Gene discovery and cellular modelling of rare autosomal recessive neurodevelopmental conditions

Marilena Elpidorou

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Leeds Institute of Medical Research
Faculty of Medicine and Health

September 2019

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

The past four years have been an amazing learning experience that would not have never been the same without my three supervisors, Professor Colin A Johnson, Professor Eamonn G Sheridan and Dr James A Poulter, who I want to deeply acknowledge. The first big thank you goes to Colin, who has always been a very supportive supervisor, always giving constructive criticism and guidance to help me grow into an independent researcher. The second big thank you goes to Eamonn, who apart from his scientific support, I am also grateful for his entertaining distractions from work, like teaching me how to read an MRI scan, arguing about genes and of course football. Last but not least, my dear friend James who offered endless help and encouragement with anything I could possibly ask, and for that I am deeply indebted. He was an extremely motivating supervisor, always being positive and very patient in several instances, like teaching me bioinformatics or listening to me complaining when an experiment had failed. I am forever grateful for having these three gentlemen as my mentors and I wouldn't change a thing.

I would also like to express my gratitude to all the members of Team Meckel, who made the past four years a real pleasure. A special thanks goes to my friend Kasia for being my lab mentor and possibly the best teacher someone could have during this learning experience. I would also like to thank all the people on level 8 and all of our project collaborators.

I am also incredibly thankful for the limitless support of my family throughout this bumpy ride of completing a PhD. Firstly, my parents Costas and Nicoletta for always being there for me and for everything they have done to help me achieve my goals. Secondly, my siblings Pavlina and Thoukis for their love and for visiting me in Leeds during these years. Also, the love of my grandma Maroulla and the support of my best friends in Cyprus were treasured during this time. Last but not least, a huge thank you goes to my boyfriend George for his love and support and for always putting a smile on my face, even during the most stressful periods.

This thesis and the award of Doctor of Philosophy is dedicated to my two grandfathers, Thoukidides and Pavlos, who have sadly passed away in 2019 while my thesis was being written. You are both deeply missed every single day and I am sure you feel extremely proud for me right now. You both wanted to call your little granddaughter Dr Elpidorou
and I am sure you are both smiling down on me. Love you always and forever.

Abstract

Neurodevelopmental disorders (NDD) are a group of heterogenous conditions characterised by global developmental delay with additional neurological defects. While some NDDs display non-specific brain phenotypes, some NDDs may have particular distinguishing neurological phenotypes, such as agenesis of the corpus callosum or hypomyelination. This project aimed to delineate novel genetic causes of neurodevelopmental conditions using next generation sequencing (NGS) technologies in a family-based approach.

As part of this study, thirteen families were recruited, with at least three individuals per family sequenced using whole exome sequencing (WES). Standardized bioinformatics pipeline analysis was carried out, prioritising variants based on segregation, *in silico* pathogenicity prediction tools and biological relevance. Known and candidate genes were successfully identified in the majority of the cases.

A nonsense variant was identified in *HERC2* in a family with severe global developmental delay. This finding provided a molecular diagnosis for this patient and expanded the known phenotype-genotype correlation associated with *HERC2* mutations. In addition, a *de novo* variant in *TUBA1A* was discovered in a family with a clinical diagnosis of a Complex Moebius syndrome and perisylvian polymicrogyria and slight callosal dysmorphism. This finding further expanded the known phenotype caused by *TUBA1A* mutations. This study was able to establish new genotype-phenotype correlations that will be beneficial for future clinical diagnosis and patient care.

Novel genetic causations for NDDs were established for the first time by the discovery of mutations in two disease genes. A nonsense variant in *KLHL7* was identified as a cause of NDD. Specifically, the affected individuals presented with a phenotype similar to 4H syndrome with additional features of myopathy, stoke-like episodes, microcephaly and abnormal sweating. The second genetic discovery was a missense variant in *MAL* identified in a family with hypomyelinating leukodystrophy similar to Pelizaeus-Merzbacher disease.

The variants in *HERC2*, *KLHL7* and *MAL* were modelled *in vitro* and functionally characterized as pathogenic by using cellular and biochemical approaches. The functional studies have provided novel insights into the disease mechanisms of NDD. For instance, the functional characterization of the variant in *MAL* determined its role in the process of myelination. This finding expands our current knowledge of the

mechanisms and proteins involved in myelin development. In the post-genomic era, it is crucial to characterise variants identified in rare autosomal recessive conditions, as they offer us a unique opportunity to gain further understanding of disease mechanisms and biological processes that would otherwise remain ambiguous.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The advancement of gene discovery techniques

This study is focused on novel gene discovery of rare autosomal recessive neurodevelopmental disorders. Gene discovery is a process by which pathogenic or causative variants in certain genes are identified and linked to a particular phenotypic trait or disease. The process of gene discovery has been revolutionized by the recent introduction of next generation sequencing, with techniques that have accelerated the process of disease gene identification. Part of this study concentrates on using next generation techniques to identify novel variants in rare autosomal recessive neurodevelopmental conditions, with the majority of cases being from consanguineous families. It is known that there is an increased probability for the causative variant to be segregating within an autozygous region in consanguineous families (Alkuraya, 2012). Autozygosity mapping is a linkage analysis technique that was used in conjunction with Sanger sequencing, to establish regions where recessive variants could be located. The advancement of gene discovery techniques along with the use of older methods, such as autozygosity mapping, has been utilized in this project to help identify the causative variants.

1.1.1 Autozygosity mapping

Autozygosity mapping, also known as homozygosity mapping, is a key method for identifying autosomal recessive variants in consanguineous families by identifying homozygous haplotypes where the disease loci will be. Even though autosomal recessive conditions do not necessarily arise from consanguineous unions, the incidence of autosomal recessive traits is significantly increased in populations where these unions are common. For instance, 6% (1/16) of the genome of the offspring of a first cousin union is predicted to be homozygous, by inheriting genomic segments from their related parents. Accessing the incidence of homozygosity in the offsprings of first-cousin unions, Woods et al found that the homozygous regions composed of 11% of the genome instead of 6%, highlighting the increased probability of a disease allele segregating in these homozygous haplotypes (Woods et al., 2006). This phenomenon is known as autozygosity, where the same haplotype is inherited from the parental genomes. In

relation to disease, deleterious alleles may be segregating in an autozygous pattern, where the affected individual carries two deleterious recessive alleles that have segregated from the same ancestral haplotype as a result of consanguinity. The disease alleles that share a common ancestral origin are also known as identical-by-descent (IBD) alleles (Johnson, 2012). Identical-by-state (IBS) alleles exist as well, where certain part of the genome is homozygous, potentially carrying a recessive disease variant, but the haplotypes were inherited from unrelated parents (Powell et al., 2010). An example of autozygosity in a first cousin union is summarised below.

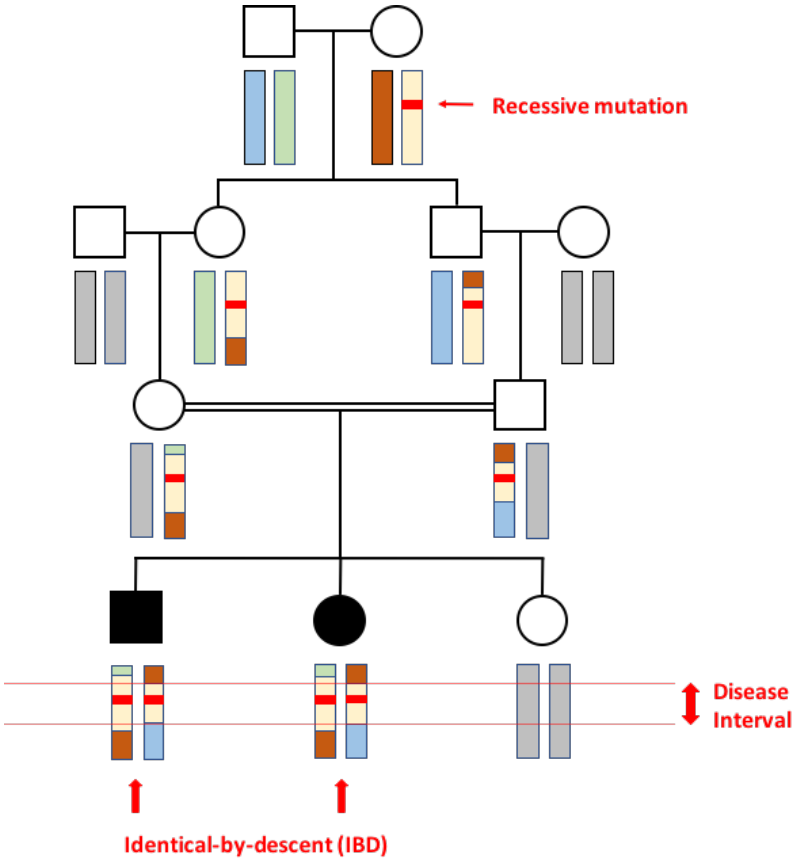


Figure 1-1 Autozygosity mapping

The above pedigree of a first cousin consanguineous family represents the principle of autozygosity. The great-grandmother is a carrier of a pathogenic recessive mutation (indicated in red), which is inherited to both of her children, along with the chromosomal haplotype surrounding it. Even though the gene is transmitted from generation to generation, the chromosomal region around it co-segregates but can decrease in size as a result of meiotic recombination. The affected individuals in generation IV carry both copies of the recessive mutation, and are also homozygous for the haplotype surrounding the disease gene. The minimal region that is inferred to contain the mutated gene is indicated on the diagram by “disease interval”.

This idea was first proposed by Sir Archibald Garrod in 1902, when he demonstrated that the incidence of patients affected with autosomal recessive conditions, such as alkaptonuria, is higher in consanguineous marriages of first cousins (Garrod, 1902).

Taking this idea further, William Bateson explained at the Evolution Committee of the Royal Society in 1902 that the mating of first cousins could significantly increase the frequency of homozygous individuals who will manifest a recessive condition. He described the genetic explanation of this as homozygosity-by-decent, another term for autozygosity. In 1953, Cedric Smith described the fundamental hypothesis that an affected child from a consanguineous mating will not only be homozygous for a specific gene, but will also be homozygous for the genetic markers surrounding this gene as shown in the diagram above (Johnson, 2012). From this hypothesis, Smith suggested that these offspring can be used for mapping disease loci by genetic linkage, and thereby map the gene of interest. However, the means of doing such a study were unavailable at that time (Mueller and Bishop, 1993).

1.1.1.1 Restriction Fragment Length Polymorphism

All of the previous innovative ideas on autozygosity mapping offered the key to Lander and Botstein in 1987 to propose a way in which the linkage study described by Smith could actually be feasible. Their idea was based on Restriction Fragment Length Polymorphisms (RFLP), and it relied on the fact that an affected offspring of a consanguineous mating will be homozygous for a chromosomal region across a discrete genetic distance around the disease locus that will be shared with other affected members of the family. Searching for these identical-by-decent (IBD) regions would provide a powerful tool for mapping a recessive gene.

Lander and Botstein proposed a mathematical model to predict the autozygosity. Firstly, the coefficient of inbreeding (*F*) was introduced and defined as the likelihood that a particular genetic locus will be homozygous-by-descent or IBD. For example, the value of *F* for siblings, first cousin and second cousin matings are expected to be 1/4, 1/16, and 1/64 respectively. Secondly, it was suggested that if a recessive disease allele occurs at a frequency *q* (assuming that the alleles are in Hardy-Weinberg equilibrium within a population), the probability that the particular disease locus being IBD is given by *Fq*. In addition, the probability of an affected offspring not being IBD is given by $(1-F)q^2$, and therefore the overall probability (α) that a disease locus in an affected individual is IBD is given by $\alpha = Fq/[Fq + (1-F)q^2]$.

Furthermore, the odds ratio (P1:P2) known as the LOD score was introduced, where P1 was defined as the probability that the disease locus is IBD, and P2 the probability that the disease locus is unlinked. The LOD score is an indication (based on statistical estimates) of the proximity of two genes, or a genetic marker and a disease gene on a

particular chromosome, hence an indication of inheritance used for autozygosity mapping. Based on LOD scores, Lander and Green introduce an algorithm to assist with the genetic analysis on linkage studies. They proposed a computer program called HOMMAP, which allowed quick analysis of sibling and cousin marriages (Lander and Botstein, 1987).

1.1.1.2 Microsatellite markers

Autozygosity mapping provided a novel way for gene localisation, particularly in consanguineous families where genetic heterogeneity is lower than non-consanguineous families. However, the advances in autozygosity mapping in the following years allowed rapid progress in gene identification. Utilizing microsatellite markers for autozygosity mapping in genome-wide linkage studies was the next step forward in the field of genetic screening and gene discovery (Johnson, 2012).

Microsatellites are di-, tri- or tetra nucleotide tandem repeats of genomic sequences that are used for mapping genes of interest as well as unravelling inheritance patterns. The frequency of repetition of a particular tandem repeat in a microsatellite can be extremely variable, a feature that makes microsatellites a very useful genetic marker. The variation in the number of repeats in a given case can indicate the dissimilarity between alleles, hence distinguishing a possible pattern of inheritance. The region around microsatellites can be amplified using fluorescently labeled primers and then the size of the DNA amplified can be determined for each individual in a family. Various different software programs exist that aid genotyping using microsatellite markers, such as ALLEGRO (Gudbjartsson et al., 2000) and MERLIN (Abecasis et al., 2002).

1.1.1.3 Shifting from microsatellite markers to SNPs

A single nucleotide polymorphism (SNP) is a type of genetic variation where a single nucleotide is altered. SNPs are the most common type of genetic variation, and arise on average once in every 300-400 nucleotides of the human genome. SNPs can occur anywhere in the genome with no pathogenic characteristics. However, it is possible that a SNP localised within an exon or in the regulatory region of a gene can have a key role in a disease (Woods et al., 2004).

Notably, SNPs can be utilised as genetic markers for gene identification, and have been proven a more successful tool for genotyping than microsatellite markers. Microsatellite markers are multiallelic and more informative compared to biallelic SNPs, but they are

highly mutable, thus making them very good markers for genetic linkage studies as allelic identity-by-decent can be easily determined. Despite that, the high mutability factor of microsatellites can be challenging to interpret, as PCR reactions with microsatellites tend to be problematic, and therefore SNPs are usually preferred as they can be easily optimised with reduced cost and time. In addition, high-density SNPs can provide more inheritance information, as SNPs are abundant in the human genome and can be readily genotyped using high-throughput techniques like SNP microarrays (Gunderson et al., 2004). High-density SNPs also allow genetic analysis of haplotypes, and are therefore very advantageous for autozygosity mapping.

1.1.2 Sequencing of nucleic acids

1.1.2.1 First Generation Sequencing

The first attempt to sequence nucleic acids was done on transfer RNA, as this was a single-stranded sequence of nucleic acids making it easier to attempt sequencing. In 1965, Robert Holley and colleagues succeeded in obtaining the first nucleic acid sequence, that of alanine, from tRNA from *Saccharomyces cerevisiae* by using selective ribonuclease treatments (Holley et al., 1965). In the 1970s, scientists worked to find the best technique to tackle double stranded nucleic acid sequences. Two protocols became available that utilize separation by polyacrylamide gel electrophoresis based on polynucleotide length. Alan Coulson and Fred Sanger developed the “plus and minus” system in 1975 (Sanger and Coulson, 1989), whereas Allan Maxam and Walter Gilbert proposed a chemical cleavage technique (Maxam and Gilbert, 1977). The “plus and minus” technique used DNA polymerase to synthesize a new DNA strand from a primer by performing a polymerization reaction and incorporating radiolabeled nucleotides. In a “plus” reaction, only a particular kind of nucleotide will be present and therefore end the extension of the stand with that nucleotide. In a “minus” reaction, all the other three types will be present and the sequence will continue until a “plus” reaction is achieved. In Maxam and Gilbert’s approach, radiolabeled DNA was treated with chemicals that will cleave the chain at certain bases, that was then run on the gel and based on the cleaved fragments the sequence was inferred (Heather and Chain, 2016).

The real breakthrough in sequencing happened in 1977, with Sanger’s chain-termination protocol. His techniques involved the use of chemical analogues of deoxyribonucleotides (dNTPs) known as dideoxyribonucleotides (ddNTPs), that lack the 3’ hydroxyl group and therefore unable to bond with the 5’ phosphate of the next dNTP, causing a chain termination. Sanger et al. mixed radiolabeled ddNTPs and performed four DNA

polymerase reactions, one for each nucleotide, and ran them on polyacrylamide gel (Sanger et al., 1977). Each radioactive band based on size corresponded to a specific nucleotide as the sequence was progressing. This technique has been used for many years now, with improvements made upon it such as removal of radiolabeling and replacement with fluorophores that can be detected by capillary based electrophoresis (Prober et al., 1987).

1.1.2.2 Next Generation Sequencing

1.1.2.2.1 Whole Exome Sequencing

As sequencing technologies advanced, several revolutionary methods were developed, known as next-generation sequencing (NGS), that allow high throughput or massively parallel sequencing, that can produce up to 6 billion reads. These provide a powerful tool for unravelling genetics variants that could possibly be pathogenic as well as translating scientific research into clinical diagnosis, especially for rare cases. NGS has significantly increased the pace of gene discovery while offering a hypothesis-free approach for genetic research.

The NGS era was revolutionized with Whole Exome Sequencing (WES), which is the technique used to sequence the coding region of the human genome (approx. 1.5% of the total). WES is achieved by initial DNA fragmentation of specific adaptors, followed by an enrichment capture approach whereby RNA libraries are utilised to capture all the coding sequence of the DNA (see section 2.2.10). Most commercial systems for NGS sequencing use the sequencing-by-synthesis approach, based on Sanger's protocol, whereby the machine detects fluorescent reversible terminators (Buermans and den Dunnen, 2014). Massive sequencing data sets are then analysed using bioinformatics and a case-specific pipeline to unravel any putative mutation (Buermans and den Dunnen, 2014).

Interestingly, with certain software packages, WES data can also be used in autozygosity mapping to search for autozygous regions in members of a consanguineous family (Carr et al., 2013). Variant call format (vcf) files, prior to SNP filtration steps, are mapped to the human genome and compared with other members of a family to highlight autozygous regions where deleterious mutations are more likely to be found. Software with capabilities to perform autozygosity mapping include SNPviewer (https://sourceforge.net/p/snpviewer/wiki/Getting%20Started/) and the in-house software AgileMultiIdeogram (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/).

1.1.2.2.2 Whole Genome Sequencing

NGS has further evolved into Whole Genome Sequencing (WGS), a technique that sequences every single nucleotide in the genome, providing a powerful tool for genomics research. WGS is superior to WES as it allows for large CNV to be picked up as well as non-coding pathogenic variants that otherwise would have been missed (Illumina Inc, 2019). WGS has successfully identified SNVs (Yuen et al., 2017) and rare CNVs related to neurodevelopment (Costain et al., 2019). As WGS lacks the exome capture step, it allows for even better coverage of the exomic regions compared to WES (Belkadi et al., 2015). Nevertheless, using WGS was much more expensive and more challenging than WES during the duration of this study. WGS has a number of limitations compared to WES. For instance, storage of large data files, lack of optimised analysis pipelines and shortage of fully comprehensive genome datasets for comparisons consisted of some key disadvantages of WGS. These issues are currently being overcome by the accumulation of more genomic datasets, such as gnomAD (Karczewski et al., 2019) and the 100k Genome Project (Genomic England). In addition, the cost of WGS has significantly decreased and the processing power enhanced, making it a more appealing approach for gene discovery nowadays. Furthermore, our increased understanding of the role of non-coding regions to biological processes and the effect of a mutation in intronic or promoter regions on a disease phenotype can be beneficial in variant interpretation (Cardoso et al., 2019, Lozano-Urena and Ferron, 2019).

1.2 Rare autosomal recessive disorders

Autosomal recessive disorders are defined as those that can manifest if the affected individual has inherited two copies of the same gene, one from each parent, carrying a deleterious variant. If the same variant is inherited from each parent, then the affected individual is considered homozygous for a particular condition, whereas the parents are heterozygous and referred to as carriers of this trait. Autosomal recessive disorders can also arise from compound heterozygosity, which is the inheritance of two different heterozygous recessive mutations in the same gene (one on each chromosome, described to be in *trans*) that can therefore cause the genetic disorder.

In some rare cases, autosomal recessive disorders can result from the inheritance of a heterozygous variant and the occurrence of a *de novo* mutation in the same gene (in *trans*), or from unipaternal disomy, where two copies of a chromosome come from the same parent carrying a pathogenic variant.

1.2.1 Rare autosomal recessive disorders in consanguineous communities

Most of the families studied in this report are from Bradford, West Yorkshire. Bradford is an ethnically diverse region of the Northern England, with a very large South Asian community, of which 90% of individuals are of Pakistani origin. As a cultural tradition, first cousin unions are very common in this community (37% of South Asian marriages) (Sheridan et al., 2013). For medical purposes, a consanguineous union is defined as a union between second degree cousins or closer. In these populations where the incidence of consanguinity is common, a higher rate of autosomal recessive diseases has been observed. This is explained by the increased probability of identity-by-descent for disease alleles.

In particular, the Born in Bradford study identified that the overall rate of birth defects in neonates was approximately 3%, a number that is almost double the average national rate (Sheridan et al., 2013). Therefore, Bradford is a hotspot for rare genetic conditions, as first cousin unions increase the prevalence of autosomal recessive conditions in this local community. This allows rare conditions to manifest through the inheritance of autozygous haplotype regions, which is a very powerful tool for researchers, as patients with rare diseases can provide unique insights into human biology and disease mechanisms.

1.2.2 Neuromuscular disorders

Neuromuscular disorder is a broad term that characterises several diseases that affect the functioning of the muscles. These diseases can arise from either a direct impairment of the voluntary muscle, or by an indirect damaging of the neuromuscular junction or peripheral nerves that results in impaired muscle function (Castro-Gago et al., 2014). Neuromuscular disorders can also be the result of central nervous system (CNS) damage. The symptoms of neuromuscular disorders fall within a wide range, with the most common ones being muscular weakness, muscle atrophy, and loss of muscular control (Engel et al., 1961). In addition, apart from some autoimmune cases, most neuromuscular disorders have a genetic cause that can manifest in both the direct and the indirect forms of neuromuscular disease. The disorders that will be discussed in this report include a Mitochondrial Myopathy, Congenital Muscular Dystrophy, Arthrogryposis, and Motor and Sensory Neuropathy.

1.2.3 Neurodevelopmental disorders

Neurodevelopmental disorders are characterised by growth and developmental impairments of the brain and the central nervous system. These diseases can have various causes, including genetic, metabolic, or immune defects. Specifically, the genetic background of a family with a 4H-like Syndrome is analysed. 4H Syndrome (MIM number 607694) stands for hypomyelination, hypogonadotropic, hypogonadism and hypodontia, which describes the essential clinical features that comprise global developmental delay, lack of myelin in the CNS, abnormal puberty development, and tooth hypoplasia (Orcesi et al., 2010). Specifically, the family discussed in this study (family ND12) is characterised by additional features such as progressive motor decline, ataxia, mild cognitive regression, microcephaly and stroke-like episodes, and excessive sweating. The family was further reassessed upon new genetic findings (Angius et al., 2016), and was thought to have a Crisponi-like syndrome. The causative variant for this case lies within an E3-ubiquitin ligase. Proteins associated with the ubiquitin-proteasome system are often associated with neurodevelopmental disorders and this will be further discussed below (see section 1.3.1).

1.2.4 Hypomyelinating Leukodystrophies

Hypomyelinating leukodystrophies (HLDs) are a group of extremely heterogenous neurodevelopmental disorders whereby the myelin sheath formation is disrupted leading to white matter abnormalities in the brain. The white matter defects are often classified

by different MRI patterns into different kinds of HLDs, ranging in severity and clinical manifestations. However, most HLDs are characterised by severe developmental delay, intellectual disability, spasticity, hypotonia and usually movement disorders (Charzewska et al., 2016).

The first reported familial case of white matter disorder was described by Pelizaeus and Merzbacher over a century ago, where they separately identified and described the occurrence of chronic progressive sclerotic hardening of white matter and lack of myelin (Merzbacher, 1909). This kind of hypomyelinating leukodystrophy subsequently received the medical eponym Pelizaeus-Merzbacher disease (PMD), and this study reports a potentially novel cause of this condition.

PMD is characterised by CNS hypomyelination, developmental delay, spasticity, hypotonia, ataxia, and intellectual disability. MRI scans of this leukodystrophy shows diffuse hypomyelination, dilation of the lateral ventricles and thinning of the corpus callosum. PMD is caused by *PLP1* mutations, where the affected proteolipid protein 1 (PLP1) comprises the main component of myelin sheath. Most of the variants reported are point mutations and they cause a clinically more severe form of PMD. Point mutations lead to protein misfolding and aggregation of PLP1 in the ER. There are also some null mutation and deletions that cause null PMD syndrome and spastic paraplegia type 2 (van der Knaap and Bugiani, 2017).

1.2.5 Clinical impact of gene identification in rare diseases

Neurodevelopmental disorders are extremely heterogeneous with a wide range of clinical symptoms and inheritance patterns. Identifying new genetic causes associated with neurodevelopmental conditions can significantly help in clinical diagnostics and patient care. This study focuses on the discovery of novel monogenic causes of neurodevelopmental conditions that will potentially assist in future clinical testing for rare disorders. Establishing clear genotype-phenotype correlations will provide clinical risk assessment opportunities and counseling for future pregnancies for the affected families. This will also allow relative screening, prenatal diagnosis and potential interventions. The ultimate goal for any gene discovery study is to hopefully inform translational approaches in targeted curative therapy for devastating genetic diseases.

1.2.6 Therapeutic approaches in genetic disorders

In the recent years there have been various different approaches in developing targeted therapies for genetic disorders. In well-characterised neuromuscular disorders, such as

Duchenne muscular dystrophy (DMD) or spinal muscular atrophy (SMA), there have been numerous attempts to develop gene therapies. Adeno-associate virus (AAV) gene therapy has been used in SMA in a gene replacement approach to reintroduce the SMN protein(Foust et al., 2009). In summary, AAV9 carrying SMN cDNA was able to cross the blood-brain barrier (Foust et al., 2010) offering a successful intravenous therapy commercially available as Spinraza. An additional approach to SMA treatment is by AAV-mediated antisense oligonucleotide (AON) therapy, also used in DMD. This approach was very important for DMD, as AAV gene therapy was not successful for DMD due to the size of the mutated dystrophin protein. Nonetheless, AAV-mediated AON therapy was able to alleviate the phenotype in a portion of the patients by restoring the disrupted reading frame by exon-skipping strategy. The two studies on AONs and currently available therapeutic agents are Eteplirsen (Lim et al., 2017) and Drisapersen (Goemans et al., 2016).

Furthermore, when mutations result in abolition of enzymatic activity there are enzyme replacement therapies (ERT) that are now in mainstream clinical use. An example of such therapy is enzyme replacement therapy for Anderson-Fabry disease, where α-galactosidase A is inactive. Fabrazyme and Replagal are two approved ERTs used for this genetic disorder (Bengtsson et al., 2003).

Additionally, small molecule based therapies are becoming more popular with Ivacaftor being the first genotype-specific therapy for cystic fibrosis (Thursfield, 2013). Several studies *in vitro* and *in vivo* are using small molecule as therapeutic agents, but further experimental studies will be required to take more small molecules to clinical trials. Similarly, chaperone-based therapies are being examined, for instance in Krabbe disease, as a potential pharmacological agent promoting correct protein folding and trafficking, but further validation is required to proceed to clinical settings (Graziano et al., 2016).

Lastly, when the CRISPR-Cas9 system was firstly used for genome editing in 2013, it provided a very appealing therapeutic option for gene editing and correcting mutation causing diseases (Cong et al., 2013). However, this system is currently only being used in animal models and pre-clinical settings for genome editing purposes. There are safety concerns that must be proven due to potential undesired off-target effects that can obstruct its clinical translation (Aguti et al., 2018, Nelson et al., 2017).

1.3 Disease mechanisms and pathways

This section will outline the major pathways that have been associated with the disease phenotypes of the cases discussed in this study. These pathways include the ubiquitin-proteasome system (UPS), as two families were identified with pathogenic variants in an E3-ubiquitin ligase and an E3-ubiquitin ligase adaptor, and the intracellular trafficking of proteins and its role in myelin formation, as the disease mechanism of family ND13. In addition, mitochondrial defects were identified as a common cellular phenotype in some of the families in this study, and the role of mitochondria in disease will be also summarised herein.

1.3.1 Ubiquitination and other post-translational modifications

Ubiquitination is an important form of post-translational modification that cells utilise to transmit signals by various proteins engaging in complex protein networks to facilitate the transmission of a particular signal. Ubiquitin is a small, 8.6kDa protein, that is covalently bound to proteins via the ubiquitin pathway, thereby determining their degradation, localisation, interacting partners, structure or activity (Pickart and Eddins, 2004). The addition of ubiquitin molecules can have a signaling role since these are recognised by other proteins carrying Ubiquitin Binding Domains (UBDs) that further transduce the signal. The modification may also be reversible by the removal of ubiquitin by deubiquitinating enzymes (DUBs). Targeted proteins can either be mono-ubiquitinated, where a single ubiquitin molecule is added, or poly-ubiquitinated where a polyubiquitin chain is attached. Each addition of ubiquitin molecule results to a different signal being transmitted ranging from gene expression to protein degradation (Woelk et al., 2007).

1.3.1.1 The Ubiquitin-Proteasome system

E1 enzymes are the first part of the UPS system and E1s are involved in activating ubiquitin and transferring it to the E2 active site. There are only two E1 enzymes in humans and they have an extremely crucial role in the maintenance of homeostasis (Allan and Phillips, 2017). Inhibition of this initial step of E1 activation of ubiquitin results to the complete shutdown of the UPS. The process by which E1s activate ubiquitin involve the ATP-dependent adenylation of the C-terminal carboxyl group of ubiquitin, forming a thioester bond between the C-terminus of ubiquitin and the catalytic cysteine residue of the E1 enzyme. During this initial process of the UPS pathway, conformational

changes take place during the activation and transfer of ubiquitin from E1 to E2 (Kleiger and Mayor, 2014).

The second step of the pathway comprises the transfer of the ubiquitin molecule from the E1 enzyme onto the E2 ubiquitin conjugating enzyme by transferring ubiquitin to the active site cysteine of E2s forming an E2-Ub thioester bond. The ubiquitin can be transferred by either transferring the thioester bond from E1-Ub to a thiol group (transthiolation) or by transferring the thioester bond to an amino group (aminolysis). There are approximately 40 different E2s in humans, each of them showing a downstream specificity to the E3 ligase enzyme that they will bind to, often having the ability to bind to more than just one E3. E2 ubiquitin conjugating enzymes are key enzymes of the UPS as they are responsible for the transfer of ubiquitin molecules from the E1s to the targeted substrate by forming E3-ligases complex with E3 ligases and other complex adaptors (Stewart et al., 2016).

In order to ubiquitinate a target substrate, E2 and E3 enzymes need to simultaneously interact, as part of a complex, whereby the ubiquitin molecule will be transferred from the E2 ubiquitin conjugating enzyme to the substrate, by interacting with the E3 ligase. In this case E3 ligases act as both a catalyst for the reaction and as a recognition molecule for targeting specific substrates. This process happens by the formation of an iso-peptide bond between the ubiquitin molecule and the substrate (Berndsen and Wolberger, 2014).

There are over 600 identified E3 ligases in humans, with each E3 ligase presenting very specific binding affinities to specific targets. The E3 ligases are mainly categorised in two groups based on their structure: the RING type and the HECT type of E3s. The RING E3s make up the largest group of E3 ligases, and they have a very distinct U-box fold or RING catalytic domain that ubiquitinates targets by a direct ubiquitin transfer. They can exist as single polypeptide units that interact with E2s, or as multicomponent complexes that interact with E2s and other adaptor proteins to form an E3-ligase complex. Examples include the cullin-RING ligase complex and the heterodimer of BRCA1 and BARD1 complex. On the other hand, HECT-type of E3s do not transfer ubiquitin directly to substrates. Instead they form a thioester-linked intermediate with ubiquitin molecule, by attaching a ubiquitin molecule to an active site cysteine on the E3, before transferring it to the substrate. This is done by the process of transthiolation (Zheng and Shabek, 2017). A simplified UPS pathway is outlined in **figure 1-2.**

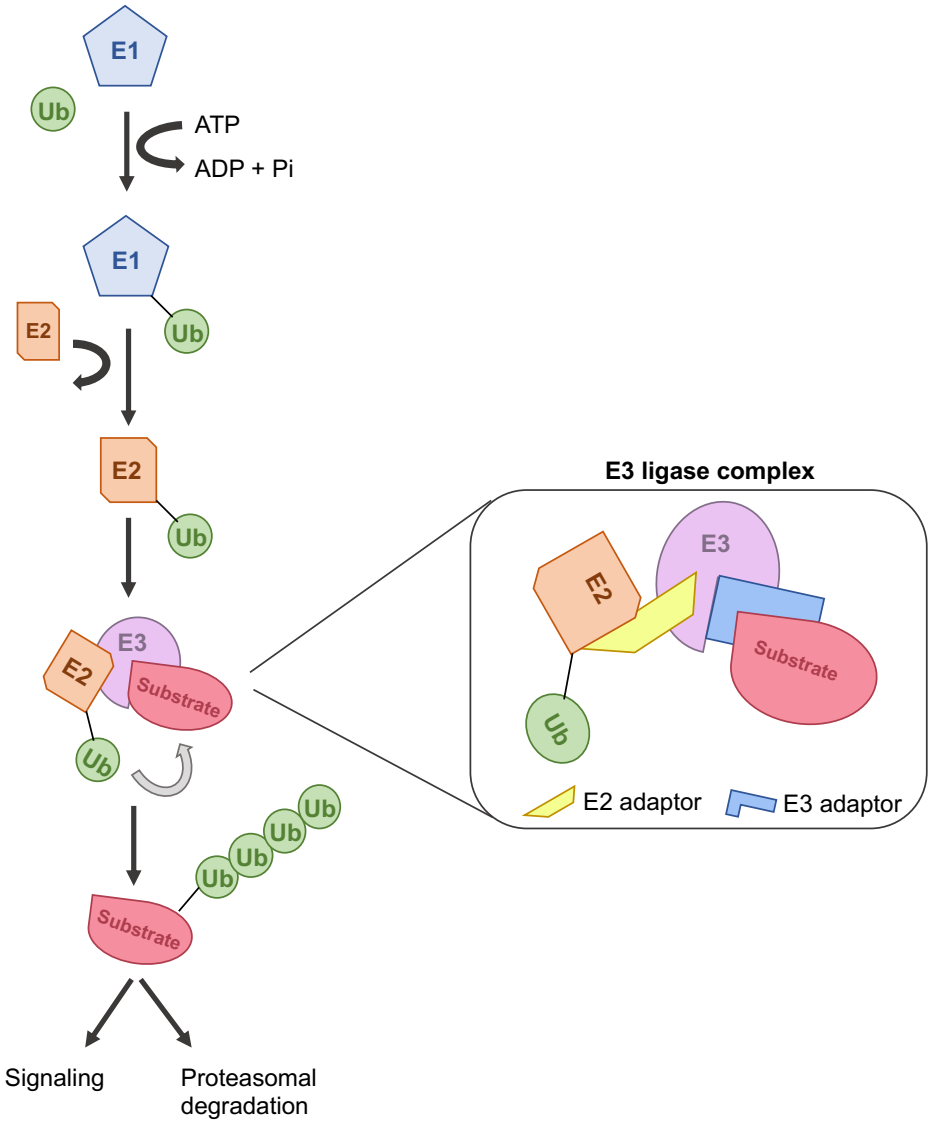


Figure 1-2 The Ubiquitin-Proteasome system

Outline of the ubiquitin-proteasome system, from ubiquitin activation by E1s to substrate ubiquitination. The ubiquitination of the substrate can lead to very specific downstream processes, either involved in cellular singaling or targeting certain substrates for proteasomal degradation. Focusing on the E3-ligase complex, a representation of specific E2 and E3 adaptors is outlined, as these proteins also participate in the complex formation to bring E2s and E3s together. One of the cases outlined in the results sections includes a variant in an E3-adaptor protein.

Ubiquitinated substrates at the end of the pathway are either targeted to the proteasome for protein degradation or act as signals for downstream processes. Substrates that are targeted for degradation have a poly-ubiquitinated chain attached to them, most commonly joined at lysine-48. It has also been reported that lysine-11, -29 and -63 can also act as signals for proteasome degradation, sometimes after being modified by DUB enzymes to distinguish between a proteasomal signal and other cellular signals (Ohtake et al., 2018, Kleiger and Mayor, 2014).

Nonetheless, ubiquitination also serves as a signaling cue for a range of downstream processes including endocytosis, DNA repair, gene expression, NF-kB activation, ribosomal functions and others. Substrates being mono-ubiquitinated are only aimed for downstream signaling purposes, whereas poly-ubiquitinated chains can also serve as signaling cues. There is a range of poly-ubiquitinated chains that serve for cellular signaling. Recent studies are also identifying more complex forms of ubiquitination, known as branched poly-ubiquitination, that seems to be involved in NF-kB regulation (Ohtake et al., 2016). However, the role of branched poly-ubiquitination is largely unknown (Woelk et al., 2007). A brief summary of the localisation of the ubiquitin chain formation and its relevant roles is summarised in **figure 1-3**.

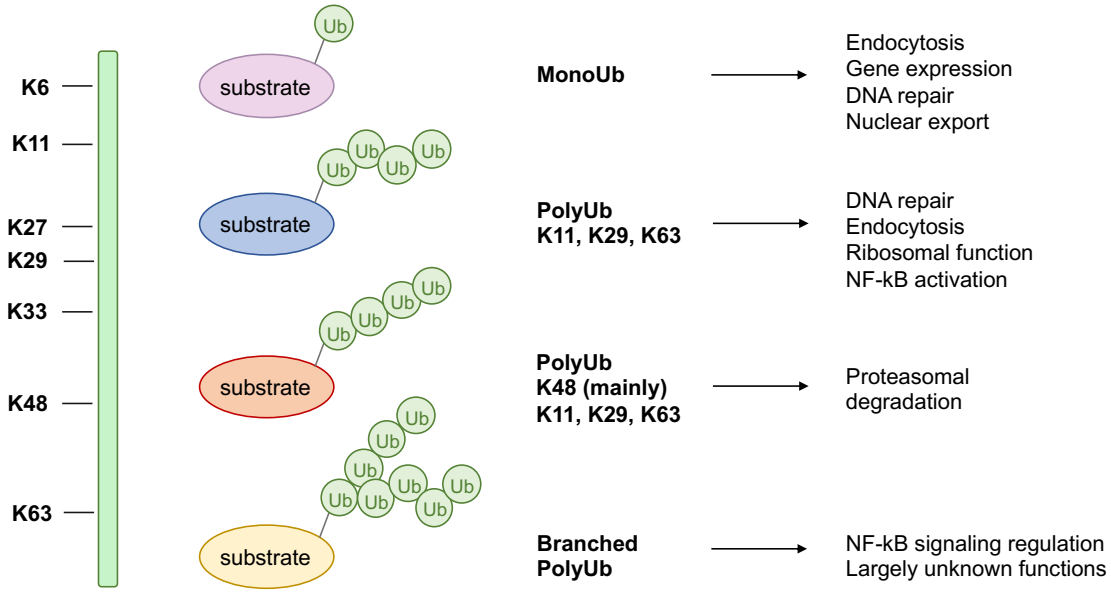


Figure 1-3 Main types of ubiquitination

Representation of the main types of ubiquitination and their key roles. On the left, the main binding sites of lysine residues on ubiquitin molecules are outlined. On the right, the ways by which a substrate can be mono- or poly- ubiquitinated and the corresponding downstream processes of each type of ubiquitination. It is worth emphasizing that the way poly-ubiquitin chains are being formed and how each chain dictates the downstream mechanism of action is still largely unknown and currently being investigated by the field. In addition, ubiqutination can also occur on residues other than lysines (methionine, valine, isoleucine and cysteine) but with much less affinity compared to lysine.

In this study, two independent cases with a neurodevelopmental delay phenotype were identified with variants in an E3-ligase and in an E3-ligase adaptor protein, *HERC2* and *KLHL7* respectively. Interestingly, *HERC2* is an E3-ligase of the HECT-type (Garcia-Cano et al., 2019) and KLHL7 is an adaptor to Cullin-RING ligases through the BTB domain (Canning et al., 2013). *HERC2* is an extremely large protein that seems to play a role in a range of cellular functions including cell proliferation, DNA damage repair and neurodevelopment (Garcia-Cano et al., 2019). On the other hand, *KLHL7* seems to be implicated in ubiquitination of target proteins for proteasomal degradation (Kigoshi et al., 2011).

These findings provide more evidence that support the crucial role of E3 ubiquitin ligases, and of the whole UPS, in neurodevelopmental diseases (George et al., 2018). Examples of such cases include the Angelman syndrome, with variants in *UBE3A* (Sadikovic et al., 2014), Charcot-Marie-Tooth Type 2, with variants in *LRSAM1* (Zhao et al., 2018) and *TRIM2* (Pehlivan et al., 2015), and Limb-girdle muscular dystrophy 2H with variants in TRIM32 (Saccone et al., 2008), as well as Bardet-Biedl syndrome type 11 with a private mutation in *TRIM32* (Chiang et al., 2006).

1.3.1.2 Membrane-associated post-translational modifications

Membrane proteins often undergo lipid modifications post-translationally, in order to be incorporated into the correct sides of the membrane bilayer. These modifications include prenylation, palmitoylation and myristoylation, and have been discussed in this study with regards to the MAL protein and its potential modifications as a membrane protein.

Prenylation is a post-translational modification that provides membrane proteins with a hydrophobic C terminus, allowing them to interact or be attached to the plasma membrane. This process happens by the attachment of a geranylgeranyl isoprenoid or a farnesyl group to the C-terminal of proteins. This process is vital for a number of signaling proteins to maintain their cellular activity and localisation. Some of these proteins include the Ras family of GTPases and the hererotrimeric G-proteins (Palsuledesai and Distefano, 2015).

Palmitoylation is a lipid modification process that is reversible. It acts as a lipid anchor for membrane localization of proteins, but it also has a role in shuttling modified proteins between different cellular compartments or different regions of the membrane into lipid rafts. This role has a significant downstream signaling effect (Guan and Fierke, 2011). This process comprises of the reversible addition of fatty acids (16 carbon fatty acid)

onto cysteines, and less frequently serine and threonine residues. This reversible mechanism occurs rapidly and close to the cytosol-membrane interface. The cysteines that get modified are often surrounded by basic or hydrophobic amino acids and are close to myristoylation or prenylation sites (Salaun et al., 2010).

Myristoylation is another form of lipid modification that plays a role in targeting proteins to the plasma membrane or endomembrane systems, as well as being involved in cellular signaling and protein-protein interactions. Myristoylation is the attachement of a myristic acid to the N-terminus of a protein, a reaction catalyzed by the N-myristoyltransferase enzyme. This process usually occurs cotranslationally on newly synthesized proteins, after the initial methionine is cleaved off by methionine aminopeptidase and myristic acid is attached. Myristoylation can also occur on proteins at post-translational stages, specifically on glycine residues. A proteolytic cleavage can take place at the N-terminus of the protein on a glycine residue, followed by the attachment of myristic acid (Udenwobele et al., 2017).

1.3.2 Nonsense mediated decay in disease

Some of the variants identified in this study are nonsense variants, leading to premature termination codons (PTCs) in the mRNAs and potentially leading to either the translation of a truncated protein or the complete abolition of that particular protein as a result of nonsense mediated decay (NMD). The NMD pathway is a surveillance mechanism of the cell that can selectively identify and degrade faulty mRNAs carrying PTCs, by recruiting RNA helicases to promote mRNA decay (Hug et al., 2016).

The first step of the NMD pathway is to correctly distinguish a PTC from a normal termination codon. This is usually linked to pre-mRNA splicing, and any mRNAs that have a PTC approximately 50 nucleotides upstream of the last exon-exon junction will be degraded. This process is supported by the signaling of a multi-subunit protein complex, known as the exon-junction complex, which is deposited 20-24 nucleotides away from an exon-exon junction during splicing (Le Hir et al., 2001). Once a PTC is identified, RNA helicases (UPF1/SMG2 complex) are recruited to the site and clamp to mRNA in an ATP-dependent fashion. The RNA helicase complex undergo cycles of phosphorylation, recruiting SMG phospho-binging proteins, such as SMG6 and SMG7 to promote mRNA decay. SMG6 is an endonuclease that will initiate the NMD-mediated mRNA degradation (Hug et al., 2016).

In some instances, the nonsense variant might be close to the exon-exon junction, thereby easily missed by the NMD machinery. In cases such as these, nonsense-mediated altered splicing (NAS) takes place, whereby the exon carrying the mutation is skipped during pre-mRNA splicing (Wang et al., 2002). Sometimes more than one exon can also be skipped, either way leading to a shorter version of the protein that will no longer carry the variant. The protein may no longer be functional, but this is an attempt to eliminate a potentially damaging variant from the translating protein. The mechanisms that dictate whether NMD or NAS will take place as a result of the PTC process are not yet fully understood (Liu et al., 2001).

1.3.3 Myelin formation and intracellular protein trafficking

1.3.3.1 Myelin formation

During cellular evolution in vertebrates, the formation of an insulating layer of myelin sheath wrapped around neuronal axons was established, transforming the way neural signals were being transmitted across axons. Myelin sheath is an insulating structure of highly compacted layers of cell membrane wrapped around axons, with periodic break points known as Nodes of Ranvier, forcing the neural impulses to jump between nodes. This increased the speed of neuronal signaling and thereby enhanced neural function in vertebrates (Fields, 2014). The process of myelination occurs relatively late in development. In mice, myelination starts at birth and it is completed at postnatal day 60. In humans, myelination starts during the third trimester of embryonic development but peaks during the first year of life, and in some cortical areas it continues until young adulthood (Snaidero and Simons, 2014).

In the CNS, myelin is formed by oligodendrocyte cells, whereas in the PNS myelin is formed by Schwann cells. The complex development of myelin sheath requires the formation of huge quantities of specialized cell membranes, precise cell-cell recognition and cell motility. Damage of myelin leads to disruption of neuronal signal transmission and thereby the source of a wide range of neurological diseases, ranging from late onset Multiple Sclerosis to congenital leukodystrophies (Fields, 2014).

In the CNS, myelin is formed by oligodendrocytes that are able to extend their membranes and spiral them around axons forming a compact multiple layer of overlapping membranes. This process is initiated once an oligodendrocyte cell makes contact with an axon forming a membrane junction known as the “spot weld” (Luse, 1959). This membrane domain promotes intercellular communication between

oligodendrocytes and neurons, initiating encirculation of membranes around axons. During the multilayer membrane formation, the cytoplasm becomes expelled from all the layers, except the outermost and innermost layer of myelin which provide insulation. The intervening layers are then joined by the myelin basic protein (MBP) to make a compact layer (Fields, 2014).

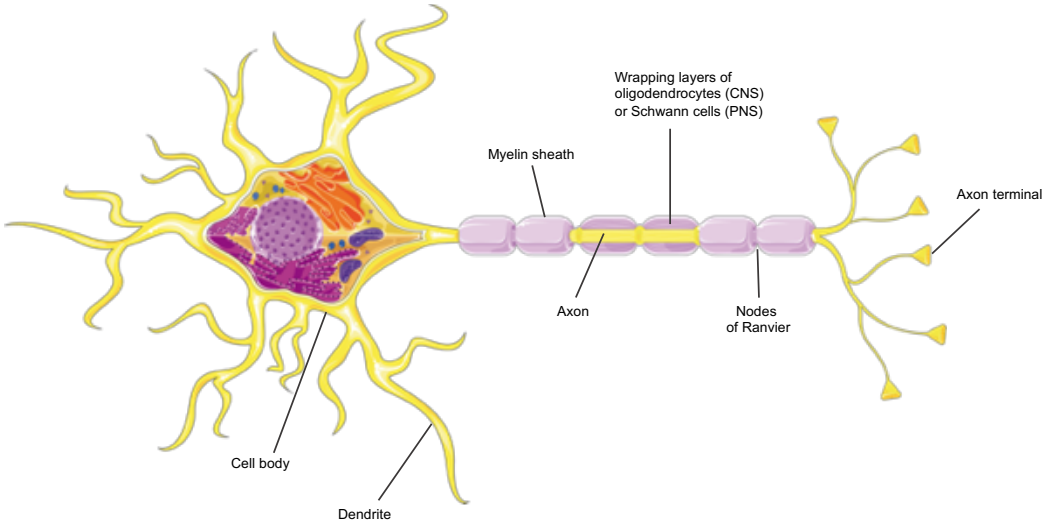


Figure 1-4 Myelinated neuron

A representation of a myelinated neuron, showing how myelin sheath wraps around the axon, either by oligodendrocytes or Schwann cells, and the formation or Nodes of Ranvier.

During myelin formation, oligodendrocyte progenitor cells become highly polarized driving extensive membrane trafficking of vesicles, proteins and mRNAs towards the leading edge of myelin biosynthesis. The constant delivery of proteins to the leading edge of the inner tongue of uncompacted membrane layers is essential for the wrapping and maturation of oligodendrocytes into compacted myelin membranes (Snaidero et al., 2014). The maturation of oligodendrocyte progenitor cells is a complex process that starts with the cell cycle exit and necessitates the coordination of expression of several genes and signaling pathways. For instance the activation of the PI3K/Akt pathway is essential for oligodendrocyte differentiation and myelination (Flores et al., 2008). Activation of PI3K can also promote myelination via receptor coupling, such as the insulin-like growth factor 1 receptor (IGF-1), which itself promotes Akt phosphorylation and activation in oligodendrocytes during myelination (Goebbels et al., 2010). In addition, the mammalian target of rapamycin (mTOR) pathway regulates oligodendrocyte progenitor cell maturation and myelination. mTOR has a key role in Akt Ser 473 phosphorylation and activation that promotes myelination. mTOR seems to act both as a regulator of Akt activity, as well as a signaling effector of PI3K/Akt pathway, having a central role in oligodendrocyte differentiation (Tyler et al., 2009).

1.3.3.2 Intracellular protein trafficking

As mentioned above, efficient and constant delivery of proteins and vesicles to the plasma membrane by intracellular trafficking is vital for the progression and maturation of myelination. One key aspect of the protein trafficking discussed in this study is the involvement of MAL in regulating intracellular transcytosis of PLP1 (Bijlard et al., 2016) and its correct redirection to the basolateral membranes for myelin development. PLP1, or proteolipid protein 1, is the major constituent of myelin in the CNS (Diehl et al., 1986) and it is required for the formation and maintenance of the multilamellar structure of myelin. PLP1 also has an essential role in adult life, as it is required to preserve myelin integrity in the CNS (Luders et al., 2019).

In comparison to other membranes, myelin membranes have an extremely high lipid to protein ratio primarily made from cholesterol, galactolipids, sulfatide and galactosylceramide. These lipids are often involved in the assembly of membrane microdomains, known as lipid rafts, that are necessary for protein trafficking and sorting during myelin development (Ozgen et al., 2014). With regard to PLP1, galactolipids have a fundamental role in facilitating vesicle-mediated transport of PLP1 to the myelinating membranes (Baron and Hoekstra, 2010). Baron *et al*. described that the transport of PLP1 to the myelin membrane is governed by a transcytotic mechanisms, involving a complex network of vesicular trafficking and lipid-protein interactions (Baron and Hoekstra, 2010). The transport of PLP1 to the membrane involved the vesicle-soluble N-ethylmaleimide-sensitive factor attachment protein receptors (v-SNAREs), specifically VAMP3 and VAMP7. Once PLP1 reaches the apical or the basolateral membrane, membrane-localized target SNAREs (t-SNAREs) are involved in the docking and fusion of transport vesicles to the membrane. In particular, syntaxin 3 and 4 seem to be implicated in the process and they bind VAMP7 and VAMP3 respectively (Baron et al., 2015).

A closely related protein to PLP1 is the myelin and lymphocyte protein (MAL), that is highly expressed during active myelination (Schaerenwiemers et al., 1995) and suggested to be implicated with PLP1 trafficking. MAL is a known regulator of direct apical sorting (Cheong et al., 1999) and it is associated with galactosylceramide and sulfatide in forming membrane microdomains; both strongly linked to PLP1 transport and membrane internalization (Ozgen et al., 2014). MAL also seems to be involved in biogenesis and maintenance of myelin membrane (Kim et al., 1996) and has a role in axon-glia interactions promoting myelination (Schaeren-Wiemers et al., 2004).

Bijlard *et al.* proposed that once PLP1 is translated and exits the Golgi apparatus, it then follows the transcytotic pathway proposed by Baron *et al.* to reach the apical membrane of polarized oligodendrocytes, whereby it interacts with MAL. Upon this interaction, PLP1 seems to undergo a conformational change, prior to reaching the myelin membrane (Bijlard et al., 2016). It is thought that MAL and PLP1 interact at the apical membrane and that MAL plays a key role in PLP1 conformational change and recruitment to membrane microdomains. Bijlard *et al.* proposed that MAL has a fundamental role in PLP1 trafficking, but that it mainly facilitates the lateral diffusion of PLP1 to the myelin membranes at the stage when the myelin starts to be compacted. They proposed that from that stage onwards the transcytosis pathway is shifted to a lateral diffusion governed by MAL, as the myelin membranes get tightly packed into forming the myelin sheath (Bijlard et al., 2016).

In this study, a potentially pathogenic variant has been identified in *MAL* that prevents the correct localisation of the protein to the apical membrane of the cell, and could thereby impact the accurate distribution of PLP1 during myelination. The patients identified with this variant were diagnosed with a hypomyelinating leukodystrophy, similar to Pelizaeus-Merzbacher disease which is caused by variants in *PLP1* (Osorio and Goldman, 2018). Considering the role of MAL in regulating PLP1, it will be reasonable to assume that variants in both genes can lead to a hypomyelinating phenotype as myelin membrane formation will be probably disrupted. Another interesting observation discussed in this study is how missense variants in both genes, *MAL* and *PLP1*, cause aggregates in the endoplasmic reticulum leading to mis-localized proteins and initiation of the unfolded protein response as an additional cause of disease (Inoue, 2017).

1.3.3.3 Unfolded protein response in disease

The unfolded protein response (UPR) is a complex signaling response that occurs in the endoplasmic reticulum (ER), if and when homeostasis is disrupted, in an attempt to alleviate ER stress. The endoplasmic reticulum is an important and multifunctioning organelle that is vital for protein synthesis, folding and processing. The ER is also a master regulator of calcium homeostasis (Bravo et al., 2013). In some instances, genetic variants can severely impact protein folding, leading to the formation of protein aggregates in the ER. These aggregates can accumulate and cause ER stress, that will then activate the UPR to restore homeostasis (Schroder and Kaufman, 2005). If the UPR fails to restore physiological ER function and reduce stress, it will then mediate UPR-dependent apoptosis (Tabas and Ron, 2011). This process is often associated with

disease pathogenesis in various disease, and is a common cause in disorders of myelinating glia such as Pelizaeus-Merzbacher disease, Vanishing White Matter disease, Charcot-Marie-Tooth and Multiple Sclerosis (Clayton and Popko, 2016).

ER stresses are sensed by the immunoglobin-binding protein (BiP) that is originally bound to the three ER transmembrane proteins that regulate the UPR. Upon stress, BiP dissociates from the three regulators of UPR, PKR-like ER kinase (PERK), Inositol Requiring 1 (IRE1) and Activating Transcription Factor 6 (ATF6), and thereby initiates UPR. This response will then activate a cascade of signaling events that will attempt to restore stress by an adaptive response, restore homeostasis by a feedback control, and determine cell fate based on whether the ER stress has been relieved (Oslowski and Urano, 2011). The adaptive response of the UPR is associated with the up-regulation of molecular chaperones that will promote folding and handling of unfolded protein aggregates. It is also involved in attenuating translation in order to reduce ER load and prevent additional accumulation of unfolded proteins. Furthermore, the adaptive response promotes increased expression of proteins linked to clearance of unwanted proteins, a response known as ER-associated protein degradation (ERAD). The feedback response of the UPR is a homeostatic control that will negatively regulate the UPR once ER stress is reduced in order to avoid hyperactivation of the UPR. However, if homeostasis cannot be achieved the cell fate response will switch to promote apoptosis, a response that plays a crucial role in pathogenesis of ER stress-related diseases (Oslowski and Urano, 2011).

The first regulator of the UPR is PERK, which is an ER kinase that oligomerizes upon BiP dissociation as a response to ER stress. PERK has a luminal domain and a cytoplasmic domain that possesses the kinase activity. When PERK oligomerizes it can then autophosphorylate and then further phosphorylate the α subunit of eukaryotic initiation factor (eIF2α) (Harding et al., 1999). Once eIF2α is phosphorylated by PERK, it binds to eIF2β, and prevents the formation of an active translation-initiation complex (Dever, 2002). This result in a global inhibition of mRNA translation in an attempt to release ER load. In addition, eIF2α phosphorylation promotes translation of activating transcription factor 4 (ATF4) that regulates the expression of proteins that initiate ER stress-mediated apoptosis such as C/EBP homologous protein (CHOP). In case homeostasis is not restored, certain proteins like CHOP can also initiate cell death (Oslowski and Urano, 2011).

The second regulator of the UPR is IRE1 which becomes activated by dimerization and autophosphorylation upon ER stress. IRE1 has two isoforms, IRE1α and IRE1β, with

IRE1α ubiquitously expressed and well-studied. IRE1α acts by splicing X-box binding protein 1 (XBP1) mRNA. The spliced form of XBP1 regulates the expression of certain genes involved in protein folding, such as disulfide isomerase (PDI). It also up-regulates expression of specific proteins involved in ER-associated degradation such as like the ER-degradation-enhancing- α-mannidose-like protein (Oslowski and Urano, 2011). In prolonged ER stress, IRE1α activates c-Jun N-terminal protein kinase (JNK) by the recruitment of TNF-receptor-associated factor 2 (TRAF2). This process requires the activation of apoptosis-signaling-kinase 1 (ASK1) that is upstream of JNK and will induce apoptosis and increase expression of BiP (Shinkai et al., 2010).

The last regulator of the UPR is ATF6, an ER transmembrane transcription factor that transmits stress signals from the ER directly to the nucleus as a response of stress (Bravo et al., 2013). ATF6 has two isoforms, ATFα and ATFβ, with two and one Golgi localisation sequences respectively. ATF6 is maintained normally in the ER by binding to calreticulin and BiP. Upon ER stress, ATF6 translocates to the Golgi apparatus where site-1 and site-2 proteases cleave the protein. The N-terminal portion of ATF6α then moves to the nucleus where it promotes transcription of UPR genes (Oslowski and Urano, 2011). ATF6α seems to up-regulate the expression of XBP1 (Yoshida et al., 2001) and CHOP (Ma et al., 2002) that cross-talk with the other UPR regulating pathways. ATF6α also promotes transcription of BiP and GRP94 chaperones (Schroder and Kaufman, 2005), and other proteins associated with ER homeostasis such as SERCA (Thuerauf et al., 2001) and p58IPK/DNAJC3 (van Huizen et al., 2003). ATF6β seems to act as an antagonist to ATF6α and represses the transcriptional signal of ATF6α. This process acts as a negative regulator of this UPR pathway that terminates the response once homeostasis is reached (Thuerauf et al., 2004).

The diagram in **figure 1-5** summarises the three regulating pathways of the UPR, upon sensing ER-stress and the potential outcomes of these responses. The mechanism discussed in this section was observed as a result of the missense variant identified in *MAL*. The mutant protein formed ER aggregates and thereby triggered the UPR as a result of ER stress. This is a very common pathomechanism in Pelizaeus-Merzbacher disease observed by missense variants in *PLP1*.

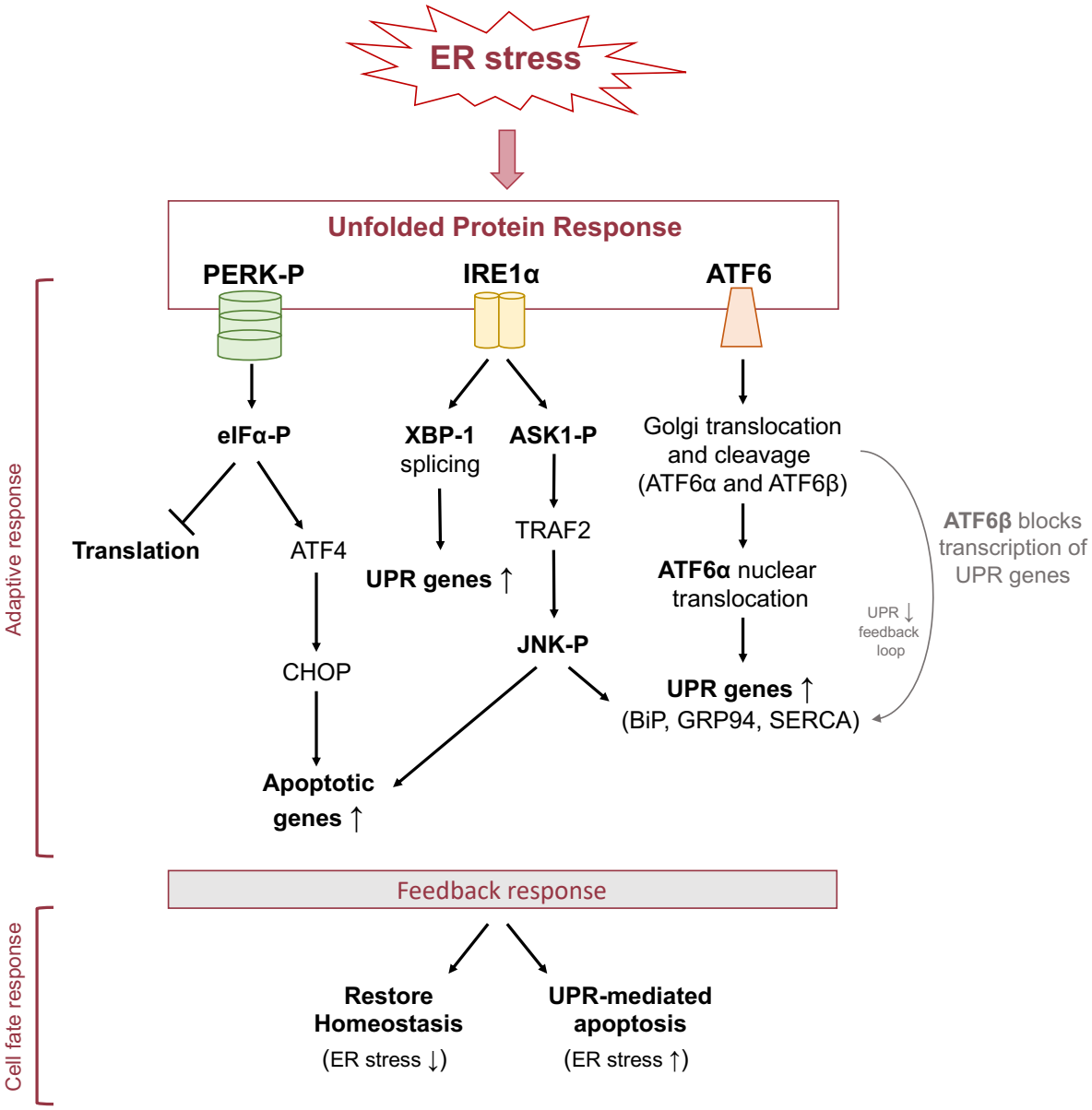


Figure 1-5 The Unfolded Protein Response

The diagram above summarises the UPR with the three responses: adaptive, feedback and cell fate response. In the adaptive response, the three main regulators of the UPR initiate downstream signaling cascades in order to restore ER stress.

1.3.4 Mitochondria in disease

Mitochondria are very important organelles for a functioning eukaryotic cell. They form an extensive and dynamic network across a cell, and maintain mitochondrial homeostasis by fission and fusion mechanisms, mitophagy and mitochondrial biogenesis. They possess a complex proteome made up of proteins encoded by mitochondrial genes as well as nuclear. Mitochondria are multifunctional and are associated with other cellular processes apart from metabolism and aerobic energy production. Mitochondria are closely associated with cellular ion homeostasis, Ca^{2+} storage and Ca^{2+} dependent cell signalling and apoptosis (Herst et al., 2017).

As part of this study, different aspects of mitochondrial biology are covered and their involvement in inherited disease. Initially, a previously identified variant in *MICU1* was further functionally characterised in this research project. The mitochondrial calcium uptake 1 protein (MICU1) is an important subunit of the mitochondrial calcium uniporter (MCU) that it is an essential channel for regulating calcium ions in the mitochondria and maintaining cellular homeostasis (Antony et al., 2016). MCU is composed primarily of MICU1 and MICU2, along with the calcium sensing regulators. MICU1 is the main gatekeeper of the channel, closing the uniporter under low calcium cytosolic levels and opening the channel in high calcium cytosolic levels. Loss-of-function mutations in *MICU1*, as well as *MICU2* (Shamseldin et al., 2017), abolish the correct gating of the uniporter leading to pathogenesis by the disruption of homeostasis and induction of mitochondrial stress (Logan et al., 2014). Patients with mutations in *MICU1* and *MICU2* are affected by proximal myopathy and movement disorder. This outlines a direct impairment of mitochondrial function by the calcium overload that can lead to mitochondrial stress and induction of cell death.

Additionally, mutations in other genes that are not localised in the mitochondria seem to have a significant impact on mitochondrial function. A well-known example of such association is loss-of-function mutations in *PRKN*, an E3-ubiquitin ligase, that causes early-onset Parkinson's disease. PRKN seems to be implicated in the elimination of impaired mitochondria by promoting autophagy, a process that is compromised in Parkinson's disease (Narendra et al., 2008). There seems to be an emerging theme as part of this research project, whereby mutations in members of the UPS seem to affect mitochondrial functions. For instance, part of this project revealed that variants in UPS members are able to impact mitochondrial OXPHOS chains, potentially by disrupting cellular homeostasis, and interrupting the healthy mitochondrial network leading to mitochondrial fragmentation.

Many E3-ligases complex are known to localize at the mitochondrial outer membrane, likely to be involved in ubiquitinating various mitochondrial proteins for degradation. The UPS is known to have a key role in quality control and maintenance of damage proteins in the mitochondria. It is also associated with regulating mitochondrial dynamics by controlling fission and fusion proteins, a process that sustains cellular functions. Fission and fusion work antagonistically to each other and it is a very active process that requires several proteins. When Mitofusin1 (Mfn1) and Mitofusin 2 (Mnf2) are ubiquitinated for degradation, and Dynamin related protein (Drp1) and Mitochondrial fission 1 protein (Fis1) accumulate on the outer mitochondrial membrane fission occurs. On the other hand, when Drp1 and Fis1 are tagged for proteasomal degradation but Mfn1 and Mfn2 accumulate fusion occurs (Bragoszewski et al., 2017). A potential disruption in the UPS activity, perhaps by loss-of-function variants in UPS proteins, can severely dysregulate the processes of fission and fusion leading to fragmented or hyperfused mitochondria respectively. A fragmented mitochondrial network is a common feature observed in primary patient cells carrying variants in proteins comprising the E3-ligase complex (Karbowski et al., 2007).

Another potential explanation of the observed fragmented mitochondrial network caused by UPS-related mutations, might be due to disrupted homeostasis and induced stress response in the mitochondria. The UPS is involved in many signaling pathways in addition to regulating protein turnover. Loss-of-function variants or even complete loss of a member of the UPS pathway may lead to disruption of a downstream signaling pathway and cellular balance. Mitochondria are sensitive to cellular stress as they can either act in an attempt to restore cellular homeostasis and ion balance or trigger autophagy and apoptosis if stress levels are not restored. The exact association of the identified UPS variants of this study and their role in mitochondria is largely unknown, but interesting observations have been established that support the significant role of the UPS in mitochondrial function.

1.4 Aims and objectives

The overall aim of this study was to use next generation sequencing technologies to identify novel disease genes in rare autosomal recessive conditions. Subsequent cellular modelling of identified variants was used to prove pathogenicity and gain insights into cellular mechanisms of disease. The research project was mainly focused on neuromuscular and neurodevelopmental disorders that remained unsolved by regional NHS genetic testing. Patient recruitment was accomplished by close collaborations with clinical geneticist and diagnostic labs in the Leeds-Bradford area, with the full inclusion criteria given in appendix A. In addition, by identifying variants in novel genes, this project also aimed to establish novel genotype-phenotype correlations for neurodevelopmental conditions and provide patients with a molecular diagnosis. Any research findings from this study would also be helpful in expanding clinical diagnostic panels and promote accurate diagnosis of future cases. Lastly, cellular modelling of identified variants were used to confirm pathogenicity and to provide useful knowledge on potential pathomechanisms that could form the basis for characterization of novel druggable targets in the future.

The following were the specific objectives of the research project:

1) Use Whole Exome Sequencing on selected individuals from the recruited families to identify novel genes that cause rare autosomal recessive disorders, specifically neuromuscular or neurodevelopmental disorders
2) Perform bioinformatics analysis on the WES data to identify novel disease genes according to family pedigree and pattern of inheritance.
3) Screen cohorts of patients with similar phenotypes in a specific family, to enable independent replication of a WES result in order to prove pathogenicity of variants in a candidate gene.
4) Identify additional families carrying putative mutations in any disease gene by sharing data with worldwide datasets, such as Decipher.
5) To validate pathogenicity of variants by using established biochemical and cell biology techniques on primary patient cells, to assess protein levels and phenotypic cellular effects as a result of the variant.
6) To validate pathogenicity of variants by functional characterization using cloning techniques to model disease variants *in vitro* and further assessment using biochemical techniques.
7) To study protein localization and protein-protein interactions in the mutant models to gain insight into disease mechanisms.

Chapter 2

Materials and Methods

2.1 Materials

Room temperature (RT) is defined as 20-25°C.

2.1.1 General Reagents

Water	Millipore
Nuclease-free Water	Ambion
Ethanol	Sigma-Aldrich
Methanol	Sigma-Aldrich
Isopropanol	Sigma-Aldrich
Phosphate Buffered Saline (PBS) 1x	Sigma-Aldrich

2.1.2 Solutions

2.1.2.1 PBST 1x

1x	PBS	
0.1% [v/v]	Tween-20	Sigma-Aldrich

2.1.2.2 TBS 10x

0.2M	Tris HCl (pH 7.4)	Sigma-Aldrich
1.5M	NaCl	Sigma-Aldrich

2.1.2.3 Tris-Acetate-EDTA (TAE) Buffer 50x

2M	Tris HCl (pH 7.4)	Sigma-Aldrich
50mM	EDTA (pH 8.0)	Ambion
0.97M	Glacial acetic acid	Sigma-Aldrich

2.1.2.4 Tris-EDTA (TE) Buffer 1x

10mM	Tris HCl (pH 7.4)	Sigma-Aldrich
1mM	EDTA (pH 8.0)	Ambion

2.1.2.5 Tris-Borate-EDTA (TBE) Buffer 10x

890mM	Tris HCl	Sigma-Aldrich
890mM	Boric Acid	Sigma-Aldrich
20mM	EDTA	Ambion

2.1.2.6 Agarose Gel Loading Dye 2x

50% [v/v]	Glycerol	Sigma-Aldrich
10% [v/v]	TBE 10x	
0.1% [v/v]	Orange G	Sigma-Aldrich
0.1% [v/v]	Xylene Cyanol	Sigma-Aldrich

2.1.2.7 NP-40 Cell Lysis Buffer

50mM	Tris-HCl (pH 8.0)	Sigma-Aldrich
150mM	NaCl	Sigma-Aldrich
1% [v/v]	NP-40	Sigma-Aldrich
1x	Protease/Phosphatase Inhibitors	Thermo Fisher Scientific

2.1.2.8 Radio Immunoprecipitation Assay (RIPA) Lysis Buffer

50mM	Tris-HCl (pH 8.0)	Sigma-Aldrich
150mM	NaCl	Sigma-Aldrich
0.1% [v/v]	SDS	Melford Laboratories Ltd
1% [v/v]	NP-40	Sigma-Aldrich
0.5% [v/v]	Sodium Deoxycholate	BDH

2.1.2.9 Cell Lysis Buffer for Genomic DNA Extraction

10mM	Tris HCl (pH 8.0)	Sigma-Aldrich
100mM	EDTA	Ambion
0.25% [v/v]	SDS	BDH
20µg/ml	RNAase A	Sigma-Aldrich

2.1.2.10 IP Incubation Buffer for Transmembrane Proteins

25mM	NaCl	Sigma-Aldrich
20mM	Tris-HCl (pH 8.0)	Sigma-Aldrich
2mM	EDTA	Ambion
10% [v/v]	Ethanol	Sigma-Aldrich
10% [v/v]	Glycerol	Sigma-Aldrich
1x	Protease Inhibitors	Thermo Fisher Scientific

2.1.2.11 IP Wash Buffer

150mM	NaCl	Sigma-Aldrich
50mM	Tris-HCl (pH 8.0)	Sigma-Aldrich
0.5mM	EDTA	Ambion
0.1% [v/v]	NP-40	Sigma-Aldrich

2.1.2.12 BioID Cell Lysis Buffer

500mM	NaCl	Sigma-Aldrich
50mM	Tris-HCl (pH 7.4)	Sigma-Aldrich
0.2% [v/v]	SDS	BDH
1mM	DTT	Sigma-Aldrich
1x	Protease Inhibitors	Thermo Fisher Scientific

2.1.2.13 BioID Wash Buffer

0.1% [v/v]	NP-40	Roche
10x	TBS	
1% [v/v]	Protease Inhibitors	Thermo Fisher Scientific
1% [v/v]	Phosphatase Inhibitors	Thermo Fisher Scientific

2.1.2.14 Bovine Serum Albumin Blocking Buffer for BioID

1% [w/v]	Bovine Serum Albumin	Sigma-Aldrich
0.2% [v/v]	Triton X-100	Sigma-Aldrich
1x	PBS – bring up to final volume	

2.1.2.15 Normal Donkey Serum (NDS) Blocking Buffer for BioID

10% [v/v]	Normal Donkey Serum	Alpha Diagnostics
1% [v/v]	Triton X-100	Sigma-Aldrich
1x	PBS – bring up to final volume	

2.1.2.16 Quenching Solution

3% [w/v]	Sodium Azide	Sigma-Aldrich
4.5% [v/v]	Hydrogen Peroxide	Thermo Fisher Scientific
1x	PBS – bring up to final volume	

2.1.2.17 On-Beads Digest Buffer 1

2M	Urea (pH 8.0)	Invitrogen
50mM	Tris-HCl (pH 7.5)	Sigma-Aldrich
50µg/ml	Trypsin NB Sequencing Grade	Serva Electrophoresis

2.1.2.18 On-Beads Digest Buffer 2

2M	Urea (pH 8.0)	Invitrogen
50mM	Tris-HCl (pH 7.5)	Sigma-Aldrich
1mM	DTT	Sigma-Aldrich

2.1.2.19 Seahorse Assay Culture Medium for Mito Stress Test

1x	Seahorse XF Base Medium	Agilent Technologies
1mM	Pyruvate	Invitrogen
2mM	Glutamine	Sigma-Aldrich
10mM	Glucose	Sigma-Aldrich

2.1.2.20 Seahorse Assay Culture Medium for Glycolysis Stress Test

1x	Seahorse XF Base Medium	Agilent Technologies
1mM	Glutamine	Sigma-Aldrich

2.1.2.21 Crystal Violet Solution

0.5% [w/v]	Crystal Violet	Alfa Aesar
20% [v/v]	Methanol	Sigma-Aldrich
80% [v/v]	dH_2O	Millipore

2.1.3 Cell Lines

The cell lines used for all the experiments of this project are summarised in **table 2.1**. Primary cell lines were obtained from skin biopsies through the NHS Cytogenetics Laboratory.

Name	Origin	Source
hTERT- RPE1	hTERT immortalised human retinal pigmented epithelium	ATCC
MDCK	Dog kidney epithelium	ATCC
hTERT- HDFneo	hTERT immortalised neonatal human dermal fibroblasts	Genlantis
HDF Bet23	Human dermal fibroblasts (adult)	Primary/NHS
HDF Adult	Human dermal fibroblasts (adult)	Primary/NHS
MICU1 - HDF	Human dermal fibroblasts from patient with *MICU1* mutation (c.547C>T)	Primary/NHS
HERC2 - HDF	Human dermal fibroblasts from patient with HERC2 mutation (c.13767_13770delTGAA)	Primary/NHS
KLHL7- HDF	Human dermal fibroblasts from patient with *KLHL7* mutation (c.947G>A)	Primary/NHS

Table 2-1 List of Cell Lines

Table includes the organism and the tissue that each cell line originates from and the source from which it was obtained.

2.1.4 Primers

Gene	Exon	Forward	Reverse	Tm°
ADAMTS15	8	TTCTATCTGCCCAAAGAGCC		63.2
ADAMTS15	8		GGCAGAACGCTGGCATATC	65.7
EPHB2	12	TGGTTTCCCATTATGAGGATG		63.4
EPHB2	12		GTCTTGCAGGGAGAGATGGA	64.4
EPHB2	14	GTAAGATGGGCCTGGCAGAG		65.9
EPHB2	14		AGGAAGAGCATTTGCCTGAA	63.8
HERC2	90	AGATGCACTTGAGGCTGACC		64.5
HERC2	90		TGGAGCCAAATCCATTACTTT	62.1
KLHL7	8	TCTTTGCATTGTCCTTTTCAG		61.8
KLHL7	8		AAGAAAGTATCTGGACAAATAATACCA	60.8
MAL	1	GAGCCAGCGAGAGGTCTG		64.1
MAL	1		CAGCTAGCGGATCCCGTC	66.0
MAL	2	CACTCCAGTCACCCCATGT		63.7
MAL	2		ATCCTACTGGGCCAGGGAC	65.2
MAL	3	GGGACCTCAGCTCTGCATCT		65.9
MAL	3		GAAAACAGACCAAGGGCCTA	62.9
MAL	4	CTGGATGCAGTGCAGACG		64.7
MAL	4		ACCATCAAGGGCATTTCTGT	63.3
NCOR2	34	CCGAGGTCCCTATCCTCAA		64.0
NCOR2	34		TGATGCCTAAGGAGTCCCTG	64.1
PNPT1	6+7	GCCATATAGTGCTCTTGTATTTTAGG		61.8
PNPT1	6+7		CCATTGCTGTAACATGGTCACTA	63.6
SH3TC2	8	CTGGGAACCCCTCATTTTCT		64.0
SH3TC2	8		GCAAATCTGCTCAAAGAGGG	63.8
SHETC2-short	8	CACTTGTCCTGAACTTGTCG		61.0
SHETC2-short	8		TGTTCATGTCTTCTGTAGCTGA	60.1
SUPV3L1	9	TAGTCACAAGAAGTTGAGTTGG		62.8
SUPV3L1	9		TTACCAGGTGGGAGACTGC	58.5
TTN	359	TTCTAAATTCAGCTTCCCAAAA		61.5
TTN	359		CAAATTGTATTCTGGAATTAGGAAAA	61.9
TUBA1A	4	GGACACAATTTGACCTATTAACCTA		60.9
TUBA1A	4		ATGAACTATTTGATGTCATTTTGTA	58.1
UNKL	8	AAGTCCTCTGTGTCCTCCTGA		62.9
UNKL	8		ATGGAACTGCTGCTCCGT	63.6

Table 2-2 List of primers used for confirming variant segregation

Table includes the gene name, exon number and nucleotide sequence of each primer, as well as the primer melting temperature (T_m).

Gene	Primer Sequence	Tm°
BioID KLHL7 EcoRI_side	ATCCTGGTGGAGCTGACA	62
BioID KLHL7 HindIII_side	GCAGGTATCTTTCTTTCCTGG	61.6
InFusion KLHL7 Forward	GATATCTGCAGAATTCCTGGGAGGGACTGATTGCA	61.5
InFusion KLHL7 Reverse	GTTTAAACTTAAGCTTTCATGTTTCAAGGGTCTCTTCA	59.4
KLHL7 Internal Forward	AGGAAAGAAAGATACCTGCT	56.2
KLHL7 Internal Reverse	ACCCACAGCAAATATCTTGT	58.5
Q5 SDM MAL_pENTR223 Forward	GAGGCCCTGGACACCATCACG	73.3
Q5 SDM MAL_pENTR223 Reverse	CAGGACTGAGGCGCTGAG	65.2

Table 2-3 List of primers used for cloning

Table includes the gene name and the nucleotide sequence of each primer, as well as the specific melting temperature (T_m).

2.1.5 Antibodies

Primary and secondary antibodies were used as detailed in **tables 2.4** and **2.5** respectively. Specific dilutions are provided for both Western Blotting and Immunofluorescence microscopy were appropriate.

