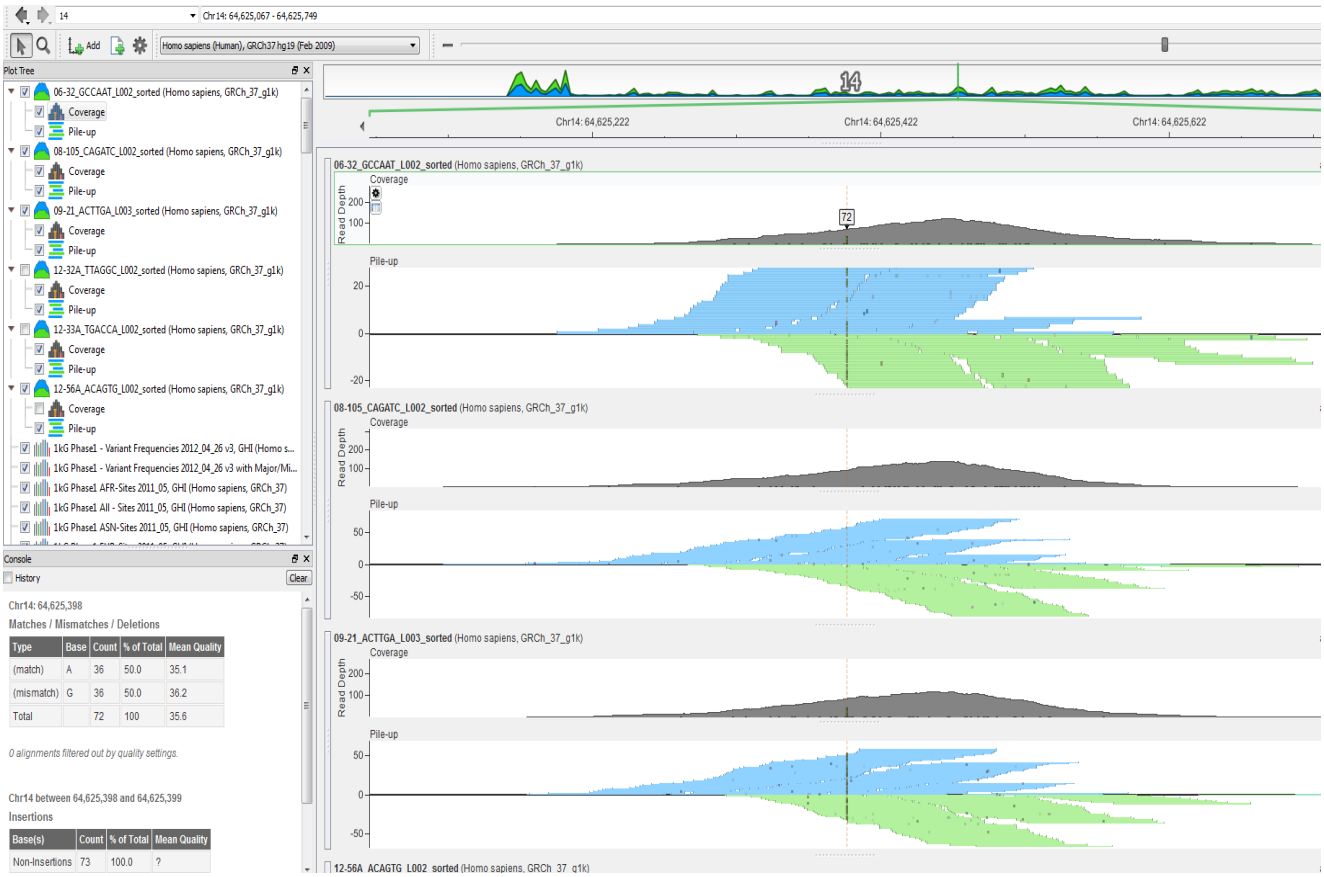

Figure 4.2: Screenshot of WES data visualized in the "Golden Helix Genome Browser".

A variant of *SYNE2* gene are shown in a heterozygous state in the Proband B (06-32) and his mother (09-21). The figure displays both the forward and reverse strands.

4.2.3 Sanger sequencing

Primers were designed for each selected variant using the "Primer3 software". After PCR optimization in controls, PCR on genomic DNAs of each family member followed by PCR clean-up were performed. Next, direct PCR product sequencing was conducted using Bigdye terminator V3.1 cycle sequencing kits, and subsequently analyzed on ABI 3130XL genetic analyzer.

4.2.4 qPCR

Three different replicates of RNAs were used to prepare cDNAs from LCLs of each Proband A/B, their two mothers and two controls according to manufactured protocols explained in chapter 3. TaqMan Gene Expression Assay for *VPS13B* (Hs00215450) and *SYNE2* (Hs00794881) genes provided by Life Technology Company (http://www.lifetechnologies.com) were used for qPCR. Quantification of expression level of each gene was performed in comparison to *Beta-actin* (Hs99999903), according to the protocol discussed in chapter 3. The mean expression of each gene was determined in the specific proband and their mother and compared to controls.

4.2.5 Western blotting

Western blotting was performed using specific antibodies for each experiment including target antibodies, anti-VPS13B (ab139814) and anti-Nesprin 2 (ab57397), and loading control antibodies, anti-Beta-actin (A2066) and anti-HSP60 (ab6530), according to the protocol discussed in chapter 3.

4.3 Results

4.3.1 WES and variant follow-up in Family A

Table 4.2 shows the variants identified in Proband A. There were eight SNVs in AR (homozygous), 80 SNVs in AR (compound heterozygous), and four SNVs in XL categories. No *de novo* variant was found. Variants in five genes were conserved and damaging ("pre-candidate genes), but variants of four genes were with no known function or role in the brain (*COL6A3*, *SYAP1*, *MAP7D2* and *RIBC1*). Only, two novel compound heterozygous mutations in *VPS13B* (8q22.2) fulfilled all filtering criteria (Table 4.3), and were subsequently confirmed by Sanger sequencing in Proband A (Figure 4.3). A sequence variant of c.1426-1G>A located in the conserved acceptor splice site of intron 10 was identified in Proband A and his mother. The second variant is a nucleotide change of G>T at c.4157+1 situated in the conserved donor site of intron 27 and was inherited from his father. Mutations and/or copy number variations in the *VPS13B* (*COH1*) gene lead to a rare autosomal recessive condition called Cohen syndrome ((CS); MIM216550) [216].

Table 4.2: Number of filtered variants in each inheritance category in Proband A.

Inheritance (MAF)	Coding, splicing	Classification	Damaging	Conserved	Pre-candidate gene	Known gene function	Relevant publication	Match to clinical data
De novo (< 0.01)	0	-	-	-	0	-	-	-
Autosomal recessive Homozygous (< 0.05)	8	6 (non-syn)	0	-	0	-	-	-
		2 (frameshift)	-	-	0	-	-	-
Autosomal recessive Compound heterozygous (< 0.05)	80	72 (non-syn)	26	16*	*COL6A3*(2)	Yes	No	-
		3 (splicing)	-	2	***VPS13B* (2)**	**Yes**	**Yes**	**Yes**
		1 (frameshift)	-	-	0	-	-	-
		3 (stop gain)	0	2*	0	-	-	-
		1(deletion)	-	-	0	-	-	-
X-linked (< 0.01)	4	4 (non-syn)	3	3	*SYAP1*	No	No	-
					MAP7D2	No	No	-
					RIBC1	No	No	-

Abbreviation: MAF: minor allele frequency, non-syn: non-synonymous; *: variants were excluded from compound heterozygous inheritance because only one variant passed previous filters, or both variants (from one gene) were inherited from the same parent.

Table 4.3: *VPS13B* variants in family A.

Gene	Chromosome-position	Classification	Exon/intron	Variation	Amino acid change	Damaging score	GERP	PhyloP
***VPS13B*-1**	8:100147823	Splicing	11	c.1426-1G>A	-	-	5.21	2.58
***VPS13B*-2**	8:100515179	Splicing	27	c.4157+1G>T	-	-	5.41	2.52

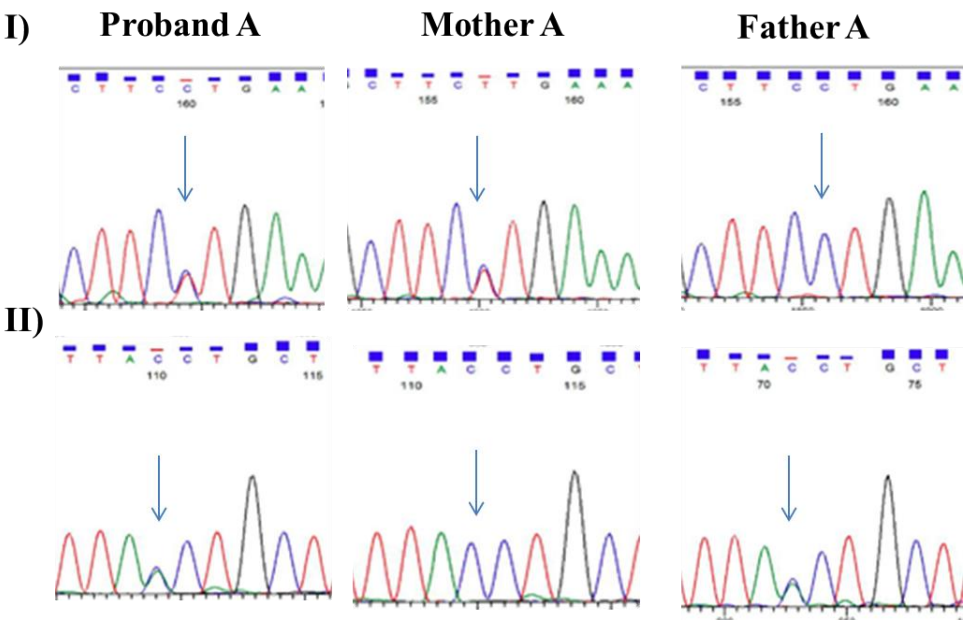

Figure 4.3: Sanger sequencing analysis of *VPS13B* variants in family A.

I) Proband A and his mother are carriers of splicing mutation of c.1426-1G>A. II) Proband A and his father are carriers of splicing mutation of c.4157+1G>T. (Sequences of reverse strands are shown).

4.3.1.1 In-silico functional prediction of *VPS13B* splicing mutations

Functional prediction tools used in GH software anticipate the effect of non-synonymous variants (coding region). However, both variants of *VPS13B* are located at canonical splice sites. Therefore, I used "Alamut software" comprising five splicing prediction algorithms including SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer and Human Splicer Finder for analyzing the splicing effect of *VPS13B* variants. Results indicated that two intronic variants of *VPS13B* would cause skipping of exons 11 and 27 (Figure 4.4).

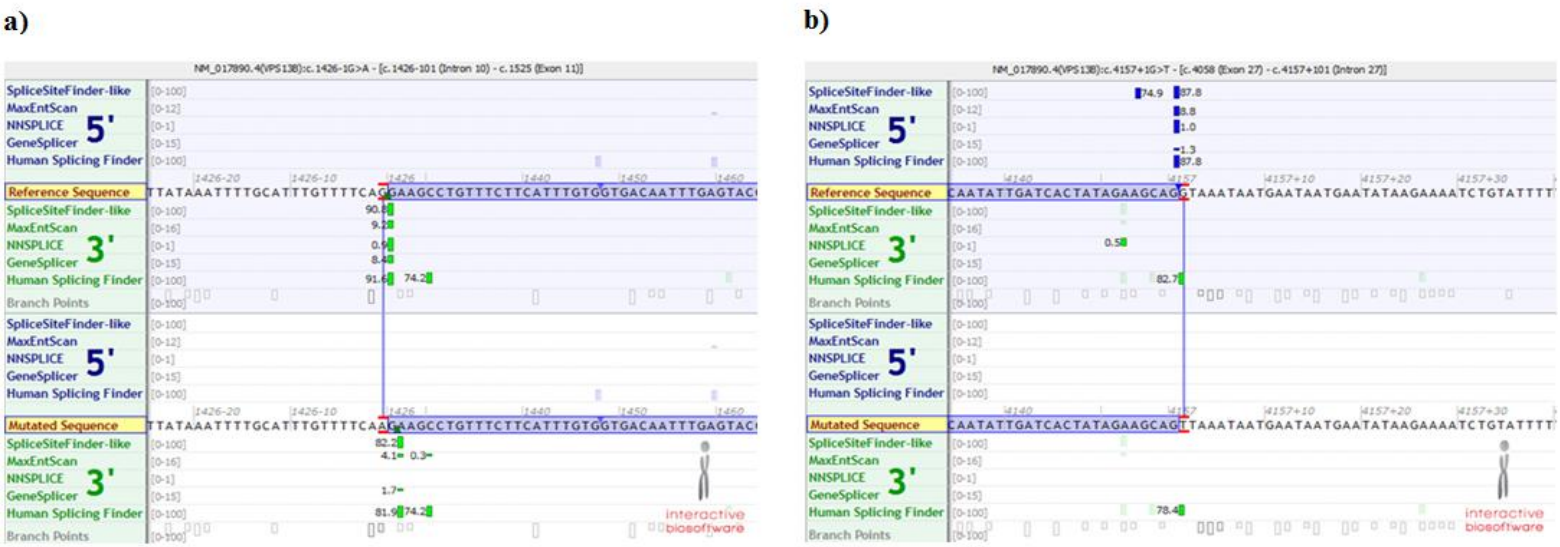

Figure 4.4: The effect of *VPS13B* splicing variants shown using Alamut software.

Numbers next to green/blue boxes at 3' UTR and 5' UTR represent the affinity score for exon-exon conjunction given to the specific nucleotide by five different tools before and after mutation. a) *VPS13B* variant is located at acceptor site (3' UTR) of intron 10 has completely lost its candidacy for exon/exon conjunction, but with presence of a close candidate nucleotide afterward; b) *VPS13B* variant is located at donor site (5' UTR) of intron 27 has completely lost its candidacy for exon/exon conjunction with no candidate nucleotide afterward.

4.3.1.2 In-vitro validation of splicing effects of *VPS13B* variants

To confirm the altered splicing as predicted by Alamut software, I designed two sets of primers (Table 4.4), covering exons 9-12 and 26-29, and performed PCR on cDNA samples of Proband A and one normal control, followed by Sanger sequencing of PCR products. The sequencing confirmed that both variants abolish the canonical splice sites, disrupted the following sequences and create a frameshift (Figure 4.5), which could lead to nonsense mediated mRNA decay.

Table 4.4: Primers sets used for in-vitro validation of splicing effect of *VPS13B* variants.

