Identification of known and new candidate molecular causes of inherited neurodevelopmental disorders

Submitted by Stanya Gardner to the University of Exeter

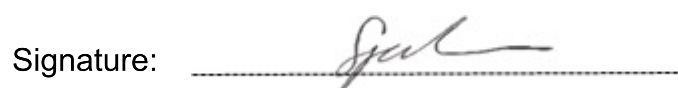
as a thesis for the degree of

Masters by Research in Medical Studies

In August 2018

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: 

ACKNOWLEDGEMENTS

I would like to thank the families within this study, as well as our local clinical collaborators, whose recruitment and support to families has facilitated this collaboration with the University of Exeter.

I am very grateful for all the guidance and advice provided by my supervisors Professor Andrew Crosby and Dr Emma Baple. Particular thanks must go to Dr Barry Chioza and Dr Guarav Harlalka, as well as all the other members of the Crosby/Baple Community Genetics Group, who have been so helpful and patient and made this year thoroughly enjoyable.

I am enormously grateful for the constant support and motivation from my family and boyfriend, who have encouraged me on all my endeavours throughout this year, and without whom this thesis would not have come to fruition.

Crosby/Baple Community Genetics Group:
Ilaria D'Atri, James Fasham, Beth Hilditch, Hannah Jones, Simon Lam, Serene Lin, Seun Olusanya, Lettie Rawlins, Olivia Rickman, Shamim Saleha, Fatema Al Salmi and Claire Salter.

ABSTRACT

Intellectual disability (ID) is a rare neurodevelopmental disorder estimated to affect at least 1% of the population. ID is defined as “a disorder with onset during the developmental period that includes both intellectual and adaptive functioning deficits in conceptual, social and practical domains. ID is a complex and highly variable disorder, often associated with a number of other conditions such as microcephaly and neurological impairment. Identifying genes responsible for ID will not only aid diagnosis and genetic counselling efforts, but also elevate our understanding of the complex mechanisms and genes crucial for normal brain development.

Consanguineous populations are often enriched for autosomal recessive disorders such as ID. With this in mind, an investigative strategy was designed to identify the underlying genetic cause of ID within four families identified through an ongoing translational community research program based at UEMS. Each family has at least two individuals affected with varying ID phenotypes. Single nucleotide polymorphism mapping and whole exome sequencing was used to identify potential candidate pathogenic variants within shared autozygous regions and genome wide.

A putative diagnosis has been made within two of the families, with one family harbouring a known variant within *ALG1*; which is known to cause a rare autosomal recessive congenital disorder of glycosylation, and the other family harbouring a novel variant within an *ARHGAP* gene; which has recently been implicated in neurodevelopmental disorders. The other two families require further research, however potential novel variants as well as chromosomal

variations have been identified. Through this collaborative study we have shown how important community-based research programmes are in defining the genetic causes of rare disorders.

CONTENTS

List of Tables 8

List of Figures 9

1. INTRODUCTION 11

1.1 Classification of Intellectual Disability 11

1.2 The Cost of Intellectual Disability 14

1.3 Microcephaly 16

1.4 The Genetics of Intellectual Disability and Microcephaly 18

1.5 The Role of Community-Based Programs to Define Genetic Causes of ID 21

1.6 Aims 24

2. SUBJECTS AND METHODS 25

2.1 Single nucleotide polymorphism (SNP) genotyping 25

2.2 Primer Design 25

2.3 Polymerase Chain Reaction (PCR) 26

2.4 Primer optimisation 27

2.5 Agarose gel electrophoresis 27

2.6 EXOSAP Purification of PCR product 29

2.7 DNA Sequencing 29

3. RESULTS 30

3.1 Sequence alteration of *ALG1* associated with a congenital disorder of glycosylation 30

3.2 Defining mutations within an ARHGAP molecule as a likely new cause of ID 37

3.3 Exome sequencing identifies new candidate genetic causes of ID in an extended family with individuals with this disorder 42

3.4 Further genetic studies are required to determine the cause of ID in an extended family with multiple individuals with this disorder 49

4. DISCUSSION 54

5. FUTURE WORK 67

6. CONCLUSION 69

7. APPENDICIES 71

7.1 Primer Sequences 71

8. BIBLIOGRAPHY 72

List of Tables

Table 2-1 Components of standard 10 µl PCR tube 27

Table 2-2. 50x LAB buffer solution composition 28

Table 3-1 Largest regions of autozygosity common to affected members in Family 2 32

Table 3-2 Candidate/ Primary ID genes identified within the largest region of autozygosity from SysID for Family 1 33

Table 3-3 Novel and rare homozygous variants identified from whole exome sequencing of VI:3 35

Table 3-4 Largest regions of autozygosity common to affected members in Family 2 38

Table 3-5. Novel and rare homozygous variants identified from whole exome sequencing of VI:1 40

Table 3-6 Largest regions of autozygosity common to affected members in Family 3 44

Table 3-7 Candidate/ Primary ID genes identified within the largest region of autozygosity from SysID for Family 3 46

Table 3-8 Novel and rare homozygous variants identified from whole exome sequencing of VI:1 47

Table 3-9 Novel and rare compound heterozygous variants identified from whole exome sequencing of VI:1 47

Table 3-10. Largest regions of autozygosity common to affected members of Family 4 51

Table 3-11. Novel and rare homozygous variants identified from whole exome sequencing of VI:8 53

Table 3-12. Novel and rare compound heterozygous variants identified from whole exome sequencing of VI:8 53

List of Figures

Figure 1-1. DSM-V classification of neurodevelopmental disorders (APA 2013) 12

Figure 1-2 Typical Head Size, Microcephaly and Severe Microcephaly Comparison. Taken from (Petersen et al. 2016)... 17

Figure 3-1. Family 1 Pedigree 30

Figure 3-2. Visual depiction of largest autozygous regions of the genome (shown in red) common to all three affected siblings in Family 1 31

Figure 3-3 *Figure 3-3 Sequencing chromatogram from a parental heterozygous carrier for the ALG1 c.826C>T gene variant, showing both the alternate (T) and wildtype (C) alleles* 36

Figure 3-4 Family 2 Pedigree 37

Figure 3-5. Visual depiction of largest autozygous regions of the genome (shown in red) common to affected siblings in Family 2 38

Figure 3-6 Sequencing chromatograms for ARHGAP c.1673C>T gene variant41

Figure 3-7 Family 3 pedigree 42

Figure 3-9 Visual depiction of largest autozygous regions of the genome (shown in red) common to all three affected siblings in Family 3 43

Figure 3-8 Karyotyping of three affected individuals within Family 3 to identify shared CNV's (highlighted by red box) 43

Figure 3-10 Family 4 pedigree 49

Figure 3-11 Karyotyping of three affected individuals within Family 4 to identify shared CNV's (highlighted by red box) 50

Figure 3-12 Visual depiction of largest autozygous regions of the genome (shown in red) common to affected siblings in Family 4 51
Figure 4-1 Expression of ARHGAP as per UCSC Genome Browser (Website: https://genome.ucsc.edu/cgi-bin/hgGateway)..58
Figure 4-2 ARHGAP interaction as per STRING version 10.5...........................59
Figure 4-3 Visual representation of the overlap between the reported 17q21.31 duplications and the duplication identified within Family 3.................................62
Figure 4-4 Expression of SALL3 as per UCSC Genome Browser (Website: https://genome.ucsc.edu/cgi-bin/hgGateway)..63
Figure 4-5 Chromosome 14 duplication location comparison between Family 4 and Patient 331321...65

1 INTRODUCTION

1.1 Classification of Intellectual Disability

Intellectual disability (ID) is estimated to affect at least 1% of the population (Harris 2006), with The American Association on Intellectual and Developmental Disabilities defining it as ‘a disability characterised by significant limitations in both intellectual functioning and adaptive behaviour, originating before the age of 18’ (Vannest 2014) . The nomenclature of ID has evolved over time, with terms such as “mental retardation” previously being used. Intellectual disability is currently recommended as the best term to describe a disability resulting from the impairment of cognitive functions and adaptive behaviour with an early onset (Chiurazzi and Pirozzi 2016). Classification of mental disorders (psychiatric nosology) has typically been done by psychiatrists. The two most widely used classification systems for mental disorders are ICD and DSM.

The International Statistical Classification of Diseases and Related Health Problems (ICD) is readily available on the worldwide web, and was founded by the World Health Organisation (WHO) in 1948. The most recent edition, ICD-11, was released on 18th June 2018, with the aim of being implemented by the 1st January 2022 following endorsement after the 72nd World Health Assembly in 2019. More than 100 countries use ICD, allowing for a wealth of information to be shared globally regarding health trends and statistics. The use of ICD is becoming increasingly popular within the field of clinical care and research in defining and studying diseases.

The American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders (DSM) is published by the American Psychiatric Association

(APA). In 1844 APA published a statistical classification of institutionalised patients with mental disorders, with the aim to improve communication between institutions regarding the different disorders they saw. This first publication has evolved over time into the DSM diagnostic classification system known today. The DSM is now in its fifth edition, the DSM-5, which was published in May 2013. Unlike the ICD, which covers health as a whole, the DSM only aims to classify mental health disorders. The classification of intellectual disability within DSM-V is summarised within Figure 1-1 (APA 2013).

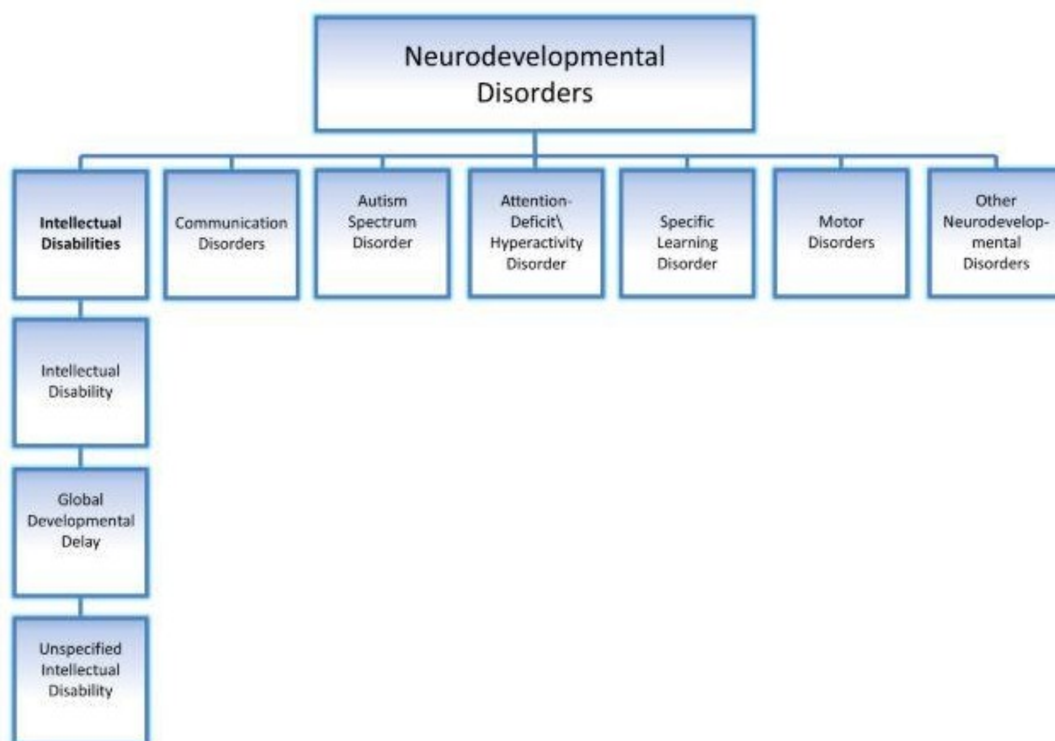


Figure 1-1. DSM-V classification of neurodevelopmental disorders (APA 2013)

Although both the ICD and DSM have proven enormously beneficial in providing a common language for communication, discrepancies within the definitions of categories between the two classification systems may actually be impeding international communication and research (First 2009). Both APA and WHO have agreed that a harmonisation between the two systems should be implemented, with a “DSM–ICD Harmonisation Coordinating Group” consisting of both DSM and

ICD members being tasked with this effort.

Within ICD-11 intellectual disability is referred to as a “disorder of intellectual development” and can be found within chapter 6 - Mental, behavioural or neurodevelopmental disorders, subsection: 6A00 Disorders of intellectual development. A disorder of intellectual development is defined as a condition “originating during the developmental period characterized by significantly below average intellectual functioning and adaptive behaviour that are approximately two or more standard deviations below the mean, based on appropriately normed, individually administered standardised tests”. Disorders of intellectual development can be subdivided into mild, moderate, severe, profound, and provisional, dependent on the number of standard deviations the individual differs from the mean, as well as age.

Conversely, DSM-5 emphasises the importance of using other methods of diagnosis, such as adaptive functioning, instead of relying on standardised tests such as IQ. By placing less weight on IQ scores and focusing more on conceptual, social, and practical domains, it was hoped that a more accurate and comprehensive description of how an individual’s deficit in mental abilities may affect their ability to function in day-to-day life would be established. DSM-5 states that three criteria must be met for a diagnosis of intellectual disability: (A) deficits in general intellectual functioning, (B) deficits in everyday adaptive functioning and (C) onset of these deficits in the developmental period. In keeping with ICD- 11, DSM-5 also subdivides intellectual disability into mild, moderate, severe and profound, with the severity determined by adaptive functioning. Comparing these

two systems and how they classify intellectual disability, it is clear that the DSM-5 places a lot more value on how a deficiency in intellectual functioning may impact the lives of the individuals.

Another term often confused with intellectual disability is developmental delay. Developmental delay can either be specific (only one area of development is affected), or global (more than one developmental domain is significantly affected). As stated in DSM-5, developmental delay is a term used only for individuals bellow the age of five. These children may be slow to reach certain developmental milestones, such as speech and language, social skills and motor skills, however as a child under the age of five may not be able to complete all the necessary assessments to ascertain a diagnosis of intellectual disability, the term developmental delay is used. A diagnosis of developmental delay does not preclude intellectual disability. Children with intellectual disability will experience developmental delay, however not all children with developmental delay will have intellectual disability.

1.2 The Cost of Intellectual Disability

The average cost of living for an older person diagnosed with ID is estimated to be around £790 a week in the UK (Strydom et al. 2010), which is more than double the basic income of £300 a week for a full time employee earning a minimum wage. Almost 90% of the participants of the study by Strydom et al relied on benefits for their income. This dependency on benefits leads to around 5% of the total care budget being used to care for older adults with ID. ID and psychiatric disorders represent the biggest share of healthcare costs in the Netherlands (J. J. Polder et al. 2002), a representation most likely paralleled

within other developed countries. The cost of living with an autism spectrum disorder increases by more than 1.5 times when also diagnosed with intellectual disability (Knapp, Romeo, and Beecham 2009), further highlighting the cost implications of living with ID.

The increased costs associated with intellectual disability may be due in part to inequalities in healthcare. Emerson et al identified the five main sources of health inequalities for people with ID (Emerson et al. 2011):

1. Greater risk of exposure to social determinants of poorer health such as poverty, poor housing, unemployment and social disconnectedness.
2. Increased risk of health problems associated with specific genetic and biological causes of learning disabilities.
3. Communication difficulties and reduced health literacy.
4. Personal health risks and behaviours such as poor diet and lack of exercise.
5. Deficiencies relating to access to healthcare provision.

These inequalities, although resulting in increased costs and a decreased standard of care and quality of life for ID patients, may also place many NHS Trusts in breach of the Equality Act 2010, the Mental Capacity Act 2005 and the Health and Social Care Act 2008.

As pointed out by Emerson et al, ID also comes with increased risks of other health issues. It has been shown that dementia occurs at a much higher rate among people with ID compared to older adults without (Cooper 1997) (Strydom et al. 2007). Epilepsy is a condition that can greatly affect the lives of the individuals and their families/ carers, and is more prevalent in people with ID

(Bowley and Kerr 2000). Having an ID also increases the likelihood of either being misdiagnosed for epilepsy or not being diagnosed at all (Chapman et al. 2011) (Lhatoo and Sander 2001), an issue which would lead to costly and ineffective health care.

Many of the inequalities faced by patients with ID could be prevented or at the very least identified with improved education of practitioners regarding the care of ID patients, as well as consistent and personal health care. The Annual Health Check is a scheme aimed at people registered by their GP as having ID. This scheme entitles registered patients to a free health check every year. The possible benefits of this scheme are clear, however for there to be a real impact on the lives of patients and better cost effective care, a more robust database of patients diagnosed with ID needs to be collated, as well as ensuring all patients attend their annual health check. In addition, further research into understanding ID will allow for better diagnosis and hopefully treatments for individuals.

1.3 Microcephaly

Neurodevelopmental disorders such as ID can be characterised by defective neuronal development, in particular, abnormalities within dendritic spines (Kaufmann and Moser 2000). Correct brain development relies on a complicated and delicate balance between gene expression and environmental inputs (Stiles and Jernigan 2010). The basic structures of the brain and central nervous system are established by the end of the embryonic period, with neuron production beginning 42 days post conception (Bystron, Blakemore, and Rakic 2008). An individual neuron can connect to more than 1,000 other neurons via synapses, and these neuron connection networks process all the information relating to our

thoughts, feelings, memories, sensations and actions. It is therefore unsurprising that abnormal brain growth and development have been linked to a vast amount of disabilities (Mirzaa and Paciorkowski 2014).

Primary Microcephaly is a defect identifiable at birth in which the baby's head is much smaller than expected (Figure 1-2), and can result in the formation of defective dendrites and dendritic structures. Microcephaly is therefore a common phenotype of children with intellectual disability (Watemberg et al. 2002). There are a number of different environmental factors that can lead to microcephaly. Exposure to alcohol or toxic chemicals during pregnancy, severe malnutrition, viruses/ infections and genetic changes are all aetiologies of microcephaly.