ANTIGEN	SPECIES RAISED IN	SPECIES REACTIVITY	POLYCLONAL/ MONOCLONAL; ISOTYPE	IF DILUTION (1/X)	WB DILUTION (1/X)	SUPPLIER	CATALOGUE NUMBER
PLP1	Rabbit	Human	Polyclonal ; IgG	200	1000	Novus Biological	C106511
ET3-PLP	Rabbit	N/A	Polyclonal; IgG	100	N/A	Kind gift from Dr Wia Baron	N/A
MAL	Rabbit	Human, Mouse, Rat, Dog, Frog	Polyclonal; IgG	200	500	Abcam	Ab15418
Calreticulin	Rabbit	Human, Mouse, Rat, Dog, Rabbit	Polyclonal; IgG	200	N/A	Abcam	Ab2907
HERC2	Rabbit	Human, Mouse	Polyclonal; IgG	N/A	1000	Bethyl Laboratories Inc	A301-905A-T
IDH3G	Rabbit	Human, Mouse	Polyclonal; IgG	N/A	500	Proteintech	25848-1-AP
KLHL7	Rabbit	Human, Mouse, Rat, Rabbit, Dog, Chicken, Cow	Polyclonal; IgG	N/A	500	Abcam	Ab90915
MICU1 (CBARA1)	Goat	Human, Mouse, Rat, Rabbit, Horse, Chicken, Cow, Dog, Pig	Polyclonal; IgG	200	500	Abcam	Ab115025
MTCO2	Mouse	Human, Mouse, Rat, Rabbit	Monoclonal; IgG1	200	N/A	Abcam	Ab3298
MCU	Rabbit	Human, Rat	Polyclonal; IgG	100	N/A	Abcam	Ab121499
UBE3A	Mouse	Human	Monoclonal; IgG2a	200	1000	Abcam	Ab58266
GFP	Mouse	wtGFP, rGFP, eGFP	Monoclonal; IgG1	100	1000	Abcam	Ab1218
C-myc	Mouse	Human	Monoclonal; IgG1	200	1000	Sigma-Aldrich	M4439
PCM1	Rabbit	Human, Mouse	Polyclonal; IgG	N/A	1000	Proteintech	19856-1-AP
CEP170	Rabbit	N/A	Polyclonal; IgG	N/A	1000	N/A	N/A
XPA	Rabbit	N/A	Polyclonal; IgG	N/A	500	Kind gift from Prof Majlinda Lako	N/A
Flag	Mouse	N/A	Monoclonal; IgG1	N/A	1500	Sigma-Aldrich	F3165
V5	Mouse	V5 tag	Monoclonal; IgG2a	100	1000	ThermoFisher Scientific	R960-25
β-actin	Mouse	Human, Mouse, Rat, Rabbit, Dog, Chicken, Cow	Monoclonal; IgG1	N/A	10000	Abcam	Ab6276

Table 2-4 List of Primary Antibodies

Table includes the species that the antibody was raised in and its reactivity, the type of isotype, specific dilutions for western blotting (WB) and immunofluorescence (IF) staining, as well as the supplier.

SPECIES REACTIVITY	SPECIES RAISED IN	CLASS; ISOTYPE	CONJUGATE	IF DILUTION (1/X)	SUPPLIER	CATALOGUE NUMBER
Rabbit	Goat	Polyclonal; IgG	Alexa Fluor® 488	1000	ThermoFisher Scientific	A11034
Rabbit	Goat	Polyclonal; IgG	Alexa Fluor® 568	1000	ThermoFisher Scientific	A11011
Mouse	Goat	Polyclonal; IgG	Alexa Fluor® 488	1000	ThermoFisher Scientific	A11029
Mouse	Goat	Polyclonal; IgG	Alexa Fluor® 568	1000	ThermoFisher Scientific	A11031
Mouse	Goat	Polyclonal; IgG	Alexa Fluor® 633	1000	ThermoFisher Scientific	A21052
Rabbit	Donkey	Polyclonal; IgG	Alexa Fluor® 488	1000	ThermoFisher Scientific	A21206
Goat	Donkey	Polyclonal; IgG	Alexa Fluor® 488	1000	ThermoFisher Scientific	A11055

Table 2-5 List of Secondary antibodies for Immunofluorescence microscopy

Table includes the species that the antibody was raised in and its reactivity, specific dilutions for IF and conjugated fluorophore, as well as the supplier.

2.2 Methods

2.2.1 Patient Identification

The affected children presented with rare forms of autosomal recessive disorders, including myopathic, neurological or developmental delay phenotypes. Clinical geneticists and pathologists performed the clinical evaluations at Bradford Teaching Hospitals NHS Foundation Trust or Leeds Teaching Hospitals NHS Trust.

2.2.2 Ethical Approval and Consent

Blood samples were taken from consanguineous and non-consanguineous families that were recruited for research studies, funded by a Sir Jules Thorn Award for Biomedical Research (ref. JTA/09). All DNA samples used for this research project were obtained with informed consent for research from the participants or their families (Appendix B) under ethical approval from the NRES Committee Yorkshire & The Humber, South Yorkshire (REC reference 11/H1310/1) (Appendix C).

2.2.3 DNA samples

DNA samples were obtained from affected children and their parents. Where feasible, DNA samples were also obtained from unaffected siblings or affected individuals from different sibships within the same family. DNA sampling included either blood or saliva. Blood collection was performed by standard methods, and Oragene collection kits were used for saliva collection.

2.2.4 Tissue Biopsies

Skin punch biopsies (4-6mm) were obtained under local anesthesia from a single affected child in three different families with homozygous mutations in *MICU1*, *HERC2* and *KLHL7*.

2.2.5 DNA Extraction

2.2.5.1 Peripheral Blood Samples

Genomic DNA extraction from blood lymphocytes was carried out by the Yorkshire Regional Genetics Service, using a standard salt precipitation protocol. DNA samples were re-diluted and stored in 1x Tris-EDTA buffer (pH 8.0).

2.2.5.2 Saliva Samples

DNA extraction from saliva was performed using the prepIT-L2P DNA extraction kit, followed by ethanol precipitation as advised by the manufacturer's protocol. DNA was re-suspended in 1x TE buffer for long-term storage.

2.2.5.3 Primary cell cultures

DNA extraction from primary cell cultures was by standard phenol/chlorophorm extraction. Pelleted cells were resuspended in 750µl of Lysis Buffer (See 2.1.2.9) and incubated at 37°C for one hour. 3µl of Proteinase K (23mg/ml) was added and incubated at 55°C for one hour. Phenol/chloroform (1:1) was added and the solution was vortexed until it turned milky and then centrifuged for 10 min at 13,400 x *g* . The aqueous phase was then collected into a new Eppendorf tube and chloroform 1:1 [v/v] was added, mixed by inverting the tube and centrifuged for 5 min at 13,400 x *g*. DNA was precipitated from the aqueous phase in 20µl of 5M NaCl and 2 volumes of 100% ethanol centrifuged for 5 min at 13,400 x *g*. The supernatant was discarded and 2 volumes of 75% [v/v] freshly prepared ethanol were added, mixed by inverting and centrifuged for 5 min at 13,400 x *g*. The supernatant was discarded and the pellet was allowed to air-dry and re-suspended in 50µl of EB (Qiagen).

2.2.6 Polymerase Chain Reaction (PCR)

2.2.6.1 Primer Design

Primer design was performed using the Exon Primer software (https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html), which is a Perl script that designs intronic primers for exonic PCR amplification. Exon Primer retrieves gene information from the UCSC (University of California Santa Cruz) Genome Browser and designs the primer sequence per exon using Primer 3 (http://bioinfo.ut.ee/primer3/). All primers were designed with specific parameters such as an optimum annealing temperature of 58-65 ^{o}C, a minimum of 15bp of flanking intronic sequence, excluding common SNPs and producing PCR products of 200-600bp in size. All primer sequences were checked using the BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast), to confirm that they uniquely bound to the gene of interest. Exons with small introns (<150bp) were combined and amplified together, whereas larger exons (>450bp) were amplified using more than one set of overlapping primers. Primer sequences used for PCR amplifications in this project are listed in **table 2.2**.

2.2.6.2 PCR Reaction

PCR reactions were performed in a total volume of 10µl, containing 1µl of genomic DNA (at approx. 20ng/µl), 1µl of 10µM primer mixture of forward and reverse primers (Sigma-Aldrich), 3µl of HotShot Master Mix (Clent Life Sciences) and 5µl of dH_2O. For GC rich regions, the PCR reaction was supplemented with 5x Combinational Enhancer Solution (5x CES) containing 6.7mM DTT, 2.7M Betaine, 6.7% DMSO and 5µg/ml BSA.

Reactions were run in a Veriti Thermal Cycler (Thermo Fisher Scientific) under the following conditions: denaturation was at 95°C for 3 min, followed by 35 cycles of 94°C for 30s; reaction specific annealing temperature was for 30s; extension was at 72°C for 1 min; and lastly, the reaction finished at 72°C for 5 min. Final PCR products were visualised by agarose gel electrophoresis (see section 2.2.7).

2.2.7 Agarose Gel Electrophoresis

Post-PCR or plasmid DNA samples were mixed in a 1:1 [v/v] ratio with 2x Loading Dye (see section 2.1.2.6). The gel was made by dissolving molecular biology grade agarose powder (Bioline, London, UK) in 1x TAE to a final concentration of 1.5-4% [w/v] depending on the separation required. 0.5µg/ml of ethidium bromide (Sigma-Aldrich) or 5µl per 100ml gel of Midori Green (Geneflow, Staffordshire, UK) was added to the melted agarose gel. A standard DNA size ladder (Easy Ladder I, Bioline), or a 2-log DNA ladder (New England Biolabs) for larger sized fragments, was used. 120V was then applied for 40 min and the gel was visualised on a UV transIluminator (Bio-Rad, Hemel Hempstead, UK) and displayed on Image Lab (v. 4.0) software for analysis (Bio-Rad, Life Science, Berkley, California, USA).

2.2.8 Exonuclease I – Shrimp Alkaline Phosphatase (ExoSAP) PCR purification

PCR products were purified by enzymatic treatment with Exonuclease I and Shrimp Alkaline Phosphatase (ExoSAP-IT®) (Affymetrix, Thermo Fisher Scientific) prior to sequencing. The clean-up was achieved by the addition of 1µl of ExoSAP-IT to 2.5µl (approx. 2 to 5µg) of amplicon, followed by an incubation of 37°C for 30 min and inactivation at 80°C for 15 min.

2.2.9 Sanger Sequencing

Sequencing of the purified PCR products was achieved using the BigDye® Terminator v3.1 Sequencing Kit (Applied Biosystems). The sequencing reaction was performed in a total volume of 10µl and consisted of 1µl of purified PCR product, 1µl Big Dye®

Terminator v3.1, 1.5µl Big Dye® Sequencing Buffer (5x), 1µl of either the 5' or the 3' primer at 2µM, and 5.5µl of dH_2O. The reaction mixtures were then placed in a Veriti Thermal Cycler (Thermo Fisher Scientific) for the following temperature incubations: 96°C for 1 min, followed by 45 cycles of 96°C for 10s, 50°C for 5s, 60°C for 4 min, and then hold at 4°C until precipitation.

Ethanol precipitation was performed by adding 5µl of 125mM EDTA and 60µl of 100% ethanol to each sequencing reaction. These were then centrifuged at 3100 x *g* for 30min at 22°C, followed by an inverted spin for 15s at 18 x *g*. 60µl of freshly prepared 70% ethanol was then added and the products were centrifuged again at 800 x *g* for 15 min at 4°C, followed by an inverted spin for 15s at 18 x *g*. The precipitated pellets were then left to air dry for 15 min at room temperature out of light. Dry pellets were resuspended in 10µl of "Hi-Di" deionized formamide (Applied Biosystems, Thermo Fisher Scientific) and run on an ABI 3130xl Genetic Analyzer (Applied Biosystems) using standard protocols and a POP7 polymer (Applied Biosystems). Sequencing data was analysed using 4Peaks (Mek&Tosj.com) or Seqscape v2.5 (Applied Biosystems).

2.2.10 Next Generation Sequiencing

2.2.10.1 DNA quantification

DNA samples were quantified using the manufacturer's protocol for the Quant-iT dsDNA BR assay (Life Technologies, Waltham, MA, USA). Samples were read in a Qubit fluorometer using the Qubit ds BR assay reading. DNA samples further underwent serial dilutions in order to ensure that 25ng of DNA per sample was used for downstream shearing.

2.2.10.2 Whole Exome Sequencing using SureSelect QXT method

Whole Exome Sequencing (WES) was performed using the Sure Select^{QXT} Target Enrichment kit according to the manufacturer's instructions (Agilent Technologies). Briefly, 25ng of genomic DNA was fragmented and adaptors were attached in a single enzymatic step using SureSelect QXT Enzyme Mix ILM. The reaction mixture was incubated for 10 min at 45^{o}C, and purification of the adaptor-tagged library was performed using AMPure XP beads (Beckman Coulter). Following purification the library was amplified in an 8 cycle PCR reaction using Herculase II fusion DNA polymerase, and purified again with AMPure XP beads. The size and quality of the fragmented DNA was evaluated using an Agilent 2100 Bioanalyzer and the DNA 1000 Assay (Agilent Technologies), according to manufacturer's instructions (see Appendix D).

The fragments were hybridised to the SureSelectQXT Human All Exon V6 biotinylated baits (Agilent Technologies), to enrich for exonic sequences, followed by capturing the

hybridised libraries using Dynabeads® MyOne™ Streptavidin T1 magnetic beads (Life Technologies). The captured libraries were then amplified and indexed using Herculase II Fusion DNA polymerase in a 10 cycle PCR, followed by a purification step using AMPure XP beads (Beckman Coulter). The quantity and quality of the indexed libraries was evaluated using the Agilent 2100 Bioanalzer and the High Sensitivity DNA standard procedure (Agilent Technologies) (see Appendix D). Equal molar quantities of each captured library were pooled and sequenced on the Illumina HiSeq 3000 platform (Illumina, San Diego, CA), using a 150bp paired end protocol (**figure 2-1**). Nine or ten samples were pooled per lane.

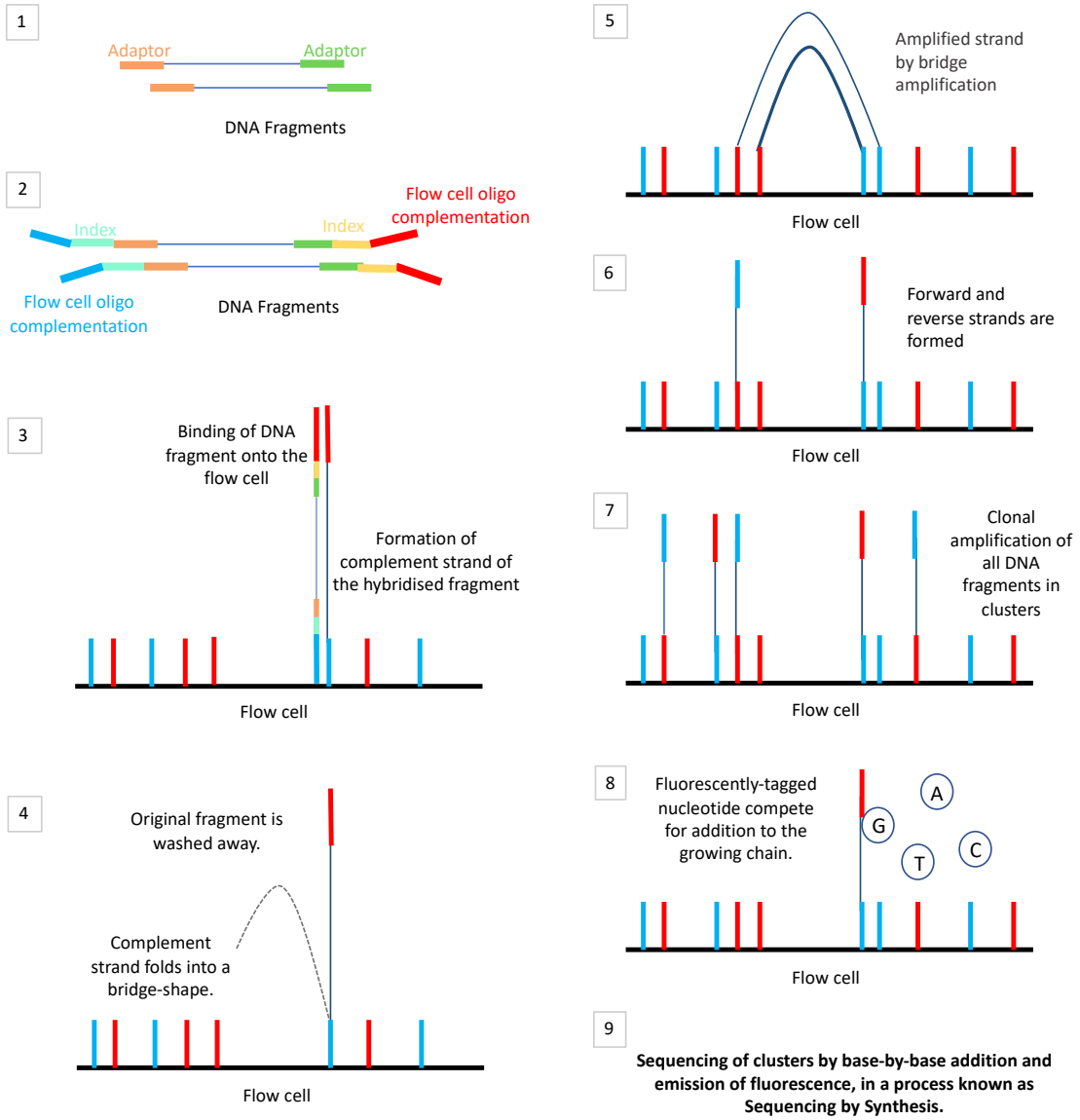


Figure 2-1 Illumina Sequencing protocol

A representation of Illumina's Sequencing protocol showing binding to flow cell and cluster amplification of fragments by forming a bridge between the two adaptors. Fluorescently-tagged terminator nucleotides are added to each strand in a "base-by-base" addition and upon laser excitation each nucleotide emits a color that is detected by the sequencer. Following the detection of one nucleotide, the terminator becomes chemically inactivated so that the next cycle of sequencing-by-synthesis can commence.

2.2.10.3 WES Data Analysis

A standard in-house bioinformatics pipeline was used for all samples. The pipeline is summarised in **figure 2.2** and the linux commands are presented in Appendix E. HiSeq sequencing data for each sample was stored in a forward (r1) and reverse (r2) FASTQ file, containing both the raw data and quality scores in a text based format. Burrows-Wheeler Aligner software package (BWA-MEM) (Li and Durbin, 2009) was used to successfully align both forward and reverse FASTQ files to the human reference genome (GRCh37, Genome Reference Consortium, human reference assembly build 37) creating a binary alignment map (bam) file. Following alignment, the bam file was sorted in chromosome order using Picard tools v.2.5.0 (http://picard.sourceforge.net). Picard tools were also used to mark PCR and optical duplicates present in the data, and the Genome Analysis Toolkit (GATK, Broad Institute, USA) was utilised for IndelRealignment, where small mismatching bases were better realigned, and for BaseRecalibration, that pre-processed data for systematic errors and estimated the accuracy per base call (DePristo et al., 2011).

HaplotypeCaller (GATK) was then used for variant calling, resulting in a variant call format (vcf) file output. HaplotypeCaller was run on each sample individually producing separate gvcf files. For each family, the gvcf files were subsequently combined and genotyped creating a combined vcf for more effective filtering and segregation assessment. The combined vcf was then passed through hard filtering utilizing the VariantFiltration programme (GATK). To achieve this, SNPs and indels were separated before proceeding to hard filtering. Hard filtering removes any variants that have annotations or statistical values above or below certain thresholds. The parameters for the statistical annotation are summarised in **table 2.6**.

Variation	Abbreviation	Character	Parameter threshold
Indel	QD	Quality by Depth	<2.0 reads
	FS	Fisher Strand PHRED; scaled p-value for strand bias	>200.0
	ReadPosRankSum	Mann-Whitney test for distant of alternate allele from the end of read	<-20
SNP/MNP	QD	Quality by Depth	<2.0 reads
	FS	Fishe Strand PHRED	>60.0
	ReadPosRankSum	Mann-Whitney test for distant of alternate variant from the end of read	<-8
	MappingQualityRankSum	Mann-Whitney Rank Test for Mapping Qualities (reference vs alternate allele)	<-12.5

Table 2-6 Summary of parameters of hard filtering

Variants passing the hard filtering were further filtered using various databases at a minor allele frequency (MAF) of either ≤1% when modeling rare recessive inheritance in a subject family or ≤ 0.1% for suspected *de novo* cases. The first database used for filtering was the National Center for Biotechnology Information's (NCBI) database for SNPs (dbSNP) version 146 (Sherry et al., 2001), while cross-referencing and maintaining any variants classified as pathogenic on ClinVar. Variants were also filtered on the NHLBI Exome Sequencing Project (ESP) database (http://evs.gs.washington.edu/EVS/) and the Broad Institute's Exome Aggregation Consortium (ExAC) version 0.3 database (http://exac.broadinstitute.org). Variants were also checked manually on the Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/) when it became available.

Variants were subsequently annotated using the Variant Effect Predictor (Ensembl) to determine any non-synonymous, likely functional variants (McLaren et al., 2010). The resulting vcf was then filtered using adapted perl scripts 'vcf hacks' (v.0.2.0) (written by Dr David Parry, https://github.com/gantzgraf/vcfhacks) to filter for biallelic or homozygous variants. In most of the cases, a family pedigree was incorporated in the analysis by the use of a .ped file allowing for identification of variants based on segregation. The resulting putative mutations were assessed for pathogenicity predictions using SIFT (http://sift.jcvi.org) (Ng and Henikoff, 2003), PolyPhen2 ((http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2010) and Condel

(http://bg.upf.edu/fannsdb/), (Gonzalez-Perez and Lopez-Bigas, 2011) and ranked according to the Combined Annotation Dependent Depletion (CADD) score version 1.3 (http://cadd.gs.washington.edu) (Kircher et al., 2014). The variants that passed all the filtering steps were annotated with summaries of gene function and mouse model phenotypes where applicable. The resulting file was then exported in to Microsoft Excel format, with all the pathogenicity predictions and annotations attached. Variants with CADD scores ≤15 were excluded from further variant assessment and interpretation. Variants with CADD scores above 15 represent the top 3% of deleterious variants, and variants with CADD scores above 30 represent the top 0.1% of deleterious variants in the human genome. The workflow of the bioinformatics analysis of WES is summarised in **figure 1.2** and the actual command line is presented in Appendix E.

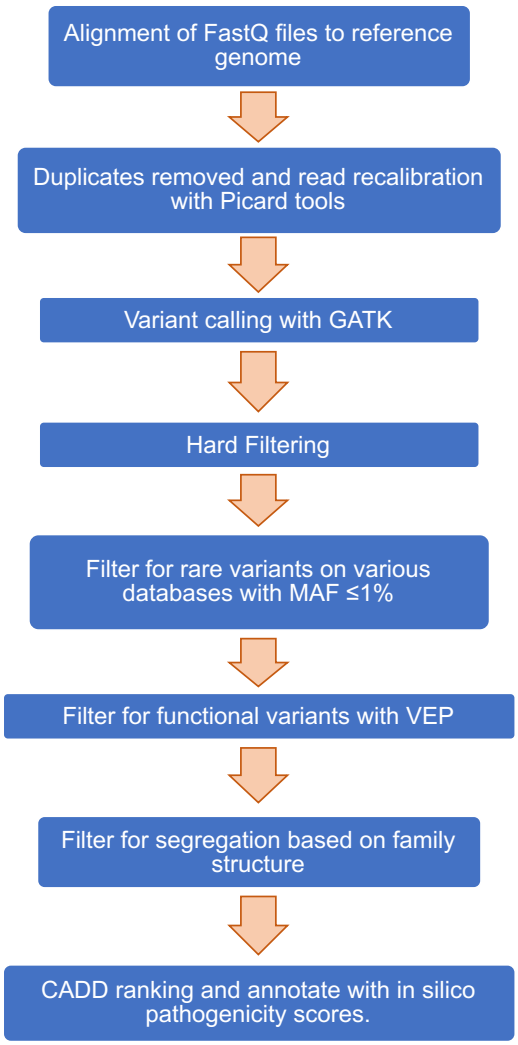


Figure 2-2 Bioinformatics pipeline for WES data analysis.

The flow chart above summarises the optimised bioinformatics pipeline used for WES data analysis for all of the cases outlined in this study.

2.2.10.4 WES Quality Control (QC)

The first level of Quality Control (QC) was applied to the raw FastQ sequence data that was collected from high-throughput sequencing. This was run on the FastQC program (Babraham Bioinformatics) and it provided a simple checkpoint for the quality of data obtained. **Figure 1.3** illustrates how a high quality data set should look compared to a low quality one.

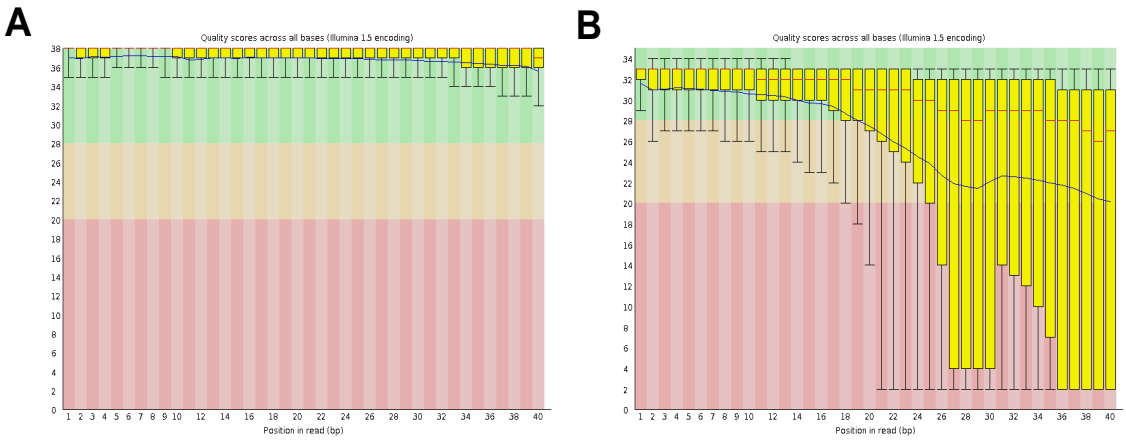


Figure 2-3 Representation of FastQC outputs

A – Representation of high quality data. B – Representation of low quality data. In both cases the yellow boxes represent the interquartile range of the data, the red line marks the median value, the blue line marks the mean quality and the upper and lower whiskers indicate the 10% and 90% range of the data points. The y-axis that indicates the per base sequence quality score is separated into three coloured regions. High quality scores are within the green region, reasonable quality scores in the orange, and poor quality scores in the red.

The second check point for QC assessment was performed on the final bam file using Picard tools CollectMultipleMetrics (https://broadinstitute.github.io/picard/command-line-overview.html#CollectMultipleMetrics). This command runs various QC metrics including insert size metrics, quality score distribution, mean quality by cycle and base distribution by cycle for sequencing artifact control. The QC output of this method is in a graphical representation that can be compared across samples and runs. **Figure 1.4** demonstrates the difference in quality between QC outputs for the same sample sequenced on two different runs. The QC metrics for each sequencing run throughout this project are summarised in Appendix F.

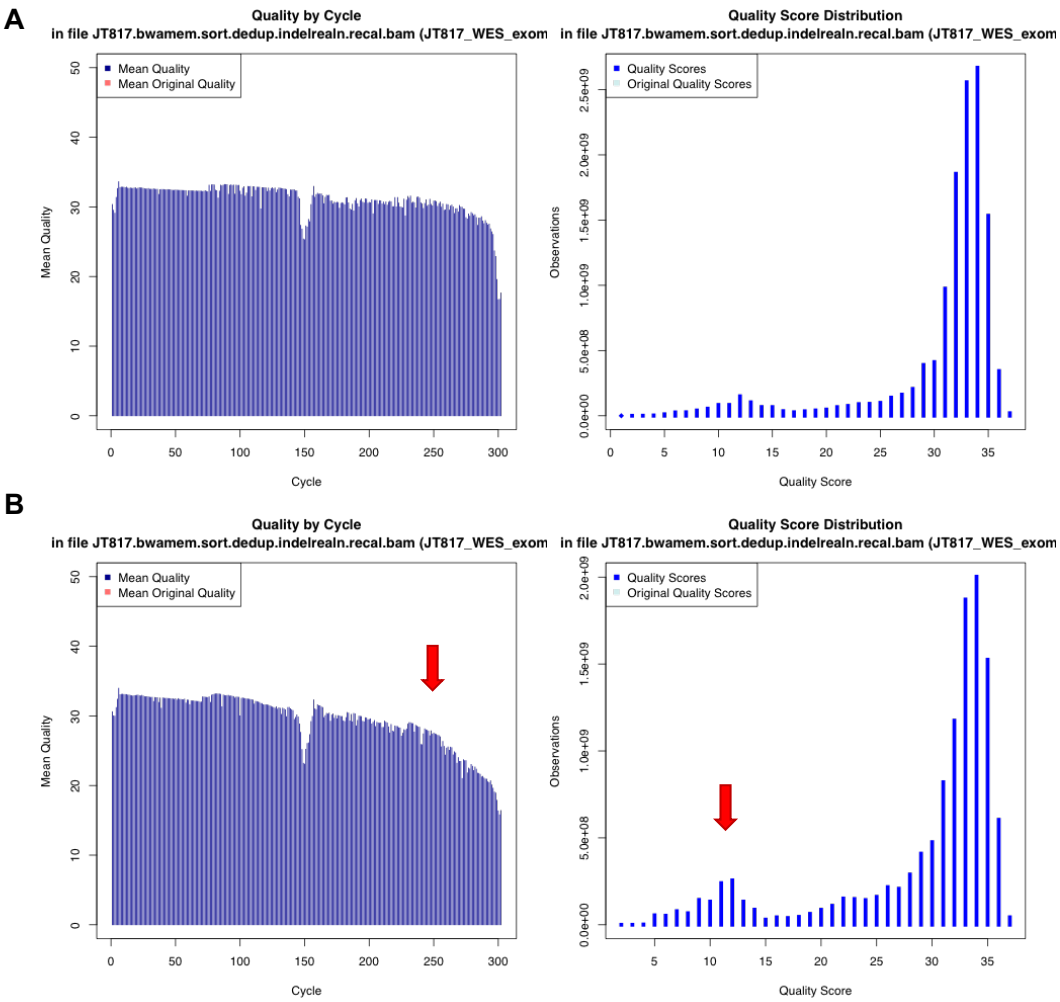


Figure 2-4 Representation of QC metrics from CollectMultipleMetrics command

A – Representation of good quality QC with mean Quality by Cycle staying above 30 and quality score distribution is mostly above QS of 30. B – Representation of bad quality QC with mean of Quality by cycle dropping below 30 early on in the second cycle and increased reads with QS less than 30 (marked with red arrow). The data with bad quality QC were obtained from a unsuccessful run where the machine failed due to temperature issue.

Finally, the Depth of Coverage tool (GATK) was used to assess average read depth across the exome and evaluate the percentage of regions with coverage <5 reads. A table outlining the output of Depth of Coverage for each sequencing run throughout this project can be found in Appendix G.

2.2.10.5 Variant Interpretation and Pathogenicity Assessment

As stated above, the final gene list with CADD scores ≥10 was inspected for variant interpretation. Pathogenicity assessment was initially evaluated from the Excel output of the bioinformatics pipeline, by looking at CADD scores, and other *in silico* pathogenicity prediction tools that included PolyPhen2, SIFT and Condel. Nonsense and splice-site

variants were initially flagged for the most damaging and potentially pathogenic variants, but other missense variants were also interpreted by other means. For example, a literature review was conducted for the best candidate genes to investigate a potential role in neurodevelopment or other organ-specific function based on the family's phenotype. In addition, interacting partners were explored using STRING (https://string-db.org/) and protein expression levels and localization were inspected on UNIPROT (https://www.uniprot.org/). Furthermore, the phenotypes of any available mutant models were inspected on MGI (http://www.informatics.jax.org/) for mouse models or ZFIN for zebrafish models (https://zfin.org/) to assess any phenotypic overlap. Where possible, protein modelling was also used in order to compare the impact of a variant on the overall structure of the protein. Programs used for this purpose included i-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), TmHelix TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and Swiss Pdb-viewer (https://spdbv.vital-it.ch/).

2.2.10.6 Autozygosity Mapping using WES data

Autozygosity mapping was performed when the family pedigree suggested consanguinity. Homozygous regions identified were then used for prioritising variants within them. Autozygosity mapping was performed using AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (written by Dr Ian Carr, University of Leeds) that identifies shared autozygous regions between affected individuals in a pedigree and displays the output in a circular ideogram of chromosomes 1-22. The regions can be identified from WES data in the form of a vcf file prior to SNP filtration.

2.2.10.7 Copy Number Variant Identification using WES data

Most of the cases recruited to the study have been previously tested with Array CGH as part of standard diagnostic workflows in Clinical Genetics. However, CNVs smaller than 100kb cannot be picked up by Array CGH. All the cases that have been recruited to this project have therefore been further analysed for CNVs using the WES data, regardless if any Array CGH had already been performed.

Exome Depth was the program used in order to identify CNVs from WES data (Plagnol et al., 2012). The program was run in R and it allowed for comparison of read depth between the test sample and 8-10 control samples. The control samples used for comparison had to be unrelated to the test sample and each other, and with the least possible technical variations during the library preparation. For example, unrelated samples run on the same sequencing run and prepared on the same day as the test

sample were considered ideal controls. Also, the controls samples had to be exome sequenced using the same exome libraries as the test sample.

The Exome Depth output was a summary table in the form of a .csv file, with all identified variations. Common CNVs identified in control populations were annotated (Conrad et al., 2010). All calls were ranked by the Bayes Factor, a statistical index defined by the log10 likelihood ratio of the reads of a CNV call divided by the normal copy number reads. CNV calls were then manually reviewed to evaluate read ratio and genes involved. Any candidate calls were further inspected using the Broad Institute's Integrative Genomics Viewer (IGV) (Robinson et al., 2011) and crossed-reference with any available literature for possible connections to neurodevelopmental or neuromuscular disorders.

2.2.11 Microbiology

Microbiology experiments were carried out in a dedicated laboratory. Luria Bertani (LB) broth was made up by dissolving 20g of LB powder (Sigma-Aldrich) in 1l of dH_2O and autoclaved using a bench top autoclave (Prestige Medical, Coventry UK). SOC medium was commercially obtained (New England Biolabs). Agar was made by dissolving 10g of LB powder (Sigma-Aldrich) and 10g of agar powder (Merck) in 500ml of dH_2O, and then autoclaved in a bench top autoclave (Prestige Medical, Coventry UK). After autoclaving, agar was allowed to cool to about 50^{o}C in a water bath and a relevant antibiotic was added at the appropriate concentration (see **table 2.6**). 25ml of agar was poured in microbiology plates in a hood, left to cool and stored at 4^{o}C up to 3 months.

Antibiotic	Working Concentration	Supplier
Ampicillin	50µg/ml	Melforth labs (Suffolk, UK)
Kanamycin	50µg/ml	Sigma-Aldrich
Spectinomycin	100µg/ml	Sigma-Aldrich

Table 2-7 List of antibiotics used for microbiology purposes

Table includes the working concentrations of each antibiotic and the supplier.

2.2.11.1 Gateway Cloning

Gateway cloning is a highly efficient and fast method of cloning that is based on the recombination properties of bacteriophage lambda I to integrate its DNA in *E. coli*. This cloning method is based on the specific recombination sites known as *att*P in bacteriophage lambda and the *att*B site in *E. coli*. *In vitro* cloning reactions based on these properties were made two directional, or reversible, using specific *att* site development. The reactions are known as either the LR and the BP reaction and are summarised in **figure 2.5**

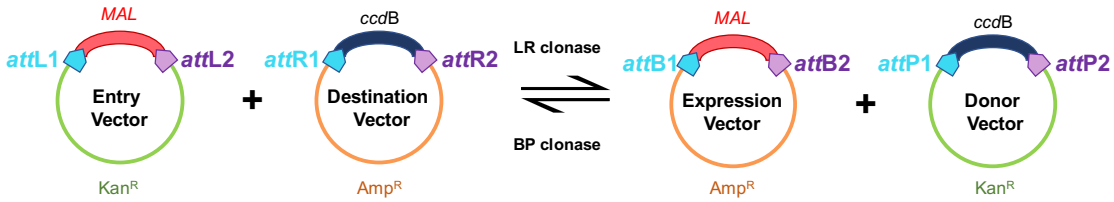


Figure 2-5 Gateway Cloning technology

Gateway cloning is an efficient and reversible method of transferring DNA fragments between plasmids. This method utilises the site-specific recombination between the '*att*' sites. As illustrated above, the LR reaction recognises the *att*L and *att*R sites between an entry vector and a destination vector respectively, producing the expression vector of interest. The reversible reaction is known as the BP reaction and occurs between *att*B and *att*P sites between an expression vector and a donor vector respectively, giving an entry clone of interest.

For the purposes of this project, only the LR Clonase Gateway reaction (Thermo Fisher Scientific) was performed, as described in the manufacturer's instructions. 150ng of Gateway Entry (pENTR) vector and 150ng of Destination vector (pDEST) were added to 2µl 5xLR Clonase Reaction Buffer and 4µl TE buffer. The LR Clonase enzyme mix was thawed on ice for 2 min and briefly vortexed. 1µl of the enzyme mix was added to the reaction mixture, vortexed briefly and incubated for 1hr at 25 °C. Upon completion of incubation, 2µg of proteinase K was added to the reaction mixture and incubated for 10 min at 37°C to terminate the reaction. The resulting plasmids were transformed in *E. coli* DH5-Alpha Competent cells (New England Biolabs) (see section 2.2.11.4) and DNA extraction was obtained by mini or maxi preps from bacterial cultures (see section 2.2.11.5 and 2.2.11.6 respectively). All cDNA clones were confirmed by Sanger sequencing, using the primers listed in **Table 1.3**, or other commonly used plasmid primers.

2.2.11.2 In-Fusion Cloning

In-Fusion Cloning technology allows for a fast and directional method of cloning of one or more fragments of DNA into any vector of interest. The In-Fusion HD Cloning Kit (Takara Bio USA, Inc) offers efficient fusion between the DNA fragments and the linearised vector by precisely recognizing 15bp overhangs at each end. The overhanging 15bp can be engineered by designing appropriate primers for PCR amplification of a desired sequence, that will be complementary to the sites where the vector has been linearised. The procedure of In-Fusion Cloning is summarised in **figure 2.3**.

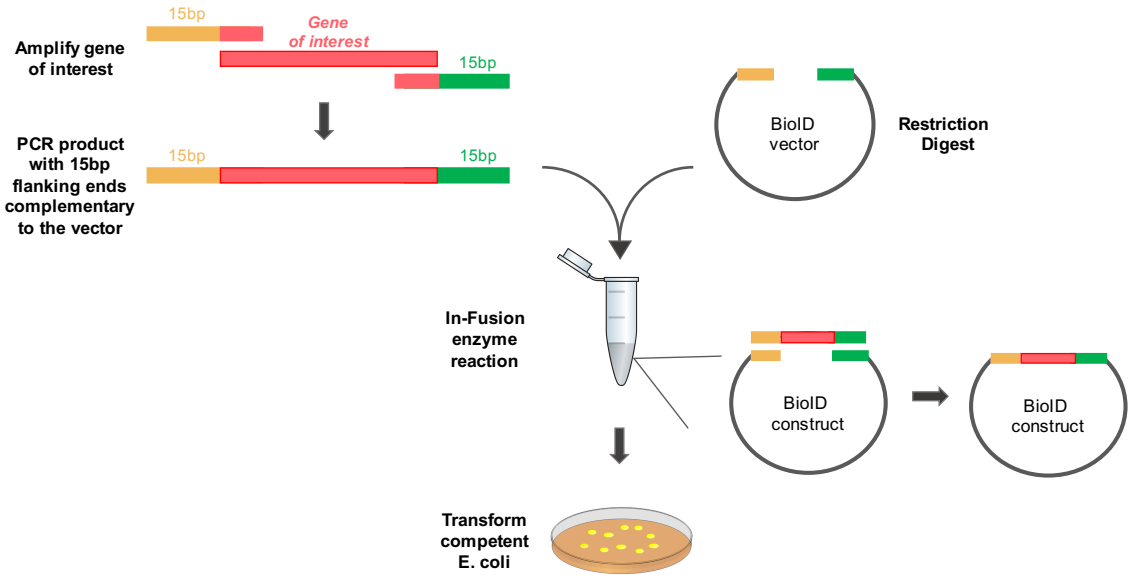


Figure 2-6 In-Fusion Cloning technology

In-Fusion Cloning is a highly efficient method of cloning that is comprised of two separate steps. Initially, the desired vector needs to be linearised by restriction enzyme digest. Secondly, the gene of interest is amplified by specific primers designed to have 15bp extensions homologous to the vector ends. Once the PCR product is purified, both products are added to a single tube reaction where the In-Fusion Enzyme allows for the homologous overhangs to join and form the new plasmid. In-Fusion cloning was used to construct the BioID vectors.

In-Fusion Cloning was carried as described by the manufacturer's protocol. The desired vector (pcDNA3.1_mycBioID_N-term – see appendix H) was linearised by restriction enzyme digest (see section 2.2.12) using *Eco*RI and *Hin*dII. The linearised vector was then run on an agarose gel and the correct size band was extracted from the gel and purified using a Gel Extraction Kit (Qiagen). Depending on the specific sites of enzymatic digest, the appropriate primers were designed for the PCR amplification step, including the 15bp extensions that are complementary to the linearised vector ends. Primers were designed manually, but confirmed using the online In-Fusion PCR primer design tool (http://www.takarabio.com). The PCR amplification was performed from a pENTR clone, thus 1ng of DNA was used for the PCR reaction. PCR conditions were set up according to the Fusion HF Enzyme Guidelines and the PCR product was purified using a PCR Clean up Kit (New England Biolabs). Once both fragments were purified, the cloning reaction was set up as follows: 1µl of 50-100ng of linearised vector, 1µl of purified PCR fragment, 2µl of 5X In-Fusion HD Enzyme Premix (Clontech Laboratories), and 6µl of dH_2O. The reaction mixture was incubated at 50° C for 15 min and the newly formed plasmid was then transformed into competent cells (see section 2.2.11.4).

2.2.11.3 Site Directed Mutagenesis

Site-directed mutagenesis (SDM) was achieved using the Q5 site-directed mutagenesis kit (New England Biolabs). The procedure was carried out as described by the manufacturer's instructions. Primers were designed using the NEBaseChanger website (http://nebasechanger.neb.com), and are listed in **table 1.3**. Briefly, the reaction comprised of 1µl plasmid DNA (25ng/µl), 12.5µl Hot Start High-Fidelity Master mix (2x), 1.25µl of forward primer (10µM), 1.25µl of reverse primer (10µM), and 9µl dH_2O. The reaction mixture was then placed in a Veriti Thermal Cycler (Thermo Fisher Scientific) for a PCR with the following conditions: 98°C for 30s, 25 cycles of 98°C for 10s, 'X'°C for 30s and 72°C for 'Y's, then 72°C for 2 min and hold at 4°C. For these conditions 'X' represents the appropriate temperature for primer annealing, as calculated by NEBaseChanger, and 'Y' stands for the elongation time for each plasmid which is equal to 30s per kb of plasmid DNA. At PCR completion, the 'KLD' step was performed, which stands for Kinase, Ligase and DpnI. For this reaction, 1µl of PCR product was mixed with 5µl 2x KLD reaction buffer, 1µl 10x KLD enzyme mix and 3µl dH_2O. The reaction mixture was pipette-mixed and incubated at room temperature for 5 min. The final mutated plasmid was transformed into competent cells (see section 2.2.11.4).

2.2.11.4 Bacterial Transformation and cultures

An aliquot of *E. coli* α-Select Gold DH5-alpha chemically competent cells (Bioline) was allowed to thaw on ice. 1µl of the plasmid was added to 25µl of competent cells, flicked briefly to mix gently and then incubated for 30 min on ice. After incubation was completed, a heat shock at 42 °C for 45s was performed in a water bath. The reaction mixture was then immediately transferred on ice for 2 min, followed by recovery in SOC medium (New England Biolabs). 200µl of prewarmed SOC medium was added to each transformation mixture and then placed in an orbital shaker (Excella E25, Eppendorf, Hamburg, Germany) at 250rpm, 37°C for 1 hr. 50µl of the bacterial mixture was spread on an LB Agar plate, with the appropriate antibiotic based on the resistance gene in each plasmid. To increase colony numbers, if required, the bacterial mixture could be spun down at 800 xg for 1 min, most of the supernatant discarded and the entire transformation mixture spread on an LB Agar plate. The plates were incubated for 10-16 hr at 37 °C to allow for bacterial growth.

After the 10-16 hr incubation, 5ml cultures were prepared in universal tubes. 5ml of LB broth and the appropriate antibiotic were added first and then single colonies were selected using a pipette tip. The pipette tip with the bacterial colony was left in a universal container and allowed to grow. The cultures were incubated in an orbital shaker (Excella E25, Eppendorf, Hamburg, Germany) at 250rpm at 37°C overnight, to ensure adequate aeration of the culture.

DNA was then extracted from the bacterial samples using a Mini Prep Kit (Qiagen) (see section 2.2.11.5) and DNA sequence was verified using Sanger sequencing. Samples with the correct sequence were further grown into 200ml LB cultures in microbiology conical flasks with the appropriate antibiotic and purified DNA was extracted using a Maxi Prep Kit (Qiagen) (see section 1.2.11.6).

2.2.11.5 Mini Preps of Plasmid DNA

Bacterial colonies were grown into cultures in 5ml LB broth at 37°C overnight in an orbital shaker. Mini prep (Qiagen) was performed according to the manufacturer's instructions. 1ml of the bacterial culture was pelleted by centrifugation at 10,000 x g for 3 min, and pellets were resuspended in 250µl of resuspension buffer (P1). 250µl of alkaline lysis buffer (P2) was added and the reaction was mixed by inverting the tube 4-6 times, followed by the addition of 350µl of neutralizing buffer (N3). The mixture was again mixed by inverting the tube 4-6 times, and then centrifuged for 10 min at 17,900 x g. The supernatant was added to a QIA prep spin column and centrifuged for 60s at 17,900 x g to bind DNA to the column. A first wash with 0.5ml of buffer PB and centrifuged for 60 s was performed. Columns were further washed with 0.75ml of buffer PE and centrifuged again for 60 s. For the DNA to be eluted from the spin column, 30µl of buffer EB was added, left to stand for 1 min. and then centrifuged for 1 min. at 17,900 x g. A Nanodrop-1000 Instrument (Thermo Fisher Scientific) was used to measure the DNA concentration prior to Sanger sequencing.

2.2.11.6 Maxi Preps of Plasmid DNA

Maxi preps (Qiagen) were used for large-scale plasmid growth and DNA extraction. For maxi preps, bacterial cultures were grown in 200ml of LB broth at 37°C for 12-16 hr and the procedure was performed according to the manufacturer's protocol. Bacterial cells from the whole culture were pelleted by centrifugation at 4000 x g for 25 min at 4°C, and resuspended in buffer P1. 10ml of alkaline lysis buffer (P2) was added, mixed by inverting 4-6 times and incubated at room temperature for 5 min. After incubation, 10ml of neutralizing buffer (P3) was added and mixed by inverting again. The lysate was poured into a QIAfilter Cartridge and incubated for 10 min at room temperature. Following this, 10ml of buffer QBT was utilised to equilibrate the Qiagen-tip and allowed to drip through. The lysate was discharged through the QIAfilter Cartridge into the Qiagen-tip and allowed to enter the resin by gravity. The DNA was then bound and washed twice with 30ml of buffer QC. 15ml of buffer QF was used to elute the DNA, followed by precipitation with 10.5ml of isopropanol and centrifugation at 15,000 x g for 30 min at 4 °C. The supernatant was discarded and 5ml of 70% ethanol was added and centrifuged for 10 min at 15,000 x g. The supernatant was discarded again and the DNA pellet was

left to air-dry for 10 min. The pellet was re-dissolved in 300µL of sterile TE buffer, and DNA concentration was measured using the Nanodrop-1000 Instrument (Thermo Fisher Scientific).

2.2.12 Restriction Enzyme Digest

Restriction enzyme digest was carried out in order to confirm the correct organisation within a plasmid, or to verify that an insert had been successfully incorporated in a vector after a cloning reaction. Restriction enzyme digest was also used for In-Fusion cloning to create specific ends for cloning. The digest was performed using a specific restriction enzyme according to the site within the construct. SNAP gene viewer (http://www.snapgene.com) was used to visualise the map of the construct and choose the correct site for enzymatic digestion. The digestion was achieved using the correct concentration of the enzyme and the appropriate buffer (New England Biolabs), and incubated on a thermocycler according to the manufacturer's instructions. The resulting fragments were checked using a 2% agarose gel electrophoresis (see section 1.2.7).

2.2.13 Cell Culture

Routine cell culture was performed using standard aseptic techniques in a dedicated tissue culture room. Primary cultures of patient fibroblasts were grown in DMEM/F12 (1:1) + GlutaMAX™ (ThermoFisher Scientific) nutrient mixture containing 10% Fetal Bovine Serum (FBS) and 1% penicillin/streptomycin. MDCK cells were grown using Dulbecco's Modified Eagle's Medium High Glucose (Sigma-Aldrich) supplemented with 10% FBS, and hTERT-RPE1 cells were cultured in DMEM/F12 (1:1) containing GlutaMAX™ (ThermoFisher Scientific) nutrient mixture and 10% FBS. Cell cultures were kept in 5% CO_2 humidified tissue culture incubators at 37°C, and primary cell lines were kept in a dedicated incubator under the same conditions. The cells were passaged twice a week, unless required otherwise to avoid over-confluent cultures (see section 2.2.14).

2.2.14 Cell Passage and Harvesting

Cell lines were usually grown in T75 canted neck cell culture flasks with vented cap (Corning ®) and were passaged when the cells were around 80-90% confluency. During cell passaging, growth media was removed and the cells were washed in 5ml 1x PBS. Once the PBS was removed, 2ml of 1x Trypsin/EDTA was added to the culture and left for 2-5 min in the 37°C/ 5% CO_2 incubator, for the cells to lift into suspension. 8ml of growth media was then added to the culture and the appropriate volume of cells was transferred into a new T75 flask, supplemented with fresh warmed media. All cell lines were split according to the supplier's instructions: MDCK cells 1:12 twice weekly, hTERT-

RPE1 cells 1:10 twice weekly, HDF 1:8 weekly, *MICU1*-HDF, *HERC2*-HDF and *KLHL7*-HDF 1:5 weekly. When cells were seeded for specific experiments, a specific cell number was required to be seeded to accomplish the correct confluency for the experiment. Cell number was determined using a Trypan Blue stain (Invitrogen) and Countess® cell counter and slides (Invitrogen). Cell lines were passaged up to a maximum of 30 passages before being discarded. Primary cell lines were only used up to passage number 15.

When cell lines were harvested for storage the same procedure was carried out, but once the cells were in suspension in Trypsin/EDTA, the growth media was added and the cells were collected in a 15ml Falcon tube. The cell suspensions were then pelleted by centrifugation at 200 x *g* for 5 min. The cell pellets were resuspended in freeze-down media (45% DMEM/F12 GlutaMAX™ or Dulbecco's Modified Eagle's Medium High Glucose (Sigma-Aldrich) according to the cell type, 45% FBS and 10% DMSO) and stored in 1ml volumes in cryovials. The cryovials were allowed to cool overnight at -80°C in a "Mr Frosty" cryo-cooling container and then stored for long-term use in liquid nitrogen stores.

2.2.15 Transient Transfection for Over-expression

Cell cultures were grown in 24 well plates for immunofluorescence experiments, 6-well plates for optimization of protein experiments and T75 flasks for experiments for which larger amounts of proteins were needed such as co-Immunoprecipitation assays or BioID experiments. Independent of where the cells were cultured, cells were allowed to reach 60-70% confluency before transfection. If immunofluorescence microscopy was going to be performed, the base of the wells was covered with glass coverslips treated with acetone and ethanol. For preparing the transfection complexes, Opti-MEM™ was mixed with Lipofectamine®2000 in a fresh Eppendorf tube, mixed gently by flicking and then incubated for 5 min. After incubation, plasmid DNA was added to the Opti-MEM™/Lipofectamine®2000 mixture, vortexed, spun down and incubated for another 20 min for the plasmid to be encapsulated in the lipid bilayer of Lipofectamine®2000. The specific amounts of each reagent required for each plate or flask are summarised in **table 1.7**.

	Optimem (µl)	Lipofectamine (µl)	Plasmid DNA (µg)
24 well plate	62.5	1.5	0.25
6 well plate	250	6	1
T75 flask	1500	45	6

Table 2-8 List of specific amounts of reagents used for transfections

Table includes the amount of each reagent used for transfections in 24 and 6 well plates, as well as T75 flask.

Before transfections, the growing media was replaced with Opti-MEM™ and the transfection complexes were added. The cells were left to incubate for 3-5 hr and then the media was changed back to normal growing media. Assays were performed after 48-72 hr, but for immunofluorescence microscopy a 24 hr period was allowed for cell growth in order to avoid over-confluent coverslips.

2.2.16 Phenylbutyrate Treatment

MDCK cells transfected with *MAL* constructs (wildtype and mutant) were further treated with sodium 4-phenylbutyrate, a well-described chemical chaperone (Brookes et al., 2014). Sodium 4-phenylbutyrate was diluted in culture media at a final concentration of 0.5mM and incubated for 24 hr.

2.2.17 Immunofluorescence and Confocal Microscopy

For immunofluorescence staining, transfected or non-transfected cells were fixed on coverslips at 80% confluency. Fixation was performed either with *para*-formaldehyde (PFA) or with methanol. For PFA fixation, 4% [w/v] of PFA was added to the cells for 20 min at room temperature and then permeabilised with 0.01% [v/v] Triton X-100 for 5 min. For methanol fixation, ice-cold methanol was added to the cells and incubated at -20°C for 5 min. Fixed coverslips were then blocked with either 1% [w/v] Marvel milk solution or 1% [w/v] BSA depending on the antibodies used downstream. All solutions were spun before use to remove particulates. Blocking solutions were added to each well containing a coverslip and incubated for 30 min on a rocker. Primary antibodies (see **table 1.4**) were made up in 1% Marvel milk solution or 1% BSA according to manufacturer's instructions and incubated with coverslips in a humid chamber overnight at 4°C. Coverslips were then washed three times in 1x PBS, followed by incubation with the suitable AlexaFluor® conjugated secondary antibodies (Life Technologies) and DAPI for nuclear staining. Secondary antibodies (see **table 1.5**) and DAPI were added to 1% Marvel milk solution or 1% BSA and incubated in the humid chamber for 1 hr out of light. Once the incubation

was completed, the coverslips were washed three times in 1x PBS and twice in dH_2O. Coverslips were allowed to set on slides using 10µL Mowiol® (Sigma-Aldrich).

Confocal imaging was performed on the Nikon A1R confocal microscope and all images were processed by NIS-Elements Confocal 4.5 (Nikon) software. For post-capture image processing and analysis, Fiji (https://fiji.sc) was used. Corrected Total Cell Fluorescence (CTCF) or co-localisation assessments were done using Fiji's plug-ins. For co-localisation analysis between two different colour channels, the colocalization threshold plug-in was used by setting the region of interest and the channels of interest. Further statistical analysis was performed on the Rcoloc values obtained from each cell, in three biological replicates. For CTCF, a designated region was chosen for all the measurements and the following formula was used:

Corrected total cell fluorescence (CTCF) = Integrated Density – (Area of selected cell X Mean fluorescence of background readings)

Following the same pattern of analysis, statistical analysis was performed on the CTCF values obtained from each cell, in three biological replicates. More details on statistical analysis are summarized in section 2.2.23.

2.2.18 Live cell Imaging

2.2.18.1 Live cell imaging using MitoTracker® Green FM

Primary fibroblasts were seeded in 35mm 4-well imaging microplates (Ibidi®) 18-20 hours prior to imaging, aiming for about 70-80% confluency. The MitoTracker® Green FM (Invitrogen) was initially dissolved in DMSO to obtain a final concentration of 1mM, and was then diluted with DMEM/F12 GlutaMAX™ (Sigma-Aldrich) to reach a working concentration of 100nM. The culture media was then removed and replaced with pre-warmed (37°C) staining media containing the MitoTracker probe, and incubated at 37°C for 45 min. After incubation was completed, the staining media was replaced with fresh pre-warmed growth media and the microplate was placed in a Nikon BioStation IM instrument for live cell imaging. Imaging was usually set to a 24-36 hr period. The videos and images captured were analysed using Image J software (https://imagej.nih.gov/ij/).

2.2.18.2 Live cell imaging for GFP-tagged protein

MDCK cells were seeded in 35mm 4-well imaging microplates (Ibidi®) 18-20 hours prior to imaging, aiming for about 70-80% confluency. Once this was reached, the cells were transfected using Lipofectamine2000 (see section 1.2.15) in the same DNA to Lipofectamine ratio as for the 24-well plates. Cell were incubated with the transfection complexes for 3-4 hr and then the media was changed to fresh Dulbecco's Modified Eagle's Medium High Glucose (Sigma-Aldrich), and the plate was placed in a Nikon BioStation IM instrument for live cell imaging. Imaging was set to a range of 24-48 hr. The videos and images captured were analysed using Image J software (https://imagej.nih.gov/ij/).

2.2.19 Western Blotting

2.2.19.1 Whole cell extract

Whole cell extracts (WCE) were collected from cell cultures following cell lysis. Prior to cell lysis, the cells were washed twice with ice-cold PBS. For cell lysis, according to the protein of interest the appropriate lysis buffer was added to the cells at a volume of 300µl for a T75 and 50µl for a 6-well plate. Cells were then incubated with lysis buffer for 5 min on ice and scraped to collect WCE in Eppendorf tubes. Samples were further incubated on ice for 30 min, while pipette mixing every 10 min. After the incubation, the lysates were centrifuged at 14000 x *g* for 15 min, collecting only the soluble fraction of the WCE. At this stage, protein concentration was calculated using a Bradford Assay following the manufacturer's instructions (Bio-Rad, Life Science, Berkley, California, USA). For protein determination, sample absorbance was measured on a spectrophotometer at 595nm and compared to a range of BSA standards (Sigma).

2.2.19.2 SDS-PAGE and Western Blotting

Equal amount of protein per sample (around 10-20µg) was mixed with 2xSDS loading buffer and 10x Reducing agent (DTT). Samples were then heated at 95 °C for 5 min on a heat block, and electrophoresed in NuPAGE™ 4-12% Bis-Tris gels or NuPAGE™ 3-8% Tris-Acetate gels (ThermoFisher Scientific) for proteins over 200kDa. All samples were run alongside Precision Plus Protein™ All Blue Protein Standards (Bio-Rad) or HiMark™ Pre-Stained Standard (ThermoFisher Scientific) for higher molecular weights. Protein gels were run in NuPAGE™ MES SDS Running Buffer or NuPage™ Tris-Acetate SDS Running Buffer (ThermoFisher Scientific) for 1.5 hr at 120V. Transfer of proteins to PVDF membranes (ThermoFisher Scientific) was performed in NuPAGE™ Transfer

Buffer (ThermoFisher Scientific) supplemented with 10% methanol. Transfer was run for 1.5 hr at 30V, or 4 hr at 30V at 4°C for larger proteins.

2.2.19.3 Antibody staining and membrane visualization

Membranes were blocked in 5% Marvel milk solution in 1xPBST (see section 2.1.2.1) or 1% BSA solution for 1 hr, followed by incubation of membranes in primary antibody solutions for 1 hr at room temperature or overnight at 4°C. Membranes were then washed four times with 1xPBST in 10min intervals, followed by 1 hr incubation at RT with the appropriate HRP-tagged secondary antibody solution (Dako, Agilent Technologies) at a concentration of 1:5000. Membranes were washed four times again with 1xPBST in 10-min intervals, followed by membrane developing using SuperSignal West Femto kit (ThermoFisher Scientific). Bio-Rad molecular image ChemiDoc™ MP Imaging System with a UV transilluminator was used to acquire membrane images. Analysis of membranes was performed on Image Lab (v. 4.0) software (Bio-Rad, Life Science, Berkley, California, USA) and all band intensity measurements were compared to beta-actin loading control for quantification.

In order to re-probe the membrane with different antibodies, membrane stripping was required. For this purpose, the membrane was incubated in 5ml of Restore™ Plus stripping buffer (ThermoFisher Scientific) for 5-10 min, followed by three 1xPBST washes prior to further blocking and staining.

2.2.20 Co-Immunoprecipitation using GFP-Trap® Magnetic beads

GFP-tagged protein was co-transfected with another protein of interest containing a V5 tag. WCEs were prepared using the same protocol as for western blotting (see section 2.2.18). Once the WCEs were collected, they were further sonicated for 10s using a Sanyo Soniprep 150 Sonicator and centrifuged at 20,000x g for 10 min at 4°C to collect the soluble fragment of the lysate. Incubation buffer (see section 2.1.2.10) was then added to each sample in equal volume as the lysis buffer used. Alongside, 25µl of Dynabeads™ MyOne™ Streptavidin C1 bead slurry (ThermoFisher Scientific) were added to 500µl of ice-cold dilution buffer to be equilibrated. The magnetic beads were then collected using a magnetic stand and the supernatant was discarded. This was repeated two more times to achieve equilibrated GFP-Trap® Magnetic beads, that were then added to the diluted lysates. Samples were incubated with beads by gentle tumbling end-over-end for 2 hr at 4° C. After incubation, the beads were magnetically separated on a stand until the supernatant was clear. At this point, the supernatant was discarded and beads were washed three times with ice-cold wash buffer. The beads were then resuspended in 80µl of 2x SDS-sample buffer and boiled for 10 min at 95° C to dissociate

immunocomplexes from the magnetic beads. SDS-PAGE and western blotting was performed using the resulting supernatant (see section 2.2.18.2). Potential protein interactions were identified by staining with anti-V5 antibody and pull-down verification was confirmed with anti-GFP staining.

2.2.21 BioID for identification of protein-protein interactions

The BioID assays were developed and optimised for both the original version of BioID, that uses an engineered BirA* biotin ligase from *E. coli* (Roux et al., 2012), and the updated version known as BioID2 that utilizes the biotin ligase activity from *A. aeolicus* (Kim et al., 2016). KLHL7 was cloned into a BioID plasmid using In-Fusion cloning (see section 2.2.11.2), and MAL was cloned into a BioID2 plasmid that is compatible with Gateway cloning (see section 2.2.11.1). Maps of each expression plasmid used can be found in Appendix I.

2.2.21.1 Biotinylation and Cell lysis

hTERT-RPE1 and MDCK cells were transfected with BioID and BioID2 plasmids, respectively, using the transfection protocol described in section 2.2.15. Cells were allowed to recover for 16-24 hr post-transfection and were then incubated in 1x Biotin Media (50µM biotin in culture media) for 24 hr for the BioID plasmid, and 4 hr for the BioID2 plasmid to induce biotinylation of interacting/ proximal proteins (see **figure 2.4**). Following biotin incubation, the cells were washed twice with 1x PBS and 300µl of Lysis buffer (see section 2.1.2.12) was added to each flask and incubated for 5 min on ice. Cells were then scraped and whole cell lysate collected. To each sample, 120µl of 20% Triton X-100 was added followed by two sessions of sonication (using a Sanyo Soniprep 150 Sonicator), each session consisting of pulses of 40µm for 10 seconds. 1ml of prechilled 50mM Tris-Cl, pH 7.4 was then added to each sample, followed by one more sonication session. Each sample was then spun down for 10 min at 16,500 x *g* at 4^{o} C. 30µl of each sample were taken forward for SDS-PAGE electrophoresis and western blotting for validation of biotinylation. Membranes were blocked with BSA blocking buffer (see section 2.1.2.14) for 30 min and then agitated with streptavidin-HRP (Vector Laboratories) using a dilution of 1:1000 for 45 min. Membranes were then washed twice with 1xPBST and then agitated again in NDS blocking buffer (see section 2.1.2.15) for 5 min to reduce background signal. Visualisation of membranes was performed using standard procedures described in (2.2.18.3). For staining of membranes with additional antibodies for specific tags, membranes had to be agitated in quenching solution (see section 2.1.2.16) for 20 min.

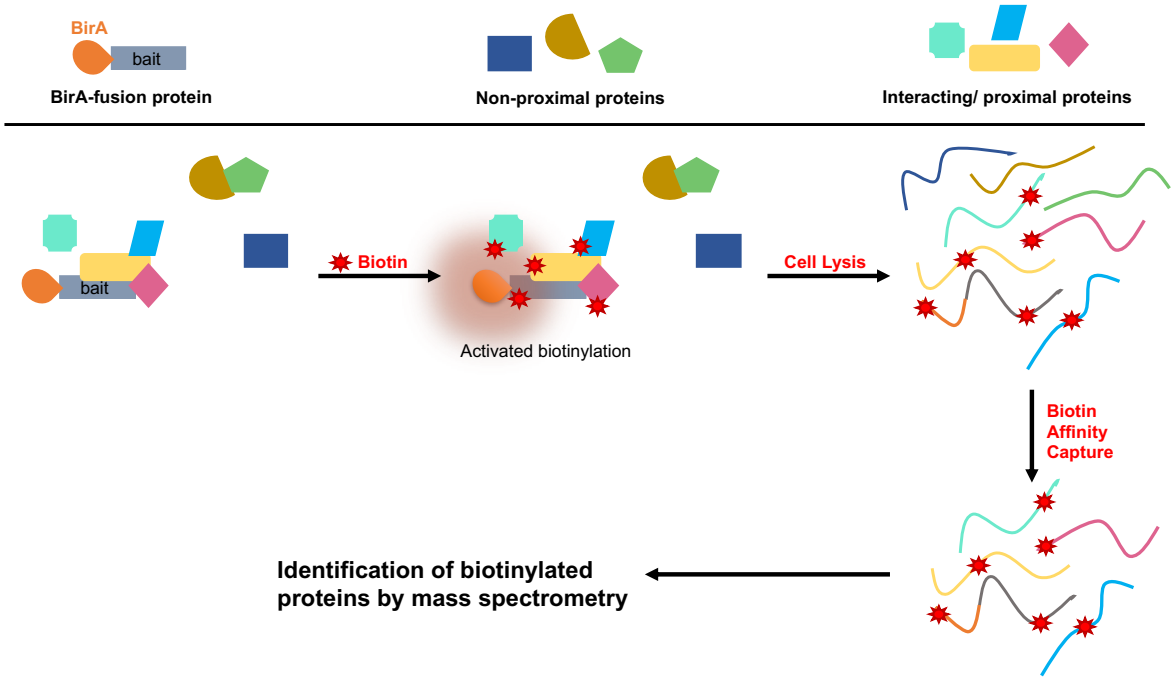


Figure 2-7 Outline of the BioID procedure

The diagram outlines the mechanism of action of the BioID technology, indicating the tagging of proximal proteins by inducing biotinylation through the BirA-fusion protein. Proximal or interacting proteins will get biotinylated once incubated with biotin, therefore allow for affinity capture using streptavidin or streptactin beads and further identification using mass spectrometry.

Prior to performing the immunoprecipitation (IP), 50µl of Strep-Tactin-Beads per sample were washed with 600µl of 1xTBS. Bead solutions were centrifuged at 5,000 xg for 30s at 4° C, and the supernatant was discarded. Beads were then mixed with 500µl of Lysis Buffer (see section 2.1.2.12), and centrifuged at 5,000 xg for 30s at 4° C to discard the supernatant. Beads were then washed twice with 500µl of Wash buffer (see section 2.1.2.13), discarding the supernatant. After the last wash, the beads were mixed with 2x bead volume Wash buffer and transferred to a clean tube. Lysates were mixed with bead solution and incubated for 2 hr at 4° C on an end-over-end shaker. Samples were then centrifuged at 5,000 xg for 1 min at 4° C and supernatant was discarded. At this stage, the beads were washed four times with 500µl 1x TBS before proceeding to on-bead trypsin digest. Standard operating procedures for on-bead digest and sample preparation for mass spectrometry analysis was performed by collaborators in Dr Karsten Boldt's laboratory, University of Tübingen. C-MS/MS analysis was performed on Ultimate3000 nanoRSLC systems (Thermo Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) by a nano spray ion source.

2.2.22 Seahorse Metabolism Assays

2.2.22.1 Mito Stress Test

The Seahorse Mito Stress Test (Agilent) is an assay that measures key parameters of mitochondrial function. This is achieved by the measurement of oxygen consumption rate (OCR) of the cells before and after sequential injection of toxins that block different processes of oxidative phosphorylation. This allows for identification of important measures of mitochondrial function such as ATP production, basal and maximal respiration, spare respiratory capacity, proton leak and non-mitochondrial respiration. This is represented in **figure 2.8**.

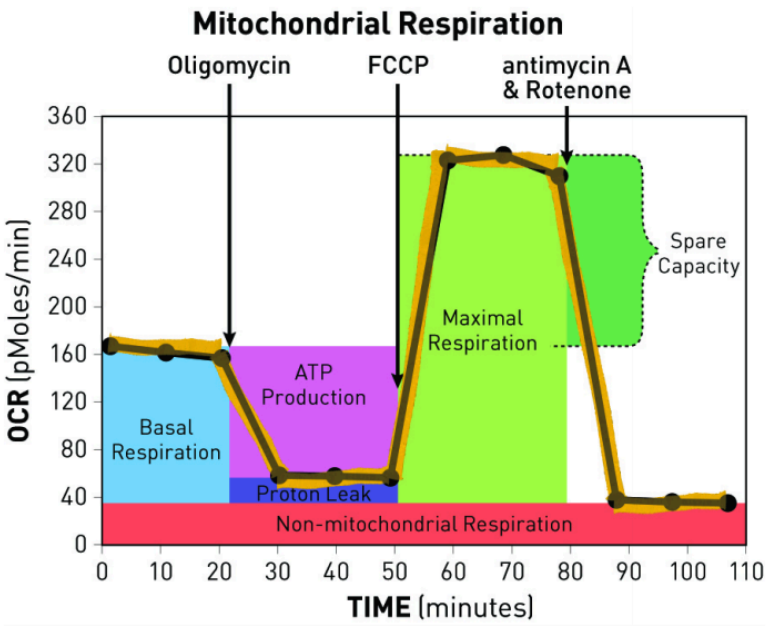


Figure 2-8 Mito Stress Test trace

A representation of the Mito Stress Test trace, indicating the three injection points and the measurements obtained from each one. Adapted from Agilent Seahorse.

One day prior to the assay, primary patient and control cells were seeded in the Seahorse XF Cell Culture Microplate at a seeding density of 15,000 cells per well. In addition, the Seahorse Sensor Cartridge was hydrated using the Seahorse XF Calibrant and incubated at 37° C in a non-CO_2 incubator overnight. On the day of the assay, the normal culture media was changed to the Seahorse Mito Stress Assay Medium (see section 2.1.2.19) and incubated for 1 hr at 37° C in a non-CO_2 incubator. During incubation, the toxins were loaded on the sensor cartridge plate at the following concentrations; Oligomycin at 1.0µM, FCCP at 2.0µM and Rotenone/Antimycin A at 0.5µM, in ports A, B and C respectively. Oligomycin blocks the ATP synthase (complex V) leading to

minimum respiration, FCCP is an uncoupling agent that activates proton conductance and allows maximal respiration. Rotenone blocks NADH Dehydrogenase (complex I) and Antimycin A blocks Cytochrome C reductase (complex III), leading to complete block of the electron transport chain and oxidative phosphorylation.

At the end of the assay, the differences between cell numbers in each well were normalised using a Crystal Violet Assay (see section 2.2.23). Data analysis was performed using Wave Software, Microsoft Excel and Prism for calculations and statistical tests.

2.2.22.2 Glycolysis Stress Test

The Seahorse Glycolysis Stress Test (Agilent Tehnologies) is an assay that measures the basic glycolytic function of the cells by correlating extracellular acidification rate (ECAR) measurements before and after specific inhibitor treatments. These sequential injections of glucose, oligomycin and 2-Deoxy-d-glucose (2-DG) affect different parameters of the glycolysis pathway allowing for measurements of the following functions: Glycolysis, Glycolytic Reserve, Glycolytic Capacity and Non-Glycolytic Acidification. This is summarised by the Glycolysis Stress Test trace in **figure 2.9**.

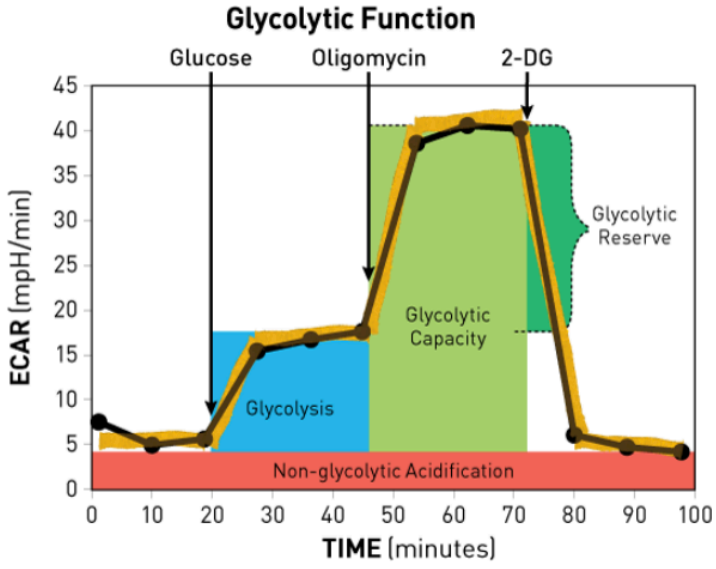


Figure 2-9 Glycolysis Stress Test trace

A representation of the Glycolysis Stress Test trace, indicating the three injection points and the measurements obtained from each one. Adapted from Agilent Seahorse.

One day prior to the assay, primary patient and control cells were seeded and the Seahorse Sensor Cartridge was hydrated as stated in section 2.2.21.1. On the day of the assay, the normal culture media was changed to the Seahorse Glycolysis Stress Assay Medium (see section 2.1.2.20) and incubated for 1 hr at 37° C in a non-CO_2 incubator. During incubation, the toxins were loaded on the sensor cartridge plate at the following concentrations: glucose at 10.0mM, oligomycin at 1.0µM and 2-DG at 50.0mM, in ports A, B and C respectively. Glucose was added to the cells initially to promote glycolysis. Oligomycin blocks the ATP synthase (complex V) leading to minimum respiration, thereby increasing dependence on glycolysis for energy production. 2-DG is a competitive inhibitor of glucose that therefore completely blocks glycolysis.

At the end of the assay, the differences between cell numbers in each well were normalised using a Crystal Violet Assay (see section 2.2.23). Data analysis was performed using Wave Software, Microsoft Excel and Prism for calculations and statistical tests.

2.2.23 Crystal Violet Assay

This assay was performed for normalisation purposes for the Seahorse experiments. Culture media was aspirated off and the cells were washed twice in gentle stream of tap water. Excess water was removed by gently tapping the plate on tissue. 50µl of 0.5% [w/v] crystal violet solution (see section 2.1.2.21) was added to each well and incubated on a bench rocker at 20 oscillations per minute for 20 min. After incubation, the plate was washed five times in a stream of tap water and let to air-dry for 16 hr at room temperature. 200µl of methanol was then added to each well and incubated on a bench rocker at 20 oscillations per minute for 20 min. This solubilised the crystals into a purple solution allowing the optical density of each well to be measured using a plate reader at 570nm absorbance. The absorbance values were used to normalise the seahorse measurements on the Wave software.

2.2.24 MTT Assay

Primary fibroblasts ($1x10^4$) were plated in a 96-well plate and grown to 90% confluency. MTT reagent was freshly prepared at a working concentration of 1mg/ml using 1x PBS. 100µl was added to each well, and the plate was wrapped in foil and incubated at 37°C for 3-4 hours. After incubation, the MTT reagent was removed and replaced by 100µl of propan-1-ol, which was mixed by pipetting to solubilise the violet crystals. The quantification of the colour change was done by an Opsys MR plate reader (Dynex Technologies) at 570nm.

2.2.25 Statistical Analysis

Statistical analysis was performed using functions in Microsoft Excel and Prism 7 (https://www.graphpad.com/). In large samples, normal distribution was assumed based on the Assumption of Normality, where n>30. Where necessary, normal distribution was tested by plotting a normal Q-Q plot. Results were represented in the form of bar charts or box plots where variability and clustering of data had to be illustrated. Error bars were also plotted on the graphs, representing the Standard Error of the Mean. Statistical significance between the means was tested using the Independent Sample T test, and significance was described as $p \leq 0.05$. All experiments of this study, unless stated otherwise, were performed in three biological replicates, where cell samples and reagents were different between each replicate.

2.2.26 Mass Spectroscopy Analysis

Mass spectroscopy analysis was performed by Dr Karsten Boldt; a collaborator on the BioID projects. Specifically, MS/MS data were analyzed using the MaxQuant software (version 1.6.1.0) (Cox and Mann, 2008, Cox et al., 2009). As a digesting enzyme, Trypsin/P was selected with maximal 2 missed cleavages. The data were analyzed by label-free quantification with the minimum ratio count of 3. The first search peptide tolerance was set to 20, the main search peptide tolerance to 4.5 ppm and the re-quantify option was selected. For peptide and protein identification the human subset of the SwissProt database (release 2014_11) was used and contaminants were detected using the MaxQuant contaminant search. A minimum peptide number of 2 and a minimum length of 7 amino acids was tolerated. Unique and razor peptides were used for quantification. The match between run option was enabled with a match time window of 0.7 min and an alignment time window of 20 min. The statistical analysis including ratio, t-test and significance A calculation was done using the Perseus software (version 1.5.5.3) (Tyanova et al., 2016). Significance A is a Q-function that detects outliers from a normal distribution. The background binders are the normal distribution and the interactors are the outliers.

Chapter 3

Identification and functional characterisation of variants in genes already associated with neurodevelopmental and neuromuscular disorders

3.1 Introduction

This chapter outlines the genetic investigation of three families that resulted in the identification of causative mutations in known disease genes. Despite the fact that no variants were identified in any novel disease genes, this part of the study describes how whole exome sequencing done on research basis can also be utilised in the clinical setting, offering a molecular diagnosis to the patiens. The affected individuals of the first family (ND1) presented with a very severe form of arthrogryposis, with one of the three affected individuals dying in the neonatal period. DNA samples from all members of the family were processed for WES. In-house bioinformatics analysis led to the identification of a nonsense variant in *CHRNG*, a known arthrogryposis gene (Morgan et al., 2006). The findings of family ND1 were not taken forward for functional studies, as mutations in this gene are well-established as a cause of arthrogryposis. The second family (ND2) presented with muscular dystrophy similar to Limb Girdle Muscular Dystrophy. WES for this family was performed by collaborators in Newcastle University and revealed a pathogenic nonsense variant in *MICU1*, a novel disease gene that was identified in the same year by Dr Clare Logan (Logan et al., 2014). Variants in *MICU1* were novel at the time of this finding, so functional studies were designed in order to elucidate the effect of *MICU1* null mutations on mitochondrial function and the possible pathomechanism of this condition. The third family (ND3) presented with a severe form of developmental delay, and three affected individuals as well as an unaffected sibling were taken forward for WES. In-house bioinformatics analysis revealed a frameshift variant in *HERC2*, a gene that is mutated as a cause of Angelman-like syndrome (Harlalka et al., 2013). Functional studies on the missense variant in *HERC2* are also be described in this chapter, in order to confirm the pathogenic impact of this variant and to extend the phenotype-genotype correlations of *HERC2* mutations. This work highlights the variable phenotypic spectrum of Angelman-like syndrome. Finally, all variants identified in known disease genes have been reported back to the referring clinician in order to inform the families and to offer potential diagnostic testing.

3.2 WES identifies a nonsense variant in *CHRNG*, a known arthrogryposis gene

3.2.1 Clinical Phenotype

Family ND1 is a consanguineous South Asian family with three affected individuals presenting with arthrogryposis at birth. One of the affected children died in the neonatal period from respiratory failure. The other two children have significant arthrogryposis with multiple contractures of both upper and lower limbs; at elbows, wrists, proximal and distal interphalangeal joints of hands and feet. The family pedigree is outlined in **figure 3-1**.

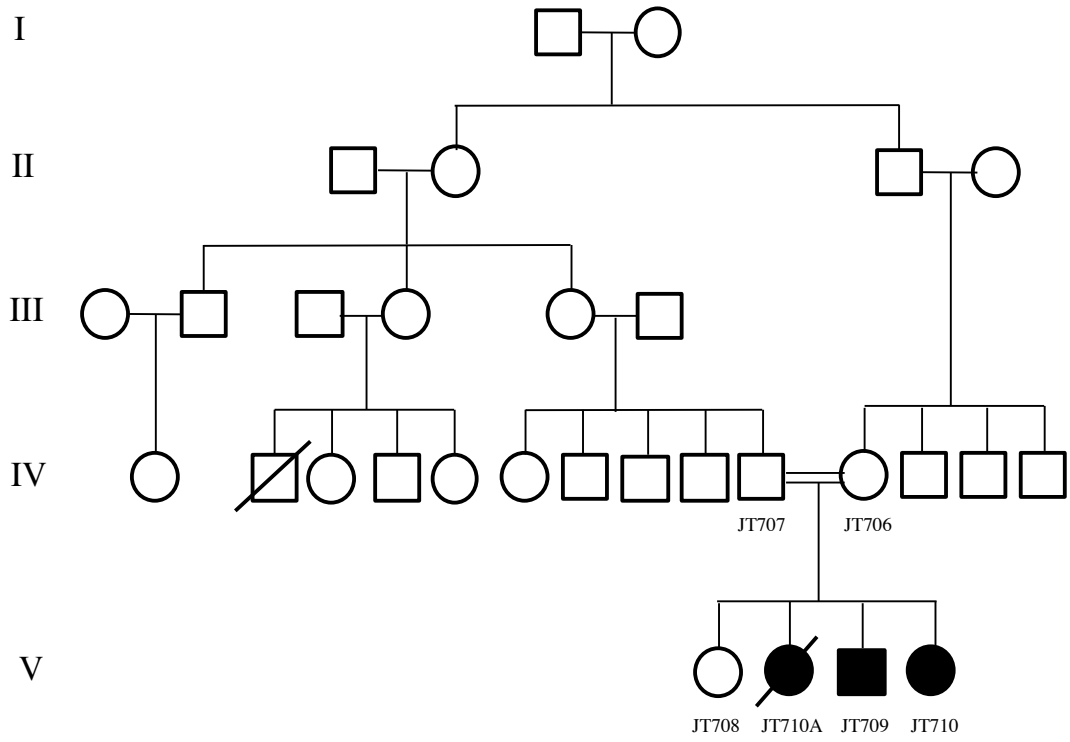


Figure 3-1 Pedigree of family ND1

Pedigree outlining five generations of the family and the consanguineous union. DNA samples indicated by anonymized codes have been used for genetic investigations using WES.