Variant	Exons	Primers
c.1426-1G>A	9-12	F:AAGGAACTACAGTTGAGGCCC
		R:CCCAAACCGTTGCATTCCAG
c.1426-1G>A	26-29	F:TTGGAGGAACCAGTGACGTG
		R:TGTCCTGACTGCCAACCTTC

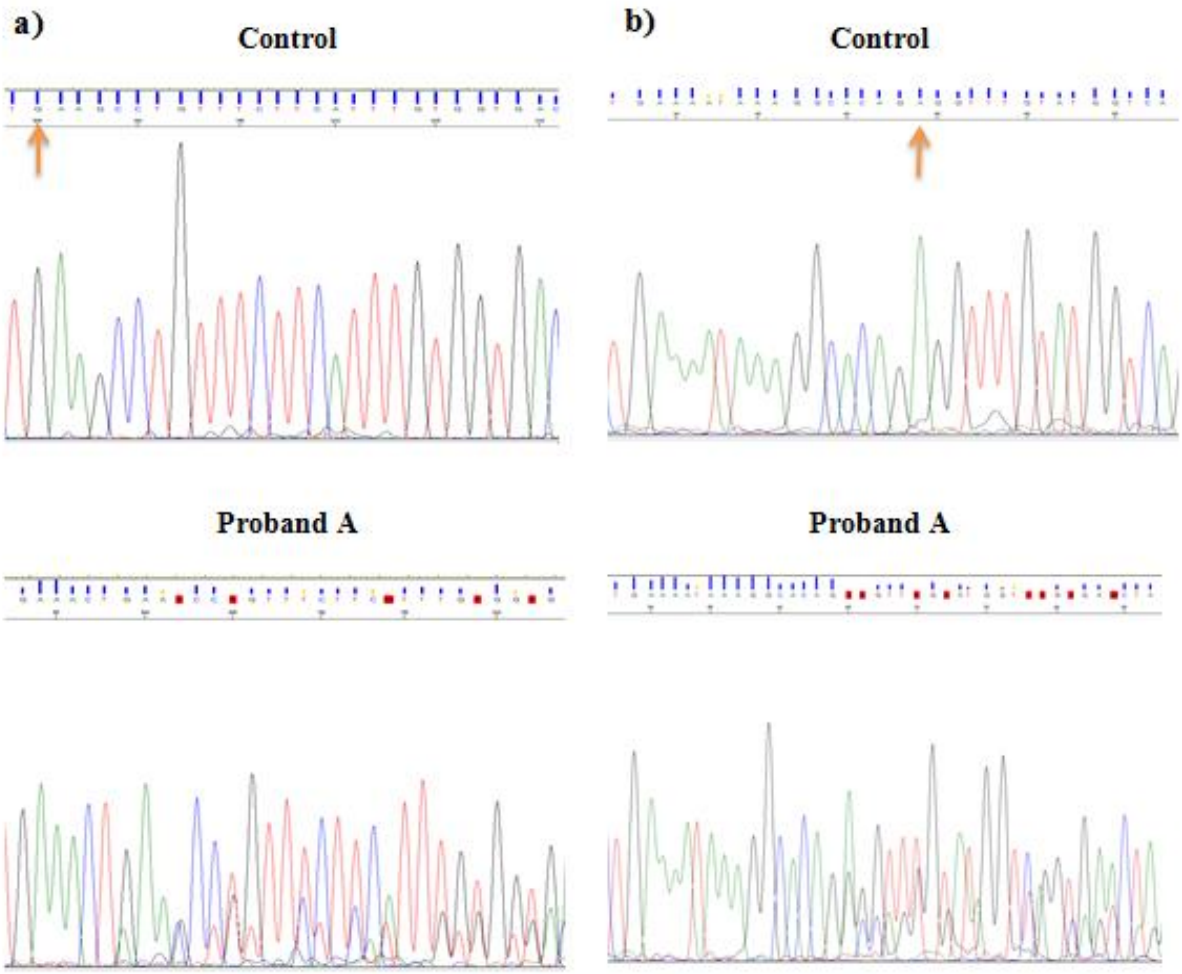

Figure 4.5: Sanger sequencing of RT-PCR products of proband A and control, using primers covering exons 9-12 and 26-29 of *VPS13B*.

a) The variant of c.1426-1G>A disrupted the following sequences and caused frameshift in the proband. The orange arrow shows the first bp of exon 11 in the normal control.

b) The variant of 4157+1G>T disrupted following sequences, and caused frameshift in the proband. The orange arrow shows the first bp of exon 27 in the normal control.

4.3.1.3 Re-evaluation of Proband A

Absence of dup16p11.2-related phenotypes in the mother of Proband A (chapter 2), presence of some of CS criteria in the proband at the age of four (Table 4.5), and more importantly the discovery of *VPS13B* mutations warranted follow up clinical evaluation of our patient at 10 years of age. Prior to that, I reviewed all publications for CS (August 2014), extracted and categorized the phenotypes of CS patients and then prepared a table of diagnostic criteria to be re-evaluated in Proband A. Upon reassessment by both Dr. Lewis and me, Proband A unequivocally demonstrated a pattern of features consistent with CS, including; ID, microcephaly, facial gestalt, retinal dystrophy, truncal obesity with slender extremities, joint hypermobility and episodic neutropenia (Table 4.6). Similar to other reports [217], our study showed that some CS features are age-dependent and evolve later in childhood (Table 4.7).

Table 4.5: Clinical diagnostic criteria of CS (Kolehmainen et al. 2004), and positive features of CS in Proband A at the age of four.

Cohen Syndrome Criteria		**Proband A (4y/o)**
1- Developmental delay/ID		Yes
2- Microcephaly		Yes
3- Typical facial gestalt	Wave shaped eyelids	No
	Short philtrum	No
	Thick hair, eyebrow, eyelashes	Yes
	Low hairline	Yes
4- Truncal obesity with slender extremities		No
5- Cheerful, friendly disposition		Yes
6- Joint hypermobility		Yes
7- High myopia and/or retinal dystrophy		Yes
8- Neutropenia (Isolated, intermittent)		Yes

Table 4.6: Clinical phenotypes of Proband A at the age of 10 were compared with reported features of cases with mutations/CNVs of *VPS13B* gene.

Features in bold are diagnostic criteria of CS.

	Reported findings in patients with Cohen syndrome	Proband A
Pregnancy/ birth	Reduced fetal activity	Yes
	Low birth weight	Yes
	Feeding difficulties	Yes
Growth	Underweight (early childhood)	Yes
	Short stature	Yes
	Truncal obesity (late childhood)	Yes
	Delayed puberty (age-dependent)	No
Neurocognitive	**DD/ID**	Yes
	Autistic/stereotypic behaviors	Yes
	Nonverbal/speech delay	Yes
	Poor motor coordination	Yes
	Hypotonia	Yes
	Seizure	No
	Friendly/sociable personality	Yes
	Brisk reflexes	Yes
	Intracranial abnormalities	No
Cranium	**Microcephaly**	Yes
	Low hairline	Yes
	Thick hair	Yes
Face	Narrow forehead	Yes
	Square/full cheeks (early childhood)	Yes
	Elongated face (late childhood)	Yes
	Hypotonic facial appearance	Yes
	Malar hypoplasia	Yes
	Maxillary prognathia	No
	Micrognathia	Yes
	Short/smooth philtrum	Yes
Mouth/oral region	Small open mouth	Yes
	Thick upper/lower lips	Yes
	High arch palate	Yes
	Prominent upper central incisors	Yes
	Grimacing smile	NE
Eyes/associated	**Thick eyebrow/eyelashes**	Yes

	Reported findings in patients with Cohen syndrome	Proband A
structures	**Downward-slanting /wave-shaped palpebral fissure**	Yes
	Hypertelorism	Yes
	Epicanthic folds	Yes
	Strabismus/ptosis	Yes
	High myopia/ retinal dystrophy	Yes
	Lens opacities/subluxation/optic atrophy/blindness (age-dependent)	No
Nose	Large/beak-shaped nose (age-dependent)	No
Musculoskeletal	**Slender extremities and/**or tapered fingers	Yes
	Joint hypermobility/dislocation	Yes
	Small hands/feet	-
	Sandal gap	Yes
	Joint valgus/varus/pes planus	Yes
	Kyphosis/scoliosis	Yes
Immune system	**Isolated intermittent neutropenia**	Yes
	Recurrent infections	Yes
	Periodontal disease	NE
Miscellaneous	IDDM	No
	Premature aging (age-dependent)	No
	Microcytic anemia	Yes
	Sensorineural hearing loss	Yes
	Cryptorchidism	Yes

Abbreviation: NE: not able to examine

Table 4.7: Evolving clinical features of Proband A.

Comparison of CS Features with Age for Proband A		
Age	**4y/o**	**10y/o**
Weight	$<3^{rd}$ %ile	$5\text{-}10^{th}$ %ile
Oval face	No	Yes
Truncal obesity	No	Yes
Down-slanting, wavy palpebral fissures	No	Yes
Short and smooth philtrum	No	Yes
Long slender distal extremities /fingers	No	Yes
Spine abnormality	No	Yes

4.3.1.4 Expression study of *VPS13B*

qPCR for *VPS13B* using three different replicates of cDNA samples from the proband, his mother and two controls, showed decrease of RNA expression in Proband A compared to controls. The mean expression in Mother A was intermediate between proband and one control (Figure 4.6). Next, I used Western blotting to further study the effect of *VPS13B* mutations on the protein level in our proband, and detected normal intensity of bands at 97 and 159 KDa as expected for this antibody (Figure 4.6). However, two large isoforms (445 KDa and 448 KDa) corresponding to two full-length transcripts of *VPS13B* could not be detected as per my communication with two companies (Abcam (http://www.abcam.com/) and EMD Millipore (http://www.emdmillipore.com/)) providing the commercial anti-VPS13B antibody.

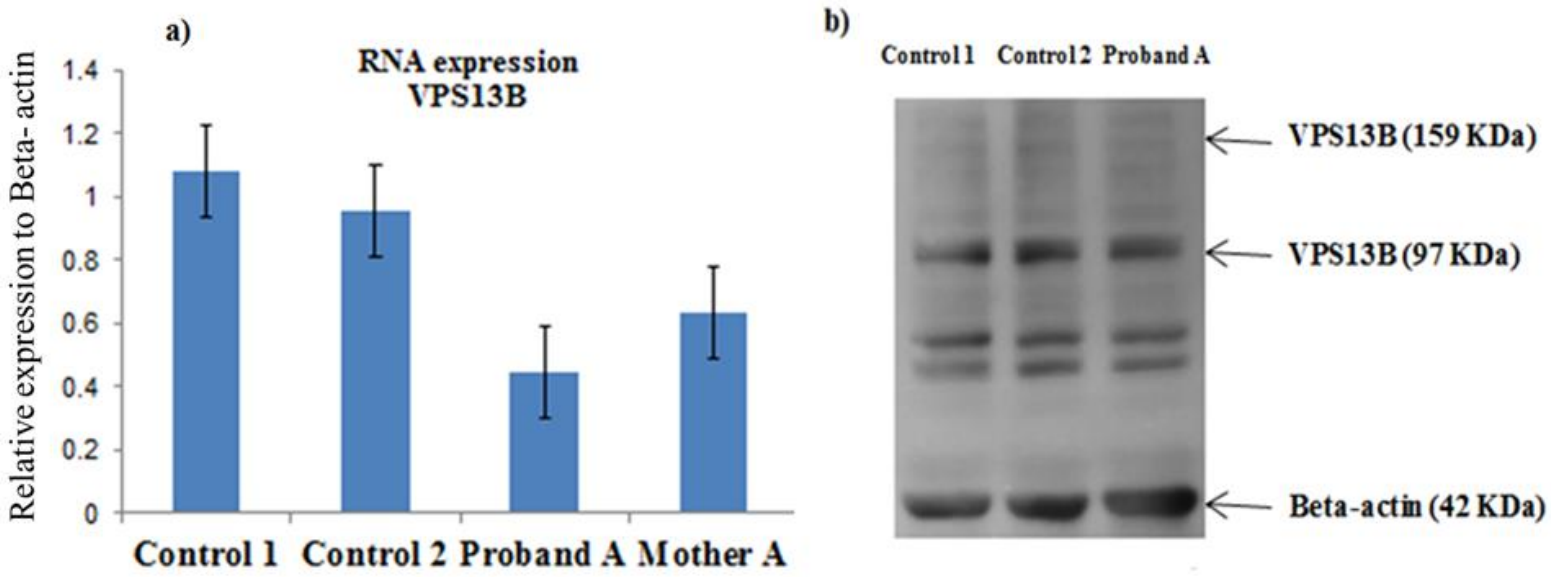

Figure 4.6: Expression study of *VPS13B* gene.

a) The mean RNA expression of *VPS13B* calculated from three different time-series of RNA extraction in Proband A, his mother and two normal controls. The relative expression of *VPS13B* is <0.5 fold in the proband and >0.6 fold in his mother. Error bars indicate standard errors from three replicates.

b) Western blot shows that the intensity of bands at 97 and 159 KDa corresponding to VPS13B are normal in Proband A compared to controls. The band of 42 KDa is corresponding to the loading control antibody (Beta-actin).

4.3.2 WES and variant follow-up in family B

WES analysis revealed one SNV in AR (homozygous), 99 SNVs in AR (compound heterozygous) and 11 SNVs in XL categories. No *de novo* variant was found. In addition to the categories mentioned above, the variants with maternal inheritance identified in proband B were further studied considering that his mother had a positive history of a mild learning disorder in childhood. After filtering out the synonymous variants, 236 variants located at coding/splicing regions remained in this category (maternal heterozygous, MAF of <1%). Although the maternal heterozygous category seems the most relevant inheritance pattern in this family, choosing the best candidate variant(s) from the huge list of potentially pathogenic variants was very difficult. Therefore, I only focused on other inheritance categories. The majority of SNVs were excluded due to lack of damaging effect or high conservation score (Table 4.8). Out of five "pre-candidate genes" that had damaging and conserved variants (*COPS7A*, *ACACB*, *DNAH2*, *SYNE2* and *DIAPH2*), only *SYNE2* (14q23.2) with two compound heterozygous variants fulfilled additional criteria for being disease causing (known function, publication, match to clinical data). The variants mapping to exon 91 (c.16639G>A) were inherited from father B, and to exon 86 (c.15848A>G) were inherited from mother B (Table 4.9). The two variants were subsequently confirmed in the proband (but not present in the brother) by Sanger sequencing (Figure 4.7).

So far, Emery-Dreifuss Muscular Dystrophy (EDMD) is a known consequence of heterozygous mutation of *SYNE2* in humans [218], however, *Syne2* double-mutant mice display significantly smaller brain with severe defects in learning and memory [219].

Table 4.8: Number of filtered variants in each inheritance category in Proband B.

Inheritance (MAF)	Coding, splicing sites	Classification	Damaging	Conserved	Pre-candidate	Known gene function	Relevant publication	Match to clinical data
De novo (< 0.01)	0	-	-	-	0	-	-	-
Autosomal recessive Homozygous (< 0.05)	1	1 (non-syn)	1	1	*COPS7A*	Yes	No	-
Autosomal recessive Compound heterozygous (< 0.05)	99	94 (non-syn)	36	26*	*ACACB* (2)	Yes	No	-
					DNAH2 (2)	Yes	No	-
					***SYNE2* (2)**	**Yes**	**Yes**	**Yes**
		1 (splicing)	-	1	0	-	-	-
		2 (frameshift)	-	0	0	-	-	-
		2 (deletion)	-	0	0	-	-	-
X-linked (< 0.01)	11	10 (non-syn)	3	1	*DIAPH2*	Yes	No	-
		1 (insertion)	-	0	0	-	-	-

Abbreviation: MAF: minor allele frequency, non-syn: non-synonymous *: variants were excluded from compound heterozygous inheritance because only one variant passed previous filters or two variants (from one gene) were inherited from the same parent.