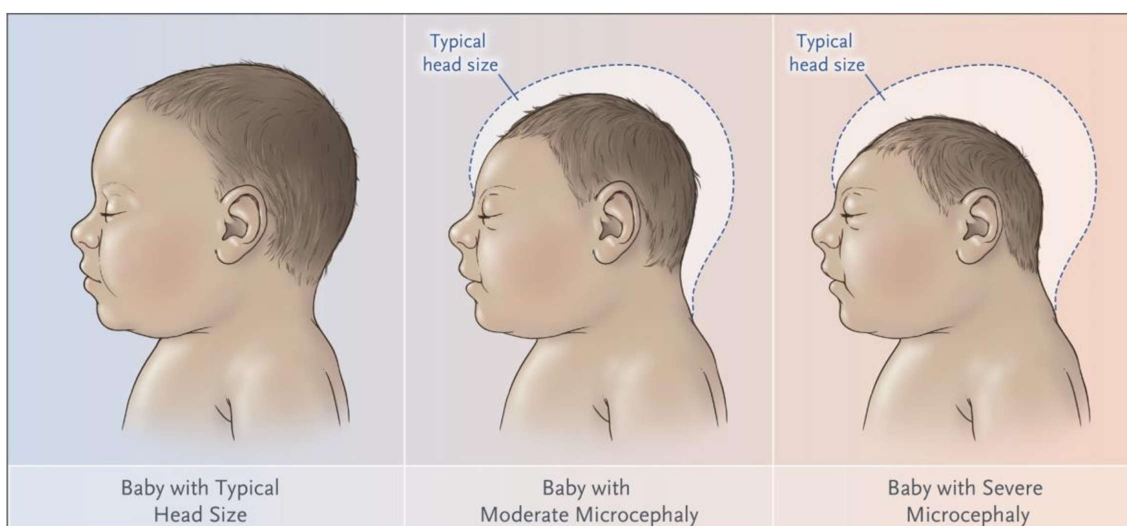


Figure 1-2 Typical Head Size, Microcephaly and Severe Microcephaly Comparison. Taken from (Petersen et al. 2016)

Environmental factors are an important cause of ID and may typically be preventable with a healthy lifestyle and access to a good standard of healthcare, as shown by a population study in Mumbai, India carried out by U. Dave et al in 2005. Their study identified 511 confirmed ID cases, 251 of which were caused by potentially preventable environmental factors, such as birth asphyxia, infections and low birth weight (Dave, Shetty and Mehta 2005). This highlights the necessity

to improve education and healthcare availability in developing countries.

1.4 The Genetics of Intellectual Disability and Microcephaly

~40-50% of moderate to severe cases of ID are thought to be linked to a genetic cause (Pavlowsky, Chelly, and Billuart 2012). The five most common types of ID are; Down syndrome, Williams syndrome, Angelman syndrome, Prader-Willi syndrome and Fragile X syndrome. All of these syndromes involve a known genetic cause. Down syndrome is a result of trisomy 21, which causes each cell in the body to have 47 chromosomes instead of the usual 46. Williams' syndrome is an autosomal dominant condition caused by a deletion mutation in chromosome 7. Angelman and Prader-Willi syndromes are a result of genomic imprinting; a phenomenon caused by parentally specific differential gene expression, and Fragile X syndrome is linked to a mutation in the *FMR1* gene on the X chromosome.

It has been estimated that the number of genes affecting cognition is 3.5 times greater on the X chromosome than on other chromosomes (Ulrich Zechner etal. 2001). One line of research has focused towards X-linked ID genetic mutations (Hu et al. 2016) (Piton, Redin, and Mandel 2013) due to the often recessive nature of ID and males being hemizygous for that chromosome. By 2012, 102 X-linked genes had been linked to 81 X-linked ID syndromes (Lubs, Stevenson, and Schwartz 2012), although this number is likely to increase year by year. Any apparent increased density of "intelligence genes" within the X-chromosome may be due to sexual selection during evolution, and may explain why there is greater

variability in male IQ scores. (Graves, Gecz, and Hameister 2002) (Johnson, Carothers, and Deary 2008). However, until every ID gene has been identified, it is clearly difficult to know for certain whether there is in fact a higher concentration of intelligence and neurological development genes located on the X-chromosome.

Through research centred predominantly on large, consanguineous families, many monogenic genetic causes of ID have been identified. The most recent update of SysID on the 14th June 2018 (a database cataloguing ID genes, Website: https://sysid.cmbi.umcn.nl/) lists 1,108 primary ID genes (genes that are published and have been confirmed to cause intellectual disability), and 808 candidate ID genes (Kochinke et al. 2016). Identifying monogenic causes of ID relies on there being a large sample size of individuals all with the same ID disorder. As the presentation of ID may be highly variable and overlapping, it can be difficult to definitively determine whether patients, often even within the same family, may have ID due to the same genetic cause. These issues may be circumvented to a degree when investigating the cause of ID in large (often consanguineous) families, where it is more likely that affected individuals will harbour the same pathogenic variant(s).

Research into the genetic causes of microcephaly has helped shed light on genes and proteins vital for normal brain development. Research into autosomal recessive primary microcephaly (MCPH) has indicated that the genes involved in this disorder may have played a role in the evolutionary expansion of the primate brain (Cox et al. 2006). To date, twelve MCPH loci have been mapped (Faheem

et al. 2015), however there will be many more syndromes in which microcephaly is a feature.

Precise regulation of the mitotic activity of neuronal progenitor cells is critical for correct development of neurons (Rakic 1988), therefore disruption of mitosis is likely to cause cortical disorders such as microcephaly. Many of the MCPH genes identified play a role in mitosis. For example, microcephalin (*MCPH1*) a known genetic cause of microcephaly, is involved in chromosomal condensation (Cox et al. 2006), *ASPM* is important for the orientation of mitotic spindles (do Carmo Avides and and Glove 1999) and *CENPJ* controls the length of centrioles (Hung et al. 2004), to name a few. Increased research into the genetic aetiology of MCPH has improved genetic testing and patient management of the condition, as well as improved scientific understanding of human brain development.

In 2002, Kelley *et al* reported a recessive form of microcephaly found within the Old Order Amish of Lancaster County (Kelley et al. 2002), due to a substitution (c.530G>C) mutation within *SLC25A19*. The *SLC25A19* gene encodes a protein required for the transport of thiamine pyrophosphate into mitochondria. Children affected by this condition have a head circumference of up to 12 standard deviations less than the general population, elevated levels of 2-ketoglutarate within the urine and early neonatal death. Further research into microcephaly identified a mutation within the sodium-dependent lysophosphatidylcholine transporter *MFSD2A* which caused a non-lethal microcephaly syndrome within a large Pakistani family (Alakbarzade et al. 2015).

The investigation of inherited forms of microcephaly has enabled the identification of many new genes important for normal brain development, which when perturbed may give rise to ID (Gund et al. 2016) (Vinkler et al. 2014) (Leshinsky-Silver et al. 2017) (Ji et al. 2015) (Tanaka et al. 2015).

1.5 The Role of Community-Based Programs to Define Genetic Causes of ID

Within clinical genetics, a marriage is traditionally considered as consanguineous when the two individuals are related as second cousins or closer (the inbreeding coefficient of their offspring is F ≥ 0.0156) (Bittles 2001), although over more recent years the term has been used to refer to marriages between individuals descended from the same ancestor. Occurrence of rare recessive syndromes, such as ID, has been shown to increase within consanguineous families (Teeuw et al. 2014). This is due to the higher chance of homozygosity for genetic variants carried by both parents, which in turn leads to the increased likelihood of recessive disorders occurring due to homozygosity for 'founder' gene mutations.

Over recent years, genetic studies have been successful in identifying a number of new inherited disorders, in particular genetically isolated populations such as the Amish community. For example, the Windows of Hope Project (WoH) is a UK (Exeter)-led translational research program working in close partnership with clinical workers in the USA who care for families within the Anabaptist communities across North America. The program aims to define the genetic basis of inherited disorders that occur in the community, with a particular (but not exclusive) focus on ID, to aid disease diagnosis and clinical care locally while improving medical-scientific knowledge of human inherited disease. Results

obtained from this program alone have enabled the clinical characterisation of 20 new inherited disorders.

One example of the work conducted by WoH involves a form of ID involving macrocephaly (a term used to describe a head circumference of more than two standard deviations above average for their age), neurodevelopmental delay and seizures due to two distinct founder mutations, both present in the Amish community, in the gene that encodes kaptin (*KPTN*) (Baple, Maroofian, et al. 2014). The importance of this finding to families is highlighted by the recent identification of numerous non- Amish families with the condition, stemming from the inclusion of the *KPTN* gene in gene sequencing diagnostic panels worldwide. Despite being published only recently, numerous clinicians have contacted the WoH research group for advice about the condition, for families harbouring the Amish mutations identified, as well as other candidate mutations in *KPTN* (Pajusalu, Reimand, and Ounap 2015). Taken together, this research not only highlighted kaptin as a molecule important for human neuromorphogenesis, but also highlighted the benefits of early diagnosis leading to improved clinical outcomes for patients worldwide with this disorder.

In another example of WoH studies, a homozygous missense sequence alteration of *PCNA* (proliferating cellular nuclear antigen), a cofactor that orchestrates genome duplication and maintenance, was identified within the Amish community of Ohio (Baple, Chambers, et al. 2014). A principal feature of this novel syndrome was neurodegeneration, and the condition shares clinical and molecular features common with other DNA repair conditions. Defining the molecular basis of the disease has provided a better understanding of the

biological processes responsible for other related DNA damage and repair disorders.

These and other similar studies highlight the importance of investigating the biological basis of rare diseases which have a direct impact on the wider population, as increased knowledge into gene function can help aid investigations into more common diseases and syndromes and may lead to possible treatments/cures.

This thesis derives from studies that form part of the University of Exeter Community Genetics Group translational genomic medicine program to aid genetic diagnosis and clinical care in families with inherited neurodevelopmental disorders. The families that are included in this thesis are all from the Nablus region of the West Bank, an area that has been, and continues to be, a place of conflict between Israel and Palestine (Gelvin 2014). The history of Israel and Palestine is a long and complex one, which has led to a unique genetic demographic within the area. The years of conflict and migration have created certain historical population bottlenecks (Kacowicz and Lutomski 2007), contributing to the occurrence of specific gene pools throughout this region. This has resulted in the accumulation of specific disease-associated alleles within particular communities.

A 'population bottleneck' occurs when a population becomes notably smaller very quickly. This can be due to migration/emigration patterns, a natural disaster, disease, war or other events. The sudden decrease in population size limits the amount of genetic variation within the population, which may lead to previously

rare ‘founder’ alleles becoming more common (the founder effect), once the population size recovers. The Amish community exemplifies how the founder effect has led to important repercussions for medical care based on the increased occurrence of otherwise rare inherited disorders within the community, although it is important to note that certain disorders that are common in other populations may occur at low frequency or even be absent in the Amish. Similarly, the high rate of consanguinity and specific marriage patterns within Middle Eastern societies has also led to increased frequencies of certain otherwise rare recessive disorders (Zlotogora 1997; Khalailah A et al. 2018). Due to the history of the Nablus population, it is likely that a number of the genetic mutations found within the community represent historical founder mutations.

1.6 Aims

This research study stems from an ongoing collaborative translational genomic medicine program between researchers in Exeter and clinicians in Nablus to aid diagnosis of families with intellectual disability. Four families with individuals who have complex intellectual disability disorders have been identified through the program. The aim of this research was to identify known and potentially new candidate genes and genetic variants to define the genetic causes of intellectual disability amongst these families.

2 SUBJECTS AND METHODS

As part of an ongoing collaboration between clinicians in Nablus (Israel) and the University of Exeter (UK), multiple families were identified with varying clinical phenotypes of undiagnosed neurodevelopmental disorders. Blood samples were collected with informed consent by the Israeli clinician and sent to the University of Exeter for genetic investigations. A genetic testing strategy was employed to determine the genetic basis of these conditions, which involved a combination of genome-wide single nucleotide polymorphism (SNP) genotyping and whole exome sequencing (WES).

2.1 Single nucleotide polymorphism (SNP) genotyping

SNP microarray genotyping was performed by Dr. Barry Chioza, University of Exeter, using Illumina Human CytoSNP v.2.1 arrays, following the manufacturer's protocol.

2.2 Primer Design

The UCSC Genome Browser (GRCh37/hg19 Feb 2009 assembly) was used to acquire the nucleotide base sequences of the genes of interest (website https://genome.ucsc.edu/index.html). Primers for PCR amplification were designed using Primer3 software version 0.4.0 (website http://bioinfo.ut.ee/primer3-0.4.0/primer3/).

The criteria used for primer selection was as followed:

1. Primer sizes were between 20-30 bases in length, to ensure binding specificity and annealing temperatures were accurate.

2. Melting temperatures of both flanking primers were kept as similar as possible (between 56- 61°C), to maintain PCR optimization and efficiency.

3. Where possible the guanine-cytosine (GC) base content was kept between 40-60%

4. Primers with inter or intra-primer efficiency extending for more than 3 bases were avoided, to prevent the formation of secondary structures and primer dimers

5. Primers were specific and matched 100% with the target region. This was checked by using *in silico* PCR analysis and BLAST searches using the UCSC Genome Browser website.

2.3 Polymerase Chain Reaction (PCR)

PCR is an *in vitro* method of amplifying targeted lengths of DNA exponentially. Each PCR reaction undergoes a number of cycles consisting of three stages: denaturation, primer annealing and DNA synthesis. PCR increases the DNA concentration of a sample to enable purified PCR products to then be sequenced.

For each PCR, a master mix containing all listed components (except DNA) was created (Table 2-1). A DNA control was used to ensure PCR efficiency, as well as a negative control of ddH_2O to ensure that the desired DNA template was being amplified, and not a contaminant. 9.2 µl of master mix was then aliquoted into individual 0.2 ml Eppendorf tubes. 0.8 µl of the DNA sample/control or ddH_2O

was added into sample, control and negative control tubes. A standard protocol of DNA amplification was followed.

Table 2-1 Components of standard 10 µl PCR tube

Component	Volume (µl)
Forward Primer (Sigma)	0.4
Reverse Primer (Sigma)	0.4
dNTPs (10mM)	0.4
Dream Taq Green Buffer	1.0
Molecular grade ddH_2O	6.9
Dream Taq DNA Polymerase	0.1
DNA sample, control or ddH_2O	0.8

An Eppendorf 96 well mastercycler, thermal cycler machine was used for the PCR reactions.

2.4 Primer optimisation

A gradient PCR program, running from 55-67°C was used to elicit the optimum temperature for primer pair annealing.

2.5 Agarose gel electrophoresis

Agarose gel electrophoresis is a method of visualising whether the desired DNA regions have been amplified. Agarose gel electrophoresis works by separating PCR products by their molecular size, therefore indicating the size of the PCR product.

1.7g agarose powder was added to 100ml 1x LAB buffer to make a 1.7% agarose gel. 1x LAB buffer was made by adding 20ml of 50x LAB buffer to 980ml of dH_20 (Table 2-2).

Table 2-2. 50x LAB buffer solution composition

51g Lithium acetate dehydrate
31g Boric acid
Make up to 1L with dH_20

The powder and buffer mixture was heated in a microwave on high heat until all powder had dissolved and the mixture was clear. 5µl of 10mg/ml ethidium bromide was carefully pipetted into the agarose solution, which was then poured into a casting tray with combs in place to make wells. Ethidium bromide is a DNA intercalating agent that fluoresces brightly when exposed to ultraviolet light, making the DNA bands easy to visualise.

Once set, the combs were removed and the gel placed within an electrophoresis tank containing enough 1x LAB buffer to submerge the gel. 5µl of 1kb plus ladder from Invitrogen was pipetted into a well, and 2µl of PCR product loaded into each subsequent well. The gel was left to run for at least 25 minutes at 150 volts, and then visualised under a UV transilluminator and photographs taken. If the negative control showed a DNA band, PCR products were discarded and PCR attempted again.

2.6 EXOSAP Purification of PCR product

Prior to sequencing it was necessary to remove unincorporated primers and dNTP's. This was done using EXOSAP, a product that contains Exonuclease-I (EXO) and Shrimp Alkaline Phosphatase (SAP). EXO degrades single stranded DNA in a 3' to 5' direction, releasing deoxyribonucleoside 5 prime-monophosphates. This releases dNTPs from the unincorporated primers. SAP catalyses the release of 5' and 3' phosphate groups from left over nucleotides. 2µl of EXOSAP was added to 5µl of each PCR sample product. The samples were then placed back into the thermal cycler and the EXOSAP program run. This program incubates the samples at 37°C for 15mins (the optimal temperature for the enzyme activity) and then denatures the enzyme by heating the samples to 85°C for 15mins.

2.7 DNA Sequencing

Purified PCR products with 5 of forward primer per sample were then sent to Source Bioscience Lifesciences, Cambridge, UK for dideoxy DNA sequencing. Whole genome exome (WES) analysis (NextSeq500: Illumina) involved: Agilent Sureselect Whole Exome v6 for targeting, with read alignment with BWA-MEM (v0.7.12), mate-pairs fixed and duplicate removal (Picard v1.129), InDel realignment / base quality recalibration (GATK v3.4-46), SNV/ InDel detection (GATK HaplotypeCaller and annotation (Alamut v1.4.4) and read depth (GATK DepthOfCoverage) to provide files (Excel) comprising all genetic variants identified for analysis in this project

3 RESULTS

Genome-wide SNP mapping and exome sequencing were used to identify putative copy number variants (CNVs), as well as candidate genomic regions of interest assuming initially autozygosity for putative founder mutations, although also considering other genetic causes.

3.1 Sequence alteration of *ALG1* associated with a congenital disorder of glycosylation

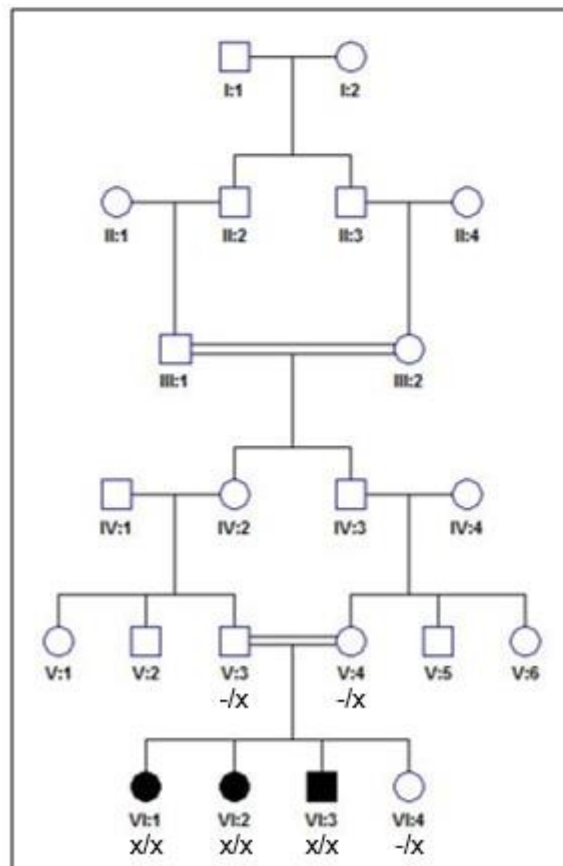


Figure 3-1. Family 1 Pedigree

Family 1 (Figure 3-1) includes 3 affected individuals presenting with ID, learning difficulties, hyperactivity, low attention span, speech delay, lower limb spasticity, mild hearing loss and non-dysmorphic microcephaly at -2SD.

Genome-wide SNP mapping was undertaken using DNA from all affected siblings. In order to identify CNV's, all affected individuals underwent whole genome microarray scan using Illumina Human CytoSNP-12v2.0 and 2.1 arrays. Data analysis was performed using karyostudio and genomestudio software (Illumina) to identify CNVs shared between affected individuals. No CNV's were identified, however regions of autozygosity common to all affected individuals were identified on chromosomes 2, 6, 12 and 16 (Figure 3-2, Table 3-1).