3.2.2 Autozygosity mapping

In the presence of consanguinity autozygosity mapping was performed using WES data. Precisely, the vcf file prior to SNP filtering (.raw.vcf) from all members of the family that have been taken forward for WES were used (see section 2.2.10.6). Homozygous regions shared by the three affected individuals, but absent from the parents and the unaffected sibling, are summarised in **Table 3-1**.

Chromosome	Start	End	Length
2	228560800	236433161	7872361
16	103423	732287	628864
2	219920037	220284779	364742
11	56143198	56143730	532

Table 3-1 Common regions of homozygosity amongst affected individuals.

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

3.2.3 Whole exome sequencing

DNA samples from all the individuals with a JT identifier were taken forward for WES library preparation using the Illumina QXT protocol (section 2.2.10). During library preparation, Bioanalyzer traces were used to assess the size and quantity of the DNA fragments at post-shearing and post-hybridization stages (Appendix D). The samples were pooled, aiming for a total of ten samples per lane, and sequencing was performed on the Illumina HiSeq 3000 platform.

Prior to data analysis, the fastq files and subsequent bam files generated were also assessed for the quality of sequencing. An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3), assuming a recessive mode of inheritance that allowed variants to be prioritized in previously identified homozygous regions. In addition, segregation analysis enabled the filtering out of variants that were not compatible with Mendelian segregation of an autosomal recessive condition. **Table 3-2** summarises the only variant identified after filtering with a CADD score >15. **Figure 3-2** illustrates the presence of the *CHRNG* variant in the largest homozygous region shared by the affected individuals.

Variant Information			
Gene	**Location**	**Variant**	**Protein change**
CHRNG	2:233404782	c.136C>T	p.R46*
Pathogenicity prediction			
Condel	**Polyphen2**	**SIFT**	**CADD score**
___	___	___	35
Frequency and function			
MAF gnomAD	**Protein Function**	**OMIM**	**In Homozygous region**
0.000054	Gamma subunit of acetylcholine receptor	Escobar Syndrome	Yes

Table 3-2 Homozygous variant identified in family ND1.
Homozygous variant c.136C>T in *CHRNG* (NM_005199.5) was annotated based on the human genome assembly GRCh37/hg19.

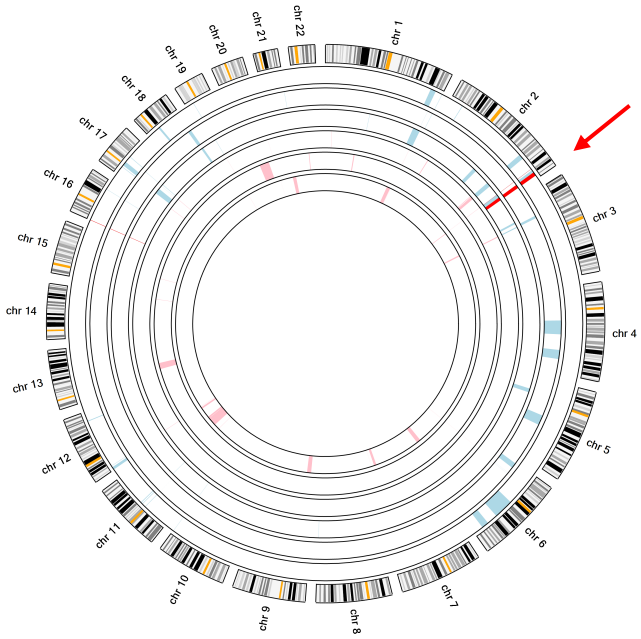


Figure 3-2 Ideogram illustrating the homozygous region shared by the affected individuals in family ND3.
The ideograms in light pink are from the unaffected individuals, whereas the three outer ideograms in light blue are from the affected children. The regions highlighted in red represent the shared homozygous regions. The homozygous *CHRNG* variant lies within chromosome 2 (2q36.3-q37.3), indicated by the red arrow. The human genome assembly GRCh37/hg19 was used.

3.2.4 Variant confirmation using IGV and allele depth

Since all available members of the family were sequenced by WES, the variant was confirmed by visualizing the bam file on Integrative Genomics Viewer (IGV). This allows for visualization of the region where the variant is located, as well as giving an indication

of the read depth over this region (major allele = C, variant allele = T). IGV images are shown in **figure 3-3** following the structure of the family's pedigree. The allele depth for this variant in all sequencing individuals is summarised in **table 3-4**, showing a segregation pattern compatible with autosomal recessive inheritance with all affected individuals homozygous for the variant. Following this conformation, this research finding in *CHRNG* was reported back to the referring clinician. Sanger sequencing for diagnostic confirmation was performed by the Yorkshire Regional Genetics Service diagnostics lab, prior to reporting the finding to the patients.

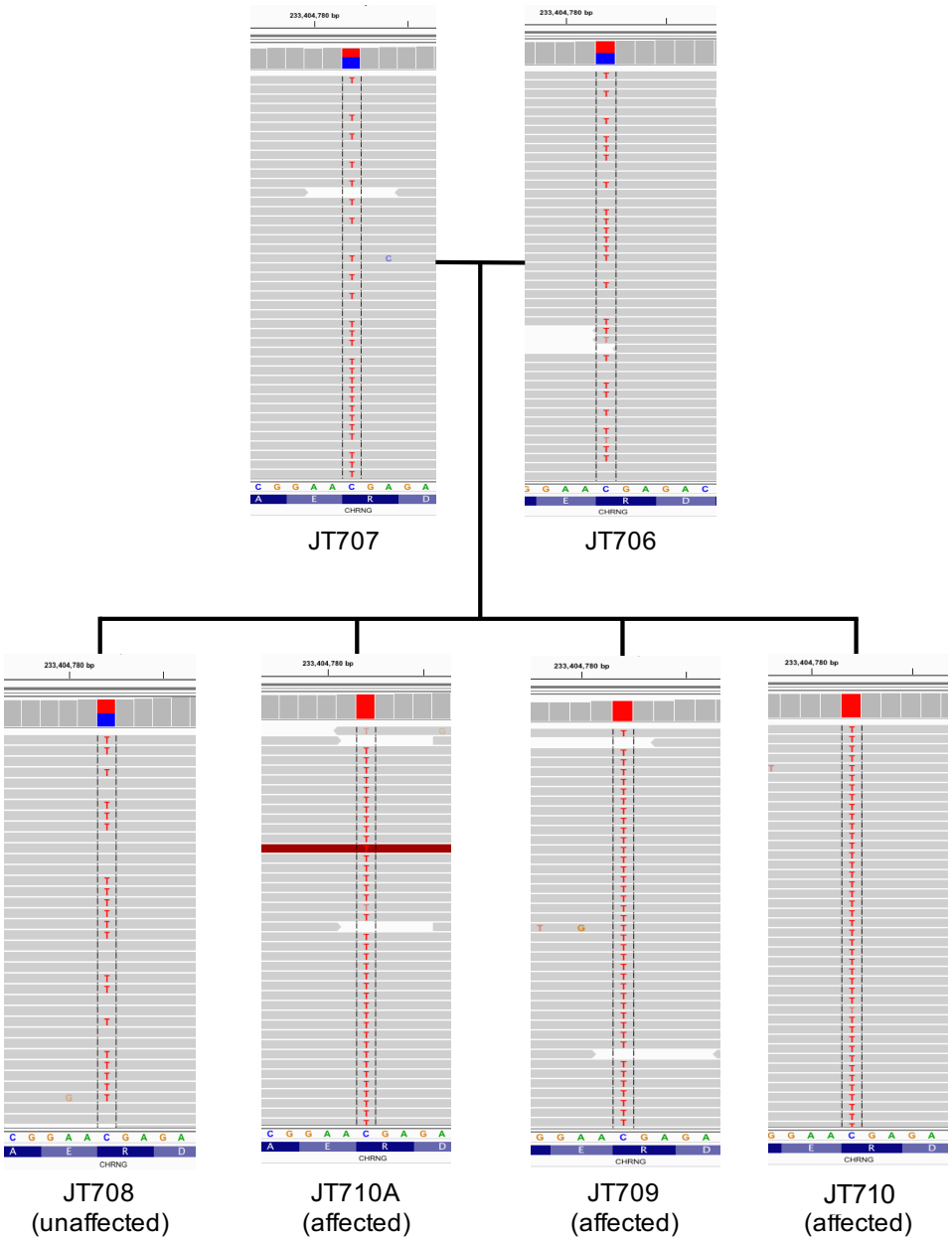


Figure 3-3 IGV images confirming segregation of the *CHRNG* variant in family ND1.

IGV screenshots illustrate the presence of the variant T allele (red) in the affected individuals with very high coverage, and the presence of the variant in approximately half the reads in the parents and unaffected sibling, consistent with autosomal recessive segregation.

	JT706	JT707	JT708	JT709	JT710	JT710A
Segregation	C/T	C/T	C/T	T/T	T/T	T/T
Allele Depth	C=89/T=73	C=126/T=105	C=90/T=91	C=0/T=157	C=0/T=133	C=0/T=221
Phred Genotype Confidence	99	99	99	99	99	99

Table 3-3 Summary of allele depth per WES sample and outlined segregation pattern.

The allele depth for every individual is very high, with approximately equal reads for each nucleotide in the heterozygous individuals and no reads for the cytosine nucleotide in the affected individuals, confirming the segregation pattern visualized by IGV.

3.2.5 CHRNG and Arthrogryposis

The *CHRNG* gene encodes for the gamma subunit of the acetylcholine receptor and it is known to cause autosomal recessive, lethal and non-lethal Escobar syndrome (MIM number: 265000) (Morgan et al., 2006). The clinical synopsis of this syndrome matches the phenotype of the family, with the key feature being the incidence of arthrogryposis.

The WES data for family ND1 were of limited scientific interest, as mutations in the gene have already been described as a cause of this condition and this specific nonsense mutation has already been reported in the literature (Morgan et al., 2006). However, the outcome has been reported to the referring clinician and is likely to benefit the family in the future through accurate molecular diagnostic testing and genetic counselling.

3.3 Mutation in *MICU1* as a cause of muscular dystrophy

3.3.1 Clinical Phenotype

Family ND2 is a consanguineous family of Saudi Arabian origin with two affected boys who presented with elevated creatine kinase levels and mild proximal weakness. No learning difficulties or abnormal movement was reported by the clinician responsible for this family's care, but muscle biopsy revealed minimal myopathic changes. The family pedigree is shown in **figure 3-4**.

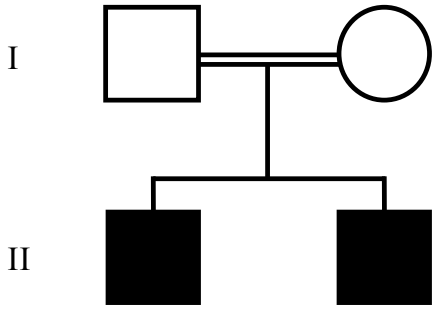


Figure 3-4 Pedigree of family ND2.

Pedigree outlining two generations of the family and the consanguineous union.

3.3.2 Whole Exome Sequencing

WES was performed by our collaborators at the Wellcome Centre for Mitochondrial Research at Newcastle University. DNA samples from the two affected individuals were prepared using the Illumina QXT kit and the Agilent v5 library. Samples were pooled on a lane and run on the Illumina HiSeq2500 platform. Data analysis was carried out using their in-house bioinformatics pipeline, revealing a homozygous nonsense mutation c.547C>T (p.Q183*) in exon 6 of the *MICU1* gene (NM_001195518.2). Both parents were screened for this variant and were heterozygous, confirming the recessive mode of inheritance.

3.3.3 The Mitochondrial Calcium Uptake 1 (*MICU1*) gene

The Mitochondrial Calcium Uptake 1 (*MICU1*) gene encodes for a protein that is a key regulator of the Mitochondrial Calcium Uniporter (MCU) (Marchi and Pinton, 2014). MCU is a calcium channel localised at the mitochondrial inner membrane, which is regulated by its various subunits including MICU1, MICU2 and EMRE (Sancak et al., 2013). MICU1 is an important regulator of the opening of the channel under increased Ca^{2+} levels, whereas MICU2 acts antagonistically to MICU1 and inhibits MCU activity under low Ca^{2+} levels (Patron et al., 2014). Both MICU1 and MICU2 interact with the uniporter by forming

heterodimers, which comprise of EF-hand domains that can sense calcium levels, allowing instant responses to calcium transients.

The MICU1 interaction with the MCU, via EMRE, is crucial in maintaining basal calcium conditions. A disruption of the gating of the uniporter can lead to pathological Ca^{2+} overload, which can thereby lead to oxidative stress and perhaps apoptosis. Such disruptions of MCU gating have been reported in cases where the *MICU1* gene was mutated and could no longer bind and interact with the uniporter. Splicing mutations, c.1078-1G>C and c.741+1G>A (Logan et al., 2014), and a 2755 bp homozygous deletion in exon 1 (Lewis-Smith et al., 2016), are causative for a myopathy phenotype with extrapyramidal signs.

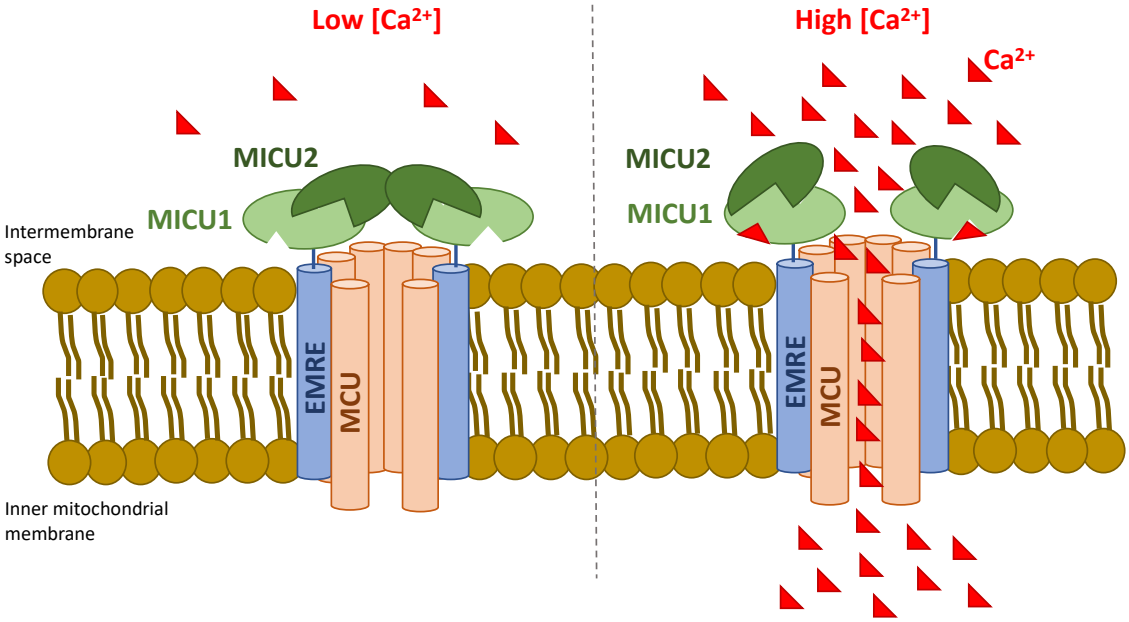


Figure 3-5 The MCU complex under low or high calcium concentrations.

A representation of the MCU complex showing all components of the complex, their interactions, and how calcium levels can regulate the MICU1 and MICU2 interactions and thereby the gating of the channel (Kamer and Mootha, 2015).

3.3.4 Functional characterisation using patient fibroblasts

3.3.4.1 Confirming the c.547C>T variant in patient fibroblasts

A skin biopsy was obtained from one of the affected children under informed consent for research as described in section 2.2.2 and 2.2.4. Primary fibroblast cultures were grown from the skin biopsy at the Yorkshire Regional Genetic Service cytogenetics lab, St James' University Hospital. Genomic DNA extraction and Sanger sequencing (section 2.2.5.3 and 2.2.9) were used to confirm the nonsense mutation identified by WES (see **figure 3-6**).

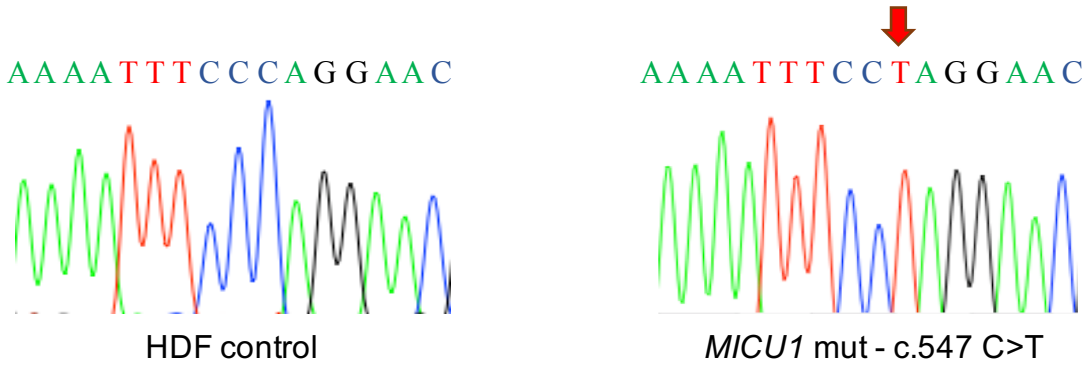


Figure 3-6 The electropherogram confirming the *MICU1* c.547C>T, p.Q183* nonsense mutation.

Genomic DNA was extracted from affected patient primary fibroblasts and normal control human primary fibroblasts to confirm the mutation.

3.3.4.2 Mitochondrial fragmentation revealed by IF microscopy

Initially, an IF experiment was set up in order to assess how the mutation affected the mitochondrial network. To achieve this, wildtype and mutant primary human dermal fibroblasts were utilised. For the purposes of the IF staining, an antibody against MTCO2 was used as a marker for mitochondria in addition to either MICU1 or MCU to visualize the localization of the uniporter complex. The antibody staining for MICU1 and MCU did not reveal any obvious differences between wildtype and mutant (see **figure 3-7**). However, mitochondrial staining for MTCO2 revealed a disrupted mitochondrial network in the affected patient cells compared to the normal controls. The mitochondrial fragmentation is an observation that was expected because the MICU1 mutation is likely to directly affect the calcium homeostasis of the mitochondria by uncoupling the calcium uniporter (Logan et al., 2014). This observation could also indicate that the fragmentation might be due to disrupted mitochondrial trafficking via the microtubules, which is also calcium-dependent. Disrupted calcium buffering around the mitochondria might have an effect on the calcium-dependent binding of the Miro/Milton complex to the kinesin motor protein of the microtubules (Wang and Schwarz, 2009). To investigate this further, we used live cell imaging to visualize mitochondrial trafficking over time (section 3.3.4.3).

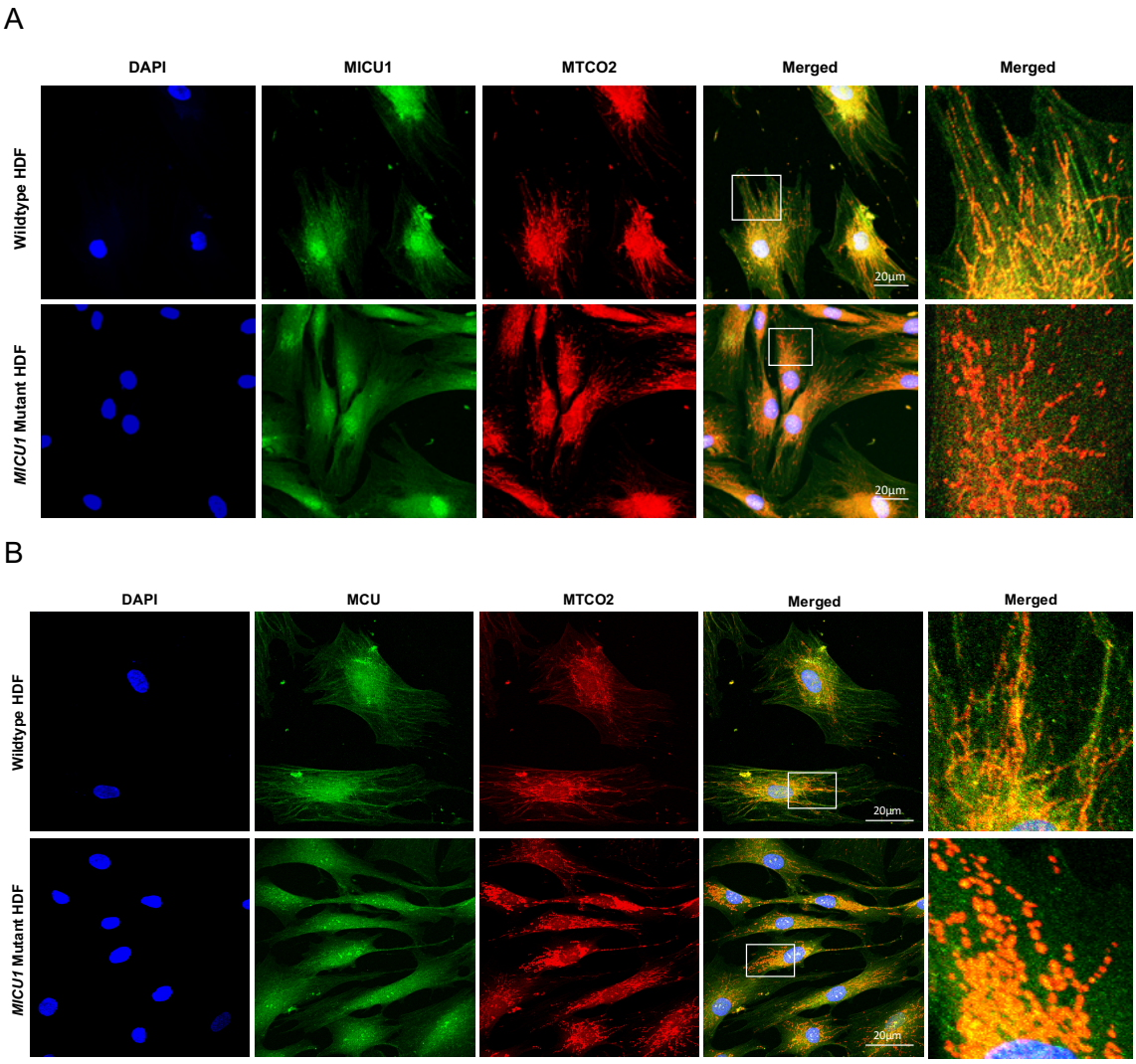


Figure 3-7 IF confocal microscopy revealing mitochondrial fragmentation.

A: Immunofluorescence confocal microscopy of MICU1 (green) and the mitochondrial marker MTCO2 (red) in *MICU1* c.547 C>T (p. Q183X) primary fibroblasts compared to wildtype normal control fibroblasts (HDF). B: Immunofluorescence confocal microscopy of MCU (green) and the mitochondrial marker MTCO2 (red) in *MICU1* c.547 C>T (p. Q183X) primary fibroblasts compared to wildtype control fibroblasts (HDF). Scale bars = 20 μm.

3.3.4.3 Live cell imaging investigating the mitochondrial distribution

Live cell imaging was used in an attempt to investigate the hypothesis of mitochondrial fragmentation arising from disrupted mitochondrial trafficking via the microtubules. Live cell mitochondrial staining was accomplished using MitoTracker® Green FM (Invitrogen), which is a green fluorescent stain that stains mitochondria in live cells, regardless of the mitochondrial membrane potential. MitoTracker® Green FM is used in this instance as a measure of mitochondrial distribution in order to test whether or not there is any difference in mitochondrial numbers, morphology and overall function between the control and mutant *MICU1* p.Q183* fibroblasts. **Figure 3-8** summarises a few of the replicates acquired from live cell imaging.

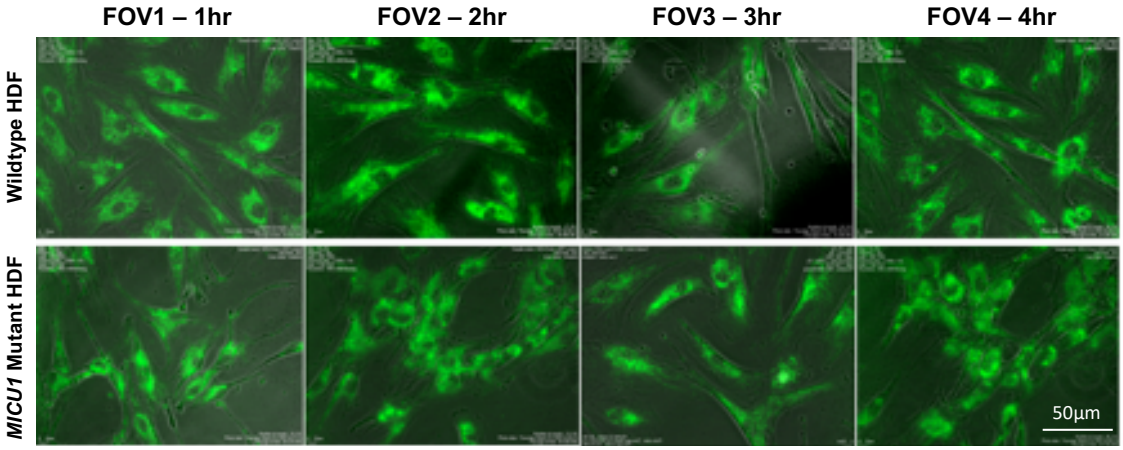


Figure 3-8 Live cell imaging investigating mitochondrial distribution

Summary of images obtained at specific time intervals, showing the MitoTracker® Green FM (Invitrogen) staining and phase contrast to identify the cell boundaries.

Individual tiff files where collected and analysis was performed for each field of view at 20 minutes intervals. Analysis was carried out on FIJI by using the protocol for "Measuring Cell Fluorescence" using the following formula:

Corrected total cell fluorescence (CTCF) = *Integrated Density – (Area of selected cell X Mean fluorescence of background readings)*

For each field of view, every cell was selected using the phase contrast view to visualize the cell boundaries and marked as "Region of Interest" (ROI). For each ROI the Integrated Density of the green fluorescence was measured. Subsequently, three background readings were collected around each ROI and were used to subtract any background fluorescence. Three biological replicates, of three technical replicates each, were imaged (see section 2.2.18.1) and analysed using the above protocol. The results of the live cell imaging experiment are summarised in **figure 3-9**, comparing the CTCF value between wildtype and mutant *MICU1* p.Q183* human dermal fibroblasts.

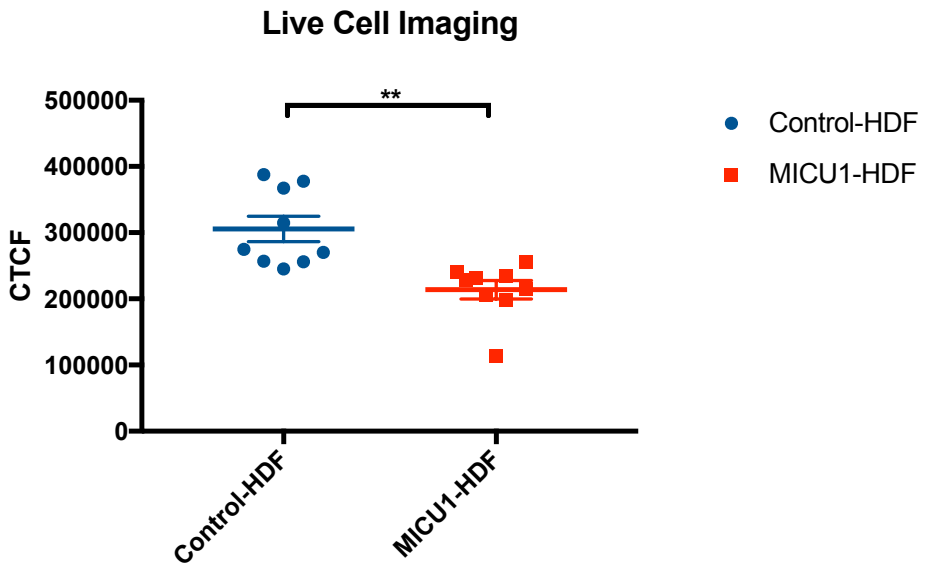


Figure 3-9 Graphical representation for the averages of the Corrected Total Cell Fluorescence (CTCF) between wildtype and *MICU1* p.Q183* mutant fibroblasts.

A representation of the average values of CTCF from nine fields of view taken from each biological replicate, indicating a reduction in CTCF in the mutant fibroblasts compared to the control.

The experiment revealed reduced CTCF in the mutant fibroblasts compared to the control. Statistical analysis using a two-tailed t-test indicated a significant reduction in CTCF (p-value: 0.0013). However, reduction of CTCF correlates with reduced mitochondrial levels. This could be interpreted as a disruption in mitochondrial trafficking on the microtubules (Wang and Schwarz, 2009), with most mitochondrial clustering around the nucleus in the mutant fibroblasts, thereby giving a reduced CTCF value. In addition, a disturbance of the balance between fission and fusion (Scott, 2016) could also lead to a more fragmented mitochondrial network. In particular, if fission is occurring more frequently than fusion, then the overall mitochondrial levels may also seem reduced as the network fragments. Lastly, the significant disruption of calcium homeostasis may lead to stressed mitochondria and potential initiation of apoptotic events (Elmore, 2007).

3.3.4.4 MTT Assay

Another key aspect that was investigated was the effect of the *MICU1* c.547 C>T p.Q183* mutation on cell metabolic activity. As the mutation is implicated in uncoupling the mitochondrial calcium uniporter, it was a reasonable assumption that it would have a downstream impact on calcium buffering, ionic homeostasis and the proton-motive force that drives oxidative phosphorylation. The first attempt to obtain more insights about metabolism was by using the MTT assay, which is a colorimetric test for measuring

cell metabolic activity. The tetrazolium dye MTT 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide is reduced to formazan, an insoluble purple crystal, by dehydrogenase enzymes (Sylvester, 2011). This experiment was performed in a dose-response manner to eliminate any false positive results due to cell number differences.

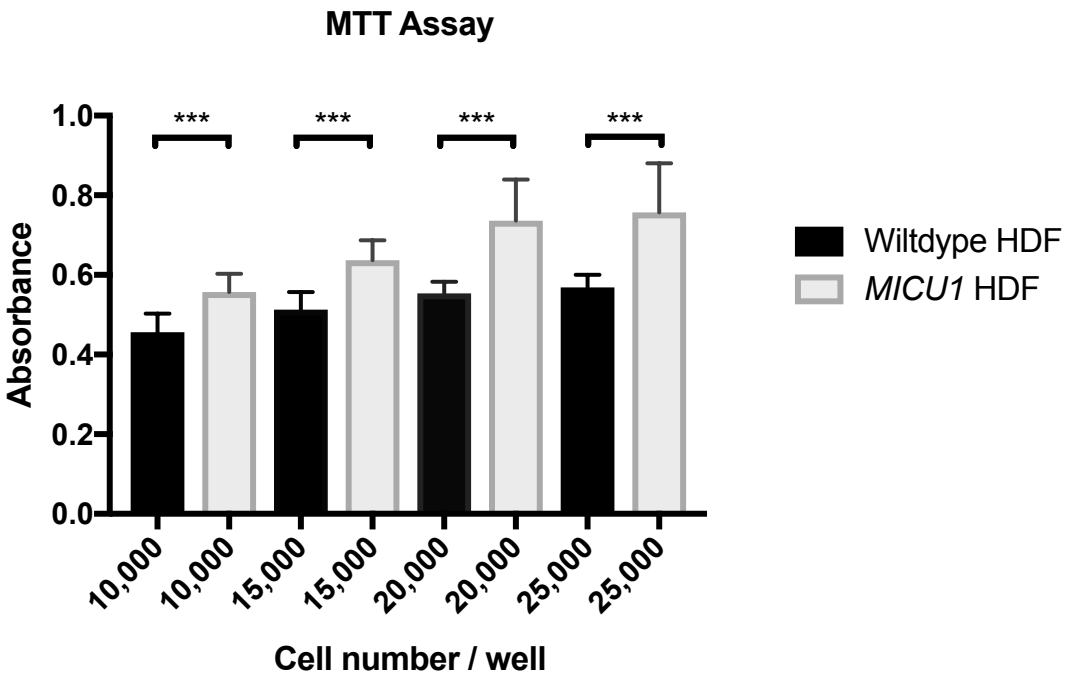


Figure 3-10 MTT assay in a dose-response experiment

A representation of a dose-response MTT experiment with cell numbers ranging from 10,000 to 25,000 cells per well. In all cases, the reduction of MTT to formazan was significantly increased in the mutant fibroblasts compared to the control.

The statistical analysis of the results, using a two-tailed t-test, indicated that the metabolic activity of the mutant cells is significantly higher compared to the controls. This assay is based on the activity of dehydrogenase enzymes, which are involved in energy production via the Kreb's cycle and electron transport chain in the mitochondria. Increased activity could be explained by the disrupted calcium buffering caused by the *MICU1* mutation, because calcium is a key metabolic regulator of the citric acid cycle with important roles in activating specific dehydrogenases to promote energy production (Denton, 2009). In particular, calcium can activate pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and NAD^+-dependent isocitrate dehydrogenase, all key enzymes of the Kreb's cycle, thereby activating ATP production by the mitochondria (Traaseth et al., 2004). High calcium concentrations in the mitochondria, as a result of the *MICU1* mutation, can potentially result to overstimulated dehydrogenases activity and thereby increased reduction of MTT in this experiment.

These descriptive observations are interesting and suggest possible new avenues to investigate the pathomechanism of *MICU1* mutations as a cause of muscular dystrophy. However, at the time of study, *MICU1* variants were no longer novel, with a number already published in the literature. Furthermore, Lee et al had solved part of the structure of the calcium uniporter (Lee et al., 2015) and, particularly for *MICU1*, Liu et al had rescued the phenotype seen in *MICU1*- knockout mice by introducing a heterozygous knockout mutation in *EMRE* (Liu et al., 2016) which significantly rescued the gating of MCU. The scientific novelty of further functional characterization of MICU1 was therefore limited, and we prioritized the investigations presented in chapters 5 and 6.

3.4 Expanding the genotype-phenotype correlation in developmental delay disorders and functional characterization of a novel *HERC2* frameshift variant

3.4.1 Clinical Phenotype

Family ND3 is a consanguineous South Asian family with three affected individuals presenting severe developmental delay. Two of the affected children died in very early childhood. All of the affected children shared the same clinical symptoms, including movement difficulties and profound developmental delay. A more detailed clinical picture was obtained from the affected boy, from whom we also obtain a skin biopsy. The affected child has severe hypotonia, with some involuntary movements, choreoathetoid seizures and no head control. He is also blind, even though his eyes are structurally normal. The family pedigree is outlined in **figure 3-11**.

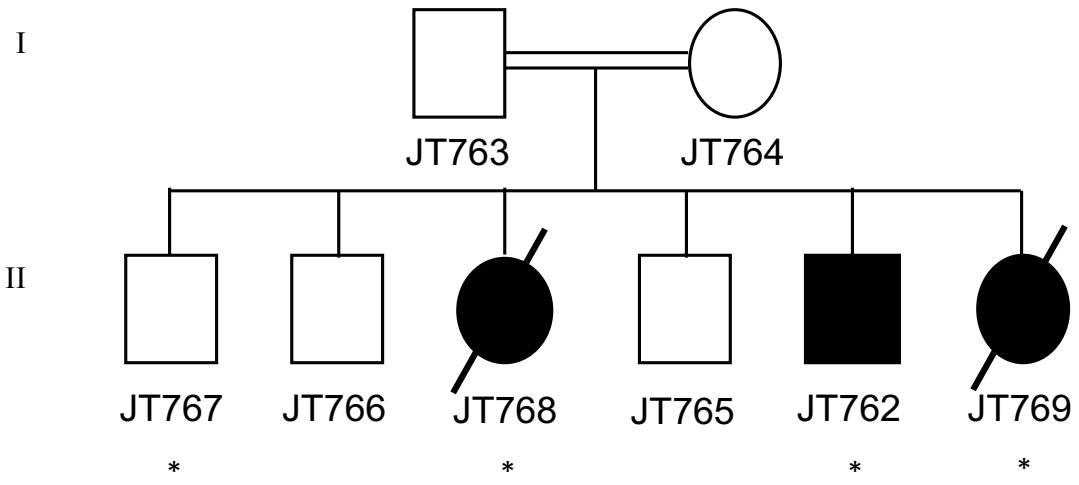


Figure 3-11 Pedigree of family ND3

Pedigree outlining two generations of the family and the consanguineous union. DNA samples of individuals marked with star have been used for genetic investigations using WES.

3.4.2 Autozygosity Mapping

In the presence of consanguinity autozygosity mapping was performed using WES data. Precisely, the vcf file prior to SNP filtering (.raw.vcf) from the four members of the family that have been taken forward for WES were used (see section 2.2.10.6). Homozygous regions shared by the three affected individuals, but absent from the parents and the unaffected sibling, are summarised in **Table 3-4**.

Chromosome	Start	End	Length
15	28200408	34640378	6439970
12	9450357	10149851	699494
20	60572663	60966318	393655
6	32487209	32546838	59629
6	29856633	29858530	1897

Table 3-4 Common regions of homozygosity amongst affected individuals.

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

3.4.3 Whole Exome Sequencing

DNA samples from the three affected individuals and the unaffected sibling were taken forward for WES library preparation using the Illumina QXT protocol (section 2.2.10). During library preparation, Bioanalyzer traces were used to assess the size and quantity of the DNA fragments at post-shearing and post-hybridization stages (Appendix D). The samples were pooled together aiming for a total of ten samples per lane and sequencing was performed on the Illumina HiSeq 3000 platform.

Prior to data analysis, the fastq files and subsequent bam files generated were also assessed for the quality of sequencing. An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance that allowed variants to be prioritized in previously identified homozygous regions. In addition, segregation analysis enabled the filtering out of variants that were not compatible with Mendelian segregation of an autosomal recessive condition. **Table 3-5** summarises the only variant identified after filtering with a CADD score >15. **Figure 3-12** illustrates the presence of a homozygous frameshift *HERC2* variant in the largest homozygous region shared by the affected individuals.

Variant Information			
Gene	**Location**	**Variant**	**Protein change**
HERC2	15:28359900	c.13767_13770delTGAA	p.N4589KTer4598
Pathogenicity prediction			
Condel	**Polyphen2**	**SIFT**	**CADD score**
___	___	___	34
Frequency and function			
MAF gnomAD	**Protein Function**	**OMIM**	**In Homozygous region**
Absent	E3 ubiquitin-protein ligase	Angelman Syndrome	Yes

Table 3-5 Homozygous variant identified in family ND2.
Homozygous variant c.13767_13770delTGAA in *HERC2* (NM_004667.6) was annotated based on the human genome assembly GRCh37/hg19.

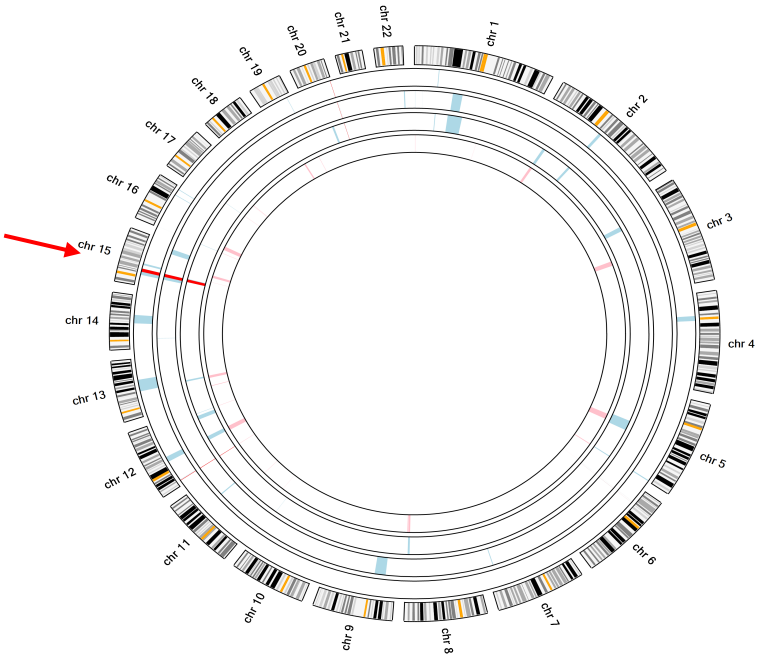


Figure 3-12 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND3.

The ideogram in light pink is from the unaffected individual, whereas the three outer ideograms in light blue are from the affected children. The regions highlighted in red represent the homozygous regions. The homozygous variant lies within the largest region on chromosome 15 (q13.1-q14) and it is indicated by the red arrow. The human genome assembly GRCh37/hg19 was used.

3.4.4 Variant confirmation using Sanger sequencing

Sanger sequencing was used to confirm the deletion identified by WES in the affected individuals. All members of the family were sequenced over this region using the relevant primers (see section 2.2.4). The c.13767_13770delTGAA variant in *HERC2* segregated with the condition, consistent with autosomal recessive inheritance. This research finding was reported back to the referring clinician. The electropherograms that summarise this research finding are shown in **Figure 3-13**.

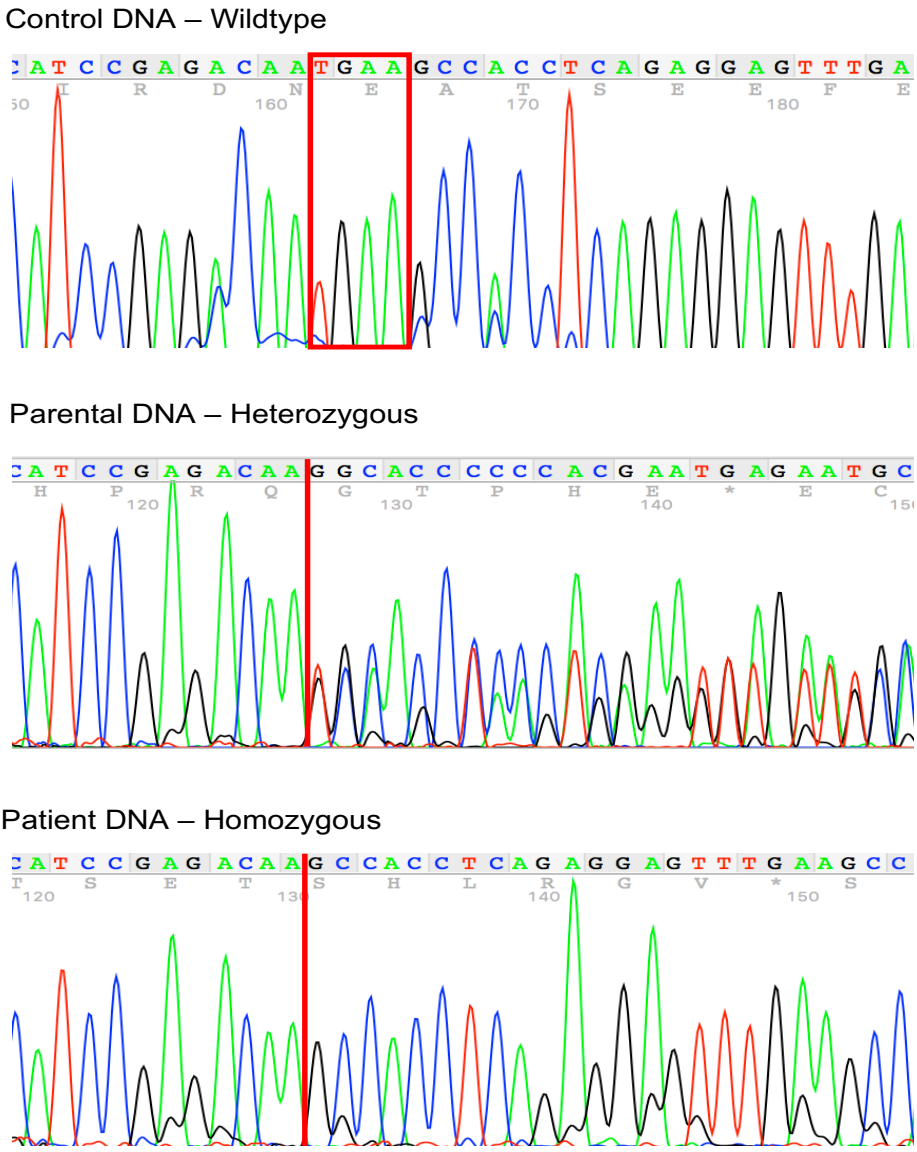


Figure 3-13 Electropherograms of the c.13767_13770delTGAA frameshift deletion in *HERC2*.

Electropherograms illustrating wildtype sequence from control DNA, heterozygous sequence from parental DNA and the homozygous mutant sequence from the patients DNA. The red frame marks the 4bp region that has been deleted and the red line highlights the break point of the deletion in the parental carrier and affected individual.

3.4.5 Variants in *HERC2* lead to severe developmental delay

3.4.5.1 *HERC2* and its biological function

The *HERC2* gene codes for a 527kDa E3 ubiquitin protein ligase consisting of multiple structural domains. Some of its key domains include the RCC1-like domain (RLD) that is predicted to act as a guanine nucleotide exchange factor, and the HECT domain that is shared by a number of E3 ubiquitin protein ligases and has enzymatic activity (**figure 3-14**) (Kuhnle et al., 2011). The main function of HERC2 is to control the ubiquitin-dependent retention of DNA repair proteins and specifically promote the ubiquitination and proteasomal degradation of XPA (Lee et al., 2014). XPA plays a crucial role in the nucleotide excision repair (NER) mechanism and its correct turnover by HERC2 can critically affect the circadian oscillation of the NER activity, particularly in the brain (Kang et al., 2010). Apart from XPA, HERC2 appears to interact with proteins implicated in mitosis and cell cycle regulation. Possible interactors of HERC2 include UBE3A, ASPM, PCM1 and CEP170 (Galligan et al., 2015).

UBE3A encodes another E3 ubiquitin protein ligase that is involved in the ubiquitin protein degradation system (Philpot et al., 2011). This gene is paternally imprinted and the maternally inherited allele is expressed in the brain (Sell and Margolis, 2015). Maternal inheritance of a deletion on chromosome 15q11.2 – q13, encompassing the *UBE3A* gene, has been associated with Angleman syndrome. This is characterised by intellectual disability, movement disorder, typical abnormal behaviours and speech limitations (Williams et al., 2010). These clinical features tend to overlap with many Angelman-like phenotypes, covering a wide range of genotypes, that are otherwise clinically indistinguishable. However, the genotype-phenotype correlation for specific genes such as *HERC2* have not been determined, comprising an important clinical need.

Abnormal Sprindle-like Microcephaly associated protein, ASPM, is a large protein involved in the regulation of mitotic spindle formation and microtubule dynamics. Expression studies of *Aspm* mRNA in mice revealed that *ASPM* seems to be significantly expressed at sites of active neurogenesis (Bond et al., 2002). Mutations in this gene are linked with Type 5 primary microcephaly, with mild to moderate intellectual disability (Bond et al., 2003).

In addition, *PCM1* encodes pericentriolar material 1 protein that is required for centrosome assembly and function. In particular, PCM1 is a component of the centriolar satellites and it is involved in correct anchoring of microtubules to the centrosome (Dammermann and Merdes, 2002). It has also been shown that PCM1 forms granules

that are concentrated around centrioles, which tend to disassemble during mitosis and reassemble once the cells enter interphase in a cell-cycle dependent manner (Kubo and Tsukita, 2003). PCM1 seems to also be implicated in regulating primary cilia disassemble before entering mitosis (Villumsen et al., 2013).

Furthermore, *CEP170* is a gene encoding centrosomal protein 170 which is another component of the centrosome. Immunoelectron microscopy revealed that CEP170 specifically localizes to the mother centriole, at the subdistal appendage. CEP170 seems to be implicated with spindle formation, and maintaining microtubular organisation and cellular morphology (Guarguaglini et al., 2005).

3.4.5.2 The *HERC2* frameshift variant leads to complete loss of protein

The c.13767_13770delTGAA four nucleotide deletion in *HERC2* lies within the HECT domain, and the frameshift results in a stop codon eight amino acids after the breakpoint. The impact of this variant on the encoded protein was further investigated, in order to clarify if the stop codon results in a truncated protein with impaired E3 ligase activity or if the mRNA undergoes nonsense-mediated decay. The premature stop codon could also lead to exon-skipping with retention of minor protein isoforms, that was also worth investigating.

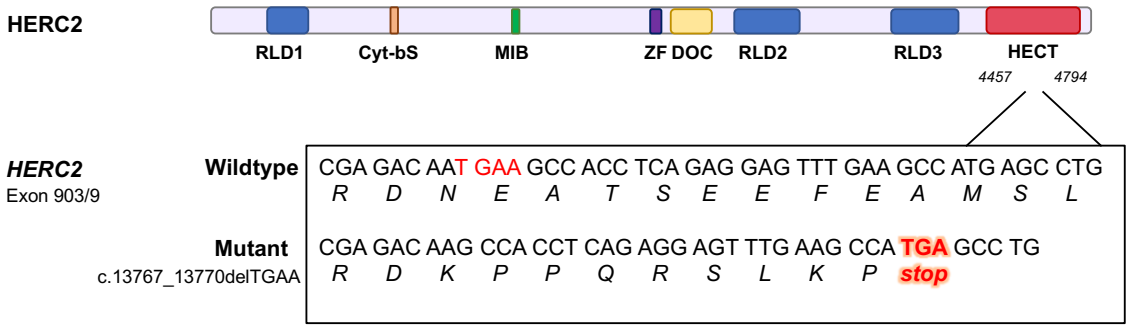


Figure 3-14 Illustration of the HERC2 protein and its multiple domains

Schematic diagram illustrating the different domains of the HERC2 protein (Galligan et al., 2015) and outlining the impact of the four nucleotide deletion within exon 90, resulting in a downstream stop codon.

To accomplish this, soluble protein was extracted from primary control HDFs and primary patient fibroblasts obtained from a skin biopsy. Protein was then denatured and run on an SDS-PAGE and analysed using Western blotting (see section 2.2.19) staining with an anti-HERC2 antibody (**Table 2.4**). The staining revealed complete loss of full-length HERC2 protein as a result of the frameshift deletion. However, there is a lower size band appearing as a result of either non-specific binding of the antibody or of a possible

retention of a smaller 270kDa isoform that has not been identified yet. (**Figure 3-15**). Based on this result the patient fibroblasts were treated as *HERC2*-null and were subsequently used to further characterise the impact of the mutation on cellular phenotypes and protein interactions.

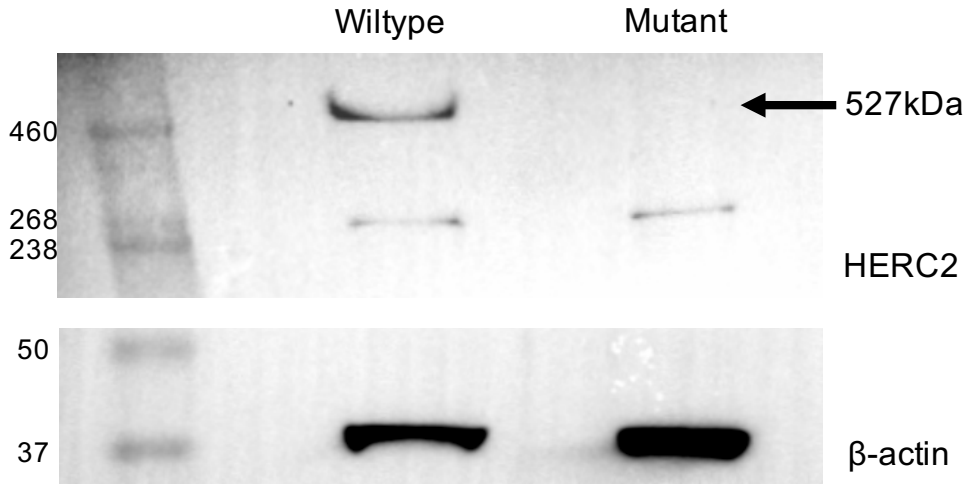


Figure 3-15 Complete loss of HERC2 protein revealed by Western blot.

Western blot analysis showed absence of a band at the expected size of 527kDa for the mutant protein when stained with the anti-HERC2 antibody. The epitope recognised by the antibody maps to the RLD2 domain and between 4784-4834 of the HECT domain. There is a lower 270kDa band on the membrane, but this is likely to be due to non-specific binding of the antibody rather than a minor protein isoform because there is no known HERC2 coding transcript that encodes a protein of that size.

3.4.5.3 Loss of HERC2 elicits impaired mitochondria

Once it was established that the four nucleotide deletion caused protein loss, the *HERC2* null fibroblasts were used as a cellular model of disease. This is currently the best model to study the pathomechanisms of HERC2-related disease, considering the size of this protein and the obvious limitations of over-expression studies.

3.4.5.3.1 *HERC2* null cells exhibit striking mitochondrial fragmentation

Based on some of the clinical features of the affected children, particularly muscle atrophy and hypotonia, we considered it important to investigate the impact of the *HERC2* c.13767_13770delTGAA frameshift variant on mitochondrial function. Loss of HERC2 is likely to affect many cellular processes, specifically cell cycle regulation and DNA repair mechanisms, resulting in cellular stress and defective cellular homeostasis

as a downstream consequence. Unbalanced homeostasis can be critical in mitochondrial dynamics and function (Willems et al., 2015).

In an attempt to assess these cellular phenotypes, the fibroblasts were seeded on coverslips and mitochondrial number and morphology visualized by immunofluorescence microscopy (section 2.2.17). Striking mitochondrial fragmentation was apparent in the mutant fibroblast compared to the control when visualized for the MTCO2 mitochondrial marker(**Figure 3-16**).

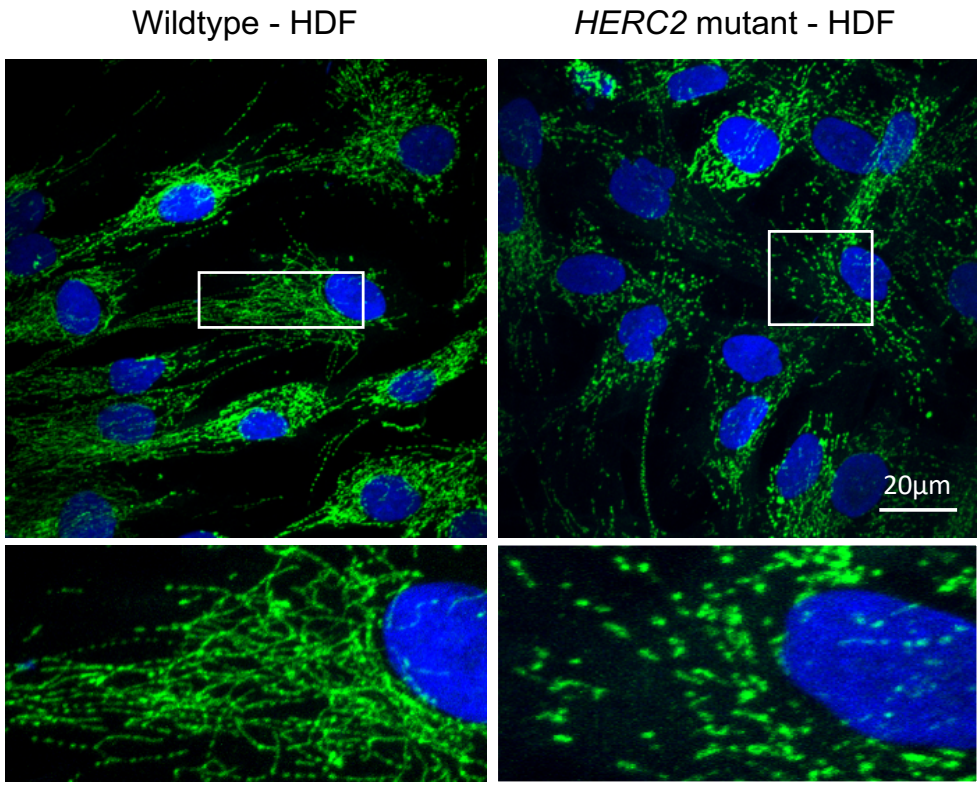


Figure 3-16 IF confocal microscopy of *HERC2* null fibroblasts revealed mitochondrial fragmentation.

Mitochondrial staining with MTCO2 antibody (green) in wildtype and *HERC2* mutant fibroblasts, showing mitochondrial fragmentation (detail shown in magnified insets). Scale bar = 20 µm.

3.4.5.3.2 Mitochondrial dysfunction due to *HERC2* mutation

In order to further elucidate the effect of a *HERC2* null mutation on mitochondrial function and cellular phenotypes, the oxidative phosphorylation pathway of energy production was tested using a Seahorse XFe96 Extracellular Flux Analyzer running the the MitoStress test (section 2.2.22). Control and mutant fibroblasts (annotated as HDF and HERC2 respectively) were challenged with various toxins that block different processes of the oxidative phosphorylation pathway, in three biological replicates. The oxygen consumption rate (OCR) of the cells was measured before and after each toxin injection allowing for identification of important measures of mitochondrial function. The trace below shows the changes in OCR in real time (**Figure 3-17**).

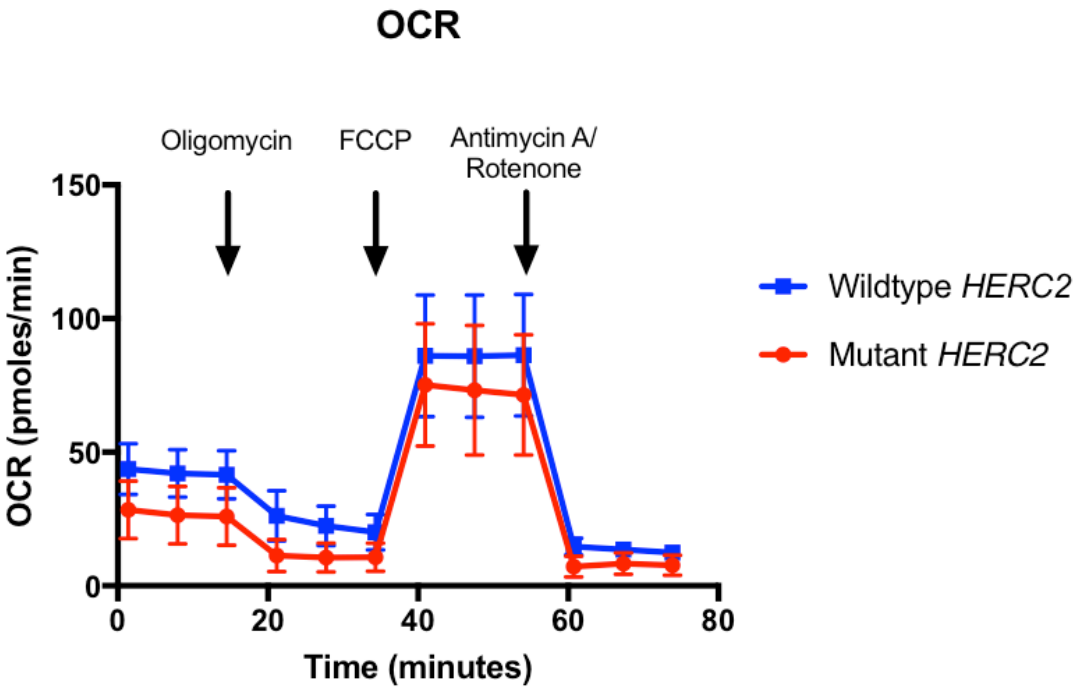


Figure 3-17 Seahorse trace for the MitoStress test of *HERC2* null fibroblasts

Traces of real-time OCR changes between wildtype and mutant fibroblasts. The first injection was with oligomycin to block ATP synthase, the second injection was with FCCP to uncouple the electron transport chain and, lastly, Antimycin A/ Rotenone completely blocked electron transport.

Further analysis of the OCR measurements at certain time points of the experiment allow for identification of key aspects of mitochondrial function, such as basal and maximal respiration, ATP production, proton leak and spare respiratory capacity. The analysis of each one of these mitochondrial characteristics is summarised in **Figure 3-18**. Basal respiration, ATP production, proton leak and coupling efficiency were significantly different in mutant *HERC2* cells compared to wildtype.

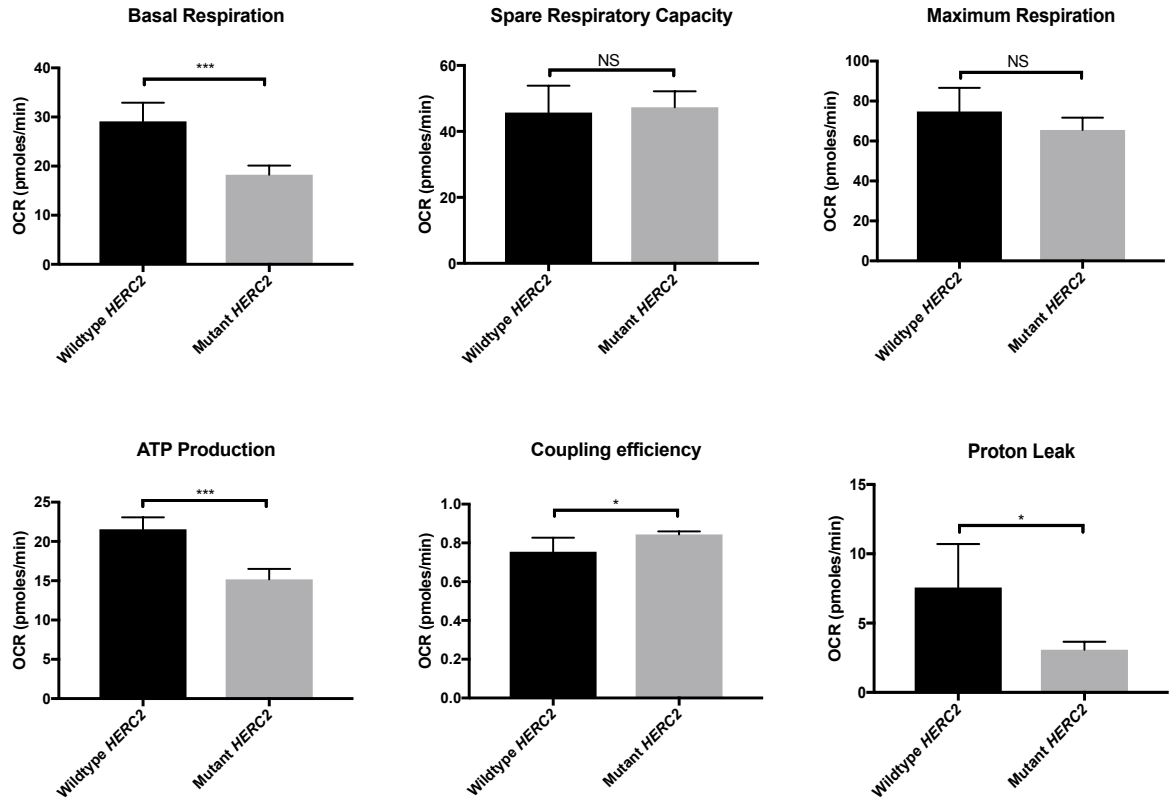


Figure 3-18 Metrics of mitochondrial function in mutant *HERC2* and normal wildtype fibroblasts.

Bar graphs that quantitate the metrics obtained from the MitoStress Test. Measurements at specific time points before or after toxin injections allowed for the calculation of each of the key aspects of mitochondrial function. *HERC2* mutant fibroblasts were significantly under-respiring at basal levels, producing less ATP by oxidative phosphorylation and had a disrupted proton-motive force. All statistical tests were performed by a two-tailed Student t-test (n=3; ns, not significant; * $p<0.05$; *** $p<0.001$). Error bars indicate s.e.m.

The results of the MitoStress test indicated that, when forced, the mitochondria of the *HERC2* mutant cells can reach almost the same maximum respiration level as in the wildtype control cells. This is also validated by the fact that there is no significant difference between their spare respiratory capacity and their coupling efficiency. However, under normal respiratory conditions the basal respiration of the *HERC2* mutant cells is significantly less than the basal respiration seen in normal control cells. This was one of the indications that suggests the *HERC2* mutant cells have switched their energy production to other means rather than oxidative phosphorylation. To test this hypothesis, we performed a Glycolysis stress test (**Figure 3-19**). In addition, the MitoStress test also revealed that the overall ATP production from oxidative phosphorylation in the *HERC2* mutant cells is significantly lower than in normal controls, supporting the hypothesis that mutant cells switch to other less efficient processes for energy production. One of the reasons for doing so could be a defect in the ionic homeostasis due to altered proton

leak in the mitochondria, affecting the proton-motive force and therefore the production of ATP (Papa et al., 2012). The observations from the MitoStress test were informative, but further experiments are needed to understand the molecular mechanism of the defect and thereby the origin of the mitochondrial fragmentation phenotype (**Figure 3-16**).

In order to further investigate this observation, a Glycolysis Stress test was performed. Normal control and mutant fibroblasts were challenged with various toxins that block different processes of either glycolysis or the oxidative phosphorylation pathway, in three biological replicates. The extracellular acidification rate (ECAR) of the cells was measured before and after each injection, allowing for the identification of key aspects of glycolysis. The trace below shows the changes in ECAR in real time (**Figure 3-19**).

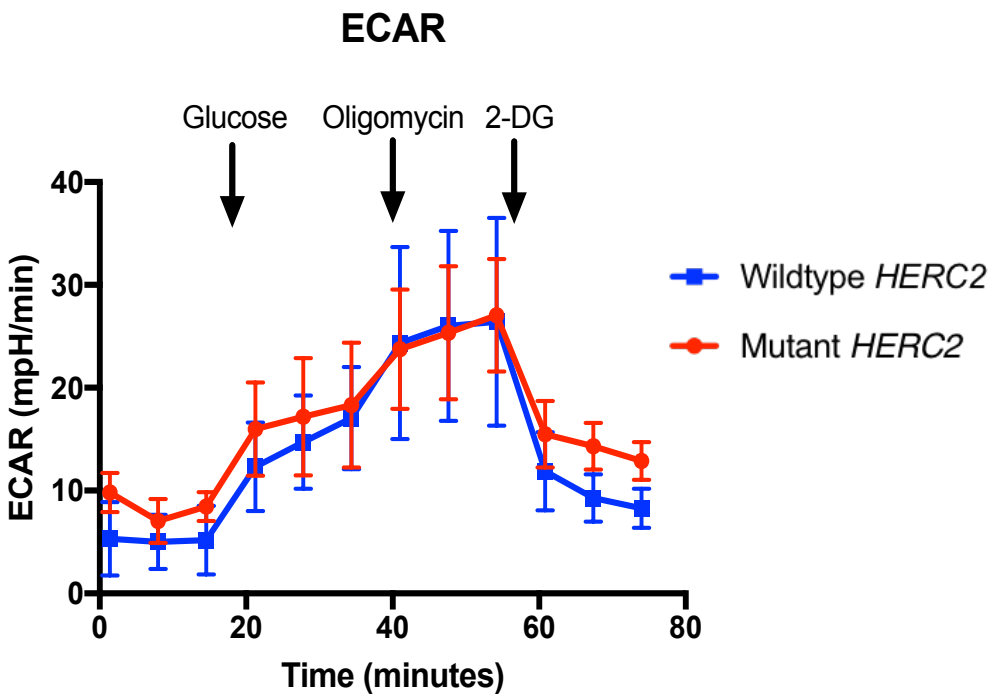


Figure 3-19 Seahorse trace for the Glycolysis Stress test of *HERC2* null fibroblasts

Representation of the changes in ECAR in real time between normal control and mutant fibroblasts. The first injection was with glucose to promote glycolysis, then oligomycin was added to block the ATP synthase and any ATP production from oxidative phosphorylation, and the last injection was with 2-deoxyglucose (2-DG), a structural analogue of glucose that completely blocks glycolysis.

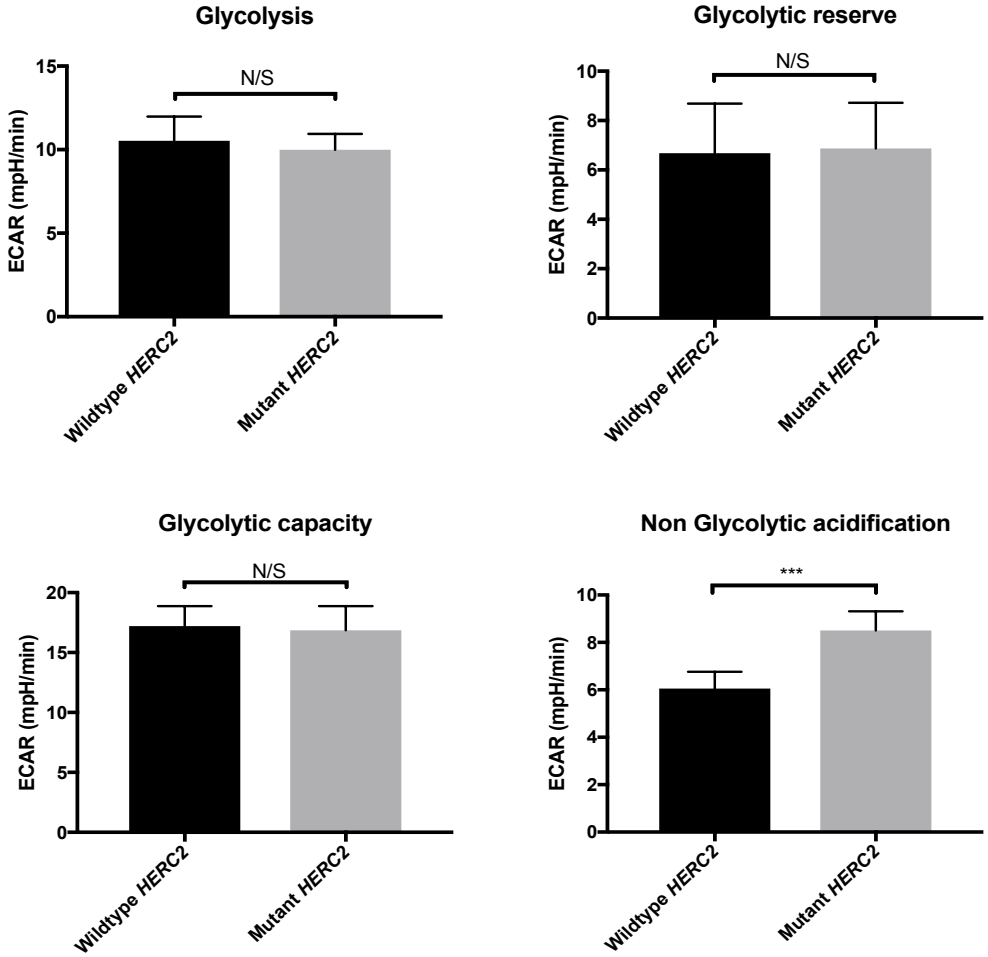


Figure 3-20 ECAR metrics accessing glycolysis.

Representation of all the metrics obtained from the Glycolysis Stress test. Measurements at specific time points before or after each injection allowed for the calculation of each of the above properties to be obtained (Agilent protocol). *HERC2* mutant fibroblasts seem to have the same glycolytic phenotype as the control HDF cells, except from the non-glycolytic acidification measurement. All statistical tests were performed by a two-tailed Student t-test (n=3; ns, not significant; * $p<0.05$; ***$p<0.001$). Error bars indicate s.e.m.

The results from the Glycolysis Stress test did not reveal any great differences between the normal control and *HERC2* mutant fibroblasts. The level of glycolysis (or in other words the conversion of glycose to pyruvate) was almost the same between the two cell types. A non-significant difference was also observed in the glycolytic reserve and the glycolytic capacity between the wildtype and mutant cell line. This indicates that under normal conditions both cell types have the same rate of glycolysis and, when forced, both cell types have the potential to reach maximum glycolysis. The only significant difference observed was in the non-glycolytic acidification rate observed in the wildtype and mutant fibroblast. *HERC2* mutant fibroblasts had a significantly increased non-glycolytic acidification level compared to the wildtype fibroblasts, indicating a source of acidification separate to the glycolysis pathway. This can potentially arise from the

conversion of pyruvate to lactate that is further reduced to lactic acid inducing acidification (Mookerjee et al., 2015). The diagram in **figure 3-21** outlines the two possible outcomes following glycolysis.

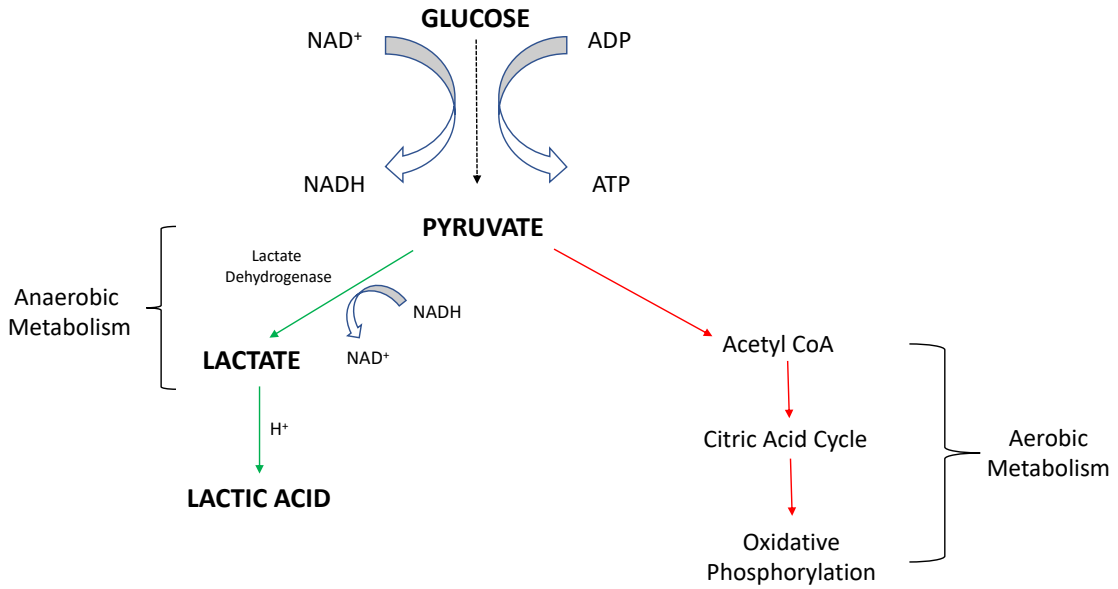


Figure 3-21 Outline of aerobic and anaerobic metabolism pathways

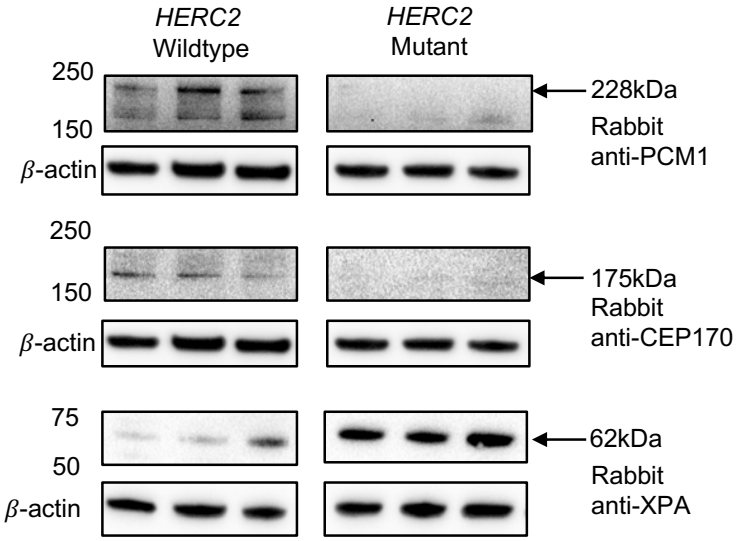
The diagram illustrates the two potential uses of pyruvate in aerobic and anaerobic metabolism. Pyruvate can be converted into Acetyl CoA and fed into the Kreb's cycle and the oxidative phosphorylation chain for maximum production of ATP by aerobic respiration. Pyruvate can also be converted to lactate for immediate energy demands when aerobic metabolism cannot produce enough ATP at any given instance, for example during exercise. Lactate can be further reduced to lactic acid leading to acidification.

Overall, this investigation has revealed that there is a defect in the oxidative phosphorylation chain resulting in reduced basal aerobic respiration and ATP production in *HERC2* null fibroblasts. This could be due to the proton-motive force, or more specifically due to one of the complexes of the oxidative phosphorylation chain, being impaired (Ghezzi and Zeviani, 2018). In addition, the mitochondria could be under excessive stress resulting in impaired aerobic metabolism and fragmentation of the network. This is also supported by the fact that the mutant cells tend to have increased acidification levels, perhaps due to the conversion of pyruvate to lactate in order to maintain their energy requirements. These are interesting descriptive observations, but further experiments are required in order to confirm that lactate is the source of acidification observed, and to further try and understand how the *HERC2* mutation can have such a severe effect on mitochondrial function.

3.4.5.4 Loss of HERC2 impacts other protein-protein interactions

Once it was established that the *HERC2* frameshift mutation led to complete loss of the protein, it was also worth investigating the impact that HERC2 loss had on potential interacting proteins (discussed in section 3.4.5.1). Soluble protein was extracted from normal control and HERC2 null patient fibroblasts and run on western blots to assess the impact on the protein levels of interacting proteins and potential HERC2 substrates. Since HERC2 is an E3 Ubiquitin-ligase protein, we hypothesized that HERC2 loss would increase the protein levels of potential substrates. For example, XPA appears to be ubiquitinated by the HERC2 E3 ligase at K48 for regulated degradation through the ubiquitin-proteasome system (Lee et al., 2014). Additionally, other interacting proteins that may be ubiquitinated by HERC2 for signalling purposes, via K11 or K63 for example, could be potentially affected by the loss of HERC2. The result of this investigation is summarised in **figure 3-22**.

A



B

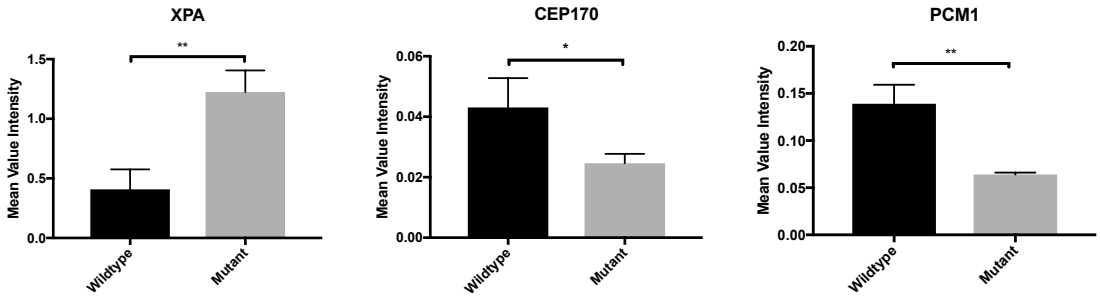


Figure 3-22 Western blot analysis to investigate the impact of HERC2 deficit to other interacting proteins.

A – Representation of protein levels on western blot membranes for the indicated interacting protein compared to loading control β-actin. Western blots include all three biological replicated of each experiment. B - Bar graphs quantitating the results from all three biological replicates normalised to β-actin. All statistical tests were performed by a two-tailed Student t-test (n=3; ns, not significant; * p<0.05; ***p<0.001). Error bars indicate s.e.m.

The western blot analysis of three interacting proteins of HERC2 revealed significant alterations in their protein levels. For example, levels of XPA are significantly higher in mutant cells compared to the controls, supporting the fact that XPA could be one of the substrates of HERC2 that is tagged for degradation. Loss of HERC2 then potentially causes XPA accumulation in the cells that could either affect its function in DNA repair mechanisms or promote a pathogenic outcome due to aggregating protein complexes. In addition, levels of CEP170 and PCM1 are significantly reduced in the mutant cells compared to the controls, potentially disrupting downstream processes of mitosis and cell-cycle regulation. In these instances, HERC2 could mediate ubiquitin signalling cascades that are separate to protein degradation pathways. These observations are worth investigating further to understand more about the function of HERC2 and its interacting partners, and how loss of HERC2, can have such a dramatic impact on health. Due to time constraints and reagent limitations, this study only investigated the protein levels of a few candidate interactors that appear to mediate cell cycle regulation. However, HERC2 is a massive protein and is likely to be implicated in many other protein-protein interactions and to have diverse roles on cellular processes.

In addition, the only variant in *HERC2* that has been associated with an Angelman-like phenotype is the homozygous missense p.Pro594Leu variant (Puffenberger et al., 2012), that seems to be better tolerated resulting in a less severe developmental phenotype. However, a 286kb homozygous deletion over a region comprising the *HERC2* gene, appears to cause a much more severe form of developmental delay and lethality (Morice-Picard et al., 2017). The complete loss of a large and important E3 ligase will have a detrimental effect on diverse cellular processes, including DNA damage repair and cell cycle regulation, which is supported by the observation that the *HERC2* homozygous null mouse model is embryonically lethal (Cubillos-Rojas et al., 2016). From a clinical perspective, this work establishes, for the first time, a clear genotype-phenotype correlation for the *HERC2* gene. This finding will allow clinicians to offer diagnostic testing to affected families and to include *HERC2* in the genetic screening of patients with developmental delay but without a molecular diagnosis. This is important for the future stratification and management of inherited disorders with non-syndromic, profound developmental delay because, without a molecular diagnosis, these conditions are often clinically indistinguishable.

Chapter 4

Candidate gene discovery using whole exome sequencing for autosomal recessive neuromuscular or neurodevelopmental disorders

4.1 Introduction

This chapter summarises the library preparation and data analysis for Whole Exome Sequencing of eight families that were recruited to the “Jules Thorn Biomedical Award” genetic research study. For each affected individual presented in this chapter, a clinical overview and variant filtering strategy is given. Candidate variants are then discussed in detail where appropriate, and for any unsolved families other alternative methods are proposed for gene discovery purposes.