Table 4.9: *SYNE2* variants in family B.

Gene	Inheritance	Chromosome-position	Classification	Exon	Variation	Amino acid change	Damaging score	GERP	PhyloP
SYNE2-1	Paternal	14:64633984	Non-syn	91	c.16639G>A	p.Asp5547Asn	2/5	4.89	1.588
SYNE2-2	Maternal	14:64625398	Non-syn	86	c.15848A>G	p.Asp5283Gly	4/5	4.63	1.113

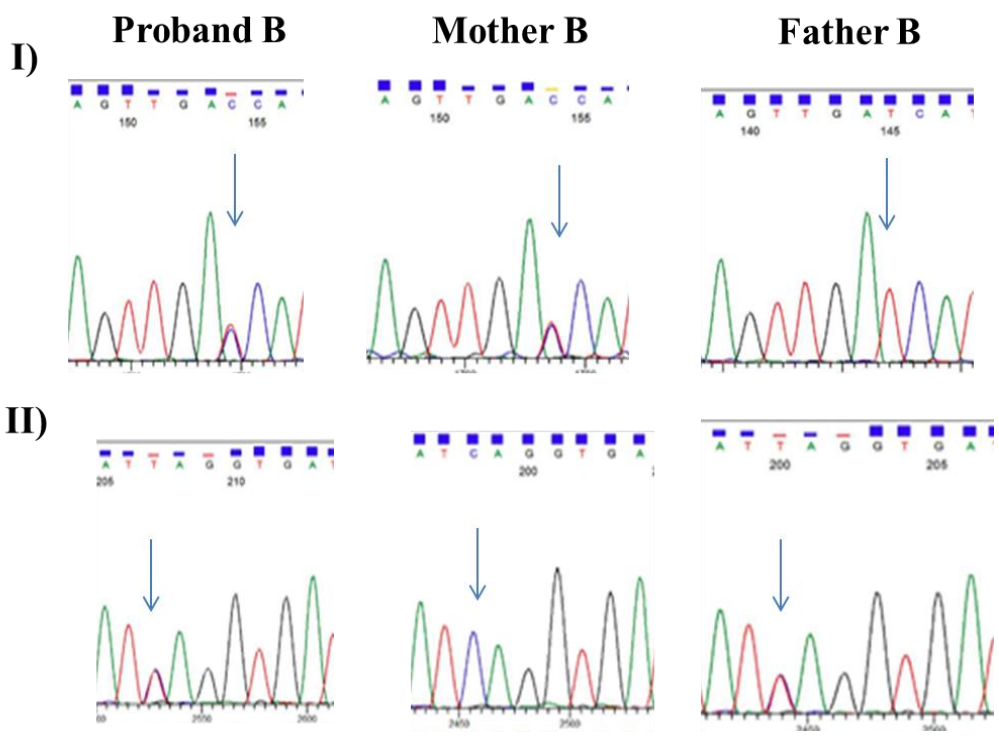

Figure 4.7: Sanger sequencing analysis of *SYNE2* variants in family B.

I) Proband B and his mother are carriers of mutation c.15848A>G. II) Proband B and his father are carriers of mutation at c.16639G>A (Sequences of reverse strands are shown).

In addition to the *SYNE2* mutations, I detected a *DMD* mutation of c.726T>G located in exon 9, which had initially been excluded due to low conservation score. Mutations of *DMD* cause Duchenne and Becker muscular dystrophy characterized by progressive muscle atrophy and elevated serum creatine phosphokinase (CPK). Co-morbid features of *DMD* mutations include cognitive impairment and ADHD, sometimes without muscular dystrophy [220, 221]. Because Proband B suffers from some degree of myopathy with muscle biopsy showing non-specific type-1 fiber predominance, and his brother has a learning deficit and ADHD, I validated the DMD variant in both brothers (Figure 4.8) and determined that it is conserved in most mammals in UCSC genome browser search (Figure 4.9). However, given that I did not have access to brother B's LCLs, and Proband B's CPK level was in normal range, no further studies of the DMD variant were pursued, instead focusing on the expression analyses for *SYNE2*.

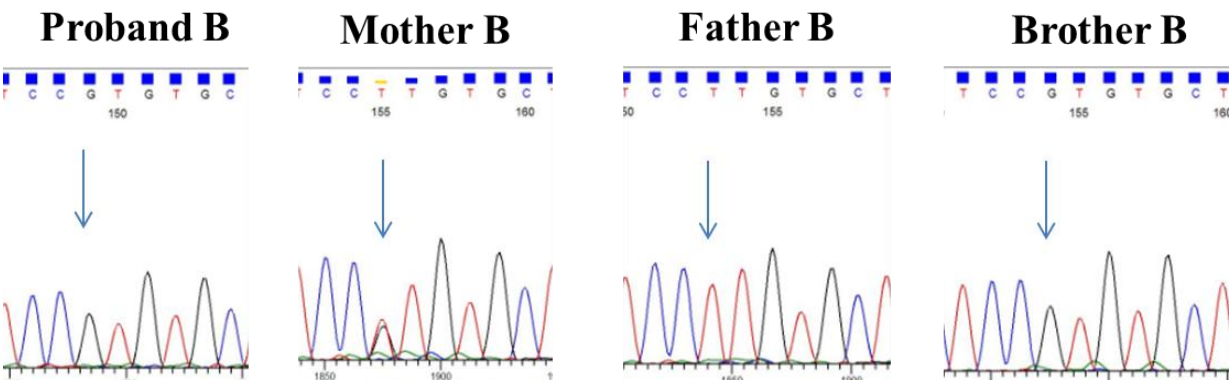

Figure 4.8: Sanger sequencing analysis of *DMD* variant in family B.

The variant of c.726T>G is positive in Proband B, Mother B and brother B (Sequences of forward strands are shown).

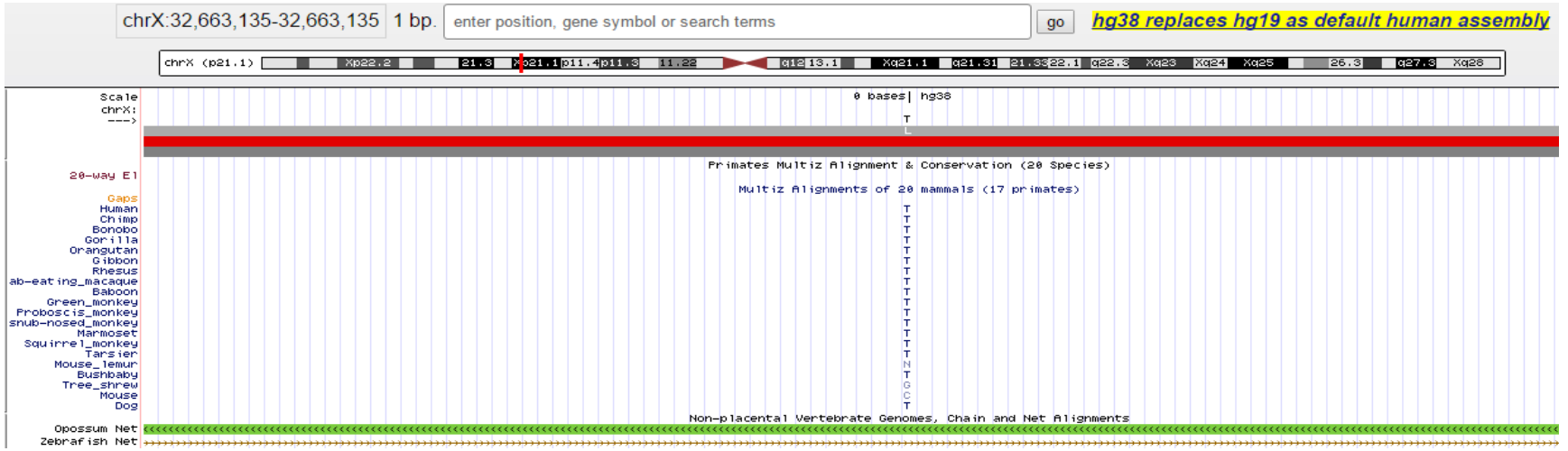

Figure 4.9: The screenshot of the UCSC genome browser shows the position of *DMD* variant, and the specific nucleotide present in each mammal at this position.

4.3.2.1 Expression study of *SYNE2*

Overall, the mean RNA and protein expression of *SYNE2* displayed over-expression of this gene in Proband B and his mother compared to two controls (Figure 4.10 and Figure 4.11). The protein encoded by *SYNE2*, Nesprin 2, at the band of ~20 KDa showed stronger intensity in Proband B and his mother (Figure 4.12). Of note, according to the company's web site, the anti-Nesprin 2 antibody had a molecular weight of 37 KDa but was tested on the purified recombinant proteins (http://www.abcam.com).

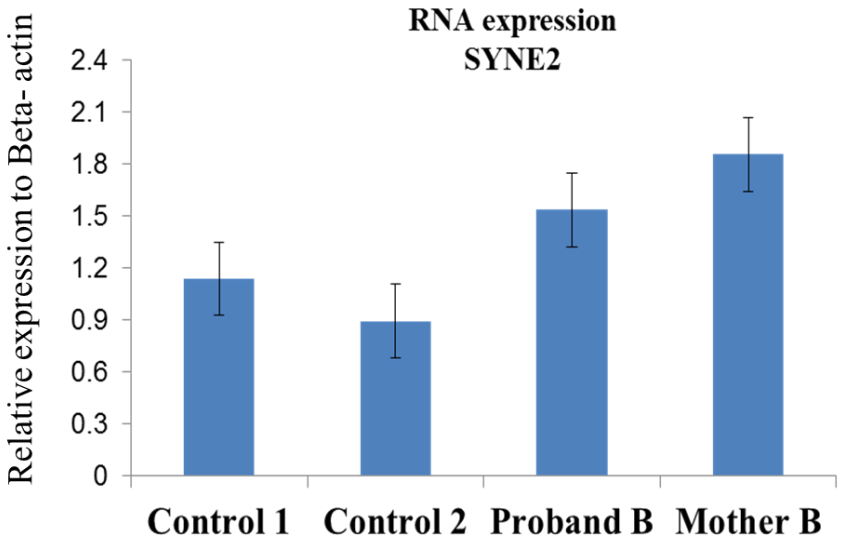

Figure 4.10: The mean RNA expression study of *SYNE2* in family B.

The Mean RNA expression of *SYNE2* calculated from qPCR results of three replicates indicates that the expression level of *SYNE2* is higher in Proband B and his mother compared to controls. Error bars indicate standard errors from three replicates.

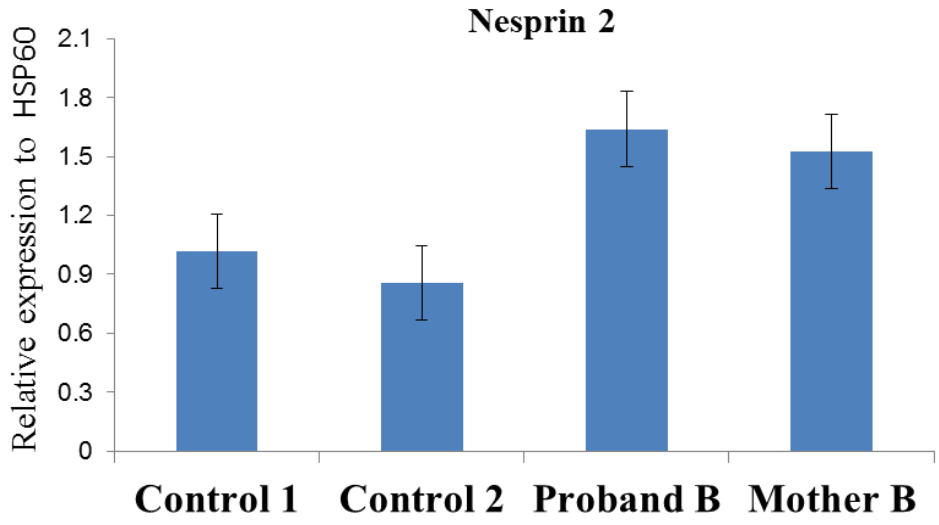

Figure 4.11: The mean expression of Nesprin 2 in family B.

The mean expression of Nesprin 2 is calculated from three replicates of western blotting in family B. The expression of Nesprin relative to HSP60 is higher in Proband B and his mother compared to controls. Error bars indicate standard errors from three replicates.

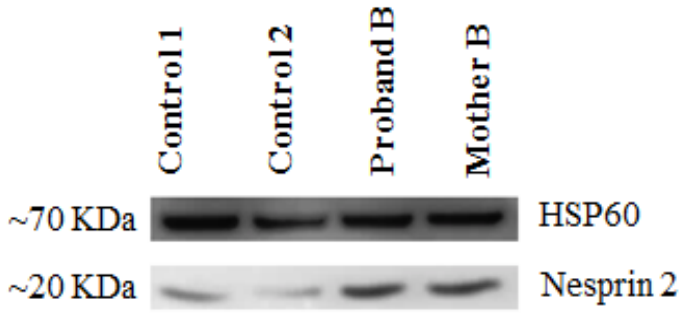

Figure 4.12: Western blotting of Nesprin 2 in family B.

Bands of ~20 KDa (Nesprin 2) and ~70 KDa (HSP60) with higher intensity of Nesprin 2 were detected in Proband B and Mother B.

4.4 Discussion

4.4.1 Family A

In Proband A, two novel deleterious mutations of *VPS13B* located at conserved splicing sites were discovered. Proband A's phenotype (ID, microcephaly, facial gestalt, retinal dystrophy, joint hypermobility and episodic neutropenia), mRNA sequence aberration, and *VPS13B* RNA expression findings were consistent with CS.

The *VPS13B* gene, also known as *COH1* (OMIM: 607817), is approximately 864 kb in length and located on chromosome 8q22.2. It consists of 62 exons encoding a transmembrane protein of 4022 amino acids [216]. VPS13B protein carries two short loci homologous to yeast vacuolar protein sorting-associated protein 13 (Vps13p) [222, 223]. VPS13B is a peripheral membrane protein that is required for function, orientation and structural integrity of the Golgi apparatus and thus plays a role in vesicle-mediated sorting and intracellular protein transport [224, 225].

Homozygous or compound heterozygous mutations/CNVs of *VPS13B* cause CS [216]. Recent study suggests that the underlying mechanism of CS is linked to defective tissue-specific glycosylation and endosomal–lysosomal trafficking [226]. CS has a broad clinical phenotype spectrum including ID, microcephaly, hypotonia, dysmorphic facial features, truncal obesity, slender extremities, joint hypermobility, myopia, retinal dystrophy, intermittent isolated neutropenia, and happy personality. Neutropenia is characterized as a neutrophil count of $<1.5 \times 10^9$/L in children and $<1.8 \times 10^9$/L in adults [227]. The facial gestalt includes down-slanting palpebral fissures, wave-shaped eyelids, thick eyebrows and eyelashes, low hairline, prominent and beak-shaped nose, malar hypoplasia, short philtrum, high-arched palate, maxillary prognathia and prominent central incisors [217, 228, 229]. Patients with CS grimace when they

are asked to smile [33, 216, 230]. Other signs and symptoms include short stature and scoliosis [33, 216]. In addition, individuals with CS have high rates of ASD or autistic features [34]. Although reported Finnish cases presented with a homogeneous phenotype as result of one specific founder mutation (c.3348_3349delCT), patients of other ethnicities demonstrate a broad clinical spectrum with age dependent features [222, 230, 231].