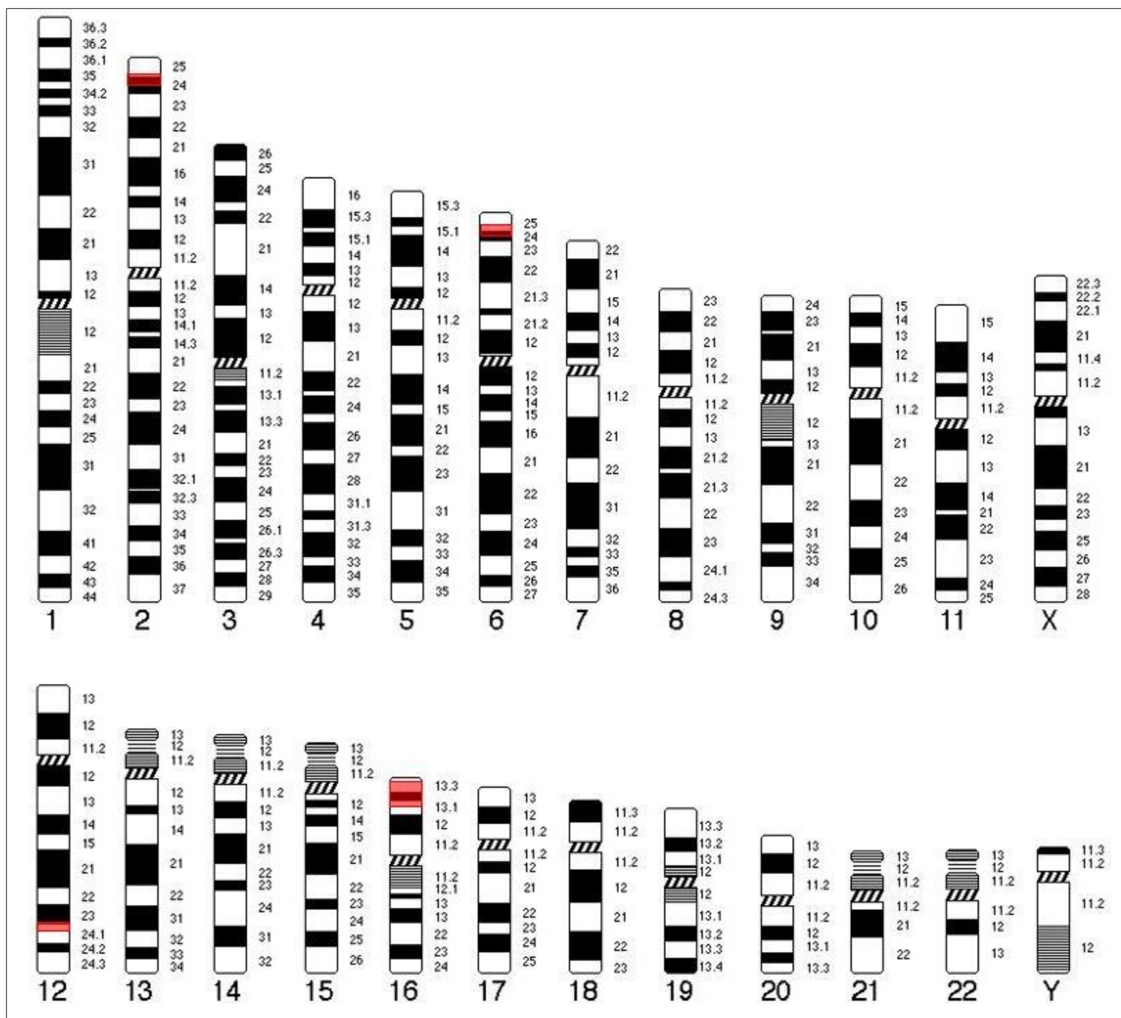


Figure 3-2. Visual depiction of largest autozygous regions of the genome (shown in red) common to all three affected siblings in Family 1

Table 3-1 Largest regions of autozygosity common to affected members in Family 2

Chromosome Location	Number of Genes within Region
Chromosome 16: 1,120,512-5,968,603	202
Chromosome 6: 4,600,524-9,283,079	30
Chromosome 2: 8,529,701-11,096,993	24
Chromosome 12: 111,057,714-112,825,713	21

Cross referencing the largest region of autozygosity on chromosome 16 with the SysID Database (Website: https://sysid.cmbi.umcn.nl/), 14 candidate or primary ID genes were identified to be located within this region (Table 3-2).

Whole exome sequencing of VI:3 was used to identify rare or novel candidate gene variants within regions of autozygosity, as well as genome-wide. The data was filtered to an allele frequency of <0.005% (with an initial frequency of <0.01% considered) to identify rare variants. Frameshift, nonsense, missense and essential splice site variants were all included in the filtering process to identify potentially deleterious variants. Those without any data available for their SIFT or POLYPHEN predictions of severity, or without a published allele frequency, were included in the shortlisting. Genes for which an individual was reported in homozygous state on gnomAD were excluded (website: http://gnomad.broadinstitute.org), to leave initially only a very selective list of variants. The genomic location of each gene was cross-referenced with the genome-wide SNP data, to identify the putative genotypings of each affected family member (indicated in table 3-3; 'haplotype region' column in which "All

three" indicates a SNP haplotype across each gene common to all three affected family members).

Table 3-2 Candidate/ Primary ID genes identified within the largest region of autozygosity from SysID for Family 1.

Candidate/ Primary ID genes	Phenotype
ALG1	Congenital disorders of glycosylation
ROGDI	Kohlschutter-Tonz syndrome
TBC1D24	DOORS syndrome
THOC6	Beaulieu-Boycott-Innes syndrome
TELO2	You-Hoover-Fong syndrome
TRAP1	Moderate ID
MRPS34	Combined oxidative phosphorylation deficiency 32
GNPTG	Mucolipidosis type III gamma
CCDC78	Centronuclear myopathy
CREBBP	Rubinstein-Taybi syndrome
TSC2	Tuberous sclerosis-2
RBFOX1	Epilepsy
NPRL3	Focal Epilepsy
USP7	Developmental delay, hypotonia and seizures

* Disorders inherited in an autosomal dominant pattern, not consistent with the pattern of inheritance in this family, have been greyed out

Whole exome sequencing identified only a limited number of candidate sequence variants (Table 3-3). Most notably, this included a homozygous variant within a gene located in the chromosome 16 region of homozygosity; *ALG1* (highlighted through SysID) (Table 3-2, Table 3-3). The variant identified within *ALG1* has previously been reported to be one of a number of *ALG1* mutations associated

with a rare autosomal recessive disorder termed ALG1- CDG (Ng et al.

Individuals with ALG1-CDG present with a neurological disorder characterised by neurodevelopmental delay, hypotonia, microcephaly and varying dysmorphic facial features. These features display significant overlap with the clinical characteristics of the affected individuals within family 1. Thus while the role of the other sequence variants identified in C16orf89, CACNA1H, CASKIN1, and PKD1 remains unclear, the ALG1 variant seems most likely to represent the genetic cause of the condition in Family 1.

Table 3-3 Novel and rare homozygous variants identified from whole exome sequencing of VI:3

Gene Name	ID	cNomen	pNomen	SIFT	Known Phenotype	Haplotype region
ALG1	rs151173406	NM_019109.4:c.826C>T	p.Arg276Trp	Deleterious	Congenital Disorders of Glycosylation	All three
C16orf89	rs371481724	NM_152459.4:c.514G>T	p.Asp172Tyr	Deleterious		All three
CACNA1H	rs61018135	NM_021098.2:c.2069C>T	p.Ala690Val	Tolerated	Familial Hyperaldosteronism Type IV	All three
CASKIN1	.	NM_020764.3:c.2525G>A	p.Gly842Asp	Deleterious		All three
PKD1	rs144512402	NM_001009944.2:c.3947A>T	p.Tyr1316Phe	Tolerated	Polycystic kidney disease	All three
PKD1	.	NM_001009944.2:c.3695G>C	p.Ser1232Thr	Tolerated	Polycystic kidney disease	All three
GLI2	rs150170739	NM_005270.4:c.1418G>A	p.Arg473His	Deleterious	Culler-Jones syndrome and Holoprosencephaly 9	VI:2 and VI:3
LRRC31	rs187788865	NM_001277128.1:c.1370C>T	p.Pro457Leu	Deleterious		VI:2 and VI:3
LRTOMT	rs138915488	NM_001318803.1:c.529C>T	p.Arg177Cys	Tolerated	Deafness	VI:2 and VI:3
SYNE1	rs373935459	NM_182961.2:c.23314C>T	p.Arg7772Trp	Deleterious	Emery-Dreifuss muscular dystrophy-4	VI:2 and VI:3
ATP2A2	rs140234740	NM_170665.3:c.3064A>G	p.Ile1022Val	Tolerated	Darier-White disease	VI:3
COL7A1	rs146783561	NM_000094.3:c.2666C>T	p.Pro889Leu	Tolerated	Epidermolysis bullosa dystrophica	VI:3
MYO7A	.	NM_000260.3:c.2467C>T	p.Arg823Cys	Tolerated	Usher syndrome type 1B and Deafness	VI:3
MYO7B	rs373648776	NM_001080527.1:c.5080G>A	p.Asp1694Asn	Deleterious		VI:3
NARS2	rs146900529	NM_024678.5:c.844G>A	p.Ala282Thr	Tolerated	Combined oxidative phosphorylation deficiency 24	VI:3
PNRC1	.	NM_006813.2:c.211C>A	p.Gln71Lys	Tolerated		VI:3
PTRHD1	.	NM_001013663.1:c.196C>T	p.Pro66Ser	Tolerated		VI:3
ABCB11	.	NM_003742.2:c.2238A>T	p.Lys746Asn	Tolerated	Cholestasis	VI:3

Primers were designed to amplify the *ALG1* gene variant (See Appendix 7.1) to validate the presence of the variant, and confirm cosegregation. As expected from the SNP haplotype data, the three affected siblings (VI:1, VI:2 and VI:3) were all homozygous for the c.826C>T variant. To ascertain if this variant cosegregated within the wider family, the parents (V:3 and V:4) and the unaffected sister (VI:4) were also sequenced. This confirmed that the parents were both heterozygous carriers for the mutation, as is the unaffected sister (Figure 3-3).

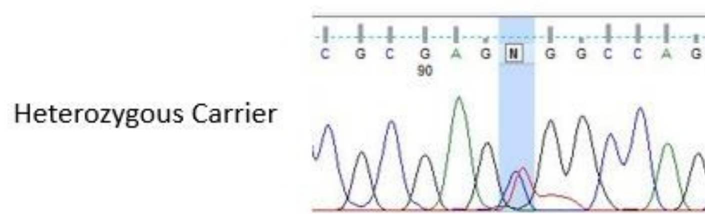


Figure 3-3 Sequencing chromatogram from a parental heterozygous carrier for the ALG1 c.826C>T gene variant, showing both the alternate (T) and wildtype (C) alleles

3.2 Defining mutations within an ARHGAP molecule as a likely new cause of ID

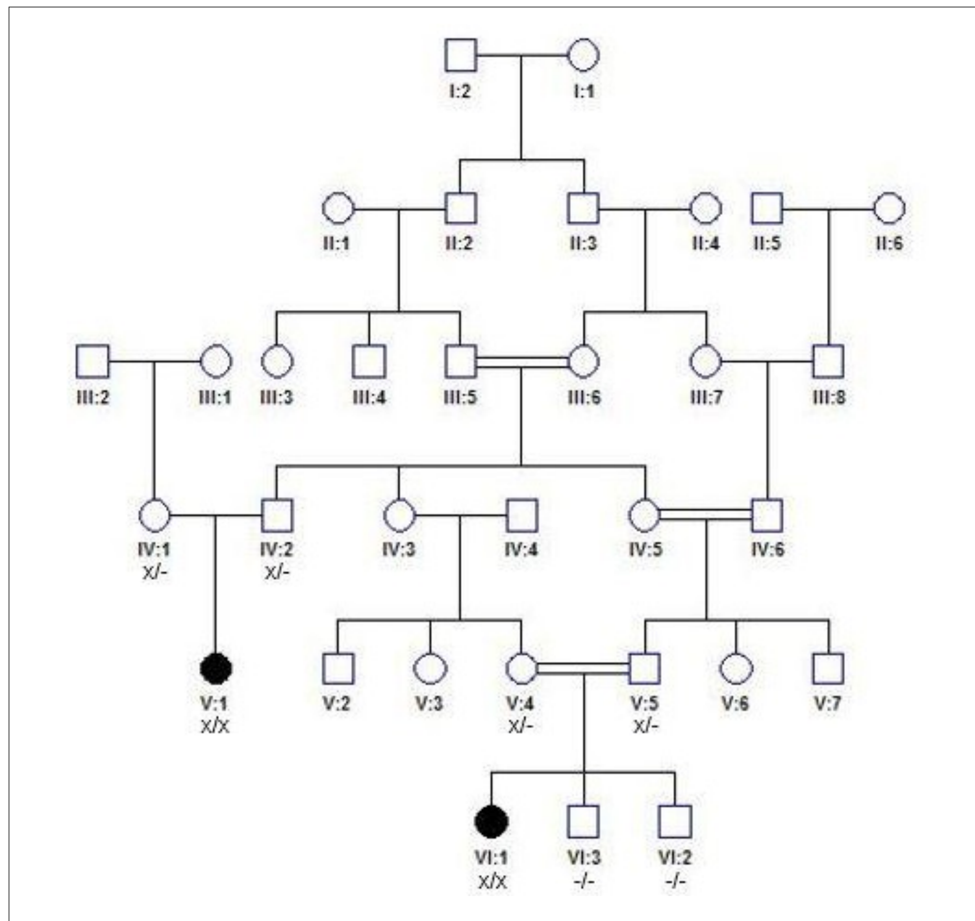


Figure 3-4 Family 2 Pedigree

Family 2 comprises an extended family involving 2 affected individuals present in two nuclear families (Figure 3-4). Both affected individuals are females presenting with global developmental delay, absent-speech, a failure to thrive, axial hypotonia, peripheral spasticity, are bed-ridden, and have microcephaly of up to -3SD. Both also have dysmorphic features including prominent ears, cleft soft palate and left cupped ear.

Assuming that a founder mutation is responsible, genome-wide SNP mapping of both the affected individuals was undertaken alongside whole exome sequencing of individual VI:1, as described in family 1. SNP mapping identified no CNV's, however two regions of autozygosity on chromosome 19 and 7 were identified (Figure 3-5, Table 3-4).

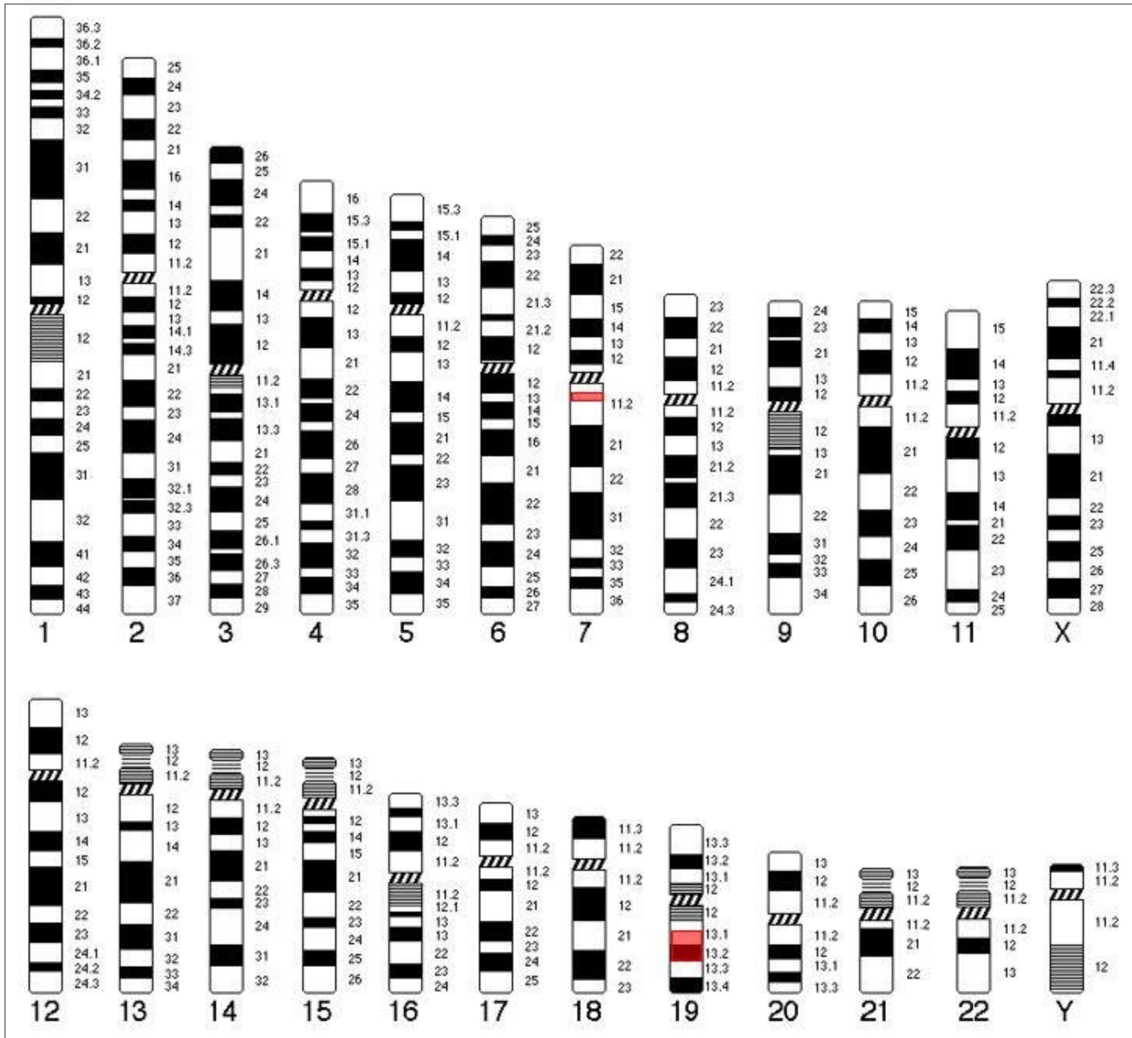


Figure 3-5. Visual depiction of largest autozygous regions of the genome (shown in red) common to affected siblings in Family 2

Table 3-4 Largest regions of autozygosity common to affected members in Family 2

Chromosome Location	Number of Coding Genes within Region
Chromosome 19: 34,610,341-47,658,320	475
Chromosome 7: 64,924,420-66,437,208	16

Whole exome sequencing was filtered to identify novel/rare (allele frequency <0.005%) deleterious homozygous variants, initially prioritising the largest region of autozygosity on chromosome 19, although also considering homozygous and compound heterozygous candidate variants genome-wide (Table 3-5). (Haplotype region column “Both homozygous” indicates a SNP haplotype across each gene common to the two affected family members.) This identified 8 candidate variants, with only one in *ARHGAP* being predicted deleterious by SIFT and PolyPhen. Notably, examination of the literature involving each gene identified highlighted a candidate variant within the same ARHGAP gene implicated as a candidate genetic cause of a neurodevelopmental disorder (See discussion section 4).