Candidate variants presented in this chapter have not been functionally characterised as part of this study. Cases where the candidate variants have been further functionally characterised are discussed in detail in chapters 5 and 6.

A number of possible disease-causing variants identified from these cases were then uploaded on the Deciphering Developmental Delay (DDD) project, in an attempt to discover additional individuals and families with similar phenotypes and mutations in the same gene.

4.2 Methodology for gene discovery

4.2.1 Patient Recruitment and Phenotyping

From the families recruited for the “Jules Thorn Biomedical Award” genetic research study, eight were investigated as part of this chapter. All family members that were included in the study gave informed consent for research (see section 2.2.2) and their DNA was collected either from blood or saliva samples. Affected individuals were phenotyped by Clinical Geneticists in regional hospitals (see section 2.2.1).

4.2.2 Whole Exome Sequencing

Genomic DNA samples from affected individuals, parents or unaffected sibling were taken forward for WES library preparation using the Illumina QXT protocol (section 2.2.10). In addition to sequencing family trios (both parents and affected child) used in sequencing, additional affected individuals or unaffected siblings were also sequenced to narrow down the candidate genes by segregation.

During library preparation, Bioanalyzer traces were used to assess the size and quantity of the DNA fragments at post-shearing and post-hybridization stages (Appendix D). The samples were pooled together aiming for a total of ten samples per lane and sequencing was performed on the Illumina HiSeq 3000 platform.

Prior to data analysis, the .fastq files generated and subsequent .bam files were also evaluated for the quality of sequencing (Appendix F). Subsequently, the depth of coverage across the exome was also assessed as an additional quality check of the sequencing libraries. An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance and prioritizing variants in previously identified homozygous regions (see section 2.2.10.6). In non-consanguineous families, all biallelic variants were considered without taking into consideration the homozygous regions. In addition, segregation analysis was run, filtering out any variants that did not segregate with the family pedigree. CNV analysis was also carried out for all cases outlined in this chapter, to look for larger duplications or deletions that were not detected by the standard pipeline used for analysis (see section 2.2.10.3). In cases with only a single affected individual, in a family with multiple siblings, analysis to identify potential *de novo* variants was also used to check this possibility.

Candidate genes were then prioritized based on the function of the encoded protein and the relevance to disease phenotype. Variants with convincing pathogenicity scores (CADD scores $\geq$10) are outlined in the first section of this chapter, whereas cases with no candidate genes are discussed at the last section. Novel potentially pathogenic variants will require further validation work to provide sufficient evidence that the variant identified affects the function of a protein and downstream processes. On the other hand, cases with no identified candidate variants that could be interpreted as pathogenic are appropriate cases for further NGS investigations using Whole Genome Sequencing (WGS). Intronic or regulatory region variants might be the cause of the disease, but WES often does not provide coverage of these regions to detect such variants.

4.3 Families with variants in Candidate Genes

4.3.1 Family ND4

4.3.1.1 Clinical Phenotype

Family ND4 is a multiply consanguineous South Asian family with several affected individuals in two different generations. The clinical phenotype shared by all affected individuals include microcephaly, central hypotonia and four limb spasticity. Individuals in generation V have been extensively examined by the Clinical Geneticists revealing right convergent squint and reduced white matter volume on MRI scans. Individual JT807 was assessed at a later stage. The affected individuals (in generation V) were also screened for a panel of genes known to cause white matter abnormalities, but no variants were found. The details of the diagnostic screening panel are described here: (https://www.leedsth.nhs.uk/a-z-of-services/the-leeds-genetics-laboratory/constitutional-genetics/molecular-genetics/by-disorder/leukodystrophy/).

DNA samples have been collected from all individuals and assigned a study identifier ("JT number") as indicated on the pedigree in **figure 4.1**. Samples from all labelled individuals were processed for WES, except from individual JT807 as the DNA sample was not available at the time. However, DNA sample from JT807 was further used for segregation purposes.

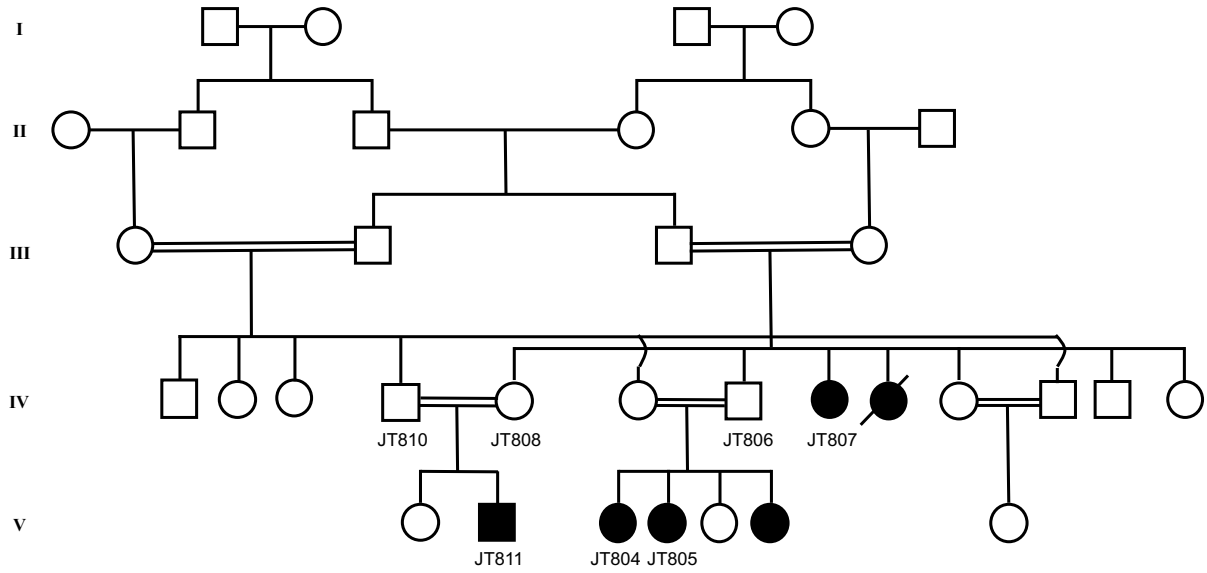


Figure 4-1 Pedigree of family ND4

Pedigree outlining five generations of the family and the consanguineous unions, indicated with a double line. DNA samples of individuals with study identifiers ("JT numbers") have been used for genetic investigations.

4.3.1.2 Autozygosity Mapping

Autozygosity mapping was performed on all available WES data (vcf files prior to filtering out of common SNPs) from all members of the family. Homozygous regions shared by the affected individuals (JT804, JT805, JT811), but not by the unaffected parents are summarized in **table 4-1** and presented on the ideogram in **figure 4-2**.

Chromosome	Start	End	Length
3	124646705	133191453	8544748
10	64944317	73403490	8459173

Table 4-1 Common regions of homozygosity amongst affected individuals of family ND4.

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

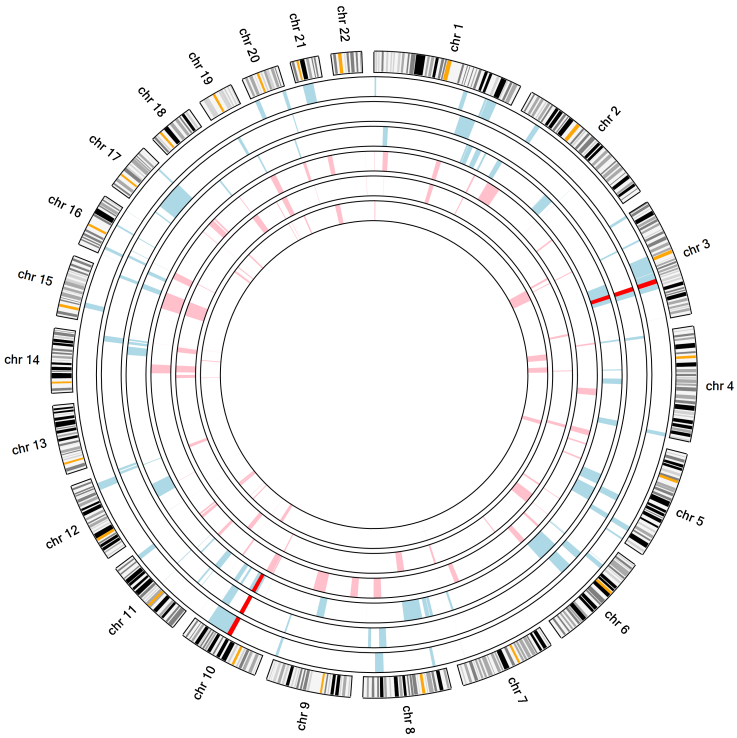


Figure 4-2 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND4.

The ideograms in light pink are from the unaffected individuals, whereas the outer ideograms in light blue are from the affected children. The regions highlighted in red represent the homozygous regions in chromosome 3 and 10. The human genome assembly GRCh37/hg19 was used.

4.3.1.3 Candidate gene

An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance and prioritizing variants in previously identified homozygous regions. In addition, segregation analysis was performed, filtering out any variants that did not segregate in a pattern consistent with the family pedigree. **Table 4-2** summarizes the only homozygous variant identified with a CADD score above 15. The variant also lies within the homozygous region on chromosome 10, supporting our initial hypothesis that, for a multiply consanguineous family such as ND4, it is more likely for the pathogenic variant to be within a homozygous region.

Variant Information			
Gene	**Location**	**Variant**	**Protein change**
SUPV3L1	10:70958197	c.1093C>T	p.R365W
Pathogenicity prediction			
Condel	**Polyphen2**	**SIFT**	**CADD score**
deleterious (0.645)	possibly_damaging (0.582)	deleterious (0)	34
Frequency and function			
MAF gnomAD	**Protein Function**	**OMIM**	**In Homozygous region**
0.00001769	ATP-dependent RNA helicase	–––	Yes

Table 4-2 Homozygous variant identified in family ND4.

Homozygous variant c.1093C>T in *SUPV3L1* (NM_003171.5) was annotated based on the human genome assembly GRCh37/hg19.

In addition to the standard pipeline, CNV analysis was also carried out for this family (see section 2.2.10.7) to identify any potential CNV that could cause the phenotype. CNV analysis did not reveal anything that could be interpreted as pathogenic, increasing our confidence that the candidate gene variant in *SUPV3L1* was the causative variant.

The variant p.R365W in *SUPV3L1* can be assumed to significantly impact the structure and function of the protein as a polar amino acid, arginine, is replaced by a non-polar aromatic amino acid, tryptophan. This change is predicted by Condel, Polyphen2 and SIFT to be deleterious, with a striking CADD score of 34. The p.R365W SUP3VL1 variant has a low frequency in gnomAD but not in homozygous state. Segregation analysis for all members of family ND4 confirmed the variant in *SUPV3L1* identified by WES. Patients

were homozygous for the change and all unaffected parents were heterozygous. The segregation analysis is summarised in **figure 4-3**.

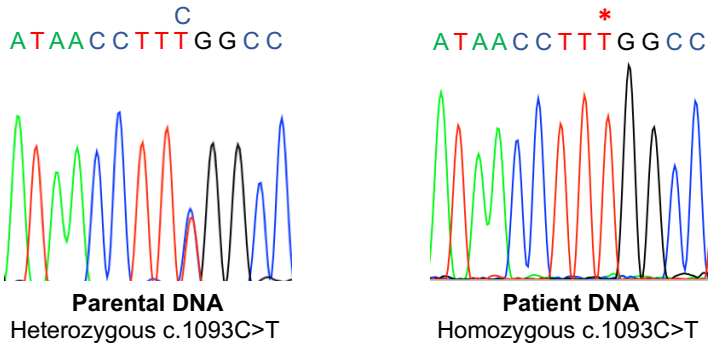


Figure 4-3 Segregation analysis for family ND4.

The electropherograms illustrate the homozygous change c.1093C>T in patient DNA and the heterozygous state in the parental DNA. Sanger sequencing was performed on all members of the family with JT identifier.

Conservation analysis revealed that the p.R365W variant is conserved in a proportion of the species outlined in **figure 4-4**. Despite the fact that the actual mutated residue is not fully conserved, in species such as *M.musculus*, the surrounding region is well conserved. Specifically, the position of the p.R365W variant lies between a predicted alpha helix and a beta strand, where both of these secondary structures seem to be conserved from *H. sapiens* to *X. tropicalis*. The addition of a big aromatic ring between these structures could impact the tertiary and quaternary structure of the overall protein, potentially leading to disruption of function. Visualization of the exact location of arginine at position 365 is shown in **figure 4-5**, in relation to the partial structure of SUPV3L1 between amino acids 47 and 772. It seems that arginine 365, a polar basic residue, is exposed on the outer part of the structure. A dramatic effect on protein structure is predicted if arginine is replaced by the non-polar and hydrophobic tryptophan.

Figure 4-4 Conservation analysis of the SUPV3L1 protein

A representation of the conservation analysis of SUPV3L1 protein around the location of the variant identified, showing protein sequences from *H. sapiens* down to *X. tropicalis*. The amino acid mutated is shown by the arrow.

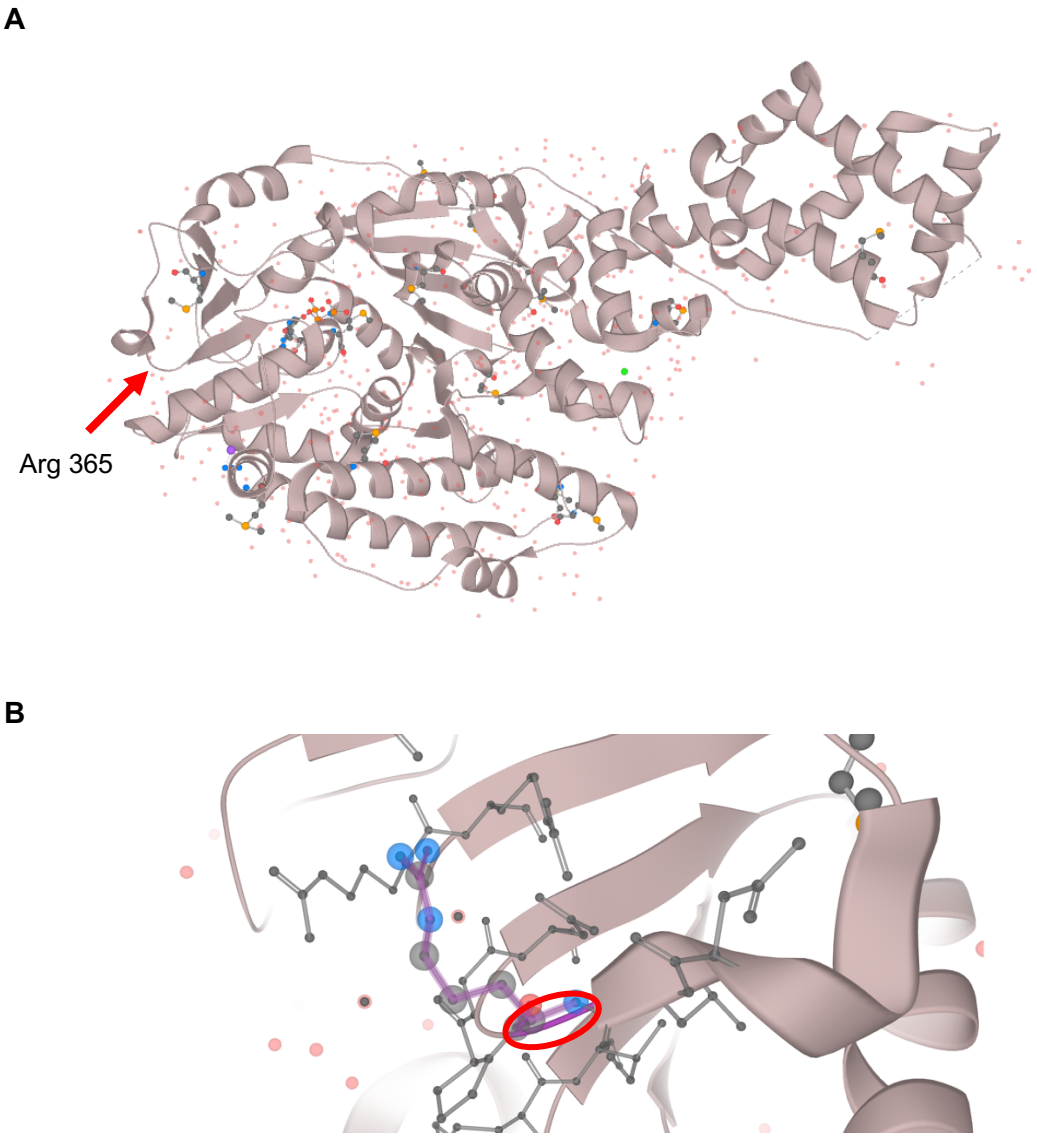


Figure 4-5 Visualisation of the position of arginine 365 on the overall structure of SUPV3L1 protein.

A – Illustration of the overall structure of SUPV3L1, indicating the position of the mutated amino acid by an arrow. B – Zoom-in version of the same structure showing that arginine 365 lies between an α-helix (grey ball-and-stick structures) and a β-sheet (beige ribbon) on the outermost part of the protein, which can be detrimental for the replaced hydrophobic tryptophan.

4.3.1.4 Variant in *SUPV3L1* as a potential cause of developmental delay

The *SUPV3L1* gene is located on chromosome 10q22.1 and it encodes an ATP-dependent RNA helicase that is localised mainly in the mitochondria (Dmochowska et al., 1998). SUPV3L1 is considered to have a major role in mitochondrial RNA metabolism and is involved in the RNA surveillance system in mitochondria, by regulating the mtRNA turnover and processing (Khidr et al., 2008, Szczesny et al., 2007). In addition, down-regulation of SUPV3L1 induces caspase-dependent apoptosis (Szczesny et al., 2007) and *Supv3l1* deletion in mice has been proven to be embryonic lethal . Another mouse

model where *Supv3l1* was progressively inactivated resulted in mutant mice that appeared normal at birth but gradually revealed signs of delayed growth, loss of adipose tissue and muscle mass, reduced life span (Paul et al., 2009).

Considering all of the above evidence on *SUPV3L1* function, there was some justification as to how this candidate gene could potentially affect growth and development as seen in family ND4. In addition, as a protein localised in the mitochondria, SUPV3L1 could be interacting with other key proteins that have been shown to be involved in developmental disorders, specifically with white matter abnormalities. In particular, SUPV3L1 is known to interact with PNPT1, an RNA binding protein involved in RNA metabolism and a component of the mitochondrial RNA degradosome. The mitochondrial RNA degradosome complex is a heteropentamer consisting of a PNPT1 trimer and a SUPV3L1 dimer, that in yeast has been demonstrated to degrade double-stranded RNA in the presence of ATP (Wang et al., 2009). If one of the two components of the heteropentamer is disrupted, the activity of the degradosome is abolished.
Interestingly, biallelic variants in *PNPT1* are also shown to be associated with abnormal myelination in patients with neurodevelopmental disease and mitochondrial dysfunction (Sato et al., 2018a). This association of *PNPT1* variants with a similar phenotype to affected individuals in family ND4 provides additional more evidence that variants in *SUPV3L1* could also lead to a neurodevelopmental condition with white matter abnormalities.

4.3.1.5 Second family with variants in *SUPV3L1* identified from the Decipher consortium

The findings from family ND4,, amongst many others, had been uploaded on the Decipher consortium website (www.decipher.sanger.ac.uk) in an attempt to identify a second family with the same phenotype having a biallelic variant in *SUPV3L1* gene (Firth et al., 2009). A second family have been matched to our case, with the homozygous variant c.329G>A, p.G110D in SUPV3L1 (Patient ID: 386801, online access: https://decipher.sanger.ac.uk/patient/386801#genotype/snv/69993/browser). The clinical features of the family include profound developmental delay, intellectual disability, CNS hypomyelination, nystagmus and seizures. Despite the fact that additional evidence is now available indicating that variants in *SUPV3L1* are likely to be the cause of a novel neurodevelopmental disorder, no further work has been performed on this project due to time constraints. The second family was identified at the end of this study and future work is required to establish the pathomechanism of *SUPV3L1* variants.

4.3.2 Family ND5

4.3.2.1 Clinical Phenotype

Family ND4 is a non-consanguineous family with twin girls born with antenatally agenesis of the corpus callosum. At delivery both infants had marked pedal oedema, and pleural oedema both of which subsequently resolved. Central hypotonia was a noticeable feature of the phenotype, the infants had reasonable anti-gravity movements of all four limbs, with poor head movement, suck and initially respiratory distress. Neither infant required intubation, and both were tube fed for six weeks. At last review (aged 8 months) central hypotonia persisted, and delayed milestones were noted (not sitting unsupported, not rolling from side to side, not reaching out for objects). Vision and hearing were normal. Clinical investigations performed include aCGH (array comparative genomic hybridization), and a focused clinical exome for neonatal hypotonia testing for variants in the following genes: *GJC2*, *VEGFC*, *FLT4*, *LMPHM2*, *EPHB4*, *PIEZO1* and *FAT4*. These and other clinical investigations were normal.

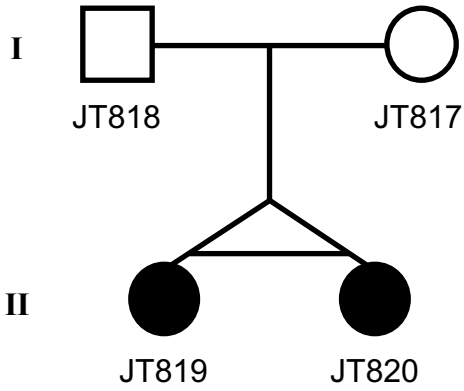


Figure 4-6 Pedigree of family ND5

Pedigree showing two generations of the family. DNA samples of the parents and one of the monozygotic twins have been used for genetic investigations.

4.3.2.2 Candidate gene

Our in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance, prioritizing both homozygous and compound heterozygous variants, as this family has no evidence of consanguinity. For the same reason, autozygosity mapping was not performed for this family as the pathogenic variant is less likely to be in an autozygous region. In addition, segregation analysis was run, filtering out any variants that had a pattern of segregation that was not compatible with

the pedigree. Separate *de novo* analysis was also performed, revealing no variants. **Table 4-3** summarises the three variants identified with a CADD score above 15.

Variant Information				
Gene	**Location**	**Variant**	**Protein change**	
EPHB2	1:23234500	het c.2191G>A	p.A731T	
EPHB2	1:23236988	het c.2616G>C	p.K872N	
CARS2	13:111340076	hom c.563G>A	p.T188M	
Pathogenicity prediction				
Gene	**Condel**	**Polyphen2**	**SIFT**	**CADD score**
EPHB2	deleterious (0.791)	possibly_damaging (0.887)	deleterious (0)	34
EPHB2	deleterious (0.516)	possibly_damaging (0.501)	deleterious (0.04)	23.6
CARS2	deleterious (0.897)	probably_damaging (0.994)	deleterious (0)	24.5
Frequency and function				
Gene	**MAF gnomAD**	**Protein Function**	**OMIM**	**In Homozygous region**
EPHB2	0.00008837	Tyrosine-protein kinase Eph receptor B2	Prostate cancer (somatic mutations)	N/A
EPHB2	Absent	Tyrosine-protein kinase Eph receptor B2	Prostate cancer (somatic mutations)	N/A
CARS2	0.0002797	Cysteinyl-tRNA synthetase	Combined oxidative phosphorylation deficiency 27	N/A

Table 4-3 Variants identified in family ND5

The table summarises the three variants identified in family ND5 with CADD score above 15. The variants include a compound heterozygous variant (marked in blue) in EPHB2 (NM_004442.7) and a homozygous variant in CARS2 (NM_024537.4). Annotation was based on the human genome assembly GRCh37/hg19.

In addition to the standard pipeline, the family was also extensively assessed for any variants in *FAT4* based on their clinical presentation. The phenotype of this family is similar to Hennekam Lymphangiectasia-Lymphedema Syndrome 2 (MIM number: 616006), that is caused by variants in *FAT4*. No variants have been identified for this family, in the coding region or splice sites of *FAT4* to the extent allowed by WES. In addition, the possibility of *de novo* variants were also assessed and excluded as a likely pathogenic cause in this family. Furthermore, CNV analysis also excluded the presence of any potential CNVs that could be interpreted as pathogenic in this family (see section 2.2.10.7). These analyses provides more evidence that one the variants summarized in **table 4-3** is likely to be causative.

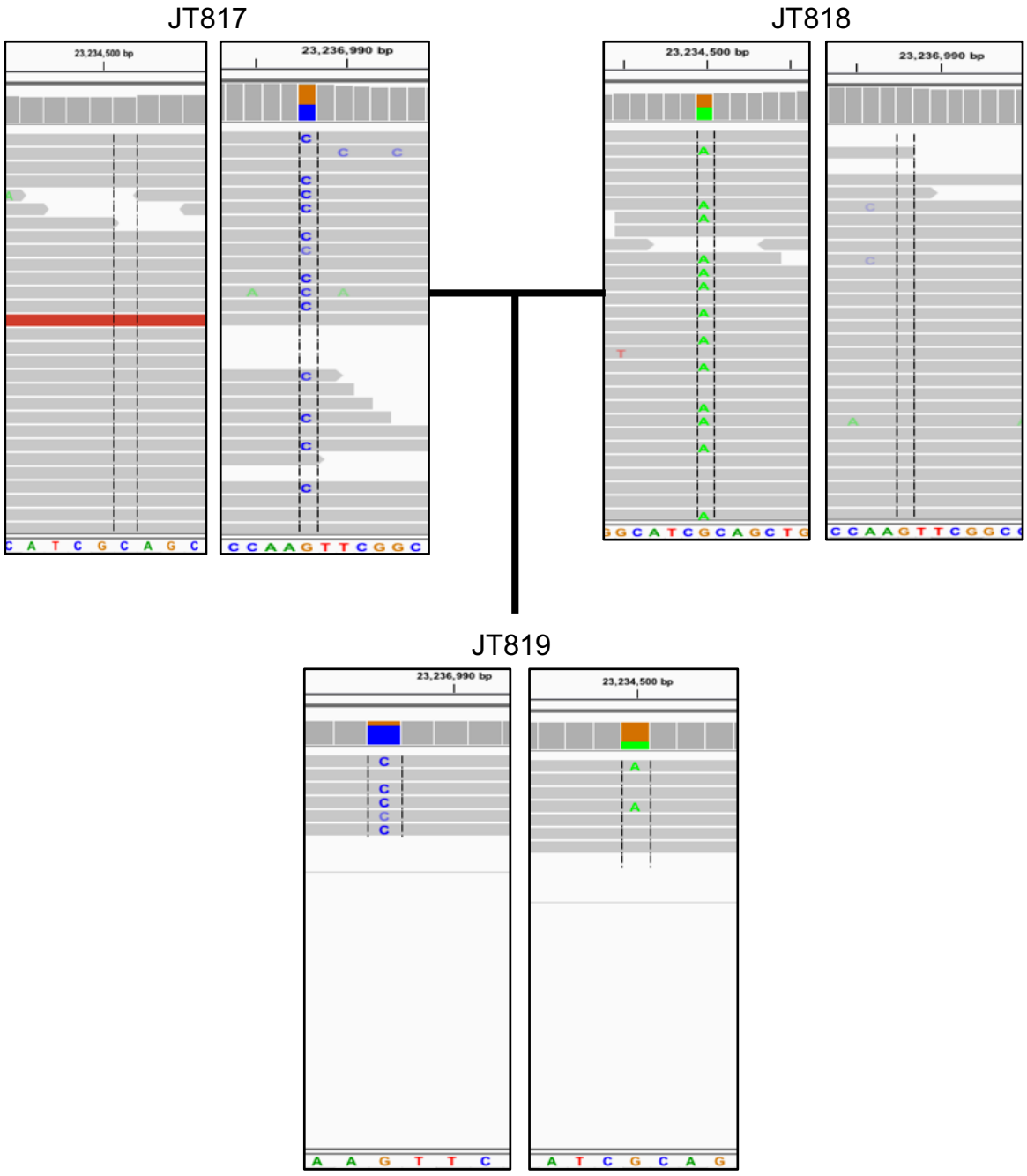


Figure 4-7 IGV analysis for *EPHB2* variant segregation

IGV analysis of the compound heterozygous variant in *EPHB2*. In blue letters, the variant c.2616G>C is shown, segregating from the mother and in green the variant c.2191G>A segregating from the father. Parents are shown here to be heterozygous for each variant and the child having a compound heterozygous variant in *EPHB2*.

Upon careful consideration of both variants and their established association with disease, the *in trans* compound heterozygous variants in *EPHB2* was considered the best candidate gene for this case. The homozygous variant in *CARS2* is predicted to be pathogenic, but variants in this genes have been associated with combined oxidative phosphorylation deficiency 27 (MIM number 616672) (Coughlin et al., 2015) which is phenotypically different to the clinical features of family ND5. This does not completely exclude the variant in *CARS2* as a candidate variant, but the compound heterozygous variant in *EPHB2* is a better candidate as the function of the EPHB2 protein fits the phenotype. Further functional characterization is still necessary to prove this hypothesis.

4.3.2.3 Variants in *EPHB2* as a novel cause of lymphoedema

EPHB2 belongs to the Eph receptor family of tyrosine kinase transmembrane glycoprotein receptors. These receptors bind ephrins by cell-cell contact (Jorgensen et al., 2009) and are involved in a range of cellular process, including axon guidance (Egea and Klein, 2007) and neural development (Klein, 2009). Somatic mutations in EPHB2 have been associated with prostate cancer (Huusko et al., 2004) and germline mutations are linked to a major platelet functional defect identified in just a single family (Berrou et al., 2018). However, heterozygous variants in *EPHB4* have been recently linked to autosomal dominant lymphatic-related (nonimmune) hydrops fetalis (LRHF) (Martin-Almedina et al., 2016). These variants seem to be devoid of the tyrosine kinase activity, suggesting that loss of ephrin signaling could contribute to LRHF pathogenesis.

Similarly, the mutations identified in family ND5 are both localized within the intracellular kinase domain of the EPHB2 receptor, potentially abolishing tyrosine kinase activity and any downstream ephrin-mediated signaling. Taking into account that EPHB4 receptor belongs to the same Eph receptor family, disruption of the EPHB2 receptor activity may lead to similar pathomechanisms. This hypothesis needs to be further investigated in order to assess whether or not the two compound heterozygous variants affect the function of the EPHB2 receptor. Initially, modelling the variants and measuring the kinase activity by determining the level of phosphorylation on EPHB2 kinase substrates could indicate whether activity is affected or not.

4.3.3 Family ND6

4.3.3.1 Clinical Phenotype

Family ND6 is a consanguineous South Asian family, with two affected sons presenting a myopathy phenotype, similar to limb-girdle muscular dystrophy. The two affected children presented with axial muscle weakness and mild peroneal nerve dysfunction. The muscle biopsy suggested core myopathy and leg muscle MRI scans showed abnormalities, principally STIR image changes in the rectus femoris muscle, and bilateral changes in the soleus muscles suggesting oedema. Neither parent was affected, suggesting a recessive mode of inheritance. DNA samples have been collected from all individuals presented in the pedigree and were analysed by WES.

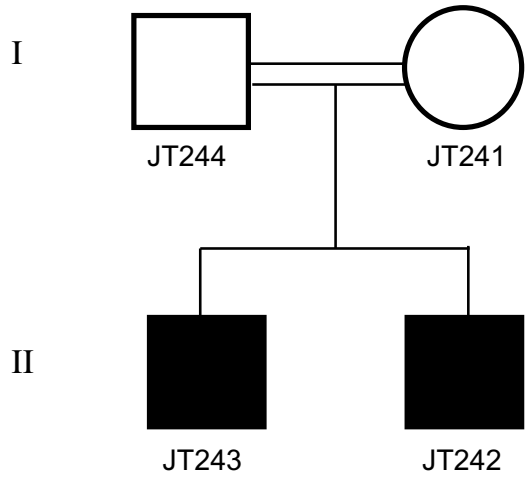


Figure 4-8 Pedigree of family ND6

Pedigree outlining two generations of the family and the consanguineous union. DNA samples of individuals with study identifiers (“JT numbers”) have been used for genetic investigations.

4.3.3.2 Autozygosity mapping

Autozygosity mapping was performed on WES data (vcf files prior to filtering of common SNPs) from all members of the family that had been taken forward for WES. Homozygous regions shared by the affected individuals, but absent from the parents are summarised in **table 4-4** and presented on the ideogram in **figure 4-8**.

Chromosome	Start	End	Length
5	19721490	102891695	83170205
2	176988290	225630435	48642145
6	62407067	102247673	39840606
12	32134638	47471439	15336801
2	272203	8943291	8671088
11	124253181	130784754	6531573
10	97127462	102506070	5378608
16	86565826	90127080	3561254
12	28605426	31244846	2639420
15	78632830	81212608	2579778
7	286402	1590443	1304041
9	131790531	132402908	612377
11	7022160	7509682	487522
5	149360877	149792337	431460
17	73125173	73552185	427012
6	30711805	31105891	394086
10	134981837	135351362	369525
11	63883985	64138905	254920
6	29691390	29798140	106750
16	1961674	2059674	98000

Table 4-4 Common regions of homozygosity amongst affected individuals of family ND6

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

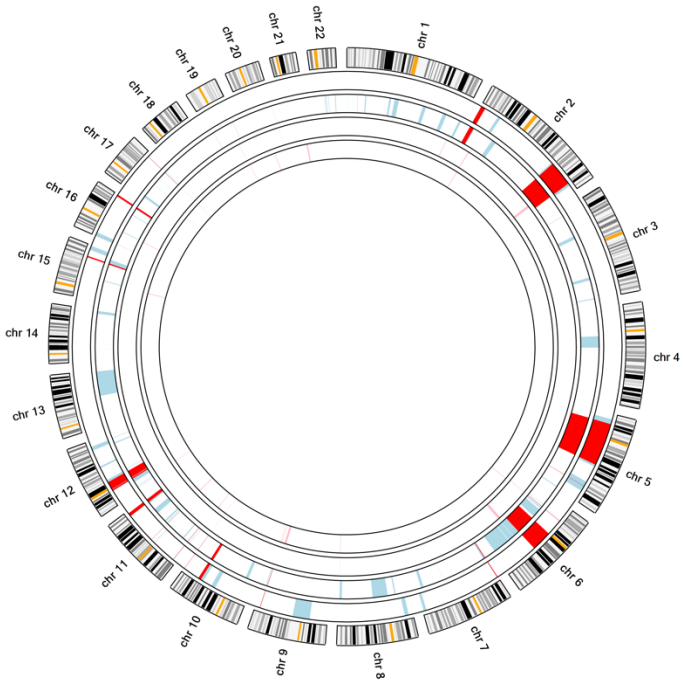


Figure 4-9 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND6.

The ideogram in light pink are from the unaffected individuals, whereas the outer ideograms in light blue are from the affected children. The regions highlighted in red represent the homozygous regions. The human genome assembly GRCh37/hg19 was used.

4.3.3.3 Candidate genes

An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance and prioritizing variants in the identified homozygous regions. In addition, segregation analysis was run, filtering out any variants that did not segregate with the family pedigree. **Table 4-5** summarises possible candidate variants identified with a CADD score above 15.

Variant Information				
Gene	**Location**	**Variant**	**Protein change**	
ADAMTS15	11:130343097	c.2234C>T	p.A745V	
TTN	2:179394686	c.101608+1C>T	splice	
PTH2R	2:209358262	c.1531G>A	p.E511K	
Pathogenicity prediction				
Gene	**Condel**	**Polyphen2**	**SIFT**	**CADD score**
ADAMTS15	deleterious (0.01)	benign (0.195)	tolerated (0.9)	35
TTN	N/A	N/A	N/A	25.7
PTH2R	tolerated (0.06)	benign (0.04)	neutral (0.325)	23.2
Frequency and function				
Gene	**MAF gnomAD**	**Protein Function**	**OMIM**	**In Homozygous region**
ADAMTS15	0.0001712	Zinc- dependent metallopeptidase	____	Yes
TTN	0.00000447	Abundant protein of striated muscle	Muscular dystrophy, limb-girdle, autosomal recessive 10	Yes
PTH2R	0.0009746	Parathyroid Hormone 2 Receptor	____	Yes

Table 4-5 Variants in candidates genes identified in family ND6.
Homozygous variants in *ADAMTS15* (NM_139055.3), *TTN* (NM_001267550.2) and *PTH2R* (NM_001309516.2) were annotated based on the human genome assembly GRCh37/hg19.

All the variants summarised above occur within a homozygous region shared by the affected individuals. The variant in *PTH2R* is predicted to be benign, and the known functions of the encoded protein do not make it a biologically plausible candidate for the phenotype of family ND6 (Dobolyi et al., 2012). *PTH2R* was therefore not prioritized in the list of candidate genes. On the other hand, both variants in *ADAMTS15* and *TTN* were predicted to be damaging and both proteins had biological relevance to the phenotype of the family. As a result, both variants were initially checked for segregation, and both variants were validated both using IGV Viewer and Sanger sequencing.

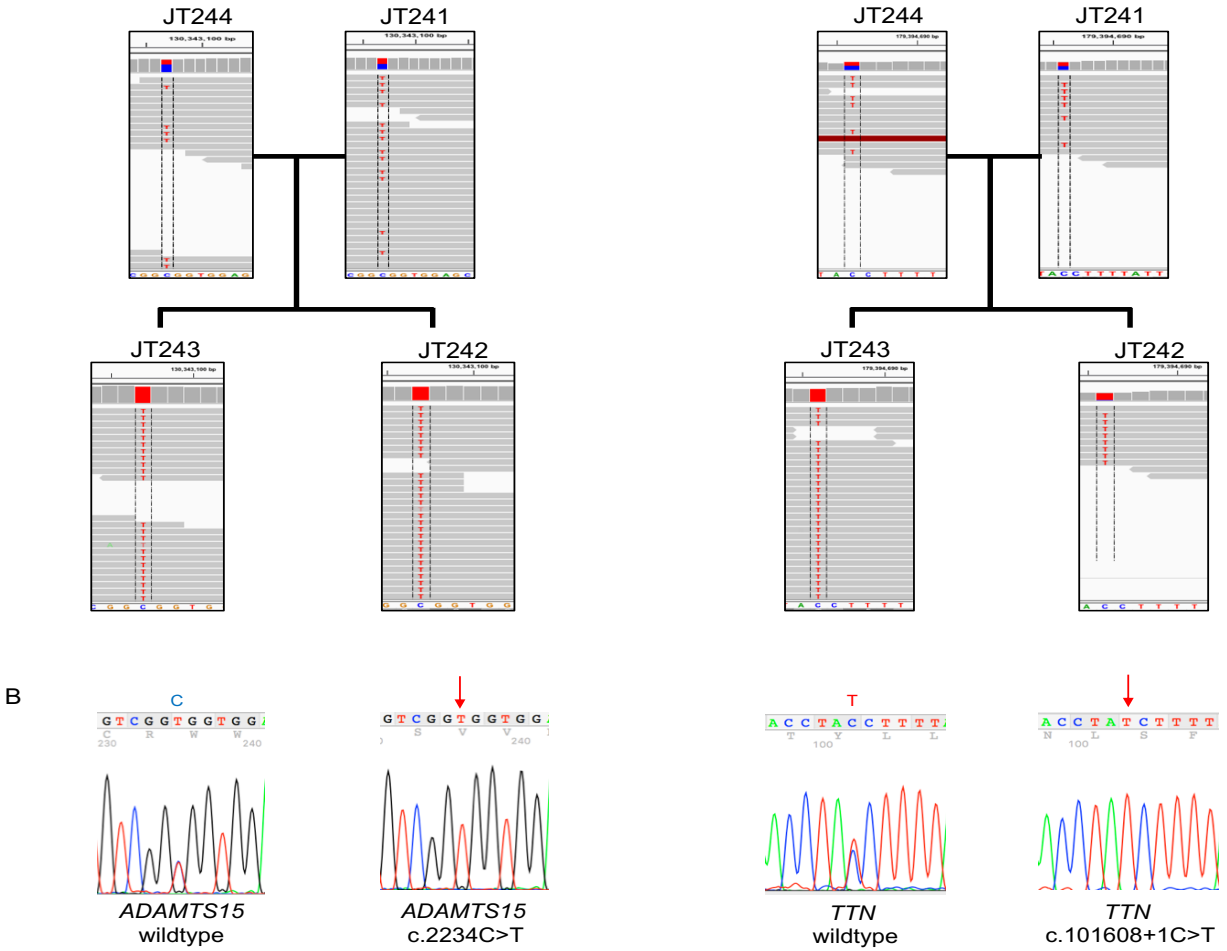


Figure 4-10 Segregation analysis of variants in *ADAMTS15* and *TTN* in family ND6

A – Segregation analysis using IGV viewer, showing the variants (in red) in *ADAMTS15* and *TTN* all members of the family (JT242 and JT243 affected individuals; JT241 and JT244 parents). B – Sanger sequencing confirming the variant in parental DNA (heterozygous traces) and patient DNA (homozygous trace).

The *ADAMTS15* missense variant has the highest CADD score and high pathogenicity predictions. The allele frequency of this variant in gnomAD is 0.0001712. In addition, the function of ADAMTS15 could potentially fit with the phenotype, as it has been reported to be involved in myoblast fusion during myogenesis (Stupka et al., 2013). *ADAMTS15* (a disintegrin and metallopeptidase with thrombospondin type 1 motif 15) belongs to the *ADAMTS* family of genes that code for proteinases that are implicated in a range of molecular processes, including myogenesis (Kelwick et al., 2015). During myogenesis, myoblasts are aligned within a versican-rich pericellular matrix, that is considered anti-adhesive due to versican properties. Enzymatic proteolysis of versican by *ADAMTS15* clears the pericellular matrix, allowing cell membrane contact between the denuded myoblasts, and thereby fusion of myoblasts to form multinucleated myotubes.

On the other hand, variants in *TTN* are mainly linked to cardiomyopathy (Itoh-Satoh et al., 2002).*TTN* is one of the largest genes in the human genome and carries a huge number of benign polymorphisms making variant interpretation challenging. *TTN* is a giant muscle protein spanning half of the sarcomere, from Z-disk to M-band (Savarese et al., 2016). *TTN* variants have been associated with several skeletal muscle diseases as well as cardiomyopathies. The *TTN* splice-site variant identified in family ND6 has a very low frequency in gnomAD (0.00000447) and is considered most likely pathogenic because of the potential effect on transcript splicing. Upon further literature review, the association of biallelic *TTN* variants and autosomal recessive myopathy similar to limb-girdle muscular dystrophy was established, sharing clinical features similar to family ND6 (Zheng et al., 2016).

At this point of the study, no further functional work was performed for this case. Both variants seem to be implicated with muscle development and function, and based on the pathogenic predictions none of the variants can be excluded. The *TTN* variant seems more likely to be the causative variant as it directly affects a splice-site and a previous association with biallelic variants in *TTN* and limb-girdle muscular dystrophy has been made. On the other hand, the *ADAMTS15* variant is a missense one that might not affect the function of the protein. This is worth further investigation. For instance, *in silico* protein modeling could provide insights into the effect of the *ADAMTS15* missense variant on predicted protein structure.

4.3.4 Family ND7

4.3.4.1 Clinical Phenotype

Family ND7 is a multiply consanguineous South Asian family with two affected individuals in two different sibships. The clinical phenotype shared by all affected individuals include congenital myopathy, severe developmental delay with diaphragmatic weakness, profound respiratory distress and generalized muscle weakness. Additionally, MRI scans of JT635 revealed frontal temporal atrophy. Clinical investigations included genetic testing by sanger sequencing for Prader Willi syndrome and for the following genes: *IGHMBP2*, SMN1, *CHRND*. DNA samples have been collected from all individuals with a JT number as indicated on the pedigree in **figure 4.10**, and processed for WES.

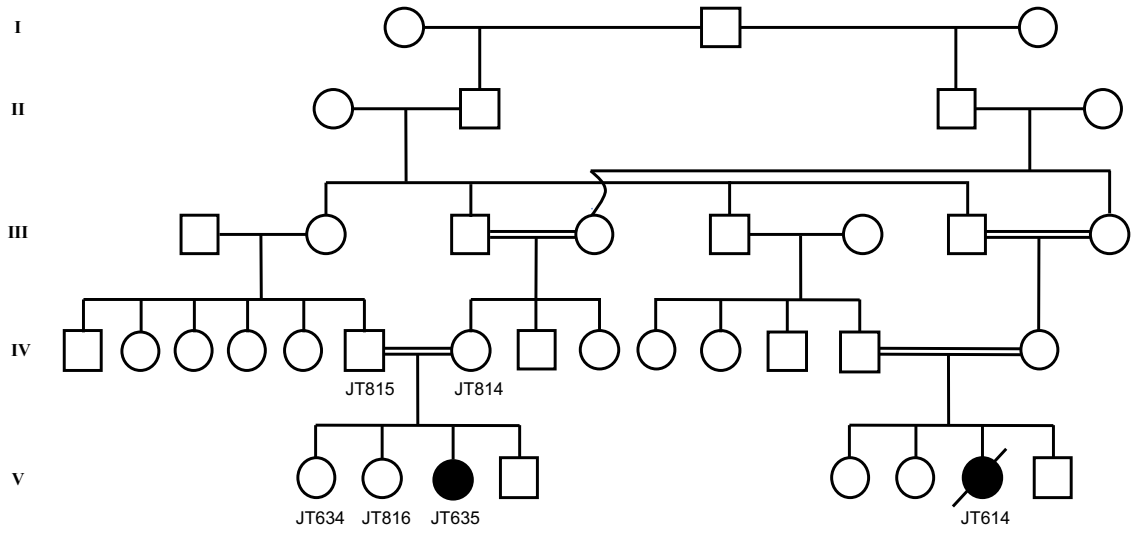


Figure 4-11 Pedigree of family ND6

Pedigree outlining five generations of the family and the consanguineous unions. DNA samples of individuals with JT numbers have been used for genetic investigations.

4.3.4.2 Autozygosity mapping

Autozygosity mapping was performed on WES data (vcf files prior to filtering of common SNPs) from all members of the family that were sequenced. Homozygous regions shared by the affected individuals (JT614, JT635), but absent from the parents (JT814, JT815) and unaffected siblings are summarized in **table 4-6** and presented on the ideogram in **figure 4-11**.

Chromosome	Start	End	Length
5	17630439	78154189	60523750
12	32716988	51752897	19035909
5	137521479	154395158	16873679
11	10655609	27077118	16421509
11	89133449	99428929	10295480
22	28316591	36661842	8345251
1	15764869	21583973	5819104
9	13183256	18907122	5723866
8	22785013	27645704	4860691
10	99991408	102771573	2780165
2	58276280	61008036	2731756
1	45205598	47904909	2699311
19	51217550	52917723	1700173
16	81639	1291608	1209969
20	60992224	62127584	1135360
1	1115994	1417696	301702
2	220356806	220494118	137312
6	29759066	29855706	96640

Table 4-6 Common regions of homozygosity amongst affected individuals of family ND6.
The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

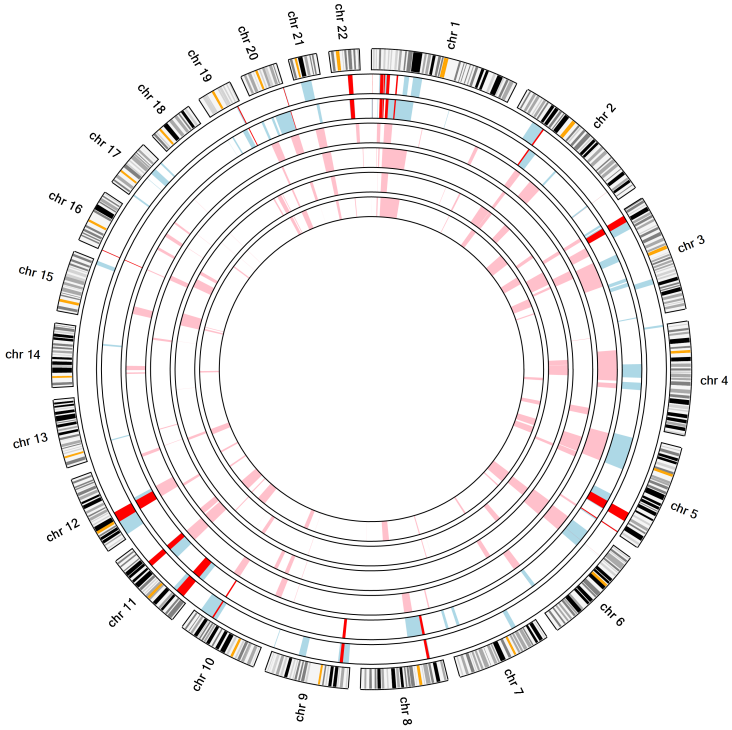


Figure 4-12 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND6.
The ideogram in light pink are from the unaffected individuals, whereas the outer ideograms in light blue are from the affected children. The regions highlighted in red represent the homozygous regions. The human genome assembly GRCh37/hg19 was used.

4.3.4.3 Candidate genes

Our in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance and prioritizing variants in previously identified homozygous regions. In addition, segregation analysis was run, filtering out any variants that did not segregate with the family pedigree. **Table 4-7** summarises possible candidate variants identified with a CADD score above 15.

Variant Information				
Gene	**Location**	**Variant**	**Protein change**	
NUP210	3:13407559	c.1819C>T	p.G607S	
TATDN2	3:10291099	c.215G>A	p.G72D	
Pathogenicity prediction				
Gene	**Condel**	**Polyphen2**	**SIFT**	**CADD score**
NUP210	neutral (0.018)	benign (0.195)	tolerated (0.9)	17.34
TATDN2	neutral (0.252)	benign (0.038)	tolerated_low_confidence (0.14)	16.72
Frequency and function				
Gene	**MAF gnomAD**	**Protein Function**	**OMIM**	**In Homozygous region**
NUP210	0.0009095 (3 homozygous South Asians)	Major component of the nuclear pore complex	____	Yes
TATDN2	0.0006097 (4 homozygous South Asians)	Putative deoxyribonuclease	____	Yes

Table 4-7 Variants in candidates genes identified in family ND7

Homozygous variants in *NUP210* (NM_024923.4) and *TATDN2* (NM_014760.4) were annotated based on the human genome assembly GRCh37/hg19.

The only variants that segregated with family pedigree are the variants summarised above in the *NUP210* and *TATDN2* genes. Both variants lie within homozygous regions, however the pathogenicity predictions are not convincing. CADD scores are just above the threshold of filtering and Condel, Polyphen2 and SIFT predict that both variants are benign. Furthermore, the variants *NUP210* c.1819C>T and *TATDN2* c.215G>A have been reported in the gnomAD database in homozygous state in South Asian populations, 3 and 4 times respectively. Based on the first output of the analysis, none of the variants identified appeared to be pathogenic and as a result we checked for compound hetergozygous variants, variants outside homozygous regions and CNV variants. The secondary analysis revealed a CNV variant that was compatible with the phenotype of family ND7.

4.3.4.4 Copy number variant in *MTMR2* identified using WES

The Exome Depth program was used in order to assess any CNVs (see section 2.2.10.7). The comparison of read depth was performed between the test sample (JT614) and 10 unrelated control samples. The analysis revealed a CNV in the *MTMR2* gene, deleting the last ten exons of the gene. Further details for the CNV are summarised in **table 4-8**, and **figure 4-13** illustrates the deletion of the ten exons in the affected individual compared to the other affected individual.

Type	Location	BF
Deletion	chr11:95568455-95591796	51.3
Gene	Number of exons	Exons
MTMR2	10	exons 6-15
Conrad.hg	Reads expected	Reads observed
N/A	573	1

Table 4-8 Details of CNV in *MTMR2* identified in family ND7

Summary of the *MTMR2* (NM_016156.6) deletion identified in patient JT614, including location of deletion, number of exons affected, reads expected and observed as well as Conrad.hg and Bayes Factor (BF). The human genome assembly GRCh37/hg19 was used.

The 23.3kb deletion at the end of *MTMR2* gene, removing the last ten exons of the gene, is predicted to either result in a truncated protein with approximately 65% of the protein missing, or lead to a complete loss of the protein by nonsense mediated decay. Either product is likely to have detrimental effect on the function of this protein.

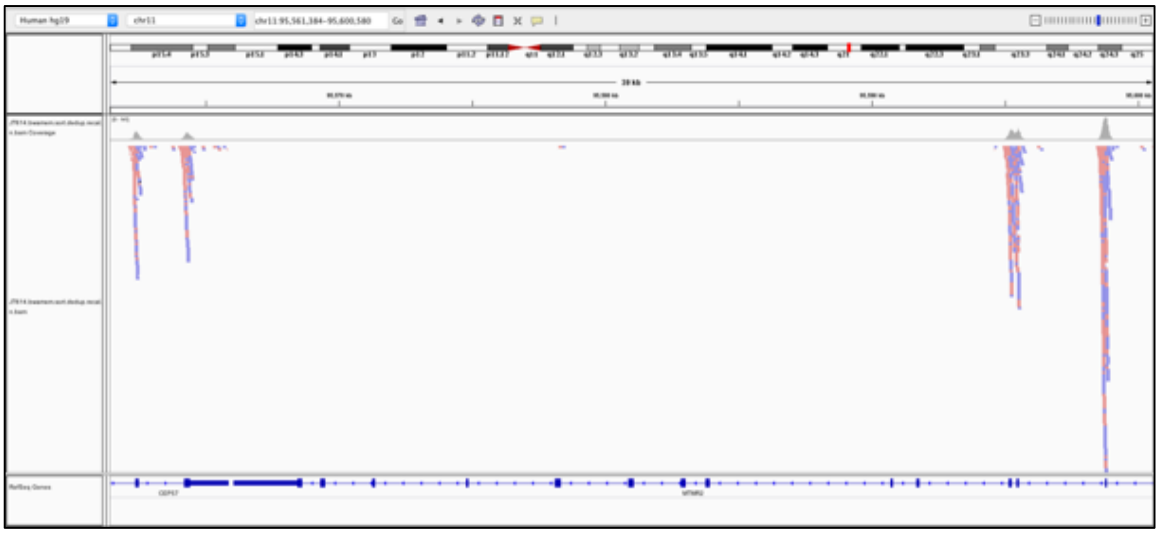
MTMR2 (myotubularin related protein 2) is a member of the myotubularin family of phosphoinositide lipid phosphatases. Berger et al. determined that phosphatidylinositol 3-phosphate (PI3P) and phosphatidylinositol 3,5-bisphophate (PI3,5P2) is dephosphorylated with high affinity by mouse Mtmr2 (Berger et al., 2002). Furthermore, MTMR2 is highly expressed in all cytoplasmic compartments of Schwann cells as well as sensory and motor neurons (Previtali et al., 2003).

Mutations in *MTMR2* have been associated with Charcot-Marie-Tooth disease, type 4B1 (MIM number 601382) (Bolino et al., 2000). The clinical features of Charcot-Marie-Tooth disease type 4B1 include peripheral motor and sensory neuropathy, muscle weakness and atrophy, delayed motor development and irregular loops and focal folding of myelin sheaths. This seems to fit with the phenotype of family ND7, however further detailed clinical assessment needs to be performed, particularly for the MRI scans, by the referring clinician.

The *MTMR2* deletion was also assessed for segregation in the other affected patient (JT635). Surprisingly, the deletion was not present in the other affected individual (see **figure 4.13**), despite this locus being identified within a homozygous region. A potential explanation for this could be due to the fact that the location of this deletion is at the end of a homozygous region identified on chromosome 11 and could potentially be falsely identified as homozygous. This could be explained as the computational tool that recognizes the regions uses common SNPs that are less accurate for identifying homozygous regions, compared to microsatellites.

In addition, this finding could be revealing a two-gene segregation within this family, with one sibship having the *MTMR2* deletion as the causative CNV for their phenotype, and the second sibship might have a different variant segregating that we failed to identified. The patient JT614 with the *MTMR2* deletion was deceased, however the research finding has been reported to the referring clinician for further clinical assessment of patient JT635, and comparison of her phenotype in relation to JT614. Further clinical investigation may be beneficial as they may have different features that will help us prove this hypothesis. Also, more analysis with adjusted pipelines can be used to investigate whether a separate variant is segregating. This observation highlights some of the weaknesses in the current pipelines and the underlying assumptions made during analysis, based on pedigree structures. In some cases, the patterns of inheritance might not be as obvious and more complex genetic inheritance might occur (Khan et al., 2013).

JT614



JT635

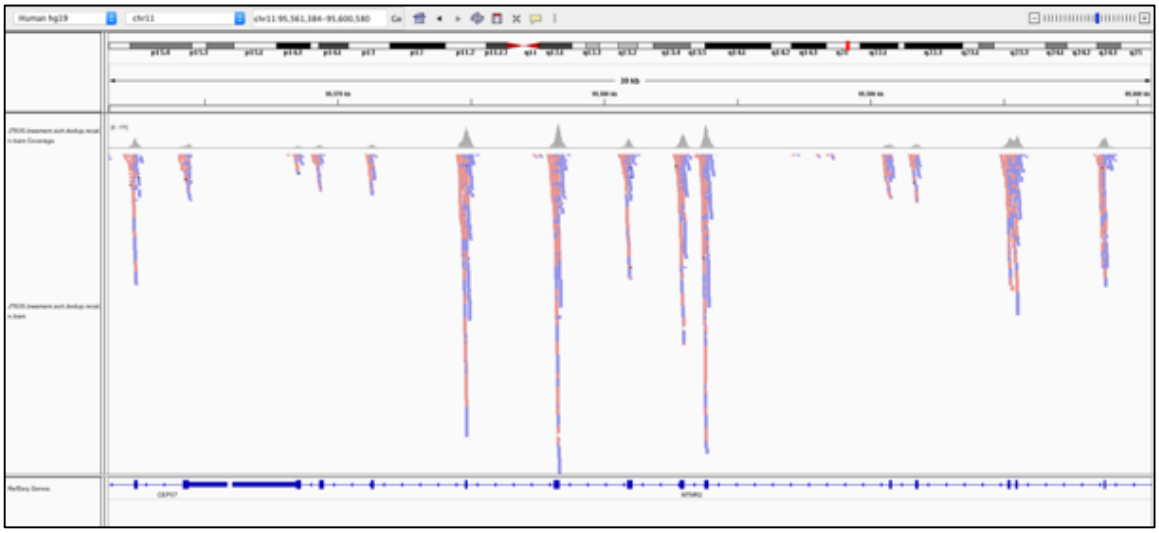


Figure 4-13 *MTMR2* segregation analysis

Segregation analysis of the *MTMR2* deletion using IGV. Top image shows the deletion of 10 exons in MTMR2 gene in JT614, as no reads are being identified within that region. Bottom image shows the absence of that deletion in JT635, as reads are being identified for each exon within that region.

4.3.5 Family ND8

4.3.5.1 Clinical Phenotype

Family ND8 is a multiply consanguineous South Asian family with affected individuals in different sibships as well as different generations. The affected individuals in generation VI presented in early childhood with delay in achieving weight-bearing posture and ability to walk independently. Individual VI-E below presented with a limping gait at age two, repeated falls and difficulty ascending the stairs at age 4. JT561 and JT562 below were reported to display similar difficulties. Individual IV-G in the pedigree below reported a similar history with increasing difficulty walking, and became wheelchair-bound in her mid 40s.

Examination of individual JT561 at age 6 revealed normal power and tone in all muscle groups except for ankle dorsiflexors. Reflexes were normal apart from absent ankle jerks, but sensation was normal in all modalities. Further investigations in patient IV-G and in JT561 revealed a demyelinating sensory and motor neuropathy.

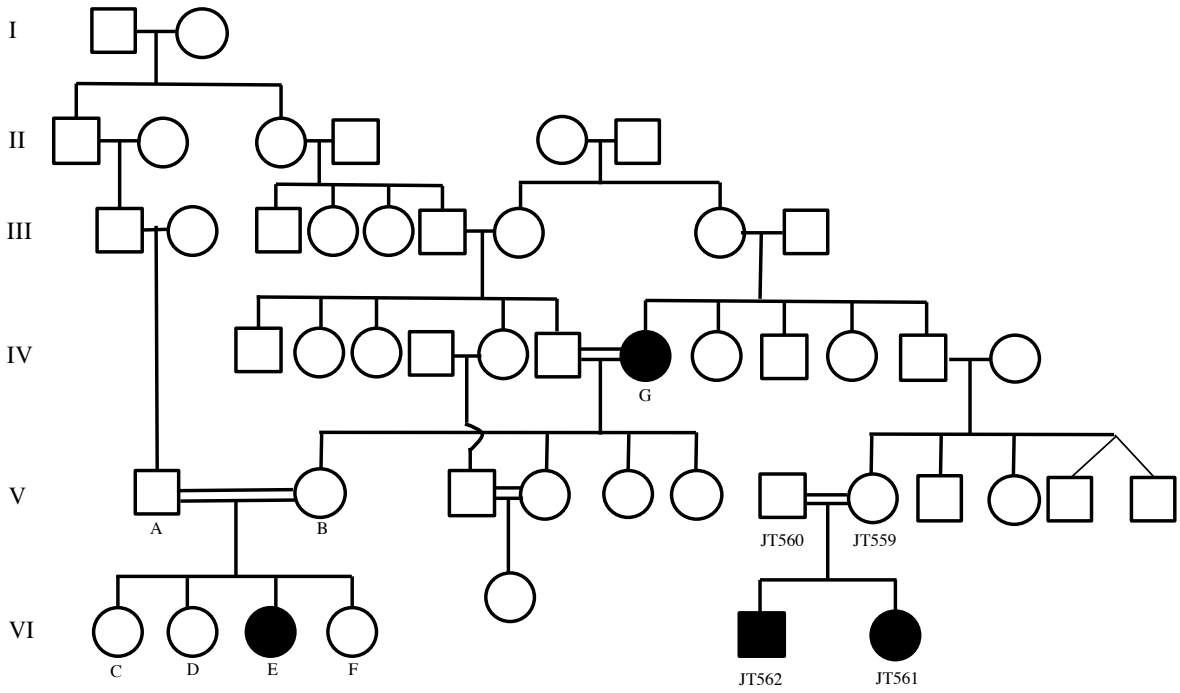


Figure 4-14 Pedigree of family ND8

Pedigree outlining six generations of the family and the consanguineous unions. DNA samples of individuals with JT numbers have been used for NGS genetic investigations. Samples from individuals A-G were acquired at a later stage for segregation purposes.

4.3.5.2 Autozygosity mapping

Autozygosity mapping was performed on WES data (vcf files prior to filtering of common SNPs) from all members of the family that have been taken forward for WES. Homozygous regions shared by the affected individuals (JT562, JT561), but absent from the parents are summarised in **table 4-9** and presented on the ideogram in **figure 4-15**.

Chromosome	Start	End	Length
11	82644904	95825797	13180893
13	19775755	30109939	10334184
6	161173946	170878886	9704940
4	55141055	62599289	7458234
4	178274565	185709891	7435326
8	119941173	125597409	5656236
15	39876495	45250756	5374261
12	122817507	125438549	2621042
9	128074807	130504070	2429263
9	122001000	124091135	2090135
11	65649774	67374581	1724807
8	145066886	145830894	764008
2	27163044	27857992	694948
7	149476666	150093701	617035
20	33867697	34457513	589816
19	3977486	4511140	533654
5	140167463	140517174	349711
10	134997480	135237196	239716
1	228399482	228560700	161218

Table 4-9 Common regions of homozygosity amongst affected individuals of family ND8.

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

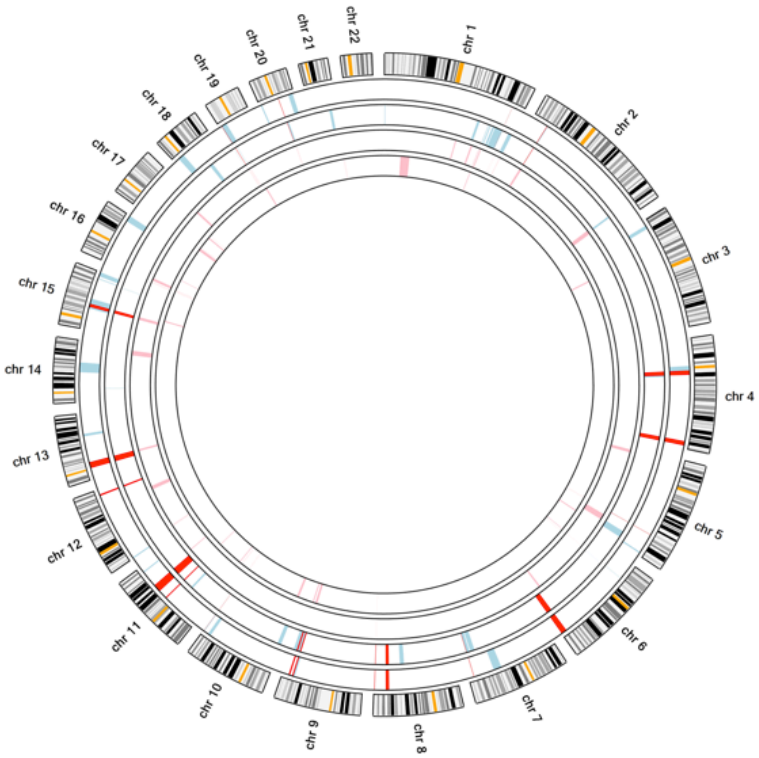


Figure 4-15 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND8.

The ideogram in light pink are from the parents, whereas the outer ideograms in light blue are from the affected children. The regions highlighted in red represent the homozygous regions. The human genome assembly GRCh37/hg19 was used.

4.3.5.3 Candidate genes

An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming a recessive mode of inheritance and prioritizing variants in previously identified homozygous regions. In addition, segregation analysis was run, filtering out any variants that did not segregate with the family pedigree. CNV analysis did not reveal any variants that could be interpreted as pathogenic. **Table 4-10** summarises possible candidate variants identified with a CADD score above 15.

Variant Information				
Gene	**Location**	**Variant**	**Protein change**	
SH3TC2	5:148418008	hom c.850_878dup	p.K284RP*	
RGS12	4:3422399	het c.2792G>A	p.R931H	
RGS12	4:3319069	het c.1172G>A	p.R391Q	
Pathogenicity prediction				
Gene	**Condel**	**Polyphen2**	**SIFT**	**CADD score**
SH3TC2	N/A	N/A	N/A	24.4
RGS12	deleterious (0.935)	probably_damaging (0.999)	deleterious (0)	29.7
RGS12	deleterious (0.646)	possibly_damaging (0.676)	deleterious (0.01)	21.9
Frequency and function				
Gene	**MAF gnomAD**	**Protein Function**	**OMIM**	**In Homozygous region**
SH3TC2	Absent	Exclusively expressed in Schwann cells and involved in endocytic membrane trafficking	Charcot-Marie-Tooth disease, type 4C	No
RGS12	0.00001997	Guanosine triphosphatase (GTPase)-activating protein as well as a transcriptional repressor.	N/A	N/A
RGS12	0.00006763			N/A

Table 4-10 Variants in candidate genes identified in family ND8.

Homozygous variant in *SH3TC2* (NM_024577.4) and compound heterozygous variant in *RGS12* (NM_198229.2) were annotated based on the human genome assembly GRCh37/hg19.

None of the variants that passed the filtering criteria were within a homozygous region. As a result, candidate genes based on pathogenicity predictions and biological relevance were prioritised. The best candidate gene was initially considered to be *SH3TC2*, as mutations in this gene are known to cause Charcot-Marie-Tooth (CMT) diseae type 4C (MIM number 608206) (Tazir et al., 2013), and the clinical features match perfectly with the phenotype of this family.

As *SH3TC2* was an ideal candidate variant for this case, Sanger sequencing analysis was performed to confirm segregation of the variant in family ND8. However, the variant did not segregate between the ascertained individuals that were sequenced. It was decided to include additional members of the family in this study. DNA was extracted

from saliva samples from individuals marked A-G, and segregation was assessed again using all eleven members of family ND8. The variant did not segregate in the family, assuming an autosomal recessive mode of inheritance with complete penetrance, because it was heterozygous in individual JT559 and all other individuals were homozygous for the duplication. To eliminate any source of uncertainty, new primers were designed surrounding the duplication, and the whole process was repeated twice, again failing to confirm segregation (**figure 4.16**).

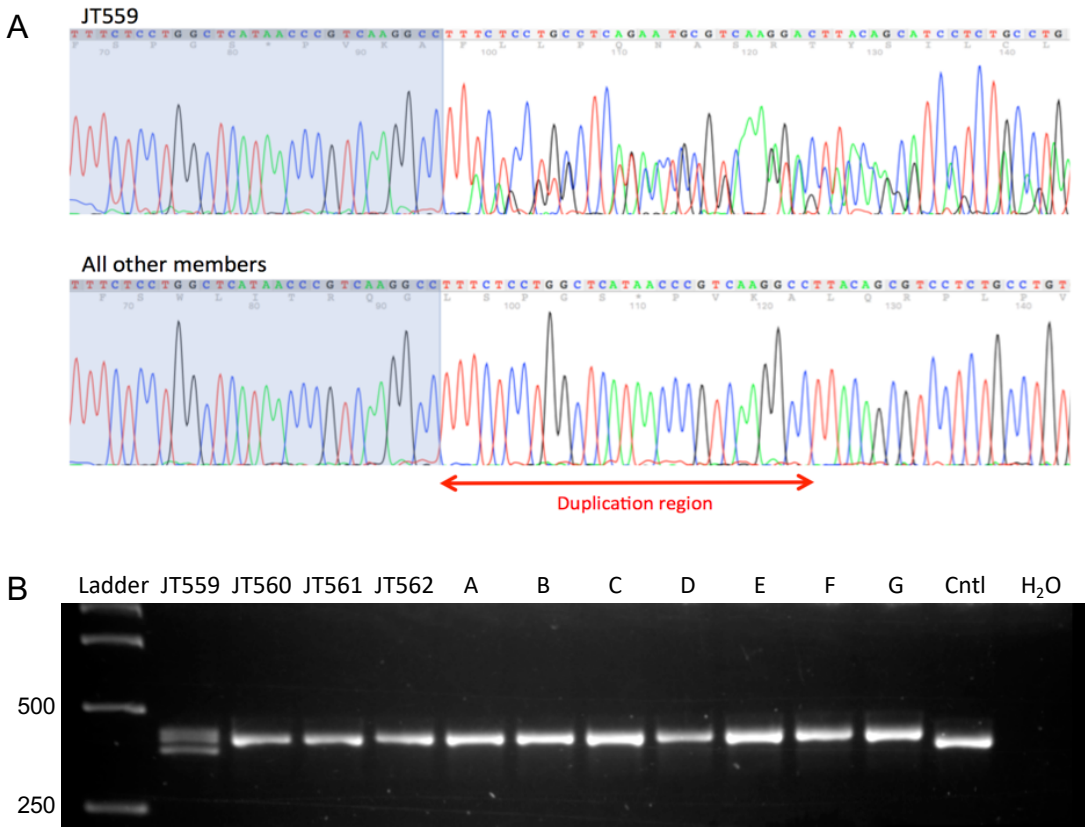


Figure 4-16 Segregation analysis for *SH3TC2* variant

Segregation analysis for duplication in SH3TC2. A – Sanger sequencing analysis showing the duplication region present in all members of the family (homozygous) apart from individual JT559 that was heterozygous for the duplication, shown by the overlapping electropherograms. B – Agarose gel image showing the size difference of the heterozygous duplication in JT599 and the homozygous state in all other members.

At this stage the duplication identified in SH3TC2 was considered non-pathogenic as healthy individuals of this family (individuals JT560, A to D, and F) were homozygous for this variant, thereby excluding it as a potential candidate gene for family ND8 despite the perfect match of SH3TC2-related phenotypes with that of family ND8.

The compound heterozygous variant in RGS12 was then assessed for segregation using IGV viewer (Thorvaldsdottir et al., 2013). Sanger sequencing was not performed due to the time constraints of the project, however IGV viewer output confirmed segregation of the compound heterozygous variants c.1172G>A and c.2792G>A, carried in *trans* for affected individuals. Segregation analysis is summarised in **figure 4.17**.

Figure 4-17 Segregation analysis for *RGS12* variant.

Segregation analysis for the compound heterozygous variant in *RGS12* using IGV viewer. Individuals JT559 and JT560 carry only one of the two variants (parents), whereas individuals JT561 and JT562 carry both variants in *RGS12* (affected).

RGS12 is a member of the R12 subfamily of RGS proteins, including RGS10 and RGS14, all involved in numerous signaling pathways. RGS12 functions as a guanosine triphosphatase (GTPase) activating protein, regulating hormone release or neurotransmitter signaling (Hepler, 1999). RGS12 is highly expressed in the human brain and myometrium (Chatterjee and Fisher, 2000). It also seems to play a role in mammalian myogenesis and neurogenesis, as it is highly expressed in dorsal root ganglia (DRG) and trigeminal neurons, as well as muscle, in the E14.5 embryonic mouse (Martin-McCaffrey et al., 2005). *RGS12* could be a good candidate gene, but further functional work will be required to prove the potential pathomechanism it is involved in.

4.4 A candidate gene identified from analysis of *de novo* variants

4.4.1 Family ND9

4.4.1.1 Clinical Phenotype

Family ND9 is a non-consanguineous family, with no previous family history. Individual JT753 presented symptoms at very early age. He had partial left third nerve palsy and left sixth nerve palsy, as well as right sided ptosis and lower limb dystonia. It was reported that he has unusual movements of his legs and had hyperekplexia in both lower limbs. He also presented ocular motility problems. The clinical diagnosis was one of Complex Moebius syndrome. MRI scans showed right perisylvian polymicrogyria and slight callosal dysmorphism. **Figure 4-18** oultines the family pedigree and **figure 4-19** shows sections of the MRI scans.

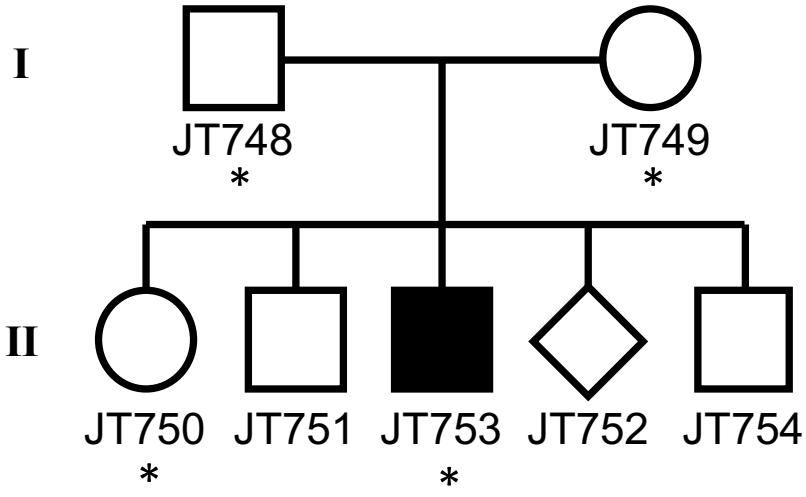


Figure 4-18 Pedigree of family ND9

Pedigree outlining two generations of the family ND9. DNA samples of individuals with JT numbers were available, but only individuals marked with an asterisks (*) have been used for NGS genetic investigations.

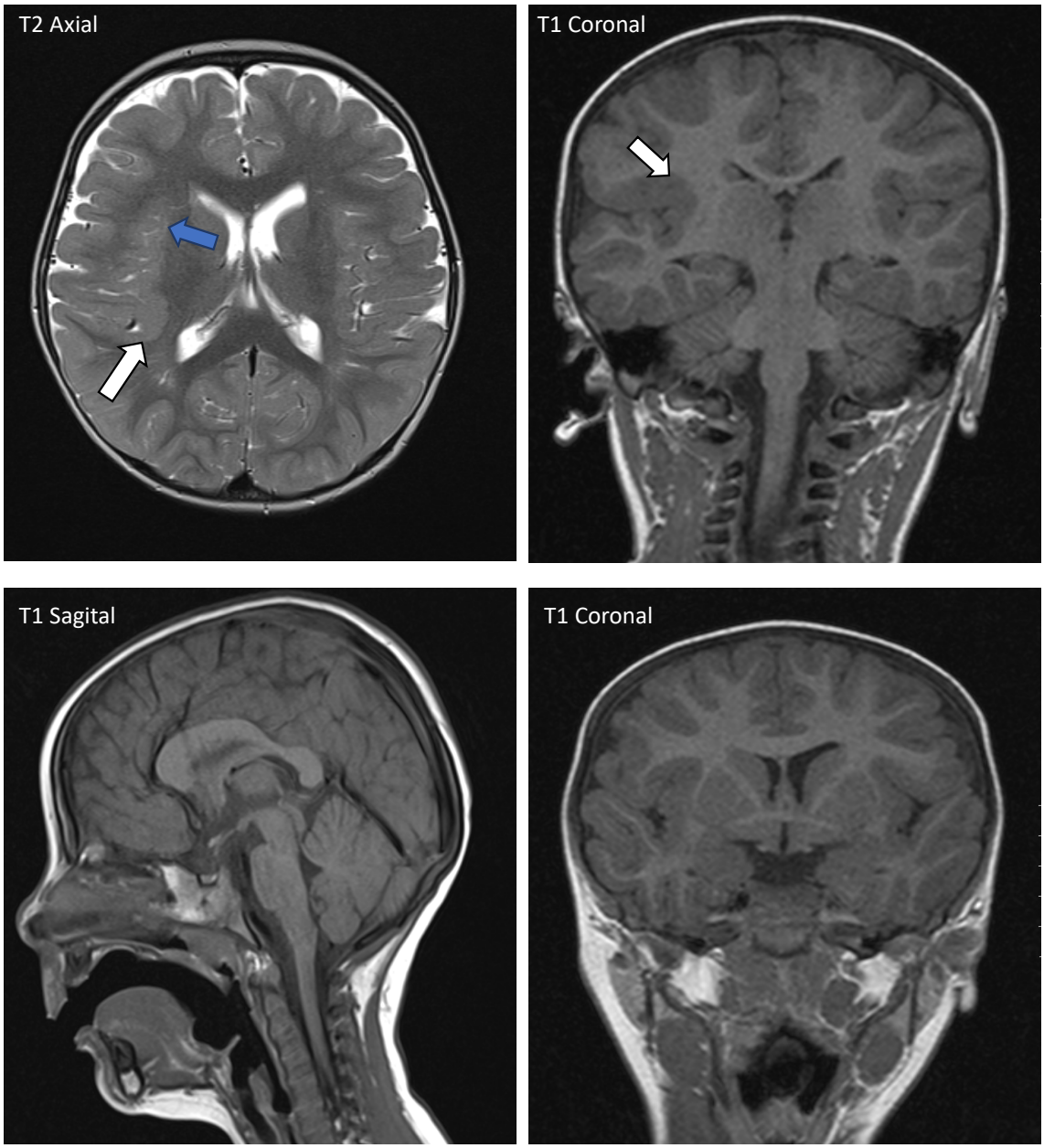


Figure 4-19 MRI scans from individual JT753.

MRI scans of individual JT753 showing brain malformations. Right perisylvian polymicrogyria as demonstrated by the indistinct grey/white matter interface; (Blue arrow). Slight callosal dysmorphism, with a bulky anterior limb and constricted splenium; (White arrows).

4.4.1.2 Candidate gene

An in house bioinformatics pipeline was used for data analysis (section 2.2.10.3), in the first instance, assuming *de novo* incidence as only a single individual was affected in the family and there was no history of consanguinity. Analysis for a recessive mode of inheritance was also performed, identifying both homozygous and compound heterozygous variants. As there was no history of consanguinity, autozygosity mapping was not performed for this family. Filtering of biallelic variants did not reveal any potential

pathogenic variant, however an interesting *de novo* variant was identified. **Table 4-11** summarizes the *de novo* variants identified in family ND9 with a CADD score above 15.

Furthermore, CNV analysis was also carried out for this family (see section 2.2.10.7) to identify potential CNVs that could cause the phenotype. CNV analysis did not reveal any variants that could be interpreted as pathogenic, increasing our confidence in the *TUBA1A* variant summarised in **table 4-11**.

Variant Information			
Gene	**Location**	**Variant**	**Protein change**
TUBA1A	12:49579706	c.443C>T	p.G148E
Pathogenicity prediction			
Condel	**Polyphen2**	**SIFT**	**CADD score**
deleterious (0.873)	probably_damaging (0.983)	deleterious_low_confidence (0)	26.8
Frequency and function			
MAF gnomAD	**Protein Function**	**OMIM**	**In Homozygous region**
Absent	Major components of microtubules	Lissencephaly 3	N/A

Table 4-11 De novo variant identified in family ND9.
De novo variant in *TUBA1A* (NM_006009.4) was annotated based on the human genome assembly GRCh37/hg19.

4.4.1.3 Novel *de novo* variant in *TUBA1A*

Alpha and beta tubulins form the heterodimers that comprise the major components of the microtubules, a major constituent of the cytoskeleton. Microtubules are essential for various cellular processes such as mitosis and intracellular transport (Forth and Kapoor, 2017). Nine α-tubulin and nine β-tubulin isotypes have been identified (Gadadhar et al., 2017) with tissue specific expression variability. Tubulin genes are also known to play a key role in central nervous system development, particularly in axonal guidance and neuronal migration, and mutations in various α- or β-tubulins genes have linked to developmental disorders with brain malformations. These malformations include defects in cortical development, polymicrogyria, gyral disorganisation, and agenesis or abnormalities of the midline commissural structures, such as the corpus callosum. Mutated tubulin genes known to cause these disorders include *TUBA1A*, *TUBA8*, *TUBB2A*, *TUBB4A*, *TUBB*, *TUBB2B* and *TUBB3* (Romaniello, 2015).