The estimated prevalence of CS is 1:105,000 [232], however, its frequency may be considerably higher due to the fact that patients are often not diagnosed until they reach their teenage or adult years. The early diagnosis of CS is challenging because facial features are less noticeable in pre-school age, truncal obesity may evolve in late-childhood, neutropenia is rarely identified due to its intermittent pattern and absence of clinical consequences, and diagnosis of retinal dystrophy usually occurs in later childhood [33, 142, 217, 222, 227, 230, 231, 233-235].

Several studies have reported that *VPS13B* encodes several transcripts [227]. The two full-length transcripts, NM_017890 and NM_152564, contain exons 1-62, utilizing exon 28 or exon 28b, respectively. Both full-length transcripts encode functional proteins and are equally expressed in both human fetal and adult brain as well as adult retina, whereas the variant containing exon 28b is the major transcript ubiquitously expressed in all tested human tissues (kidney, liver, placenta, small intestine, and lung), at levels five times higher than transcript NM_017890. Other shorter transcripts are less frequently expressed [227].

Intronic point mutations within donor and acceptor sites at mRNA splice junctions typically cause mRNA mis-splicing, leading to subsequent nonsense-mediated mRNA decay (NMD), and altered protein with effect on the clinical phenotype [236, 237]. Indeed, Sanger sequencing of RT-PCR products corresponding to each specific *VPS13B* variant demonstrated

that both variants create aberrant RNA sequences and frameshift and thus probably lead to NMD.

The association of *VPS13B* with Golgi apparatus functions depends on GTPase RAB6. Down-regulation of either *VPS13B* or *RAB6* results in significant reduction in neurite outgrowth. In fact, 50% reduction of *VPS13B* mRNA significantly reduced the length of neurites to ~40% [225]. Interestingly, mRNA level of *VPS13B* gene was reduced in proband A (<50%) compared to two controls, whereas its level of RNA expression in his heterozygous mother was intermediate (>60%) between proband and one control, suggesting that partial loss-of-function in heterozygous carriers of autosomal recessive disorders is not sufficient to produce a complete disease phenotype. Although it would be expected to see the effect of deleterious mutations on protein level, I was not able to assess the translational effect of *VPS13B* mutations on two large isoforms (445 KDa and 448 KDa) corresponding to two full-length transcripts, as the available anti-VPS13B antibody produces only two bands at 97 and 159 KDa, corresponding to shorter transcripts.

4.4.2 Family B

In Proband B, I identified compound heterozygous variants in *SYNE2*, not present in his brother. The RNA and protein expression of *SYNE2* was up-regulated in Proband B and his mother. The mother has a past history of learning and attention deficits, and more recently suffers from memory problems. I have also identified a variant in *DMD* found in the proband, his mother and brother, which could be a candidate for the milder learning and attention deficits in this family.

The *SYNE2* (OMIM: 608442) contains 115 exons with approximately 370 kb in length [238]. There are four Nesprins described in mammals, each encoded by a different gene (*SYNE1*,

SYNE2, *SYNE3*, *SYNE4*) [238-241]. Nesprins are intracellular linkers and scaffold proteins with highly versatile tissue-specific functions, including cellular stiffness, ciliogenesis, organelle positioning, endocytosis, Wnt-signaling and cell adhesion [219, 242-249].

Nuclear movement is an essential element of neurogenesis and neuronal migration [219], and is crucial for development and patterning of the mammalian brain [250]. Several proteins including Lamin A and C, Emerin, SUN 1/2, and Nesprin 1/2 are located at the inner nuclear membrane (INM) and form an evolutionary conserved physical link between nucleus and cytoplasm called LINC (Linkers of Nucleoskeleton and Cytoskeleton) complex [251]. LINC complex not only mediates the mechanical interactions across the nuclear envelope [242, 252], but also functions in signaling pathways and gene regulation [251]. Dysregulation of any of LINC proteins causes detachment of the nucleus from the cytoplasmic cytoskeleton and leads to nuclear deformity [253, 254].

Two large Nesprins, Nesprin 1 and Nesprin 2, are conserved in the worm and fly [255, 256] and are characterized by a C-terminal transmembrane KASH (Klarsicht–ANC–Syne-homology) domain, N-terminal actin binding CH (calponin homology) domains (also known as ABDs (actin binding domains)), and a rod domain composed of multiple spectrin repeats (SRs) which separate KASH and ABD domains [257] (Figure 4.13).

Nesprin 2 contains 6,884 amino acids, and 14 different isoforms with different sizes (different length of rod domain) ranging from 20 to 796 KDa [238, 258]. Giant isoforms of Nesprin 2 link outer nuclear membrane (ONM) with intracellular organelles such as Golgi apparatus [245, 259, 260], whereas shorter isoforms are at the nucleus and INM [261].

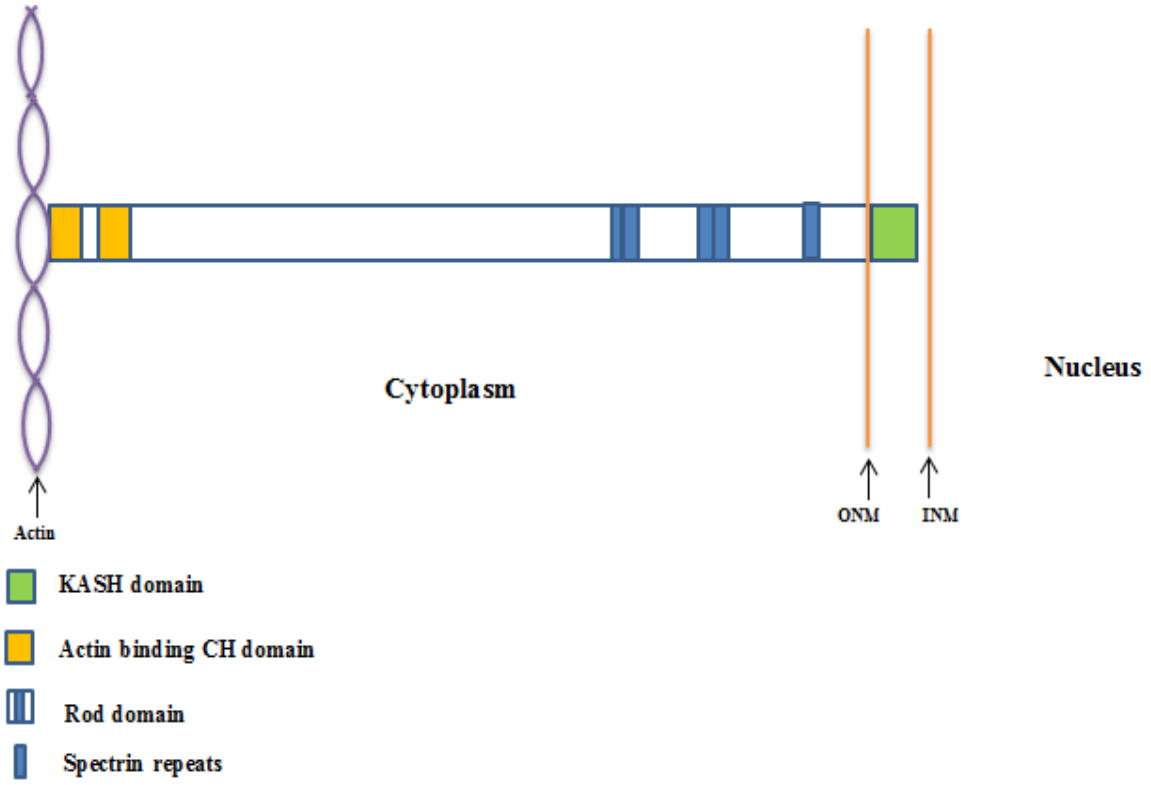

Figure 4.13: Schematic of Nesprin 1 and 2.

Abbreviation: ONM: outer nuclear membrane, INM: inner nuclear membrane. (Adopted from both NCBI and paper published by Meinke et al. 2011[251]).

Although the function of Nesprin 2 is not clear, this protein is involved in a variety of cellular functions. Nesprin 2 interacts with two Meckel-Gruber syndrome (MKS) proteins, Meckelin and MKS1, and mediates ciliogenesis [249]. Meckelin is frequently mutated in MKS and Joubert syndrome [262, 263]. The depletion of Meckelin or Nesprin 2 disrupts correct centrosome migration and thus inhibits cilia formation [249]. Alteration of ciliogenesis, which is essential for brain development, particularly brain patterning [264], can cause pleiotropic disorders including ID [265]. Notably, a recent study identified a significant pathway enrichment relevant to neurodevelopmental and primary ciliary functions in carriers of 16p11.2 CNVs [194], however *SYNE2* was not amongst the genes implicated. Nesprin 2 is also crucial for DNA repair and regulates Wnt signaling at the nuclear envelop through its interaction with α-Catenin [266, 267].

Mutations affecting different isoforms of Nesprins contribute to their diverse disease phenotypes. For instance, heterozygous mutations of C-terminal regions of Nesprin 2 in humans cause Emery-Dreifuss Muscular Dystrophy (EDMD). EDMD is characterized by joint contracture, progressive muscle wasting and cardiomyopathy [218, 268]. However, *Syne2* mice with double-deletion of KASH domain die at birth, whereas double KASH-mutations affect neuronal migration in the cerebral cortex and the hippocampus and produce mice with smaller brain size and severe learning and memory deficit compared to controls or heterozygous mutant mice [219]. Further, ABD domain-knockout of Nesprin 2 in mice leads to complete loss of the giant isoform of Nesprin 2, and although they are viable and do not display physical anomalies, they suffer from epidermal thickness and wound healing defects [247, 269].

It is also interesting that over-expression of different isoforms of Nesprin 2 have different effects on the nucleus. For example, up-regulation of the giant isoform expands the nuclear area,

whereas over-expression of the small isoform, containing both ABD and KASH domains and no rod domain, decreases nuclear volume [247, 270].

In the family B, the variant of *SYNE2* inherited from the mother (but not the father) is affecting one SR of the rod domain in Nesprin 2 (Figure 4.14). SRs anchor proteins and orchestrate specific protein–protein interactions which facilitate several cellular processes including structural elasticity [271]. Mutations in the rod domain of Nesprin 2 have not yet been reported in human diseases. Without access to an LCL sample from father B, it is difficult to predict the effect of paternal mutation on expression of *SYNE2*, especially because it is located outside of rod domains and has no known phenotypic consequence. However, based on mice data [219], I speculate that maternal missense mutation of *SYNE2* may induce thermodynamic instability and protein aggregation with dominant negative effect, and thus cause some of the observed NDD phenotype in the proband and his mother. Of note, mutation of *SYNE2* also affects the skeletal muscle and causes myopathy [218, 268] which may explain mild myopathy with nonspecific type-1 fiber predominance present in the Proband B. However, further systematic functional validation studies of downstream biological pathways affected by *SYNE2* mutations are required to confirm their role in the phenotype.

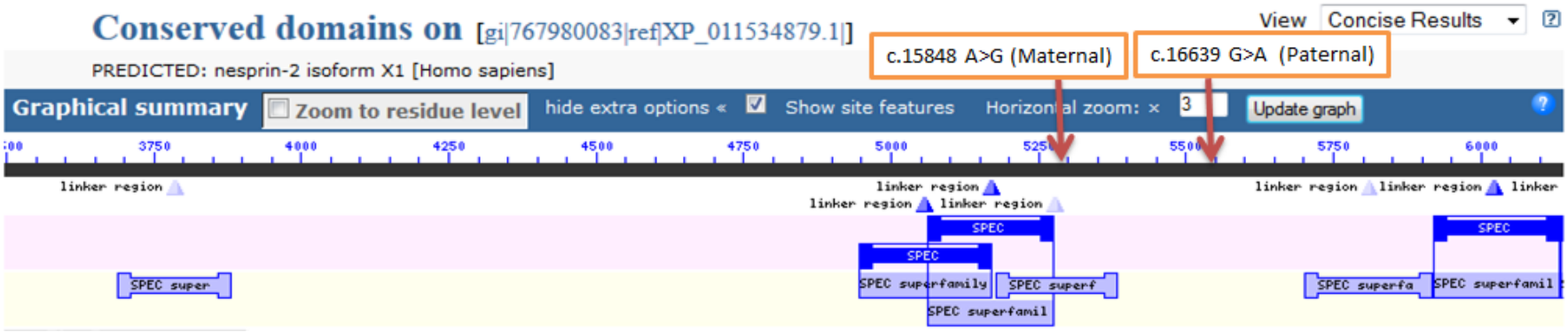

Figure 4.14: The position of *SYNE2* mutations shown in Nesprin 2 (from NCBI).

The position of two *SYNE2* mutations in family B has shown on the Nesprin 2. Only the maternal variant affects spectrin repeats.

The presence of a *DMD* variant in the proband, brother and his mother is also of interest. *DMD* (MIM: 300377) is the largest human gene with a size of >2.2 Mb containing 86 exons. The protein encoded by the *DMD* gene, dystrophin, is disrupted in Duchenne Muscular Dystrophy (DMD) and Becker Muscular dystrophy (BMD). Dystrophin creates a link between the intracellular cytoskeleton and the extracellular matrix through F-actin; and possesses multiple SRs with 21–28% homology to the rod domain of Nesprin 2 [261, 272-274]. It has been suggested that mutations located downstream of exon 44 of *DMD* gene disrupt Dp140 and Dp71 isoforms (expressed in the central nervous system) and cause cognitive impairment in patients with DMD [275, 276]. About 25% of DMD cases with mutations upstream of exon 44 present with ADHD, some without ID [221]. The estimated frequency of ADHD among cases with DMD is 12-50% [277-279] compared to 3-7% in the general population [280, 281].

It is worth noting, that the variant of *DMD* in this family is located in exon 9 (upstream of exon 44) and disrupts one of many SRs of dystrophin. Although contribution of the *DMD* mutation to the phenotype of family B is of uncertain significance, it is possible that the *DMD* mutation is responsible for ADHD as well as some degree of learning problems in this family. Oligogenic theory may explain the collective clinical phenotypes of family B. It is also important to note that other variants could still be implicated because the filtering strategy used for WES analysis in this study excluded variants that were damaging and conserved, but with no known function at the moment.

In conclusion, in this chapter I identified that disease-causing genomic variants outside of the 16p11.2 region contribute to the variable expressivity observed in Proband A and B. These findings serve as an example that the extreme clinical variability seen among carriers of

dup16p11.2 could be due to the presence of diverse pathogenic genome variation other than this CNV.