Table 3-5. Novel and rare homozygous variants identified from whole exome sequencing of VI:1

Gene Name	ID	c.Nomen	p.Nomen	SIFT prediction	Known Phenotype	SNP Region
ARHGAP	.	c.1673C>T	p.Ser558Phe	Deleterious	Identified as candidate gene for intellectual disability.	both homozygous
FDFT1	rs71711801	NM_001287750.1:c.226_231del	p.His76_Ser77del			both homozygous
HNRNPL	.	NM_001533.2:c.604C>T	p.Pro202Ser	Tolerated		both homozygous
PDCD2L	.	NM_032346.1:c.449C>A	p.Ala150Asp	Tolerated		both homozygous
PTGIR	.	NM_000960.3:c.268G>A	p.Ala90Thr	Tolerated		both homozygous
SARS2	.	NM_017827.3:c.422G>A	p.Arg141His	Tolerated	HUPRA syndrome	both homozygous
ZNF181	rs370173129	NM_001029997.3:c.407A>C	p.Asp136Ala	Tolerated		both homozygous
ZNF781	.	NM_152605.3:c.444G>C	p.Gln148His	Tolerated		both homozygous
AGTPBP1	.	NM_001286715.1:c.42_47dup	p.Ala18_Ala19dup			not shared
AKAP2	.	NM_001136562.2:c.2396A>G	p.Gln799Arg	Tolerated		not shared
AKAP2	.	NM_001198656.1:c.2663A>G	p.Gln888Arg	Tolerated		not shared
AKNA	.	NM_001317952.1:c.2923A>T	p.Ser975Cys	Deleterious		not shared
AKNA	.	NM_030767.5:c.3280A>T	p.Ser1094Cys	Deleterious		not shared
AKNAD1	rs141526254	NM_152763.4:c.961G>C	p.Gly321Arg	Tolerated		not shared
AP2A2	.	NM_001242837.1:c.2657C>T	p.Pro886Leu	Deleterious		not shared
ARHGAP45	.	NM_012292.4:c.1534A>G	p.Met512Val			not shared
ASPN	rs111419727	NM_017680.4:c.150_152dup	p.Asp50dup		D14 allele associated with Osteoarthritis	not shared
CCDC114	.	NM_144577.3:c.448A>T	p.Arg150Trp	Deleterious	Ciliary dyskinesia	not shared
CCDC177	rs147747237	NM_001271507.1:c.531_536del	p.Ala179_Ala180del			not shared
CTSD	.	NM_001909.4:c.190G>A	p.Glu64Lys	Tolerated	Neuronal ceroid lipofuscinosis	not shared
EP400	.	NM_015409.4:c.7410C>A	p.Asp2470Glu	Deleterious		not shared
EP400	.	NM_015409.4:c.8223_8225dup	p.Gln2748dup			not shared
GDF7	.	NM_182828.3:c.141_149dup	p.Gly48_Gly50dup			not shared
IFITM10	.	NM_001170820.3:c.541C>T	p.Arg181*			not shared
KRTAP5-6	.	NM_001012416.1:c.251G>T	p.Gly84Val	Deleterious		not shared
KRTAP9-1	rs11283848	NM_001190460.1:c.483_484ins27	p.Pro161_Cys162ins9			not shared
NDUFS7	rs201112782	NM_024407.4:c.31G>A	p.Gly11Ser	Tolerated	Leigh syndrome	not shared
NPIPB6	rs2411937	NM_001282524.1:c.917C>T	p.Ala306Val	Tolerated		not shared
PALM2-AKAP2	.	NM_007203.4:c.3089A>G	p.Gln1030Arg	Deleterious		not shared
PCDHA10	.	NM_018901.3:c.476C>A	p.Ala159Glu	Tolerated		not shared
PHOX2B	rs17879189	NM_003924.3:c.756_776del	p.Ala254_Ala260del		Central hypoventilation syndrome and Neuroblastoma susceptibility	not shared
RNF212	rs138488801	NM_001193318.2:c.721_722ins21	p.Ser241delins8			not shared
RPS6KA2	rs201444480	NM_021135.5:c.1237G>A	p.Asp413Asn	Tolerated		not shared
RSF1	.	NM_016578.3:c.2350C>A	p.Pro784Thr	Tolerated		not shared
SLC10A2	.	NM_000452.2:c.162_203dup	p.Ile55_Ile68dup			not shared
SLC3A2	.	NM_001012662.2:c.70T>C	p.Ser24Pro	Tolerated		not shared
SPATA31A3	.	NM_001083124.1:c.3943G>T	p.Val1315Leu	Tolerated		not shared
TENM3	.	NM_001080477.3:c.5223G>C	p.Glu1741Asp	Tolerated	Microphthalmia	not shared
USP20	.	NM_001008563.4:c.2101G>A	p.Ala701Thr	Tolerated		not shared
ZFP42	rs148426653	NM_174900.4:c.89G>A	p.Gly30Asp	Tolerated		not shared

Primers were designed to amplify and validate the *ARHGAP* gene variant, and confirm cosegregation. As expected the two affected girls (V:1 and VI:1) both harboured the same c.1673C>T/ p.Ser558Phe variant in homozygous form. To ascertain if this variant cosegregated within the wider family, both sets of parents (IV:1+IV:2 and V:4+V:5) and the two unaffected brothers (VI:3 and VI:2) were also sequenced, confirming that the parents were both heterozygous carriers for the mutation, whereas the unaffected brothers were both wildtype (Figure 3-6).

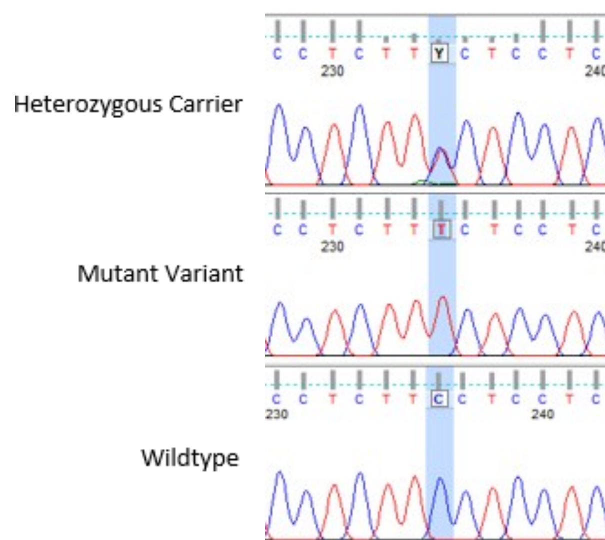


Figure 3-6 Sequencing chromatograms for ARHGAP c.1673C>T gene variant

To check whether any of the other Nablus families may also harbour the ARHGAP mutation, a single affected individual from each family was also sequenced for the *ARHGAP* variant. All were wildtype.

3.3 Exome sequencing identifies new candidate genetic causes of ID in an extended family with individuals with this disorder

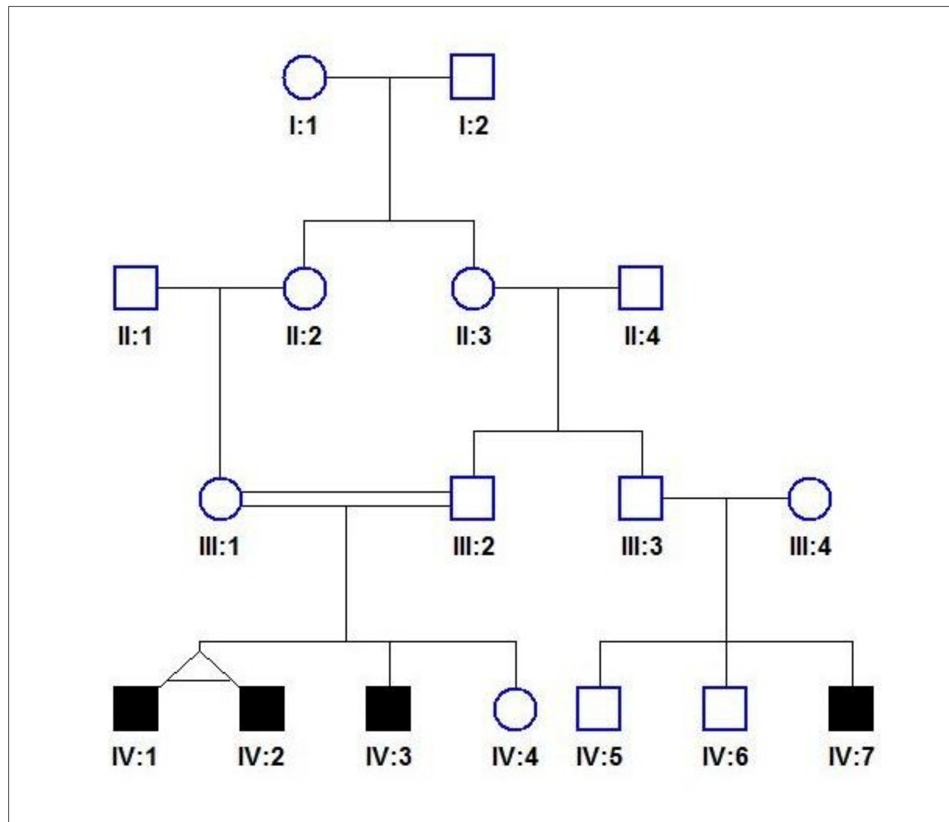


Figure 3-7 Family 3 pedigree

Family 3 includes 4 affected individuals (including a pair of monozygotic twins; Figure 3-7), presenting with intellectual disability, speech disturbances, delayed gross motor skills, short stature, dysmorphic features and generalized seizure disorder, which are controlled by antiepileptic drugs.

Genome-wide SNP mapping was undertaken using DNA from the three affected brothers (IV:1, IV:2, IV:3). A possible micro-duplication, common to all three brothers was identified on chromosome 17 (Figure 3-8). Four regions of autozygosity common to all three were also identified; two large regions within chromosome 16, as well as two smaller regions on chromosomes 1 and 13 (Figure 3-9, Table 3-6).

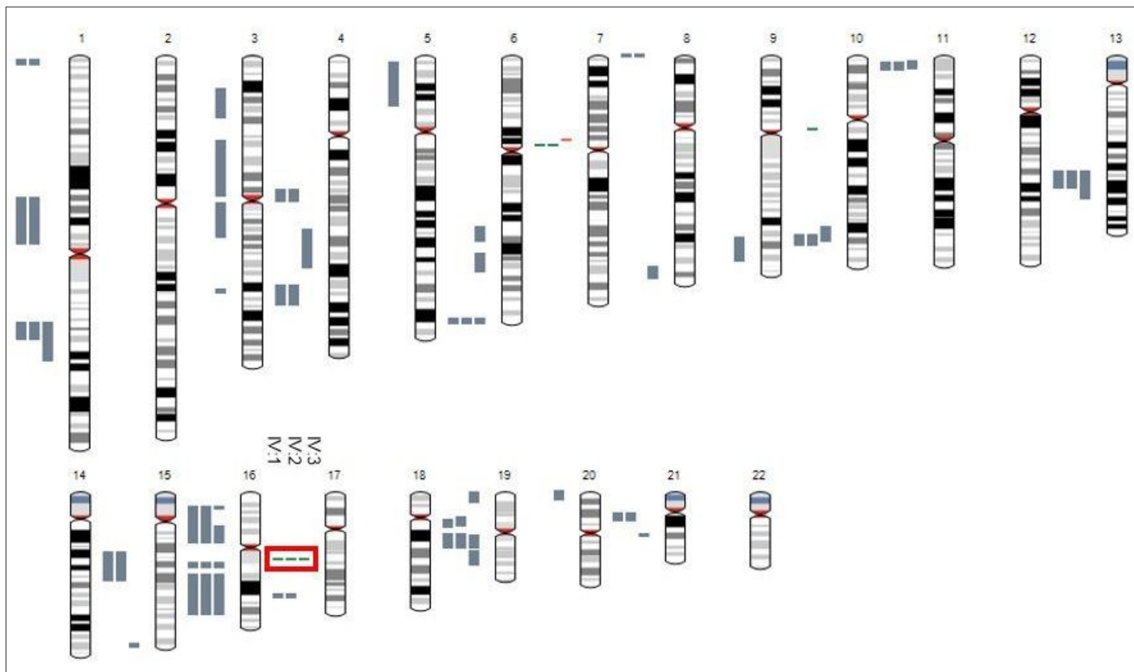


Figure 3-9 Karyotyping of three affected individuals within Family 3 to identify shared CNV's (highlighted by red box)

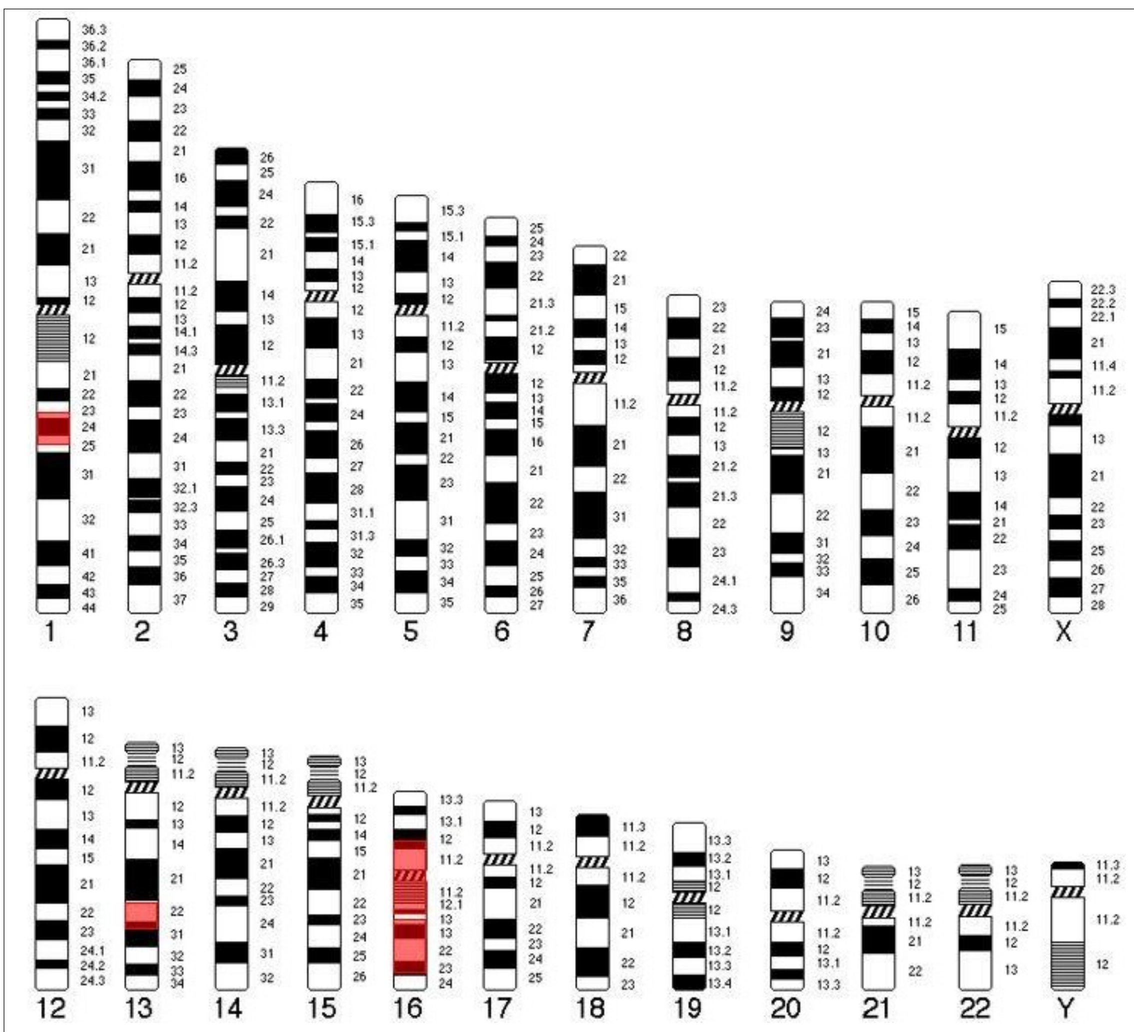


Figure 3-8 Visual depiction of largest autozygous regions of the genome (shown in red) common to all three affected siblings in Family 3

Table 3-6 Largest regions of autozygosity common to affected members in Family 3

Chromosome Location	Number of Genes within Region
Chromosome 16: 22,859,387-51,131,980	296
Chromosome 16: 54,080,912-81,177,309	308
Chromosome 1: 168,412,762-180,263,615	132
Chromosome 13: 73,269,968-85,330,291	50

The duplication region within 17q21.31 contains one gene (*KANSL1*), however further research into this duplication is required as the SNP data cannot confirm conclusively that this duplication is genuine. Duplications within 17q21.31 have been reported within a handful of intellectual disability patients with the term 17q21.31 micro-duplication syndrome now being used.

To identify any candidate sequence variants in genes of interest within the regions of autozygosity, the largest region was cross-referenced with the SysID database (Website: https://sysid.cmbi.umcn.nl/), which identified 15 candidate primary ID genes (Table 3-7).

Whole exome sequencing of VI:1 was used to identify rare (<0.005%) or novel candidate gene variants within the regions of autozygosity, as well as genome-wide (Table 3-8). The data was filtered as described previously (Section 3.1). (Tables 3-8 and 3-9 haplotype region column “All three” indicates a SNP haplotype across each gene common to all three affected family members.) No candidate homozygous variants were found within any of the ID genes identified by SysID within the largest region of autozygosity, however 7 homozygous

variants were identified genome wide. The exome data was also inspected for candidate compound heterozygous variants (Table 3-9). Any variant identified in homozygous form as per gnomeAD Browser (website: http://gnomad.broadinstitute.org) were discarded from the initial analysis as potential candidate variants. This identified a number of sequence variants likely present in homozygous (or compound heterozygous) state in all three affected siblings, with one gene known to cause autosomal dominant mental retardation (Table 3-8). As a number of novel candidate variants were identified, further work including the genotyping of additional family members (once samples become available) may identify a new genetic cause of this condition.

Table 3-7 Candidate/ Primary ID genes identified within the largest region of autozygosity from SysID for Family 3.