TUBA1A is a highly-conserved gene that expresses predominantly in the developing brain, and decreases in postnatal and adult periods (Gardner et al., 2018). Heterozygous mutations in *TUBA1A* have been mainly associated with lissencephaly type 3 (MIM number: 611603). Most of the cases reported consist of patients with *de novo* variants in *TUBA1A* (Keays et al., 2007), but a hereditary form has been identified where in two sisters who inherited a heterozygous *TUBA1A* variant from their mother, who had somatic mutation in the gene (Jansen et al., 2011).

Recent genetic studies have expanded the clinical spectrum of *TUBA1A*-associated disorders, with cases presenting overlapping brain malformations but not necessarily lissencephaly. An interesting study reporting a *de novo TUBA1A* p.R2H variant identified in four unrelated patients but without presenting lissenchephaly. The patients described in this study had mild phenotypic variability, but shared phenotypes included developmental delay, microcephaly, dysplasia or thinning of the corpus callosum, dysmorphic basal ganglia and hypoplasia of the cerebellar vermis (Gardner et al., 2018). Two of the patients also had bilateral perisylvian polymicrogyria, resembling the phenotypes observed in family ND9.

Another recently published study describes a different *de novo* variant in *TUBA1A* (p.D127E) but the described patient does not have lissenchephaly. Clinical features of this case include motor delay, occipital polymicrogyria, hypoplasia of the corpus callosum, cerebellar hypoplasia, dysplastic brainstem, on-separative basal ganglia and volume loss of the white matter (Sato et al., 2018b).

Sato *et al.* outline two variants (p.R2H and p.D127E) that appear to have a minimal change of amino acid residue in terms of physicochemical properties, yet sufficient to cause a pathogenic phenotype. It is important to appreciate that a highly conserved protein such as TUBA1A has critical functions in cell cycle and maintenance that could be disrupted even with subtle changes of amino acid residues. Such changes could possibly affect the way in which α- and β-tubulins form heterodimers and polymerise to form the microtubule network. Disruption in the microtubule formation or stability could potentially affect downstream cellular mechanisms that are key to development. It is essential to establish a new genotype-phenotype correlation with variants in *TUBA1A*, as many recent studies are identifying variants that cause a wide range of brain malformations that may or may not include lissencephaly.

4.5 Unsolved cases with no variants that are interpreted as pathogenic in the coding DNA sequence.

This section summarizes the genetic investigation of two consanguineous families, ND10 and ND11, for whom WES was unable to detect any potentially pathogenic variants. More precisely, the same in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3) assuming an autosomal recessive mode of inheritance and prioritizing variants in previously identified homozygous regions. This first line of analysis revealed no homozygous variants above CADD score of 15. There were no biallelic variants outside of the homozygous regions that could be interpreted as pathogenic. In addition, CNV analysis excluded the involvement of potentially pathogenic CNVs as the cause of disease in families ND10 and ND11.

The sections below summarize the clinical phenotype of the two cases, along with the family pedigree and discuss further options of investigating the genetic basis of these diseases, such as whole genome sequencing and/or RNAseq.

4.5.1 Family ND10

4.5.1.1 Clinical Phenotype

Family ND10 is a consanguineous South Asian family, with no previous family history. Parents of the two affected individuals are first cousins. The affected children presented global developmental delay, learning difficulties and delayed speech. They were reported to have delayed language development and unsteady gait.

Diagnostic array CGH revealed no pathogenic variation, and a gene panel for early onset epileptic encephalopathy was normal (https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/epileptic-encephalopathies-early-infantile-72-gene-panel-670/), and as a result this family was investigated further using NGS.

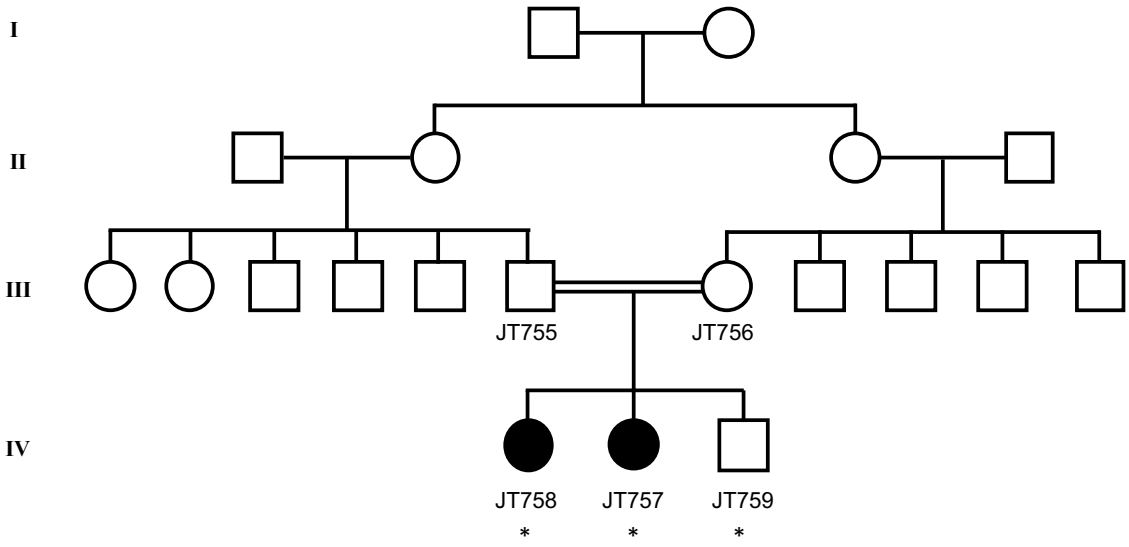


Figure 4-20 Pedigree of family ND10

Pedigree outlining four generations of the family ND10, and the consanguineous union between first cousins. DNA samples of individuals with study identifiers ("JT numbers") were available, but only individuals marked with an asterisks (*) have been used for NGS genetic investigations.

4.5.2 Family ND11

4.5.2.1 Clinical phenotype

Family ND10 is a consanguineous South Asian family, with all affected members of this family presented with a consistent clinical phenotype of four limb spasticity, and marked developmental delay, with very limited developmental progress made. No children developed speech or independent ambulation. All remained dependent on parents for simple activities of daily living including, washing, feeding, toileting. All had a consistent MRI phenotype of progressive frontal atrophy, but with very little in the way of localising features. White matter appearances were normal. aCGH and a gene panel for cerebral malformations were normal (https://www.leedsth.nhs.uk/a-z-of-services/the-leeds-genetics-laboratory/constitutional-genetics/molecular-genetics/by-disorder/cerebral-malformations/).

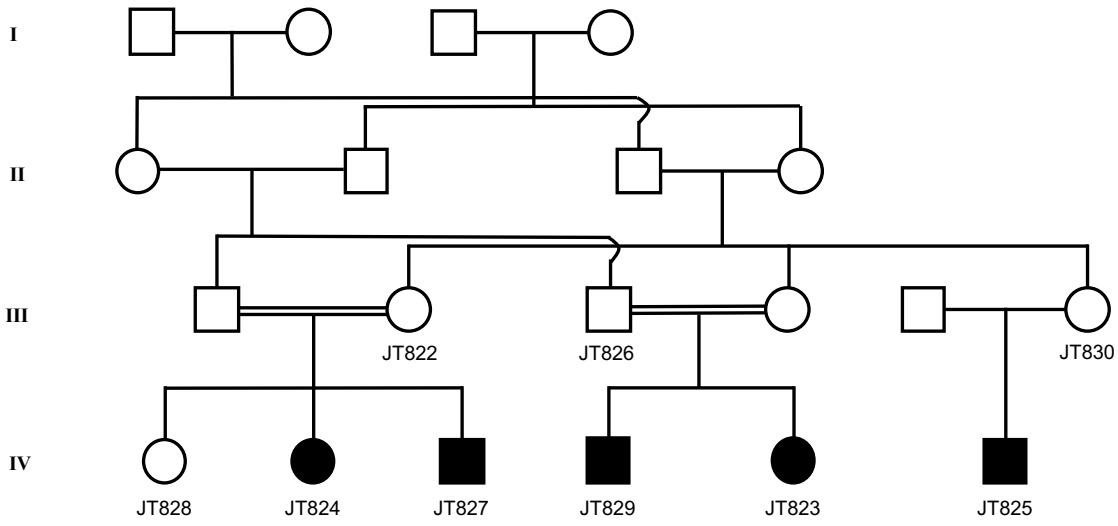


Figure 4-21 Pedigree of family ND11

Pedigree outlining four generations of the family ND11, and the consanguineous unions. DNA samples of individuals with study identifiers ("JT numbers") were used for NGS genetic investigations.

4.5.3 Genetic Findings

Both of the families ND10 and ND11 were considered unsolved at this stage of the project, as no variants have been identified that would segregate with family pedigree. Both homozygous and compound heterozygous variants were checked, along with autozygosity mapping, but no variant passed the filtering strategy. In addition, CNV analysis also revealed no potential variant that could be interpreted as pathogenic, thereby classifying these cases as unsolved.

4.5.4 Further genetic investigations of the unsolved cases

These unsolved cases have been taken forward by a different study for further genetic investigations, where a combined approach using WGS and RNA sequencing will be performed in an attempt to identify the genetic basis of these diseases. WGS will be initially carried out to check for deep intronic variants or complex CNVs that WES did not capture. Alongside WGS, RNA sequencing will also be used to assess whether any intronic variants identified might disrupt the splicing or expression of a particular gene, thereby affecting downstream function of the protein. RNA sequencing will be performed by extracting RNA from patient cells. Increasing numbers of Mendelian disorders that were not be solved using WES, are now being solved using RNAseq, with analyses revealing cryptic splice sites, private exons, or aberrant splicing events that are proving to be pathogenic (Kremer et al., 2017).

Chapter 5

A novel homozygous null variant in *KLHL7* as a cause of a recessive neurodevelopmental condition

5.1 Introduction

This chapter summarises the identification of a null variant in *KLHL7* that is associated with an autosomal recessive neurodevelopmental condition in a single consanguineous family. The chapter outlines library preparation for whole exome sequencing of individuals in this family, the follow-up functional characterization of the *KLHL7* variant and the association of the encoded protein with neurodevelopment. A clinical overview and variant filtering strategy is given, in the same format as discussed previously, and the candidate variant is then functionally assessed using patient-derived dermal fibroblasts and cell-lines.

5.2 Variant identification by whole exome sequencing

5.2.1 Patient Recruitment

Family ND12 was recruited and processed as part of the Jules Thorn Biomedical Award study for molecular genetic research. All family members that were included in the study gave informed consent for research and their genomic DNA was collected from blood by using standard methods. Skin biopsy was also collected from affected individual JT728 (see method 2.2.4). Affected individuals were phenotyped by Clinical Geneticists in regional hospitals (see section 2.2.1).

5.2.2 Clinical Ascertainment

Family ND12 is a consanguineous South Asian family with three affected sons with profound developmental delay. In these children the initial assessment revealed abnormal movements, and a suspected white matter abnormality. The boys also presented with hypogenitalism raising the suspicion of hypogonadism. This led to an initial clinical diagnosis of 4H Syndrome (MIM number: 607694). This is a childhood onset autosomal recessive neurodegenerative disorder. It is characterized by progressive motor decline that can manifest as tremor, ataxia, spasticity, and cerebellar signs, as well as mild cognitive regression (Wolf et al., 2007). Other features may include hypogonadotropic hypogonadism and hypodontia or oligodontia. Causative variants

have been reported in *POLR3A* and *POLR3B* (Cayami et al., 2015). Clinical investigations included a focused clinical exome for Hypomyelinating Leukodystrophy, testing for 12 genes, including *POLR3A* and *POLR3B* (https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/details/6736/).

Later review revealed that the children's symptoms included abnormal movement, ptyalism, failure to thrive, myopathy, developmental delay, poor feeding and evidence of and stroke-like episodes.. They also had abnormal brain scans, demonstrating hypomyelination and microcephaly. Excessive sweating was a distinct feature noted for the younger boy once he was admitted to hospital. Further examinations revealed elevated blood lactate levels, with normal CSF lactate, and a skeletal muscle biopsy showed a severe combined deficiency of mitochondrial respiratory chain complexes I and IV.

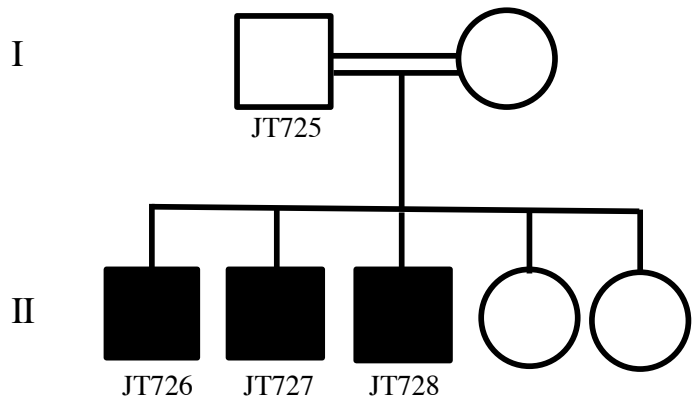


Figure 5-1 Pedigree of family ND12

Pedigree outlining two generations of the family and the consanguineous union. DNA samples indicated by anonymized codes have been used for genetic investigations using WES.

5.2.3 Autozygosity mapping

Autozygosity mapping was performed using WES data. Precisely, the vcf file prior to SNP filtering (.raw.vcf) from the four members of the family that have been taken forward for WES were used (see section 2.2.10.6). Homozygous regions shared by the three affected individuals, but absent from the parents and the unaffected sibling, are summarised in **Table 5-1**.

Chromosome	Start	End	Length
7	16834551	24758645	7924094
10	93945	3182941	3088996
9	131689361	132625602	936241
17	73089524	73565262	475738
17	40811781	41167957	356176

Table 5-1 Common regions of homozygosity amongst affected individuals.

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

5.2.4 Whole exome sequencing

DNA samples from all the individuals with a JT identifier (**Figure 5.1**) were taken forward for WES library preparation using the Illumina QXT protocol (section 2.2.10). During library preparation, Bioanalyzer traces were used to assess the size and quantity of the DNA fragments at post-shearing and post-hybridization stages (Appendix D). The samples were pooled, aiming for a total of ten samples per flowcell, and sequencing was performed on the Illumina HiSeq 3000 platform.

Prior to data analysis, the .fastq files, and subsequent .bam files, were also assessed for the quality of sequencing (section 2.2.10.4). An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3), assuming a recessive mode of inheritance that allowed variants to be prioritized in the shared homozygous regions identified previously. In addition, segregation analysis enabled the filtering out of variants that were not compatible with Mendelian segregation of an autosomal recessive condition. Investigation of any variants in either *POLR3A* and *POLR3B* was also carried out, based on the 4H-like phenotype of the family, with no homozygous or compound heterozygous variants being identified, confirming the previous clinical investigations. **Table 5-2** summarises the variants identified after filtering with a CADD score >15. **Figure 5-2** illustrates the presence of the best candidate variant in *KLHL7* in the homozygous region previously identified. Given the pedigree structure, further analysis for variants associated with X-linked inheritance was also performed but no variants were identified.

Variant Information				
Gene	**Location**	**Variant**	**Protein change**	
KLHL7	7:23205327	c.947G>A	p.W316*	
MAML3	4:140811083	c.1506delG	p.Q502HfsTer20	
Pathogenicity prediction				
Gene	**Condel**	**Polyphen2**	**SIFT**	**CADD score**
KLHL7	N/A	N/A	N/A	39
MAML3	N/A	N/A	N/A	26
Frequency and function				
Gene	**MAF gnomAD**	**Protein Function**	**OMIM**	**In Homozygous**
KLHL7	0.000003978	Substrate-specific adapter of E3 ubiquitin ligases	Retinitis pigmentosa 42 (AD) Cold-induced sweating syndrome 3 (AR)	Yes
MAML3	0.7481	Transcriptional coactivator for NOTCH proteins	N/A	No

Table 5-2 Homozygous variants identified in family ND12

Homozygous variants in *KLHL7* (NM_001031710.3) and *MAML3* (NM_018717.5) were annotated based on the human genome assembly GRCh37/hg19.

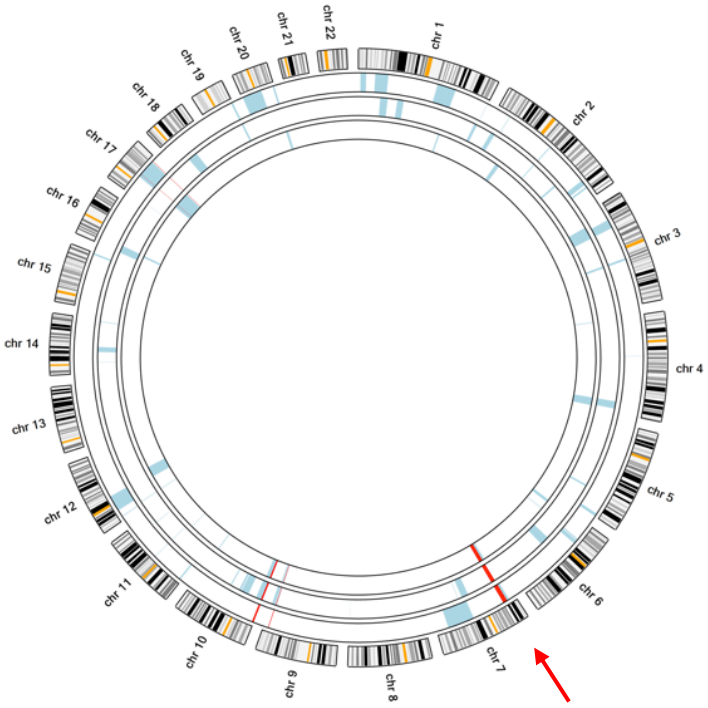


Figure 5-2 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND12

The ideogram illustrates the homozygous regions of the three affected individuals. The regions highlighted in red represent the shared homozygous regions. The homozygous *KLHL7* variant lies within chromosome 7 (7p15.3), indicated by the red arrow. The human genome assembly GRCh37/hg19 was used.

5.2.5 Variant confirmation using Sanger Sequencing

Due to the presence of consanguinity in family ND12, the variants in the homozygous regions were initially prioritized. In this case, the variant in *KLHL7* (c.947G>A, p.W316*) was considered the best candidate variant as it was within a homozygous region, had a higher CADD score compared to the variant in *MAML3*, and the minor allele frequency on gnomAD was very low, compatible with the frequency of a pathogenic rare disease allele. The allele frequency of *MAML3* was high and was therefore excluded as pathogenic and causative.

Sanger sequencing was used to confirm that the c. 947G>A variant in *KLHL7* identified by WES was in the homozygous state in all affected individuals, and heterozygous in the unaffected father. DNA from all members of the family was amplified and Sanger sequenced using exon specific primers (see section 2.1.4). The c.947G>A variant in *KLHL7* segregated with the condition, consistent with autosomal recessive inheritance. This research finding was reported back to the referring clinician. The electropherograms that summarise this research finding are shown in **Figure 5-3**.

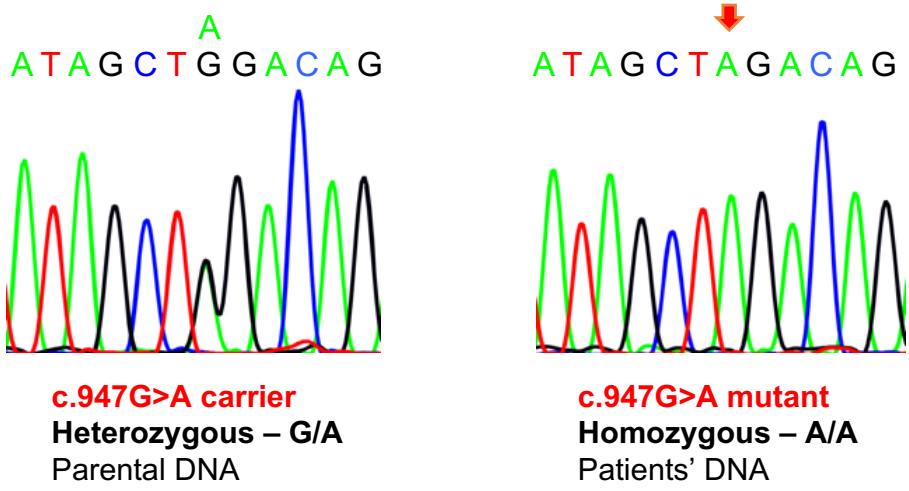


Figure 5-3 Electropherograms of the c.947G>A variant in *KLHL7*.

Electropherograms illustrating heterozygous sequence from parental DNA and the homozygous mutant sequence from the patients DNA. The red arrow indicates the changed nucleotide in the patients DNA. All affected individuals are homozygous for this variant.

5.3 Interpretation of the c.947G>A variant in *KLHL7*

5.3.1 Conservation and expression pattern of *KLHL7*

Conservation analysis revealed that the amino acid residue p.W316 and this region of the KLHL7 protein are highly conserved in vertebrate species from Humans down to frog and zebrafish (**Figure 5-4**). This suggests that the specific amino acid, as well as the region surrounding this residue, is likely to have a key role in the function of this protein in order to be so highly conserved during evolution.

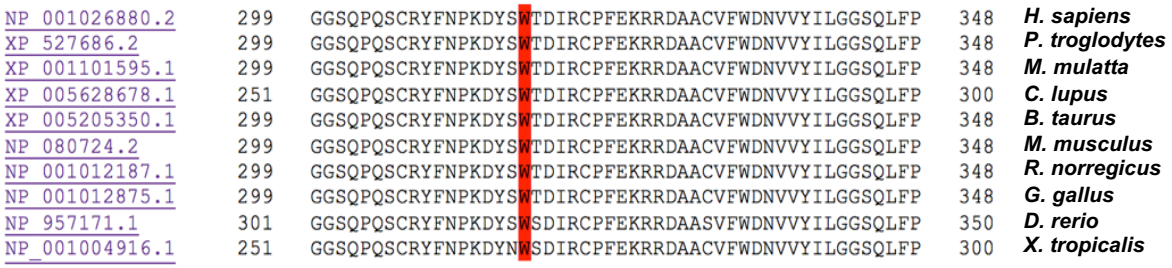


Figure 5-4 Conservation analysis of a region of KLHL7 protein

A representation of the conservation analysis of KLHL7 protein around the location of the variant identified, showing protein sequences from *H. sapiens* down to *X. tropicalis*. The mutated amino acid residue p.W316 is highlighted in red.

In addition, RNA *in situ* hybridisation assays in mouse embryos (embryonic day 14.5) indicate that *KLHL7* is expressed in the brain, spinal cord, thoracic ganglion and trigeminal ganglion and the eye. This suggests that the tissue expression pattern is consistent with the disease phenotype in humans (**figure 5-5**). This information was important at the time of the variant discovery, because *KLHL7* was at that time only associated with autosomal dominant (AD) retinitis pigmentosa type 42 (Friedman et al., 2009) and the expression profile provided further confidence in pursuing this research study into functional characterization.

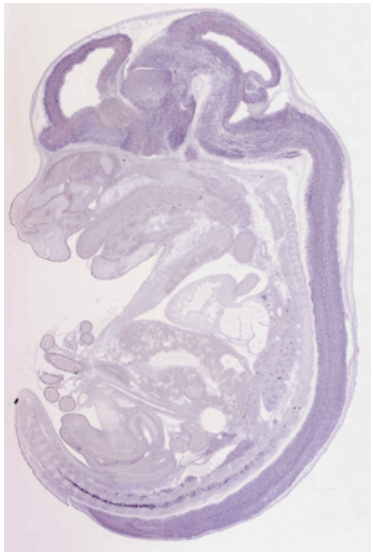


Figure 5-5 Representation of KLHL7 expression pattern in mouse models.

This image is adapted from the Mouse Genome Informatics (MGI) database illustrating the expression pattern of *KLHL7* by using RNA *in situ* hybridization assays (Finger et al., 2017). This provided evidence supporting the high expression level of *KLHL7* in the central nervous system (brain, spinal cord, thoracic ganglion and trigeminal ganglion) and the eye, and the potential relevance of *KLHL7* in an autosomal recessive neurodevelopmental disorder.
Website link: http://www.informatics.jax.org/assay/MGI:4825807#euxassay_011530_17_id [Accessed January 2016].

Clinically, this case was challenging to interpret because the children had no eye phenotype and the parents, who carried a heterozygous variant in *KLHL7*, would be expected to have symptoms of autosomal dominant retinitis pigmentosa type 42. However, the parents did not have a detectable eye phenotype, perhaps due to their young age, as retinitis pigmentosa type 42 is usually a late onset disease and the parents are in their early thirties. Interestingly, during the progress of this study the first study was published linking a *KLHL7* mutation to an autosomal recessive neurodevelopmental condition (Angius et al., 2016), providing more certainty that this is the causative variant in family ND12.

5.3.2 The biological function of KLHL7 protein in disease

KLHL7 encodes a 66kDa substrate-specific adapter of E3 ubiquitin ligase complexes, specifically targeting substrates for degradation. The KLHL7 protein is made up of a BTB and BACK domain that mediates binding to E3 ligases, followed by six Kelch repeat domains that appear to be a substrate-specific binding site (**figure 5-6**). A known E3 ubiquitin ligase complex is the CUL3-KLHL7 complex, where KLHL7 forms a dimer with CUL3 through its BTB and BACK domains. The enzymatic activity is exerted by a RING finger-based E3 ligase complex formed by bringing an E2 ubiquitin conjugating enzyme and the substrate in close proximity (Kigoshi et al., 2011). It is known that KLHL7 induces polyubiquitination at ubiquitin residue Lys48, targeting substrates for proteasomal degradation (Li and Ye, 2008). However, it is not yet known which substrates KLHL7 specifically binds to and which are the specific E2 ubiquitin conjugating enzymes and E3 ligases that form a complex with KLHL7.

Up until 2016, mutations in KLHL7 have only been associated with autosomal dominant retinitis pigmentosa type 42 (RP42), with all the variants reported being within the BACK domain of the KLHL7 protein (Friedman et al., 2009, Hugosson et al., 2010). Functional characterisation of the missense variants in the BACK domain revealed that the E3 ligase activity was diminished, as a result of the missense variant interfering with the incorporation of KLHL7 and CUL3 to form a complex. This resulted in the inappropriate accumulation of substrates that would have otherwise be tagged for degradation (Kigoshi et al., 2011). However, as these variants were heterozygous the E3 ligase activity was not completely abolished, therefore a proportion of the substrates would still be properly degraded. One of the reasons that these variants only cause a phenotype in the eye is potentially due to the fact the even a small amount of aggregated substrates would potentially cause a pathogenic phenotype in more susceptible tissues. The reason why retinal tissue is particularly susceptible in this disease context is unclear.

The same year of our identification of the candidate pathogenic variant in family ND12, the first association of *KLHL7* and a recessive neurodevelopmental disorder was established by Angius et al. Their work revealed four families with homozygous variants in *KLHL7* affected by a rare autosomal recessive disorder known as Criponi syndrome (Angius et al., 2016). The clinical features reported in this study matched the features of the affected individuals in family ND12. Two years later another case was reported with a homozygous nonsense variant, establishing a new association of KLHL7 mutations other than with autosomal dominant RP42 (Jeffries et al., 2018). A key characteristic that is shared by these cases is that all the mutations reported to be associated with the autosomal recessive disease, are all present within the Kelch domains which are

responsible for substrate binding. **Figure 5-6A** outlines the KLHL7 protein structure and indicates the location of the RP42 specific mutations as well as the Crisponi syndorme-specific mutations. In addition, the diagram indicates the nonsense variant that we have identified. This localizes at the first Kelch domain, either abolishing the protein completely, or in case of a truncated protein, entirely stopping any substrate from binding. **Figure 5-6B** illustrates the structure of the Kelch domains, showing the 6 repeats and the location of our identified variant.

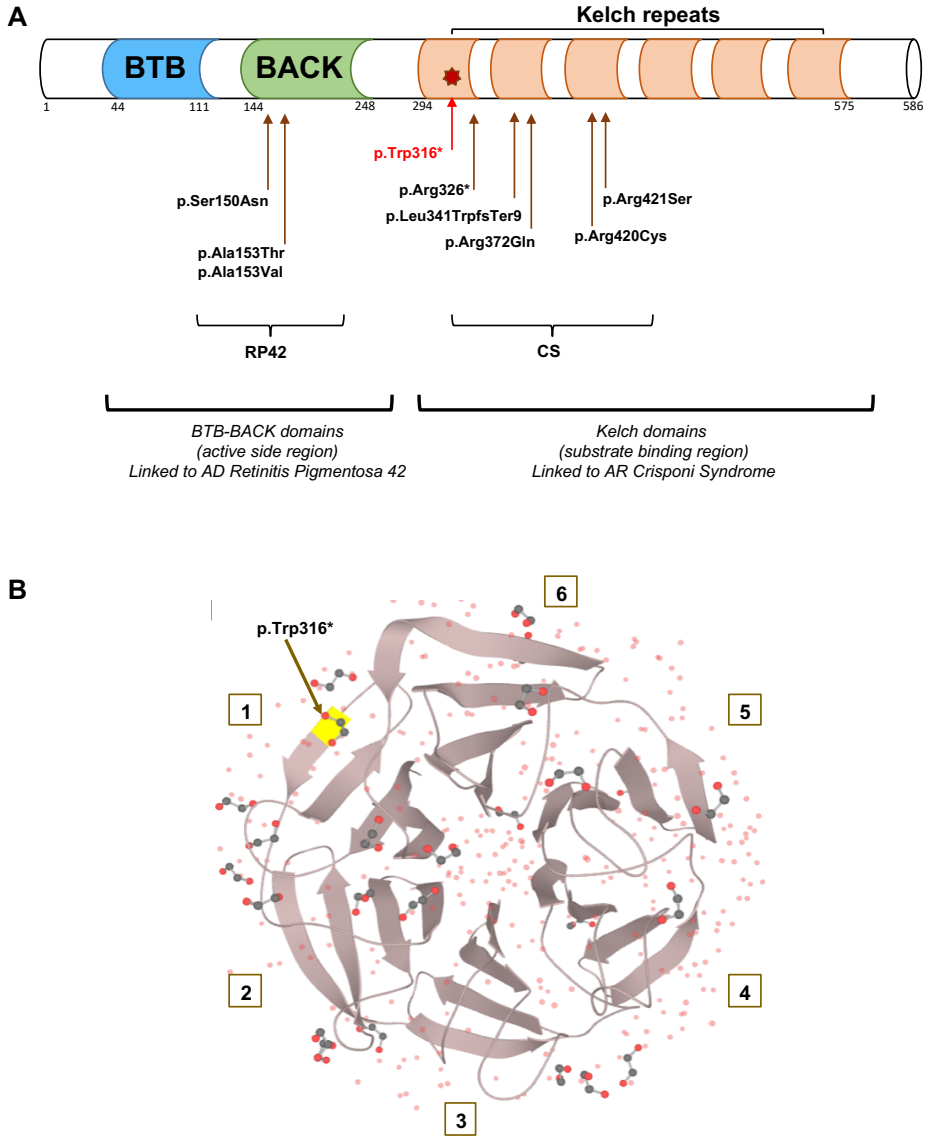


Figure 5-6 Structure outline of KLHL7 protein

A – A representation of the KLHL7 protein outlining the different domains of the protein, showing specifically how the BTB and BACK domain are binding to the active side region of the E3 ligase and how the six KELCH repeats are specific for substrate binding. Arrows also show the different mutations identified in *KLHL7*, separating the ones that are causing autosomal dominant retinitis pigmentosa type 42 and the ones causing autosomal recessive Crisponi syndrome. B – A protein structure adapted from UniProt, showing the six Kelch repeats. The location of the variant identified in family ND12 is marked in yellow at the start of the first Kelch repeat.

5.4 Functional characterisation of p.W316* in KLHL7

At the time of discovery, the homozygous nonsense variant p.W316* was novel and no initial associations had been made between *KLHL7* mutations and autosomal recessive disease that may or may not have an eye phenotype (Angius et al., 2016). This possibility was puzzling both clinically and scientifically, and further functional characterisation of the variant was necessary in order to gain more insight on the pathomechanism associated with the recessive neurodevelopmental phenotype.

5.4.1 Assessing the impact of p.W316* on metabolism

According to the NHS Highly Specialised Service for Rare Mitochondrial Disorders, a severe combined deficiency of mitochondrial respiratory chain complexes I and IV was observed in a skeletal muscle biopsy from individual JT728. The results from the biopsy are entirely compatible with the stroke-like episodes, developmental delay, failure to thrive, myopathy and elevated blood lactate levels. Clinical immunohistochemical evidence of mitochondrial biochemical defects were not conclusive, therefore a more detailed biochemical test needed to be performed to further understand the effect of the KLHL7 p.W316* mutation on metabolism.

In order to interpret the effect that the nonsense variant in *KLHL7* has on mitochondrial function, the oxidative phosphorylation pathway of energy production was initially tested using a Seahorse XFe96 Extracellular Flux Analyzer, running the MitoStress test (section 2.2.22). Control and patient fibroblasts were challenged with toxins that block different processes of the oxidative phosphorylation pathway, in three biological replicates. The oxygen consumption rate (OCR) of the cells was measured before and after each toxin injection allowing for identification of important measures of mitochondrial function. The trace in **figure 5-7** shows the changes in OCR in real time.

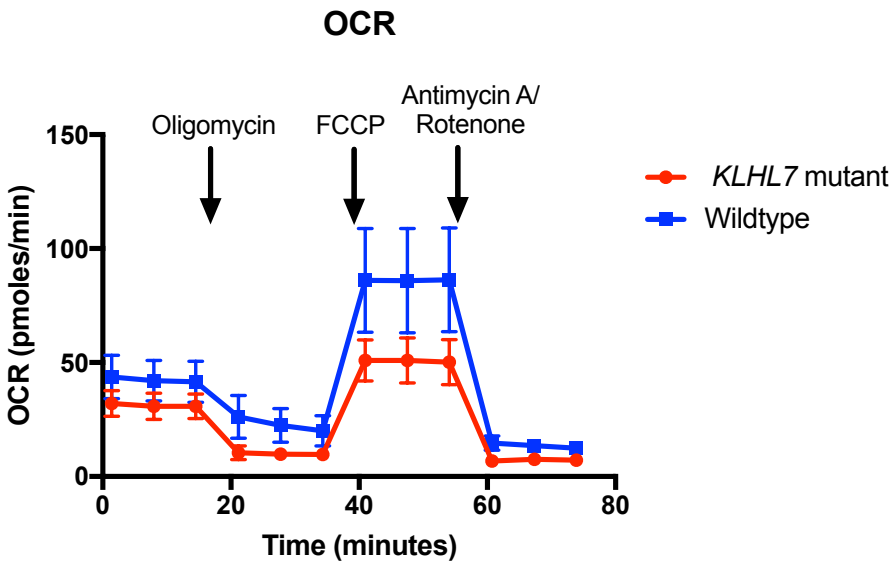


Figure 5-7 Seahorse trace for the MitoStress test of *KLHL7* p.W316* fibroblasts

Traces of real-time OCR changes between wildtype and patient fibroblasts. The first injection was with oligomycin to block ATP synthase, the second injection was with FCCP to uncouple the electron transport chain and, lastly, Antimycin A/ Rotenone to completely block electron transport.

Further analysis of the OCR measurements at certain time points of the experiment allow for identification of key aspects of mitochondrial function, such as basal and maximal respiration, ATP production, proton leak and spare respiratory capacity. The analysis of each one of these characteristics is summarised in **Figure 5-8**. There was no difference in ATP production between wildtype and patient fibroblast, but the basal and maximal respiration, spare respiratory capacity, proton leak and coupling efficiency were significantly different in mutant *KLHL7* cells compared to wildtype.

More precisely, the MitoStress test revealed that the basal respiration of patient fibroblasts is slightly reduced compared to wildtype, but when forced, the patient fibroblasts cannot reach maximum respiration. This observation validates a defect in the oxidative phosphorylation pathway, as observed by the muscle biopsy testing. In addition, the patient fibroblasts show significantly reduced spare respiratory capacity compared to the wildtype, again supporting the hypothesis of a defect in the oxidative phosphorylation chain. This is further supported by the reduced proton leak and the increased coupling efficiency in the patient fibroblasts compared to the wildtype control, suggesting a potential disruption in ionic homeostasis (Papa et al., 2012). This would lead to increased H^+ concentration in the intermembrane space, affecting the proton-motive force and the coupling efficiency, since more H^+ will flow through ATP synthase (**figure 5-8**.)

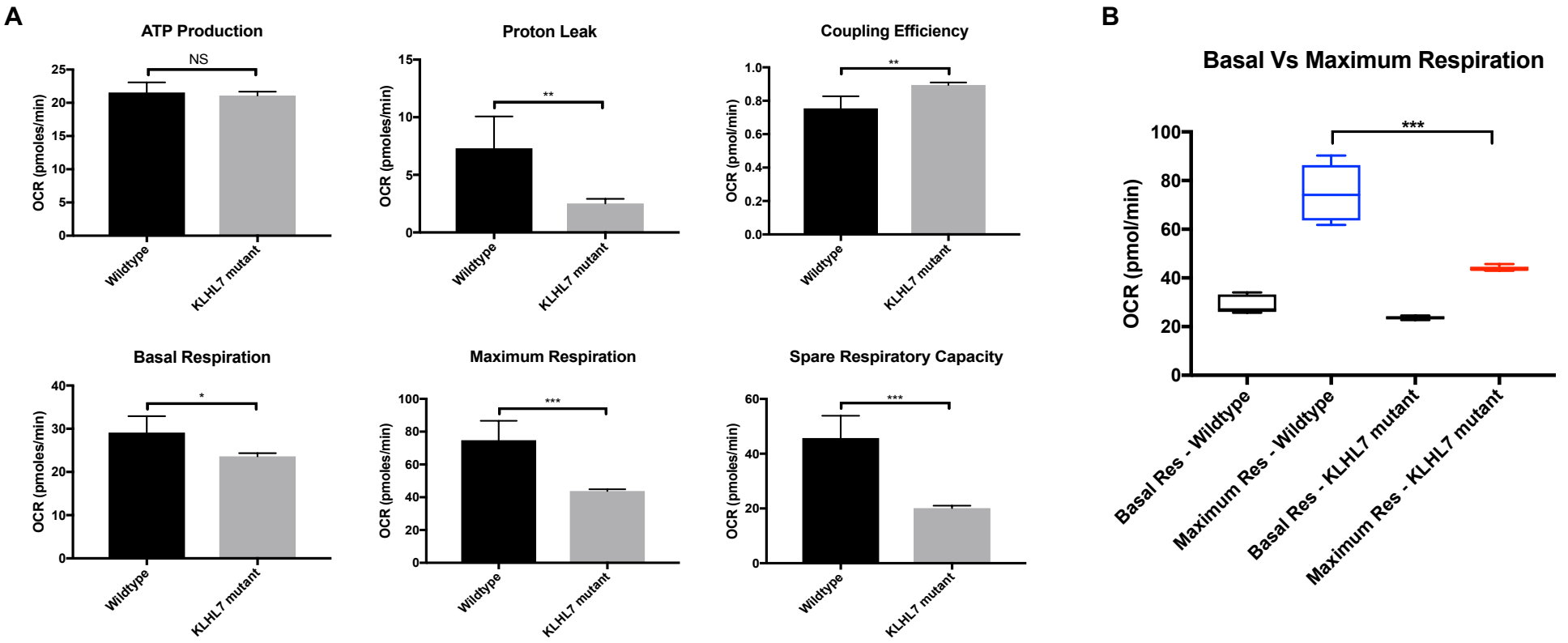


Figure 5-8 Metrics of mitochondrial function in wildtype and patient fibroblasts.

Bar graphs that quantitate the metrics obtained from the MitoStress Test. Measurements at specific time points before or after toxin injections allowed for the calculation of each of the key aspects of mitochondrial function. A) Patient fibroblasts were significantly under-respiring at basal levels and when forced to maximum respiration, had a disrupted proton-motive force, but produce similar amounts of ATP. B) This representation shows the same results, but highlights the defect in achieving maximum respiration in the patient fibroblast upon FCCP injection. All statistical tests were performed by a two-tailed Student t-test (n=3; NS, not significant; * $p<0.05$; ** $p<0.01$; ***$p<0.001$). Error bars indicate s.e.m.

The MitoStress test indicated a clear disruption of the oxidative phosphorylation chain, as respiration is significantly reduced both at basal and maximum levels. However, ATP levels were similar between wildtype and patient fibroblasts, suggesting the utilization of a different pathway for energy production such as glycolysis. In order to further investigate this observation, a Glycolysis Stress test was performed. Wildtype and patient fibroblasts were challenged with various toxins that block different processes of either glycolysis or the oxidative phosphorylation pathway, in three biological replicates. The extracellular acidification rate (ECAR) of the cells was measured before and after each injection, at regular intervals, allowing measurements to be taken during key aspects of glycolysis. The trace below shows the changes in ECAR in real time (**Figure 5-9**).

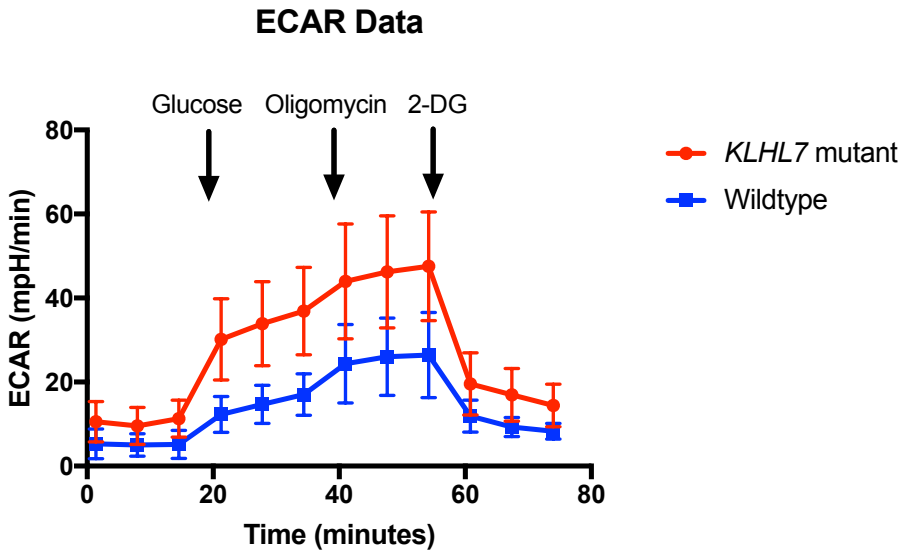


Figure 5-9 Seahorse trace for the Glycolysis Stress test of *KLHL7* p.W316* fibroblasts

Representation of the changes in ECAR in real time between normal control and patient fibroblasts. The first injection was with glucose to promote glycolysis, then oligomycin to block ATP synthase and any ATP production from oxidative phosphorylation, and lastly 2-deoxyglucose (2-DG), a structural analogue of glucose that completely blocks glycolysis, was injected.

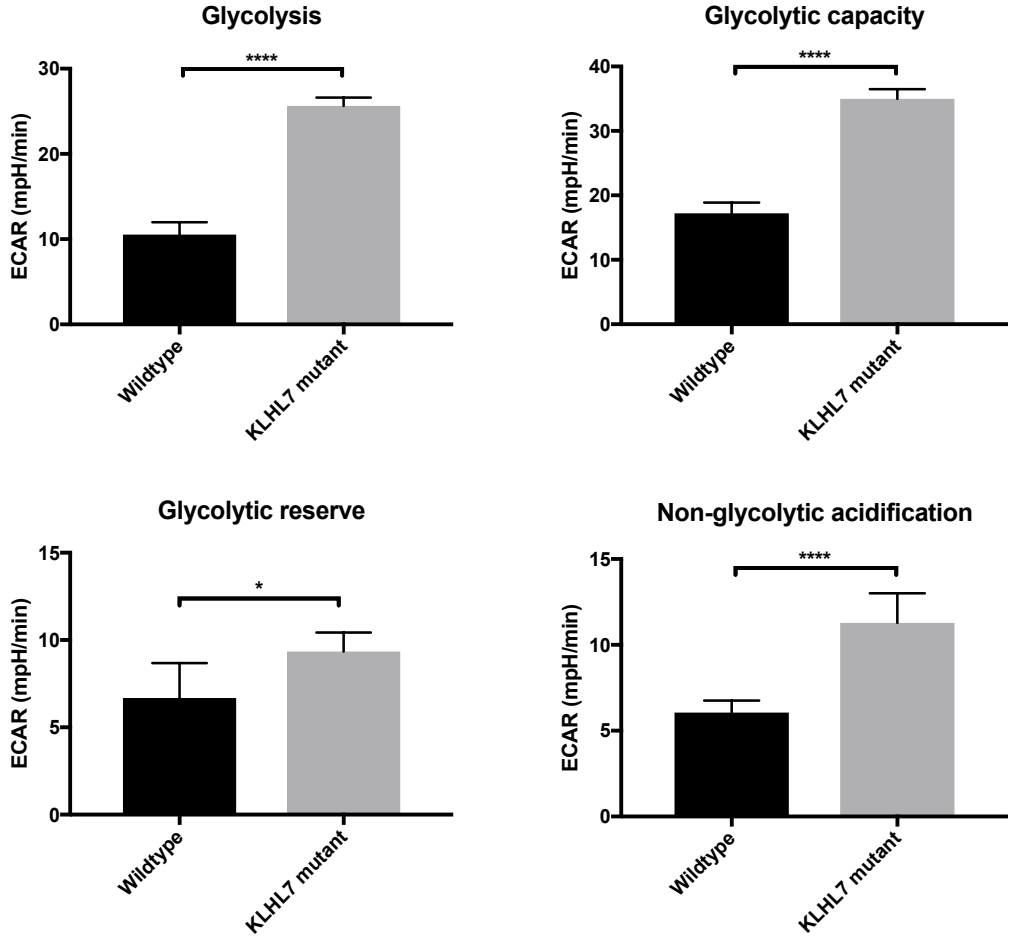


Figure 5-10 ECAR metrics accessing glycolysis in *KLHL7* p.W316* fibroblasts

Representation of all the metrics obtained from the Glycolysis Stress test. Measurements at specific time points before or after each injection allowed for the calculation of each of the above properties to be obtained (Agilent protocol). *KLHL7* patient fibroblasts seem to have an increased glycolytic phenotype compared to the control HDF cells. All statistical tests were performed by a two-tailed Student t-test (n=3; ns, not significant; * $p<0.05$; ***$p<0.001$). Error bars indicate s.e.m.

The Glycolysis Stress test revealed significant differences between wildtype and *KLHL7* patient fibroblasts. The level of glycolysis (or in other words the conversion of glucose to pyruvate) was much higher in the *KLHL7* patient fibroblasts compared to the wildtype, supporting the previous findings that oxidative phosphorylation is disrupted in the mutant cells and an alternative method for energy production had to be recruited. The same pattern was also observed in the glycolytic reserve and the glycolytic capacity between the wildtype and mutant cell line. This indicates that when forced, *KLHL7* mutant cells have the potential to reach maximum glycolysis in a much higher level compared to wildtype control, suggesting that all the pyruvate made is mostly used in glycolysis and barely used in aerobic metabolism through oxidative phosphorylation. In addition, a significant difference was also observed in the non-glycolytic acidification rate observed in the wildtype and patient fibroblasts. *KLHL7* p.W316* fibroblasts had a significantly

increased non-glycolytic acidification level compared to the wildtype fibroblasts, indicating a source of acidification separate to the glycolysis pathway. This can potentially arise from the conversion of pyruvate to lactate that is further reduced to lactic acid inducing acidification (Mookerjee et al., 2015). The diagram in **figure 5-11** outlines the two possible outcomes following glycolysis.

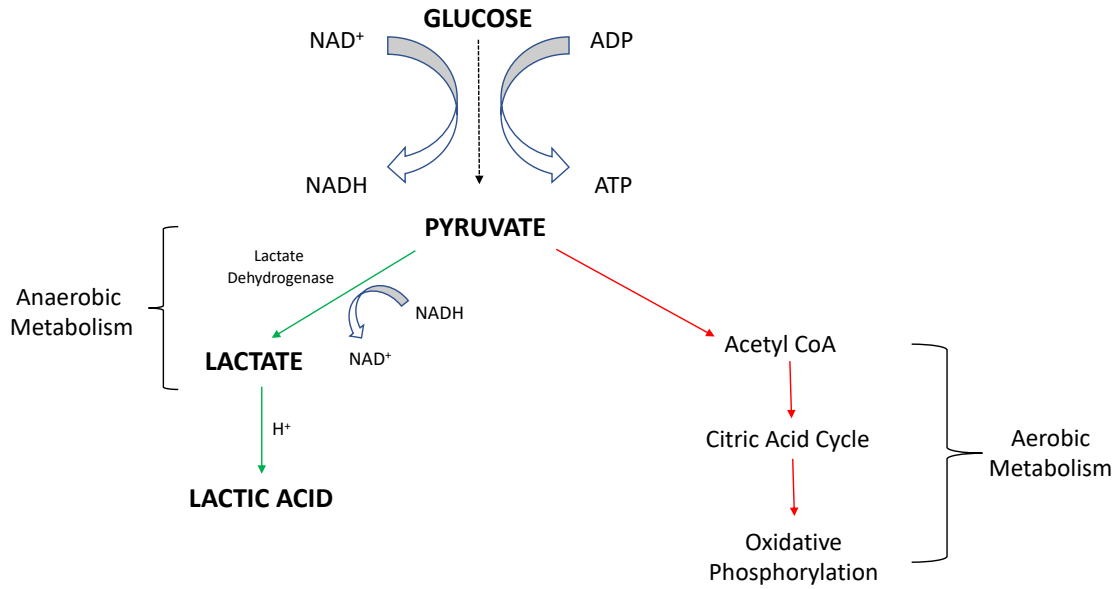


Figure 5-11 Outline of aerobic and anaerobic metabolism pathways

The diagram illustrates the two potential uses of pyruvate in aerobic and anaerobic metabolism. Pyruvate can be converted into acetyl CoA and fed into the Kreb's cycle and the oxidative phosphorylation chain for maximum production of ATP by aerobic respiration. Pyruvate can also be converted to lactate for immediate energy demands when aerobic metabolism cannot produce enough ATP at any given instance, for example during exercise. Lactate can be further reduced to lactic acid leading to acidification.

Overall, this investigation supported the results obtained from the clinical tests on the muscle biopsies, revealing a defect in the oxidative phosphorylation chain resulting in reduced aerobic respiration in *KLHL7* null fibroblasts. This could be due to the proton-motive force, or more specifically due to one of the complexes of the oxidative phosphorylation chain being impaired, as suggested by the clinical tests (Ghezzi and Zeviani, 2018). In addition, the mitochondria could be under excessive stress resulting in impaired aerobic metabolism. This is also supported by the fact that the mutant cells tend to have increased acidification levels, perhaps due to the conversion of pyruvate to lactate in order to maintain their energy requirements. These are interesting descriptive observations, but further experiments are required in order to understand how the *KLHL7* mutation can have such a severe effect on mitochondrial function.

5.4.2 Using BioID to identifying novel protein interactions of KLHL7 within the ubiquitin proteasome system.

5.4.2.1 Design and optimisation of BioID experiments

BioID was used to detect physiologically relevant protein-protein interactions associated with KLHL7, because this technique allows for identification of weak and transient interactions (Roux et al., 2018). This is key for this project as KLHL7 is involved in forming an E3-ligase complex for ubiquitinating substrates for proteasomal degradation and this process is relatively fast with many interactions being transient and some even indirect. However, BioID is powerful enough to identify these interactions using a promiscuous biotin ligase (BirA*) that is fused to the protein of interest and is capable of biotinylating proximal endogenous proteins. More precisely, in our experiment, KLHL7 was N-terminally tagged with BirA* and then transfected into hTERT-RPE1 cells to induce biotinylation upon addition of exogenous biotin. Interacting proteins are then isolated by a standard biotin-affinity capture and identified using mass spectroscopy (Roux et al., 2012). A summary of the experimental plan is illustrated in **Figure 5-12.**

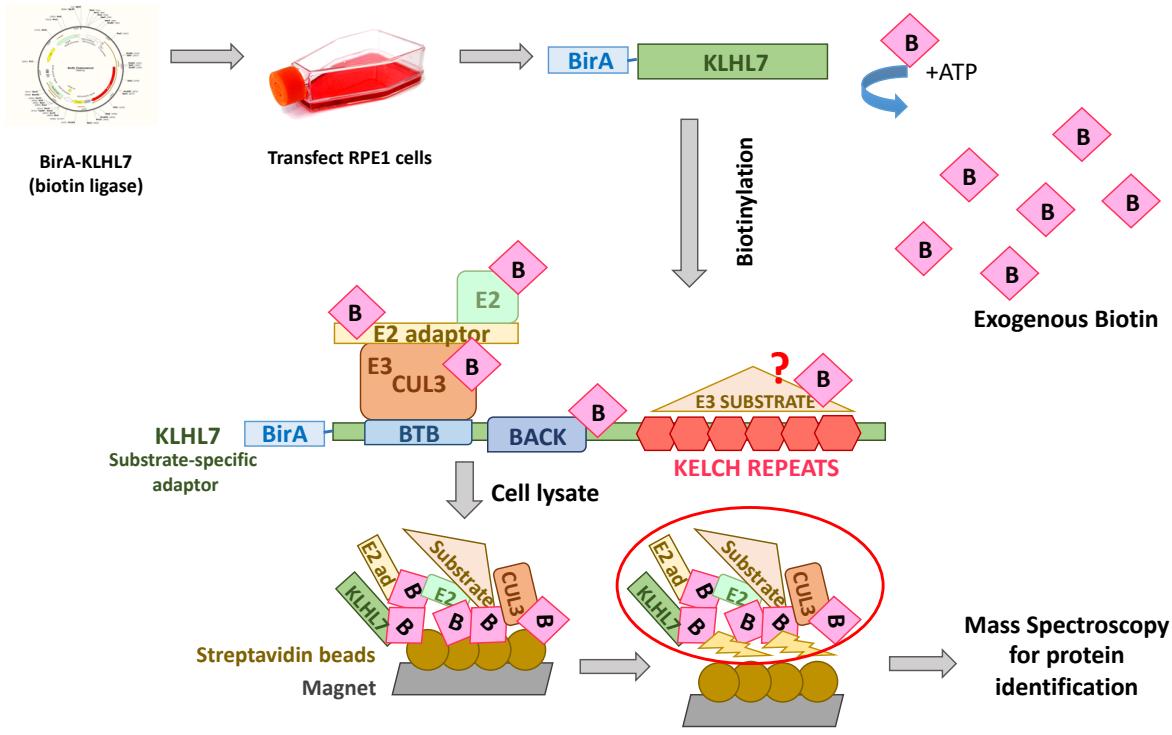


Figure 5-12 Outline of experimental plan for BioID

A summary of the key steps in the BioID experiment, starting with exogenous expression of the BirA* biotin ligase-KLHL7 fusion protein in hTERT-RPE1 cells. Addition of exogenous biotin causes biotinylation of proximal or closely interacting proteins followed by cell lysis and biotin-affinity capture of labelled proteins. Protein identification is achieved by using mass spectrometry.

The BirA tagged KLHL7 (BirA-KLHL7) protein was exogenously expressed using a standard transient transfection protocol (see section 2.2.15).The design and construction of the BirA-KLHL7 plasmid was done using the In-Fusion cloning methodology (see section 2.2.11.2). The *KLHL7* full-length ORF sequence was amplified from a pENTR221 plasmid (DNASU, plasmid HsCD00351708) using specifically designed flanking primers (see section 2.1.4) and was cloned into the pcDNA3.1 myc-BioID plasmid (Addgene, plasmid #35700) (Appendix H). The final map of the biotin ligase KLHL7 fusion protein (BirA-KLHL7 plasmid) is illustrated in **figure 5-13**, together with the PCR validations of each cloning step. The final plasmid was verified by Sanger sequencing (see section 2.2.9) using T7 and bGH primers, along with internal primers designed within the *KLHL7* sequence (see section 2.1.4).

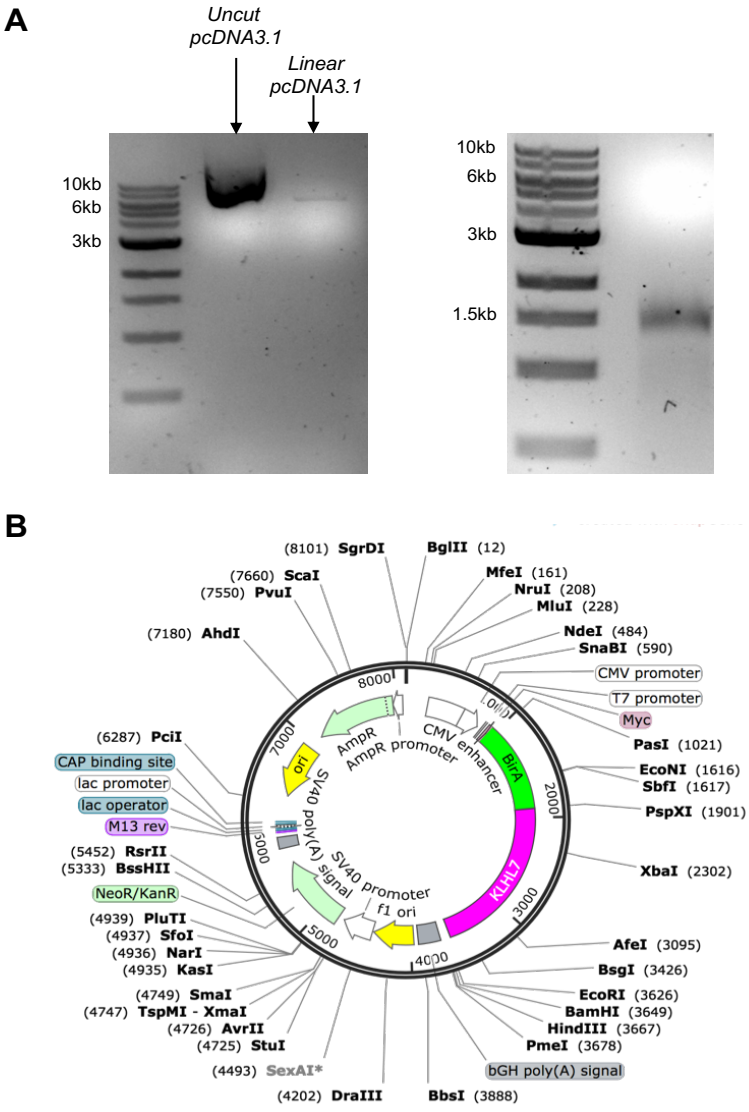


Figure 5-13 Construction of BirA-KLHL7 plasmid for BioID experiment

A – PCR validation steps for In-Fusion cloning protocol. On the left, a validation of the double digest performed using EcoRI and BamHI enzymes to obtain a linearised vector and on the right the band from the PCR amplification of the KLHL7 insert. B – The final map of BirA-KLHL7 plasmid used for the BioID experiment.

Once the plasmid was purified and sequence verified, the next step of the procedure was to optimise the conditions of the BioID experiment. Standard transfections were used (see section 2.2.15), followed by incubation with media supplemented with biotin once the transfection complexes were removed. However, incubation times with biotin-supplemented media after transfection had to be optimized in order to achieve the best possible biotinylation levels. In the first instance, we used a 24 hour incubation period, which is the maximum incubation time recommended (Roux et al., 2012), allowing sufficient time for BirA-KLHL7 to be expressed and for induction of biotinylation. These conditions were validated between three different BioID plasmids (USP39 and RBX1 are relevant to other projects) and the appropriate controls, as outlined in **figure 5-14**. This clearly showed that 24 hr are necessary for KLHL7 to reach a good expression level (shown by anti-c-myc staining) and to have sufficient biotinylation (shown by the streptavidin staining).

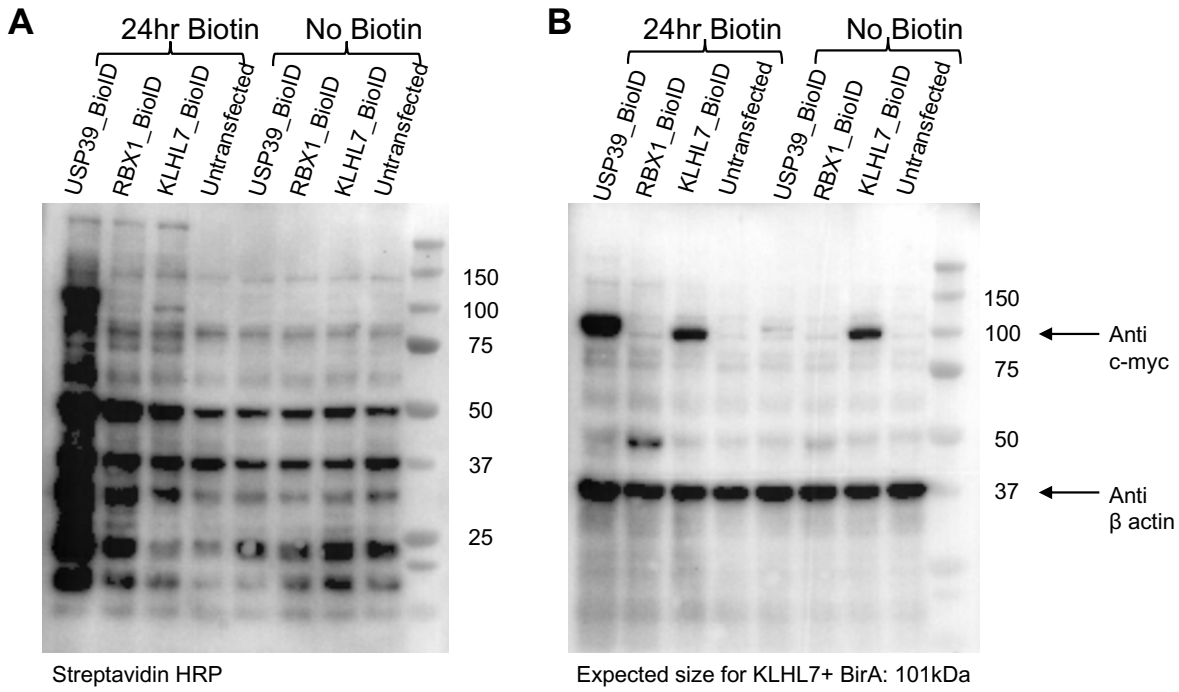


Figure 5-14 Validation of BirA-KLHL7 plasmid by western blotting.

A – Western blot showing, on the left, the induction of biotinylation in transfected cells that have been incubated with biotin media for 24 hr. On the right, the same transfections but without biotin incubation show the absence of higher molecular weight material, indicating that biotinylation is being induced with the biotin media. B – Western blot showing the transfection level of KLHL7 fusion protein during the 24 hr incubation period along, with the β-actin staining for loading control. The other proteins shown on these blots (USP39 and RBX1) were run alongside kLHL7 for validation purposes but are not relevant to this project.

Once the biotin incubation time and transfection protocol were optimised, BioID experiments were performed (see section 2.2.21) using the hTERT-RPE1 cell-line as our model system. Ideally, a neuronal cell line would be the best model system based on the phenotype of the patients, but due to the challenges of optimizing transient transfection in neuronal lines it was decided to select an easier model to work with but still relevant to KLHL7 biological mechanisms. Since *KLHL7* mutations are a known cause of retinitis pigmentosa, the hTERT-RPE1 cell line was considered an appropriate model system to resemble the mechanisms and interacting partners that KLHL7 is involved with in retinal tissue. BioID experiments were performed, lysates specifically pulled down with streptactin beads, trypsinised and then sent to our collaborator (Karsten Boldt, University of Tubingen) for mass spectroscopy analysis. For this project, three different biological replicates were used with the appropriate controls in place (**Figure 5.15**) to help with the downstream filtering of identified proteins. In addition, a proteasomal inhibitor (MG132) was also used as an additional experiment for each biological replicate. Given that KLHL7 specifically targets proteins for degradation, the proteasome was inhibited in order to enhance the amount of potential biotinylated substrates present in the final lysate. Without proteasomal inhibition, many of the tagged substrates may have been degraded and thus might have been missed.

	No Biotin	24 hours Biotin	No Biotin	24 hours Biotin
No MG132	Untransfected	Untransfected	KLHL7-BirA transfection	KLHL7-BirA transfection
Plus MG132	Untransfected	Untransfected	KLHL7-BirA transfection	KLHL7-BirA transfection
	Negative controls	*Positive controls*		*Experiment*

Figure 5-15 Experimental samples and controls for BioID experiments

The table summarises each sample prepared for each biological replicate for the BioID experiments. Untransfected samples without biotin treatment were considered as the negative control of the experiment and any hits obtained by these samples were discarded. Untransfected samples with 24 hr incubation with biotin were considered as one of the positive controls of the experiments that helped in excluding any false positive results arising from excess biotin in cells. Transfected samples without biotin were also considered as positive controls that helped in excluding any false positive results arising from having an active biotin ligase enzyme within the cells that could potentially biotinylate proteins in near proximity using endogenous biotin present within the cells. Furthermore, having the same setup with and without the proteasome inhibitor was beneficial in downstream analysis, specifically looking for any enriched proteins that could be considered as potential substrates of the E3 ligase complex that contains KLHL7 as a component.

5.4.2.2 Analysis of mass spectroscopy data

Affinity purified eluates were precipitated with chloroform and methanol followed by trypsin digestion as described before (Gloeckner et al., 2009). C-MS/MS analysis was performed on Ultimate3000 nanoRSLC systems (Thermo Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) by a nano spray ion source. MS/MS data were analyzed using the MaxQuant software (version 1.6.1.0) (Cox and Mann, 2008, Cox et al., 2009). As a digesting enzyme, Trypsin/P was selected with maximal 2 missed cleavages. The data were analyzed by label-free quantification with the minimum ratio count of 3. The first search peptide tolerance was set to 20, the main search peptide tolerance to 4.5 ppm and the re-quantify option was selected. For peptide and protein identification the human subset of the SwissProt database (release 2014_11) was used and contaminants were detected using the MaxQuant contaminant search. A minimum peptide number of 2 and a minimum length of 7 amino acids was tolerated. Unique and razor peptides were used for quantification. The match between run option was enabled with a match time window of 0.7 min and an alignment time window of 20 min. The statistical analysis including ratio, t-test and significance A calculation was done using the Perseus software (version 1.5.5.3) (Tyanova et al., 2016). Significance A is a Q-function that detects outliers from a normal distribution. The background binders are the normal distribution and the interactors are the outliers. The output of the analysis is plotted in **figure 5-16**.

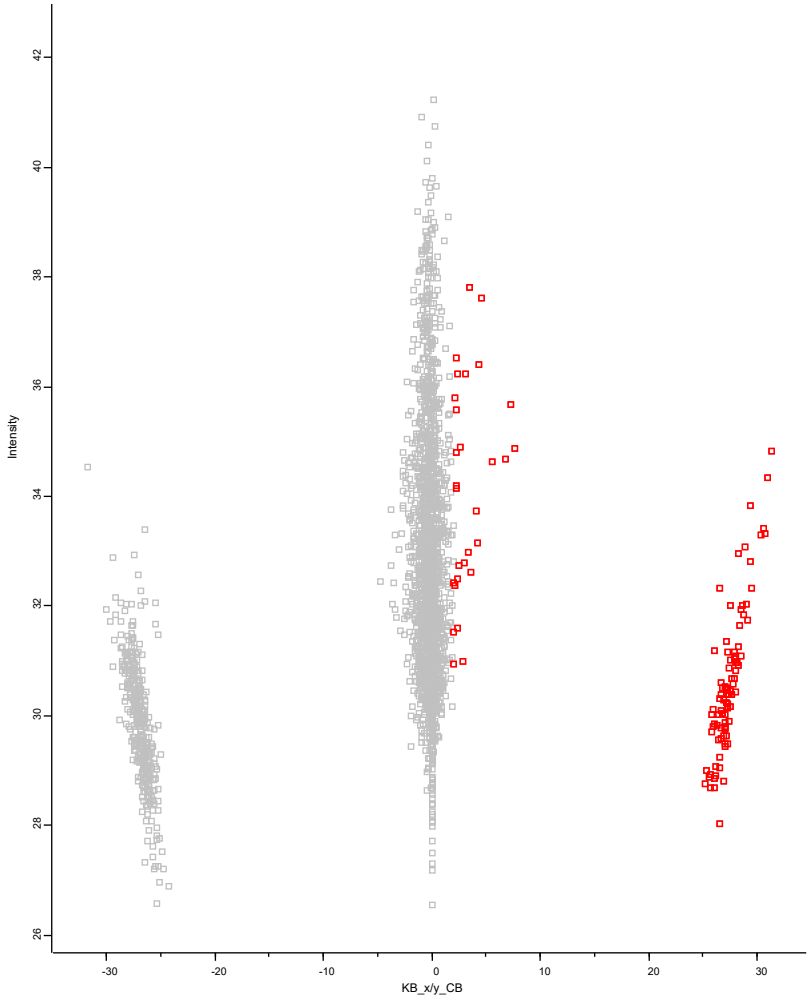


Figure 5-16 MS/MS data analysis using MaxQuant software

Plotted are the log10 ratios (x-axes) and the log10 intensities (y-axes) for all proteins quantified by the MaxQuant software. Proteins significantly enriched from the background noise are highlighted in red (Benjamini-Hochberg corrected, left sided Significance A < 0.05).This represents proteins that deviate from the normal distribution and are considered as outliers. These outliers are the proteins that are significantly enriched when KLHL7 is biotinylated. .

5.4.2.3 Identification of protein-protein interactions

Following specialized analysis of data by Dr Boldt, further manual analysis was performed on the identified hits, using different criteria at each step. Initially any hit that was not passing the A significance test was excluded from the list, and any hits enriched upon proteasomal inhibition were marked as potential substrates. As mentioned previously, alongside KLHL7 we optimized BioID experiments for other projects (RBX1 and USP39 proteins) that are also involved in the ubiquitin-proteasome system. KLHL7 data were compared to data obtained from these other projects, prioritizing hits that are unique to KLHL7. Furthermore, hits were also prioritized according to their biological relevance, specifically looking for any E3-ligases and E2 ubiquitin conjugating enzymes

to help us construct the complex for which KLHL7 is an adaptor protein, thereby understanding more about the disease mechanism. In addition, any hits that had previously been associated with similar diseases or have a neuronal function were highlighted. The figure below (**figure 5.17**) illustrates the top hits obtained by the BioID experiment, ranked based on the average number of peptides identified per biological replicate. The figure also highlights potential E2s and E3s, as well as the enriched hits that could be potential substrates.

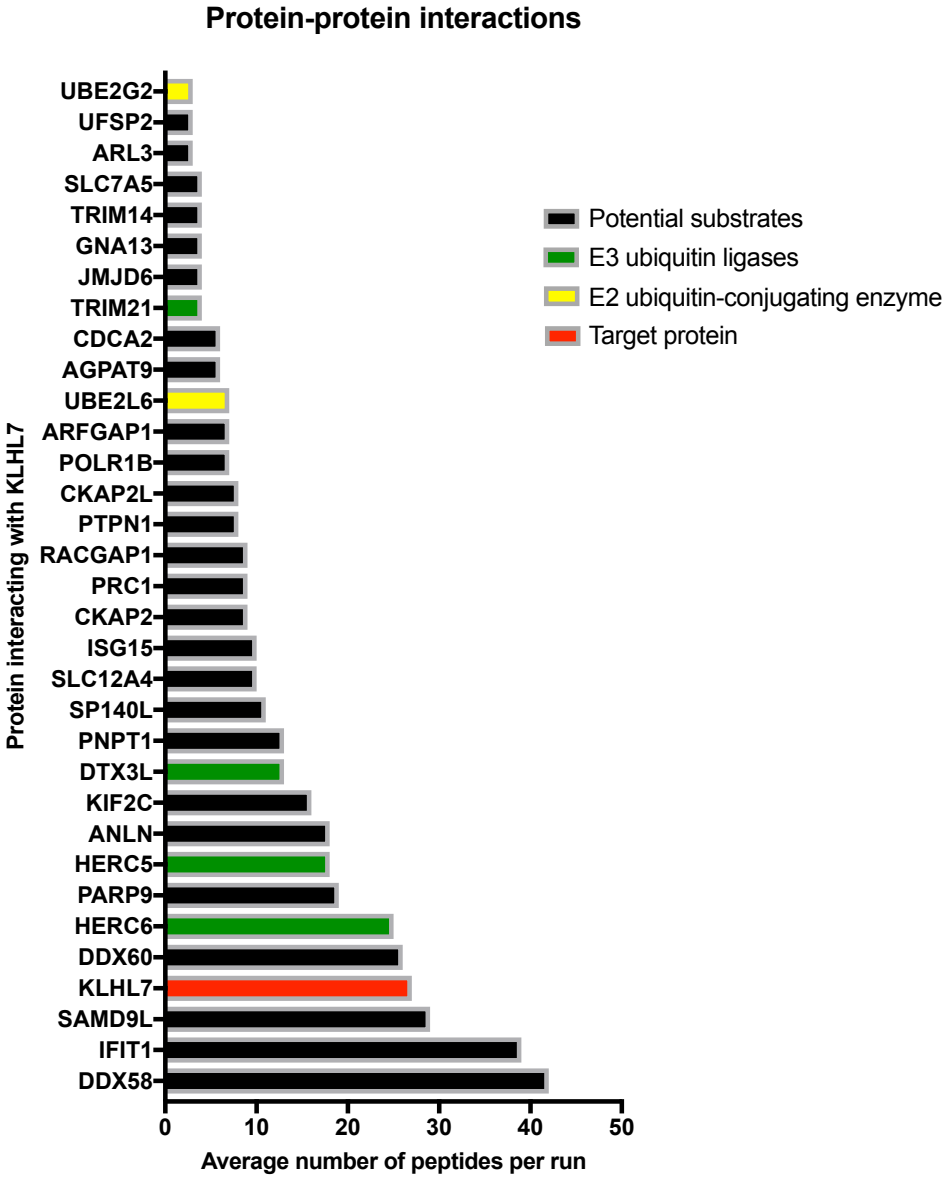


Figure 5-17 List of top hits identified from the BioID experiment

All the hits are presented in an ascending order based on the average number of peptides identified per run. All the hits listed passed the A-significance test from the analysis. Proteins coloured in green could be potential E3-ligases, and proteins coloured in yellow potential E2-ubiquitin-conjugating enzymes, all working in the same complex as KLHL7. The remaining proteins (in black) could be potential substrates of this complex.

5.4.2.4 Protein network analysis

In order to comprehend the most significant hits of the BioID experiment listed above, various computational tools were used to help with grouping proteins and building networks based on known protein-protein interactions. To achieve this a database of known and predicted protein-protein interactions, known as STRING, was used (Snel et al., 2000). This database comprises of interactions that include both physical (or direct) and functional (or indirect) associations derived from the following sources: genomic context predictions, high-throughput lab experiments, conservations and co-expression datasets, automated text-mining and previous knowledge from databases. By using STRING v11.0 (Szklarczyk et al., 2019), we were able to identify two already known networks of protein-protein interactions. The first big network illustrated on the left side of **figure 5-18** contains all the hits that are associated with forming the E2/E3 ubiquitin ligase complex, comprising of interactions between various E3 ligases (DTX3L, TRIM21, HERC5, HERC6), E2 ubiquitin-conjugating enzymes (UBE2G2, UBE2L6) and a few other interactors that may or may not have a role in ubiquitination. On the right side of the figure there is another smaller network, mainly comprising of protein associated with cell cycle, cytokinesis and cytoskeleton maintenance. This provides evidence of a potential pathomechism that implicates KLHL7 in these cellular processes. There are also a few hits that are not associated with any of the two networks, with potential novel links to functional mechanisms associated with neurodevelopment making them interesting candidates for future investigations. An example is *ARL3*, mutations in which cause both AD retinitis pigmentosa type 83 (Holtan et al., 2019) and AR Joubert syndrome type 35 (Alkanderi et al., 2018). With these established genetic links, ARL3 could be an interesting interacting partner of KLHL7 since they share similarities in both the eye phenotype of RP and a severe global developmental syndrome. Surprisingly, in this high-throughput experiment to identify protein-protein interactions, KLHL7 was not associated with any of the hits analysed using STRING, indicating an extremely novel protein with no established associations. This provides numerous possibilities for future work in understanding the role of this protein and how its disruption can either cause an AD phenotype of the eye or a severe AR global developmental delay phenotype.

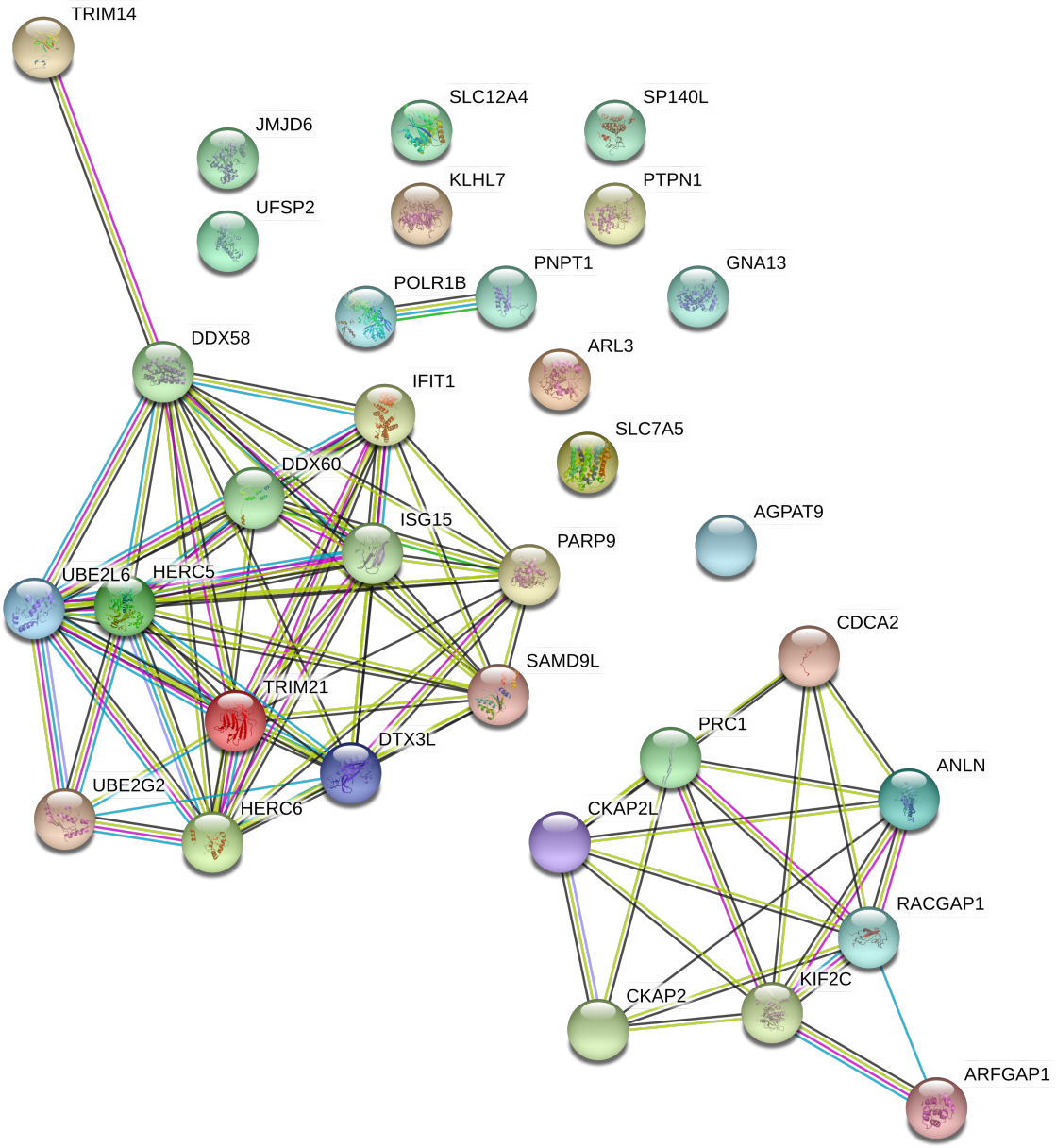


Figure 5-18 KLHL7 protein-protein interaction network from BioID experiments

The figure above illustrates the networks forming from the significant hits obtained from KLHL7 BioID experiments using STRING v11.0 software. The network on the left mostly comprises proteins involved in the ubiquitin-proteasome system, and the network on the right mostly of proteins involved in cell cycle, cytokinesis and cytoskeleton. Interestingly, KLHL7 is not associated with any of the hits, highlighting the novelty and importance of this dataset in understanding the role of this protein in neurodevelopment.

5.4.2.5 Pathway enrichment analysis

To further interpret the BioID data, a pathway enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6.8). DAVID provides a comprehensive set of functional annotation tools useful for interpreting large list of genes (Huang et al., 2009b). DAVID can functionally annotate and classify genes, mainly using GO terms for enriched biological themes (Huang et al., 2009a). Such an enrichment analysis for pathway/process was performed with the 33 significant hits from the BioID data, highlighting a few key processes that are significantly enriched in this dataset. A summary of this analysis is illustrated in **figure 5-19**.