Chapter 5: Discussion

Variable expressivity among carriers of dup16p11.2 confounds accurate genetic counselling. In this project, I investigated several factors that may contribute to the variable expressivity observed in such individuals. I performed a standardized and detailed phenotypic analysis of features observed among carriers of dup16p11.2, assessed the impact of dup16p11.2 on its integral and nearby gene expression in the probands and healthy carrier parent, and searched for disease-causing variants in other areas of the genome in two affected dup16p11.2 probands.

In this discussion, I will summarize the main findings and their significance, highlight the strengths and limitations, and discuss future directions for study in this field.

5.1 Summary and significance of findings

CMA analysis of patients with NDD has uncovered a large number of CNVs, some of which are challenging to interpret as they occur in clinically heterogeneous disorders. In particular, proximal duplication of the 16p11.2 region has been detected in patients with a broad spectrum of NDD as well as individuals without a recognized phenotype.

In chapter 2, I searched for phenotypic commonality among carriers of dup16p11.2 and for the possibility that apparently healthy carrier parents might have some unnoticed clinical features. I performed phenotypic analysis of 18 patients with dup16p11.2 (16 published and two probands from my study), and also evaluated the available data from six carrier parents (four published and two carrier from my study). Although patients with dup16p11.2 share some clinical phenotypes, no recognizable and consistent clinical patterns were found among them and/or carrier parents. Results of this analysis also showed that the frequency of known features of dup16p11.2 such as microcephaly (44%) and underweight (11%) are not as common as

generally thought. It is important to note that some clinical features (e.g. head circumference, weight, height and facial gestalt) evolve with age, and thus may be more representative of the effect of dup16p11.2 at one specific, but not all stages of life.

An emerging theme in the field of medical genetics is that 16p11.2 CNV increases the risk of NDD, based on higher frequency of this CNV in patients compared to controls. However, ascertainment bias may affect such studies. For example, a significantly higher frequency of dup16p11.2 among patients (0.25%) than controls (0.04%) was reported by Kaminsky *et al.* and led to re-classification of dup16p11.2 as a pathogenic CNV, although being previously known as a VOUS [64]. Tucker *et al.* screened 6813 cord blood samples mostly from French-Canadian ancestry for presence of rare CNVs including CNV at 16p11.2 region [187]. The authors found five dup16p11.2 carriers (0.07%) in this cohort which is higher than the results of previous studies in control populations (0.03-0.04%) [64, 147, 149]. This may suggest that population stratification can affect the measurement of susceptibility risk of dup16p11.2 [187]**.** Moreover, the presence of unaffected carrier parents/siblings suggests an ascertainment bias against clinically asymptomatic or mildly affected carriers; as they are not the target of chromosome microarray.

In brief, inter- and intra-familial variability among carriers of dup16p11.2 presented in this chapter suggests that dup16p11.2 itself is not sufficient to produce an abnormal phenotype; hence other genetic factors may influence clinical heterogeneity. These results are in keeping with variable but predominantly less affected transgenic mice carrying the dup16p11.2 compared to deletion mice. Similarly, the difference in the level of gene expression in several brain areas was much less dramatic in dup-mice compared to controls [106].

In chapter 3, I questioned whether any difference in specific dosage effects of the genes in the 16p11.2 region between the study probands and their unaffected carrier parent could explain their difference in phenotype. Luo *et al.* reported altered expression of genes within the 16p11.2 CNV and showed that the gene expression in unaffected 16p11.2 CNV carrier mothers were more similar to controls than their affected children, suggesting that expression level changes may partially explain the variable expressivity of CNV at this region [96]. I evaluated the effect of duplication on genes within and nearby the 16p11.2 region in two probands, and subsequently for selected genes in the probands and one carrier parent. The result of WGE analysis contradicted the earlier finding of a positive correlation between copy number change of "all" genes located within 16p11.2 CNV region and their gene expression [94]. Subsequent expression validation of three candidate genes in this region (*KCTD13*, *MAPK3* and *MVP*) showed that these genes were up-regulated only in Proband B, suggesting a possible and yet unknown compensatory mechanism(s) in Proband A. Nevertheless, my results are consistent with other studies [95, 96, 99] indicating that dosage compensation occurs for genes within rare CNVs . Moreover, my data showed that the expression of these three genes had a subtle change in mother A (unaffected carrier of dup16p11.2).

Golzio *et al.* demonstrated that *KCTD13* is a major driver of brain phenotype in 16p11.2 CNV zebrafish, and *MAPK3* and *MVP* are KCTD13 modifiers [117]. Although injection of *KCTD13* mRNA (over-expression) may cause microcephaly in zebrafish, it is important to note that *KCTD13* transcript was not over-expressed in Proband A, although he showed microcephaly of postnatal onset. Therefore, factors other than *KCTD13* are responsible for the microcephaly seen in this boy. Intriguingly, protein study of KCTD13 demonstrated similar expression (within normal limits) in both probands, an unaffected carrier parent, non-carreir parent and two

controls, suggesting that the *KCTD13* is not the sole cause of microcephaly in the studied subjects.

In chapter 4, using WES, I presented evidence that pathogenic variants outside of the dup16p11.2 region could contribute to the variable expressivity observed in the carriers from two families. I discovered compound heterozygous variants of *VPS13B* (8q22.2) located at conserved canonical sites in the Proband A. Mutations of *VPS13B* cause CS in keeping with the phenotype (ID, microcephaly, facial gestalt, retinal dystrophy, joint hypermobility and episodic neutropenia), mRNA sequence aberration, and low RNA expression in Proband A. In family B, I found compound heterozygous variants of *SYNE2* (14q23.2) in Proband B, and a variant of *DMD* (Xp21.1) shared between the proband, his brother and mother. The mean RNA and protein expression of *SYNE2* were up-regulated in both the Proband B and his mother. The result of recent expression study for carriers of 16p11.2 CNV suggests that ciliary gene alterations may be responsible for clinical variability of 16p11.2 CNVs [194]. Interestingly, Nesprin 2 plays a role in ciliogenesis [249]. Mutations affecting different parts of Nesprins affect different isoforms and thus may contribute to the diverse disease phenotypes associated with variants of Nesprin [282]. Heterozygous mutations of C-terminal regions of Nesprin 2 in humans cause EDMD, whereas double KASH-mutations produce mice with smaller brain size and exhibit severe learning and memory deficit compared to controls or heterozygous mutant mice [219]. Proband B does not show the expected manifestations of EDMD, however, he has ID, microcephaly and myopathy; and his mother has a history of learning problems and memory deficit. Thus, I speculate that mutant Nesprin 2 may have a dominant negative effect interfering with the wild type functional protein, leading to a more severe phenotype in the proband (compound heterozygous) and milder features in his mother (one mutation). Contribution of the *DMD* mutation to the phenotype of

family B is of uncertain significance as the CPK level of Proband B was normal, his muscle biopsy shows only nonspecific type-1 fiber predominance and I was not able to further evaluate this variant in the brother.

My study also showed that the diagnosis of syndromic NDD such as CS cannot always be ruled out in the absence of typical facial findings as some phenotypic features evolve with age. Serial genetic testing for such evolving heterogeneous disorders is time consuming and more expensive than WES [283], and risk missing earlier diagnosis and possible treatments that can lead to better outcomes, both behavioural and medical. Early application of WES leads to earlier diagnosis and more informed individual management and intervention, which for CS in family A, can possibly avert or ameliorate secondary insulin-dependent diabetes, cataracts, blindness, tooth loss and premature aging [284]. Consistent with the results of other studies [142, 143], WES in this study also emphasized that recessive-acting mutations should be considered even when non-consanguineous families are involved.

The estimated penetrance of dup16p11.2 reported by Rosenfeld group [149] might be a useful tool in prenatal genetic counseling in families at risk for dup16p11.2. Genetic counsellors can use this estimation to explain that a fetus with dup16p11.2 has ~27% chance of being affected; whereas a negative microarray result indicates that the fetus has no heightened risk above the general population risk. However, as my study indicates, this risk could be higher if the possible presence of other disease-causing variations is taken into consideration. For instance the hypothetical fetus with negative array in family A still has 25% chance of being affected by CS. The results of this study will have a great impact on genetic counselling of families with rare putatively pathogenic CNVs that harbor variable expressivity and incomplete penetrance.

5.2 Strengths and limitations

This study is focused on variable expressivity among carriers of dup16p11.2. The strength of my study is that it used a multifaceted approach to explore several possibilities that might explain this clinical variability. Phenotypic investigation was used to find consistent and recognizable features among carriers, and to explore the possibility that unaffected carrier parents might have some subtle dup16p11.2 features if carefully looked for. This study provides a deeper and more accurate characterization of dup16p11.2 phenotypes. I also considered and evaluated the possibility that there could be a difference in the expression level of genes within 16p11.2 region between Probands A and B compared to their unaffected parents. More importantly, I considered the possibility of secondary disease-causing variants among patients with dup16p11.2. This is the first study to perform WES in subjects with dup16p11.2. Reverse-phenotyping is a newly emerging approach; however the approach used in this study (detailed phenotypic analysis using WBDD) highlighted that initial carefully categorized phenotype assessments benefited a more informed WES analysis. The diagnosis of CS in Proband A also underscores the importance of longitudinally following patients for evolving phenotypic features. Overall, this study can serve as a template for analysis of other ambiguous rare CNVs that harbor variable expressivity and incomplete penetrance.

The main limitation of my study is the small sample size. Further, 16p11.2 CNV is a recurrent CNV with possibly identical breakpoints, however; my study would benefit from using uniform microarray platforms in both probands and their parents.

In addition, the presence of the dup16p11.2 and mild phenotype of brother B and his mother later came to attention after performing WES in Proband B and thus a more ideal scenario for this family would be to perform WES and expression studies in the proband, his

brother and both parents at the same time. I was also not able to evaluate the expression of *SYNE2* in the father of Proband B as the LCL sample from him was not available at the time of this study. Despite the fact that a *DMD* variant may in part explain the milder phenotype of the brother of Proband B, the pathogenicity of this variant or the influence of the identified dup1611.2 CNV remains uncertain.

Moreover, although I identified the damaging variants that most likely are causative, there are other variants that might contribute to the phenotype of family B and more forthcoming information is needed. For example, in family B, non-synonymous variant of *COPS7A* with AR inheritance (homozygous), was detected, and considered as damaging by 3/5 functional prediction tools, and also is located at a very conserved genomic location (GRP: 5.88, PhyloP: 2.24). *COPS7A* is a part of the COP9 signalosome complex which is involved in various cellular and developmental activities [285]. So far, the role of *COPS7A* in humans is unknown, and thus the variant of this gene was filtered out from the list of candidate genes. Nevertheless, future studies may provide new information relevant to human disease which could affect today's decision.

The other concern might be that the adopted LCL cellular model system could alter the expression of genes due to viral transformation, or that may not express all relevant CNS genes [286, 287]. Thus, this study may have missed the brain specific changes of KCTD13 protein that are not expressed in lymphoblasts.

Lastly, the proteins encoded by *VPS13B* and *SYNE2* are very large and have multiple isoforms, and thus cannot be easily detected by Western blotting, especially when currently available commercial antibodies cover only one or two small isoforms. Theoretically, each isoform can be detected using antibodies against isoform-specific amino acid sequences, yet such

antibodies are not commercially available [288]. Application of mass spectrometry could help to overcome this limitation in detecting diverse and large proteins with multiple isoforms [289]. In brief, establishing the functional consequence of NGS detected mutations is a major challenge in data analysis.

5.3 Conclusions and future research directions

This study suggests that future human research would benefit from uniform and standardized detailed phenotyping. Such comprehensive phenotyping not only helps elucidate a clearer and earlier clinical diagnosis, but also critically benefits and informs concordant NGS analysis. More accurate phenotype/genotype correlation is essential to accelerate the identification of disease subtypes, natural history and therapeutic implications.

The understanding of the role of 16p11.2 CNV can be improved in the future by expanding the exome sequence analysis to a larger number of familial and *de novo* 16p11.2 CNV cases that have standardized physical and neuropsychological evaluation. This work lays the framework for future directions and advances our understanding of the variable expressivity among rare CNVs.

Furthermore, the ability to better understand the influence of rare CNVs or mutations on neural development is in the use of the affected tissue type. This became possible thanks to the development of induced pluripotent stem cells (iPSc) technology [290]. Human iPSc generated from primary somatic cells (e.g: fibroblast) of patients with known genotypes can be used for this purpose. One of the strengths of using iPSc is that they can expand to unlimited number of specialized cells. Application of human iPSc-derived neuronal cells may also provide this benefit as well as an opportunity for high throughput drug screening. However, it is important to note that the differentiation of iPSc takes several months, sometimes with inconsistent results. In

addition, their maintenance and prevention from uncontrolled differentiation are very difficult [291].

The approach and the body of work presented in this thesis contribute to a more advanced understanding of the variable expressivity of dup16p11.2. My study highlights the importance of a multifaceted phenotype-genotype correlation study including detailed phenotyping, gene dosage and expression comparisons of probands and their carrier parents versus controls, and the search for cryptic pathogenic variations in other parts of genome confounding expected CNV phenotype.

In conclusion, I discovered in each proband that disease-causing variants in genes distinct from the 16p11.2 CNV region contribute to the clinical variability between parent-offspring dup16p11.2 carriers. My findings suggest discordance in phenotype of dup16p11.2 carriers warrants further study by WES and individualized clinical genetic assessment and counselling of families with dup16p11.2.

Genomic microarray is a valuable first-tier test for the postnatal evaluation of individuals with NDD including ID, ASD, and/or multiple congenital anomalies [62, 144]. However, coupling of microarray with WES or whole genome data analyses will facilitate a more comprehensive and accurate analysis of genetic causes of NDD, heighten understanding of the etiology of variable expressivity among NDD patients, and optimize clinically-informed and effective genetic counselling and personalized treatment options.

Bibliography

1. Coe, B.P., S. Girirajan, and E.E. Eichler, *The genetic variability and commonality of neurodevelopmental disease.* Am J Med Genet C Semin Med Genet, 2012. **160C**(2): p. 118-29.

2. Dietrich, K.N., et al., *Principles and practices of neurodevelopmental assessment in children: lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research.* Environ Health Perspect, 2005. **113**(10): p. 1437-46.

3. Association, A.P., *Diagnostic and statistical manual of mental disorders (5th ed.).* 2013. Washington, DC.

4. Kim, Y.S. and M.W. State, *Recent challenges to the psychiatric diagnostic nosology: a focus on the genetics and genomics of neurodevelopmental disorders.* Int J Epidemiol, 2014. **43**(2): p. 465-75.

5. Pettersson, E., et al., *Different neurodevelopmental symptoms have a common genetic etiology.* J Child Psychol Psychiatry, 2013. **54**(12): p. 1356-65.

6. Moeschler, J.B., *Genetic evaluation of intellectual disabilities.* Semin Pediatr Neurol, 2008. **15**(1): p. 2-9.

7. *World Health Organization. The world health report : report of the Director-General. World Health Organization, Geneva, pp. v.* 1995.

8. Drews, C.D., et al., *Variation in the influence of selected sociodemographic risk factors for mental retardation.* Am J Public Health, 1995. **85**(3): p. 329-34.