Candidate/ Primary ID genes	Phenotype
ALDOA	Glycogen storage disease
ATP2A1	Brody myopathy
CLN3	Neuronal ceroid lipofuscinoses
COG7	Congenital glycosylation type IIe
EARS2	Combined oxidative phosphorylation deficiency 12
GPT2	Autosomal recessive mental retardation 49
KIAA0556	Joubert syndrome
VPS35	Parkinson's disease
DCTN5	Autism
PRRT2	Episodic kinesigenic dyskinesia, seizures
SALL1	Townes-Brocks syndrome
SETD1A	Schizophrenia
SRCAP	Floating-Harbor syndrome
STX1B	Generalized epilepsy with febrile seizures
TBX6	Autism spectrum disorder

* Disorders inherited in an autosomal dominant pattern, not of relevance to the mode of inheritance in this family, have been greyed out

Table 3-8 Novel and rare homozygous variants identified from whole exome sequencing of VI:1

Gene Name	ID	cNomen	pNomen	SIFT	Known Phenotype	Haplotype Region
DHX38	.	NM_014003.3:c.889C>T	p.Arg297Cys	Deleterious	Possible role in retinitis pigmentosa	All three
KIAA0355	.	NM_014686.3:c.1585T>G	p.Ser529Ala	Deleterious		All three
SALL3	.	NM_171999.3:c.545C>G	p.Ser182Trp	Deleterious		All three
MYLPF	rs201552895	NM_001324458.1:c.350A>G	p.Lys117Arg	Tolerated		All three
NUP93	.	NM_014669.4:c.1348G>A	p.Glu450Lys	Tolerated	Nephrotic syndrome	All three
ZNF681	.	NM_138286.2:c.1081A>G	p.Thr361Ala	Tolerated		All three
BEAN1	.	NM_001178020.2:c.307T>C	p.Tyr103His	Tolerated	Spinocerebellar ataxia	All three
BAG3	.	NM_004281.3:c.557C>T	p.Ala186Val	Tolerated	Cardiomyopathy	VI:1 and VI:2
NOMO3	.	NM_001004067.3:c.2161A>G	p.Arg721Gly	Deleterious		VI:1 and VI:2
DNAH3	rs200671634	NM_017539.2:c.3070A>G	p.Lys1024Glu	Tolerated		VI:1 and VI:2
ZNF99	rs183240953	NM_001080409.2:c.419C>T	p.Thr140Ile	Tolerated		VI:1 and VI:2
ZNF91	rs372751233	ENST00000397082.2:c.2047A>G	p.Lys683Glu	Tolerated		VI:1 and VI:2
DCLK2	.	NM_001040261.4:c.106A>G	p.Ser36Gly	Tolerated		VI:1 and VI:2

Table 3-9 Novel and rare compound heterozygous variants identified from whole exome sequencing of VI:1

Gene Name	ID	cNomen	pNomen	SIFTprediction	Known Phenotype	Haplotype Region
MYT1L	.	NM_001303052.1:c.869C>T	p.Ser290Leu	Tolerated	Mental retardation	All three
MYT1L	.	NM_001303052.1:c.373G>A	p.Glu125Lys	Tolerated	Mental retardation	All three
RGPD4	.	NM_182588.2:c.80C>T	p.Thr27Met	Tolerated		All three
RGPD4	.	NM_182588.2:c.3982G>C	p.Glu1328Gln	Deleterious		All three
AMER3	rs201307390	NM_152698.2:c.538G>A	p.Glu180Lys	Tolerated		All three
AMER3	rs200166860	NM_152698.2:c.1033G>T	p.Ala345Ser	Tolerated		All three
ZNF717	rs199663998	NM_001128223.2:c.856G>A	p.Val286Ile	Tolerated		All three
ZNF717	rs200828229	NM_001128223.2:c.848A>G	p.Tyr283Cys	Tolerated		All three
CPZ	rs143281934	NM_003652.3:c.602G>A	p.Ser201Asn	Deleterious		All three
CPZ	rs139569426	NM_003652.3:c.1166A>T	p.Lys389Met	Deleterious		All three
CYP4V2	rs202204817	NM_207352.3:c.1372G>A	p.Val458Met	Deleterious	Bietti crystalline corneoretinal dystrophy	All three
CYP4V2	rs142244984	NM_207352.3:c.1493A>C	p.Glu498Ala	Deleterious	Bietti crystalline corneoretinal dystrophy	All three
TAS2R19	rs370558279	NM_176888.2:c.535G>A	p.Val179Ile	Tolerated		All three
TAS2R19	.	NM_176888.2:c.529T>A	p.Leu177Met	Tolerated		All three
TAS2R19	rs200043527	NM_176888.2:c.527G>A	p.Ser176Asn	Tolerated		All three
NBPF1	.	ENST00000430580.2:c.3200T>C	p.Val1067Ala	Tolerated		VI:1 and VI:2
NBPF1	.	ENST00000430580.2:c.1174C>T	p.Arg392Cys	Tolerated		VI:1 and VI:2
HRNR	rs113263380	NM_001009931.2:c.7570G>A	p.Gly2524Ser	Tolerated		VI:1 and VI:2
HRNR	.	NM_001009931.2:c.7397G>A	p.Arg2466His	Tolerated		VI:1 and VI:2

Gene Name	ID	cNomen	pNomen	SIFTprediction	Known Phenotype	Haplotype Region
FCRLA	rs144403707	NM_032738.3:c.523G>A	p.Ala175Thr	Deleterious		VI:1 and VI:2
FCRLA	rs144830049	NM_032738.3:c.524C>A	p.Ala175Glu	Deleterious		VI:1 and VI:2
SP8	rs185796187	NM_182700.5:c.487G>T	p.Gly163Cys	Deleterious		VI:1 and VI:2
SP8	.	NM_182700.5:c.475G>C	p.Gly159Arg	Deleterious		VI:1 and VI:2
POMT1	rs142057517	NM_001077366.1:c.1090G>A	p.Ala364Thr	Tolerated	Muscular dystrophy-dystroglycanopathy with mental retardation	VI:1 and VI:2
POMT1	rs138171526	NM_001077366.1:c.1898C>T	p.Ala633Val	Tolerated	Muscular dystrophy-dystroglycanopathy with mental retardation	VI:1 and VI:2
DRD4	.	NM_000797.3:c.860A>G	p.Gln287Arg	Tolerated	Risk factor for ADHD	VI:1 and VI:2
DRD4	.	NM_000797.3:c.863A>G	p.Asp288Gly	Tolerated	Risk factor for ADHD	VI:1 and VI:2

3.4 Further genetic studies are required to determine the cause of ID in an extended family with multiple individuals with this disorder

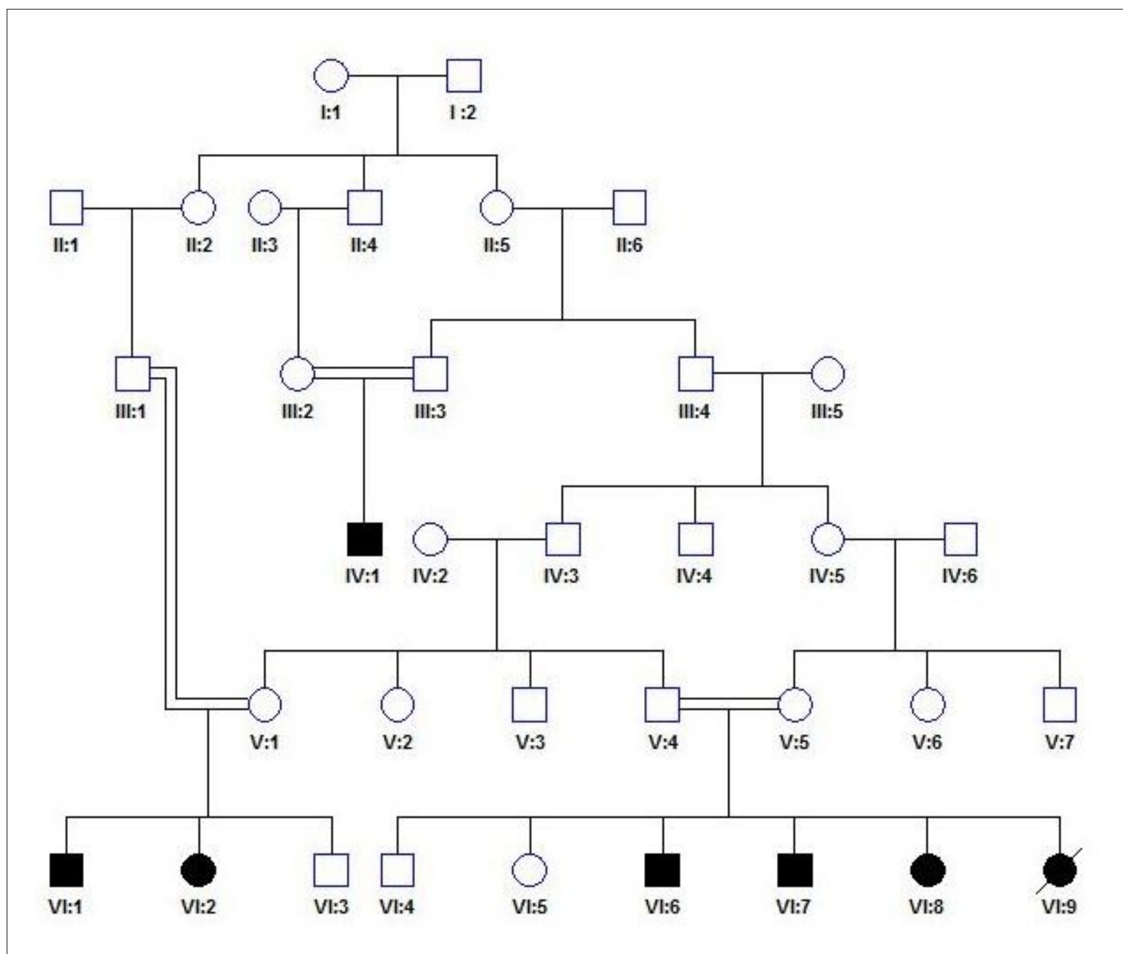


Figure 3-10 Family 4 pedigree

Family 4 incorporates multiple interlinking families with 7 affected individuals (in three nuclear pedigrees), one of whom is deceased (Figure 3-10). The siblings from the marriage between V:4 and V:5 were the research focus of this family. All three surviving siblings present with severe psychomotor retardation, speech delay, failure to thrive, can only crawl/ are wheelchair bound and have axial and peripheral hypotonia. All siblings have varying degrees of microcephaly, and sensorineural hearing loss.

Assuming that a founder mutation was most likely responsible for the condition, genome-wide SNP mapping was undertaken which identified one micro-duplication within one copy of chromosome 14 in all affected individuals (Figure 3-11). This analysis also identified autozygous regions common to the affected individuals within chromosomes 7, 12 and 17 (Figure 3-12, Table 3-10).

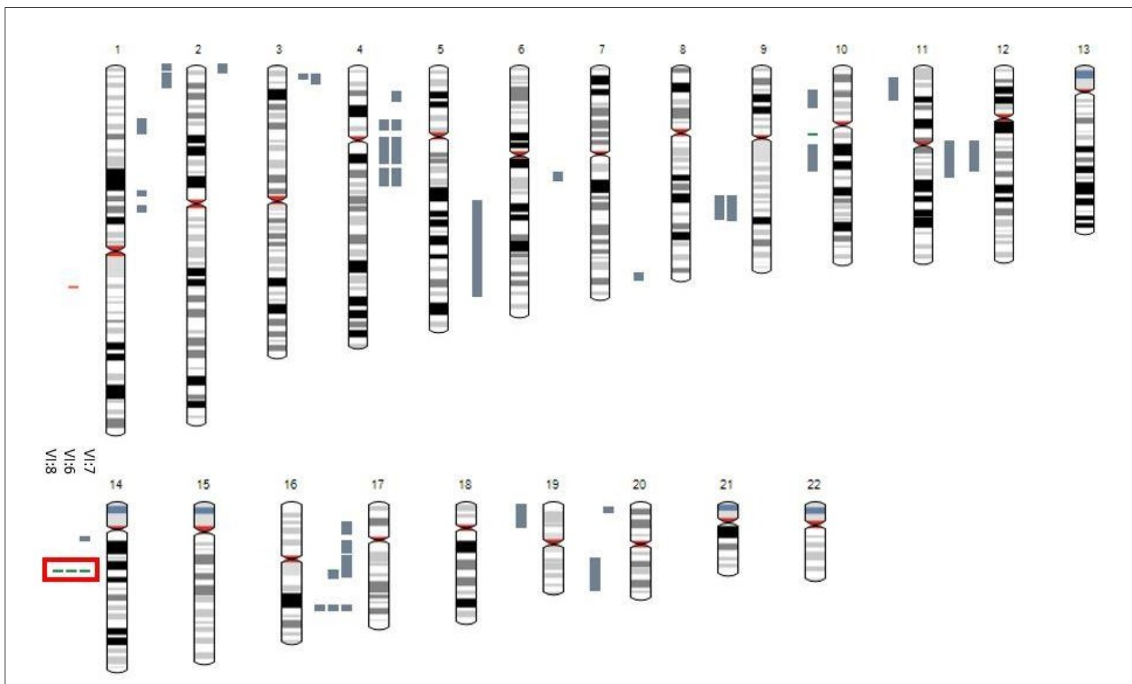


Figure 3-11 Karyotyping of three affected individuals within Family 4 to identify shared CNV's (highlighted by red box)

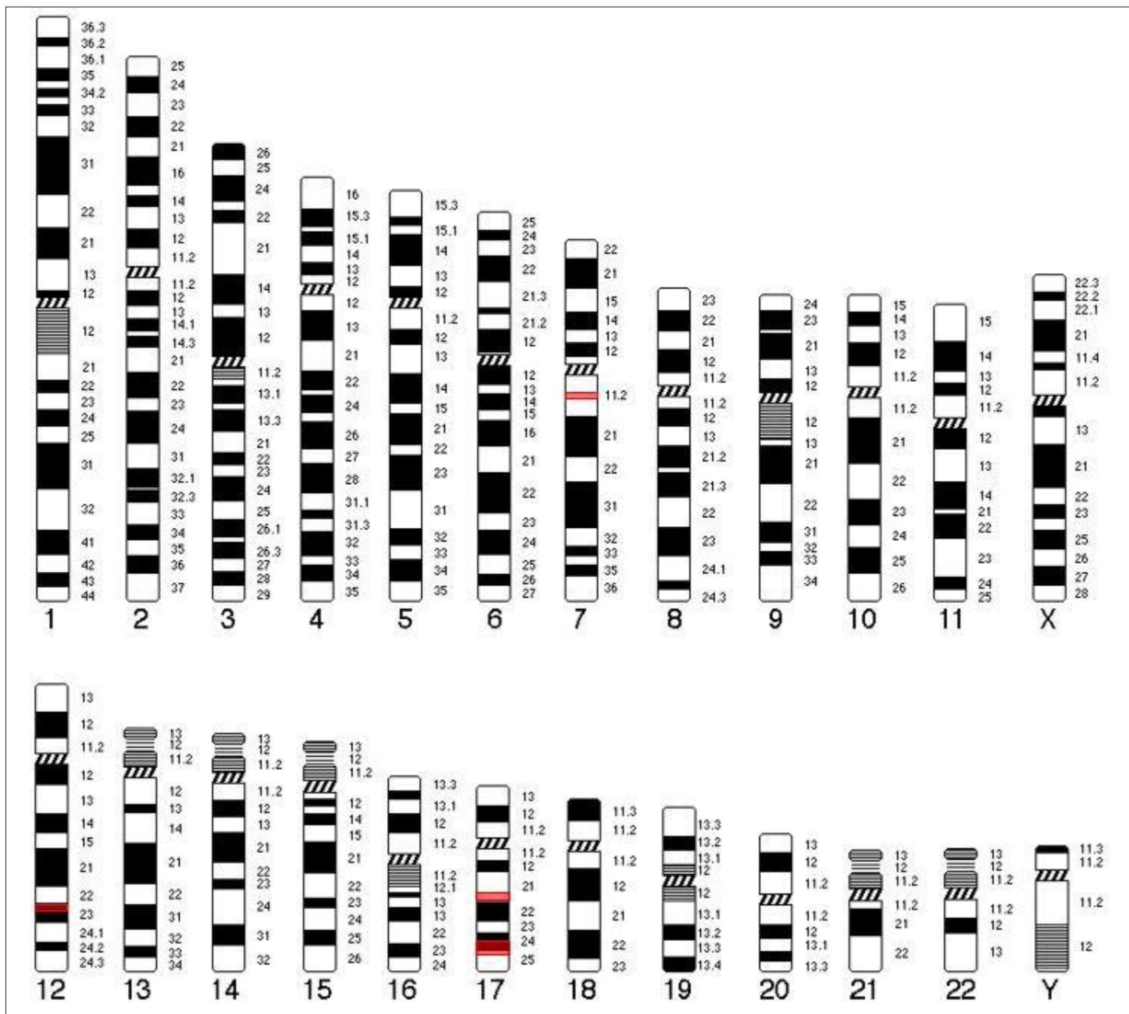


Figure 3-12 Visual depiction of largest autozygous regions of the genome (shown in red) common to affected siblings in Family 4

Table 3-10. Largest regions of autozygosity common to affected members of Family 4

Chromosome Location	Number of Genes within Region
Chromosome 17: 67,313,699-71,929,968	24
Chromosome 17: 45,221,271-46,176,007	27
Chromosome 7: 69,200,705-70,616,470	2
Chromosome 12: 108,313,346-109,372,311	11

The duplication region within 14q21.2 contains only one pseudogene (*KRT8P2*) and may therefore be unlikely to be pathogenic. However, there is a patient reported within DECIPHER v9.24 (website: https://decipher.sanger.ac.uk/) who is also heterozygous for a duplication on chromosome 14 which overlaps the coordinates of the duplication found within Family 4. The individuals' phenotype consists of microcephaly, facial abnormalities and intellectual disability.

Whole exome sequencing was also undertaken of a single affected individual (VI:8) to identify rare (<0.005%) or novel candidate gene variants within the autozygous regions, as well as genome-wide (Table 3-11). The data was filtered as described above (Section 3.1) No candidate homozygous variants in genes located within the largest shared autozygous regions or genome wide were identified. Finally, the exome data was inspected to identify candidate compound heterozygous variants (Table 3-12). Any variant identified in homozygous form as per gnomeAD (website: http://gnomad.broadinstitute.org) were discarded as potential candidate variants and removed from the table. No candidate compound heterozygous variants were identified.