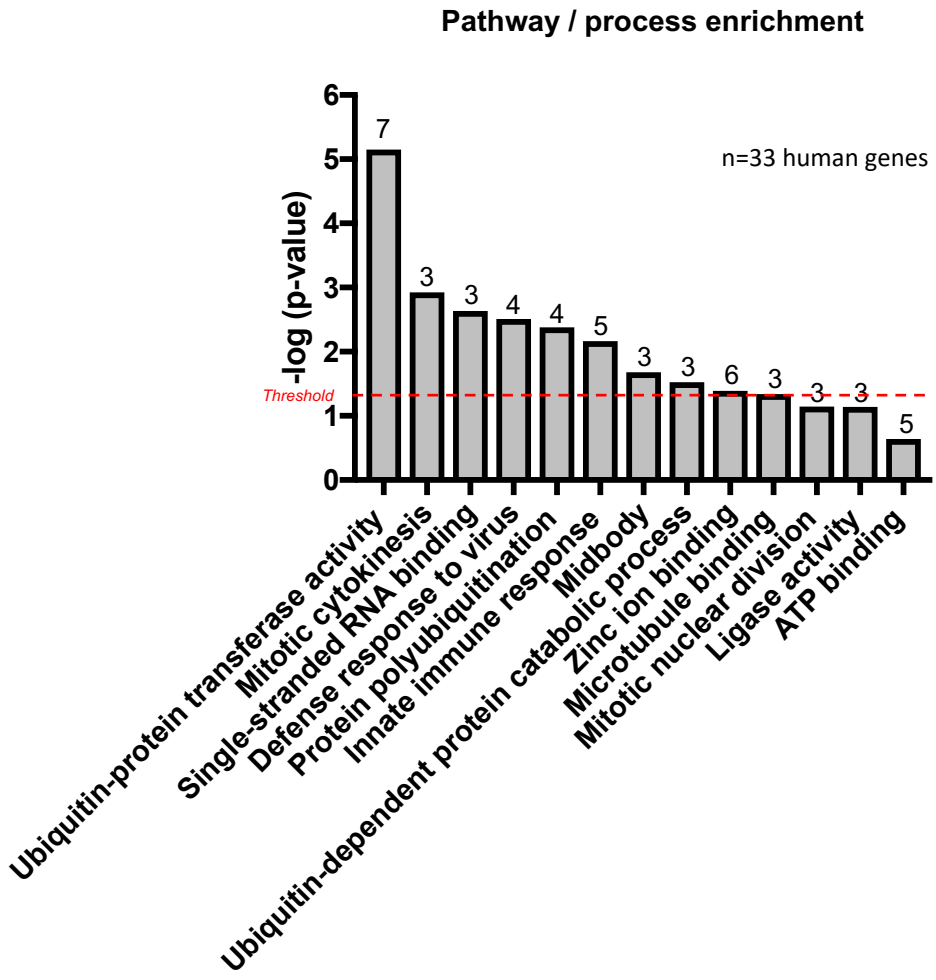


Figure 5-19 Pathway enrichment analysis for the BioID dataset.

A summary of the main pathways or processes enriched in the BioID dataset, with most of them being significant ($p \leq 0.05$ - above the red threshold line). The number above each bar represents the number of genes associated with each process.

The obvious and expected process to be significantly highlighted was the ubiquitin-protein transferase activity with the greatest number of genes being linked to this pathway. This result was expected since KLHL7 is an E3-ligase adaptor and the purpose of the BioID experiment was to identify other partners of KLHL7 that participate in forming the E2/E3 ligase complex for tagging proteins for degradation. A few other ubiquitin-dependent processes have been identified, such as protein polyubiquitination, ligase activity and ubiquitin-dependent protein catabolic processes, again supporting the existing literature on KLHL7 being an adaptor responsible for tagging substrates for proteasomal degradation.

The second most significant process is mitotic cytokinesis. This pathway was still expected based on the previous STRING analysis, as the majority of the genes identified are involved in cytokinesis and cell cycle regulation. Again, a few other relevant processes to cytokinesis are picked up, such as midbody, mitotic nuclear division and microtubule binding.

Furthermore, some novel pathways have also been detected that implicate a few interesting cellular processes. These include single-stranded RNA binding and immune responses. The single-stranded RNA binding could involve binding to various RNA species that can have a huge range of roles. Also, the immune response processes include a defense response to virus and the innate immune response.

5.4.2.6 Functional validation and future work

The output from the BioID data revealed a strong association of KLHL7 with cell cycle and cytokinesis, something worth investigating to further understand the biological function of this protein and its associated pathomechanism. To achieve this the nonsense variant in KLHL7 could be modelled using CRISPR/Cas9 technology to characterize any potential impact of the mutation on cell cycle progression and cytokinesis. The cell line produced, carrying the nonsense variant, can be used in a flow cytometry assay for cell cycle analysis or stain for different markers for spindle formation to study cytokinesis.

In addition, a number of UPS components were identified, including E2 and E3 enzymes, that could potentially be studied using *in-vitro* ubiquitination assays to gain insights into the E3-ligase complexes formed with KLHL7, and possibly verify novel substrates from the BioID dataset. Specific substrates of the KLHL7 complex could be validated by adding E1-ubiquitin activating enzyme, one of the two identified E2-ubiqtuin conjugating enzymes, one of the E3-ligases identified, KLHL7 as the adaptor, potential substrates and ATP. The substrates that get poly-ubiquitinated will be specific to the E3-ligase complex associated with KLHL7. Validating these substrates can provide novel insights into biological mechanisms of the UPS and how this specific complex ubiquitinates its substrates.

Also, the patient cells can be used as a model system to extract protein to assess whether or not the mutation in KLHL7 leads to accumulation of a specific substrate. For instance, obtaining antibodies for potential substrates and running western blots can determine whether a particular protein is accumulating in the mutant cell line.

It will also be interesting to try and understand how the missense variants that cause RP-42 differ to the biallelic variants that cause a recessive neurodevelopmental disorder in terms of biological functions and protein interactions.

Chapter 6

A novel missense mutation in *MAL* is associated with a rare leukodystrophy similar to Pelizaeus-Merzbacher disease

6.1 Introduction

This chapter summarises the identification of a missense variant in *MAL* that is associated with a recessive neurodevelopmental condition, characterised by cerebellar atrophy, neurodevelopmental delay and central hypomyelination/ leukodystrophy. Leukodystrophies (LD) are genetic disorders affecting the white matter of the central nervous system (CNS). The majority of LD present with motor symptoms, chiefly delay in the achievement of motor milestones. Cognitive impairment and seizures can also be present. Early involvement of the corticospinal tracts leads to a pattern of central weakness and evolving spasticity. Movement disorders are a common feature: loss of cerebellar volume can lead to a progressive ataxia, whilst involvement of the deep gray nuclei can result in dystonia or chorea. Genetic forms of LD display considerable locus heterogeneity with numerous different causative genes. One of the most common causes of hypomyelinating LD is mutations in *PLP1,* which results in the classic disorder, Pelizaeus-Merzbacher disease (Osorio and Goldman, 2018). In this study, a novel mutation in *MAL,* encoding a gene product that interacts with PLP1, was identified as a cause of a neurodevelopmental condition characterised by central hypomyelination and cerebellar atrophy, similar to Pelizaeus-Merzbacher disease.

This chapter outlines the library preparation for Whole Exome Sequencing of affected individuals, the follow-up functional characterisation of the variant and its association with neurodevelopment and white matter abnormalities. A clinical overview and variant filtering strategy is given, as for each case discussed previously, and the candidate variant is then functionally assessed by modelling the mutation in established cell-lines.

6.2 Variant identification by Whole Exome Sequencing

6.2.1 Patient Recruitment

Family ND13 was ascertained and recruited as part of the Jules Thorn genetics research study, as described elsewhere (see section 2.2.2). All family members that were included in the study gave informed consent for research and their DNA was collected from blood. A skin biopsy was also collected from affected individual JT778 (see method 2.2.4). Affected individuals were phenotyped by Clinical Geneticists in regional hospitals (see method 2.2.1).

6.2.2 Clinical Ascertainment

Family ND13 is a consanguineous UK family of Pakistani origin. Two affected individuals presented with significant developmental delay and learning disabilities, feeding difficulties, cerebellar developmental defects, and cortical white matter abnormalities similar to Pelizaeus-Merzbacher disease. The pedigree structure was compatible with both autosomal recessive and X-linked inheritance. The affected individuals were screened for a panel of genes known to cause white matter abnormalities, including *PLP1*, but no variants were found. The details of the diagnostic screening panel are described here: (https://www.leedsth.nhs.uk/a-z-of-services/the-leeds-genetics-laboratory/constitutional-genetics/molecular-genetics/by-disorder/leukodystrophy/).

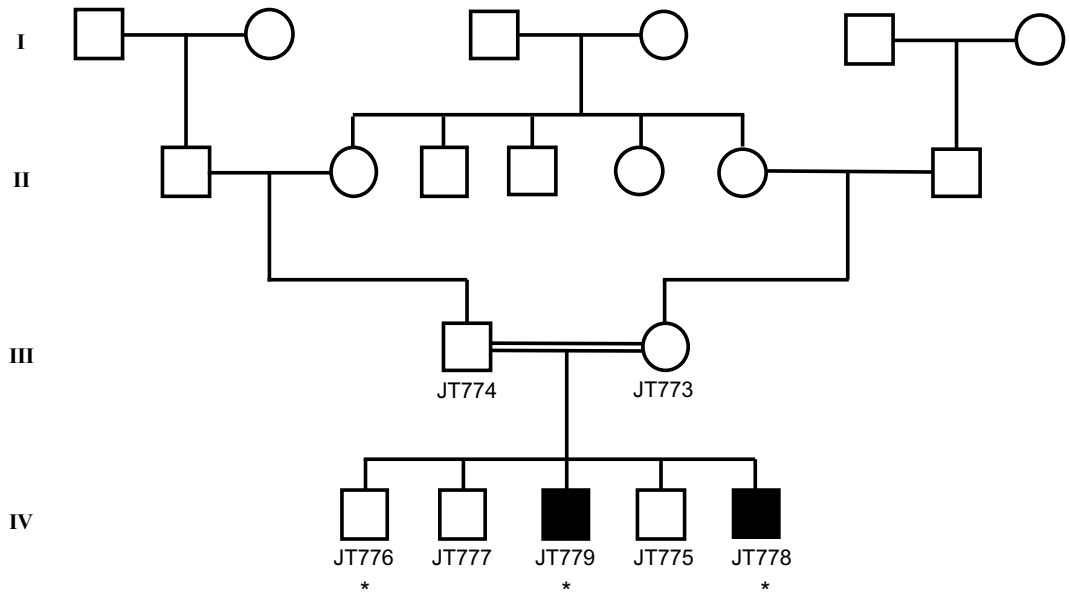


Figure 6-1 Pedigree of family ND13

Pedigree outlining four generations of the family and the consanguineous union. Individuals recruited to the study are indicated by anonymized codes and individuals whose DNA samples have been used for genetic investigations using WES are marked with a star (*).

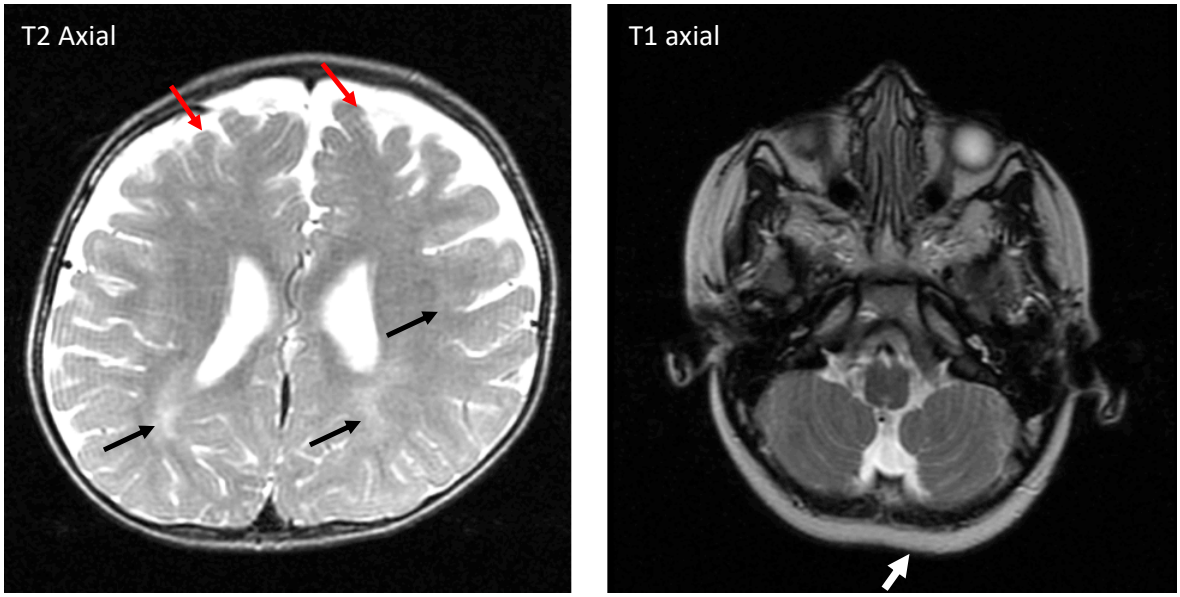


Figure 6-2 MRI scan from individual JT779

MRI scans from Individual JT779. Left panel (T2 Axial view); mild central hypomyelination (black arrows) and prominence of extra axial spaces indicative of cerebral volume loss (red arrows). Right panel (T1 Axial view); prominence of cerebellar folia, and vermian hypoplasia, indicating cerebellar volume loss. Flattening of the posterior aspect of the skull vault indicative of plagiocephaly.

6.2.3 Autozygosity mapping

Autozygosity mapping was performed using WES data. Precisely, the vcf file prior to SNP filtering (.raw.vcf) from the individuals that have been taken forward for WES were used (see section 2.2.10.6). Homozygous regions shared by the two affected individuals, but absent from the unaffected sibling, are summarised in **Table 6-1**.

Chromosome	Start	End	Length
2	91873492	121740505	29867013
7	102755330	131347374	28592044
2	33540172	61175312	27635140
2	69409029	88926729	19517700
5	135178081	147286054	12107973
1	234563060	241767930	7204870
8	11142389	15978063	4835674
7	98460824	102136550	3675726
5	178634619	180582604	1947985
11	45672261	47238522	1566261
7	1518336	2515382	997046
6	33036388	33136575	100187
6	31236853	31239010	2157

Table 6-1 Common regions of homozygosity amongst affected individuals.

The homozygous regions are listed in descending order based on the size of each region. The human genome assembly GRCh37/hg19 was used.

6.2.4 Whole Exome Sequencing

DNA samples from all the individuals indicated (**Figure 6.1**) were taken forward for WES library preparation using the Illumina QXT protocol (section 2.2.10). During library preparation, Bioanalyzer traces were used to assess the size and quantity of the DNA fragments at post-shearing and post-hybridization stages (Appendix D). The samples were pooled, aiming for a total of ten samples per lane, and sequencing was performed on the Illumina HiSeq 3000 platform.

Prior to data analysis, the fastq files and subsequent bam files generated were also assessed for the quality of sequencing. An in-house bioinformatics pipeline was used for data analysis (section 2.2.10.3), assuming a recessive mode of inheritance that allowed variants to be prioritized in previously identified homozygous regions. In addition, segregation analysis enabled the filtering out of variants that were not compatible with Mendelian segregation of an autosomal recessive condition. Additional analysis was also performed for any X-linked variants, based on the pedigree structure, with no variants identified that could be interpreted as pathogenic. **Table 6-2** summarises the variants identified after filtering with a CADD score >15. **Figure 6-2** illustrates the presence of the candidate variants in the homozygous regions shared by the affected individuals, with the best candidate variant in *MAL* being present in the largest homozygous region. All candidate variants were further processed for segregation analysis.

Variant Information				
Gene	**Location**	**Variant**	**Protein change**	
PNPT1	2:55907990	c.517G>A	p.A173T	
ZNHIT1	7:100867096	c.416G>A	p.R139H	
MAL	2:95715390	c.326C>A	p.A109D	
Pathogenicity prediction				
Gene	**Condel**	**Polyphen2**	**SIFT**	**CADD score**
PNPT1	deleterious (0.572)	possibly damaging (0.554)	deleterious (0.02)	34
ZNHIT1	deleterious (0.600)	possibly damaging (0.621)	deleterious (0.02)	32
MAL	deleterious (0.621)	possibly damaging (0.839)	deleterious (0.08)	28.7
Frequency and function				
Gene	**MAF gnomAD**	**Protein Function**	**OMIM**	**In Homozygous region**
PNPT1	0.0001355 (0.001% in South Asians)	Polyribonucleotide nucleotidyltransferase; implicated in RNA processing and degradation	Combined oxidative phosphorylation deficiency 13	Yes
ZNHIT1	0.00009172 (0.0006% in South Asians)	Zinc Finger HIT-Type Containing 1; involved in p53-mediated apoptosis	____	Yes
MAL	Absent	Highly hydrophobic integral membrane protein implicated in myelin biogenesis and function	____	Yes

Table 6-2 Homozygous variants identified in family ND13

Homozygous variants in *PNPT1* (NM_033109.5)p, *ZNHIT1* (NM_006349.3) and *MAL* (NM_002371.4) were annotated based on the human genome assembly GRCh37/hg19.

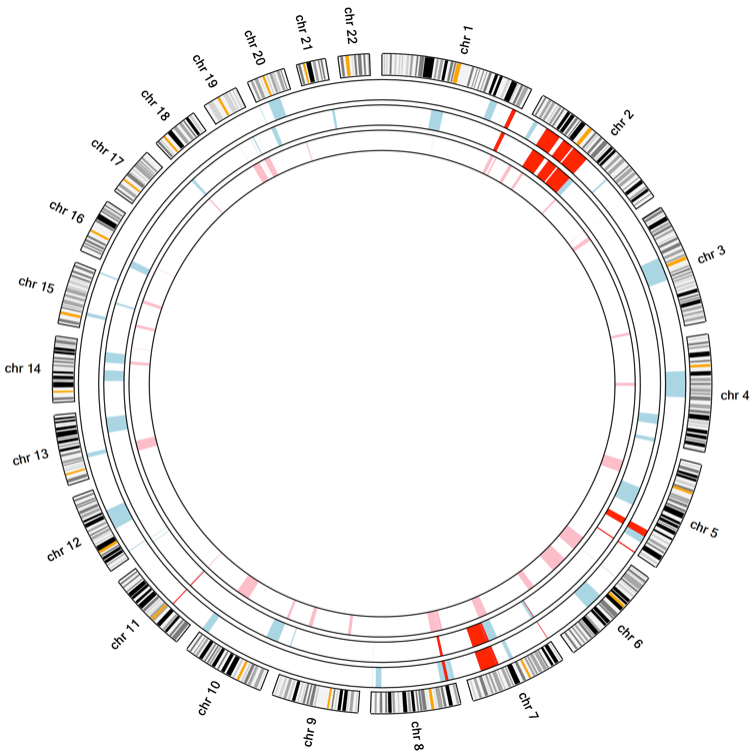


Figure 6-3 Ideogram illustrating the homozygous regions shared by the affected individuals in family ND13

The ideograms illustrate the homozygous regions of the two affected individuals. The regions highlighted in red represent the shared homozygous regions, that are absent from the unaffected individual (regions in pink). The variant in *PNPT1* localises within the first homozygous region on chromosome 2, the variant in *MAL* in the last and biggest homozygous region on chromosome 2, and the variant in *ZNHIT1* in the homozygous region on chromosome 7. The human genome assembly GRCh37/hg19 was used.

6.2.5 Variant confirmation using Sanger Sequencing

The homozygous variants summarised in **Table 5-2** were the only variants passing the biallelic filtering criteria from the bioinformatics analysis. All three variants are within a homozygous region, therefore none of them can be prioritised based on autozygosity mapping, as the causative variant is more likely to segregate within a homozygous haplotype. Segregation analysis was therefore used to determine if the segregation of all variants within the family's pedigree were compatible with autosomal recessive inheritance. Sanger sequencing was used to confirm the c.517G>A variant in *PNPT1*, c.416G>A variant in *ZNHIT1*, and c.326C>A variant in *MAL* that were identified by WES. All members of the family with the anonymised codes were sequenced using the relevant primers (see section 2.1.4). The variant in *MAL* and *ZNHIT1* segregated according to the family pedigree, consistent with autosomal recessive inheritance. The variant in *PNPT1* did not segregate and was therefore excluded as a candidate for this condition. Segregation analysis is outlined in **figure 6-4**.

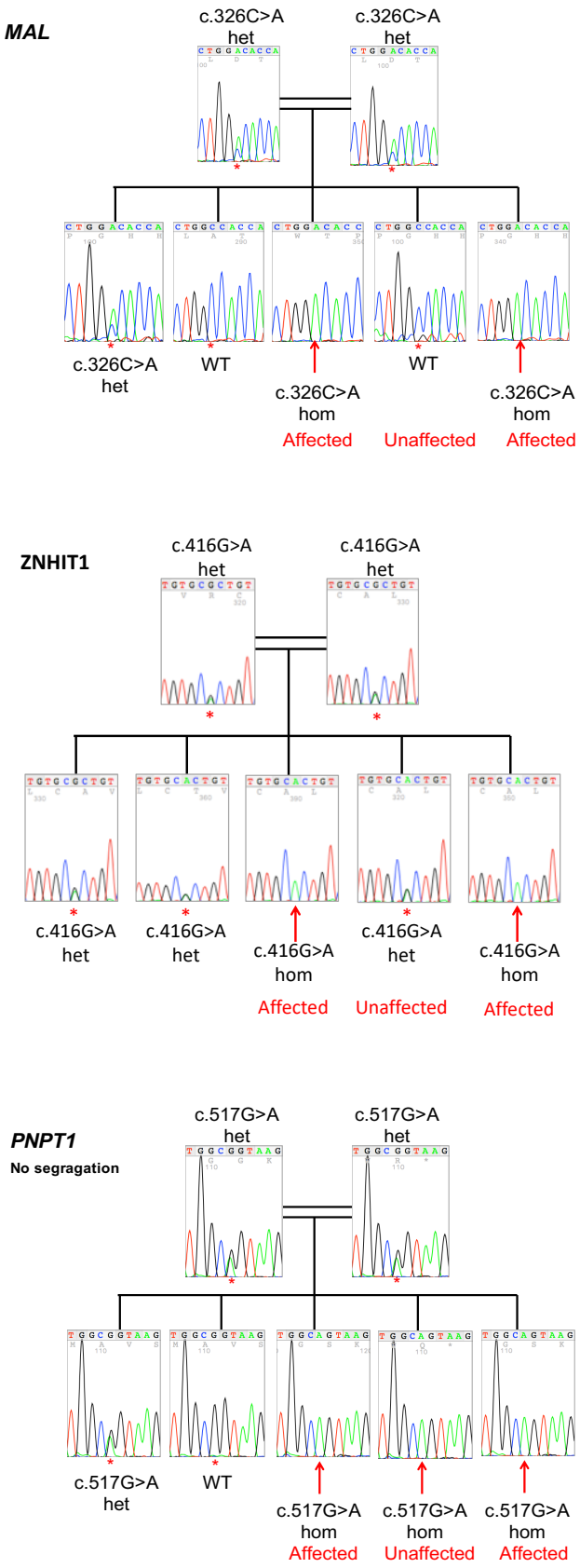


Figure 6-4 Segregation analysis for the variants identified in ND13

The electropherograms above outline the segregation analysis for all the members of the family for whom DNA sample was available. Each pedigree shows the sequencing for the variants identified by WES, with the segregation of the variant in *PNPT1* not compatible with an autosomal recessive inheritance pattern in this family pedigree.

The pathogenicity predictions for the two variants in *MAL* and *PNPT1* were very similar, therefore no conclusion could be drawn based on pathogenicity prediction scores alone. However, it is worth noting that the variant in *MAL* was completely absent from gnomAD, whereas the variant in *PNPT1* has an allele frequency of 0.006% in South Asians.

As a result, a literature search was carried out on both genes to assess whether any of the two would be a better candidate gene, based on functional relevance to the disease phenotype or protein function. With regard to *ZNHIT1*, little is known about its function but there is some indication that it is involved in cancers. It seems to be downregulated in breast cancer cell lines and overexpression seems to inhibit tumorigenesis by regulating PI3K/Akt/mTOR pathway via PTEN-mediated inactivation (Cui et al., 2019).

On the other hand, MAL (myelin and lymphocyte protein) is a 16kDa membrane proteolipid with four transmembrane domains. It seems to have a role in vesicular trafficking in polarized cells (Marazuela and Alonso, 2004), specifically involved in apical transport via a direct route from the Golgi apparatus (Puertollano and Alonso, 1999), and in transcytosis to basolateral membrane (Bijlard et al., 2016). MAL also seems to be regulating the distribution of PLP1 (Bijlard et al., 2016), the major myelin-resident protein, by targeting PLP1 to the basolateral membrane where myelin assembly initiates. The distribution and correct assembly of PLP1 protein is a key element for the mechanism of myelin formation (Baron et al., 2015).

The fact that MAL has an association with myelin assembly processes, clearly makes it a better functional candidate gene for future validation work compared to *ZNHIT1*. Furthermore, mutations in PLP1 are a known cause of Pelizaeus-Merzbacher disease, a rare X-linked recessive hypomyelination disorder (Hoffman-Zacharska et al., 2013). Family ND13 have a very similar phenotype to Pelizaeus-Merzbacher disease with a characteristic MRI scan, giving us more confidence that the variant in *MAL* is probably the disease-causing variant. In addition, the most severe forms of Pelizaeus-Merzbacher disease are caused by missense mutations (Hoffman-Zacharska et al., 2013), just like the one in *MAL*, often leading to ER aggregates as the cause of the disease (Inoue, 2017). As discussed below, we observed the same mechanism of action for the missense variant in *MAL*.

In addition, MAL has a high expression pattern in pre-mature Schwann cells as well as in oligodendrocytes and mature Schwann cells, and the protein predominantly localizes in compact myelin (Frank, 2000). Based on a major study by Zhang et al., focusing on RNA-sequencing data from different cellular classes of the brain, MAL seems to have a very high expression pattern in myelinating oligodendrocytes, but once myelination is completed *MAL* levels decrease (Zhang et al., 2015). Potentially, MAL could be essential during the neurodevelopmental period, whilst oligodendrocytes are forming myelin sheaths.

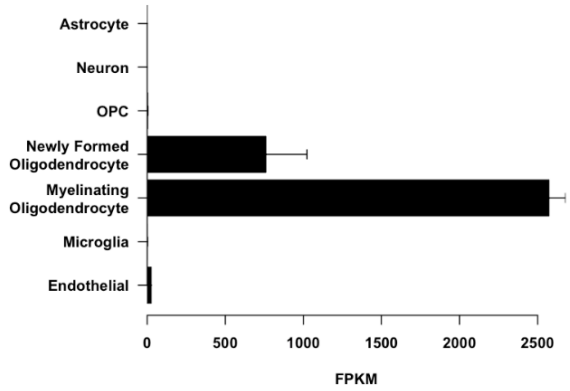


Figure 6-5 RNA-expression of *MAL* in the brain.

The figure above illustrates the expression levels of *MAL* in different types of cells in the brain, mainly being present in oligodendrocytes. The expression level is proportional to the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) value. In other words, the expression of a particular transcript in RNAseq is proportional to the number of cDNA fragments. Figure adapted from Zhang et al., 2015.

Based on the evidence found in the literature for *MAL*, as opposed to *ZNHIT1*, *MAL* was considered a better functional candidate gene for the phenotype of family ND13. MAL also appears to regulate the distribution of PLP1, encoded by a gene that is mutated as a cause of the same phenotype. For that reason the variant c.326C>A in *MAL* was modelled using computational tools (see section 6.4) and further functionally characterised using cellular modelling of the variant to assess the possible pathomechanisms of disease (see section 6.5).

6.3 Identification of additional patients with variants in *MAL*

6.3.1 Sanger sequencing local leukodystrophy patients

Regional patients, at the Leeds Teaching Hospitals NHS Trust, with clinical features similar to Pelizaeus-Merzbacher disease and a negative result on the NHS leukodystrophy panel, were recruited for *MAL* screening under the ethical approval from the NRES Committee Yorkshire & The Humber, South Yorkshire (REC reference 11/H1310/1). Thirty-three independent cases were screened for *MAL* variants using Sanger sequencing (see section 2.2.9), with primers spanning all four exons (see section 2.1.4). No biallelic variants were identified in the coding region of *MAL*.

An additional case was also sent by Dr Paul J. Benke from Joe DiMaggio Children's Hospital in Florida, with the affected individual being part of a Pakistani consanguineous family, sharing the same phenotype with patient in ND13. Screening did not reveal any variants in the coding region of *MAL*.

6.3.2 Sharing data with other centers and databases.

The genetic information regarding family ND13 (genotype and phenotype) was shared with national and international centers that specialize in white matter abnormalities, again with no additional cases with pathogenic variants in *MAL* being identified.

The genotypic and phenotypic information of ND13 were therefore uploaded on to the Decipher database (https://decipher.sanger.ac.uk/), with one match being identified. This cases was identified via Gene Matcher, but the variant in *MAL* was heterozygous and the phenotype did not match that of family ND13.

6.4 Prediction of impact of MAL p.A109D variant on protein

Prior to investing into further studies into functional characterisation of the MAL p.A109D variant, computational tools were used to predict the impact that this variant may have on the overall protein structure. Such tools may be informative and provide more evidence to support a hypothesis for further investigations. The main reason for performing such prediction tests was due to the fact that a small non-polar amino acid with hydrophobic side chain (alanine), was replaced by a slightly larger, negatively charged amino acid with a hydrophilic side chain (aspartic acid). Considering that this change occurred within a transmembrane domain of the MAL protein, the impact of the substitution on the protein structure was expected to be severe as a hydrophilic amino acid was now incorporated into a hydrophobic domain that spans the cell membrane. An illustration of the location of the change is shown in **Figure 6-6a**.

The first computational tool used was the Transmembrane Helix Prediction TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/). TMHMM allows for prediction of transmembrane helices based on a hidden Markov model. The model is trained on a set of 160 transmembrane proteins, and 645 non-membrane proteins as negative controls, giving it power to distinguish between soluble and membrane proteins with high degree of accuracy (Krogh et al., 2001). Apart from detecting the transmembrane helices, it can even differentiate which part of the protein is cytoplasmic (in) or non-cytoplasmic (outside). TMHMM output of the wildtype sequence illustrated the expected behavior of the MAL protein, with four domains spanning the membrane (**Figure 6-6b**). However, output from the mutant sequence suggested that the third transmembrane domain, that carries the missense variant, is no longer incorporated in the membrane. Instead, this transmembrane helix appears to remain on the non-cytoplasmic (outside) region. This supports the original hypothesis that the replacement of the hydrophobic alanine with a charged amino acid will have a deleterious effect on the transmembrane domain. This supports the conjecture that the p.A109D variant is likely to be the causative variant in family ND13.

The second computational tool used was HeliQuest (http://heliquest.ipmc.cnrs.fr/), a software that can identify helices by screening a particular amino acid sequence against a biobank, searching for segments that possess similar features to a helix (Gautier et al., 2008). It recognises these segments by calculating the physiochemical properties and amino acid composition of each helix from the biobank. HeliQuest uses algorithms that can also determine various properties such as amino acid composition per helix, hydrophobicity, hydrophobic phase and net charge. The server also allows the user to

manually mutate particular amino acids and to calculate the impact of the mutation on the formation of a helix. Wildtype and mutant MAL sequences were input on HeliQuest in an attempt to predict the impact of p.A109D variant on the formation of the alpha helix. The result revealed a potentially disrupted helix, as the hydrophobic phase was severely reduced as shown in **figure 6-6c**. The overall hydrophobicity (H) of the mutant sequence was decreased compared to the wildtype, the overall net charge decreased and an additional charged residue was introduced potentially affecting the formation of the alpha helix.

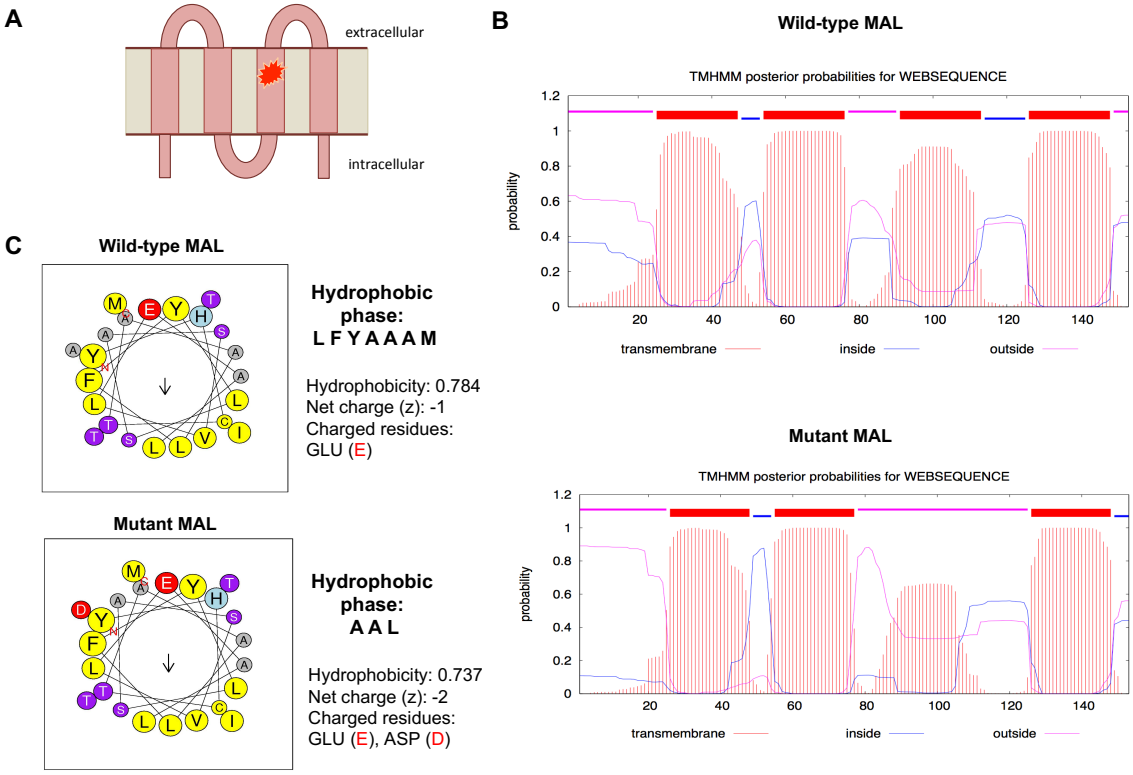


Figure 6-6 Computational prediction of p.A109D variant

A – Representation of the orientation and helix organization of the MAL protein spanning the membrane. The red star indicates the location of the p.A109D variant in the third transmembrane domain. B – Graphical output of the TMHMM Transmembrane Helix Prediction tool. Top image represents the wildtype protein, clearly showing how each domain spans the membrane in accordance to the predicated organization of MAL. Bottom image shows the prediction for the mutant protein, whereby the third transmembrane domain is impacted by the p.A109D and is no longer able to be incorporated into the membrane. C – Illustration of the output from HeliQuest showing the alpha helices that comprise the p.A109D variant. On the left side of each helix is the location of the variant that is replaced from an alanine to an aspartic acid (shown in red on the bottom helix). Some key outputs from this tool are summarised, and the hydrophobic phase is outlined on the side of each helix.

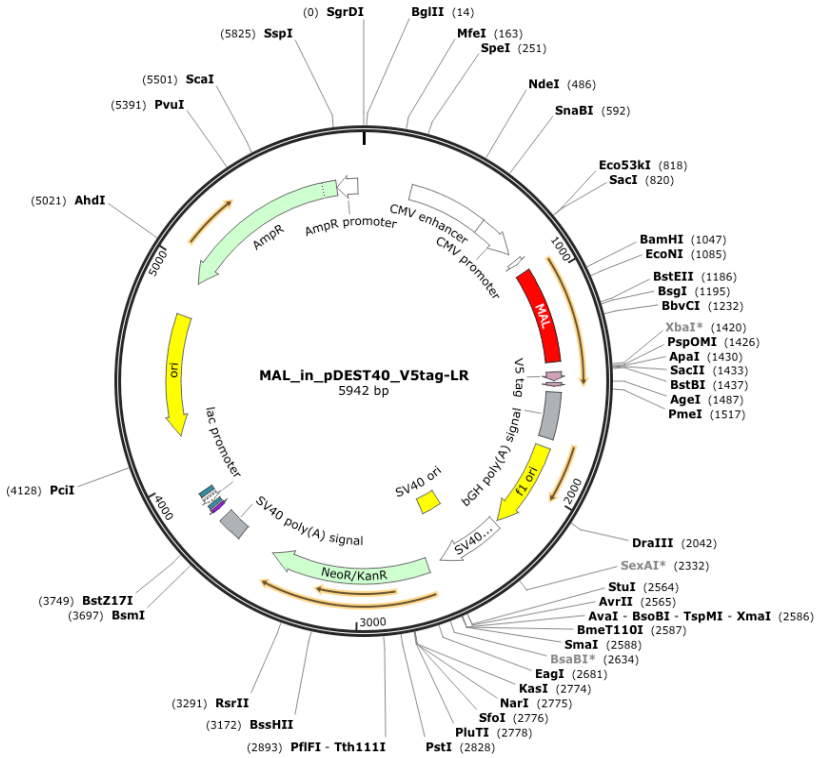
The prediction analyses suggested that the p.A109D variant is expected to have a significant impact on MAL protein organisation and function, and therefore is likely to be pathogenic. The computational analyses provided confidence to proceed to further functional studies as described below in this chapter, in order to validate p.A109D as a causative pathogenic variant for the described leukodystrophy.

6.5 Functional characterisation of MAL p.A109D variant

6.5.1 Construction of *MAL* expressing plasmids

Using the gateway cloning technique (see section 2.2.11.1) a *MAL* entry vector pENTR223 (DNASU, plasmid HsCD00505233) was cloned into destination vectors containing either eYFP and c-Myc tags (pDEST504) or V5 and 6His tags (pDEST40), as outlined in **figure 6-7**. Both of these plasmids were also mutated for *MAL* c,326C>A using site-directed mutagenesis (see section 2.2.11.3) and all plasmids were sequenced verified before used.

A



B

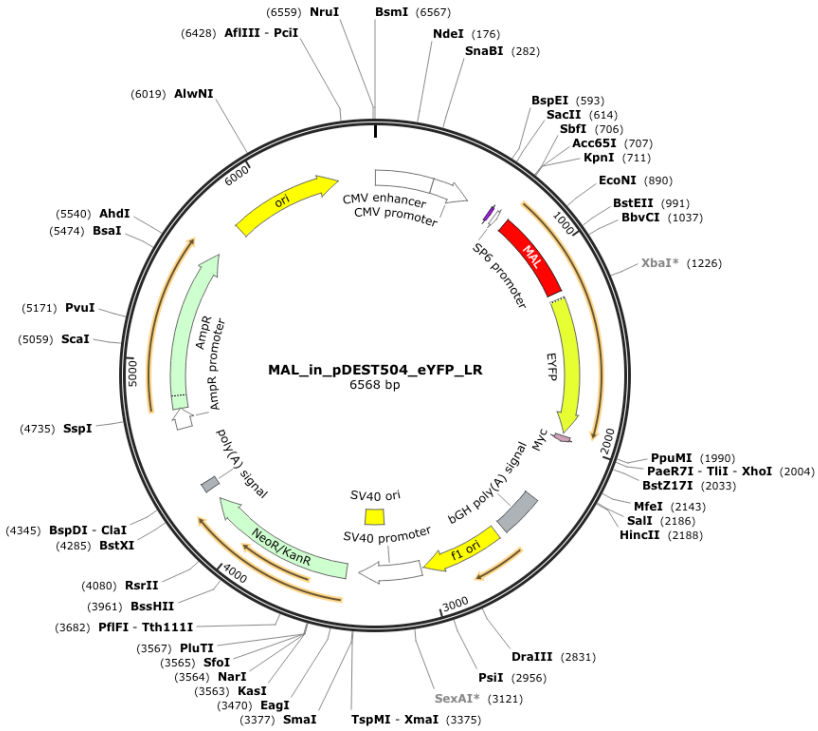


Figure 6-7 Maps for MAL expressing plasmids

A – Plasmid map of *MAL* tagged with V5 and 6 x His at the c-terminal of the protein (called MAL_*V5*). B – Plasmid map of *MAL* tagged with eYFP and c-myc at the c-terminal of the protein (called MAL_*eYFP*).

The plasmids cloned for the purpose of this study were used for overexpression of the wildtype and mutant MAL protein in the Madin-Darby Canine Kidney cells (MDCK) cell line, enabling various functional assays such as immunoprecipitation experiments and immunofluorescence microscopy. The choice of the MDCK cell line was primarily because these cells are a commonly used mammalian model of cell polarity and formation of adherens junctions (Balcarovastander et al., 1984). MDCK cells provided a good model to study our hypothesis, as transcytosis occurs following cell polarisation, enabling us to study the role of MAL in intracellular transport of PLP1 vesicles during myelin formation.

6.5.2 Mutant MAL protein aggregates in the endoplasmic reticulum

6.5.2.1 MAL p.A109D aggregation revealed by imaging

The first assessment of the MAL p.A109D variant was performed by live cell imaging (see section 2.2.18). Wildtype and mutant MAL-eYFP fusion proteins were overexpressed (see section 2.2.15) in MDCK cells cultured in 35mm 4-well imaging microplates. Four hours post-transfection, the transfecting complexes were washed off and fresh media was added to the cells prior to live cell imaging on a Nikon BioStation IM (at magnification of x20). The cells were imaged for a period of 24 hours, allowing sufficient time for MAL protein tagged with eYFP to be produced and detected. Live cell imaging revealed eYFP signal from the wildtype protein in perinuclear regions, possibly localized to the endoplasmic reticulum (ER), which is expected at a certain level due to the overexpression of this protein. Wildtype protein also revealed a distinct localisation at the membrane of the cell, in the expected localization of MAL. However, mutant MAL protein was exclusively localized as large and prominent aggregates in perinuclear regions without any membrane localization (**Figure 6.8**). This suggests that mutant MAL could aggregate in the ER, potentially due to the severe impact of the p.A109D variant on overall protein organisation.

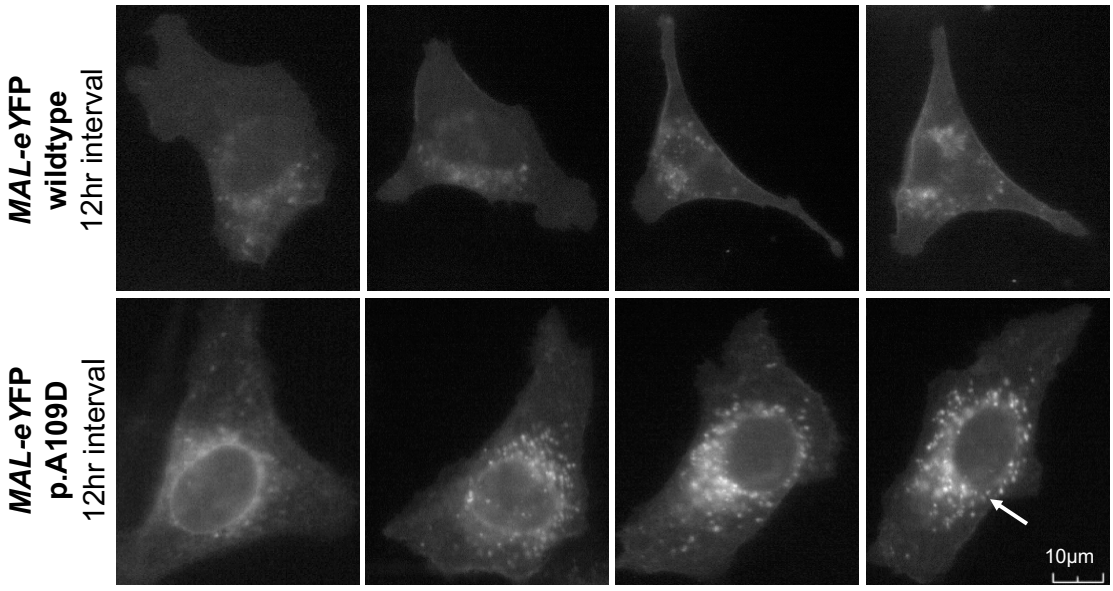


Figure 6-8 Live cell imaging of MAL wildtype and p.A109D mutant protein

MDCK cells transfected with wild-type and mutant *MAL*-eYFP, visualised using the BioStation for a total of 24 hours. The top panel shows the wildtype MAL protein localised at the membrane of the cell, and the bottom panel shows the mutant MAL protein forming aggregates, potentially in the endoplasmic reticulum, as indicated by the arrow. The figure shows representative images from two independent biological replicates.

Live cell imaging was used as an initial experiment to visualise the behaviour of wildtype and mutant MAL proteins, therefore the data were only processed as images and were not further analysed at this point. This was due to limitations such as magnification and multicolour imaging of the Nikon BioStation IM. Therefore, it was decided to use other means of imaging, that allowed higher resolution and multicolour staining, in order to prove that the aggregation observed is indeed localised in the ER.

Immunofluorescence microscopy and confocal imaging was used for this purpose (see section 2.2.17). MDCK cells were again transfected with wildtype and mutant eYFP tagged MAL protein, and 24 hours post-transfection cells were fixed with PFA and stained with antibodies against calreticulin, a marker of the ER, and DAPI (see section 2.1.5). Imaging was carried out on a Nikon A1R confocal microscope, at x100 magnification and images were analysed using the co-localisation plug-ins on Fiji. The confocal images revealed the same pattern of staining for wildtype and mutant MAL proteins, again showing the distinct formation of aggregates for the mutant protein. Using the calreticulin staining, and merging the channels, the mutant MAL_eYFP protein appeared to colocalise with the endoplasmic reticulum marker, supporting the conjecture that mutant MAL is unable to be correctly trafficked from the ER (see **figure 6.9**).

To further validate this observation, 100 cells were analysed from three independent biological replicates and the colocalization between the green channel (MAL protein) and the red channel (ER marker calreticulin) was assessed using the colocalization threshold plug-in in Fiji. For this analysis z-stacks of each image taken from the confocal had to be uploaded on Fiji and a region of interest had to be set around each cell analysed. The colocalization analysis provided an Rcoloc value for each cell, and these values were used for statistical analysis. The Rcoloc value is directly proportional to the level of colocalization between the two channels. Statistical analysis revealed a statistically significant difference in ER colocalization between the wildtype and p.A109D MAL proteins, proving the initial observation of mutant MAL forming aggregates in the ER (see **figure 6-10**).

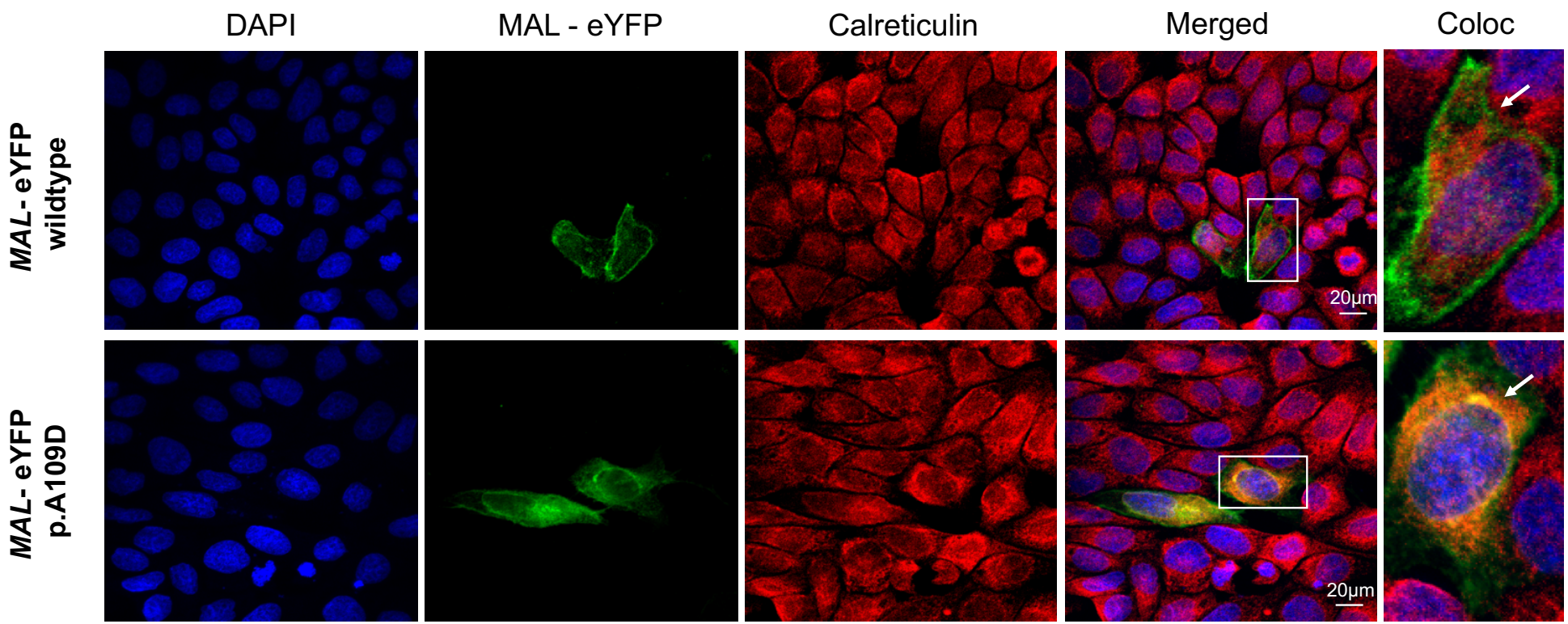


Figure 6-9 Immunofluorescence microscopy of wildtype and p.A109D MAL protein

Immunofluorescence staining (IF) and confocal microscopy imaging of wildtype and mutant MAL eYFP tagged-protein overexpression in MDCK cells. PFA fixation and co-staining with calreticulin (red) and DAPI (blue), and imaging at x100 magnification. Magnified insets are indicated by the white frames. Merging of channels allowed for colocalization of mutant MAL_eYFP and calreticulin to be visible (orange colour, indicated by the arrow), supporting the hypothesis that mutant MAL protein aggregates in the ER. Wildtype MAL localizes at the plasma membrane as expected (indicated by the arrow). IF experiments were performed in three independent biological replicates, of three technical replicates each, and a total of 100 cells were analyzed. Scale bar = 20µm.

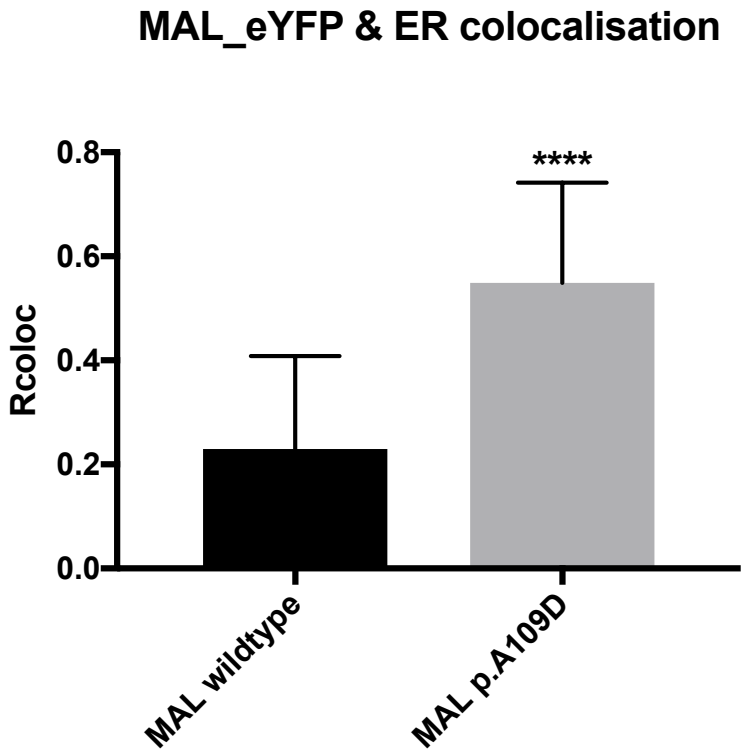


Figure 6-10 Colocalisation analysis using Fiji
Analysis of MAL protein and ER colocalization using Fiji. Rcoloc values calculated for each transfected cell, indicating a significant difference in ER colocalization with the mutant MAL protein compared to the wildtype. Analysis included a total of 100 cells, from three independent biological replicates. Statistical test was performed by a two-tailed Student t-test (****p<0.0001). Error bars indicate s.e.m.

As predicted by the computational analyses, the variant p.A109D was expected to impact the organisation of the MAL protein, as well as the correct incorporation of it within the cellular membrane. As shown by the preliminary functional work described above, the missense variant seems to severely affect protein folding thereby leading to protein aggregates being formed in the ER.

Apart from its many other roles, the ER is also responsible for assessing the quality of newly synthesized proteins for correct folding, prior to the secretion of these proteins to their functionally relevant compartment of the cell. Misfolded proteins that do not pass the quality control tend to form aggregates that lead to ER stress (Ihara et al., 2017). This signal will activate another mechanism known as the unfolded protein response (UPR) (Walter and Ron, 2011) that will try to mitigate and overcome the stress by diminishing translation of the protein, degrading the ER aggregates and increasing the expression of ER residential chaperones that often tend to resolve ER stress by promoting secretion of the aggregates (Schröder M, 2005). If this fails, the increasing ER

stress will eventually lead to apoptosis (Tabas and Ron, 2011). The proteinopathies caused by point mutations that can directly affect the ER via aggregates formation is a common disease mechanism seen in various genetic diseases (Brookes et al., 2017) as well as neurodegenerative diseases (Xiang et al., 2017).

Based on the initial observations, it was hypothesized that the disease mechanism responsible for the phenotype observed in family ND13 was a proteinopathy, whereby the p.A109D variant was causing severe misfolding of MAL. The misfolded protein was causing the formation of ER aggregates and inducing ER stress. Our disease model clearly showed the formation of ER aggregates and the endogenous UPR was potentially unable to resolve the stress from the ER. As a result, the disease model was further used to assess whether or not addition of exogenous ER chaperones would be able to relieve the stress and potentially allow release of protein in the ER aggregates in order to restore correct localization. This does not mean that the correctly trafficked mutant MAL protein will be functional, but it is a key investigation to validate our observations.

To achieve this, we used a chemical chaperone, 4-phenylbutyrate (4-PBA) that is known to relieve ER stress (Mimori et al., 2012). It was previously shown in other proteinopathies, such as amylogenesis imperfecta, that 4-PBA can alleviate symptoms in affected mice by diminishing ER stress and promoting cell survival (Brookes et al., 2014). Since it was assumed that the ER signal shown by the mutant MAL is due to protein aggregates formed as a result of the missense variant, it was decided to assess the alleviation of ER stress by using 4-phenylbutarate (4-PBA).

For this experiment, we performed initial staining to assess whether or not 4-PBA has an effect on removing the ER aggregates. MDCK cells were transfected with wildtype and mutant MAL-eYFP plasmid, showing again the same pattern of protein aggregation in the ER by the mutant MAL. Once 0.5mM of 4-PBA was used, the staining pattern of the mutant MAL-eYFP resembled that of the wildtype, with most of the ER signal now lost (**figure 6-11**). Following this observation, 4-PBA was considered an effective way to relieve ER stress, and the experiment was then repeated with ER marker calreticulin to allow for quantification of the level of ER aggregate removal (**figure 6-12**). For this experiment, 30 cells were visualized and quantified for each condition (wildtype or mutant; treated or untreated) giving a total of 120 cells from three independent biological replicates.

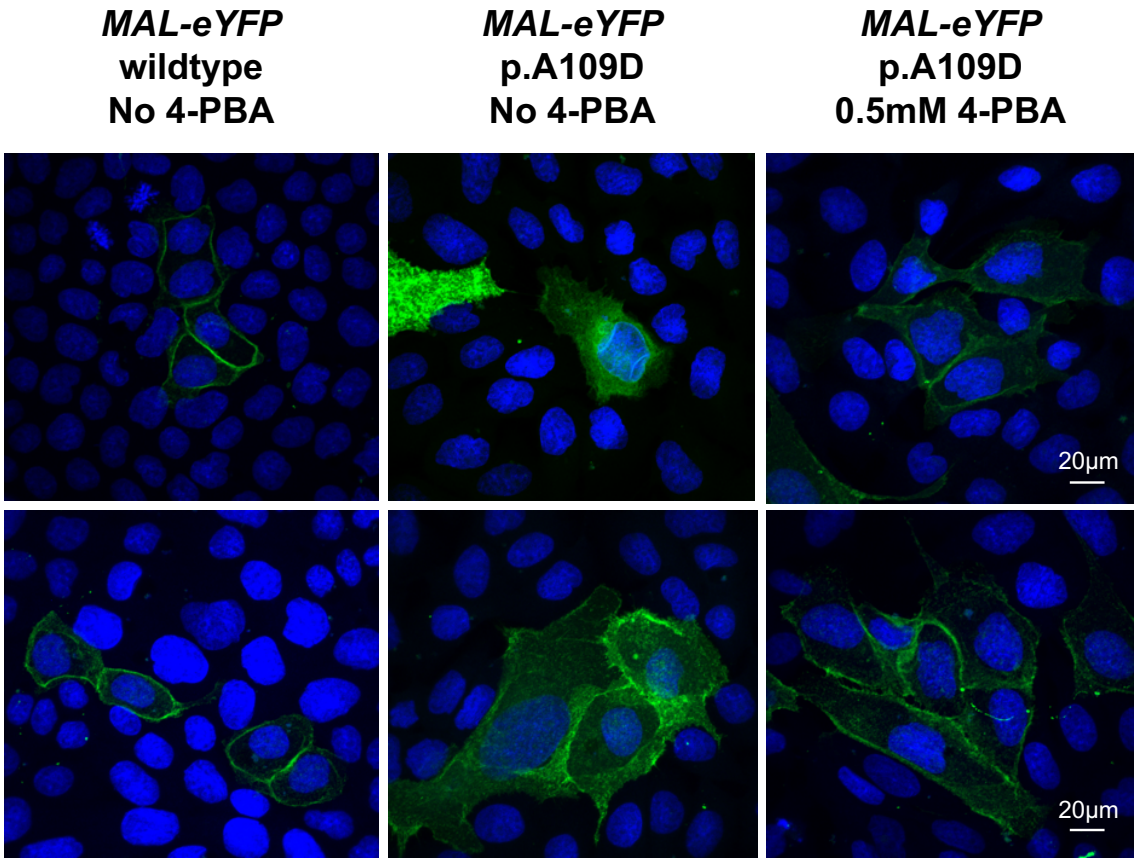


Figure 6-11 Use of 4-PBA to relieve ER stress caused by the mutant MAL.

Immunofluorescence staining (IF) and confocal microscopy imaging of wildtype and mutant MAL eYFP tagged-protein overexpression in MDCK cells. Cell expressing mutant MAL protein were also treated with 4-PBA to check whether or not the use of ER chaperones may have an effect on removing ER aggregation and stress. PFA fixation and co-staining with DAPI (blue) was carried out, and imaging at x100 magnification. Scale bar = 20µm.

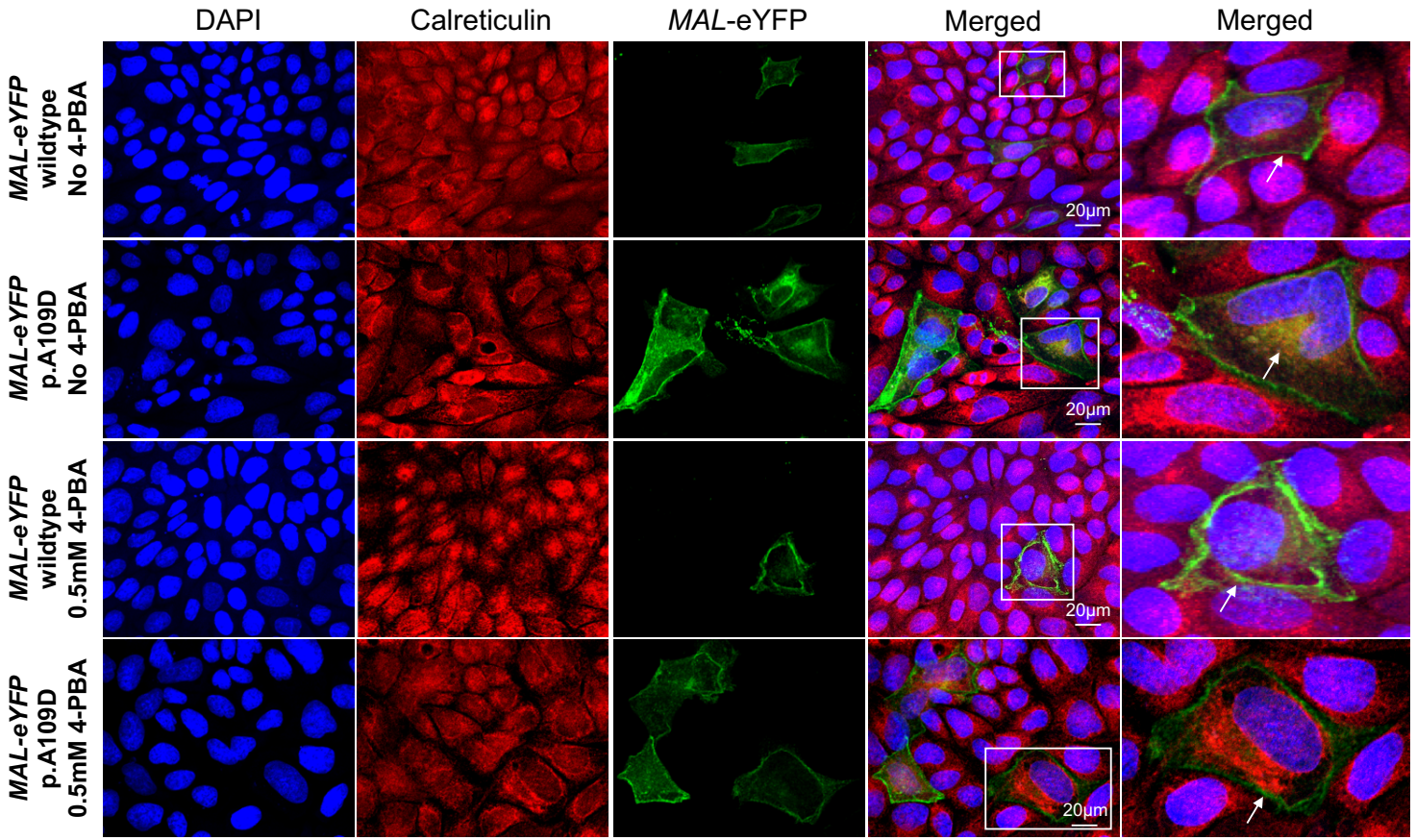


Figure 6-12 Immunofluorescence microscopy of wildtype and p.A109D MAL protein using 4-PBA to reduce ER aggregates.
Immunofluorescence staining (IF) and confocal microscopy imaging of wildtype and mutant MAL eYFP tagged-protein overexpression in MDCK cells, with and without 4-PBA treatment. PFA fixation and co-staining with calreticulin (red) and DAPI (blue), and imaging at x100 magnification. Merging of channels allowed for colocalization of mutant MAL_eYFP and calreticulin to be visible and further quantify the level of ER aggregate reduction upon 4-PBA treatment. MAL localisation shown with the arrows, and magnified cells are indicated by the box. IF experiments were performed in three independent biological replicates, of three technical replicates each, and a total of 120 cells were analyzed. Scale bar = 20µm .

The colocalization between the green channel (MAL protein) and the red channel (ER marker calreticulin) was assessed using the colocalization threshold plug-in in Fiji. For this analysis z-stacks of each image taken from the confocal had to be uploaded on Fiji and a region of interest had to be set around each cell analysed. The colocalization analysis provided an Rcoloc value for each cell, and these values were used for statistical analysis. The Rcoloc value is directly proportional to the level of colocalization between the two channels. Statistical analysis revealed a significant difference between colocalization of ER and MAL in treated and untreated cells expressing p.A109D MAL protein, supporting our initial finding. The statistical analysis for each condition is summarised in **figure 6-13**.

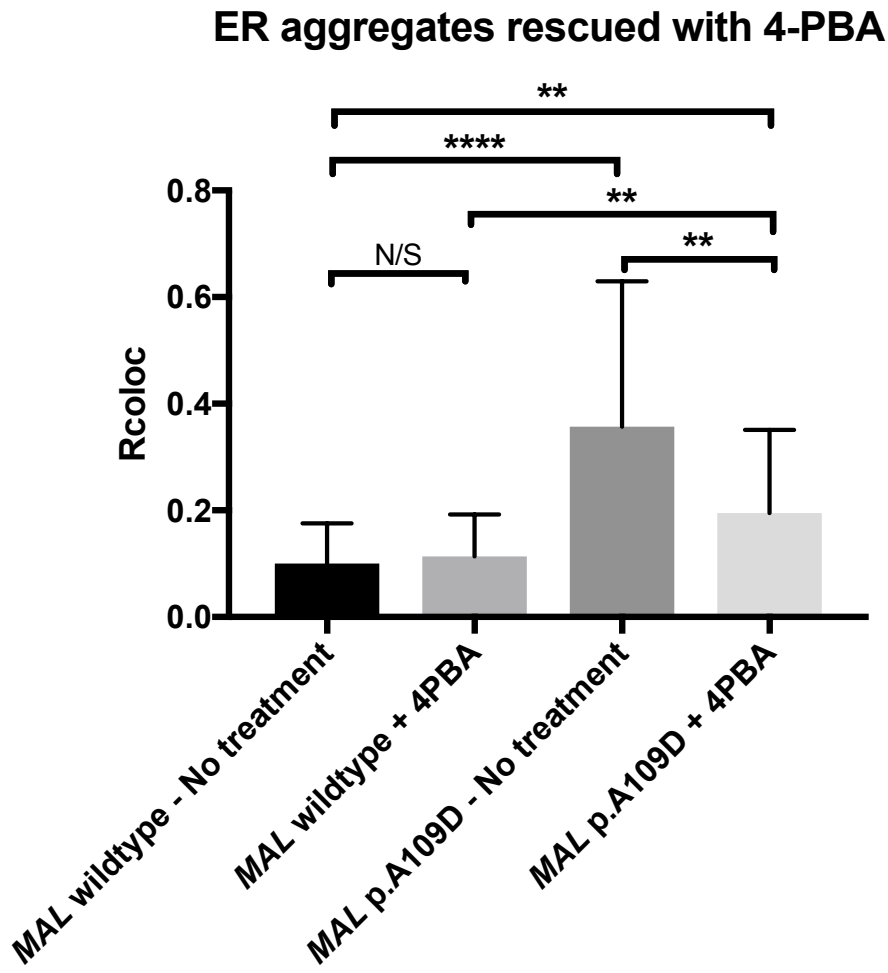


Figure 6-13 Quantification of ER aggregates rescue using 4-PBA

Analysis of MAL protein and ER colocalization, for 4-PBA treated or untreated cells. Rcoloc values for each cell analysed indicate a significant difference in ER colocalization for the treated cells expressing mutant MAL compared to the untreated cells expressing mutant MAL. There was also no significant difference in ER colocalization between wildtype untreated and wildtype treated cells, indicative that 4-PBA is having a specific effect only on ER aggregates and not other non-specific cellular impacts. Statistical tests was performed by a two-tailed Student t-test (ns=not significant; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). Error bars indicate s.e.m.

The result from this experiment revealed a significant difference between the mutant MAL and ER colocalisation in untreated and 4-PBA treated cells, and no significant difference between the wildtype MAL and ER colocalisation in untreated and treated 4-PBA cells, respectively. This means that the 4-PBA does not have any non-specific effects on the cells or overexpression of the protein. Instead it appears to act as a synthetic chaperone and promote correct trafficking of misfolded proteins from the ER, relieving the stress from the mutant aggregates (Yam et al., 2007). When comparing each condition with its corresponding control, there is still a significant difference between wildtype and mutant MAL and ER colocalisation (in treated and untreated conditions), indicating that this treatment does not completely abolish the aggregates from the ER. Despite this, there is a significant decrease in mutant MAL aggregating in the ER, an observation that could potentially lead to a treatment strategy for this family. An improved way to study this will be in modelling the MAL p.A109D variant in a more stable and physiologically relevant setting, for instance using CRISPR/Cas9 technologies, in order to access the effect of 4-PBA on promoting ER secretion of misfolded proteins. This will eliminate any effects that overexpression of a protein might have on ER itself.

The above result supports our initial hypothesis that the pathomechanism of MAL p.A109D is likely to be a proteinopathy, whereby protein misfolding leads to disease pathology (Neumann et al., 2007). In this case, the missense variant in *MAL* seems to be severely affecting protein orientation and conformation, as predicted by computational analyses, leading to ER aggregates formation as the misfolded protein is unable to be correctly trafficked. This effect can potentially lead to induced ER stress and subsequent UPR mediated apoptosis, if the UPR response cannot relieve the stress (Tabas and Ron, 2011, Walter and Ron, 2011). The proposed pathomechanism can be potentially rescued with synthetic chaperones, such as 4-PBA, possibly rescuing the phenotype by relieving ER stress and preventing apoptosis. It has been shown previously that 4-PBA is involved in restoring normal ER secretory function and cell survival over UPR mediated apoptosis (Zode et al., 2015). 4-PBA has been used in various studies and clinical trials, both in vivo and in vitro, with proven potential therapeutic capacities in a range of conformational diseases including spinal muscular atrophy, cancer, cystic fibrosis, Huntington's disease (Iannitti and Palmieri, 2011) and amelogenesis imperfecta (Brookes et al., 2014).

6.5.3 The missense variant p.A109D affects interaction between MAL and PLP1.

MAL is known to be involved in protein sorting in polarised transport, or transcytosis, between the trans-Golgi network and the plasma membrane (Cheong et al., 1999, Puertollano and Alonso, 1999). As discussed previously, it has been shown that MAL acts as a regulator of PLP1, by trafficking PLP1 to the basolateral membrane during myelin formation (Bijlard et al., 2016). MAL is not involved in PLP1 transport from the Golgi to the apical membrane, but it seems to play a key role in redirecting the transport of PLP1 from the apical membrane to the basolateral membrane where myelin begins to form in oligodendrocytes (Bijlard et al., 2016, Baron et al., 2015). It is also shown that PLP1 undergoes a conformational change at the apical membrane before being redirected to the basolateral membrane, whereby the secondary structure of the second extracellular loop is altered (Jung et al., 1996). I hypothesised that MAL could be involved in this process by interacting with PLP1 at the apical membrane, and further investigations on this interaction were undertaken. Importantly, I also investigated the impact of the missense variant in MAL on the MAL-PLP1 interaction.

To assess whether or not p.A109D affects the interaction between MAL and PLP1 proteins, immunoprecipitation methods were used (see section 2.2.20). In summary, the GFP-Trap method (Chromotek) was used that consists of anti-GFP magnetic agarose beads with high affinity for GFP. To perform this method, MDCK cells were co-transfected with a *PLP1*-GFP plasmid and a wildtype or mutant *MAL*-V5 plasmid (see section 2.2.15) and allowed to grow to confluency, to allow transcytosis to be initiated. At that stage, cell lysates were collected and incubated with the GFP-Trap magnetic beads and the beads were then isolated with a magnet. The pull-down proteins were then run on a western blot (see section 2.2.19) and stained with anti-V5 antibody (see section 2.1.5) to assess whether or not MAL is indeed an interactor of PLP1 and if the mutation abolishes this interaction. This experiment was performed in three independent biological replicates.

The quantified whole cell extract signal was normalised to β-actin levels as a loading control, and further compared to the bands from the V5 antibody staining in the immunoprecipitation lane using Image Lab. Analysis revealed 56% loss of the MAL and PLP1 interaction as a result of the missense mutation. This can potentially be explained based on the previous results, as most of the p.A109D MAL protein seems to be aggregating in the ER, thereby never reaching the apical membrane where MAL and PLP1 interact. However, judging from the immunofluorescence imaging, a portion of the overexpressed mutant MAL protein appears to escape ER aggregation, with moderate levels localizing to the membrane. This could potentially explain why this interaction is not completely abolished by the missense mutation (**figure 6.14**). Also, the overexpression of the two proteins could force a false positive interaction, and a more physiologically relevant model could be developed to further this interaction. However, a 56% loss of interaction could be enough to have a biological impact in myelin formation, as not all of PLP1 will be directed to the basolateral membrane for myelin formation, potentially leading to the hypomyelination features observed in the patients.

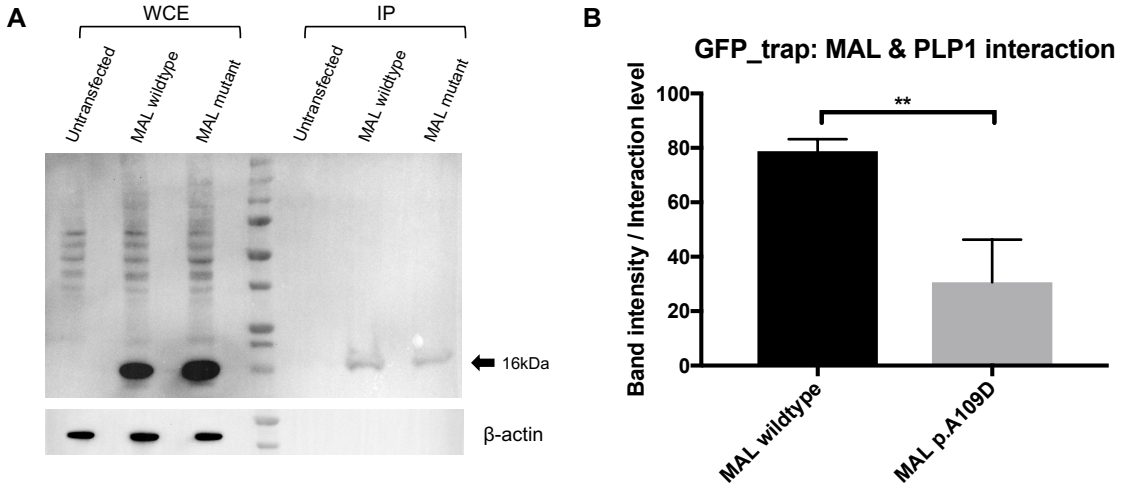


Figure 6-14 The p.A109D variant reduces MAL interaction with PLP1

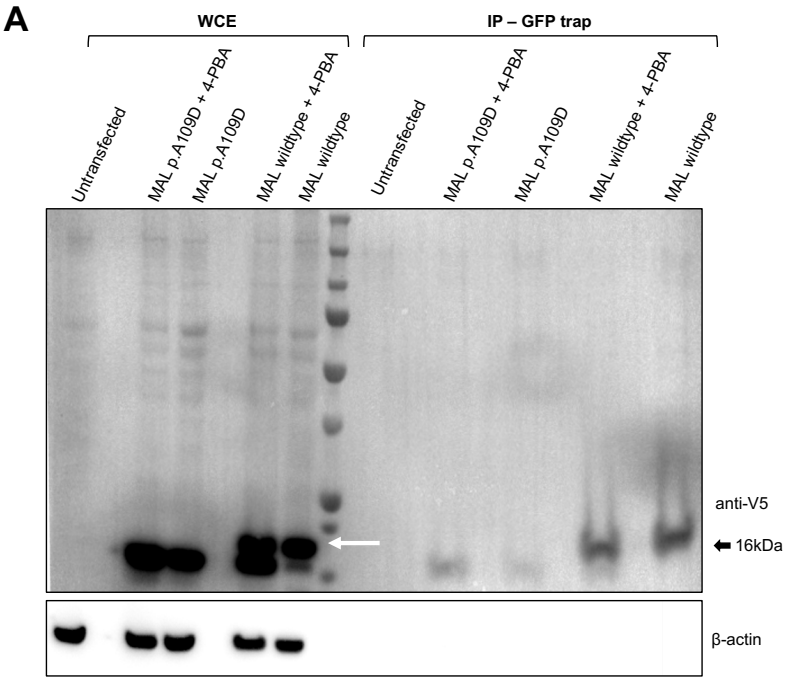
A – Western blot analysis of immunoprecipitation experiment for assessing interaction of PLP1-GFP protein with wildtype and mutant MAL-V5 protein. Pull-down was performed with GFP-beads and antibody staining was performed using V5 antibody, which targets MAL protein (expected size 16kDa). B – Quantification of MAL and PLP1 interaction in wildtype and mutant cases from three independent biological replicates. Student t-test (**p<0.01). Error bars indicate s.e.m.

Using the same technique as above, the MAL and PLP1 interaction was assessed upon treatment with ER chaperone, 4-PBA. Taking into consideration the IF microscopy results, it was hypothesized that 4-PBA would potentially recover some of the mutant MAL and PLP1 interaction, as more of the mutant protein will be released from the ER. The quantified whole cell extract signal was normalised to β-actin levels, and further compared to the bands from the V5 antibody staining in the immunoprecipitation lane using Image Lab. Analysis revealed the interaction between mutant MAL and PLP1 is slightly increased upon the treatment with 4-PBA (**figure 6.15**). This supports my previous observations using IF microscopy, whereby more mutant MAL is able to reach the membrane and thereby interact with PLP1. This experiment was only performed once due to time constraints, but should be repeated at least for three biological replicates for the purposes of quantification and statistical significance testing.

The IP experiments also resolved a second band for wildtype MAL but not for the mutant MAL protein (**figure 6.15**). This minor band seems to be 1-2 kDa larger than the major MAL band, potentially due to post-translational modifications. This is also supported by previous observations, as mutant MAL potentially never leaves the ER in order to reach the Golgi apparatus where most post-translational modifications occur. However, wildtype MAL is able to undergo post translational modifications. Many transmembrane proteins undergo post-translational modifications for functional significance, in order for the protein to be able to be incorporated in the membrane (Souda et al., 2011).

Lipoproteins are post-translationally modified in order to anchor or be targeted to the plasma membrane. These modifications include myristoylation, palmitoylation and prenylation (Bateman et al., 2019). Prenylation is a lipid modification whereby a hydrophobic group is covalently added to cysteine residues near or at the C-terminus of a protein. MAL does not have a cysteine residue at its C-terminus, therefore prenylation is not the expected post-translational modification in this instance. On the other hand, palmitoylation is the reversible addition of fatty acids (16 carbon fatty acid) onto cysteines, and less frequently serine and threonine residues. This reversible mechanism occurs rapidly, facilitating shuttling of protein between the plasma membrane and the Golgi apparatus (Salaun et al., 2010). In addition, myristoylation is another lipid modification process whereby a myristic acid is attached to glycine residues (alpha amino acid) of the N-terminus of the protein. It is thought that it occurs following cleavage of either the methionine residue, if it is followed by a glycine, or by a proteolytic cleavage of a few amino acids at the N-terminus, leading to a glycine at the alpha amino acid position (Udenwobele et al., 2017). Regarding MAL, both myristoylation and

palmitoylation processes can occur post-translationally as a requirement for membrane targeting of a lipoprotein.



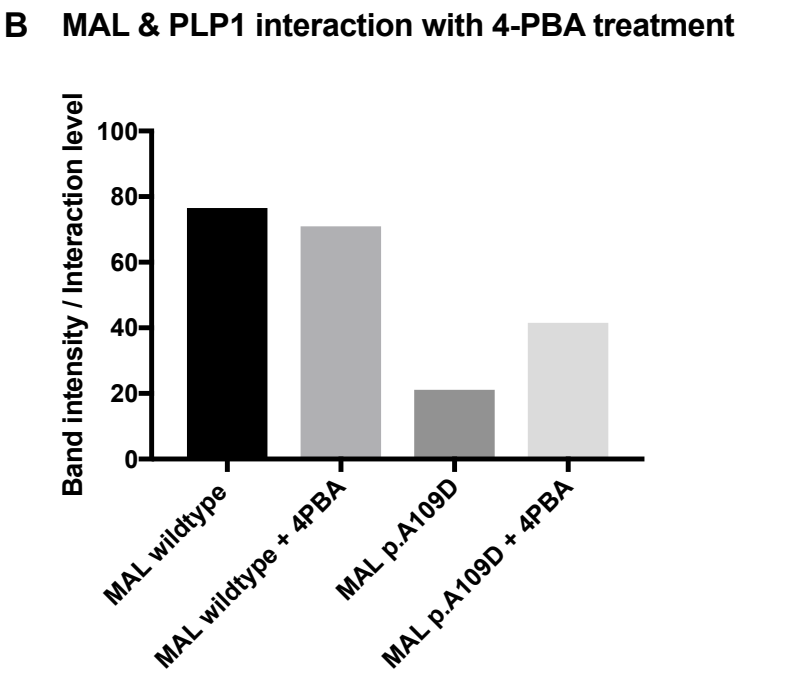


Figure 6-15 MAL – PLP1 interaction can be rescued by 4-PBA treatment.

A – Western blot analysis of immunoprecipitation experiment for assessing interaction of PLP1-GFP protein with wildtype and mutant MAL-V5 protein, with and without 4-PBA treatment. Pull-down was performed with GFP-beads and antibody staining was performed using V5 antibody, which targets MAL protein (expected size 16kDa). Arrows indicates the extra band in wildtype MAL as a result of post-translational modification, that is absent from mutant MAL. B – Quantification of MAL and PLP1 interaction in all four different conditions was achieved by Image Lab. This experiment was performed on one biological replicate.

6.5.4 Missense variant p.A109D possibly affects PLP1 activation and redirection to the basolateral membrane

I hypothesised that MAL is a regulator of the intracellular trafficking of PLP1 and drives its redirection to the basolateral membrane. I therefore decided to assess this by determining apical membrane localisation using the EZ-Link Sulfo-NHS-Biotin assay that allows for cell surface protein labelling. This assay uses sulfo-N-hydroxysuccinimide (NHS) esters of biotin that react with primary amino groups, specifically the side chain of lysine residues, to form stable amide bonds at pH 7 to 9. The reagent cannot permeate the cell membrane, thereby only exposed primary amines on the apical surface of the plasma membrane will become biotinylated.

Using the EZ-Link Sulfo-NHS-Biotin assay as a working strategy, the amino acid chains of our proteins of interest were assessed for lysine residues. MAL protein only has one lysine residue whereas PLP1 has twelve, hence PLP1 is likely to be a better biotinyation target in this assay. The assay was performed on confluent and polarized MDCK cells, over-expressing wildtype MAL, mutant MAL and PLP1 proteins. From my previous results, it was expected that PLP1 would be redirected to the basolateral membrane in the wildtype MAL cellular model, and therefore not biotinylated and detected by this assay. In contrast, PLP1 would remain on the apical cell membrane in the mutant MAL model because PLP1 transcytosis is impaired due to incorrect targeting of mutant MAL, and PLP1 would therefore be biotinylated and detected by this assay. The diagram below (**figure 6-16**) summarises the hypothesis that forms the basis of this assay.