9. 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.

10. 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.

11. *The Centers for Disease Control and Prevention (CDC). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. 63(SS02);1-21.*

12. Taylor, B., H. Jick, and D. Maclaughlin, *Prevalence and incidence rates of autism in the UK: time trend from 2004-2010 in children aged 8 years.* BMJ Open, 2013. **3**(10): p. e003219.

13. Oeseburg, B., et al., *Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review.* Intellect Dev Disabil, 2011. **49**(2): p. 59-85.

14. Schwartz, C.E. and G. Neri, *Autism and intellectual disability: two sides of the same coin.* Am J Med Genet C Semin Med Genet, 2012. **160C**(2): p. 89-90.

15. Fombonne, E., *Epidemiological surveys of autism and other pervasive developmental disorders: an update.* J Autism Dev Disord, 2003. **33**(4): p. 365-82.

16. Willemsen, M.H. and T. Kleefstra, *Making headway with genetic diagnostics of intellectual disabilities.* Clin Genet, 2014. **85**(2): p. 101-10.

17. Persico, A.M. and V. Napolioni, *Autism genetics.* Behav Brain Res, 2013. **251**: p. 95-112.

18. Bagni, C. and W.T. Greenough, *From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome.* Nat Rev Neurosci, 2005. **6**(5): p. 376-87.

19. Napolioni, V., R. Moavero, and P. Curatolo, *Recent advances in neurobiology of Tuberous Sclerosis Complex.* Brain Dev, 2009. **31**(2): p. 104-13.

20. 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.

21. Stevenson, R.E., et al., *Genetic syndromes among individuals with mental retardation.* Am J Med Genet A, 2003. **123A**(1): p. 29-32.

22. 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.

23. Van Buggenhout, G.J., et al., *Dysmorphology and mental retardation: molecular cytogenetic studies in dysmorphic mentally retarded patients.* Ann Genet, 2001. **44**(2): p. 89-92.

24. Hagerman, R., G. Hoem, and P. Hagerman, *Fragile X and autism: Intertwined at the molecular level leading to targeted treatments.* Mol Autism, 2010. **1**(1): p. 12.

25. Asato, M.R. and A.Y. Hardan, *Neuropsychiatric problems in tuberous sclerosis complex.* J Child Neurol, 2004. **19**(4): p. 241-9.

26. Gutierrez, G.C., S.L. Smalley, and P.E. Tanguay, *Autism in tuberous sclerosis complex.* J Autism Dev Disord, 1998. **28**(2): p. 97-103.

27. de Vries, P., et al., *Consensus clinical guidelines for the assessment of cognitive and behavioural problems in Tuberous Sclerosis.* Eur Child Adolesc Psychiatry, 2005. **14**(4): p. 183-90.

28. Krueger, D.A., H. Northrup, and G. International Tuberous Sclerosis Complex Consensus, *Tuberous sclerosis complex surveillance and management: recommendations of the 2012 International Tuberous Sclerosis Complex Consensus Conference.* Pediatr Neurol, 2013. **49**(4): p. 255-65.

29. Prather, P. and P.J. de Vries, *Behavioral and cognitive aspects of tuberous sclerosis complex.* J Child Neurol, 2004. **19**(9): p. 666-74.

30. Williams, P.G. and J.H. Hersh, *Brief report: the association of neurofibromatosis type 1 and autism.* J Autism Dev Disord, 1998. **28**(6): p. 567-71.

31. Superti-Furga, A., et al., *Maternal phenylketonuria syndrome in cousins caused by mild, unrecognized phenylketonuria in their mothers homozygous for the phenylalanine hydroxylase Arg-261-Gln mutation.* Eur J Pediatr, 1991. **150**(7): p. 493-7.

32. Baieli, S., et al., *Autism and phenylketonuria.* J Autism Dev Disord, 2003. **33**(2): p. 201-4.

33. Kivitie-Kallio, S. and R. Norio, *Cohen syndrome: essential features, natural history, and heterogeneity.* Am J Med Genet, 2001. **102**(2): p. 125-35.

34. Howlin, P., J. Karpf, and J. Turk, *Behavioural characteristics and autistic features in individuals with Cohen Syndrome.* Eur Child Adolesc Psychiatry, 2005. **14**(2): p. 57-64.

35. Freitag, C.M., et al., *Quantitative assessment of neuromotor function in adolescents with high functioning autism and Asperger Syndrome.* J Autism Dev Disord, 2007. **37**(5): p. 948-59.

36. Patel, S.V., et al., *Clinical manifestations of trisomy 4p syndrome.* Eur J Pediatr, 1995. **154**(6): p. 425-31.

37. Battaglia, A., et al., *Wolf-Hirschhorn syndrome (WHS): a history in pictures.* Clin Dysmorphol, 2000. **9**(1): p. 25-30.

38. Schonewolf-Greulich, B., et al., *Segregation of a 4p16.3 duplication with a characteristic appearance, macrocephaly, speech delay and mild intellectual disability in a 3-generation family.* Am J Med Genet A, 2013. **161A**(9): p. 2358-62.

39. Butler, M.G., *Prader-Willi syndrome: current understanding of cause and diagnosis.* Am J Med Genet, 1990. **35**(3): p. 319-32.

40. Fryns, J.P., et al., *Angelman's syndrome and 15q11-13 deletions.* J Med Genet, 1989. **26**(8): p. 538.

41. Freitag, C.M., *The genetics of autistic disorders and its clinical relevance: a review of the literature.* Mol Psychiatry, 2007. **12**(1): p. 2-22.

42. Castorina, P., et al., *Genotype-phenotype correlation in two sets of monozygotic twins with Williams syndrome.* Am J Med Genet, 1997. **69**(1): p. 107-11.

43. Knoll, J.H., et al., *FISH ordering of reference markers and of the gene for the alpha 5 subunit of the gamma-aminobutyric acid receptor (GABRA5) within the Angelman and Prader-Willi syndrome chromosomal regions.* Hum Mol Genet, 1993. **2**(2): p. 183-9.

44. 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.

45. Novelli, A., et al., *Diagnosis of DiGeorge and Williams syndromes using FISH analysis of peripheral blood smears.* Mol Cell Probes, 1999. **13**(4): p. 303-7.

46. Vissers, L.E., C. Gilissen, and J.A. Veltman, *Genetic studies in intellectual disability and related disorders.* Nat Rev Genet, 2016. **17**(1): p. 9-18.

47. Brown, P.O. and D. Botstein, *Exploring the new world of the genome with DNA microarrays.* Nat Genet, 1999. **21**(1 Suppl): p. 33-7.

48. Pollack, J.R., et al., *Genome-wide analysis of DNA copy-number changes using cDNA microarrays.* Nat Genet, 1999. **23**(1): p. 41-6.

49. Snijders, A.M., et al., *Assembly of microarrays for genome-wide measurement of DNA copy number.* Nat Genet, 2001. **29**(3): p. 263-4.

50. 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.

51. Girirajan, S., C.D. Campbell, and E.E. Eichler, *Human copy number variation and complex genetic disease.* Annu Rev Genet, 2011. **45**: p. 203-26.

52. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.

53. Chen, W., et al., *Copy number variation across European populations.* PLoS One, 2011. **6**(8): p. e23087.

54. 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.

55. Stankiewicz, P., et al., *Serial segmental duplications during primate evolution result in complex human genome architecture.* Genome Res, 2004. **14**(11): p. 2209-20.

56. Bailey, J.A. and E.E. Eichler, *Primate segmental duplications: crucibles of evolution, diversity and disease.* Nat Rev Genet, 2006. **7**(7): p. 552-64.

57. Dumas, L., et al., *Gene copy number variation spanning 60 million years of human and primate evolution.* Genome Res, 2007. **17**(9): p. 1266-77.

58. Hastings, P.J., et al., *Mechanisms of change in gene copy number.* Nat Rev Genet, 2009. **10**(8): p. 551-64.

59. 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.

60. Girirajan, S., et al., *Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder.* Am J Hum Genet, 2013. **92**(2): p. 221-37.

61. Kidd, J.M., et al., *A human genome structural variation sequencing resource reveals insights into mutational mechanisms.* Cell, 2010. **143**(5): p. 837-47.

62. 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.

63. Kearney, H.M., et al., *American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants.* Genet Med, 2011. **13**(7): p. 680-5.

64. Kaminsky, E.B., et al., *An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities.* Genet Med, 2011. **13**(9): p. 777-84.

65. Chelly, J., et al., *Genetics and pathophysiology of mental retardation.* Eur J Hum Genet, 2006. **14**(6): p. 701-13.

66. Niihori, T., et al., *Germline KRAS and BRAF mutations in cardio-facio-cutaneous syndrome.* Nat Genet, 2006. **38**(3): p. 294-6.

67. Cooper, G.M., et al., *A copy number variation morbidity map of developmental delay.* Nat Genet, 2011. **43**(9): p. 838-46.

68. Vissers, L.E., B.B. de Vries, and J.A. Veltman, *Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis.* J Med Genet, 2010. **47**(5): p. 289-97.

69. Wagner, A., *Selection and gene duplication: a view from the genome.* Genome Biol, 2002. **3**(5): p. reviews1012.

70. 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.

71. 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.

72. Rivera-Brugues, N., et al., *Cohen syndrome diagnosis using whole genome arrays.* J Med Genet, 2011. **48**(2): p. 136-40.

73. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.

74. Walsh, T., et al., *Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia.* Science, 2008. **320**(5875): p. 539-43.

75. Stefansson, H., et al., *Large recurrent microdeletions associated with schizophrenia.* Nature, 2008. **455**(7210): p. 232-6.

76. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

77. McCarroll, S.A., et al., *Common deletion polymorphisms in the human genome.* Nat Genet, 2006. **38**(1): p. 86-92.

78. Conrad, D.F., et al., *A high-resolution survey of deletion polymorphism in the human genome.* Nat Genet, 2006. **38**(1): p. 75-81.

79. Girirajan, S., et al., *Phenotypic heterogeneity of genomic disorders and rare copy-number variants.* N Engl J Med, 2012. **367**(14): p. 1321-31.

80. Sanders, S.J., et al., *Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism.* Neuron, 2011. **70**(5): p. 863-85.

81. Levy, D., et al., *Rare de novo and transmitted copy-number variation in autistic spectrum disorders.* Neuron, 2011. **70**(5): p. 886-97.

82. Morrow, E.M., *Genomic copy number variation in disorders of cognitive development.* J Am Acad Child Adolesc Psychiatry, 2010. **49**(11): p. 1091-104.

83. Turner, D.J., et al., *Germline rates of de novo meiotic deletions and duplications causing several genomic disorders.* Nat Genet, 2008. **40**(1): p. 90-5.

84. Koolen, D.A., et al., *Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome.* Nat Genet, 2012. **44**(6): p. 639-41.

85. Anderlid, B.M., et al., *FISH-mapping of a 100-kb terminal 22q13 deletion.* Hum Genet, 2002. **110**(5): p. 439-43.

86. Wilson, H.L., et al., *Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms.* J Med Genet, 2003. **40**(8): p. 575-84.

87. Slager, R.E., et al., *Mutations in RAI1 associated with Smith-Magenis syndrome.* Nat Genet, 2003. **33**(4): p. 466-8.

88. 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.

89. 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.

90. Merla, G., et al., *Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes.* Am J Hum Genet, 2006. **79**(2): p. 332-41.

91. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science, 2007. **315**(5813): p. 848-53.

92. 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.

93. Henrichsen, C.N., et al., *Segmental copy number variation shapes tissue transcriptomes.* Nat Genet, 2009. **41**(4): p. 424-9.

94. Jacquemont, S., et al., *Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus.* Nature, 2011. **478**(7367): p. 97-102.

95. Walters, R.G., et al., *A new highly penetrant form of obesity due to deletions on chromosome 16p11.2.* Nature, 2010. **463**(7281): p. 671-5.

96. Luo, R., et al., *Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders.* Am J Hum Genet, 2012. **91**(1): p. 38-55.

97. Dermitzakis, E.T. and B.E. Stranger, *Genetic variation in human gene expression.* Mamm Genome, 2006. **17**(6): p. 503-8.

98. Reymond, A., et al., *Side effects of genome structural changes.* Curr Opin Genet Dev, 2007. **17**(5): p. 381-6.

99. Harvard, C., et al., *Understanding the impact of 1q21.1 copy number variant.* Orphanet J Rare Dis, 2011. **6**: p. 54.

100. Ye, T., et al., *Analysis of copy number variations in brain DNA from patients with schizophrenia and other psychiatric disorders.* Biol Psychiatry, 2012. **72**(8): p. 651-4.

101. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

102. Li, H.H., et al., *Induced chromosome deletions cause hypersociability and other features of Williams-Beuren syndrome in mice.* EMBO Mol Med, 2009. **1**(1): p. 50-65.

103. Capossela, S., et al., *Growth defects and impaired cognitive-behavioral abilities in mice with knockout for Eif4h, a gene located in the mouse homolog of the Williams-Beuren syndrome critical region.* Am J Pathol, 2012. **180**(3): p. 1121-35.

104. Hoogenraad, C.C., et al., *Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice.* Nat Genet, 2002. **32**(1): p. 116-27.

105. Fejgin, K., et al., *A mouse model that recapitulates cardinal features of the 15q13.3 microdeletion syndrome including schizophrenia- and epilepsy-related alterations.* Biol Psychiatry, 2014. **76**(2): p. 128-37.

106. Horev, G., et al., *Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism.* Proc Natl Acad Sci U S A, 2011. **108**(41): p. 17076-81.

107. Portmann, T., et al., *Behavioral abnormalities and circuit defects in the basal ganglia of a mouse model of 16p11.2 deletion syndrome.* Cell Rep, 2014. **7**(4): p. 1077-92.

108. Bi, W., et al., *Rai1 deficiency in mice causes learning impairment and motor dysfunction, whereas Rai1 heterozygous mice display minimal behavioral phenotypes.* Hum Mol Genet, 2007. **16**(15): p. 1802-13.

109. Girirajan, S. and S.H. Elsea, *Distorted Mendelian transmission as a function of genetic background in Rai1-haploinsufficient mice.* Eur J Med Genet, 2009. **52**(4): p. 224-8.

110. Lindsay, E.A., et al., *Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice.* Nature, 2001. **410**(6824): p. 97-101.

111. Jerome, L.A. and V.E. Papaioannou, *DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1.* Nat Genet, 2001. **27**(3): p. 286-91.

112. Kimber, W.L., et al., *Deletion of 150 kb in the minimal DiGeorge/velocardiofacial syndrome critical region in mouse.* Hum Mol Genet, 1999. **8**(12): p. 2229-37.

113. Streisinger, G., et al., *Production of clones of homozygous diploid zebra fish (Brachydanio rerio).* Nature, 1981. **291**(5813): p. 293-6.

114. Davis, E.E., S. Frangakis, and N. Katsanis, *Interpreting human genetic variation with in vivo zebrafish assays.* Biochim Biophys Acta, 2014. **1842**(10): p. 1960-1970.

115. Dauber, A., et al., *SCRIB and PUF60 are primary drivers of the multisystemic phenotypes of the 8q24.3 copy-number variant.* Am J Hum Genet, 2013. **93**(5): p. 798-811.