Table 3-11. Novel and rare homozygous variants identified from whole exome sequencing of VI:8

Gene Name	ID	cNomen	pNomen	PolyPhen	SIFT	Known Phenotype	Haplotype Region
ANKRD36	.	NM_001164315.1:c.959C>T	p.Ser320Leu	Damaging	Tolerated		VI:8
CARD6	rs143022216	NM_032587.3:c.1340T>C	p.Leu447Pro	Damaging	Deleterious		VI:8+ VI:6
CHL1	.	NM_006614.3:c.3355G>T	p.Val1119Leu	Damaging	Deleterious	Candidate gene for an autosomal form of intellectual disability.	VI:8
ESPL1	.	NM_012291.4:c.1154G>T	p.Gly385Val	Benign	Tolerated		VI:8+ VI:7
FDFT1	rs71711801	NM_001287750.1:c.226_231del	p.His76_Ser77del				VI:8
USP17L10	.	NM_001256852.1:c.718T>A	p.Leu240Met	Damaging	Tolerated		VI:8+ VI:6

Table 3-12. Novel and rare compound heterozygous variants identified from whole exome sequencing of VI:8

Gene Name	ID	cNomen	pNomen	SIFTprediction	Known phenotype	Haplotype region
NBPF1	.	ENST00000430580.2:c.1051_1053del	p.Glu351del			VI:6+VI:7
NBPF1	.	ENST00000430580.2:c.2092C>A	p.Gln698Lys	Not scored		VI:6+VI:7
OBSCN	rs199738703	NM_052843.3:c.11750C>T	p.Thr3917Ile	Tolerated		VI:6+VI:7
OBSCN	rs200262505	NM_052843.3:c.15689C>T	p.Thr5230Met	Deleterious		VI:6+VI:7
PCNT	.	NM_001315529.1:c.1589C>G	p.Pro530Arg	Tolerated	Microcephalic osteodysplastic primordial dwarfism type II	VI:6+VI:7
PLB1	.	NM_001170585.1:c.892A>T	p.Thr298Ser	Tolerated		VI:6+VI:7
PLB1	.	NM_153021.4:c.859A>T	p.Thr287Ser	Tolerated		VI:6+VI:7
PLB1	.	NM_153021.4:c.863C>A	p.Thr288Asn	Tolerated		VI:6+VI:7
PLB1	.	NM_153021.4:c.869C>A	p.Ser290Tyr	Tolerated		VI:6+VI:7
PRSS3	.	NM_007343.3:c.257C>A	p.Thr86Asn	Tolerated		VI:6+VI:7
ALDH16A1	rs146615636	NM_153329.3:c.1484C>T	p.Ser495Phe	Tolerated	Possible role in gout	VI:8+VI:6
DGKG	.	NM_001346.2:c.2152C>G	p.Gln718Glu	Deleterious		VI:8+VI:6
DGKG	.	NM_001346.2:c.2149G>T	p.Gly717Trp	Deleterious		VI:8+VI:6
FOXO6	rs373524987	ENST00000372591.1:c.1009_1010del	p.Gln337Glyfs*39		Potentially role in neurodegenerative disorders and risk gene of bipolar	VI:8+VI:6
FOXO6	.	ENST00000372591.1:c.1013C>G	p.Ala338Gly	Not scored		VI:8+VI:6
ADAMTSL4	.	NM_001288607.1:c.419T>A	p.Ile140Asn	Tolerated	Ectopia lentis	VI:8+VI:7
ADAMTSL4	.	NM_001288607.1:c.2372C>G	p.Pro791Arg	Deleterious		VI:8+VI:7
DUOX1	rs367870737	NM_017434.4:c.2425G>A	p.Glu809Lys	Tolerated		VI:8+VI:7
DUOX1	.	NM_017434.4:c.3317G>A	p.Arg1106His	Deleterious		VI:8+VI:7
FLG	rs142483068	NM_002016.1:c.2363G>A	p.Arg788Gln	Not scored	Ichthyosis vulgaris	VI:8+VI:7
GIGYF2	.	NM_001103147.1:c.1889G>A	p.Arg630Gln	Tolerated	Possible risk factor for Parkinson’s	VI:8+VI:7
GIGYF2	.	NM_001103147.1:c.2386G>A	p.Glu796Lys	Deleterious		VI:8+VI:7

4 DISCUSSION

Identifying candidate genes for neurodevelopmental disorders is important for the diagnostic process of the families, but also to help allude to the functions of genes of evolutionary importance for brain development and function. The research described in this thesis, through the collaboration between Exeter University and our Nablus colleagues, has enabled a suspected diagnosis to be provided for two families, and also identified a novel candidate disease gene responsible for ID. While two further families require further evaluations in order to reach diagnosis, these findings highlight the importance of such collaborative clinical-genetics programmes around the world.

The first family investigated (Family 1, Section 3.1) includes 3 affected individuals, all with ID, learning difficulties, hyperactivity, low attention span, speech delay, lower limb spasticity, mild hearing loss and non-dysmorphic microcephaly at -2SD. Genome-wide SNP mapping identified no CNV's, however chromosomes 2, 6, 12 and 16 all displayed regions of autozygosity common to all affected individuals. The largest region of autozygosity within chromosome 16 contained 14 genes listed within the SysID database. Importantly, whole exome sequencing of VI:3 identified a rare (allele frequency <0.005%) homozygous mutation (c.826C>T) within one of these aforementioned genes, *ALG1*.

Mutations within *ALG1* are a well-documented cause of a rare autosomal recessive congenital disorder of glycosylation termed ALG1-CDG (Ng et al. 2016). CDG's are caused by defects in the synthesis and processing of

asparagine (ASN)-linked oligosaccharides of glycoproteins (Grunewald 2002). ALG1-CDG can be highly variable, but is usually characterised by; microcephaly, developmental delay, abnormal fat distribution, strabismus, and coagulation abnormalities (Morava. et al. 2012), with recurrent infantile seizures, cerebellar hypoplasia, hypotonia, and failure to thrive also being described in ALG1-CDG patients (Harshman et al. 2016).

The c.826C>T variant identified within the family investigated here has previously been reported in compound heterozygous form with c.1263G>A mutation within a young male from France (Dupre et al. 2010) who presented with psychomotor delay, growth retardation, hypotonia, multi-focal epilepsy and facial dysmorphic features. While it is likely that the *ALG1* alteration represents the cause of the condition in this family, there were some difficulties with obtaining pure sequencing chromatograms from the affected children. Resequencing using repositioned oligonucleotide primers will enable sequence verification and confirmation, thus reporting the first three cases of individuals with this variant in homozygous form. Abnormal isoelectric focussing of transferrins is seen in individuals with CDG disorders and this confirmatory clinical investigation is being organised by our local clinical colleagues.

Other homozygous variants were also identified within homozygous genomic regions shared between all three affected individuals, and therefore expected to be found in homozygous form. While these cannot be conclusively excluded, none of them are known to cause ID. Of these homozygous variants, only the variants within *C16orf89* and *CASKIN1* were predicted to be deleterious. The *C16orf89* gene and protein shows little to no expression within the brain and has

no known corresponding phenotype, and may therefore be unlikely to be pathogenic. *CASKIN1* however is very highly expressed within the brain, and the Gly842Asp variant is very rare with only two heterozygous carriers reported within gnomAD. It is currently unclear as to if or how this variant may contribute to the ID phenotype within the family. While the relevance of the other variants identified in family 1 remains unclear, together these findings identify the *ALG1* mutation as the likely genetic cause of the condition in this family. This therefore likely provides an opportunity for genetic screening, as if the unaffected heterozygous sister decides to have children, it would be of importance to ascertain if her partner harbours a deleterious variant in this gene, as to establish the possible risk of them having children affected by this disorder, and be counselled accordingly.

The second family investigated (Family 2, Section 3.2) comprised of two young females with microcephaly, dysmorphic features including prominent ears, cleft soft palate, left cupped ear, axial hypotonia and peripheral spasticity, as well as global developmental delay. Both girls are bed-ridden with absent speech and a failure to thrive. Genome-wide SNP mapping highlighted a notable region of autozygosity on chromosome 19 shared by the two affected children. Whole exome sequencing of this region identified a single variant within an *ARHGAP* gene (c.1673C>T), (specific gene identity withheld as requested by project sponsors Prof Crosby and Dr Baple) affecting a highly conserved serine at position 558 being replaced by phenylalanine (p.Ser558Phe). The variant is predicted to be deleterious in SIFT and POLYPHEN, and is listed at very low frequency (0.000007673) and not in homozygous form in the gnomAD browser across all populations. Both sets of parents of the affected children were

heterozygous for the mutation, and the unaffected brothers of VI:1 were wildtype (Figure 3-6), consistent with autosomal recessive inheritance for this founder mutation.

A different variant within the ARHGAP gene was recently identified by a collaborator of the UEMS research group as a candidate cause of ID (Prof Crosby, personal communication). The variant is present in homozygous form in two siblings who displayed very similar clinical characteristics to the affected individuals in family 2, highlighting this gene variant as the likely cause of the disorder and confirming this gene as a likely new genetic cause of ID. Of the other homozygous variants identified from the exome data which lie within homozygous regions shared between the affected individuals, none represent obvious candidate genes. Only *ZNF181* and *ZNF781* are known to be expressed in the brain and these gene variants were not predicted to be deleterious by SIFT, although this alone does not exclude them.

ARHGAP is highly expressed within the brain (Figure 4-1) and encodes a member of a unique subfamily of sorting nexins (SNX). Sorting nexins are a large family of proteins categorized by a conserved lipid-binding Phox (PX) domain, which have been shown to play critical roles within brain development (Nakazawa et al. 2011) (Mizutani et al. 2012) (Wang et al. 2013) (Muirhead and Dev 2014) (Mizutani et al. 2011). The subfamily of sorting nexins in which *ARHGAP* belongs, contains the usual PX domain, as well as a Rho GTPase-activating protein (RhoGAP) domain.

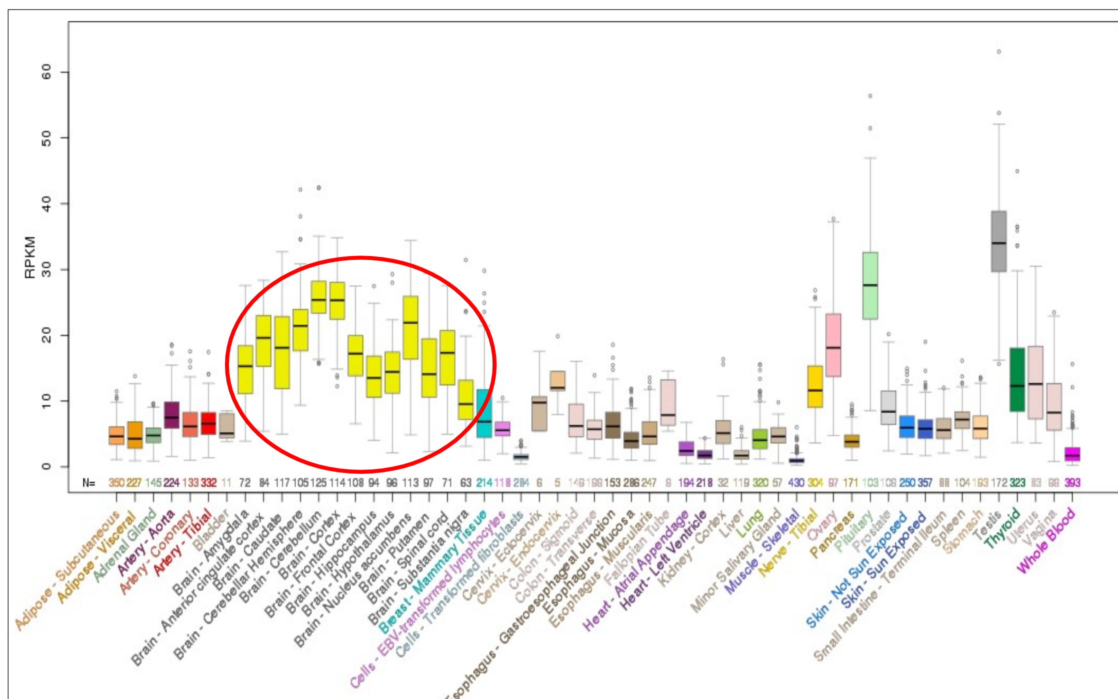


Figure 4-1 Expression of ARHGAP as per UCSC Genome Browser (Website: https://genome.ucsc.edu/cgi-bin/hgGateway)

The Rho family of GTPases (Rho GTPases) is comprised of a number of small signalling G proteins, and are best known for their involvement within the reorganisation of the actin cytoskeleton and the orientation and stabilisation of microtubules, which is very important for correct neuronal development (Govek, Newey, and Van Aelst 2005).

G proteins act as molecular switches, cycling between an active GTP bound state and inactive GDP bound state. The three key regulatory proteins involved in this cycling are guanine nucleotide exchange factors (GEFs), GTPase activating proteins (GAPs), and guanine nucleotide dissociation inhibitors (GDIs). GEFs are proteins that facilitate the exchange of GDP for GTP, therefore promoting the active GTP-bound state. GAPs deactivate Rho GTPases by promoting the hydrolysis of GTP, and GDIs have recently been shown to play both a negative and positive role in the regulation of Rho activities (Dovas and Couchman 2005). The ARHGAP molecule identified is therefore expected to have a dual function

as both a sorting nexin and a deactivator of Rho GTPases.

This ARHGAP molecule has also been shown to interact with other molecules linked to ID disorders, further indicating its potential role within ID pathogenesis (Figure 4-2). Variants within *RAC1, RHOBTB2 and CDC42* and have all been described within patients with ranging ID phenotypes (Reijnders et al. 2017) (Straub et al. 2018) (Takenouchi et al. 2015).

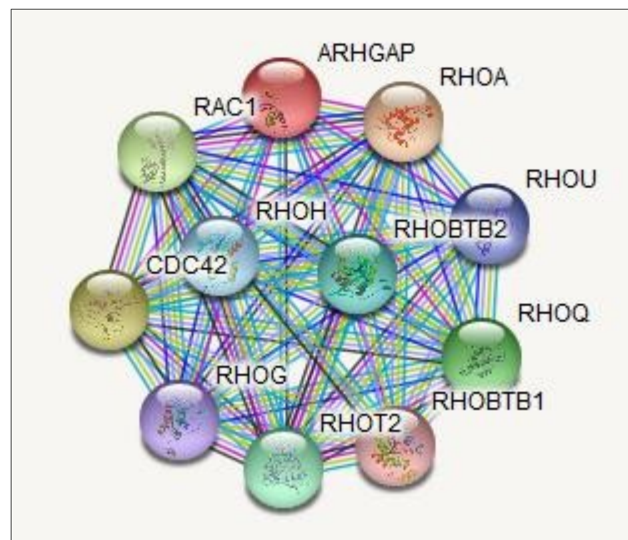


Figure 4-2 ARHGAP interaction as per STRING version 10.5

Further work needs to be conducted to discover just how the ARHGAP gene mutation may cause ID. Functional studies are currently ongoing within the UEMS research group and collaborators to validate this by investigating the outcomes of the sequence variants identified. This includes cell and protein binding partner studies, as well as studies of a mouse model of this gene.

In the third family investigated (Family 3, Section 3.3), three affected brothers (one set of monozygotic twins) presented with intellectual disability, speech disturbances, delayed gross motor skills, short stature, dysmorphic features and generalised seizure disorder. Genome-wide microarray scanning identified a 161kb micro-duplication within chromosome 17 which the three brothers shared.

Four regions of autozygosity within chromosomes 1, 13 and 16 were identified through genome-wide SNP mapping.

Considering that approximately 60% of chromosome 16 was autozygous for the three affected individuals, and that a founder mutation was the likely cause of the disorder, it was expected that the pathogenic variant(s) would be found within chromosome 16. Surprisingly, inspection of the exome data of VI:1 revealed no candidate homozygous or compound heterozygous variants within the autozygous regions. No known candidate homozygous variants were found genome wide.

Compound heterozygous variants within *MYT1L* were identified within a region on chromosome 2 in which the three affected individuals shared a haplotype. Neither of the two variants (p.Ser290Leu or p.Glu125Lys) have previously been reported. *MYT1L* has been identified as a candidate ID gene (de Ligt et al. 2012) with *de novo* heterozygous variants identified within a number of patients with autosomal dominant mental retardation (De Rocker et al. 2015) (Blanchet et al. 2017). Given that the mode of inheritance of the disorder in family 3 is most likely recessive, it is unlikely that either MYTL1 variant is pathogenic. However, it may be possible that these two variants have less penetrance, and that when inherited together in compound heterozygous form a more severe ID phenotype is caused. It may also be possible that one (or both) of these variants are pathogenic, for which a parent is germline mosaic (although this is unlikely). Currently the collaborative Nablus team are recruiting additional family members to the study, including cousins, one of whom is also thought to be affected by the disorder (individual IV:7, Figure 3-7). Sequencing studies of these individuals will be very

helpful for elucidating which gene variant may be responsible for this condition.

The 161kb duplication region within 17q21.31 contains one gene (*KANSL1*). Further research into this duplication is required (See Section 5) as the SNP data cannot confirm conclusively that this duplication is genuine and therefore cannot be considered a strong candidate for the pathogenic variant within this family yet. A deletion within *KANSL1* is known to cause Koolen-De Vries syndrome (Zollino et al. 2012) (Koolen et al. 2012), which is an autosomal dominant ID syndrome (Koolen et al. 2006) (Koolen et al. 2008). Microduplications within 17q21.31 were first reported in 2007 by Kirchhoff et al and has since been termed 17q21.31 microduplication syndrome (Kirchhoff et al. 2007). Since then only a handful of patients with 17q21.31 microduplication syndrome have been reported (Grisart et al. 2009) (Kitsiou-Tzeli et al. 2012) (Mc Cormack et al. 2014) (Natacci et al. 2016), all with intellectual disability and other characteristics such as microcephaly and behavioural abnormalities. Comparing the specific break points of the previously reported 17q21.31 microduplication syndrome cases and the approximate location of the duplication found within family 3, there is some overlap (Figure 4-3). The coding region of *KANSL1* is from chr17:44,108,842-44,249,509. The first two exons within *KANSL1* lie within the duplicated region identified within Family 3.

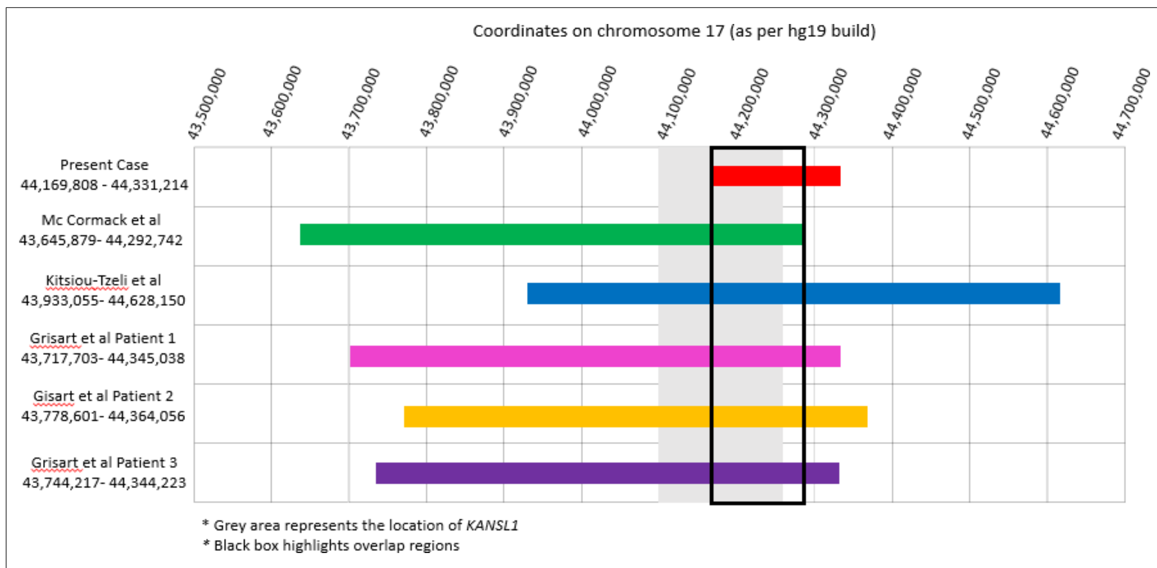


Figure 4-3 Visual representation of the overlap between the reported 17q21.31 duplications and the duplication identified within Family 3.