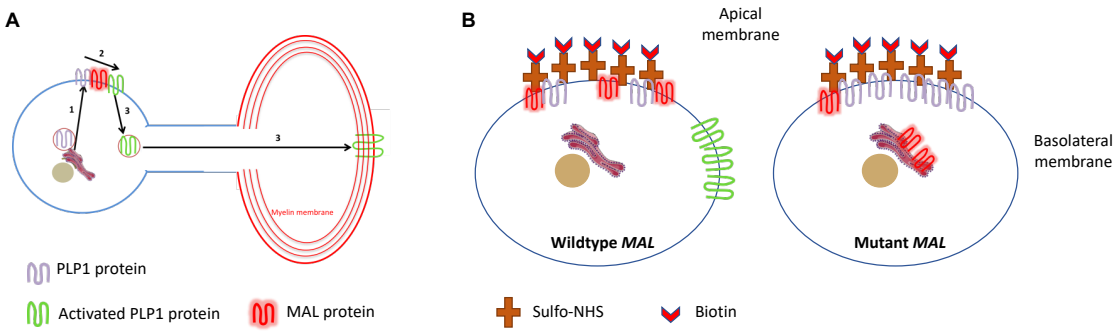


Figure 6-16 EZ-Link-Sulfo-NHS-Biotin assay hypothesis

The diagram outlines the hypothesis of assessing PLP1 localisation within a cell in the wildtype and mutant MAL cellular models. A – The overall working hypothesis of the role of MAL as a regulator of intracellular PLP1 trafficking. B – The two models whereby PLP1 correctly localizes in the basolateral membrane in the wildtype MAL model but remains in the apical membrane in the mutant MAL model.

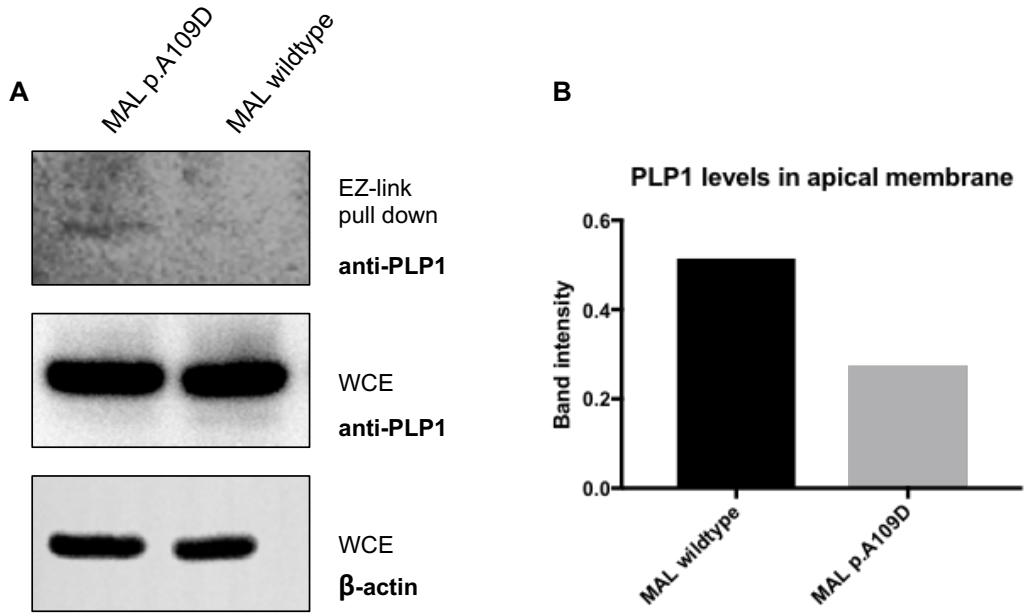


Figure 6-17 EZ-Link-Sulfo-NHS-Biotin assay for wildtype and mutant MAL model

A – Western blot results from the biotinylation assay showing a reduction in PLP1 levels from the apical membrane in the wildtype MAL model compared to the mutant MAL model. B – Quantification of western blot results using Image Lab. Anti-PLP1 band intensity from the pull-down was normalised to whole cell extract (WCE) anti-PLP1 bands, which was normalised to loading control.

For this experiment, wildtype and mutant MAL-V5 plasmids were overexpressed in MDCK cells, together with a PLP1-GFP plasmid. Cells were allowed to grow to full confluency and polarize. At that time, EZ-Link-Sulfo-NHS-Biotin reagent was added to the cells and cell lysate was then used for a pull-down assay using Streptavidin beads (collecting only biotinylated surface proteins). The pull-down proteins along with whole cell extracts were run on western blots and the membrane was stained for both the tagged PLP1 and MAL proteins. As expected, MAL was not targeted therefore the staining revealed no bands. For PLP1, bands were detected with more PLP1 protein being detected in the apical membrane in the mutant MAL cell model compared to the wildtype. This supports our previous findings and our intial hypothesis, proving that MAL has a key role in regulating intracellular transcytosis of PLP1, and potentially other proteins.

6.5.5 Investigation of novel protein-protein interactions using BioID2 and the impact of p.A109D on those interactions

6.5.5.1 Design and optimization of BioID2 experiments

As previously discussed, the BioID technique was used in this study to detect physiologically relevant protein-protein interactions, because this method allows the capture of weak and transient interactions (Roux et al., 2012). The original BioID protocol was improved to BioID2 by Kim et al. (2016), allowing for more efficient and faster biotinylation of proteins in proximity to the BirA*-tagged protein (Kim et al., 2016). This was achieved by using a smaller BirA biotin ligase from *A. aeolicus* instead of *E. coli* with a size of 25kDa instead of 35kDa, respectively (Kim et al., 2016). For the purposes of this study, the BioID protocol was updated to BioID2 in an attempt to identify key interactions that might be lost as a result of the p.A109D variant in MAL. Wildtype and mutant MAL was C-terminally tagged with BirA2* and then transfected into MDCK cells to induce biotinylation upon addition of exogenous biotin. Interacting proteins are then isolated by a standard biotin-affinity capture and identified using mass spectrometry, similar to the original protocol (Roux et al., 2012). A summary of the experimental plan is outlined in **Figure 6-18.**

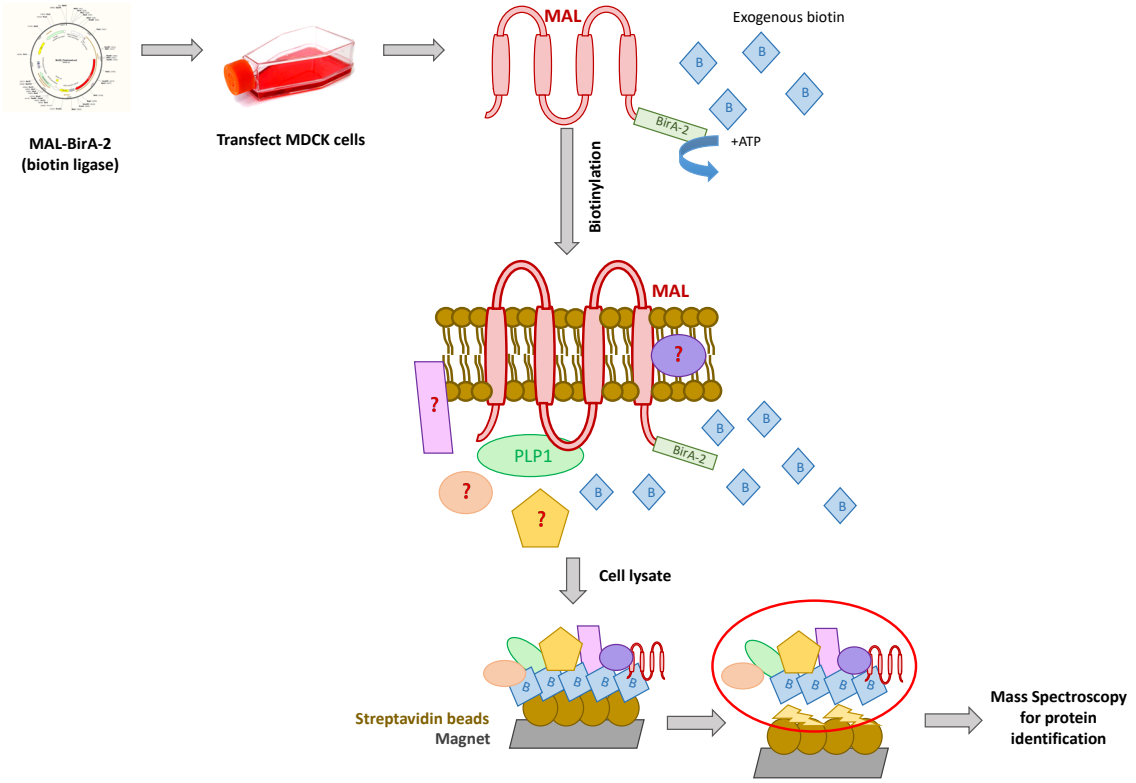


Figure 6-18 Outline of experimental plan for BioID2.

The illustration summarises the key steps in the BioID2 experiment, starting with exogenous expression of MAL-BirA2 fusion protein in MDCK cells. Addition of exogenous biotin causes biotinylation of proximal or closely interacting proteins followed by cell lysis and biotin-affinity capture of labelled proteins. Protein identification is achieved by using mass spectrometry.

The wildtype and mutant MAL-BirA2 fusion proteins were exogenously expressed using a standard transient transfection protocol (see section 2.2.15). The plasmids were constructed using Gateway technology (see section 2.2.11.1), where the *MAL* entry vector pENTR223 (DNASU, plasmid HsCD00505233) was cloned into destination vectors expressing the biotin ligase and a flag tag (a kind gift from Dr Taipale). Once the plasmid was constructed, the *MAL* c.326C>A mutation was introduced using site-directed mutagenesis (see section 2.2.11.3) and all plasmids were sequenced verified before used. The map of the plasmid is outlined below (**figure 6-19**).

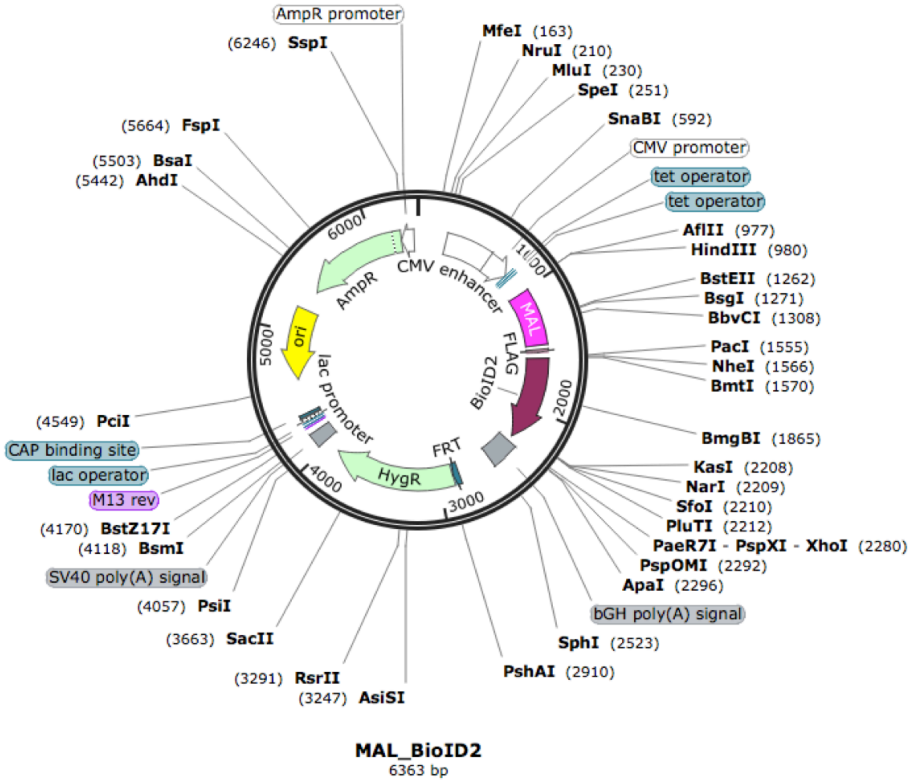


Figure 6-19 Map of MAL-BirA2 plasmid for BioID2 experiment

Once the plasmids were purified and sequence verified, the next step of the procedure was to optimise the conditions of the BioID2 experiment, as biotinylation was now occurring at a faster rate compared to the previous BioID experiment. Standard transfections were used, followed by incubation with media supplemented with biotin once the transfection complexes were removed. However, incubation times with biotin-supplemented media after transfection had to be optimized in order to achieve the best possible biotinylation levels. In the first instance, we used a range of incubation times ranging from 1 to 12 hours, allowing sufficient time for MAL-BirA2 to be expressed and for induction of biotinylation. **Figure 6-20** outlines the validation of biotinylation levels by western blotting. It was concluded that 4 hour incubation period would be the minimum time required to achieve adequate levels of biotinylation.

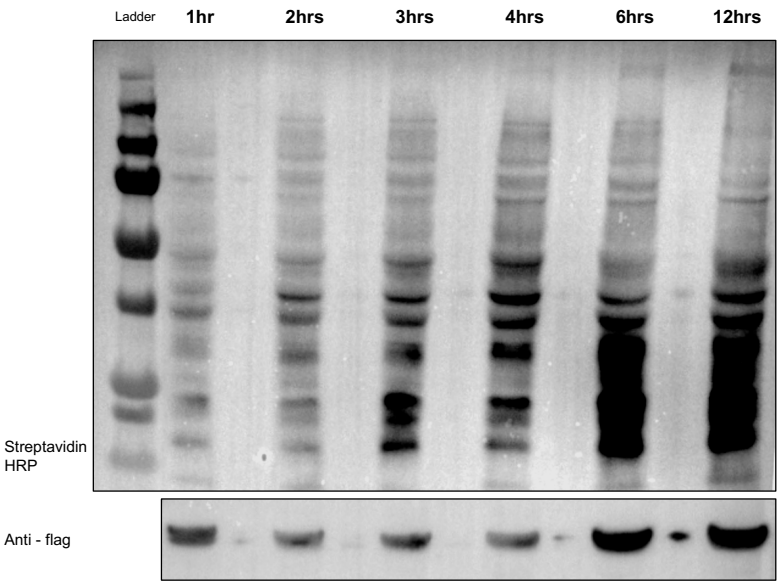


Figure 6-20 Optimisation of biotin incubation for BioID2

Western blot analysis of a range of different biotin incubation times (1-12 hours) to detect optimal levels of biotinylation. Streptavidin HRP stains for biotinylated proteins in the whole cell extracts and anti-flag antibody stains for levels of transfection (MAL-BirA2 protein is tagged with flag). Optimal incubation time before oversaturation of biotinylation was determined at 4 hours.

Once the biotin incubation time and transfection protocol was established, BioID2 experiments were performed (see section 2.2.21), using the MDCK cell-line as our model system since my previous experiments suggest that it polarizes and targets PLP1 correctly. BioID2 experiments were performed, lysates specifically pulled down with streptactin beads, trypsinised and then sent to my collaborator (Karsten Boldt, University of Tubingen) for mass spectroscopy analysis. For this project, three different biological replicates were used for wildtype and mutant MAL, with the appropriate controls in place (**figure 6-21**) to help with the downstream filtering of identified proteins.

Untransfected	Wildtype	Mutant
No Biotin	No Biotin	No Biotin
4 hours Biotin	4 hours Biotin	4 hours Biotin

Figure 6-21 Experimental samples and controls for BioID2 experiment

The table summarises each sample prepared for each biological replicate for the BioID2 experiments. Untransfected samples without biotin treatment were considered as the negative control of the experiment and any hits obtained by these samples were discarded. Untransfected samples with 4 hr incubation with biotin were considered as one of the positive controls of the experiments that helped in excluding any false positive results arising from excess biotin in cells. Transfected samples without biotin were also considered as positive controls that helped in excluding any false positive results arising from having an active biotin ligase enzyme within the cells that could potentially biotinylate proteins in near proximity using endogenous biotin present within the cells.

6.5.5.2 Analysis of mass spectroscopy data

Affinity purified eluates were precipitated with chloroform and methanol followed by trypsin digestion as described before (Gloeckner et al., 2009). C-MS/MS analysis was performed on Ultimate3000 nanoRSLC systems (Thermo Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) by a nano spray ion source. MS/MS data were analyzed using the MaxQuant software (version 1.6.1.0) (Cox and Mann, 2008, Cox et al., 2009). As a digesting enzyme, Trypsin/P was selected with maximal 2 missed cleavages. The data were analyzed by label-free quantification with the minimum ratio count of 3. The first search peptide tolerance was set to 20, the main search peptide tolerance to 4.5 ppm and the re-quantify option was selected. For peptide and protein identification the human subset of the SwissProt database (release 2014_11) was used and contaminants were detected using the MaxQuant contaminant search. A minimum peptide number of 2 and a minimum length of 7 amino acids was tolerated. Unique and razor peptides were used for quantification. The match between run option was enabled with a match time window of 0.7 min and an alignment time window of 20 min. The statistical analysis including ratio, t-test and significance A calculation was done using the Perseus software (version 1.5.5.3) (Tyanova et al., 2016). Significance A is a Q-function that detects outliers from a normal distribution. The background binders are the normal distribution and the interactors are the outliers. The output of the analysis is plotted in **figure 6-22**.

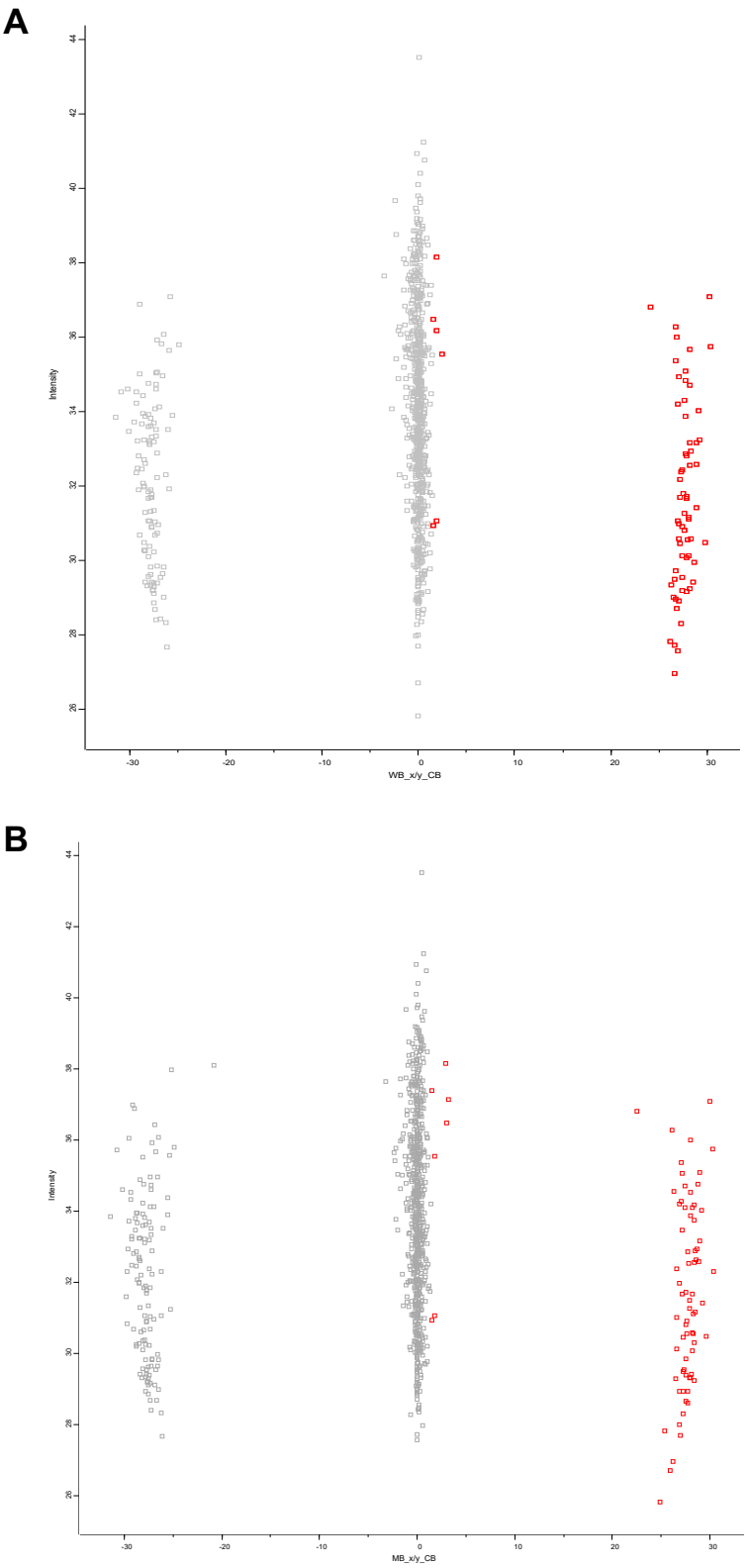


Figure 6-22 MS/MS data analysis using MaxQuant software

Plotted are the log10 ratios (x-axes) and the log10 intensities (y-axes) for all proteins quantified by the MaxQuant software. A – Wildtype MAL experiment. B – Mutant MAL experiment. Proteins significantly enriched are highlighted in red in both plots (Benjamini-Hochberg corrected, left sided Significance A < 0.05). This represents proteins that deviate from the normal distribution and are considered as outliers. These outliers are interacting proteins that are significantly enriched when wildtype and mutant MAL are biotinylated respectively.

6.5.5.3 Identification of protein-protein interactions

Following specialized analysis of data by Dr Boldt as mentioned above, further manual analysis was performed on the identified hits. Manual analysis was performed using different criteria at each step. Initially any hit that was not passing the t-test and significance A calculation was excluded from the list, and any hits enriched in negative or positive controls were also being excluded. Furthermore, any hits that were unique to either the wildtype MAL model or the mutant MAL model were used for downstream analysis, prioritising hits according to their biological relevance or cellular compartments and pathways. In addition, any hits that had previously been associated with similar diseases or have a neuronal function were highlighted.

The figure below (**figure 6.23**) illustrates the top hits obtained by the BioID2 experiment, that are unique to the wildtype MAL protein, ranked based on the average number of peptides identified per biological replicate. Some of the hits are highlighted based on their relevance to the function of MAL and disease association. None of these hits were identified with the mutant MAL protein.

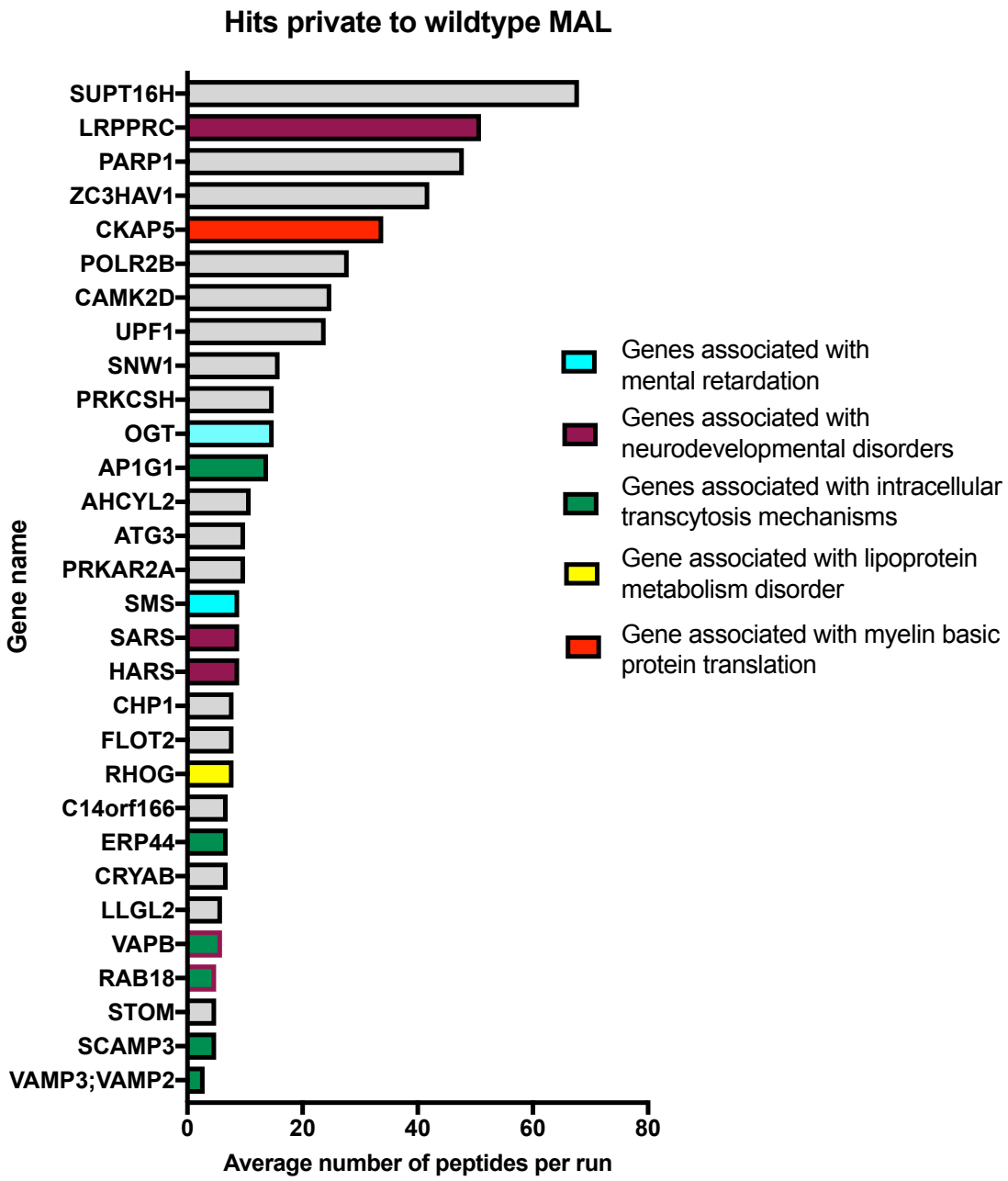


Figure 6-23 List of PPIs identified for wildtype MAL

The graph illustrates the hits obtained for wildtype MAL ranked based on the average number of peptides identified per run. None of this hits were identified in the mutant MAL pull-down. Highlighted are different genes associated with neurodevelopmental disorders or have a functional relevance to MAL and its cellular function.

MAL and PLP1 were not picked up by mass spectrometry, possibly because of the limited trypsin cleavage sites of those proteins. To validate this ExPASy PeptideCutter (Gasteiger et al., 2003) was used to perform an *in silico* trypsin cleavage for both MAL and PLP1. MAL has only got 3 cleavage sites at positions 120, 149 and 151, resulting in very few protein fragments to be read by the mass spectrometer. Protein size is not the issue, and detection is solely dependent on the tryptic sites for the cleavage products. A protein with fewer than 10 cleavage sites is assumed to be very difficult to be detected by this method. For example, VAMP3 is an 11kDa protein but has 13 tryptic sites making it possible to be identified. On the other hand, PLP1 has 19 tryptic sides so, in theory, should have been detected. However, the low endogenous expression of PLP1 in kidney epithelial cells could explain why it was not detected. Although mass spectroscopy did not detect these proteins, western blotting demonstrated that MAL and PLP1 were expressed and present in BioID2 pull-downs (**figure 6-24**).

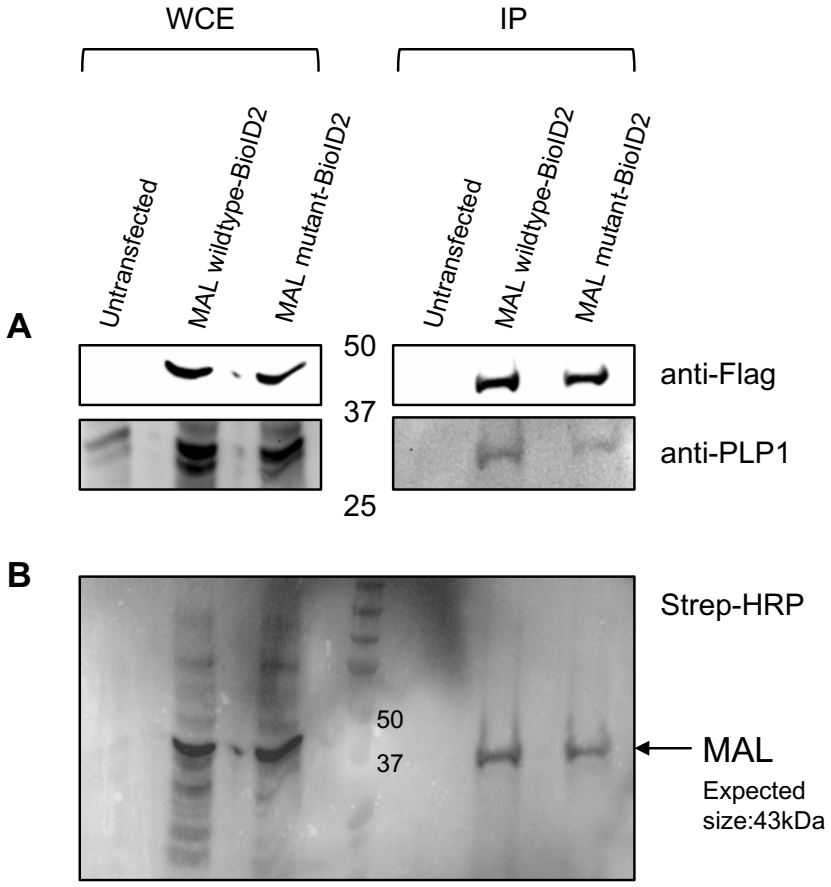


Figure 6-24 BioID2 validation for MAL and PLP1 pull downs

Pull-down validation for BioID2 by western blotting for MAL and PLP1 proteins. A – Staining with flag antibody (for BioID2-MAL plasmid with flag tag) and with PLP1 antibody, showing that both MAL and PLP1 proteins are present in the BioID2 experiment. Immunoprecipitation is done with streptavidin beads, following the BioID2 protocol. B – Strep-HRP staining showing the induction of biotinylation in the WCE lanes and the presence of MAL in the IP lanes.

6.5.5.4 Protein network analysis

In order to comprehend the most significant hits of the BioID2 experiment listed above, various computational tools were used to help us group proteins and build networks based on known protein-protein interactions (PPIs). By grouping hits based on their function, differences between wildtype and mutant MAL PPIs can be determined as well as the effect of the missense variant on these PPIs. To achieve this, I used a database of known and predicted protein-protein interactions known as STRING (Snel et al., 2000). This database comprises of interactions that include both physical (or direct) and functional (or indirect) associations derived from the following sources: genomic context predictions, high-throughput lab experiments, conservations and co-expression datasets, automated text-mining and previous knowledge from databases. By using STRING v11.0 (Szklarczyk et al., 2019), we were able to group the hits identified by both wildtype and mutant MAL in groups based on their biological processes using GO terms.

For wildtype MAL, PPI hits were grouped in cellular processes involving post-Golgi vesicle-mediated protein transport, macromolecule localization and protein localization to the membrane. Interestingly, protein-containing complex assembly were also identified potentially involved in MAL complex assembly, suggesting possible PPIs involved in the activating conformational change of PLP1 (Bijlard et al., 2016). Other interesting proteins include VAPB, a vesicle or plasma membrane protein involved in vesicle trafficking, that has been implicated in Amyotrophic Lateral Sclerosis (Kanekura et al., 2006) and Spinal Muscular Atrophy (Nishimura et al., 2004). VAMP2 and VAMP3 interact with VAPB and are involved in vesicle trafficking and SNARE complex formation. They also have a role in docking of synaptic vesicle in presynaptic neurons (Schwarz et al., 2017). Mutations in VAMP2 impact synaptic membrane fusion and neurodevelopment (Salpietro et al., 2019). Another functionally relevant hit is SCAMP3 that participates in protein trafficking to the cell surface and involved in post-Golgi recycling pathways (Thomas et al., 2016). ERP44 was also identified as an interactor of MAL. ERP44 acts as a chaperone, having a role in protein quality control that participates in the transport of proteins from the ER to the Golgi (Watanabe et al., 2017).

In addition, RAB18 was also identified which is a protein known to regulate vesicle transport and membrane trafficking, and that is required for eye and brain development (Wu et al., 2016). Mutations in this gene are a cause of Warburg Micro Syndrome Type 3 (Bem et al., 2011). Furthermore, LRPPRC is a protein that may play a role in vesicular transport and the organization of the cytoskeleton. Mutations in this gene are associated with Leigh Syndrome, French Canadian Type, with severe neurological and metabolic features (Debray et al., 2011). Lastly, a very interesting hit is CKAP5 that also encodes

a cytoskeleton-associated protein and microtubule binding. However, CKAP5 is also involved in the translation of myelin basic protein (MBP), supporting the hypothesis that MAL has a key role in myelin formation during neurodevelopment (Francone et al., 2007).

Taking into consideration disease-related hits, many of the proteins interacting with wildtype MAL are known to be mutated in a range of neurological or neurodevelopmental phenotypes, which supports a central role for MAL during neurodevelopment. Furthermore, wildtype MAL is associated with intracellular trafficking of proteins and targeting to the plasma membrane as an enriched functional process, consistent with my previous data and the existing literature on the role of MAL. The results are summarised in a STRING map (**figure 6-25**).

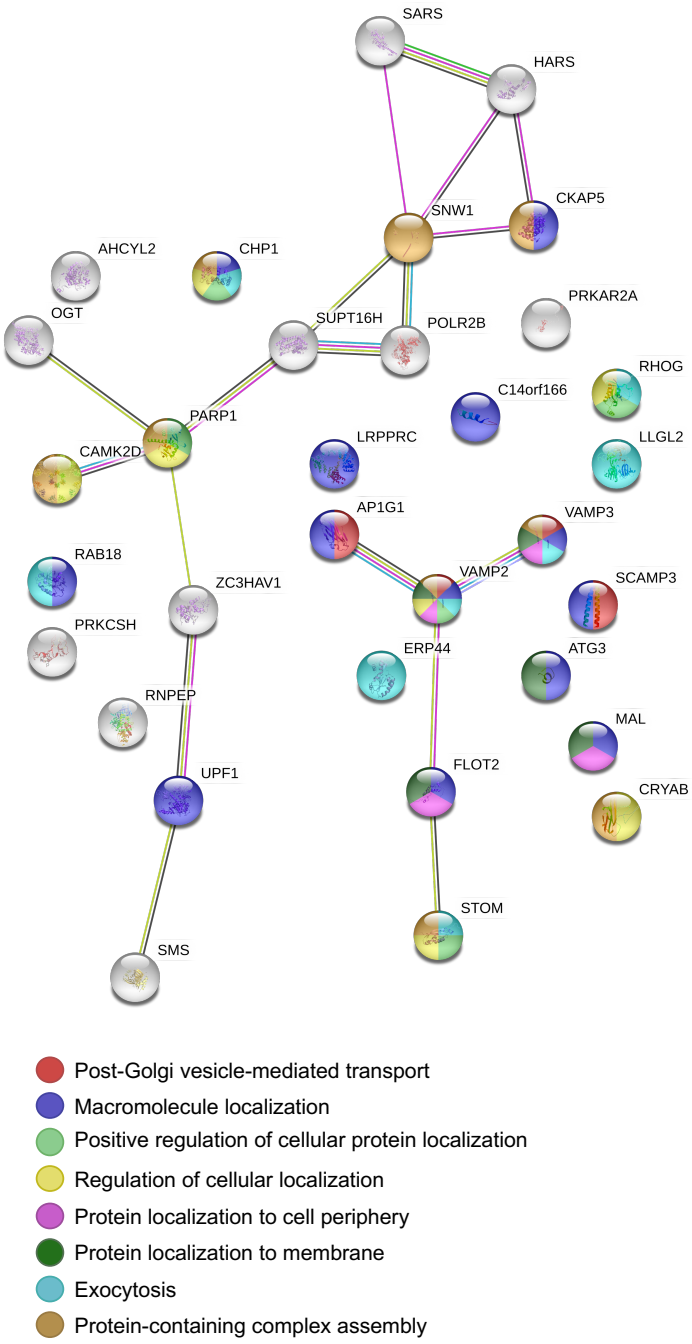


Figure 6-25 STRING map analysis for PPIs in wildtype MAL using BioID2

The map outlines the private hits obtained from the BioID2 experiment of wildtype MAL, with lines joining proteins indicating known PPIs. The different colours used correspond to specific biological processes identified between this proteins, using GO terms. It is noteworthy that MAL has not previously been described to interact with any of the above proteins, indicating the novelty of this dataset.

The same analysis was also performed for the hits obtained from the mutant MAL. In contrast with the biological processes identified for the wildtype protein, the mutant MAL dataset revealed a completely different enrichment of processes, consistent with the expected impact of the p.A109D variant on the mutant protein localization. Processes highlighted in this dataset include translation initiation, SRP-dependent protein targeting to the membrane, and RNA and heterocycle catabolic processes. These may represent proteins associated with the ER, or that are involved in ER stress and implicated in the subsequent UPR initiation, as a result of the MAL aggregates.

Firstly, a majority of proteins being identified are proteins that encode a component of the ribosomes, such as RPS18, RPS12, RLP9, RLP8 and RLP18A (Bateman et al., 2019). These proteins comprise of hits that are private only to the mutant MAL dataset, suggesting that they have been biotinylated by being in near proximity with the MAL aggregates in the rough ER.

Another ER-associated protein being identified is ERP29, which acts as a molecular chaperone for the release of proteins from the ER (McLaughlin et al., 2018). HSPBP1 protein was also a hit for mutant MAL, and it plays a role in regulating chaperone expression (Rogon et al., 2014). Increase in expression of chaperone-associated proteins may indicate that the UPR is initiated by mutant MAL protein, as a response to alleviate ER stress and to remove aggregates from the ER (Bravo et al., 2013).

To support the involvement of the UPR with processing of mutant MAL, eIF2B2, eIF3B and eIF3H proteins were all identified and are closely linked to the UPR. eIF2α is phosphorylated by the PERK pathway as part of the UPR to inhibit translation (Bravo et al., 2013). Specifically, eIF2α phosphorylation converts eIF2-GDP into a competitive inhibitor of eIF2B, which then delays protein synthesis (Bogorad et al., 2017). eIF3B, along with eIF3H, form parts of the eIF3 complex that is also required for translation initiation by interacting with mRNA and the 40S ribosomal subunit (Sharma et al., 2016). The eIF3 complex can play a role in attenuation of protein translation by regulating the eIF2 complex. This regulation is driven by pro-apoptotic stress signals and inactivation of the eIF3 complex eventually leads to apoptosis by caspases (Lasfargues et al., 2013).

Another hit, PDCD6, is a calcium-binding protein involved in ER-Golgi vesicular transport. In addition to its role in protein transport, PDCD6 is considered to be an apoptotic mediator via the caspase-3 dependent pathway (Lee et al., 2005), potentially supporting the hypothesis that UPR has been initiated in mutant MAL cells.

Another interesting hit is SRP72, that is involved in forming the signal recognition particle which is an essential for proper membrane localization of proteins. Perhaps this hit is private to the mutant MAL, as aggregates are being targeted but struggle to reach the plasma membrane, therefore keeping SRP bound to them (Akopian et al., 2013). Moreover, mutant MAL interactants include a calcium-related protein, potentially handling calcium release from the ER lumen stores to the cytosol (Bravo et al., 2013). STIM1 maintains Ca^{2+}levels by monitoring ER-luminal Ca^{2+} release (Gudlur et al., 2018). STIM1 can also be linked to UPR and apoptosis initiation, as failed UPR can lead to severe disturbance in homeostasis leading to pro-apoptotic signaling via Ca^{2+} signaling (Bahar et al., 2016).

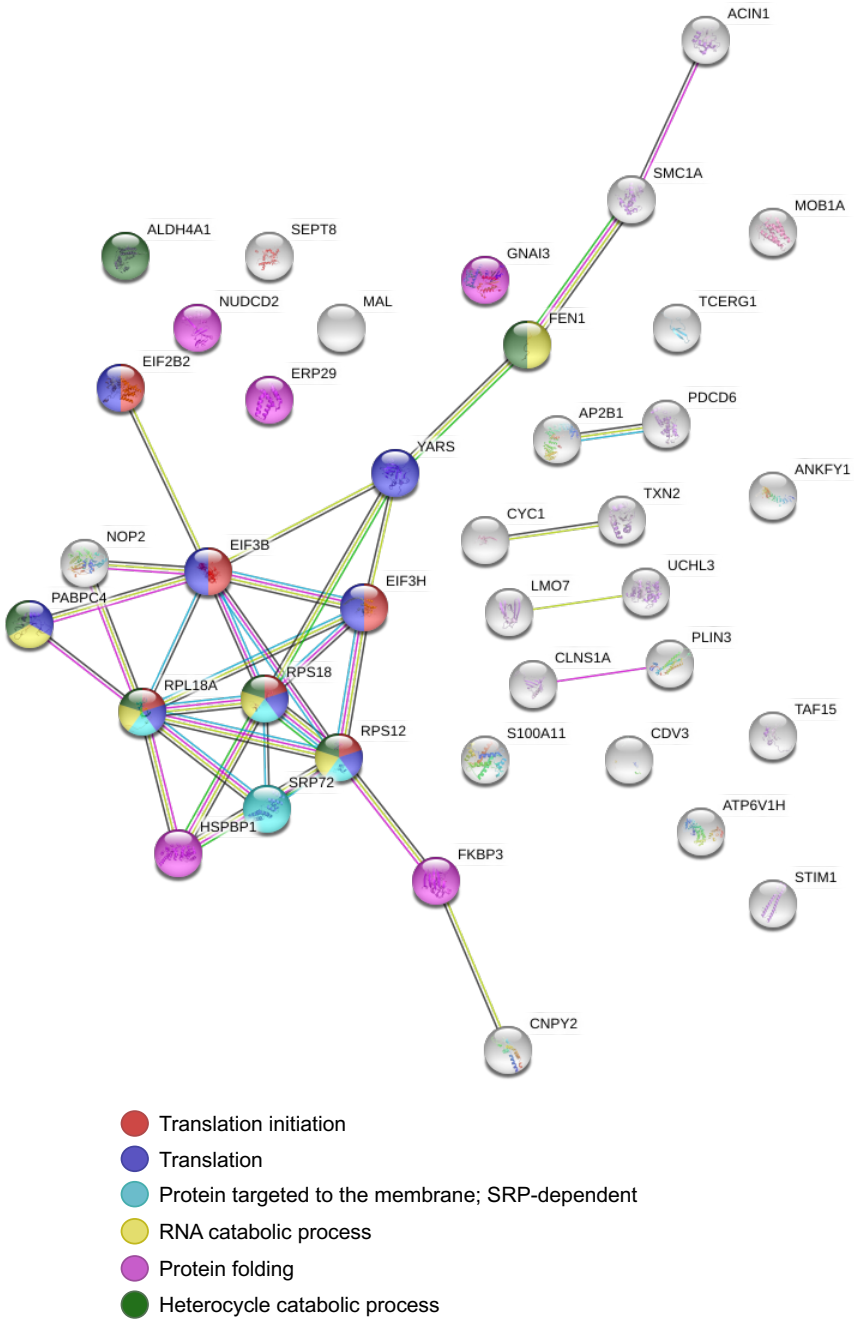


Figure 6-26 STRING map analysis for PPIs in mutant MAL using BioID2

The map outlines the private hits obtained from the BioID2 experiment of wildtype MAL, with lines joining proteins indicating known PPIs. The different colours used correspond to specific biological processes identified between this proteins, using GO terms. It is noteworthy that MAL is not associated with any of the above proteins, indicating the novelty of this dataset.

6.5.5.5 Pathway enrichment analysis

To further interpret the BioID2 data, a pathway enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6.8). DAVID provides a comprehensive set of functional annotation tools useful for interpreting large lists of genes (Huang et al., 2009b). DAVID can functionally annotate and classify genes, mainly using GO terms for enriched biological themes (Huang et al., 2009a). Such an enrichment analysis for pathway/process was performed with the significant hits from the BioID2 data, highlighting a few key processes that are significantly enriched. A summary of this analysis is illustrated in **figure 6-27**.

As expected, many of the enriched pathways identified in the wildtype MAL BioID2 dataset include intracellular vesicle-mediated transport and plasma membrane components, supporting my previous findings. Additionally, lipoprotein and membrane raft pathways are significantly enriched, highlighting the key role of MAL as a lipoprotein at the apical membrane of cells. Protein complex assembly is also enriched, a process that is presumably required for the MAL-PLP1 interaction and redistribution of PLP1, upon a conformational change, to the basolateral membrane.

In contrast, the mutant MAL BioID2 did not have any enriched pathways related to intracellular protein transport, protein complex assembly or membrane parts. This observation is consistent with the STRING analyses, suggesting that mutant MAL aggregates in the ER and is unable to act as a regulator of intracellular trafficking. The enriched pathways are linked to cell death processes such as phagosome acidification, probably of apoptotic cells as a result of failed UPR. Also, enrichment of translation associated pathways are also a part of the UPR, with certain proteins such as chaperones being overexpressed whereas the translation of other proteins is attenuated in order to release ER stress.

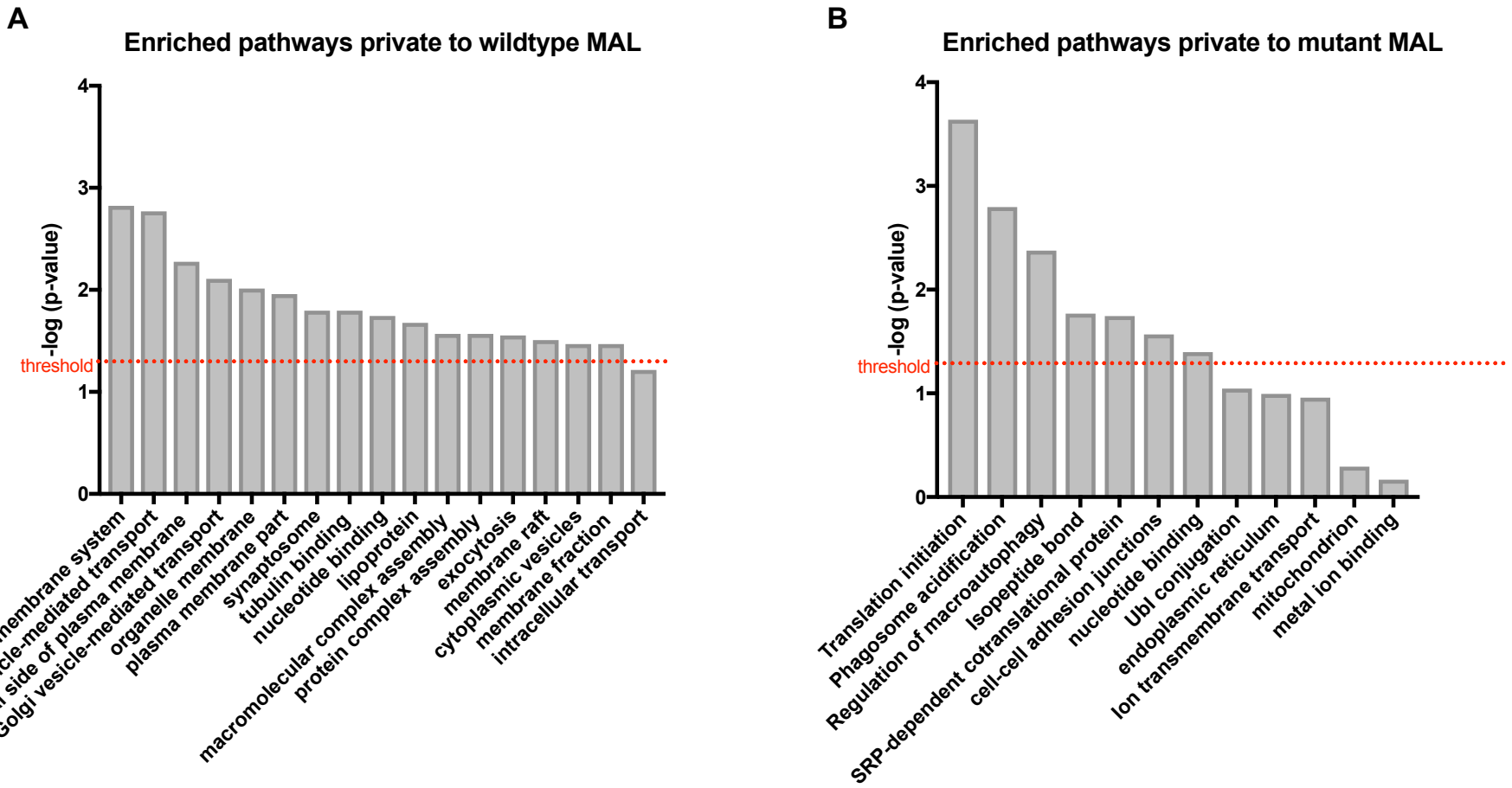


Figure 6-27 Pathway enrichment analysis for the BioID2 dataset for wildtype and mutant MAL.

The figure summarises the main pathways or processes enriched in the BioID2 dataset, between wildtype and mutant MAL. Processes above threshold are considered to be significant (p≤0.05 - above the red threshold line).

6.6 Future work

The work outlined in this chapter characterizes the identification of a variant c.326C>A in *MAL*, that establishes a novel cause for a rare form of leukodystrophy similar to Pelizaeus-Merzbacher disease. The variant identified in family ND13 was first assessed using *in silico* protein modelling to check if the variant is predicted to have any pathogenic effect on the protein. *In silico* modelling revealed that the missense variant impacts the third hydrophobic domain, probably preventing its incorporation to the membrane. As a result, the study was then extended to functional characterisation of the impact of the variant on cellular processes, by modelling wildtype and mutant MAL proteins using transient transfections.

The functional work described herein reveals that the missense variant in *MAL* causes protein aggregates to be formed in the ER, without reaching the plasma membrane where MAL is supposed to be localised to accommodate its function in intracellular trafficking. In addition, the ER aggregates, leading to induction of ER stress and potentially triggering the UPR, is a very common pathomechanism seen in a variety of disorders including Pelizaeus-Merzbacher disease (Dhaunchak and Nave, 2007), Vanishing White Matter Disease (VWMD) and Charcot-Marie-Tooth disease (Volpi et al., 2017).

Additional work on the current setup of experiments, would be to prove that UPR is induced as a result of ER stress by checking for specific markers. The UPR is initiated via IRE1, PERK and ATF6. IRE1 undergoes autophosphorylation upon ER-stress and UPR, something that can be measured by western blotting. PERK also undergoes autophosphorylation during the UPR, and also phosphorylates eIF2α to reduce global translation of mRNA. Both of these targets can be assessed with phospho-specific antibodies by western blotting. In addition, in response to ER-stress ATF6α translocates from the ER to the Golgi apparatus, something that can be tracked with immunofluorescence microscopy. Other downstream markers of the UPR activation include BiP, that acts as a sensor of ER stress and an ER chaperone, and could be assessed with western blotting or qPCR. In addition, CHOP is a marker of UPR, that is regulated by the PERK-eIF2a-ATF4 pathway and its expression level is usually low. CHOP is over-expressed upon UPR and it induces ER stress-mediated apoptosis. Checking the levels of CHOP and its targeted pro-apoptotic genes by qPCR could be another way of assessing UPR initiation and whether it leads to apoptosis or ER-associated protein degradation (Oslowski and Urano, 2011).

For future work on this study, transient expression of proteins would potentially be replaced by gene-editing technologies using CRISPR-Cas9 to obtain a more physiological expression of the wildtype and mutant MAL proteins. In addition, although MDCK cells are a good model for polarization and initiation of transcytosis, the same experiments should also be performed in a neuronal cell line, if achievable, as they are more relevant to the disease phenotype. If editing neuronal cell lines turns out to be challenging, embryonic stem cells could be potentially used for gene editing and then differentiating them into oligodendrocytes precursors or even 3D culture models (Rodrigues et al., 2017, Madhavan et al., 2018) to asses myelin formation.

As stated in this chapter, this finding is already shared with the scientific community working on rare diseases. Hopefully, identifying an independent replication of this variant in a different family with similar phenotype will help make this study stronger and benefit future clinical assessment of patients with leukodystrophies.

Chapter 7
Final Discussion

7.1 Summary of key findings

This study aimed to investigate the genetic causes of rare autosomal recessive neurodevelopmental and neuromuscular conditions using whole exome sequencing and functional characterization of identified variants in the best candidate genes. For this study, thirteen families were recruited and DNA was collected from multiple members of each family via blood or saliva samples. The affected individuals were fully assessed by clinical geneticists at regional hospitals. The sequencing strategy for each family varied according to sample availability or pedigree structure, but usually a trio was sequenced and often an unaffected sibling. Variant segregation, using multiple individuals from a family, provided a powerful tool in filtering out non-pathogenic variants, and in some cases fundamental for identifying the disease-causing variant.

In chapter 3, three families were described with mutations in known disease genes. The affected children in the first family (ND1) have significant arthrogryposis with multiple contractures of both upper and lower limbs. WES revealed a nonsense variant (p.R46X) in *CHRNG* that is a known cause of lethal and non-lethal Escobar syndrome (MIM number: 265000) (Morgan et al., 2006). The clinical synopsis of this syndrome matches the phenotype of family ND1, providing a molecular diagnosis for the family. Family ND2 presented with myopathy, abnormal movement and learning difficulties. For this family the WES investigations were performed by collaborators in Newcastle University, but the genetic finding revealed another nonsense variant (p.Q183*) in *MICU1*. At the time of this discovery *MICU1* was already established as a disease gene (Logan et al., 2014), and the variant p.Q183* was functionally characterized using patient fibroblasts. The third family of this chapter (ND3) presented with a severe form of developmental delay. WES revealed a frameshift variant in *HERC2* that caused a premature stop codon. This variant was functionally assessed using patient fibroblast, revealing complete loss of protein that is compatible with this variant being causative and pathogenic. However, *HERC2* is currently considered as a disease gene for Angelman-like syndrome (Puffenberger et al., 2012); a much milder phenotype compared to that for family ND3. The identification of this frameshift variant expands the genotype-phenotype correlation of *HERC2* mutations, whilst also providing a diagnosis for the patients. This finding was considered extremely important for aiding correct clinical diagnosis and care of future cases with severe developmental phenotypes carrying *HERC2* mutations.

In chapter 4, WES investigations of eight more families were discussed, with some of these cases still remaining unsolved. Family ND4 presented with microcephaly, central hypotonia and four limb spasticity. WES revealed a single homozygous variant (p.R365W) in *SUPV3L1*. At the end of this study, a second family with a variant in *SUPV3L1* was identified via the Decipher consortium, increasing the chances of *SUPV3L1* being a novel disease-causing gene. Furthermore, family ND5 presented with lymphoedema and antenatally agenesis of the corpus callosum. WES revealed compound heterozygous variants in EPHB2 (p.A731T and p.K872N), that could potentially be a novel cause of lymphoedema. Both of the above variants seem likely to be the causative mutations of each family, however further functional validation is necessary to establish the pathogenic nature of each variant.

In addition, family ND6 had a myopathy phenotype similar to limb-girdle muscular dystrophy. Genetic investigations identified variants in two candidate genes; a missense variant (p.A745V) in *ADAMTS15* and a splice-site variant (c.101608+1C>T) in *TTN*. Both genes are implicated in muscle formation and function, but the splice-site variant in *TTN* was considered the more likely causative variant due to the nature of the genetic variation identified. Also, biallelic variants in TTN have been previously associated with autosomal recessive myopathy similar to limb-girdle muscular dystrophy (Zheng et al., 2016). Family ND8 presented with motor and sensory neuropathy and WES revealed a homozygous duplication in *SH3TC2* that phenotypically was the best candidate gene identified. However, upon segregation analysis of the extended pedigree, including 9 individuals, the variant in *SH3TC2* did not segregate and therefore excluded. The next best candidate variant for this family was compound heterozygous variants in RGS12 (p.R392Q and p.R931H), a protein that has been suggested to have a biological role in neurogenesis and myogenesis. As mentioned above, the candidate variants identified in this chapter need further validation to prove whether they are causative and pathogenic.

Interestingly, a *de novo* case was also analyzed as part of this study. Family ND9 was clinically diagnosed with complex Moebius syndrome (MIM number:157900). MRI scans showed right perisylvian polymicrogyria and slight callosal dysmorphism. Sequencing revealed a *de novo* variant (p.G148E) in TUBA1A. This finding was important as it expands the genotype-phenotype correlation of *TUBA1A* to a wider range of brain malformations, apart from just lissencephaly. Some studies have reported similar cases to family ND8 (Gardner et al., 2018, Sato et al., 2018b), but the more variants that are reported with different brain defects to lissencephaly, the better the delineation of this novel *TUBA1A* phenotype-genotype correlation.

Lastly, families ND6 and ND10-ND11 described in chapter 4 remained unsolved. These families were multiplex and consisted of multiply consanguineous pedigree that identified clear homozygous regions, but no variants were identified in the coding exome that could be interpreted as pathogenic. Perhaps the disease causing variant lies within the intronic regions the genome and will require WGS in order to unravel the pathogenic causes of these families. On the other hand, the causative variant might be a synonymous variant within the exome that can cause cryptic splice sites. Such variants are being filtered out by our current bioinformatics pipelines and will require RNAseq along with WGS to resolve such complex genetic variations. Additional bioinformatics analysis tools with novel algorithms are being developed to detect potentially pathogenic intronic or synonymous variants that may interfere with splicing. An example of algorithm is splice AI (Jaganathan et al., 2019).

Chapter 5 describes a novel homozygous missense variant in *KLHL7* that is associated with a recessive neurodevelopmental condition. This chapter outlines the genetic investigation of family ND12 that presented with profound developmental delay and a phenotype similar to 4H syndrome (MIM number: 607694). The affected individuals also had abnormal brain scans with indications of microcephaly, and abnormal sweating. WES revealed a nonsense variant (p.W316*) in KLHL7, a gene that was at the time known to be mutated as a cause of autosomal dominant retinitis pigmentosa type 42 (RP42) (Friedman et al., 2009). A few months after this discovery, a study was published that characterized an autosomal recessive disorder associated with mutations in *KLHL7* (Angius et al., 2016). The patients in this study had the same phenotype as family ND12, confirming the finding of a disease-causing variant in *KLHL7*. KLHL7 is an E3-ligase adaptor, responsible for tagging substrates for proteasomal degradation. The intriguing part of this finding was that all the mutations associated with RP42 were localised in the BTB and BACK domain, which forms the region that binds to active side of the E3-ligasae complex, whereas mutations associated with the recessive neurodevelopmental disorder localize in the Kelch domains that form the substrate binding domain.

Patient fibroblasts were obtained from family ND12, and further functional studies were performed in order to investigate the pathogenic effects of p.W316*. Initially the impact of the null mutation on the cellular metabolism was assessed using a Seahorse XFe96. By using various toxins, the OXPHOS chain and glycolytic activity of the cells were evaluated. The results indicated a clear disruption of the OXPHOS chain and increased non-glycolytic acidification rate of the mutant cells compared to the wildtype. This may suggest that the mutation in *KLHL7* indirectly impacted on mitochondrial function and energy production, and that the mutant cells use lactate for their source of energy. This

observation supports the clinical observations of increased lactate levels in the blood of affected individuals.

Furthermore, novel protein-protein interactions of KLHL7 were identified using BioID experiments to capture stable and transient interactions. This was only performed in wildtype KLHL7 and identified potential E2s and E3s that could be binding partners with KLHL7 in the UPS pathway. The BioID experiment also revealed a clear association of KLHL7 with proteins involved in cell cycle and cytokinesis. A very interesting hit identified was ARL3, which is mutated as a known cause of autosomal dominant retinitis pigmentosa type 83 (Holtan et al., 2019) and autosomal recessive Joubert syndrome type 35 (Alkanderi et al., 2018). This could be a potential substrate of the KLHL7-associated E3-ligase complex, but further investigation will be required to establish this link.

Lastly, chapter 6 describes the identification of a homozygous missense variant (p.A109D) in MAL and the establishment of a novel disease gene associated with hypomyelinating leukodystrophy. The variant was identified in family ND13 that presented significant developmental delay and learning disabilities, feeding difficulties, cerebellar developmental defects, and cortical white matter abnormalities similar to Pelizaeus-Merzbacher disease (PMD). WES revealed variants in three candidate genes; *PNPT1*, *ZNHIT1* and *MAL*, but the variants in *PNPT1* did not segregate with family pedigree and the biological relevance of *MAL* made it the best functional candidate gene for this family. The p.A109D variant in MAL is in the third transmembrane domain and it was predicted *in silico* to impact the incorporation of that transmembrane domain in the membrane, due to the substitution of a hydrophobic to a hydrophilic amino acid.

Cellular modeling of the variant in MDCK epithelial cells revealed that the mutant MAL protein formed aggregates in the endoplasmic reticulum (ER). This was firstly observed by live cell imaging and further validated by immunofluorescence microscopy. This is a noteworthy as missense mutations in PLP1, the known cause of PMD (Hoffman-Zacharska et al., 2013), behave in the same way by forming ER aggregates leading to ER stress and triggering the unfolded protein response (UPR) (Bravo et al., 2013). This pathogenic phenotype was partly rescued by treating the cells with the chemical chaperone 4-phenylbuterate (4-PBA). Furthermore, MAL was thought to regulate the distribution of PLP1 (Bijlard et al., 2016), and as a result their potential interaction was assessed by immunoprecipitation experiments. The interaction between MAL and PLP1 was 60% reduced in the mutant MAL model compared to the wildtype. This was again partly rescued with the 4-PBA treatment.

Additionally, the overall protein-protein interaction network of MAL was assessed using BioID experiments, in an attempt to identify novel proteins potentially being involved in the process of intracellular trafficking and regulation of PLP1 during myelination. For this pupose, wildtype and mutant MAL protein were used which revealed completely distinct protein-protein interaction networks. The wildtype MAL seems to interact with proteins encoded by a few disease-causing genes, such as *OGT*, *LRPPRC* and *SARS*. Wildtype MAL also interacts with CKAP5, a protein already known to be associated with the translation of the myelin basic protein (MBP) (Francone et al., 2007). Interestingly, wildtype MAL has also revealed proteins essential for intracellular transcytosis mechanisms, such as VAPB, VAMP2/3 and RAB18. This highlighted the mechanistic role of MAL in intracellular trafficking mechanisms. On the other hand, the majority of interactions in the mutant MAL experiment were associated with the ER. Many ribosomal proteins were identified as well as ER molecular chaperones. A key observation was the identification of proteins such as EIF2B2 and EIF3B and EIF3H, all closely linked to the UPR (Bogorad et al., 2017). Also, PDCD6 was identified as an interactant of mutant MAL protein, which is a protein associated with apoptosis through the caspase-3 dependent pathway (Lee et al., 2005). This further supported the hypothesis of UPR initiation in mutant MAL cells and potential apoptotic signals as a result of unresolved ER stress.

Further research investigations will be required to further understand how MAL regulates PLP1 during the process of myelination, but the functional experiments outlined in chapter 6 provide novel insights into the pathomechanism related to the MAL mutation. There is clear evidence supporting the hypothesis that ER aggregates and unresolved ER stress lead to the pathogenic phenotype observed in family ND13. PLP1 is also not being transported to the correct membrane location, a process that is key during myelination and that could explain the leukodystrophy feature. Understanding more about this intracellular regulation mechanism during myelination will expand our current knowledge of the biology of neurogenesis and myelination.

7.2 Plans for future research

This research project can be further expanded in various other ways. Firstly, mining the 100K Genome Project dataset and screening additional patients with similar phenotype will be valuable in obtaining an independent replication of any of the private variants identified as part of this study. Achieving this will provide more supporting evidence that the variants identified herein are indeed pathogenic. Secondly, functionally characterizing the best candidate variants outlined in this project will establish whether or not they have any pathogenic effect. This could provide a confirmed diagnosis to the patients of the affected families recruited as part of this study, and will also improve future clinical diagnosis and care. Furthermore, functionally characterizing a variant can be advantageous in understanding more about the disease mechanism and gain insights into potentially novel uncharacterized biological mechanisms. For instance, cellular modeling of the variants in *EPHB2* (a gene that codes for a tyrosine kinase receptor) would be possible through subsequent identification of phosphorylation in ephrins by western blot staining with an anti-phosho ephrin antibody. Perhaps more elaborate kinase activity assays can be performed to assess any effect on phosphorylation. Purified wildtype and mutant EPHB2 protein, ephrin ligands and ATP are incubated *in vitro* (in-tube assay) to model phosphorylation. By measuring the ATP and ADP levels of the assay (ATP to ADP conversion) in wildtype and mutant proteins, the kinase activity can be determined (Peck, 2006).

In a similar approach, an *in vitro* ubiquitination assay could be used for any of the variants identified in proteins associated with the UPS (Zhao et al., 2012). HERC2 is an extremely large protein that might be problematic in purifying and performing this assay. However, KLHL7 (E3 adaptor protein) can be purified along with E1 ubiquitin-activating enzyme, potential E2-ubiqtuitin conjugating enzymes, E3-ligases and potential substrates identified from the BioID experiment. By adding all the components of the UPS along with ubiquitin and ATP in a tube and detecting ubiquitination of the substrate, specific E3-ligase complexes and their substrates can be identified expanding our current knowledge of KLHL7 function. In addition, the impact of the nonsense variant can be assessed by purifying a truncated protein and checking whether or not the ubiquitination of the substrate will be disrupted. This technique can be utilized to validate the BioID hits presented in this study.

For proteins such as HERC2, that are extremely large, transient cellular modeling via expression of a plasmid is not possible. In these instances, CRISPR-Cas9 technologies can be used instead to model and functionally assess a mutation. Using genome editing

methods to model variants is also considered a superior method than exogenous expression of plasmids, as the expression of the mutant protein will be matched to endogenous levels and will better represent physiological levels of the protein. Over the past few years, CRISPR-Cas9 technologies have significantly improved with many different options of genome editing now available. Loss of function variants, such as the one in *HERC2*, can be modelled by knock out CRISPR/Cas9 methods. These models consist of an RNA guided approach where a 20 base pair guide RNA (known as crRNA) directs the Cas9 endonuclease at the target sequence to create a double-strand cut. This will then be repaired by Non-Homologous End Joining (NHEJ), leading to disruption of the sequence's frame and complete knock out of the targeted gene (Au et al., 2019, Friedland et al., 2013).

A similar approach can be used to create knock-in models in order to model specific genetic variants, such as the missense mutation in MAL. As in the knock-out models, a guide RNA approach is used to direct Cas9 to generate a double stranded cut. However, for knock-in models a specific sequence is used as a template for genome editing through homology-directed repair (HDR) (Niccheri et al., 2017, Aslan et al., 2017). This methodology can model specific indels or SNPs but it is proven to be more challenging than knock-out models. In both cases, the efficiency of generating an edited cell line depends on the specificity and efficiency of the guide RNAs to target the correct region of the genome. This attempt may also generate off target effects, due to guide RNAs binding to other sides of the genome (Wilbie et al., 2019). This could be potentially overcome by using two guides instead of one to make the Cas9 cut more precisely. Off target effects in any generated cell-lines can be detected by sequenced with WGS.

Using genome editing of established cell lines can offer significant understanding to protein function and pathomechanism caused by genetic variants. However, editing stem cells (iPSCs) and growing them into a relevant 3D organoid model can offer great insights into biological functions of a certain protein in the overall function of an organ. This technology allows researchers to bridge the gap between two dimensional culture model and *in vivo* animal models, while offering sufficient power to recapitulate organ development and disease pathways in a 3D culture (Xu et al., 2018). With relevance to this study, development of human cortical spheroids carrying the *MAL* mutation will give us the opportunity to more efficiently model this variant and study how this variant can affect the process of myelination. Madhavan et al. proved that treating cortical spheroids with platelet-derived growth factor AA, insulin-like growth factor and thyroid hormone can promote the proliferation of oligodendrocyte progenitor cells and induce myelination in the 3D brain models. Furthermore, they also modelled three different variants in *PLP1*

(a deletion, a duplication and a missense variant), and demonstrated impaired myelination relevant to the PDM phenotype. Interestingly, the missense variant showed the most severe defects in the cortical organoid models. This study also used an inhibitor for the protein-kinase-R-like ER kinase, GSK2656157, which promoted mobilization of PLP1 away from the ER, resulting in improved myelination (Madhavan et al., 2018). This could also be used for the MAL model, as these two proteins seem to be working in a similar fashion.

7.3 Potential therapeutic approaches in genetic disorders

A wide range of genetic disorders still remain untreatable, with researchers trying to discover new methods to create treatment strategies for these patients. With regards to this study, a potential enzyme replacement therapy (Bengtsson et al., 2003) could be investigated, whereby the mutated E3 ligases enzymes would be replaced. Furthermore, chemical chaperones or the optimised PERK inhibitor mentioned above could be utilised in cases such as family ND13, where the mutant variant results in ER aggregates and induction of ER stress leading to the disease phenotype (Axten et al., 2013).

In addition, gene therapies could be examined where the mutant or abolished protein will be substituted by gene replacement approaches, using an adeno-associated virus (AAV) for delivery (Aguti et al., 2018). For cases where a nonsense variant is identified, gene therapies using antisense oligonucleotides (AON) might be a useful (Collin et al., 2012), resulting in a bypass of the premature stop and the generation of a shorter but potentially functional protein. Some of these therapies are already in clinical trials or used in clinic, like the gene therapy used in SMA and DMD (Lim et al., 2017), but developing new approaches specific to the variants identified herein requires additional research and validation of the therapeutic effect of each strategy.

Lastly, CRISPR-Cas9 genome editing approach could be used in order to correct the genetic variation that causes the disease phenotype (Papasavva et al., 2019). Currently, the first clinical trial has been launched, using gene-editing to correct a mutation in *CEP290* that causes Leber's congenital amaurosis. Preliminary *in vivo* work showed that the common IVS26 intronic mutation was corrected giving the green light from the US Food and Drug Administration to start the clinical trial (Sheridan, 2018, Maeder et al., 2019). As the editing is performed locally, only targeting somatic cells in the eye, it overcomes the ethical and safety considerations of editing the germline genome.

Others have tried to use CRISPR-Cas9 technologies in human embryos, to correct germline variants with mixed success. Tang et al. have demonstrated that they could use CRISPR-Cas9 methodology with HDR to correct point mutations in *HBB* and *G6PD* which cause thalassemia and hemolytic anemia respectively (Tang et al., 2017). These experiments were performed in dual pronuclear state human zygotes, but a poor mutation rate was observed due to low targeting efficiencies of the guide RNAs. Also, off-target effects were reported as well as high incidences of mosaicism. Schenkwein et al. tried a similar approach to correct a heterozygous mutation in *MYBPC3* that causes hypertrophic cardiomyopathy. They performed the editing at an earlier stage during fertilization (MII-phase egg cells were injected with CRISPR/Cas9 reagents and sperm) and observed a higher efficiency rate, with 72.4% of cases having the corrected version of the gene. They only had a single embryo with mosaicism and reduced off-target effects (Schenkwein and Yla-Herttuala, 2018).

In theory, similar approaches can be used as for *in vitro* fertilization (IVF) of single cell level embryos in order to correct any genetic variation that will lead to a disease. However, these approaches still remain under strong moral and ethical considerations that may or may not be resolved in the near future. A major safety consideration is the ability of this system to cause off target effects elsewhere in the genome, with unknown clinical outcomes. Potentially advancing the targeting efficiency and overcoming the off-target effects could significantly improve the safety considerations regarding genome editing approaches to therapy. In addition, there is still a major ethical unease about manipulating the human genome and rewriting the gene pool for the future generations (Lanphier et al., 2015). The recent scandal in China with He Jiankui shocked the scientific community and raises further concerns as to whether more regulations or bans should be placed regarding gene editing tools. He Jiankui unethically edited the genomes of embryos that were then implanted in women, giving birth to the first genetically edited humans (Cyranoski, 2019). I believe that this is unacceptable practice, as He didn't consider the ethical and safety considerations prior to doing this, jeopardizing life of the three babies. In the future, genome editing technologies should be strictly regulated and utilized to treat conditions as opposed to eugenic manipulation.

7.4 The impact of studying autosomal recessive diseases

Studying autosomal recessive diseases can have a great impact on both clinical and scientific settings. From the clinical perspective, this study offered molecular diagnosis for patients that would otherwise remain undiagnosed from the standard NHS genetic panel testing. This can be vital for the treating clinician, as knowing the cause of the disease can help with prescription of any available drugs that might alleviate symptoms. In addition, knowing the genetic background of the patient can help with future treatment strategies or clinical trial enrolment of potentially available gene therapies. Furthermore, obtaining a molecular diagnosis can be extremely helpful in genetic counseling of other members of the family that might be potential carriers of the disease variant, as well as prenatal screening for any future pregnancies of the affected family. In addition, if a family would consider IVF for future pregnancies, having a molecular diagnosis can help with pre-implantation genetic diagnosis to test the embryo prior to implantation in the womb.

The molecular diagnosis can be valuable for the parents and the patients, as it gives them clarity about their disease. Diagnosis often comes with a clinical syndrome's name that patients and their parents can read about further. In many instances, affected families join support groups and meet with other people affected by the same condition, offering them hope and a sense of belonging. In my opinion, offering the patients with the diagnostic clarity is precious, even if no treatment options are available for them.

Moreover, the study of autosomal recessive disorders can be beneficial for the scientific community too, offering novel knowledge about biological processes and disease mechanisms. With regards to this study, the identification of a genetic variant that causes a rare autosomal recessive condition can establish a new genotype-phenotype correlation and provide us with biological insights that would otherwise remain unknown. For instance, the discovery of the variant in *MAL* determines for the first time its role in myelination and its association with a leukodystrophy phenotype similar to PMD.

In contrast with big genomic projects, such as the 100K Genomes, this study offers higher resolution for variant identification. The hypothesis-free and family-based approached (trio or more) of this study, offers better chances of truly identifying the causative variant. As demonstrated in family ND13, more than one variant in a biologically-relevant gene can be identified via WES. However, extensive segregation analysis in additional family members can be critical in ruling out variants that do not segregate, increasing the chances of identifying the correct mutation. In contrast, with studies such as the 100K Genomes or the 5 million Genome Projects, this study also

utilized variant modelling and functional characterisation to elucidate the pathogenicity of a variant. The big genomic studies do not offer any functional interpretation of variants, something that I believe is very important in clarifying pathogenic variants from non-pathogenic ones, and understanding more about disease mechanisms. However, the advantage of the current and future big genomic projects is the establishment of large datasets that will enhance the identification of independent replications of certain variants, that currently remain private to a single family.

7.5 Conclusions

The family-based and hypothesis-free approach for the genetic investigations of this project lead to a number of key findings. Novel disease genes have been identified in two cases with global neurodevelopmental condition and a distinct hypomyelinating leukodystrophy. In addition, several candidate genes have been identified in other cases outlined in this study, that require further validation to prove pathogenicity and gain insights into disease mechanism. Most of the variants discussed in this study are private and future investigations will be needed to discover independent replication of variants in the genes mentioned herein.

Next generation sequencing techniques have been growing rapidly, offering higher quality and coverage of sequencing. The genomics field has also been developing new approaches to analyze and interpret big genomic data. These developments offer great opportunities for big studies such as the 100K Genomes and the 5 Million Genomes Projects to develop big dataset of genetic variants associated with disease. This might help researchers in the field establish novel disease genes by data-sharing and matching cases. Despite the fact that these big genome projects can provide genetic input into establishing new genotype-phenotype correlations, they lack the family-approach strategies outlined in this study that often provided the fundamental aspect in solving a case via segregation analysis. In addition, this research project was valuable as it proposed new pathomechanisms and provided insights into the biological function of certain proteins by performing numerous functional studies. Functional characterisation of variants is a crucial step in resolving the cause of genetic conditions, both because it contributes to the current knowledge of the scientific community but it also provides essential information for developing future therapies.

Bibliography

ABECASIS, G. R., CHERNY, S. S., COOKSON, W. O. & CARDON, L. R. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics,* 30**,** 97-101.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nature Methods,* 7**,** 248-249.

AGUTI, S., MALERBA, A. & ZHOU, H. Y. 2018. The progress of AAV-mediated gene therapy in neuromuscular disorders. *Expert Opinion on Biological Therapy,* 18**,** 681-693.

AKOPIAN, D., SHEN, K., ZHANG, X. & SHAN, S. O. 2013. Signal Recognition Particle: An Essential Protein-Targeting Machine. *In:* KORNBERG, R. D. (ed.) *Annual Review of Biochemistry, Vol 82.*

ALKANDERI, S., MOLINARI, E., SHAHEEN, R., ELMAGHLOOB, Y., STEPHEN, L. A., SAMMUT, V., RAMSBOTTOM, S. A., SRIVASTAVA, S., CAIRNS, G., EDWARDS, N., RICE, S. J., EWIDA, N., ALHASHEM, A., WHITE, K., MILES, C. G., STEEL, D. H., ALKURAYA, F. S., ISMAIL, S. & SAYER, J. A. 2018. ARL3 Mutations Cause Joubert Syndrome by Disrupting Ciliary Protein Composition. *American Journal of Human Genetics,* 103**,** 612-620.

ALKURAYA, F. S. 2012. Discovery of rare homozygous mutations from studies of consanguineous pedigrees. *Current protocols in human genetics,* Chapter 6**,** Unit6.12.

ALLAN, D. C. & PHILLIPS, J. C. 2017. Evolution of the ubiquitin-activating enzyme Uba1 (E1). *Physica a-Statistical Mechanics and Its Applications,* 483**,** 456-461.

ANGIUS, A., UVA, P., BUERS, I., OPPO, M., PUDDU, A., ONANO, S., PERSICO, I., LOI, A., MARCIA, L., HOHNE, W., CUCCURU, G., FOTIA, G., DEIANA, M., MARONGIU, M., ATALAY, H. T., INAN, S., EL ASSY, O., SMIT, L. M. E., OKUR, I., BODUROGLU, K., UTINE, G. E., KILIC, E., ZAMPINO, G., CRISPONI, G., CRISPONI, L. & RUTSCH, F. 2016. Bi-allelic Mutations in KLHL7 Cause a Crisponi/CISS1-like Phenotype Associated with Early-Onset Retinitis Pigmentosa. *American Journal of Human Genetics,* 99**,** 236-245.

ANTONY, A. N., PAILLARD, M., MOFFAT, C., JUSKEYICIUTE, E., CORRENTI, J., BOLON, B., RUBIN, E., CSORDAS, G., SEIFERT, E. L., HOEK, J. B. & HAJNOCZKY, G. 2016. MICU1 regulation of mitochondrial Ca2+ uptake dictates survival and tissue regeneration. *Nature Communications,* 7.

ASLAN, Y., TADJUIDJE, E., ZORN, A. M. & CHA, S. W. 2017. High-efficiency non-mosaic CRISPR-mediated knock-in and indel mutation in F0 Xenopus. *Development,* 144**,** 2852-2858.

AU, V., LI-LEGER, E., RAYMANT, G., FLIBOTTE, S., CHEN, G., MARTIN, K., FERNANDO, L., DOELL, C., ROSELL, F. I., WANG, S., EDGLEY, M. L., ROUGVIE, A. E., HUTTER, H. & MOERMAN, D. G. 2019. CRISPR/Cas9 Methodology for the Generation of Knockout Deletions in Caenorhabditis elegans. *G3-Genes Genomes Genetics,* 9**,** 135-144.

AXTEN, J. M., ROMERIL, S. P., SHU, A., RALPH, J., MEDINA, J. R., FENG, Y. H., LI, W. H. H., GRANT, S. W., HEERDING, D. A., MINTHORN, E., MENCKEN, T., GAUL, N., GOETZ, A., STANLEY, T., HASSELL, A. M., GAMPE, R. T., ATKINS, C. & KUMAR, R. 2013. Discovery of GSK2656157: An Optimized PERK Inhibitor Selected for Preclinical Development. *Acs Medicinal Chemistry Letters,* 4**,** 964-968.

BAHAR, E., KIM, H. & YOON, H. 2016. ER Stress-Mediated Signaling: Action Potential and Ca2+ as Key Players. *International Journal of Molecular Sciences,* 17.

BALCAROVASTANDER, J., PFEIFFER, S. E., FULLER, S. D. & SIMONS, K. 1984. DEVELOPMENT OF CELL-SURFACE POLARITY IN THE EPITHELIAL MADIN-DARBY CANINE KIDNEY (MDCK) CELL-LINE. *Embo Journal,* 3**,** 2687-2694.

BARON, W. & HOEKSTRA, D. 2010. On the biogenesis of myelin membranes: Sorting, trafficking and cell polarity. *Febs Letters,* 584**,** 1760-1770.

BARON, W., OZGEN, H., KLUNDER, B., DE JONGE, J. C., NOMDEN, A., PLAT, A., TRIFILIEFF, E., DE VRIES, H. & HOEKSTRA, D. 2015. The Major Myelin-Resident Protein PLP Is Transported to Myelin Membranes via a Transcytotic Mechanism: Involvement of Sulfatide. *Molecular and Cellular Biology,* 35**,** 288-302.