116. Blaker-Lee, A., et al., *Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes.* Dis Model Mech, 2012. **5**(6): p. 834-51.

117. Golzio, C., et al., *KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant.* Nature, 2012. **485**(7398): p. 363-7.

118. Piotrowski, T., et al., *The zebrafish van gogh mutation disrupts tbx1, which is involved in the DiGeorge deletion syndrome in humans.* Development, 2003. **130**(20): p. 5043-52.

119. Zhang, L.F., et al., *Effect of Tbx1 knock-down on cardiac performance in zebrafish.* Chin Med J (Engl), 2010. **123**(9): p. 1182-9.

120. Bellen, H.J., C. Tong, and H. Tsuda, *100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future.* Nat Rev Neurosci, 2010. **11**(7): p. 514-22.

121. Reiter, L.T., et al., *A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster.* Genome Res, 2001. **11**(6): p. 1114-25.

122. Wu, Y., et al., *A Drosophila model for Angelman syndrome.* Proc Natl Acad Sci U S A, 2008. **105**(34): p. 12399-404.

123. Kramer, J.M., et al., *Epigenetic regulation of learning and memory by Drosophila EHMT/G9a.* PLoS Biol, 2011. **9**(1): p. e1000569.

124. Paciorkowski, A.R., et al., *Deletion 16p13.11 uncovers NDE1 mutations on the non-deleted homolog and extends the spectrum of severe microcephaly to include fetal brain disruption.* Am J Med Genet A, 2013. **161A**(7): p. 1523-30.

125. McDonald-McGinn, D.M., et al., *Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11.2DS.* J Med Genet, 2013. **50**(2): p. 80-90.

126. Classen, C.F., et al., *Dissecting the genotype in syndromic intellectual disability using whole exome sequencing in addition to genome-wide copy number analysis.* Hum Genet, 2013. **132**(7): p. 825-41.

127. Yang, Y., et al., *Clinical whole-exome sequencing for the diagnosis of mendelian disorders.* N Engl J Med, 2013. **369**(16): p. 1502-11.

128. Soden, S.E., et al., *Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders.* Sci Transl Med, 2014. **6**(265): p. 265ra168.

129. Ng, S.B., et al., *Targeted capture and massively parallel sequencing of 12 human exomes.* Nature, 2009. **461**(7261): p. 272-6.

130. de Ligt, J., et al., *Diagnostic exome sequencing in persons with severe intellectual disability.* N Engl J Med, 2012. **367**(20): p. 1921-9.

131. O'Roak, B.J., et al., *Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations.* Nat Genet, 2011. **43**(6): p. 585-9.

132. O'Roak, B.J., et al., *Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations.* Nature, 2012. **485**(7397): p. 246-50.

133. Rauch, A., et al., *Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study.* Lancet, 2012. **380**(9854): p. 1674-82.

134. Sanders, S.J., et al., *De novo mutations revealed by whole-exome sequencing are strongly associated with autism.* Nature, 2012. **485**(7397): p. 237-41.

135. Neale, B.M., et al., *Patterns and rates of exonic de novo mutations in autism spectrum disorders.* Nature, 2012. **485**(7397): p. 242-5.

136. O'Roak, B.J., et al., *Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders.* Science, 2012. **338**(6114): p. 1619-22.

137. Iossifov, I., et al., *De novo gene disruptions in children on the autistic spectrum.* Neuron, 2012. **74**(2): p. 285-99.

138. Novarino, G., et al., *Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy.* Science, 2012. **338**(6105): p. 394-7.

139. Puffenberger, E.G., et al., *A homozygous missense mutation in HERC2 associated with global developmental delay and autism spectrum disorder.* Hum Mutat, 2012. **33**(12): p. 1639-46.

140. Chahrour, M.H., et al., *Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism.* PLoS Genet, 2012. **8**(4): p. e1002635.

141. Nava, C., et al., *Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE.* Transl Psychiatry, 2012. **2**: p. e179.

142. Yu, T.W., et al., *Using whole-exome sequencing to identify inherited causes of autism.* Neuron, 2013. **77**(2): p. 259-73.

143. Lim, E.T., et al., *Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders.* Neuron, 2013. **77**(2): p. 235-42.

144. Manning, M., et al., *Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities.* Genet Med, 2010. **12**(11): p. 742-5.

145. Marshall, C.R., et al., *Structural variation of chromosomes in autism spectrum disorder.* Am J Hum Genet, 2008. **82**(2): p. 477-88.

146. Fernandez, B.A., et al., *Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder.* J Med Genet, 2010. **47**(3): p. 195-203.

147. Weiss, L.A., et al., *Association between microdeletion and microduplication at 16p11.2 and autism.* N Engl J Med, 2008. **358**(7): p. 667-75.

148. Kumar, R.A., et al., *Association and mutation analyses of 16p11.2 autism candidate genes.* PLoS One, 2009. **4**(2): p. e4582.

149. Rosenfeld, J.A., et al., *Estimates of penetrance for recurrent pathogenic copy-number variations.* Genet Med, 2013. **15**(6): p. 478-81.

150. Martin, J., et al., *The sequence and analysis of duplication-rich human chromosome 16.* Nature, 2004. **432**(7020): p. 988-94.

151. 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.

152. Bedoyan, J.K., et al., *Duplication 16p11.2 in a child with infantile seizure disorder.* Am J Med Genet A, 2010. **152A**(6): p. 1567-74.

153. Ghebranious, N., et al., *A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation.* Am J Med Genet A, 2007. **143A**(13): p. 1462-71.

154. Kumar, R.A., et al., *Recurrent 16p11.2 microdeletions in autism.* Hum Mol Genet, 2008. **17**(4): p. 628-38.

155. Bijlsma, E.K., et al., *Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals.* Eur J Med Genet, 2009. **52**(2-3): p. 77-87.

156. Bachmann-Gagescu, R., et al., *Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity.* Genet Med, 2010. **12**(10): p. 641-7.

157. McCarthy, S.E., et al., *Microduplications of 16p11.2 are associated with schizophrenia.* Nat Genet, 2009. **41**(11): p. 1223-7.

158. McCarthy, S.E., et al., *Microduplications of 16p11.2 are associated with schizophrenia.* Nat Genet, 2009. **41**(11): p. 1223-7.

159. Al-Kateb, H., et al., *Scoliosis and vertebral anomalies: additional abnormal phenotypes associated with chromosome 16p11.2 rearrangement.* Am J Med Genet A, 2014. **164A**(5): p. 1118-26.

160. Robinson, P.N., *Deep phenotyping for precision medicine.* Hum Mutat, 2012. **33**(5): p. 777-80.

161. Kohler, S., et al., *Ontological phenotype standards for neurogenetics.* Hum Mutat, 2012. **33**(9): p. 1333-9.

162. Dorschner, M.O., et al., *Actionable, pathogenic incidental findings in 1,000 participants' exomes.* Am J Hum Genet, 2013. **93**(4): p. 631-40.

163. Schulze, T.G. and F.J. McMahon, *Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping.* Hum Hered, 2004. **58**(3-4): p. 131-8.

164. Qiao, Y., et al., *Copy number variants (CNVs) analysis in a deeply phenotyped cohort of individuals with intellectual disability (ID).* BMC Med Genet, 2014. **15**: p. 82.

165. Frebourg, T., *The challenge for the next generation of medical geneticists.* Hum Mutat, 2014. **35**(8): p. 909-11.

166. Akawi, N., et al., *Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families.* Nat Genet, 2015. **47**(11): p. 1363-9.

167. Zufferey, F., et al., *A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders.* J Med Genet, 2012. **49**(10): p. 660-8.

168. Hanson, E., et al., *The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population.* Biol Psychiatry, 2015. **77**(9): p. 785-93.

169. Raca, G., et al., *Childhood Apraxia of Speech (CAS) in two patients with 16p11.2 microdeletion syndrome.* Eur J Hum Genet, 2013. **21**(4): p. 455-9.

170. Shiow, L.R., et al., *Severe combined immunodeficiency (SCID) and attention deficit hyperactivity disorder (ADHD) associated with a Coronin-1A mutation and a chromosome 16p11.2 deletion.* Clin Immunol, 2009. **131**(1): p. 24-30.

171. Zheng, X., et al., *A rare duplication on chromosome 16p11.2 is identified in patients with psychosis in Alzheimer's disease.* PLoS One, 2014. **9**(11): p. e111462.

172. Schaaf, C.P., et al., *Expanding the clinical spectrum of the 16p11.2 chromosomal rearrangements: three patients with syringomyelia.* Eur J Hum Genet, 2010. **19**(2): p. 152-6.

173. Sahoo, T., et al., *Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems.* Genet Med, 2011. **13**(10): p. 868-80.

174. Duyzend, M.H. and E.E. Eichler, *Genotype-first analysis of the 16p11.2 deletion defines a new type of "autism".* Biol Psychiatry, 2015. **77**(9): p. 769-71.

175. Shinawi, M., et al., *Mixed gonadal dysgenesis in a child with isodicentric Y chromosome: Does the relative proportion of the 45,X line really matter?* Am J Med Genet A, 2010. **152A**(7): p. 1832-7.

176. Reinthaler, E.M., et al., *16p11.2 600 kb Duplications confer risk for typical and atypical Rolandic epilepsy.* Hum Mol Genet, 2014. **23**(22): p. 6069-80.

177. Shimojima, K., et al., *A familial 593-kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae.* Eur J Med Genet, 2009. **52**(6): p. 433-5.

178. Filges, I., et al., *Brain MRI abnormalities and spectrum of neurological and clinical findings in three patients with proximal 16p11.2 microduplication.* Am J Med Genet A, 2014. **164A**(8): p. 2003-12.

179. Barge-Schaapveld, D.Q., et al., *The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome?* Am J Med Genet A, 2011. **155A**(5): p. 1066-72.

180. Owen, J.P., et al., *Aberrant white matter microstructure in children with 16p11.2 deletions.* J Neurosci, 2014. **34**(18): p. 6214-23.

181. Maillard, A.M., et al., *The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity.* Mol Psychiatry, 2015. **20**(1): p. 140-7.

182. Nik-Zainal, S., et al., *High incidence of recurrent copy number variants in patients with isolated and syndromic Mullerian aplasia.* J Med Genet, 2011. **48**(3): p. 197-204.

183. Bardakjian, T.M., et al., *Clinical report of microphthalmia and optic nerve coloboma associated with a de novo microdeletion of chromosome 16p11.2.* Am J Med Genet A, 2010. **152A**(12): p. 3120-3.

184. Hernando, C., et al., *Comparative genomic hybridisation shows a partial de novo deletion 16p11.2 in a neonate with multiple congenital malformations.* J Med Genet, 2002. **39**(5): p. E24.

185. Ciuladaite, Z., et al., *Mental retardation and autism associated with recurrent 16p11.2 microdeletion: incomplete penetrance and variable expressivity.* J Appl Genet, 2011. **52**(4): p. 443-9.

186. Maggadottir, S.M., et al., *Rare variants at 16p11.2 are associated with common variable immunodeficiency.* J Allergy Clin Immunol, 2015. **135**(6): p. 1569-77.

187. Tucker, T., et al., *Prevalence of selected genomic deletions and duplications in a French-Canadian population-based sample of newborns.* Mol Genet Genomic Med, 2013. **1**(2): p. 87-97.

188. Shubayeva, G.S., A.A.T. , and D.K.B. , *Comorbidities of Children with Intellectual Disabilities* Asian Social Science, 2015. **11**(16): p. 168-174.

189. Prelock, P.A., H. , T., and G. , F. P., *Speech-Language Impairment: How to Identify the Most Common and Least Diagnosed Disability of Childhood.* Medscape J Med., 2008. **10**(6): p. 136.

190. O'Brien, G., *Dual diagnosis in offenders with intellectual disability: setting research priorities: a review of research findings concerning psychiatric disorder (excluding personality disorder) among offenders with intellectual disability.* J Intellect Disabil Res, 2002. **46 Suppl 1**: p. 21-30.

191. Bhaumik, S., et al., *Psychiatric service use and psychiatric disorders in adults with intellectual disability.* J Intellect Disabil Res, 2008. **52**(11): p. 986-95.

192. Mehta, D., et al., *Comprehensive survey of CNVs influencing gene expression in the human brain and its implications for pathophysiology.* Neurosci Res, 2014. **79**: p. 22-33.

193. Blumenthal, I., et al., *Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families.* Am J Hum Genet, 2014. **94**(6): p. 870-83.

194. Migliavacca, E., et al., *A Potential Contributory Role for Ciliary Dysfunction in the 16p11.2 600 kb BP4-BP5 Pathology.* Am J Hum Genet, 2015. **96**(5): p. 784-96.

195. Tsien, J.Z., P.T. Huerta, and S. Tonegawa, *The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory.* Cell, 1996. **87**(7): p. 1327-38.

196. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

197. Gross, C. and G.J. Bassell, *Neuron-specific regulation of class I PI3K catalytic subunits and their dysfunction in brain disorders.* Front Mol Neurosci, 2014. **7**: p. 12.

198. Choi, S., et al., *ARF6 and EFA6A regulate the development and maintenance of dendritic spines.* J Neurosci, 2006. **26**(18): p. 4811-9.

199. Iwakura, Y. and H. Nawa, *ErbB1-4-dependent EGF/neuregulin signals and their cross talk in the central nervous system: pathological implications in schizophrenia and Parkinson's disease.* Front Cell Neurosci, 2013. **7**: p. 4.

200. Takei, N. and H. Nawa, *mTOR signaling and its roles in normal and abnormal brain development.* Front Mol Neurosci, 2014. **7**: p. 28.

201. Jen, Y.H., M. Musacchio, and A.D. Lander, *Glypican-1 controls brain size through regulation of fibroblast growth factor signaling in early neurogenesis.* Neural Dev, 2009. **4**: p. 33.

202. Hu, V.W., et al., *Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes.* BMC Genomics, 2006. **7**: p. 118.

203. Cai, C., et al., *Is human blood a good surrogate for brain tissue in transcriptional studies?* BMC Genomics, 2010. **11**: p. 589.

204. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-93.

205. van der Spek, P.J., et al., *Are gene expression microarray analyses reliable? A review of studies of retinoic acid responsive genes.* Genomics Proteomics Bioinformatics, 2003. **1**(1): p. 9-14.

206. Futcher, B., et al., *A sampling of the yeast proteome.* Mol Cell Biol, 1999. **19**(11): p. 7357-68.

207. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-41.

208. Maier, T., M. Guell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples.* FEBS Lett, 2009. **583**(24): p. 3966-73.

209. Wu, G., L. Nie, and W. Zhang, *Integrative analyses of posttranscriptional regulation in the yeast Saccharomyces cerevisiae using transcriptomic and proteomic data.* Curr Microbiol, 2008. **57**(1): p. 18-22.

210. Gygi, S.P., et al., *Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.* Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9390-5.

211. Gorg, A., W. Weiss, and M.J. Dunn, *Current two-dimensional electrophoresis technology for proteomics.* Proteomics, 2004. **4**(12): p. 3665-85.