All of the previously reported duplications were either de novo or inherited in a dominant fashion both maternally and paternally. Although this may therefore discard 17q21.31 microduplication syndrome as the cause of the disorder within this family (as neither of the parents seem to be affected), it is important to note that 17q21.31 microduplication syndrome can show variable expressivity within the same family. It may therefore be possible that either of the parents may harbour the duplication but have a less severe phenotype. To explore this, further genetic studies could be carried out on all the affected individuals as well as unaffected siblings and parents to determine the exact break points of the duplication and determine if any unaffected individuals may also harbour the duplication. In the event that any of the "unaffected" individuals do have the duplication, more in depth clinical evaluations would be able to show if they have a milder phenotype and therefore either confirm the diagnosis or discard the duplication as a candidate for the disorder.

Of the homozygous variants identified (Table 3-8) none of them are known to

cause ID. Three of the variants (within the genes *DHX38, KIAA0355* and *SALL3*) were predicted deleterious by SIFT. Neither *DHX38* or *KIAA0355* are highly expressed within the brain, however *SALL3* is (Figure 4-4). *SALL3* is a zinc-finger protein and suspected transcription factor that has been shown to play a role in olfactory bulb dopaminergic neuron development by regulating tyrosine hydroxylase gene expression (Heng et al. 2012). This could therefore provide a possible mechanism by which a variant within this gene could result in an ID phenotype. Deletions and duplications within *SALL3* have been reported within DECIPHER with patients with ID. There is no allele frequency information available for the p.Ser182Trp variant, so it can be assumed that it is incredibly rare. Within gnomAD two different variants at this position are reported in heterozygous form at very low frequency. The serine at position 182 is also highly conserved, providing further evidence that a mutation at this position could be pathogenic.

Figure 4-4 Expression of SALL3 as per UCSC Genome Browser (Website: https://genome.ucsc.edu/cgi-bin/hgGateway)

The other homozygous variants identified that were not predicted to be deleterious by SIFT do not appear to be strong candidates for ID. *MYLPF* and *NUP93* show no expression within the brain and have no correlating ID phenotypes. Both ZNF681 and BEAN1 are expressed in the brain, however neither show a link to neurodevelopmental disorders.

The fourth family under investigation (Family 4, Section 3.4) entails three individuals presenting with severe psychomotor retardation, speech delay, failure to thrive, axial and peripheral hypotonia as well as being either wheelchair bound or can only crawl. All siblings also have varying degrees of microcephaly, and sensorineural hearing loss. Genome-wide SNP mapping identified four modest sized autozygous regions shared by all affected individuals within chromosomes 17, 7 and 12. Cross-referencing these autozygous regions with whole exome sequencing of one of the affected siblings failed to identify candidate homozygous variants within any of these regions, nor genome-wide. Candidate compound heterozygous were also not identified. The lack of definitive putative variants identified within this family, as well as in Family 3, may be due to methodological and technological limitations within this study. For example WES may not detect variants in regions in which there is low coverage. Additionally, the very nature of WES identifies sequence variants located within coding regions, so disease-causing variants located in non-coding regions (deep intronic, or intergenic regions) would not be identified. However our inability to (as yet) identify the causative variant(s) responsible for the disease in Family 4 may be due to other reasons.

Due to the consanguinity within the family, the perhaps higher than expected number of affected individuals (3 of 5 siblings being affected), and the lack of

candidate genomic variants identified by whole exome sequencing, it may be possible that other more complex genetic mechanisms may underlie the condition in this family. One possibility may be that there is more than one inherited disorder cosegregating within the family. However this may be unlikely given the closely overlapping clinical features of the affected individuals in this family. Examination for shared autozygous regions between pairs of siblings/ affected individuals may help find candidate regions of interest, alongside more in-depth clinical evaluation of all the affected individuals to ascertain the likelihood of one or more disorders. Another possibility is that a more complex mode of inheritance may be at play. Germline mosaicism may be possible, which may explain why no candidate homozygous/compound heterozygous variations were identified. Mosaicism arises when an individual carries more than one set of genetic information. In germline mosaicism, the gamete DNA differs to that of the somatic cells. To investigate this, future genetic studies could be undertaken to identify candidate deleterious variants in genes known to cause autosomal dominant forms of ID.

The SNP data identified a micro-duplication within 14q21.2 shared between the three siblings. This duplication is around 380kb in size and contains only one pseudogene (*KRT8P2*), so was assumed unlikely to be pathogenic. However, within DECIPHER v9.24 (website: https://decipher.sanger.ac.uk/), patient 331321 is reported to be heterozygous for a duplication on chromosome 14 which overlaps with the coordinates of the duplication found within Family 4 (Figure 4-5).

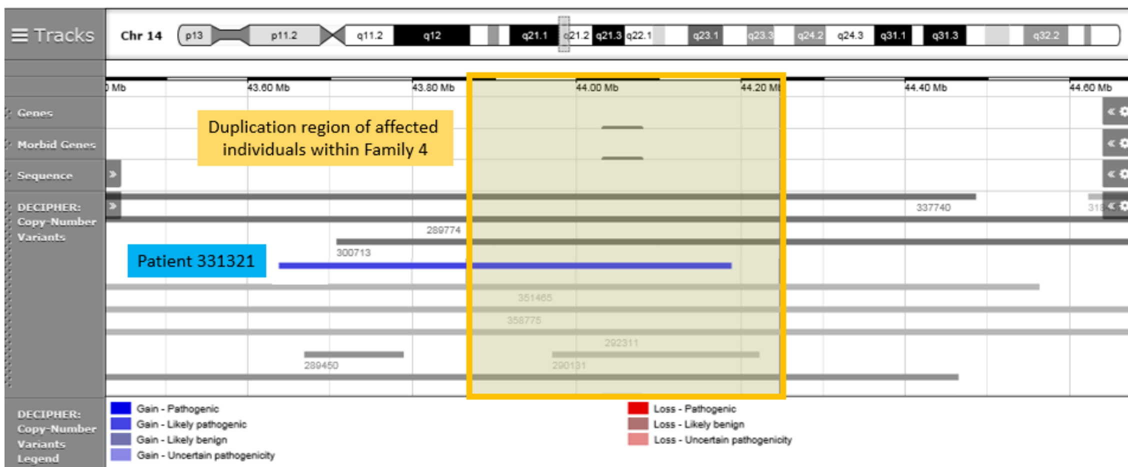


Figure 4-5 Chromosome 14 duplication location comparison between Family 4 and Patient 331321

Patient 331321 has microcephaly, facial abnormalities and intellectual disability, a phenotype similar to the affected individuals within Family 4. The duplication within patient 331321 was reported as 'likely pathogenic' by DECIPHER, and therefore requires further evidence to be conclusive. It was also reported that patient 331321 inherited the duplication maternally, yet the mother was reported to have no clinical signs of a disorder.

Chromosome 14 is known to be subjected to imprinting, with uniparental disomy (UPD) being reported on chromosome 14 (Kotzot and Utermann 2005) . Recently a duplication within 14q was reported in a patient born to a consanguineous marriage, in which the patient had mental retardation, growth retardation, dysmorphic features and muscle weakness (Sadr-Nabavi and Saeidi 2014). The exact duplication region of chromosome 14q was not specified, however it is likely that this duplication spanned a region containing important imprinting genes, such as the 14q32.2 imprinted region responsible for both maternal and paternal UPD-14 (Sutton and Shaffer 2000), which would therefore explain its pathogenic outcome. As the duplication region identified within this family doesn't span any known imprinting genes, it may be unlikely that this is the cause of the disorder.

5 FUTURE WORK

After highlighting potential candidate gene variants for ID, functional and cell studies are required to determine how any candidate mutations may affect protein structure and/or gene interactions resulting in the ID phenotypes described. Cell studies examining the functional effects of the *ARHGAP* variant identified through this work, as well as the previously reported variant, have begun at UEMS and with collaborative workers internationally. In particular mouse model studies aiming to introduce both the ARHGAP gene variant identified in family 2 as well as the distinct gene variant identified in our collaborator's family into the knock out mouse model, will provide insight into the phenotypical and molecular outcomes of each gene mutation.

Regarding families 3 and 4, in which no obvious variants within known ID genes were observed, recruitment of additional affected family members would enable most of the candidate variants to be excluded, and hopefully lead to the identification of a single primary candidate variant. In addition, it would be helpful to validate the presence of the possible duplications and specify the break points of the duplication regions proving them to be genuine and show whether or not they are present in unaffected members. To further ascertain if coding exons within genes located in the duplicate regions are disrupted, multiplex ligation-dependent probe amplification (MLPA) would confirm whether or not any exons are also duplicated.

Identifying the precise molecular cause of disease in a family is invaluable for the clinical care of the individual patient as well as enabling informed genetic counselling for the wider family. In the future it would be incredibly useful to create

a diagnostic screening panel of common founder pathogenic variants within populations, such as Nablus, to aid in the management of affected individuals and their families.

6 CONCLUSION

Through a community-based translational genetic medicine program based at the University of Exeter, four families from Nablus were recruited for genetic investigation. Each family consisted of at least two individuals with ID and a myriad of other clinical characteristics. Through genome-wide SNP mapping and whole exome sequencing, potential candidate genetic variants were identified.

A novel mutation within an *ARHGAP* gene was identified and found to cosegregate within one of the families. A different variant within a different transcript of this gene was recently implemented within neurodevelopmental disorders, furthering the belief that this novel mutation is pathogenic. Ongoing functional studies into the *ARHGAP* variant as well as mouse model studies will hopefully elucidate the biomechanical implications of the variant as well as draw phenotypical correlations.

A previously reported variant within *ALG1* was found to cosegregate within another family. Mutations within *ALG1* are known to cause a rare autosomal recessive congenital disorder of glycosylation (ALG1-CDG). ALG1-CDG is characterised by; microcephaly, developmental delay, abnormal fat distribution, strabismus, coagulation abnormalities, hypotonia, and failure to thrive. The three affected individuals within this family showed clinically similar characteristics to those of ALG1-CDG patients. Although the variant identified has been reported before, it is believed that this is the first time it has been reported in homozygous form. On-going clinical investigations will aid in confirming pathogenicity of the variant.

Two further families were found to harbor genetic duplications within one copy of a chromosome. One of these families possessed a duplication within 17q21.3, an area correlated with 17q21.31 duplication syndrome, the reciprocal Koolen- De Vries syndrome. Only a handful of patients with duplications within this region have been reported, all with similar clinical phenotypes to the affected individuals. Further research is required to reach a diagnosis within of the two families involved in this study, however through this collaborative program we have shown how important community-based research programmes are in defining the genetic causes of rare disorders.

7 APPENDICIES

7.1 Primer Sequences

Gene Name	Forward Primer	Reverse Primer
ALG1	tgctctatggcctctcacag	ctagacacccctcttccagc

8 BIBLIOGRAPHY

Alakbarzade, V., A. Hameed, D. Q. Quek, B. A. Chioza, E. L. Baple, A. Cazenave-Gassiot, L. N. Nguyen, M. R. Wenk, A. Q. Ahmad, A. Sreekantan-Nair, M. N. Weedon, P. Rich, M. A. Patton, T. T. Warner, D. L. Silver, and A. H. Crosby. 2015. 'A partially inactivating mutation in the sodium-dependent lysophosphatidylcholine transporter MFSD2A causes a non-lethal microcephaly syndrome', *Nat Genet*, 47: 814-7.

APA. 2013. "American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders"

Baple, E. L., H. Chambers, H. E. Cross, H. Fawcett, Y. Nakazawa, B. A. Chioza, G. V. Harlalka, S. Mansour, A. Sreekantan-Nair, M. A. Patton, M. Muggenthaler, P. Rich, K. Wagner, R. Coblentz, C. K. Stein, J. I. Last, A. M. Taylor, A. P. Jackson, T. Ogi, A. R. Lehmann, C. M. Green, and A. H. Crosby. 2014. 'Hypomorphic PCNA mutation underlies a human DNA repair disorder', *J Clin Invest*, 124: 3137-46.

Baple, E. L., R. Maroofian, B. A. Chioza, M. Izadi, H. E. Cross, S. Al-Turki, K. Barwick, A. Skrzypiec, R. Pawlak, K. Wagner, R. Coblentz, T. Zainy, M. A. Patton, S. Mansour, P. Rich, B. Qualmann, M. E. Hurles, M. M. Kessels, and A. H. Crosby. 2014. 'Mutations in KPTN cause macrocephaly, neurodevelopmental delay, and seizures', *Am J Hum Genet*, 94: 87-94.

Bittles, A H. 2001. 'Consanguinity and its relevance to clinical genetics', *Clin Genet*, 60: 89-98.

Blanchet, P., M. Bebin, S. Bruet, G. M. Cooper, M. L. Thompson, B. Duban-Bedu, B. Gerard, A. Piton, S. Suckno, C. Deshpande, V. Clowes, J. Vogt, P. Turnpenny, M. P. Williamson, Y. Alembik, Consortium Clinical Sequencing Exploratory Research Study, Consortium Deciphering Developmental Disorders, E. Glasgow, and A. McNeill. 2017. 'MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus', *PLoS Genet*, 13: e1006957.

Bowley, C., and M. Kerr. 2000. 'Epilepsy and intellectual disability', *Journal of Intellectual Disability Research*, 44: 529-43.

Bystron, I., C. Blakemore, and P. Rakic. 2008. 'Development of the human cerebral cortex: Boulder Committee revisited', *Nat Rev Neurosci*, 9: 110-22.

Chapman, M., P. Iddon, K. Atkinson, C. Brodie, D. Mitchell, G. Parvin, and S. Willis. 2011. 'The misdiagnosis of epilepsy in people with intellectual disabilities: a systematic review', *Seizure*, 20: 101-6.

Chiurazzi, P., and F. Pirozzi. 2016. 'Advances in understanding - genetic basis of intellectual disability', *F1000Res*, 5.

Cooper, S. 1997. 'High prevalence of dementia among people with learning disabilities not attributable to Down's syndrome', *Psychological Medicine*, 27: 609-16.

Cox, J., A. P. Jackson, J. Bond, and C. G. Woods. 2006. 'What primary microcephaly can tell us about brain growth', *Trends Mol Med*, 12: 358-66.

Dave, U., N. Shetty, and L. Mehta. 2005. 'A community genetics approach to population screening in India for mental retardation--a model for developing countries', *Ann Hum Biol*, 32: 195-203.

de Ligt, J., M. H. Willemsen, B. W. van Bon, T. Kleefstra, H. G. Yntema, T. Kroes, A. T. Vulto-van Silfhout, D. A. Koolen, P. de Vries, C. Gilissen, M. del Rosario, A. Hoischen, H. Scheffer, B. B. de Vries, H. G. Brunner, J. A. Veltman, and L. E.

Vissers. 2012. 'Diagnostic exome sequencing in persons with severe intellectual disability', *N Engl J Med*, 367: 1921-9.

De Rocker, N., S. Vergult, D. Koolen, E. Jacobs, A. Hoischen, S. Zeesman, B. Bang, F. Bena, N. Bockaert, E. M. Bongers, T. de Ravel, K. Devriendt, S. Giglio, L. Faivre, S. Joss, S. Maas, N. Marle, F. Novara, M. J. Nowaczyk, H. Peeters, A. Polstra, F. Roelens, C. Rosenberg, J. Thevenon, Z. Tumer, S. Vanhauwaert, K. Varvagiannis, A. Willaert, M. Willemsen, M. Willems, O. Zuffardi, P. Coucke, F. Speleman, E. E. Eichler, T. Kleefstra, and B. Menten. 2015. 'Refinement of the critical 2p25.3 deletion region: the role of MYT1L in intellectual disability and obesity', *Genet Med*, 17: 460-6.

do Carmo Avides, M., and D.M. and Glove. 1999. 'Abnormal Spindle Protein, Asp, and the Integrity of Mitotic Centrosomal Microtubule Organizing Centers', *Science*, 283: 1733-35.

Dovas, A., and J. R. Couchman. 2005. 'RhoGDI: multiple functions in the regulation of Rho family GTPase activities', *Biochem J*, 390: 1-9.

Dupre, T., S. Vuillaumier-Barrot, I. Chantret, H. Sadou Yaye, C. Le Bizec, A. Afenjar, C. Altuzarra, C. Barnerias, L. Burglen, P. de Lonlay, F. Feillet, S. Napuri, N. Seta, and S. E. Moore. 2010. 'Guanosine diphosphate-mannose:GlcNAc2-PP-dolichol mannosyltransferase deficiency (congenital disorders of glycosylation type Ik): five new patients and seven novel mutations', *J Med Genet*, 47: 729-35.

Emerson, E. , S. Baines, L. Allerton, and V and Welch. 2011. "Health Inequalities & People with Learning Disabilities in the UK." In, edited by Department of Health.

Faheem, M., M. I. Naseer, M. Rasool, A. G. Chaudhary, T. A. Kumosani, A. M. Ilyas, P. Pushparaj, F. Ahmed, H. A. Algahtani, M. H. Al-Qahtani, and H. Saleh Jamal. 2015. 'Molecular genetics of human primary microcephaly: an overview', *BMC Med Genomics*, 8 Suppl 1: S4.

First, M. B. 2009. 'Harmonisation of ICD-11 and DSM-V: opportunities and challenges', *Br J Psychiatry*, 195: 382-90.

Gelvin, James L. 2014. *The Israel-Palestine Conflict: One Hundred Years of War* (Cambridge University Press: 32 Avenue of the Americas, New York, USA).

Govek, E. E., S. E. Newey, and L. Van Aelst. 2005. 'The role of the Rho GTPases in neuronal development', *Genes Dev*, 19: 1-49.

Graves, J. A., J. Gecz, and H. Hameister. 2002. 'Evolution of the human X--a smart and sexy chromosome that controls speciation and development', *Cytogenet Genome Res*, 99: 141-5.

Grisart, B., L. Willatt, A. Destree, J. P. Fryns, K. Rack, T. de Ravel, J. Rosenfeld, J. R. Vermeesch, C. Verellen-Dumoulin, and R. Sandford. 2009. '17q21.31 microduplication patients are characterised by behavioural problems and poor social interaction', *J Med Genet*, 46: 524-30.

Grunewald, S. 2002. 'Congenital Disorders of Glycosylation: A Review', *Pediatric Research*, 52: 618-24.

Gund, C., Z. Powis, W. Alcaraz, S. Desai, and K. Baranano. 2016. 'Identification of a syndrome comprising microcephaly and intellectual disability but not white matter disease associated with a homozygous c.676C>T p.R226W DEAF1 mutation', *Am J Med Genet A*, 170A: 1330-2.

Harris, James C. 2006. *Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment.*

Harshman, L. A., B. G. Ng, H. H. Freeze, P. Trapane, A. Dolezal, P. D. Brophy, and J. E. Brumbaugh. 2016. 'Congenital nephrotic syndrome in an infant with ALG1-congenital disorder of glycosylation', *Pediatr Int*, 58: 785-8.