212. Hong, L.E., et al., *Specific motion processing pathway deficit during eye tracking in schizophrenia: a performance-matched functional magnetic resonance imaging study.* Biol Psychiatry, 2005. **57**(7): p. 726-32.

213. Takarae, Y., et al., *Pursuit eye movement deficits in autism.* Brain, 2004. **127**(Pt 12): p. 2584-94.

214. Jiang, Y.H., et al., *Genetic diagnosis of autism spectrum disorders: the opportunity and challenge in the genomics era.* Crit Rev Clin Lab Sci, 2014. **51**(5): p. 249-62.

215. Isakov, O., M. Perrone, and N. Shomron, *Exome sequencing analysis: a guide to disease variant detection.* Methods Mol Biol, 2013. **1038**: p. 137-58.

216. Kolehmainen, J., et al., *Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport.* Am J Hum Genet, 2003. **72**(6): p. 1359-69.

217. El Chehadeh-Djebbar, S., et al., *Changing facial phenotype in Cohen syndrome: towards clues for an earlier diagnosis.* Eur J Hum Genet, 2013. **21**(7): p. 736-42.

218. Zhang, Q., et al., *Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity.* Hum Mol Genet, 2007. **16**(23): p. 2816-33.

219. Zhang, X., et al., *SUN1/2 and Syne/Nesprin-1/2 complexes connect centrosome to the nucleus during neurogenesis and neuronal migration in mice.* Neuron, 2009. **64**(2): p. 173-87.

220. de Brouwer, A.P., et al., *A 3-base pair deletion, c.9711_9713del, in DMD results in intellectual disability without muscular dystrophy.* Eur J Hum Genet, 2014. **22**(4): p. 480-5.

221. Pane, M., et al., *Attention deficit hyperactivity disorder and cognitive function in Duchenne muscular dystrophy: phenotype-genotype correlation.* J Pediatr, 2012. **161**(4): p. 705-9 e1.

222. Horn, D., et al., *Homozygosity mapping in a family with microcephaly, mental retardation, and short stature to a Cohen syndrome region on 8q21.3-8q22.1: redefining a clinical entity.* Am J Med Genet, 2000. **92**(4): p. 285-92.

223. Mochida, G.H., et al., *Broader geographical spectrum of Cohen syndrome due to COH1 mutations.* J Med Genet, 2004. **41**(6): p. e87.

224. Seifert, W., et al., *Cohen syndrome-associated protein, COH1, is a novel, giant Golgi matrix protein required for Golgi integrity.* J Biol Chem, 2011. **286**(43): p. 37665-75.

225. Seifert, W., et al., *Cohen syndrome-associated protein COH1 physically and functionally interacts with the small GTPase RAB6 at the Golgi complex and directs neurite outgrowth.* J Biol Chem, 2015. **290**(6): p. 3349-58.

226. Duplomb, L., et al., *Cohen syndrome is associated with major glycosylation defects.* Hum Mol Genet, 2014. **23**(9): p. 2391-9.

227. Seifert, W., et al., *Expanded mutational spectrum in Cohen syndrome, tissue expression, and transcript variants of COH1.* Hum Mutat, 2009. **30**(2): p. E404-20.

228. Hurmerinta, K., et al., *Craniofacial features in Cohen syndrome: an anthropometric and cephalometric analysis of 14 patients.* Clin Genet, 2002. **62**(2): p. 157-64.

229. Garcia-Ballesta, C., et al., *New oral findings in Cohen syndrome.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2003. **95**(6): p. 681-7.

230. Chandler, K.E., et al., *Diagnostic criteria, clinical characteristics, and natural history of Cohen syndrome.* J Med Genet, 2003. **40**(4): p. 233-41.

231. Hennies, H.C., et al., *Allelic heterogeneity in the COH1 gene explains clinical variability in Cohen syndrome.* Am J Hum Genet, 2004. **75**(1): p. 138-45.

232. Kivitie-Kallio, S., et al., *Neurological and psychological findings in patients with Cohen syndrome: a study of 18 patients aged 11 months to 57 years.* Neuropediatrics, 1999. **30**(4): p. 181-9.

233. Fryns, J.P., et al., *Cohen syndrome: the clinical symptoms and stigmata at a young age.* Clin Genet, 1996. **49**(5): p. 237-41.

234. Falk, M.J., et al., *Cohen syndrome in the Ohio Amish.* Am J Med Genet A, 2004. **128A**(1): p. 23-8.

235. Katzaki, E., et al., *Clinical and molecular characterization of Italian patients affected by Cohen syndrome.* J Hum Genet, 2007. **52**(12): p. 1011-7.

236. Chen, X., et al., *Intronic alterations in BRCA1 and BRCA2: effect on mRNA splicing fidelity and expression.* Hum Mutat, 2006. **27**(5): p. 427-35.

237. Wang, G.S. and T.A. Cooper, *Splicing in disease: disruption of the splicing code and the decoding machinery.* Nat Rev Genet, 2007. **8**(10): p. 749-61.

238. Zhen, Y.Y., et al., *NUANCE, a giant protein connecting the nucleus and actin cytoskeleton.* J Cell Sci, 2002. **115**(Pt 15): p. 3207-22.

239. Zhang, Q., et al., *Nesprins: a novel family of spectrin-repeat-containing proteins that localize to the nuclear membrane in multiple tissues.* J Cell Sci, 2001. **114**(Pt 24): p. 4485-98.

240. Wilhelmsen, K., et al., *Nesprin-3, a novel outer nuclear membrane protein, associates with the cytoskeletal linker protein plectin.* J Cell Biol, 2005. **171**(5): p. 799-810.

241. Roux, K.J., et al., *Nesprin 4 is an outer nuclear membrane protein that can induce kinesin-mediated cell polarization.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2194-9.

242. Starr, D.A. and H.N. Fridolfsson, *Interactions between nuclei and the cytoskeleton are mediated by SUN-KASH nuclear-envelope bridges.* Annu Rev Cell Dev Biol, 2010. **26**: p. 421-44.

243. Schneider, M., et al., *Molecular mechanisms of centrosome and cytoskeleton anchorage at the nuclear envelope.* Cell Mol Life Sci, 2011. **68**(9): p. 1593-610.

244. Cottrell, J.R., et al., *CPG2: a brain- and synapse-specific protein that regulates the endocytosis of glutamate receptors.* Neuron, 2004. **44**(4): p. 677-90.

245. Kobayashi, Y., et al., *Identification and characterization of GSRP-56, a novel Golgi-localized spectrin repeat-containing protein.* Exp Cell Res, 2006. **312**(16): p. 3152-64.

246. Zhang, X., et al., *Syne-1 and Syne-2 play crucial roles in myonuclear anchorage and motor neuron innervation.* Development, 2007. **134**(5): p. 901-8.

247. Luke, Y., et al., *Nesprin-2 Giant (NUANCE) maintains nuclear envelope architecture and composition in skin.* J Cell Sci, 2008. **121**(Pt 11): p. 1887-98.

248. Stewart-Hutchinson, P.J., et al., *Structural requirements for the assembly of LINC complexes and their function in cellular mechanical stiffness.* Exp Cell Res, 2008. **314**(8): p. 1892-905.

249. Dawe, H.R., et al., *Nesprin-2 interacts with meckelin and mediates ciliogenesis via remodelling of the actin cytoskeleton.* J Cell Sci, 2009. **122**(Pt 15): p. 2716-26.

250. Coffinier, C., L.G. Fong, and S.G. Young, *LINCing lamin B2 to neuronal migration: growing evidence for cell-specific roles of B-type lamins.* Nucleus, 2010. **1**(5): p. 407-11.

251. Meinke, P., T.D. Nguyen, and M.S. Wehnert, *The LINC complex and human disease.* Biochem Soc Trans, 2011. **39**(6): p. 1693-7.

252. Crisp, M., et al., *Coupling of the nucleus and cytoplasm: role of the LINC complex.* J Cell Biol, 2006. **172**(1): p. 41-53.

253. Lombardi, M.L., et al., *The interaction between nesprins and sun proteins at the nuclear envelope is critical for force transmission between the nucleus and cytoskeleton.* J Biol Chem, 2011. **286**(30): p. 26743-53.

254. Mellad, J.A., D.T. Warren, and C.M. Shanahan, *Nesprins LINC the nucleus and cytoskeleton.* Curr Opin Cell Biol, 2011. **23**(1): p. 47-54.

255. Starr, D.A. and J.A. Fischer, *KASH 'n Karry: the KASH domain family of cargo-specific cytoskeletal adaptor proteins.* Bioessays, 2005. **27**(11): p. 1136-46.

256. Wilhelmsen, K., et al., *KASH-domain proteins in nuclear migration, anchorage and other processes.* J Cell Sci, 2006. **119**(Pt 24): p. 5021-9.

257. Autore, F., et al., *Large-scale modelling of the divergent spectrin repeats in nesprins: giant modular proteins.* PLoS One, 2013. **8**(5): p. e63633.

258. Zhang, Q., et al., *Nesprin-2 is a multi-isomeric protein that binds lamin and emerin at the nuclear envelope and forms a subcellular network in skeletal muscle.* J Cell Sci, 2005. **118**(Pt 4): p. 673-87.

259. Gough, L.L., et al., *Golgi localization of Syne-1.* Mol Biol Cell, 2003. **14**(6): p. 2410-24.

260. Fan, J. and K.A. Beck, *A role for the spectrin superfamily member Syne-1 and kinesin II in cytokinesis.* J Cell Sci, 2004. **117**(Pt 4): p. 619-29.

261. Warren, D.T., et al., *Nesprins: intracellular scaffolds that maintain cell architecture and coordinate cell function?* Expert Rev Mol Med, 2005. **7**(11): p. 1-15.

262. Smith, U.M., et al., *The transmembrane protein meckelin (MKS3) is mutated in Meckel-Gruber syndrome and the wpk rat.* Nat Genet, 2006. **38**(2): p. 191-6.

263. Baala, L., et al., *The Meckel-Gruber syndrome gene, MKS3, is mutated in Joubert syndrome.* Am J Hum Genet, 2007. **80**(1): p. 186-94.

264. Gorivodsky, M., et al., *Intraflagellar transport protein 172 is essential for primary cilia formation and plays a vital role in patterning the mammalian brain.* Dev Biol, 2009. **325**(1): p. 24-32.

265. Han, Y.G. and A. Alvarez-Buylla, *Role of primary cilia in brain development and cancer.* Curr Opin Neurobiol, 2010. **20**(1): p. 58-67.

266. Warren, D.T., et al., *Nesprin-2-dependent ERK1/2 compartmentalisation regulates the DNA damage response in vascular smooth muscle cell ageing.* Cell Death Differ, 2015. **22**(9): p. 1540-50.

267. Neumann, S., et al., *Nesprin-2 interacts with {alpha}-catenin and regulates Wnt signaling at the nuclear envelope.* J Biol Chem, 2010. **285**(45): p. 34932-8.

268. Puckelwartz, M.J., et al., *Nesprin-1 mutations in human and murine cardiomyopathy.* J Mol Cell Cardiol, 2010. **48**(4): p. 600-8.

269. Rashmi, R.N., et al., *The nuclear envelope protein Nesprin-2 has roles in cell proliferation and differentiation during wound healing.* Nucleus, 2012. **3**(2): p. 172-86.

270. Lu, W., et al., *Nesprin interchain associations control nuclear size.* Cell Mol Life Sci, 2012. **69**(20): p. 3493-509.

271. De Matteis, M.A. and J.S. Morrow, *Spectrin tethers and mesh in the biosynthetic pathway.* J Cell Sci, 2000. **113 (Pt 13)**: p. 2331-43.

272. Blake, D.J., et al., *Function and genetics of dystrophin and dystrophin-related proteins in muscle.* Physiol Rev, 2002. **82**(2): p. 291-329.

273. Nowak, K.J. and K.E. Davies, *Duchenne muscular dystrophy and dystrophin: pathogenesis and opportunities for treatment.* EMBO Rep, 2004. **5**(9): p. 872-6.

274. Zhou, G.Q., et al., *Current understanding of dystrophin-related muscular dystrophy and therapeutic challenges ahead.* Chin Med J (Engl), 2006. **119**(16): p. 1381-91.

275. Moizard, M.P., et al., *Are Dp71 and Dp140 brain dystrophin isoforms related to cognitive impairment in Duchenne muscular dystrophy?* Am J Med Genet, 1998. **80**(1): p. 32-41.

276. Daoud, F., et al., *Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression.* Hum Mol Genet, 2009. **18**(20): p. 3779-94.

277. Hinton, V.J., et al., *Social behavior problems in boys with Duchenne muscular dystrophy.* J Dev Behav Pediatr, 2006. **27**(6): p. 470-6.

278. Poysky, J. and D.M.D.S.G. Behavior in, *Behavior patterns in Duchenne muscular dystrophy: report on the Parent Project Muscular Dystrophy behavior workshop 8-9 of December 2006, Philadelphia, USA.* Neuromuscul Disord, 2007. **17**(11-12): p. 986-94.

279. Steele, M., et al., *Mental health of children and adolescents with Duchenne muscular dystrophy.* Dev Med Child Neurol, 2008. **50**(8): p. 638-9.

280. Rappley, M.D., *Clinical practice. Attention deficit-hyperactivity disorder.* N Engl J Med, 2005. **352**(2): p. 165-73.

281. Polanczyk, G., et al., *The worldwide prevalence of ADHD: a systematic review and metaregression analysis.* Am J Psychiatry, 2007. **164**(6): p. 942-8.

282. Rajgor, D., et al., *Multiple novel nesprin-1 and nesprin-2 variants act as versatile tissue-specific intracellular scaffolds.* PLoS One, 2012. **7**(7): p. e40098.

283. Iglesias, A., et al., *The usefulness of whole-exome sequencing in routine clinical practice.* Genet Med, 2014.

284. Douzgou, S. and M.B. Petersen, *Clinical variability of genetic isolates of Cohen syndrome.* Clin Genet, 2011. **79**(6): p. 501-6.

285. Wang, Y., et al., *Polyamine-modulated factor 1 binds to the human homologue of the 7a subunit of the Arabidopsis COP9 signalosome: implications in gene expression.* Biochem J, 2002. **366**(Pt 1): p. 79-86.

286. Carter, K.L., E. Cahir-McFarland, and E. Kieff, *Epstein-barr virus-induced changes in B-lymphocyte gene expression.* J Virol, 2002. **76**(20): p. 10427-36.

287. Caliskan, M., et al., *The effects of EBV transformation on gene expression levels and methylation profiles.* Hum Mol Genet, 2011. **20**(8): p. 1643-52.

288. Duong, N.T., et al., *Nesprins: tissue-specific expression of epsilon and other short isoforms.* PLoS One, 2014. **9**(4): p. e94380.

289. Tipton, J.D., et al., *Analysis of intact protein isoforms by mass spectrometry.* J Biol Chem, 2011. **286**(29): p. 25451-8.

290. Martinez, R.A., et al., *Genome engineering of isogenic human ES cells to model autism disorders.* Nucleic Acids Res, 2015. **43**(10): p. e65.

291. Heilker, R., et al., *iPS cell derived neuronal cells for drug discovery.* Trends Pharmacol Sci, 2014. **35**(10): p. 510-9.