Heng, X., H. Breer, X. Zhang, Y. Tang, J. Li, S. Zhang, and W. Le. 2012. 'Sall3 correlates with the expression of TH in mouse olfactory bulb', *J Mol Neurosci*, 46: 293-302.

Hu, H., S. A. Haas, J. Chelly, H. Van Esch, M. Raynaud, A. P. de Brouwer, S. Weinert, G. Froyen, S. G. Frints, F. Laumonnier, T. Zemojtel, M. I. Love, H. Richard, A. K. Emde, M. Bienek, C. Jensen, M. Hambrock, U. Fischer, C. Langnick, M. Feldkamp, W. Wissink-Lindhout, N. Lebrun, L. Castelnau, J. Rucci, R. Montjean, O. Dorseuil, P. Billuart, T. Stuhlmann, M. Shaw, M. A. Corbett, A. Gardner, S. Willis-Owen, C. Tan, K. L. Friend, S. Belet, K. E. van Roozendaal, M. Jimenez-Pocquet, M. P. Moizard, N. Ronce, R. Sun, S. O'Keeffe, R. Chenna, A. van Bommel, J. Goke, A. Hackett, M. Field, L. Christie, J. Boyle, E. Haan, J. Nelson, G. Turner, G. Baynam, G. Gillessen-Kaesbach, U. Muller, D. Steinberger, B. Budny, M. Badura-Stronka, A. Latos-Bielenska, L. B. Ousager, P. Wieacker, G. Rodriguez Criado, M. L. Bondeson, G. Anneren, A. Dufke, M. Cohen, L. Van Maldergem, C. Vincent-Delorme, B. Echenne, B. Simon-Bouy, T. Kleefstra, M. Willemsen, J. P. Fryns, K. Devriendt, R. Ullmann, M. Vingron, K. Wrogemann, T. F. Wienker, A. Tzschach, H. van Bokhoven, J. Gecz, T. J. Jentsch, W. Chen, H. H. Ropers, and V. M. Kalscheuer. 2016. 'X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes', *Mol Psychiatry*, 21: 133-48.

Hung, L. Y., H. L. Chen, C. W. Chang, B. R. Li, and T. K. Tang. 2004. 'Identification of a novel microtubule-destabilizing motif in CPAP that binds to tubulin heterodimers and inhibits microtubule assembly', *Mol Biol Cell*, 15: 2697-706.

J. J. Polder, W. J. Meerding, L. Bonneux, and and P. J. van der Maas. 2002. 'Healthcare costs of intellectual disability in the Netherlands: a cost-of-illness perspective', *Journal of Intellectual Disability Research*, 46: 168–78.

Ji, J., H. Lee, B. Argiropoulos, N. Dorrani, J. Mann, J. A. Martinez-Agosto, N. Gomez-Ospina, N. Gallant, J. A. Bernstein, L. Hudgins, L. Slattery, B. Isidor, C. Le Caignec, A. David, E. Obersztyn, B. Wisniowiecka-Kowalnik, M. Fox, J. L. Deignan, E. Vilain, E. Hendricks, M. Horton Harr, S. E. Noon, J. R. Jackson, A. Wilkens, G. Mirzaa, N. Salamon, J. Abramson, E. H. Zackai, I. Krantz, A. M. Innes, S. F. Nelson, W. W. Grody, and F. Quintero-Rivera. 2015. 'DYRK1A haploinsufficiency causes a new recognizable syndrome with microcephaly, intellectual disability, speech impairment, and distinct facies', *Eur J Hum Genet*, 23: 1473-81.

Johnson, W., A. Carothers, and I. Deary. 2008. 'Sex Differences in Variability in General Intelligence?', *Perspectives on Psychological Science*, 3: 518-31

Kacowicz, A.M., and P. Lutomski. 2007. *Population Resettlement in International Conflicts: A Comparative Study* (Lexington Books).

Kaufmann, W.E, and H.W. Moser. 2000. 'Dendritic anomalies in disorders associated with mental retardation', *Cereb Cortex*, 10: 981-91.

Kelley, R. I., D. Robinson, E. G. Puffenberger, K. A. Strauss, and D. H. Morton. 2002. 'Amish lethal microcephaly: a new metabolic disorder with severe congenital microcephaly and 2-ketoglutaric aciduria', *Am J Med Genet*, 112: 318-26.

Khalailah A, Abu-Diab A, Ben-Yosef T, Raas-Rothschild A, Lerer I, Alswaiti Y, Chowers I, Banin E, Sharon D, and Khateb S. 2018. 'The Genetics of Usher

Syndrome in the Israeli and Palestinian Populations.', *Invest Ophthalmol Vis Sci*, 59: 1095-104.
Kirchhoff, M., A. M. Bisgaard, M. Duno, F. J. Hansen, and M. Schwartz. 2007. 'A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features', *Eur J Med Genet*, 50: 256-63.
Kitsiou-Tzeli, S., H. Frysira, K. Giannikou, A. Syrmou, K. Kosma, G. Kakourou, E. Leze, C. Sofocleous, E. Kanavakis, and M. Tzetis. 2012. 'Microdeletion and microduplication 17q21.31 plus an additional CNV, in patients with intellectual disability, identified by array-CGH', *Gene*, 492: 319-24.
Knapp, M., R. Romeo, and J. Beecham. 2009. 'Economic cost of autism in the UK', *Autism*, 13: 317-36.
Kochinke, K., C. Zweier, B. Nijhof, M. Fenckova, P. Cizek, F. Honti, S. Keerthikumar, M. A. Oortveld, T. Kleefstra, J. M. Kramer, C. Webber, M. A. Huynen, and A. Schenck. 2016. 'Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules', *Am J Hum Genet*, 98: 149-64.
Koolen, D. A., J. M. Kramer, K. Neveling, W. M. Nillesen, H. L. Moore-Barton, F. V. Elmslie, A. Toutain, J. Amiel, V. Malan, A. C. Tsai, S. W. Cheung, C. Gilissen, E. T. Verwiel, S. Martens, T. Feuth, E. M. Bongers, P. de Vries, H. Scheffer, L. E. Vissers, A. P. de Brouwer, H. G. Brunner, J. A. Veltman, A. Schenck, H. G. Yntema, and B. B. de Vries. 2012. 'Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome', *Nat Genet*, 44: 639-41.
Koolen, D. A., A. J. Sharp, J. A. Hurst, H. V. Firth, S. J. Knight, A. Goldenberg, P. Saugier-Veber, R. Pfundt, L. E. Vissers, A. Destree, B. Grisart, L. Rooms, N. Van der Aa, M. Field, A. Hackett, K. Bell, M. J. Nowaczyk, G. M. Mancini, P. J. Poddighe, C. E. Schwartz, E. Rossi, M. De Gregori, L. L. Antonacci-Fulton, M. D. McLellan, 2nd, J. M. Garrett, M. A. Wiechert, T. L. Miner, S. Crosby, R. Ciccone, L. Willatt, A. Rauch, M. Zenker, S. Aradhya, M. A. Manning, T. M. Strom, J. Wagenstaller, A. C. Krepischi-Santos, A. M. Vianna-Morgante, C. Rosenberg, S. M. Price, H. Stewart, C. Shaw-Smith, H. G. Brunner, A. O. Wilkie, J. A. Veltman, O. Zuffardi, E. E. Eichler, and B. B. de Vries. 2008. 'Clinical and molecular delineation of the 17q21.31 microdeletion syndrome', *J Med Genet*, 45: 710-20.
Koolen, D. A., L. E. Vissers, R. Pfundt, N. de Leeuw, S. J. Knight, R. Regan, R. F. Kooy, E. Reyniers, C. Romano, M. Fichera, A. Schinzel, A. Baumer, B. M. Anderlid, J. Schoumans, N. V. Knoers, A. G. van Kessel, E. A. Sistermans, J. A. Veltman, H. G. Brunner, and B. B. de Vries. 2006. 'A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism', *Nat Genet*, 38: 999-1001.
Kotzot, D., and G. Utermann. 2005. 'Uniparental disomy (UPD) other than 15: phenotypes and bibliography updated', *Am J Med Genet A*, 136: 287-305.
Leshinsky-Silver, E., J. Ling, J. Wu, C. Vinkler, K. Yosovich, S. Bahar, M. Yanoov-Sharav, T. Lerman-Sagie, and D. Lev. 2017. 'Severe growth deficiency, microcephaly, intellectual disability, and characteristic facial features are due to a homozygous QARS mutation', *Neurogenetics*, 18: 141-46.
Lhatoo, S.D., and J. Sander. 2001. 'The Epidemiology of Epilepsy and Learning Disability', *Epilepsia*, 42: 6-9.

Lubs, H. A., R. E. Stevenson, and C. E. Schwartz. 2012. 'Fragile X and X-linked intellectual disability: four decades of discovery', *Am J Hum Genet*, 90: 579-90.
Mc Cormack, A., J. Taylor, L. Te Weehi, D. R. Love, and A. M. George. 2014. 'A case of 17q21.31 microduplication and 7q31.33 microdeletion, associated with developmental delay, microcephaly, and mild dysmorphic features', *Case Rep Genet*, 2014: 658570.
Mirzaa, G. M., and A. R. Paciorkowski. 2014. 'Introduction: Brain malformations', *Am J Med Genet C Semin Med Genet*, 166C: 117-23.
Mizutani, R., K. Nakamura, N. Kato, K. Aizawa, Y. Miyamoto, T. Torii, J. Yamauchi, and A. Tanoue. 2012. 'Expression of sorting nexin 12 is regulated in developing cerebral cortical neurons', *J Neurosci Res*, 90: 721-31.
Mizutani, R., K. Nakamura, S. Yokoyama, A. Sanbe, S. Kusakawa, Y. Miyamoto, T. Torii, H. Asahara, H. Okado, J. Yamauchi, and A. Tanoue. 2011. 'Developmental expression of sorting nexin 3 in the mouse central nervous system', *Gene Expr Patterns*, 11: 33-40.
Morava., Eva, Julia Vodopiutz., Dirk J. Lefeber., Andreas R. Janecke., Wolfgang M. Schmidt., Silvia Lechner., Chike B. Item., Jolanta Sykut-Cegielska., Maciej Adamowicz., Jolanta Wierzba., Zong H. Zhang., Ivana Mihalek., Sylvia Stockler., Olaf A. Bodamer., Ludwig Lehle., and Ron A. Wevers. 2012. 'Defining the Phenotype in Congenital Disorder of Glycosylation Due to ALG1 Mutations', *American Academy of Pediatrics*, 130: e1034-9.
Muirhead, G., and K. K. Dev. 2014. 'The expression of neuronal sorting nexin 8 (SNX8) exacerbates abnormal cholesterol levels', *J Mol Neurosci*, 53: 125-34.
Nakazawa, S., N. Gotoh, H. Matsumoto, C. Murayama, T. Suzuki, and T. Yamamoto. 2011. 'Expression of sorting nexin 18 (SNX18) is dynamically regulated in developing spinal motor neurons', *J Histochem Cytochem*, 59: 202-13.
Natacci, F., E. Alfei, L. Tarara, S. D'Arrigo, O. Zuffardi, B. Gentilin, and C. Pantaleoni. 2016. 'Chromosome 17q21.31 duplication syndrome: Description of a new familiar case and further delineation of the clinical spectrum', *Eur J Paediatr Neurol*, 20: 183-7.
Ng, B. G., S. A. Shiryaev, D. Rymen, E. A. Eklund, K. Raymond, M. Kircher, J. E. Abdenur, F. Alehan, A. T. Midro, M. J. Bamshad, R. Barone, G. T. Berry, J. E. Brumbaugh, K. J. Buckingham, K. Clarkson, F. S. Cole, S. O'Connor, G. M. Cooper, R. Van Coster, L. A. Demmer, L. Diogo, A. J. Fay, C. Ficicioglu, A. Fiumara, W. A. Gahl, R. Ganetzky, H. Goel, L. A. Harshman, M. He, J. Jaeken, P. M. James, D. Katz, L. Keldermans, M. Kibaek, A. J. Kornberg, K. Lachlan, C. Lam, J. Yaplito-Lee, D. A. Nickerson, H. L. Peters, V. Race, L. Regal, J. S. Rush, S. L. Rutledge, J. Shendure, E. Souche, S. E. Sparks, P. Trapane, A. Sanchez-Valle, E. Vilain, A. Vollo, C. J. Waechter, R. Y. Wang, L. A. Wolfe, D. A. Wong, T. Wood, A. C. Yang, Genomics University of Washington Center for Mendelian, G. Matthijs, and H. H. Freeze. 2016. 'ALG1-CDG: Clinical and Molecular Characterization of 39 Unreported Patients', *Hum Mutat*, 37: 653-60.
Pajusalu, S., T. Reimand, and K. Ounap. 2015. 'Novel homozygous mutation in KPTN gene causing a familial intellectual disability-macrocephaly syndrome', *Am J Med Genet A*, 167A: 1913-5.
Pavlowsky, A., J. Chelly, and P. Billuart. 2012. 'Emerging major synaptic signaling pathways involved in intellectual disability', *Mol Psychiatry*, 17: 682-93.
Petersen, L. R., D. J. Jamieson, A. M. Powers, and M. A. Honein. 2016. 'Zika Virus', *N Engl J Med*, 374: 1552-63.

Piton, A., C. Redin, and J. L. Mandel. 2013. 'XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing', *Am J Hum Genet*, 93: 368-83.
Rakic, P. 1988. 'Specification of Cerebral Cortical Areas', *Science*, 241: 170-76.
Reijnders, M. R. F., N. M. Ansor, M. Kousi, W. W. Yue, P. L. Tan, K. Clarkson, J. Clayton-Smith, K. Corning, J. R. Jones, W. W. K. Lam, G. M. S. Mancini, C. Marcelis, S. Mohammed, R. Pfundt, M. Roifman, R. Cohn, D. Chitayat, Study Deciphering Developmental Disorders, T. H. Millard, N. Katsanis, H. G. Brunner, and S. Banka. 2017. 'RAC1 Missense Mutations in Developmental Disorders with Diverse Phenotypes', *Am J Hum Genet*, 101: 466-77.
Sadr-Nabavi, A., and M. Saeidi. 2014. 'Chromosome duplication (14q) and the genotype phenotype correlation', *Int J Fertil Steril*, 8: 95-98.
Stiles, J., and T. L. Jernigan. 2010. 'The basics of brain development', *Neuropsychol Rev*, 20: 327-48.
Straub, J., E. D. H. Konrad, J. Gruner, A. Toutain, L. A. Bok, M. T. Cho, H. P. Crawford, H. Dubbs, G. Douglas, R. Jobling, D. Johnson, B. Krock, M. A. Mikati, A. Nesbitt, J. Nicolai, M. Phillips, A. Poduri, X. R. Ortiz-Gonzalez, Z. Powis, A. Santani, L. Smith, A. P. A. Stegmann, C. Stumpel, M. Vreeburg, Study Deciphering Developmental Disorders, A. Fliedner, A. Gregor, H. Sticht, and C. Zweier. 2018. 'Missense Variants in RHOBTB2 Cause a Developmental and Epileptic Encephalopathy in Humans, and Altered Levels Cause Neurological Defects in Drosophila', *Am J Hum Genet*, 102: 44-57.
Strydom, A., G. Livingston, M. King, and A. Hassiotis. 2007. 'Prevalence of dementia in intellectual disability using different diagnostic criteria', *The British Journal of Psychiatry*, 191: 150-57.
Strydom, A., R. Romeo, N. Perez-Achiaga, G. Livingston, M. King, M. Knapp, and A. Hassiotis. 2010. 'Service use and cost of mental disorder in older adults with intellectual disability', *Br J Psychiatry*, 196: 133-8.
Sutton, V., and L. Shaffer. 2000. 'Search for Imprinted Regions on Chromosome 14: Comparison of Maternal and Paternal UPD Cases with Cases of Chromosome 14 Deletion', *Am J Med Genet*, 93: 381-87.
Takenouchi, T., R. Kosaki, T. Niizuma, K. Hata, and K. Kosaki. 2015. 'Macrothrombocytopenia and developmental delay with a de novo CDC42 mutation: Yet another locus for thrombocytopenia and developmental delay.', *Am J Med Genet*, 167: 2822-5.
Tanaka, A. J., M. T. Cho, F. Millan, J. Juusola, K. Retterer, C. Joshi, D. Niyazov, A. Garnica, E. Gratz, M. Deardorff, A. Wilkins, X. Ortiz-Gonzalez, K. Mathews, K. Panzer, E. Brilstra, K. L. van Gassen, C. M. Volker-Touw, E. van Binsbergen, N. Sobreira, A. Hamosh, D. McKnight, K. G. Monaghan, and W. K. Chung. 2015. 'Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss', *Am J Hum Genet*, 97: 457-64.
Teeuw, M. E., G. Loukili, E. A. Bartels, L. P. ten Kate, M. C. Cornel, and L. Henneman. 2014. 'Consanguineous marriage and reproductive risk: attitudes and understanding of ethnic groups practising consanguinity in Western society', *Eur J Hum Genet*, 22: 452-7.
Ulrich Zechner, Monika Wilda, Hildegard Kehrer-Sawatzki, Walther Vogel, Rainald Fundele, and and Horst Hameister. 2001. 'A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution?', *Trends in Genetics*, 17: 697-701.
Vannest, K.J. 2014. "AAIDD (American Association on Intellectual Developmental Disabilities) Classification Systems." In *Encyclopedia of Special Education*.

Vinkler, C., E. Leshinsky-Silver, M. Michelson, D. Haas, T. Lerman-Sagie, and D. Lev. 2014. 'A newly recognized syndrome of severe growth deficiency, microcephaly, intellectual disability, and characteristic facial features', *Eur J Med Genet*, 57: 288-92.

Wang, X., Y. Zhao, X. Zhang, H. Badie, Y. Zhou, Y. Mu, L. S. Loo, L. Cai, R. C. Thompson, B. Yang, Y. Chen, P. F. Johnson, C. Wu, G. Bu, W. C. Mobley, D. Zhang, F. H. Gage, B. Ranscht, Y. W. Zhang, S. A. Lipton, W. Hong, and H. Xu. 2013. 'Loss of sorting nexin 27 contributes to excitatory synaptic dysfunction by modulating glutamate receptor recycling in Down's syndrome', *Nat Med*, 19: 473-80.

Watemberg, N., S. Silver, S. Harel, and T and Lerman-Sagie. 2002. 'Significance of microcephaly among children with developmental disabilities', *Journal of Child Neurology*, 17: 117-22.

Zlotogora, Joel. 1997. 'Genetic Disorders Among Palestinian Arabs: 1. Effects of Consanguinity', *American Journal of Medical Genetics*, 68: 472-75.

Zollino, M., D. Orteschi, M. Murdolo, S. Lattante, D. Battaglia, C. Stefanini, E. Mercuri, P. Chiurazzi, G. Neri, and G. Marangi. 2012. 'Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype', *Nat Genet*, 44: 636-8.