BATEMAN, A., MARTIN, M. J., ORCHARD, S., MAGRANE, M., ALPI, E., BELY, B., BINGLEY, M., BRITTO, R., BURSTEINAS, B., BUSIELLO, G., BYE-A-JEE, H., DA SILVA, A., DE GIORGI, M., DOGAN, T., CASTRO, L. G., GARMIRI, P., GEORGHIOU, G., GONZALES, D., GONZALES, L., HATTON-ELLIS, E., IGNATCHENKO, A., ISHTIAQ, R., JOKINEN, P., JOSHI, V., JYOTHI, D., LOPEZ, R., LUO, J., LUSSI, Y., MACDOUGALL, A., MADEIRA, F., MAHMOUDY, M., MENCHI, M., NIGHTINGALE, A., ONWUBIKO, J., PALKA, B., PICHLER, K., PUNDIR, S., QI, G. Y., RAJ, S., RENAUX, A., LOPEZ, M. R., SAIDI, R., SAWFORD, T., SHYPITSYNA, A., SPERETTA, E., TURNER, E., TYAGI, N., VASUDEV, P., VOLYNKIN, V., WARDELL, T., WARNER, K., WATKINS, X., ZARU, R., ZELLNER, H., BRIDGE, A., XENARIOS, I., POUX, S., REDASCHI, N., AIMO, L., ARGOUD-PUY, G., AUCHINCLOSS, A., AXELSEN,

K., BANSAL, P., BARATIN, D., BLATTER, M. C., BOLLEMAN, J., BOUTET, E., BREUZA, L., CASALS-CASAS, C., DE CASTRO, E., COUDERT, E., CUCHE, B., DOCHE, M., DORNEVIL, D., ESTREICHER, A., FAMIGLIETTI, L., FEUERMANN, M., GASTEIGER, E., GEHANT, S., GERRITSEN, V., GOS, A., GRUAZ, N., HINZ, U., HULO, C., HYKA-NOUSPIKEL, N., JUNGO, F., KELLER, G., KERHORNOU, A., LARA, V., LEMERCIER, P., LIEBERHERR, D., LOMBARDOT, T., MARTIN, X., MASSON, P., MORGAT, A., NETO, T. B., PAESANO, S., PEDRUZZI, I., PILBOUT, S., POZZATO, M., et al. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research,* 47**,** D506-D515.

BELKADI, A., BOLZE, A., ITAN, Y., COBAT, A., VINCENT, Q. B., ANTIPENKO, A., SHANG, L., BOISSON, B., CASANOVA, J. L. & ABEL, L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 5473-5478.

BEM, D., YOSHIMURA, S. I., NUNES-BASTOS, R., BOND, F. C., KURIAN, M. A., RAHMAN, F., HANDLEY, M. T. W., HADZHIEV, Y., MASOOD, I., STRAATMAN-IWANOWSKA, A. A., CULLINANE, A. R., MCNEILL, A., PASHA, S. S., KIRBY, G. A., FOSTER, K., AHMED, Z., MORTON, J. E., WILLIAMS, D., GRAHAM, J. M., DOBYNS, W. B., BURGLEN, L., AINSWORTH, J. R., GISSEN, P., MULLER, F., MAHER, E. R., BARR, F. A. & ALIGIANIS, I. A. 2011. Loss-of-Function Mutations in RAB18 Cause Warburg Micro Syndrome (vol 88, pg 499, 2011). *American Journal of Human Genetics,* 88**,** 678-678.

BENGTSSON, B. A., JOHANSSON, J. O., HOLLAK, C., LINTHORST, G. & FELDTRASMUSSEN, U. 2003. Enzyme replacement in Anderson-Fabry disease. *Lancet,* 361**,** 352-352.

BERGER, P., BONNEICK, S., WILLI, S., WYMANN, M. & SUTER, U. 2002. Loss of phosphatase activity in myotubularin-related protein 2 is associated with Charcot-Marie-Tooth disease type 4B1. *Human Molecular Genetics,* 11**,** 1569-1579.

BERNDSEN, C. E. & WOLBERGER, C. 2014. New insights into ubiquitin E3 ligase mechanism. *Nature Structural & Molecular Biology,* 21**,** 301-307.

BERROU, E., SOUKASEUM, C., FAVIER, R., ADAM, F., ELAIB, Z., KAUSKOT, A., BORDET, J. C., BALLERINI, P., LOYAU, S., FENG, M., DIAS, K., MUHEIDLI, A., GIRAULT, S., NURDEN, A. T., TURRO, E., OUWEHAND, W. H., DENIS, C. V., JANDROT-PERRUS, M., ROSA, J. P., NURDEN, P. & BRYCKAERT, M. 2018. A mutation of the human EPHB2 gene leads to a major platelet functional defect. *Blood,* 132**,** 2067-2077.

BIJLARD, M., DE JONGE, J. C., KLUNDER, B., NOMDEN, A., HOEKSTRA, D. & BARON, W. 2016. MAL Is a Regulator of the Recruitment of Myelin Protein PLP to Membrane Microdomains. *Plos One,* 11.

BOGORAD, A. M., LIN, K. Y. & MARINTCHEV, A. 2017. Novel mechanisms of eIF2B action and regulation by eIF2 alpha phosphorylation. *Nucleic Acids Research,* 45**,** 11962-11979.

BOLINO, A., MUGLIA, M., CONFORTI, F. L., LEGUERN, E., SALIH, M. A. M., GEORGIOU, D. M., CHRISTODOULOU, K., HAUSMANOWA-PETRUSEWICZ, I., MANDICH, P., SCHENONE, A., GAMBARDELLA, A., BONO, F., QUATTRONE, A., DEVOTO, M. & MONACO, A. P. 2000. Charcot-Marie-Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2. *Nature Genetics,* 25**,** 17-19.

BOND, J., ROBERTS, E., MOCHIDA, G. H., HAMPSHIRE, D. J., SCOTT, S., ASKHAM, J. M., SPRINGELL, K., MAHADEVAN, M., CROW, Y. J., MARKHAM, A. F., WALSH, C. A. & WOODS, C. G. 2002. ASPM is a major determinant of cerebral cortical size. *Nature Genetics,* 32**,** 316-320.

BOND, J., SCOTT, S., HAMPSHIRE, D. J., SPRINGELL, K., CORRY, P., ABRAMOWICZ, M. J., MOCHIDA, G. H., HENNEKAM, R. C. M., MAHER, E. R., FRYNS, J. P., ALSWAID, A., JAFRI, H., RASHID, Y., MUBAIDIN, A., WALSH, C. A., ROBERTS, E. & WOODS, C. G. 2003. Protein-truncating mutations in ASPM cause variable reduction in brain size. *American Journal of Human Genetics,* 73**,** 1170-1177.

BRAGOSZEWSKI, P., TUREK, M. & CHACINSKA, A. 2017. Control of mitochondrial biogenesis and function by the ubiquitin - proteasome system. *Open Biology,* 7.

BRAVO, R., PARRA, V., GATICA, D., RODRIGUEZ, A. E., TORREALBA, N., PAREDES, F., WANG, Z. V., ZORZANO, A., HILL, J. A., JAIMOVICH, E., QUEST, A. F. G. & LAVANDERO, S. 2013. Endoplasmic Reticulum and the Unfolded Protein Response: Dynamics and Metabolic Integration. *In:* JEON, K. W. (ed.) *International Review of Cell and Molecular Biology, Vol 301.*

BROOKES, S. J., BARRON, M. J., BOOT-HANDFORD, R., KIRKHAM, J. & DIXON, M. J. 2014. Endoplasmic reticulum stress in amelogenesis imperfecta and phenotypic rescue using 4-phenylbutyrate. *Human Molecular Genetics,* 23**,** 2468-2480.

BROOKES, S. J., BARRON, M. J., SMITH, C. E. L., POULTER, J. A., MIGHELL, A. J., INGLEHEARN, C. F., BROWN, C. J., RODD, H., KIRKHAM, J. & DIXON, M. J. 2017. Amelogenesis imperfecta caused by N-terminal enamelin point mutations in mice and men is driven by endoplasmic reticulum stress. *Human Molecular Genetics,* 26**,** 1863-1876.

BUERMANS, H. P. J. & DEN DUNNEN, J. T. 2014. Next generation sequencing technology: Advances and applications. *Biochimica Et Biophysica Acta-Molecular Basis of Disease,* 1842**,** 1932-1941.

CANNING, P., COOPER, C. D. O., KROJER, T., MURRAY, J. W., PIKE, A. C. W., CHAIKUAD, A., KEATES, T., THANGARATNARAJAH, C., HOJZAN, V., AYINAMPUDI, V., MARSDEN, B. D., GILEADI, O., KNAPP, S., VON DELFT, F. & BULLOCK, A. N. 2013. Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases (vol 288, pg 7803, 2013). *Journal of Biological Chemistry,* 288**,** 28304-28304.

CARDOSO, A. R., LOPES-MARQUES, M., SILVA, R. M., SERRANO, C., AMORIM, A., PRATA, M. J. & AZEVEDO, L. 2019. Essential genetic findings in neurodevelopmental disorders. *Human Genomics,* 13.

CARR, I. M., BHASKAR, S., SULLIVAN, J. O., ALDAHMESH, M. A., SHAMSELDIN, H. E., MARKHAM, A. F., BONTHRON, D. T., BLACK, G. & ALKURAYA, F. S. 2013. Autozygosity Mapping with Exome Sequence Data. *Human Mutation,* 34**,** 50-56.

CASTRO-GAGO, M., DACRUZ-ALVAREZ, D., PINTOS-MARTINEZ, E., BEIRAS-IGLESIAS, A., DELMIRO, A., ARENAS, J., MARTIN, M. A. & MARTINEZ-AZORIN, F. 2014. Exome sequencing identifies a CHKB mutation in Spanish patient with Megaconial Congenital Muscular Dystrophy and mtDNA depletion. *European Journal of Paediatric Neurology,* 18**,** 796-800.

CAYAMI, F. K., LA PIANA, R., VAN SPAENDONK, R. M. L., NICKEL, M., BLEY, A., GUERRERO, K., TRAN, L. T., VAN DER KNAAP, M. S., BERNARD, G. & WOLF, N. I. 2015. POLR3A and POLR3B Mutations in Unclassified Hypomyelination. *Neuropediatrics,* 46**,** 221-227.

CHARZEWSKA, A., WIERZBA, J., IZYCKA-SWIESZEWSKA, E., BEKIESINSKA-FIGATOWSKA, M., JUREK, M., GINTOWT, A., KLOSOWSKA, A., BAL, J. & HOFFMAN-ZACHARSKA, D. 2016. Hypomyelinating leukodystrophies a molecular insight into the white matter pathology. *Clinical Genetics,* 90**,** 293-304.

CHATTERJEE, T. K. & FISHER, R. A. 2000. Novel alternative splicing and nuclear localization of human RGS12 gene products. *Journal of Biological Chemistry,* 275**,** 29660-29671.

CHEONG, K. H., ZACCHETTI, D., SCHNEEBERGER, E. E. & SIMONS, K. 1999. VIP17/MAL, a lipid raft-associated protein, is involved in apical transport in MDCK cells. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 6241-6248.

CHIANG, A. P., BECK, J. S., YEN, H. J., TAYEH, M. K., SCHEETZ, T. E., SWIDERSKI, R. E., NISHIMURA, D. Y., BRAUN, T. A., KIM, K. Y. A., HUANG, J., ELBEDOUR, K., CARMI, R., SLUSARSKI, D. C., CASAVANT, T. L., STONE, E. M. &

SHEFFIELD, V. C. 2006. Homozygosity mapping with SNP arrays identifies TRIM32 an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 6287-6292.

CLAYTON, B. L. L. & POPKO, B. 2016. Endoplasmic reticulum stress and the unfolded protein response in disorders of myelinating glia. *Brain Research,* 1648**,** 594-602.

COLLIN, R. W. J., DEN HOLLANDER, A. I., VAN DER VELDE-VISSER, S. D., BENNICELLI, J., BENNETT, J. & CREMERS, F. P. M. 2012. Antisense Oligonucleotide (AON)-based Therapy for Leber Congenital Amaurosis Caused by a Frequent Mutation in CEP290. *Molecular Therapy-Nucleic Acids,* 1.

CONG, L., RAN, F. A., COX, D., LIN, S. L., BARRETTO, R., HABIB, N., HSU, P. D., WU, X. B., JIANG, W. Y., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science,* 339**,** 819-823.

CONRAD, D. F., PINTO, D., REDON, R., FEUK, L., GOKCUMEN, O., ZHANG, Y. J., AERTS, J., ANDREWS, T. D., BARNES, C., CAMPBELL, P., FITZGERALD, T., HU, M., IHM, C. H., KRISTIANSSON, K., MACARTHUR, D. G., MACDONALD, J. R., ONYIAH, I., PANG, A. W. C., ROBSON, S., STIRRUPS, K., VALSESIA, A., WALTER, K., WEI, J., TYLER-SMITH, C., CARTER, N. P., LEE, C., SCHERER, S. W., HURLES, M. E. & WELLCOME TRUST CASE, C. 2010. Origins and functional impact of copy number variation in the human genome. *Nature,* 464**,** 704-712.

COSTAIN, G., WALKER, S., ARGIROPOULOS, B., BARIBEAU, D. A., BASSETT, A. S., BOOT, E., DEVRIENDT, K., KELLAM, B., MARSHALL, C. R., PRASAD, A., SERRANO, M. A., STAVROPOULOS, D. J., TWEDE, H., VERMEESCH, J. R., VORSTMAN, J. A. S. & SCHERER, S. W. 2019. Rare copy number variations affecting the synaptic gene DMXL2 in neurodevelopmental disorders. *Journal of Neurodevelopmental Disorders,* 11.

COUGHLIN, C. R., SCHARER, G. H., FRIEDERICH, M. W., YU, H. C., GEIGER, E. A., CREADON-SWINDELL, G., COLLINS, A. E., VANLANDER, A. V., VAN COSTER, R., POWELL, C. A., SWANSON, M. A., MINCZUK, M., VAN HOVE, J. L. K. & SHAIKH, T. H. 2015. Mutations in the mitochondrial cysteinyl-tRNA synthase gene, CARS2, lead to a severe epileptic encephalopathy and complex movement disorder. *Journal of Medical Genetics,* 52**,** 532-540.

COX, J. & MANN, M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology,* 26**,** 1367-1372.

COX, J., MATIC, I., HILGER, M., NAGARAJ, N., SELBACH, M., OLSEN, J. V. & MANN, M. 2009. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature Protocols,* 4**,** 698-705.

CUBILLOS-ROJAS, M., SCHNEIDER, T., HADJEBI, O., PEDRAZZA, L., DE OLIVEIRA, J. R., LANGA, F., GUENET, J. L., DURAN, J., DE ANTA, J. M., ALCANTARA, S., RUIZ, R., PEREZ-VILLEGAS, E. M., AGUILAR-MONTILLA, F. J., CARRION, A. M., ARMENGOL, J. A., BAPLE, E., CROSBY, A. H., BARTRONS, R., VENTURA, F. & ROSA, J. L. 2016. The HERC2 ubiquitin ligase is essential for embryonic development and regulates motor coordination. *Oncotarget,* 7**,** 56083-56106.

CUI, C. G., LI, S. J. & WU, D. 2019. Znhit1 inhibits breast cancer by up-regulating PTEN to deactivate the PI3K/ Akt/mTOR pathway. *Life Sciences,* 224**,** 204-211.

CYRANOSKI, D. 2019. WHAT'S NEXT FOR CRISPR BABIES? *Nature,* 566**,** 440-442.

DAMMERMANN, A. & MERDES, A. 2002. Assembly of centrosomal proteins and microtubule organization depends on PCM-1. *Journal of Cell Biology,* 159**,** 255-266.

DEBRAY, F. G., MORIN, C., JANVIER, A., VILLENEUVE, J., MARANDA, B., LAFRAMBOISE, R., LACROIX, J., DECARIE, J. C., ROBITAILLE, Y., LAMBERT, M., ROBINSON, B. H. & MITCHELL, G. A. 2011. LRPPRC mutations cause a phenotypically distinct form of Leigh syndrome with cytochrome c oxidase deficiency. *Journal of Medical Genetics,* 48**,** 183-189.

DENTON, R. 2009. Regulation of mitochondrial dehydrogenases by calcium ions. *Biochimica Et Biophysica Acta-Bioenergetics,* 1787.

DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNELL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics,* 43**,** 491-+.

DEVER, T. E. 2002. Gene-specific regulation by general translation factors. *Cell,* 108**,** 545-556.

DHAUNCHAK, A. S. & NAVE, K. A. 2007. A common mechanism of PLP/DM20 misfolding causes cysteine-mediated endoplasmic reticulum retention in oligodendrocytes and Pelizaeus-Merzbacher disease. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 17813-17818.

DIEHL, H. J., SCHAICH, M., BUDZINSKI, R. M. & STOFFEL, W. 1986. INDIVIDUAL EXONS ENCODE THE INTEGRAL MEMBRANE DOMAINS OF HUMAN

MYELIN PROTEOLIPID PROTEIN. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 9807-9811.

DMOCHOWSKA, A., STANKIEWICZ, P., GOLIK, P., STEPIEN, P. P., BOCIAN, E., HANSMANN, I. & BARTNIK, E. 1998. Assignment of SUPV3L1 to human chromosome band 10q22.1 by in situ hybridization. *Cytogenetics and Cell Genetics,* 83**,** 84-85.

DOBOLYI, A., DIMITROV, E., PALKOVITS, M. & USDIN, T. B. 2012. The neuroendocrine functions of the parathyroid hormone 2 receptor. *Frontiers in endocrinology,* 3**,** 121.

EGEA, J. & KLEIN, R. 2007. Bidirectional Eph-ephrin signaling during axon guidance. *Trends in Cell Biology,* 17**,** 230-238.

ELMORE, S. 2007. Apoptosis: A Review of Programmed Cell Death. *Toxicologic Pathology* 35.

ENGEL, W. K., MAHLER, R., HUXLEY, H. E., FOSTER, J. B. & HUGHES, B. P. 1961. CENTRAL CORE DISEASE-AN INVESTIGATION OF A RARE MUSCLE CELL ABNORMALITY. *Brain,* 84**,** 167-&.

FIELDS, R. D. 2014. Myelin Formation and Remodeling. *Cell,* 156**,** 15-17.

FINGER, J. H., SMITH, C. M., HAYAMIZU, T. F., MCCRIGHT, I. J., XU, J. X., LAW, M., SHAW, D. R., BALDARELLI, R. M., BEAL, J. S., BLODGETT, O., CAMPBELL, J. W., CORBANI, L. E., LEWIS, J. R., FORTHOFER, K. L., FROST, P. J., GIANNATTO, S. C., HUTCHINS, L. N., MIERS, D. B., MOTENKO, H., STONE, K. R., EPPIG, J. T., KADIN, J. A., RICHARDSON, J. E. & RINGWALD, M. 2017. The mouse Gene Expression Database (GXD): 2017 update. *Nucleic Acids Research,* 45**,** D730-D736.

FIRTH, H. V., RICHARDS, S. M., BEVAN, A. P., CLAYTON, S., CORPAS, M., RAJAN, D., VAN VOOREN, S., MOREAU, Y., PETTETT, R. M. & CARTER, N. P. 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics,* 84**,** 524-533.

FLORES, A. I., NARAYANAN, S. P., MORSE, E. N., SHICK, H. E., YIN, X. H., KIDD, G., AVILA, R. L., KIRSCHNER, D. A. & MACKLIN, W. B. 2008. Constitutively active Akt induces enhanced myelination in the CNS. *Journal of Neuroscience,* 28**,** 7174-7183.

FORTH, S. & KAPOOR, T. M. 2017. The mechanics of microtubule networks in cell division. *Journal of Cell Biology,* 216**,** 1525-1531.

FOUST, K. D., NURRE, E., MONTGOMERY, C. L., HERNANDEZ, A., CHAN, C. M. & KASPAR, B. K. 2009. Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes. *Nature Biotechnology,* 27**,** 59-65.

FOUST, K. D., WANG, X. Y., MCGOVERN, V. L., BRAUN, L., BEVAN, A. K., HAIDET, A. M., LE, T. T., MORALES, P. R., RICH, M. M., BURGHES, A. H. M. & KASPAR, B. K. 2010. Rescue of the spinal muscular atrophy phenotype in a mouse model by early postnatal delivery of SMN. *Nature Biotechnology,* 28**,** 271-U126.

FRANCONE, V. P., MAGGIPINTO, M. J., KOSTURKO, L. D. & BARBARESE, E. 2007. The microtubule-associated protein tumor overexpressed gene/cytoskeleton-associated protein 5 is necessary for myelin basic protein expression in oligodendrocytes. *Journal of Neuroscience,* 27**,** 7654-7662.

FRANK, M. 2000. MAL, a proteolipid in glycosphingolipid enriched domains: functional implications in myelin and beyond. *Progress in Neurobiology,* 60**,** 531-544.

FRIEDLAND, A. E., TZUR, Y. B., ESVELT, K. M., COLAIACOVO, M. P., CHURCH, G. M. & CALARCO, J. A. 2013. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nature Methods,* 10**,** 741-+.

FRIEDMAN, J. S., RAY, J. W., WASEEM, N., JOHNSON, K., BROOKS, M. J., HUGOSSON, T., BREUER, D., BRANHAM, K. E., KRAUTH, D. S., BOWNE, S. J., SULLIVAN, L. S., PONJAVIC, V., GRANSE, L., KHANNA, R., TRAGER, E. H., GIESER, L. M., HUGHBANKS-WHEATON, D., COJOCARU, R. I., GHIASVAND, N. M., CHAKAROVA, C. F., ABRAHAMSON, M., GORING, H. H. H., WEBSTER, A. R., BIRCH, D. G., ABECASIS, G. R., FANN, Y., BHATTACHARYA, S. S., DAIGER, S. P., HECKENLIVELY, J. R., ANDREASSON, S. & SWAROOP, A. 2009. Mutations in a BTB-Kelch Protein, KLHL7, Cause Autosomal-Dominant Retinitis Pigmentosa. *American Journal of Human Genetics,* 84**,** 792-800.

GADADHAR, S., BODAKUNTLA, S., NATARAJAN, K. & JANKE, C. 2017. The tubulin code at a glance. *Journal of Cell Science,* 130**,** 1347-1353.

GALLIGAN, J. T., MARTINEZ-NOEL, G., ARNDT, V., HAYES, S., CHITTENDEN, T. W., HARPER, J. W. & HOWLEY, P. M. 2015. Proteomic Analysis and Identification of Cellular Interactors of the Giant Ubiquitin Ligase HERC2. *Journal of Proteome Research,* 14**,** 953-966.

GARCIA-CANO, J., MARTINEZ-MARTINEZ, A., SALA-GASTON, J., PEDRAZZA, L. & ROSA, J. L. 2019. HERCing: Structural and Functional Relevance of the Large HERC Ubiquitin Ligases. *Frontiers in Physiology,* 10.

GARDNER, J. F., CUSHION, T. D., NIOTAKIS, G., OLSON, H. E., GRANT, P. E., SCOTT, R. H., STOODLEY, N., COHEN, J. S., NAIDU, S., ATTIE-BITACH, T., BONNIERES, M., BOUTAUD, L., ENCHA-RAZAVI, F., PALMER-SMITH, S. M., MUGALAASI, H., MULLINS, J. G. L., PILZ, D. T. & FRY, A. E. 2018. Clinical and Functional Characterization of the Recurrent TUBA1A p.(Arg2His) Mutation. *Brain Sciences,* 8.

GARROD, A. E. 1902. About Alkaptonuria. *Medico-chirurgical transactions,* 85**,** 69-78.

GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research,* 31**,** 3784-3788.

GAUTIER, R., DOUGUET, D., ANTONNY, B. & DRIN, G. 2008. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics,* 24**,** 2101-2102.

GEORGE, A. J., HOFFIZ, Y. C., CHARLES, A. J., ZHU, Y. & MABB, A. M. 2018. A Comprehensive Atlas of E3 Ubiquitin Ligase Mutations in Neurological Disorders. *Frontiers in Genetics,* 9.

GHEZZI, D. & ZEVIANI, M. 2018. Human diseases associated with defects in assembly of OXPHOS complexes. *In:* GARONE, C. & MINCZUK, M. (eds.) *Mitochondrial Diseases.*

GLOECKNER, C. J., BOLDT, K. & UEFFING, M. 2009. Strep/FLAG tandem affinity purification (SF-TAP) to study protein interactions. *Current protocols in protein science,* Chapter 19**,** Unit19.20.

GOEBBELS, S., OLTROGGE, J. H., KEMPER, R., HEILMANN, I., BORMUTH, I., WOLFER, S., WICHERT, S. P., MOBIUS, W., LIU, X., LAPPE-SIEFKE, C., ROSSNER, M. J., GROSZER, M., SUTER, U., FRAHM, J., BORETIUS, S. & NAVE, K. A. 2010. Elevated Phosphatidylinositol 3,4,5-Trisphosphate in Glia Triggers Cell-Autonomous Membrane Wrapping and Myelination. *Journal of Neuroscience,* 30**,** 8953-8964.

GOEMANS, N. M., TULINIUS, M., VAN DEN HAUWE, M., KROKSMARK, A. K., BUYSE, G., WILSON, R. J., VAN DEUTEKOM, J. C., DE KIMPE, S. J., LOURBAKOS, A. & CAMPION, G. 2016. Long-Term Efficacy, Safety, and Pharmacokinetics of Drisapersen in Duchenne Muscular Dystrophy: Results from an Open-Label Extension Study. *Plos One,* 11.

GONZALEZ-PEREZ, A. & LOPEZ-BIGAS, N. 2011. Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, Condel. *American Journal of Human Genetics,* 88**,** 440-449.

GRAZIANO, A. C. E., PANNUZZO, G., AVOLA, R. & CARDILE, V. 2016. Chaperones as Potential Therapeutics for Krabbe Disease. *Journal of Neuroscience Research,* 94**,** 1220-1230.

GUAN, X. M. & FIERKE, C. A. 2011. Understanding protein palmitoylation: Biological significance and enzymology. *Science China-Chemistry,* 54**,** 1888-1897.

GUARGUAGLINI, G., DUNCAN, P. I., STIERHOF, Y. D., HOLMSTROM, T., DUENSING, S. & NIGG, E. A. 2005. The forkhead-associated domain protein

Cep170 interacts with polo-like kinase 1 and serves as a marker for mature centrioles. *Molecular Biology of the Cell,* 16**,** 1095-1107.

GUDBJARTSSON, D. F., JONASSON, K., FRIGGE, M. L. & KONG, A. 2000. Allegro, a new computer program for multipoint linkage analysis. *Nature Genetics,* 25**,** 12-13.

GUDLUR, A., ZERAIK, A. E., HIRVE, N., RAJANIKANTH, V., BOBKOV, A. A., MA, G. L., ZHENG, S. S., WANG, Y. J., ZHOU, Y. B., KOMIVES, E. A. & HOGAN, P. G. 2018. Calcium sensing by the STIM1 ER-luminal domain. *Nature Communications,* 9.

GUNDERSON, K. L., KRUGLYAK, S., GRAIGE, M. S., GARCIA, F., KERMANI, B. G., ZHAO, C. F., CHE, D. P., DICKINSON, T., WICKHAM, E., BIERLE, J., DOUCET, D., MILEWSKI, M., YANG, R., SIEGMUND, C., HAAS, J., ZHOU, L. X., OLIPHANT, A., FAN, J. B., BARNARD, S. & CHEE, M. S. 2004. Decoding randomly ordered DNA arrays. *Genome Research,* 14**,** 870-877.

HARDING, H. P., ZHANG, Y. H. & RON, D. 1999. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature,* 397**,** 271-274.

HARLALKA, G. V., BAPLE, E. L., CROSS, H., KUHNLE, S., CUBILLOS-ROJAS, M., MATENTZOGLU, K., PATTON, M. A., WAGNER, K., COBLENTZ, R., FORD, D. L., MACKAY, D. J. G., CHIOZA, B. A., SCHEFFNER, M., ROSA, J. L. & CROSBY, A. H. 2013. Mutation of HERC2 causes developmental delay with Angelman-like features. *Journal of Medical Genetics,* 50**,** 65-73.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.

HEPLER, J. R. 1999. Emerging roles for RGS proteins in cell signalling. *Cell Press,* 20**,** 376-382.

HERST, P. M., ROWE, M. R., CARSON, G. M. & BERRIDGE, M. V. 2017. Functional Mitochondria in Health and Disease. *Frontiers in Endocrinology,* 8.

HOFFMAN-ZACHARSKA, D., MIERZEWSKA, H., SZCZEPANIK, E., POZNANSKI, J., MAZURCZAK, T., JAKUBIUK-TOMASZUK, A., MADRY, J., KIERDASZUK, A. & BAL, J. 2013. The spectrum of PLP1 gene mutations in patients with the classical form of the Pelizaeus-Merzbacher disease. *Medycyna wieku rozwojowego,* 17**,** 293-300.

HOLLEY, R. W., APGAR, J., EVERETT, G. A., MADISON, J. T., MARQUISEE, M., MERRILL, S. H., PENSWICK, J. R. & ZAMIR, A. 1965. STRUCTURE OF A RIBONUCLEIC ACID. *Science,* 147**,** 1462-+.

HOLTAN, J. P., TEIGEN, K., AUKRUST, I., BRAGADOTTIR, R. & HOUGE, G. 2019. Dominant ARL3-related retinitis pigmentosa. *Ophthalmic genetics,* 40**,** 124-128.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research,* 37**,** 1-13.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols,* 4**,** 44-57.

HUG, N., LONGMAN, D. & CACERES, J. F. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research,* 44**,** 1483-1495.

HUGOSSON, T., FRIEDMAN, J. S., PONJAVIC, V., ABRAHAMSON, M., SWAROOP, A. & ANDREASSON, S. 2010. Phenotype Associated With Mutation in the Recently Identified Autosomal Dominant Retinitis Pigmentosa KLHL7 Gene. *Archives of Ophthalmology,* 128**,** 772-778.

HUUSKO, P., PONCIANO-JACKSON, D., WOLF, M., KIEFER, J. A., AZORSA, D. O., TUZMEN, S., WEAVER, D., ROBBINS, C., MOSES, T., ALLINEN, M., HAUTANIEMI, S., CHEN, Y. D., ELKAHLOUN, A., BASIK, M., BOVA, G. S., BUBENDORF, L., LUGLI, A., SAUTER, G., SCHLEUTKER, J., OZCELIK, H., ELOWE, S., PAWSON, T., TRENT, J. M., CARPTEN, J. D., KALLIONIEMI, O. P. & MOUSSES, S. 2004. Nonsense-mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. *Nature Genetics,* 36**,** 979-983.

IANNITTI, T. & PALMIERI, B. 2011. Clinical and experimental applications of sodium phenylbutyrate. *Drugs in R&D,* 11**,** 227-49.

IHARA, S., NAKAYAMA, S., MURAKAMI, Y., SUZUKI, E., ASAKAWA, M., KINOSHITA, T. & SAWA, H. 2017. PIGN prevents protein aggregation in the endoplasmic reticulum independently of its function in the GPI synthesis. *Journal of Cell Science,* 130**,** 602-613.

INOUE, K. 2017. Cellular Pathology of Pelizaeus-Merzbacher Disease Involving Chaperones Associated with Endoplasmic Reticulum Stress. *Frontiers in Molecular Biosciences,* 4.

ITOH-SATOH, M., HAYASHI, T., NISHI, H., KOGA, Y., ARIMURA, T., KOYANAGI, T., TAKAHASHI, M., HOHDA, S., UEDA, K., NOUCHI, T., HIROE, M., MARUMO, F., IMAIZUMI, T., YASUNAMI, M. & KIMURA, A. 2002. Titin mutations as the molecular basis for dilated cardiomyopathy. *Biochemical and Biophysical Research Communications,* 291**,** 385-393.

JAGANATHAN, K., PANAGIOTOPOULOU, S. K., MCRAE, J. F., DARBANDI, S. F., KNOWLES, D., LI, Y. I., KOSMICKI, J. A., ARBELAEZ, J., CUI, W. W., SCHWARTZ, G. B., CHOW, E. D., KANTERAKIS, E., GAO, H., KIA, A., BATZOGLOU, S., SANDERS, S. J. & FARH, K. K. H. 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell,* 176**,** 535-+.

JANSEN, A. C., OOSTRA, A., DESPRECHINS, B., DE VLAEMINCK, Y., VERHELST, H., REGAL, L., VERLOO, P., BOCKAERT, N., KEYMOLEN, K., SENECA, S., DE MEIRLEIR, L. & LISSENS, W. 2011. TUBA1A mutations From isolated lissencephaly to familial polymicrogyria. *Neurology,* 76**,** 988-992.

JEFFRIES, L., OLIVIERI, J. E., JI, W., SPENCER-MANZON, M., BALE, A., KONSTANTINO, M. & LAKHANI, S. A. 2018. Two siblings with a novel nonsense variant provide further delineation of the spectrum of recessive KLHL7 diseases. *European journal of medical genetics*.

JOHNSON, C. A. M., E. R.; 2012. Autosomal recessive conditions and autozygosity mapping. *In:* KUMAR, D. (ed.) *Genomics & Health in the Developing world.* Oxford University Press.

JORGENSEN, C., SHERMAN, A., CHEN, G. I., PASCULESCU, A., POLIAKOV, A., HSIUNG, M., LARSEN, B., WILKINSON, D. G., LINDING, R. & PAWSON, T. 2009. Cell-Specific Information Processing in Segregating Populations of Eph Receptor Ephrin-Expressing Cells. *Science,* 326**,** 1502-1509.

JUNG, M., SOMMER, I., SCHACHNER, M. & NAVE, K. A. 1996. Monoclonal antibody O10 defines a conformationally sensitive cell-surface epitope of proteolipid protein (PLP): Evidence that PLP misfolding underlies dysmyelination in mutant mice. *Journal of Neuroscience,* 16**,** 7920-7929.

KAMER, K. J. & MOOTHA, V. K. 2015. The molecular era of the mitochondrial calcium uniporter. *Nature Reviews Molecular Cell Biology,* 16**,** 545-553.

KANEKURA, K., NISHIMOTO, I., AISO, S. & MATSUOKA, M. 2006. Characterization of amyotrophic lateral sclerosis-linked P56S mutation of vesicle-associated membrane protein-associated protein B (VAPB/ALS8). *Journal of Biological Chemistry,* 281**,** 30223-30233.

KANG, T. H., LINDSEY-BOLTZ, L. A., REARDON, J. T. & SANCAR, A. 2010. Circadian control of XPA and excision repair of cisplatin-DNA damage by cryptochrome and HERC2 ubiquitin ligase. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 4890-4895.

KARBOWSKI, M., NEUTZNER, A. & YOULE, R. J. 2007. The mitochondrial E3 ubiquitin ligase MARCH5 is required for Drp1 dependent mitochondrial division. *Journal of Cell Biology,* 178**,** 71-84.

KARCZEWSKI, K. J., FRANCIOLI, L. C., TIAO, G., CUMMINGS, B. B., ALFÖLDI, J., WANG, Q., COLLINS, R. L., LARICCHIA, K. M., GANNA, A., BIRNBAUM, D. P., GAUTHIER, L. D., BRAND, H., SOLOMONSON, M., WATTS, N. A., RHODES, D., SINGER-BERK, M., ENGLAND, E. M., SEABY, E. G., KOSMICKI, J. A., WALTERS, R. K., TASHMAN, K., FARJOUN, Y., BANKS, E., POTERBA, T., WANG, A., SEED, C., WHIFFIN, N., CHONG, J. X., SAMOCHA, K. E., PIERCE-

HOFFMAN, E., ZAPPALA, Z., O'DONNELL-LURIA, A. H., MINIKEL, E. V., WEISBURD, B., LEK, M., WARE, J. S., VITTAL, C., ARMEAN, I. M., BERGELSON, L., CIBULSKIS, K., CONNOLLY, K. M., COVARRUBIAS, M., DONNELLY, S., FERRIERA, S., GABRIEL, S., GENTRY, J., GUPTA, N., JEANDET, T., KAPLAN, D., LLANWARNE, C., MUNSHI, R., NOVOD, S., PETRILLO, N., ROAZEN, D., RUANO-RUBIO, V., SALTZMAN, A., SCHLEICHER, M., SOTO, J., TIBBETTS, K., TOLONEN, C., WADE, G., TALKOWSKI, M. E., NEALE, B. M., DALY, M. J. & MACARTHUR, D. G. 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210.

KEAYS, D. A., TIAN, G., POIRIER, K., HUANG, G. J., SIEBOLD, C., CLEAK, J., OLIVER, P. L., FRAY, M., HARVEY, R. J., MOLNAR, Z., PINON, M. C., DEAR, N., VALDAR, W., BROWN, S. D. M., DAVIES, K. E., RAWLINS, J. N. P., COWAN, N. J., NOLAN, P., CHELLY, J. & FLINT, J. 2007. Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans. *Cell,* 128**,** 45-57.

KELWICK, R., DESANLIS, I., WHEELER, G. N. & EDWARDS, D. R. 2015. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biology,* 16.

KHAN, K. N., ALI, M., POULTER, J. A., MCKIBBIN, M. & INGLEHEARN, C. F. 2013. Patterns of inheritance, not always easily visible. *Bmj-British Medical Journal,* 347.

KHIDR, L., WU, G. K., DAVILA, A., PROCACCIO, V., WALLACE, D. & LEE, W. H. 2008. Role of SUV3 helicase in maintaining mitochondrial Homeostasis in human cells. *Journal of Biological Chemistry,* 283**,** 27064-27073.

KIGOSHI, Y., TSURUTA, F. & CHIBA, T. 2011. Ubiquitin Ligase Activity of Cul3-KLHL7 Protein Is Attenuated by Autosomal Dominant Retinitis Pigmentosa Causative Mutation. *Journal of Biological Chemistry,* 286**,** 33613-33621.

KIM, D. I., JENSEN, S. C., NOBLE, K. A., BIRENDRA, K. C., ROUX, K. H., MOTAMEDCHABOKI, K. & ROUX, K. J. 2016. An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell,* 27**,** 1188-1196.

KIM, T., FIEDLER, K., MADISON, D. L., KRUEGER, W. H. & PFEIFFER, S. E. 1996. Cloning and characterization of MVP17: A developmentally regulated myelin protein in oligodendrocytes. *Journal of Neurochemistry,* 66**,** S5-S5.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics,* 46**,** 310-+.

KLEIGER, G. & MAYOR, T. 2014. Perilous journey: a tour of the ubiquitin-proteasome system. *Trends in Cell Biology,* 24**,** 352-359.

KLEIN, R. 2009. Bidirectional modulation of synaptic functions by Eph/ephrin signaling. *Nature Neuroscience,* 12**,** 15-20.

KREMER, L. S., BADER, D. M., MERTES, C., KOPAJTICH, R., PICHLER, G., IUSO, A., HAACK, T. B., GRAF, E., SCHWARZMAYR, T., TERRILE, C., KONARIKOVA, E., REPP, B., KASTENMULLER, G., ADAMSKI, J., LICHTNER, P., LEONHARDT, C., FUNALOT, B., DONATI, A., TIRANTI, V., LOMBES, A., JARDEL, C., GLASER, D., TAYLOR, R. W., GHEZZI, D., MAYR, J. A., ROTIG, A., FREISINGER, P., DISTELMAIER, F., STROM, T. M., MEITINGER, T., GAGNEUR, J. & PROKISCH, H. 2017. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications,* 8.

KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E. L. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology,* 305**,** 567-580.

KUBO, A. & TSUKITA, S. 2003. Non-membranous granular organelle consisting of PCM-1: subcellular distribution and cell-cycledependent assembly/disassembly. *Journal of Cell Science,* 116**,** 919-928.

KUHNLE, S., KOGEL, U., GLOCKZIN, S., MARQUARDT, A., CIECHANOVER, A., MATENTZOGLU, K. & SCHEFFNER, M. 2011. Physical and Functional Interaction of the HECT Ubiquitin-protein Ligases E6AP and HERC2. *Journal of Biological Chemistry,* 286**,** 19410-19416.

LANDER, E. S. & BOTSTEIN, D. 1987. HOMOZYGOSITY MAPPING - A WAY TO MAP HUMAN RECESSIVE TRAITS WITH THE DNA OF INBRED CHILDREN. *Science,* 236**,** 1567-1570.

LANPHIER, E., URNOV, F., HAECKER, S. E., WERNER, M. & SMOLENSKI, J. 2015. Don't edit the human germ line. *Nature,* 519**,** 410-411.

LASFARGUES, C., MARTINEAU, Y., BOUSQUET, C. & PYRONNET, S. 2013. Changes in Translational Control after Pro-Apoptotic Stress. *International Journal of Molecular Sciences,* 14.

LE HIR, H., GATFIELD, D., IZAURRALDE, E. & MOORE, M. J. 2001. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *Embo Journal,* 20**,** 4987-4997.

LEE, J. H., RHO, S. B. & CHUN, T. 2005. Programmed cell death 6 (PDCD6) protein interacts with death-associated protein kinase 1 (DAPk1): additive effect on apoptosis via caspase-3 dependent pathway. *Biotechnology Letters,* 27**,** 1011-1015.

LEE, T. H., PARK, J. M., LEEM, S. H. & KANG, T. H. 2014. Coordinated regulation of XPA stability by ATR and HERC2 during nucleotide excision repair. *Oncogene,* 33**,** 19-25.

LEE, Y., MIN, C. K., KIM, T. G., SONG, H. K., LIM, Y., KIM, D., SHIN, K., KANG, M., KANG, J. Y., YOUN, H. S., LEE, J. G., AN, J. Y., PARK, K. R., LIM, J. J., KIM, J. H., KIM, J. H., PARK, Z. Y., KIM, Y. S., WANG, J. M., KIM, D. H. & EOM, S. H. 2015. Structure and function of the N-terminal domain of the human mitochondrial calcium uniporter. *Embo Reports,* 16**,** 1318-1333.

LEWIS-SMITH, D., KAMER, K. J., GRIFFIN, H., CHILDS, A.-M., PYSDEN, K., TITOV, D., DUFF, J., PYLE, A., TAYLOR, R. W., YU-WAI-MAN, P., RAMESH, V., HORVATH, R., MOOTHA, V. K. & CHINNERY, P. F. 2016. Homozygous deletion in MICU1 presenting with fatigue and lethargy in childhood. *Neurology. Genetics,* 2**,** e59.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-1760.

LI, W. & YE, Y. 2008. Polyubiquitin chains: functions, structures, and mechanisms. *Cellular and Molecular Life Sciences,* 65**,** 2397-2406.

LIM, K. R. Q., MARUYAMA, R. & YOKOTA, T. 2017. Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Design Development and Therapy,* 11**,** 533-545.

LIU, H. X., CARTEGNI, L., ZHANG, M. Q. & KRAINER, A. R. 2001. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nature Genetics,* 27**,** 55-58.

LIU, J. C., LIU, J., HOLMSTROM, K. M., MENAZZA, S., PARKS, R. J., FERGUSSON, M. M., YU, Z. X., SPRINGER, D. A., HALSEY, C., LIU, C., MURPHY, E. & FINKEL, T. 2016. MICU1 Serves as a Molecular Gatekeeper to Prevent In Vivo Mitochondrial Calcium Overload. *Cell Reports,* 16**,** 1561-1573.

LOGAN, C. V., SZABADKAI, G., SHARPE, J. A., PARRY, D. A., TORELLI, S., CHILDS, A. M., KRIEK, M., PHADKE, R., JOHNSON, C. A., ROBERTS, N. Y., BONTHRON, D. T., PYSDEN, K. A., WHYTE, T., MUNTEANU, I., FOLEY, A. R., WHEWAY, G., SZYMANSKA, K., NATARAJAN, S., ABDELHAMED, Z. A., MORGAN, J. E., ROPER, H., SANTEN, G. W. E., NIKS, E. H., VAN DER POL, W. L., LINDHOUT, D., RAFFAELLO, A., DE STEFANI, D., DEN DUNNEN, J. T., SUN, Y., GINJAAR, I., SEWRY, C. A., HURLES, M., RIZZUTO, R., DUCHEN, M. R., MUNTONI, F., SHERIDAN, E. & CONSORTIUM, U. K. 2014. Loss-of-function mutations in MICU1 cause a brain and muscle disorder linked to primary alterations in mitochondrial calcium signaling. *Nature Genetics,* 46**,** 188-+.

LOZANO-URENA, A. & FERRON, S. R. 2019. Beyond protein-coding genes. *Elife,* 8.

LUDERS, K. A., NESSLER, S., KUSCH, K., PATZIG, J., JUNG, R. B., MOBIUS, W., NAVE, K. A. & WERNER, H. B. 2019. Maintenance of high proteolipid protein level in adult central nervous system myelin is required to preserve the integrity of myelin and axons. *Glia,* 67**,** 634-649.

LUSE, S. A. 1959. The fine structure of the morphogenesis of myelin. *Progress in neurobiology,* 4**,** 59-95.

MA, Y. J., BREWER, J. W., DIEHL, J. A. & HENDERSHOT, L. M. 2002. Two distinct stress signaling pathways converge upon the CHOP promoter during the mammalian unfolded protein response. *Journal of Molecular Biology,* 318**,** 1351-1365.

MADHAVAN, M., NEVIN, Z. S., SHICK, H. E., GARRISON, E., CLARKSON-PAREDES, C., KARL, M., CLAYTON, B. L. L., FACTOR, D. C., ALLAN, K. C., BARBAR, L., JAIN, T., DOUVARAS, P., FOSSATI, V., MILLER, R. H. & TESAR, P. J. 2018. Induction of myelinating oligodendrocytes in human cortical spheroids. *Nature Methods,* 15**,** 700-+.

MAEDER, M. L., STEFANIDAKIS, M., WILSON, C. J., BARAL, R., BARRERA, L. A., BOUNOUTAS, G. S., BUMCROT, D., CHAO, H. S., CIULLA, D. M., DASILVA, J. A., DASS, A., DHANAPAL, V., FENNELL, T. J., FRIEDLAND, A. E., GIANNOUKOS, G., GLOSKOWSKI, S. W., GLUCKSMANN, A., GOTTA, G. M., JAYARAM, H., HASKETT, S. J., HOPKINS, B., HORNG, J. E., JOSHI, S., MARCO, E., MEPANI, R., REYON, D., TA, T., TABBAA, D. G., SAMUELSSON, S. J., SHEN, S., SKOR, M. N., STETKIEWICZ, P., WANG, T. Y., YUDKOFF, C., MYER, V. E., ALBRIGHT, C. F. & JIANG, H. Y. 2019. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nature Medicine,* 25**,** 229-+.

MARAZUELA, M. & ALONSO, M. A. 2004. Expression of MAL and MAL2, two elements of the protein machinery for raft-mediated transport, in normal and neoplastic human tissue. *Histology and Histopathology,* 19**,** 925-933.

MARCHI, S. & PINTON, P. 2014. The mitochondrial calcium uniporter complex: molecular components, structure and physiopathological implications. *Journal of Physiology-London,* 592**,** 829-839.

MARTIN-ALMEDINA, S., MARTINEZ-CORRAL, I., HOLDHUS, R., VICENTE, A., FOTIOU, E., LIN, S., PETERSEN, K., SIMPSON, M. A., HOISCHEN, A., GILISSEN, C., JEFFERY, H., ATTON, G., KARAPOULIOU, C., BRICE, G., GORDON, K., WISEMAN, J. W., WEDIN, M., ROCKSON, S. G., JEFFERY, S., MORTIMER, P. S., SNYDER, M. P., BERLAND, S., MANSOUR, S., MAKINEN, T. & OSTERGAARD, P. 2016. EPHB4 kinase-inactivating mutations cause

autosomal dominant lymphatic-related hydrops fetalis. *Journal of Clinical Investigation,* 126**,** 3080-3088.

MARTIN-MCCAFFREY, L., HAINS, M. D., PRITCHARD, G. A., PAJAK, A., DAGNINO, L., SIDEROVSKI, D. P. & D'SOUZA, S. J. A. 2005. Differential expression of regulator of G-protein signaling R12 subfamily members during mouse development. *Developmental Dynamics,* 234**,** 438-444.

MAXAM, A. M. & GILBERT, W. 1977. NEW METHOD FOR SEQUENCING DNA. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 560-564.

MCLAREN, W., PRITCHARD, B., RIOS, D., CHEN, Y. A., FLICEK, P. & CUNNINGHAM, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics,* 26**,** 2069-2070.

MCLAUGHLIN, T., FALKOWSKI, M., WANG, J. J. & ZHANG, S. X. 2018. Molecular Chaperone ERp29: A Potential Target for Cellular Protection in Retinal and Neurodegenerative Diseases. *In:* ASH, J. D., ANDERSON, R. E., LAVAIL, M. M., RICKMAN, C. B., HOLLYFIELD, J. G. & GRIMM, C. (eds.) *Retinal Degenerative Diseases: Mechanisms and Experimental Therapy.*

MERZBACHER, L. 1909. New Contributions on "Abnormal Myellnum Seperation" in Cerebral Cortex. *Monatsschrift Fur Psychiatrie Und Neurologie,* 26**,** 1-4.

MIMORI, S., OKUMA, Y., KANEKO, M., KAWADA, K., HOSOI, T., OZAWA, K., NOMURA, Y. & HAMANA, H. 2012. Protective Effects of 4-Phenylbutyrate Derivatives on the Neuronal Cell Death and Endoplasmic Reticulum Stress. *Biological & Pharmaceutical Bulletin,* 35**,** 84-90.

MOOKERJEE, S. A., GONCALVES, R. L. S., GERENCSER, A. A., NICHOLLS, D. G. & BRAND, M. D. 2015. The contributions of respiration and glycolysis to extracellular acid production. *Biochimica Et Biophysica Acta-Bioenergetics,* 1847**,** 171-181.

MORGAN, N. V., BRUETON, L. A., COX, P., GREALLY, M. T., TOLMIE, J., PASHA, S., ALIGIANIS, I. A., VAN BOKHOVEN, H., MARTON, T., AL-GAZALI, L., MORTON, J. E. V., OLEY, C., JOHNSON, C. A., TREMBATH, R. C., BRUNNER, H. G. & MAHER, E. R. 2006. Mutations in the embryonal subunit of the acetylcholine receptor (CHRNG) cause lethal and escobar variants of multiple pterygium syndrome. *American Journal of Human Genetics,* 79**,** 390-395.

MORICE-PICARD, F., BENARD, G., REZVANI, H. R., LASSEAUX, E., SIMON, D., MOUTTON, S., ROORYCK, C., LACOMBE, D., BAUMANN, C. & ARVEILER, B. 2017. Complete loss of function of the ubiquitin ligase HERC2 causes a severe neurodevelopmental phenotype. *European Journal of Human Genetics,* 25**,** 52-58.

MUELLER, R. F. & BISHOP, D. T. 1993. AUTOZYGOSITY MAPPING, COMPLEX CONSANGUINITY, AND AUTOSOMAL RECESSIVE DISORDERS. *Journal of Medical Genetics,* 30**,** 798-799.

NARENDRA, D., TANAKA, A., SUEN, D. F. & YOULE, R. J. 2008. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *Journal of Cell Biology,* 183**,** 795-803.

NELSON, C. E., ROBINSON-HAMM, J. N. & GERSBACH, C. A. 2017. Genome engineering: a new approach to gene therapy for neuromuscular disorders. *Nature Reviews Neurology,* 13**,** 647-661.

NEUMANN, M., KWONG, L. K., SAMPATHU, D. M., TROJANOWSKI, J. Q. & LEE, V. M. Y. 2007. TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis - Protein misfolding diseases without amyloidosis. *Archives of Neurology,* 64**,** 1388-1394.

NG, P. C. & HENIKOFF, S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research,* 31**,** 3812-3814.

NICCHERI, F., PECORI, R. & CONTICELLO, S. G. 2017. An efficient method to enrich for knock-out and knock-in cellular clones using the CRISPR/Cas9 system. *Cellular and Molecular Life Sciences,* 74**,** 3413-3423.

NISHIMURA, A. L., MITNE-NETO, M., SILVA, H. C. A., RICHIERI-COSTA, A., MIDDLETON, S., CASCIO, D., KOK, F., OLIVEIRA, J. R. M., GILLINGWATER, T., WEBB, J., SKEHEL, P. & ZATZ, M. 2004. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *American Journal of Human Genetics,* 75**,** 822-831.

OHTAKE, F., SAEKI, Y., ISHIDO, S., KANNO, J. & TANAKA, K. 2016. The K48-K63 Branched Ubiquitin Chain Regulates NF-kappa B Signaling. *Molecular Cell,* 64**,** 251-266.

OHTAKE, F., TSUCHIYA, H., SAEKI, Y. & TANAKA, K. 2018. K63 ubiquitylation triggers proteasomal degradation by seeding branched ubiquitin chains. *Proceedings of the National Academy of Sciences of the United States of America,* 115**,** E1401-E1408.

ORCESI, S., TONDUTI, D., UGGETTI, C., LARIZZA, D., FAZZI, E. & BALOTTIN, U. 2010. New Case of 4H Syndrome and a Review of the Literature. *Pediatric Neurology,* 42**,** 359-364.

OSLOWSKI, C. M. & URANO, F. 2011. MEASURING ER STRESS AND THE UNFOLDED PROTEIN RESPONSE USING MAMMALIAN TISSUE CULTURE SYSTEM. *In:* CONN, P. M. (ed.) *Methods in Enzymology: Unfolded Protein Response and Cellular Stress, Vol 490, Pt B.*

OSORIO, M. J. & GOLDMAN, S. A. 2018. Neurogenetics of Pelizaeus-Merzbacher disease. *Handbook of clinical neurology,* 148**,** 701-722.

OZGEN, H., SCHRIMPF, W., HENDRIX, J., DE JONGE, J. C., LAMB, D. C., HOEKSTRA, D., KAHYA, N. & BARON, W. 2014. The Lateral Membrane Organization and Dynamics of Myelin Proteins PLP and MBP Are Dictated by Distinct Galactolipids and the Extracellular Matrix. *Plos One,* 9.

PALSULEDESAI, C. C. & DISTEFANO, M. D. 2015. Protein Prenylation: Enzymes, Therapeutics, and Biotechnology Applications. *Acs Chemical Biology,* 10**,** 51-62.

PAPA, S., MARTINO, P. L., CAPITANIO, G., GABALLO, A., DE RASMO, D., SIGNORILE, A. & PETRUZZELLA, V. 2012. The Oxidative Phosphorylation System in Mammalian Mitochondria. *In:* SCATENA, R., BOTTONI, P. & GIARDINA, B. (eds.) *Advances in Mitochondrial Medicine.* Berlin: Springer-Verlag Berlin.

PAPASAVVA, P., KLEANTHOUS, M. & LEDERER, C. W. 2019. Rare Opportunities: CRISPR/Cas-Based Therapy Development for Rare Genetic Diseases. *Molecular Diagnosis & Therapy,* 23**,** 201-222.

PATRON, M., CHECCHETTO, V., RAFFAELLO, A., TEARDO, E., REANE, D. V., MANTOAN, M., GRANATIERO, V., SZABO, I., DE STEFANI, D. & RIZZUTO, R. 2014. MICU1 and MICU2 Finely Tune the Mitochondrial Ca2+ Uniporter by Exerting Opposite Effects on MCU Activity. *Molecular Cell,* 53**,** 726-737.

PAUL, E., CRONAN, R., WESTON, P. J., BOEKELHEIDE, K., SEDIVY, J. M., LEE, S. Y., WIEST, D. L., RESNICK, M. B. & KLYSIK, J. E. 2009. Disruption of Supv3L1 damages the skin and causes sarcopenia, loss of fat, and death. *Mammalian Genome,* 20**,** 92-108.

PECK, S. C. 2006. Analysis of protein phosphorylation: methods and strategies for studying kinases and substrates. *Plant Journal,* 45**,** 512-522.

PEHLIVAN, D., AKDEMIR, Z. C., KARACA, E., BAYRAM, Y., JHANGIANI, S., YILDIZ, E. P., MUZNY, D., ULUC, K., GIBBS, R. A., ELCIOGLU, N., LUPSKI, J. R., HAREL, T. & BAYLOR-HOPKINS CTR, M. 2015. Exome sequencing reveals homozygous TRIM2 mutation in a patient with early onset CMT and bilateral vocal cord paralysis. *Human Genetics,* 134**,** 671-673.

PHILPOT, B. D., THOMPSON, C. E., FRANCO, L. & WILLIAMS, C. A. 2011. Angelman syndrome: advancing the research frontier of neurodevelopmental disorders. *Journal of Neurodevelopmental Disorders,* 3**,** 50-56.

PICKART, C. M. & EDDINS, M. J. 2004. Ubiquitin: structures, functions, mechanisms. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1695**,** 55-72.

PLAGNOL, V., CURTIS, J., EPSTEIN, M., MOK, K. Y., STEBBINGS, E., GRIGORIADOU, S., WOOD, N. W., HAMBLETON, S., BURNS, S. O.,

THRASHER, A. J., KUMARARATNE, D., DOFFINGER, R. & NEJENTSEV, S. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics,* 28**,** 2747-2754.

POWELL, J. E., VISSCHER, P. M. & GODDARD, M. E. 2010. Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics,* 11**,** 800-805.

PREVITALI, S. C., ZEREGA, B., SHERMAN, D. L., BROPHY, P. J., DINA, G., KING, R. H. M., SALIH, M. M., FELTRI, L., QUATTRINI, A., RAVAZZOLO, R., WRABETZ, L., MONACO, A. P. & BOLINO, A. 2003. Myotubularin-related 2 protein phosphatase and neurofilament light chain protein, both mutated in CMT neuropathies, interact in peripheral nerve. *Human Molecular Genetics,* 12**,** 1713-1723.

PROBER, J. M., TRAINOR, G. L., DAM, R. J., HOBBS, F. W., ROBERTSON, C. W., ZAGURSKY, R. J., COCUZZA, A. J., JENSEN, M. A. & BAUMEISTER, K. 1987. A SYSTEM FOR RAPID DNA SEQUENCING WITH FLUORESCENT CHAIN-TERMINATING DIDEOXYNUCLEOTIDES. *Science,* 238**,** 336-341.

PUERTOLLANO, R. & ALONSO, M. A. 1999. MAL, an integral element of the apical sorting machinery, is an itinerant protein that cycles between the trans-Golgi network and the plasma membrane. *Molecular Biology of the Cell,* 10**,** 3435-3447.

PUFFENBERGER, E. G., JINKS, R. N., WANG, H., XIN, B. Z., FIORENTINI, C., SHERMAN, E. A., DEGRAZIO, D., SHAW, C., SOUGNEZ, C., CIBULSKIS, K., GABRIEL, S., KELLEY, R. I., MORTON, D. H. & STRAUSS, K. A. 2012. A homozygous missense mutation in HERC2 associated with global developmental delay and autism spectrum disorder. *Human Mutation,* 33**,** 1639-1646.

ROBINSON, J. T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nature Biotechnology,* 29**,** 24.

RODRIGUES, G. M. C., GAJ, T., ADIL, M. M., WAHBA, J., RAO, A. T., LORBEER, F. K., KULKARNI, R. U., DIOGO, M. M., CABRAL, J. M. S., MILLER, E. W., HOCKEMEYER, D. & SCHAFFER, D. V. 2017. Defined and Scalable Differentiation of Human Oligodendrocyte Precursors from Pluripotent Stem Cells in a 3D Culture System. *Stem Cell Reports,* 8**,** 1770-1783.

ROGON, C., ULBRICHT, A., HESSE, M., ALBERTI, S., VIJAYARAJ, P., BEST, D., ADAMS, I. R., MAGIN, T. M., FLEISCHMANN, B. K. & HOHFELD, J. 2014. HSP70-binding protein HSPBP1 regulates chaperone expression at a posttranslational level and is essential for spermatogenesis. *Molecular Biology of the Cell,* 25**,** 2260-2271.

ROMANIELLO, R. A., F; BASSI, M.T; BORGATTI, R. 2015. Mutations in α- and β-tubulin encoding genes: implications in brain malformations. *Brain Development* 37**,** 273-280.

ROUX, K. J., KIM, D. I., BURKE, B. & MAY, D. G. 2018. BioID: A Screen for Protein-Protein Interactions. *Current protocols in protein science,* 91**,** 19.23.1-19.23.15.

ROUX, K. J., KIM, D. I., RAIDA, M. & BURKE, B. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology,* 196**,** 801-810.

SACCONE, V., PALMIERI, M., PASSAMANO, L., PILUSO, G., MERONI, G., POLITANO, L. & NIGRO, V. 2008. Mutations that impair interaction properties of TRIM32 associated with limb-girdle muscular dystrophy 2H. *Human Mutation,* 29**,** 240-247.

SADIKOVIC, B., FERNANDES, P., ZHANG, V. W., WARD, P. A., MILOSLAVSKAYA, I., RHEAD, W., ROSENBAUM, R., GIN, R., ROA, B. & FANG, P. 2014. Mutation Update for UBE3A Variants in Angelman Syndrome. *Human Mutation,* 35**,** 1407-1417.

SALAUN, C., GREAVES, J. & CHAMBERLAIN, L. H. 2010. The intracellular dynamic of protein palmitoylation. *Journal of Cell Biology,* 191**,** 1229-1238.

SALPIETRO, V., MALINTAN, N. T., LLANO-RIVAS, I., SPAETH, C. G., EFTHYMIOU, S., STRIANO, P., VANDROVCOVA, J., CUTRUPI, M. C., CHIMENZ, R., DAVID, E., DI ROSA, G., MARCE-GRAU, A., RASPALL-CHAURE, M., MARTIN-HERNANDEZ, E., ZARA, F., MINETTI, C., BELLO, O. D., DE ZORZI, R., FORTUNA, S., DAUBER, A., ALKHAWAJA, M., SULTAN, T., MANKAD, K., VITOBELLO, A., THOMAS, Q., MAU-THEM, F. T., FAIVRE, L., MARTINEZ-AZORIN, F., PRADA, C. E., MACAYA, A., KULLMANN, D. M., ROTHMAN, J. E., KRISHNAKUMAR, S. S., HOULDEN, H., DECIPHERING DEV DISORDERS, S. & GRP, S. S. 2019. Mutations in the Neuronal Vesicular SNARE VAMP2 Affect Synaptic Membrane Fusion and Impair Human Neurodevelopment. *American Journal of Human Genetics,* 104**,** 721-730.

SANCAK, Y., MARKHARD, A. L., KITAMI, T., KOVACS-BOGDAN, E., KAMER, K. J., UDESHI, N. D., CARR, S. A., CHAUDHURI, D., CLAPHAM, D. E., LI, A. A., CALVO, S. E., GOLDBERGER, O. & MOOTHA, V. K. 2013. EMRE Is an Essential Component of the Mitochondrial Calcium Uniporter Complex. *Science,* 342**,** 1379-1382.

SANGER, F. & COULSON, A. R. 1989. *A RAPID METHOD FOR DETERMINING SEQUENCES IN DNA BY PRIMED SYNTHESIS WITH DNA POLYMERASE*.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA SEQUENCING WITH CHAIN-TERMINATING INHIBITORS. *Proceedings of the National Academy of Sciences of the United States of America,* 74**,** 5463-5467.

SATO, R., ARAI-ICHINOI, N., KIKUCHI, A., MATSUHASHI, T., NUMATA-UEMATSU, Y., UEMATSU, M., FUJII, Y., MURAYAMA, K., OHTAKE, A., ABE, T. & KURE, S. 2018a. Novel biallelic mutations in the PNPT1 gene encoding a mitochondrial-RNA-import protein PNPase cause delayed myelination. *Clinical Genetics,* 93**,** 242-247.

SATO, T., KATO, M., MORIYAMA, K., HARAGUCHI, K., SAITSU, H., MATSUMOTO, N. & MORIUCHI, H. 2018b. A case of tubulinopathy presenting with porencephaly caused by a novel missense mutation in the TUBA1A gene. *Brain & Development,* 40**,** 819-823.

SAVARESE, M., SARPARANTA, J., VIHOLA, A., UDD, B. & HACKMAN, P. 2016. Increasing Role of Titin Mutations in Neuromuscular Disorders. *Journal of neuromuscular diseases,* 3**,** 293-308.

SCHAEREN-WIEMERS, N., BONNET, A., ERB, M., ERNE, B., BARTSCH, U., KERN, F., MANTEI, N., SHERMAN, D. & SUTER, U. 2004. The raft-associated protein MAL is required for maintenance of proper axon-glia interactions in the central nervous system. *Journal of Cell Biology,* 166**,** 731-742.

SCHAERENWIEMERS, N., VALENZUELA, D. M., FRANK, M. & SCHWAB, M. E. 1995. CHARACTERIZATION OF A RAT GENE, RMAL, ENCODING A PROTEIN WITH 4 HYDROPHOBIC DOMAINS IN CENTRAL AND PERIPHERAL MYELIN. *Journal of Neuroscience,* 15**,** 5753-5764.

SCHENKWEIN, D. & YLA-HERTTUALA, S. 2018. Gene Editing of Human Embryos with CRISPR/Cas9: Great Promise Coupled with Important Caveats. *Molecular Therapy,* 26**,** 659-660.

SCHRODER, M. & KAUFMAN, R. J. 2005. The mammalian unfolded protein response. *Annual Review of Biochemistry,* 74**,** 739-789.

SCHRÖDER M, K. R. 2005. The mammalian unfolded protein response. *Annual Review of Biochemistry* 74**,** 739-89.

SCHWARZ, Y., ZHAO, N., KIRCHHOFF, F. & BRUNS, D. 2017. Astrocytes control synaptic strength by two distinct v-SNARE-dependent release pathways. *Nature Neuroscience,* 20**,** 1529-+.

SCOTT, I. Y., R.J. 2016. Mitochondrial fission and fusion *Essays Biochemistry*

SELL, G. L. & MARGOLIS, S. S. 2015. From UBE3A to Angelman syndrome: a substrate perspective. *Frontiers in Neuroscience,* 9**,** 6.

SHAMSELDIN, H. E., ALASMARI, A., SALIH, M. A., SAMMAN, M. M., MIAN, S. A., ALSHIDI, T., IBRAHIM, N., HASHEM, M., FAQEIH, E., AL-MOHANNA, F. &

ALKURAYA, F. S. 2017. A null mutation in MICU2 causes abnormal mitochondrial calcium homeostasis and a severe neurodevelopmental disorder. *Brain,* 140**,** 2806-2813.

SHARMA, D. K., BRESSLER, K., PATEL, H., BALASINGAM, N. & THAKOR, N. 2016. Role of Eukaryotic Initiation Factors during Cellular Stress and Cancer Progression. *Journal of Nucleic Acids*.

SHERIDAN, C. 2018. Go-ahead for first in-body CRISPR medicine testing
. *Nature Biotechnology*.

SHERIDAN, E., WRIGHT, J., SMALL, N., CORRY, P. C., ODDIE, S., WHIBLEY, C., PETHERICK, E. S., MALIK, T., PAWSON, N., MCKINNEY, P. A. & PARSLOW, R. C. 2013. Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. *Lancet,* 382**,** 1350-1359.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res,* 29**,** 308-11.

SHINKAI, Y., YAMAMOTO, C. & KAJI, T. 2010. Lead Induces the Expression of Endoplasmic Reticulum Chaperones GRP78 and GRP94 in Vascular Endothelial Cells via the JNK-AP-1 Pathway. *Toxicological Sciences,* 114**,** 378-386.

SNAIDERO, N., MOBUS, W., CZOPKA, T., HEKKING, L. H. P., MATHISEN, C., VERKLEIJ, D., GOEBBELS, S., EDGAR, J., MERKLER, D., LYONS, D. A., NAVE, K. A. & SIMONS, M. 2014. Myelin Membrane Wrapping of CNS Axons by PI(3,4,5) P3-Dependent Polarized Growth at the Inner Tongue. *Cell,* 156**,** 277-290.

SNAIDERO, N. & SIMONS, M. 2014. Myelination at a glance. *Journal of Cell Science,* 127**,** 2999-3004.

SNEL, B., LEHMANN, G., BORK, P. & HUYNEN, M. A. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research,* 28**,** 3442-3444.

SOUDA, P., RYAN, C. M., CRAMER, W. A. & WHITELEGGE, J. 2011. Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry. *Methods,* 55**,** 330-336.

STEWART, M. D., RITTERHOFF, T., KLEVIT, R. E. & BRZOVIC, P. S. 2016. E2 enzymes: more than just middle men. *Cell Research,* 26**,** 423-440.

STUPKA, N., KINTAKAS, C., WHITE, J. D., FRASER, F. W., HANCIU, M., ARAMAKI-HATTORI, N., MARTIN, S., COLES, C., COLLIER, F., WARD, A. C., APTE, S. S. & MCCULLOCH, D. R. 2013. Versican Processing by a Disintegrin-like and Metalloproteinase Domain with Thrombospondin-1 Repeats Proteinases-5 and-15 Facilitates Myoblast Fusion. *Journal of Biological Chemistry,* 288**,** 1907-1917.

SYLVESTER, P. W. 2011. Optimization of the Tetrazolium Dye (MTT) Colorimetric Assay for Cellular Growth and Viability. *In:* SATYANARAYANAJOIS, S. D. (ed.) *Drug Design and Discovery: Methods and Protocols.*

SZCZESNY, R. J., OBRIOT, H., PACZKOWSKA, A., JEDRZEJCZAK, R., DMOCHOWSKA, A., BARTNIK, E., FORMSTECHER, P., POLAKOWSKA, R. & STEPIEN, P. P. 2007. Down-regulation of human RNA/DNA helicase SUV3 induces apoptosis by a caspase- and AIF-dependent pathway. *Biology of the Cell,* 99**,** 323-332.

SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H., BORK, P., JENSEN, L. J. & MERING, C. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research,* 47**,** D607-D613.

TABAS, I. & RON, D. 2011. Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nature Cell Biology,* 13**,** 184-190.

TANG, L. C., ZENG, Y. T., DU, H. Z., GONG, M. M., PENG, J., ZHANG, B. X., LEI, M., ZHAO, F., WANG, W. H., LI, X. W. & LIU, J. Q. 2017. CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein. *Molecular Genetics and Genomics,* 292**,** 525-533.

TAZIR, M., BELLATACHE, M., NOUIOUA, S. & VALLAT, J. M. 2013. Autosomal recessive Charcot-Marie-Tooth disease: from genes to phenotypes. *Journal of the Peripheral Nervous System,* 18**,** 113-129.

THOMAS, P., WOHLFORD, D. & AOH, Q. L. 2016. SCAMP 3 is a novel regulator of endosomal morphology and composition. *Biochemical and Biophysical Research Communications,* 478**,** 1028-1034.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 14**,** 178-192.

THUERAUF, D. J., HOOVER, H., MELLER, J., HERNANDEZ, J., SU, L., ANDREWS, C., DILLMANN, W. H., MCDONOUGH, P. M. & GLEMBOTSKI, C. C. 2001. Sarco/endoplasmic reticulum calcium ATPase-2 expression is regulated by ATF6 during the endoplasmic reticulum stress response - Intracellular signaling of calcium stress in a cardiac myocyte model system. *Journal of Biological Chemistry,* 276**,** 48309-48317.

THUERAUF, D. J., MORRISON, L. & GLEMBOTSKI, C. C. 2004. Opposing roles for ATF6 alpha and ATF6 beta in endoplasmic reticulum stress response gene induction. *Journal of Biological Chemistry,* 279**,** 21078-21084.

THURSFIELD, R. M. D., J.C. 2013. Genotype-specific small-molecule therapy for cystic fibrosis. *Breathe*.

TRAASETH, N., ELFERING, S., SOLIEN, J., HAYNES, V. & GIULIVI, C. 2004. Role of calcium signaling in the activation of mitochondrial nitric oxide synthase and citric acid cycle. *Biochimica Et Biophysica Acta-Bioenergetics,* 1658**,** 64-71.

TYANOVA, S., TEMU, T., SINITCYN, P., CARLSON, A., HEIN, M. Y., GEIGER, T., MANN, M. & COX, J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods,* 13**,** 731-740.

TYLER, W. A., GANGOLI, N., GOKINA, P., KIM, H. A., COVEY, M., LEVISON, S. W. & WOOD, T. L. 2009. Activation of the Mammalian Target of Rapamycin (mTOR) Is Essential for Oligodendrocyte Differentiation. *Journal of Neuroscience,* 29**,** 6367-6378.

UDENWOBELE, D. I., SU, R. C., GOOD, S. V., BALL, T. B., SHRIVASTAV, S. V. & SHRIVASTAV, A. 2017. Myristoylation: An Important Protein Modification in the Immune Response. *Frontiers in Immunology,* 8.

VAN DER KNAAP, M. S. & BUGIANI, M. 2017. Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms. *Acta Neuropathologica,* 134**,** 351-382.

VAN HUIZEN, R., MARTINDALE, J. L., GOROSPE, M. & HOLBROOK, N. J. 2003. P58(IPK), a novel endoplasmic reticulum stress-inducible protein and potential negative regulator of eIF2 alpha signaling. *Journal of Biological Chemistry,* 278**,** 15558-15564.

VILLUMSEN, B. H., DANIELSEN, J. R., POVLSEN, L., SYLVESTERSEN, K. B., MERDES, A., BELI, P., YANG, Y. G., CHOUDHARY, C., NIELSEN, M. L., MAILAND, N. & BEKKER-JENSEN, S. 2013. A new cellular stress response that triggers centriolar satellite reorganization and ciliogenesis. *Embo Journal,* 32**,** 3029-3040.

VOLPI, V. G., TOUVIER, T. & D'ANTONIO, M. 2017. Endoplasmic Reticulum Protein Quality Control Failure in Myelin Disorders. *Frontiers in Molecular Neuroscience,* 9.

WALTER, P. & RON, D. 2011. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science,* 334**,** 1081-1086.

WANG, D. D. H., SHU, Z., LIESER, S. A., CHEN, P. L. & LEE, W. H. 2009. Human Mitochondrial SUV3 and Polynucleotide Phosphorylase Form a 330-kDa Heteropentamer to Cooperatively Degrade Double-stranded RNA with a 3 '-to-5 ' Directionality. *Journal of Biological Chemistry,* 284**,** 20812-20821.

WANG, J., CHANG, Y. F., HAMILTON, J. I. & WILKINSON, M. F. 2002. Nonsense-associated altered splicing: A frame-dependent response distinct from nonsense-mediated decay. *Molecular Cell,* 10**,** 951-957.

WANG, X. N. & SCHWARZ, T. L. 2009. The Mechanism of Ca2+-Dependent Regulation of Kinesin-Mediated Mitochondrial Motility. *Cell,* 136**,** 163-174.

WATANABE, S., HARAYAMA, M., KANEMURA, S., SITIA, R. & INABA, K. 2017. Structural basis of pH-dependent client binding by ERp44, a key regulator of protein secretion at the ER-Golgi interface. *Proceedings of the National Academy of Sciences of the United States of America,* 114**,** E3224-E3232.

WILBIE, D., WALTHER, J. & MASTROBATTISTA, E. 2019. Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing. *Accounts of Chemical Research,* 52**,** 1555-1564.

WILLEMS, P., ROSSIGNOL, R., DIETEREN, C. E. J., MURPHY, M. P. & KOOPMAN, W. J. H. 2015. Redox Homeostasis and Mitochondrial Dynamics. *Cell Metabolism,* 22**,** 207-218.

WILLIAMS, C. A., DRISCOLL, D. J. & DAGLI, A. I. 2010. Clinical and genetic aspects of Angelman syndrome. *Genetics in Medicine,* 12**,** 385-395.

WOELK, T., SIGISMUND, S., PENENGO, L. & POLO, S. 2007. The ubiquitination code: a signalling problem. *Cell Division,* 2.

WOLF, N. I., HARTING, I., INNES, A. M., PATZER, S., ZEITLER, P., SCHNEIDER, A., WOLFF, A., BAIER, K., ZSCHOCKE, J., EBINGER, F., BOLTSHAUSER, E. & RATING, D. 2007. Ataxia, delayed dentition and hypornyelination: A novel Leukoencephalopathy. *Neuropediatrics,* 38**,** 64-70.

WOODS, C. G., COX, J., SPRINGELL, K., HAMPSHIRE, D. J., MOHAMED, M. D., MCKIBBIN, M., STERN, R., RAYMOND, F. L., SANDFORD, R., SHARIF, S. M., KARBANI, G., AHMED, M., BOND, J., CLAYTON, D. & INGLEHEARN, C. F. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *American Journal of Human Genetics,* 78**,** 889-896.

WOODS, C. G., VALENTE, E. M., BOND, J. & ROBERTS, E. 2004. A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR. *Journal of Medical Genetics,* 41.

WU, Q. W., SUN, X. Q., YUE, W. H., LU, T. L., RUAN, Y. Y., CHEN, T. D. & ZHANG, D. 2016. RAB18, a protein associated with Warburg Micro syndrome, controls neuronal migration in the developing cerebral cortex. *Molecular Brain,* 9.

XIANG, C. C., WANG, Y. J., ZHANG, H. & HAN, F. 2017. The role of endoplasmic reticulum stress in neurodegenerative disease. *Apoptosis,* 22**,** 1-26.

XU, H. X., JIAO, Y., QIN, S., ZHAO, W. H., CHU, Q. & WU, K. M. 2018. Organoid technology in disease modelling, drug development, personalized treatment and regeneration medicine. *Experimental Hematology & Oncology,* 7.

YAM, G. H. T., GAPLOVSKA-KYSELA, K., ZUBER, C. & ROTH, J. 2007. Sodium 4-phenylbutyrate acts as a chemical chaperone on misfolded myocilin to rescue cells from endoplasmic reticulum stress and apoptosis. *Investigative Ophthalmology & Visual Science,* 48**,** 1683-1690.

YOSHIDA, H., MATSUI, T., YAMAMOTO, A., OKADA, T. & MORI, K. 2001. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell,* 107**,** 881-891.

YUEN, R. K. C., MERICO, D., BOOKMAN, M., HOWE, J. L., THIRUVAHINDRAPURAM, B., PATEL, R. V., WHITNEY, J., DEFLAUX, N., BINGHAM, J., WANG, Z. Z., PELLECCHIA, G., BUCHANAN, J. A., WALKER, S., MARSHALL, C. R., UDDIN, M., ZARREI, M., DENEAULT, E., D'ABATE, L., CHAN, A. J. S., KOYANAGI, S., PATON, T., PEREIRA, S. L., HOANG, N., ENGCHUAN, W., HIGGINBOTHAM, E. J., HO, K., LAMOUREUX, S., LI, W. L., MACDONALD, J. R., NALPATHAMKALAM, T., SUNG, W. W. L., TSOI, F. J., WEI, J., XU, L. Z., TASSE, A. M., KIRBY, E., VAN ETTEN, W., TWIGGER, S., ROBERTS, W., DRMIC, I., JILDERDA, S., MODI, B. M., KELLAM, B., SZEGO, M., CYTRYNBAUM, C., WEKSBERG, R., ZWAIGENBAUM, L., WOODBURY-SMITH, M., BRIAN, J., SENMAN, L., IABONI, A., DOYLE-THOMAS, K., THOMPSON, A., CHRYSLER, C., LEEF, J., SAVION-LEMIEUX, T., SMITH, I. M., LIU, X. D., NICOLSON, R., SEIFER, V., FEDELE, A., COOK, E. H., DAGER, S., ESTES, A., GALLAGHER, L., MALOW, B. A., PARR, J. R., SPENCE, S. J., VORSTMAN, J., FREY, B. J., ROBINSON, J. T., STRUG, L. J., FERNANDEZ, B. A., ELSABBAGH, M., CARTER, M. T., HALLMAYER, J., KNOPPERS, B. M., ANAGNOSTOU, E., SZATMARI, P., RING, R. H., GLAZER, D., PLETCHER, M. T. & SCHERER, S. W. 2017. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience,* 20**,** 602-+.

ZHANG, Y., CHEN, K. N., SLOAN, S. A., BENNETT, M. L., SCHOLZE, A. R., O'KEEFFE, S., PHATNANI, H. P., GUARNIERI, P., CANEDA, C., RUDERISCH, N., DENG, S. Y., LIDDELOW, S. A., ZHANG, C. L., DANEMAN, R., MANIATIS, T., BARRES, B. A. & WU, J. Q. 2015. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex (vol 35, pg 11929, 2014). *Journal of Neuroscience,* 35**,** 864-866.

ZHAO, G. H., SONG, J., YANG, M., SONG, X. H. & LIU, X. M. 2018. A novel mutation of LRSAM1 in a Chinese family with Charcot-Marie-Tooth disease. *Journal of the Peripheral Nervous System,* 23**,** 55-59.

ZHAO, Q., LIU, L. & XIE, Q. 2012. In vitro protein ubiquitination assay. *Methods in molecular biology (Clifton, N.J.),* 876**,** 163-72.

ZHENG, N. & SHABEK, N. 2017. Ubiquitin Ligases: Structure, Function, and Regulation. *In:* KORNBERG, R. D. (ed.) *Annual Review of Biochemistry, Vol 86.*

ZHENG, W., CHEN, H., DENG, X., YUAN, L. M., YANG, Y., SONG, Z., YANG, Z. J., WU, Y. & DENG, H. 2016. Identification of a Novel Mutation in the Titin Gene in a Chinese Family with Limb-Girdle Muscular Dystrophy 2J. *Molecular Neurobiology,* 53**,** 5097-5102.

ZODE, G. S., KUEHN, M. H., NISHIMURA, D. Y., SEARBY, C. C., MOHAN, K., GROZDANIC, S. D., BUGGE, K., ANDERSON, M. G., CLARK, A. F., STONE, E. M. & SHEFFIELD, V. C. 2015. Reduction of ER stress via a chemical chaperone prevents disease phenotypes in a mouse model of primary open angle glaucoma (vol 121, pg 3542, 2011). *Journal of Clinical Investigation,* 125**,** 3303-3303.

List of Abbreviations

2-DG	2-Deoxy-D-glucose
4-PBA	4-Phenylbutyrate
AON	Antisense Oligonucleiotide
ATP	Adenosine Triphosphate
BR	Broad Range
BSA	Bovine Serum Albumin
BWA	Burrows-Wheeler Aligner
Ca2+	Calcium ions
CADD	Combined Annotation Dependent Depletion
CGH	Comparative Genomic Hybridisation
CNS	Central Nervous System
CNV	Copy Number Variation
Condel	Consensus Deleteriousness
CSF	Cerebrospinal fluid
CTCF	Corrected Total Cell Fluoerescence
DAPI	4',6-diamidino-2-phenylindole
DDD	Deciphering Developmental Delay
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
ECAR	Extracellular Acidification Rate
ER	Endoplasmic Reticulum
ERT	Enzyme Replacement Therapy
FCCP	Carbonyl cyanide-4-phenylhydrazone
GATK	Genome Analysis Toolkit
GFP	Green Fluorescent Protein
GO	Gene Ontology
HDF	Human Dermal Fibroblasts
HDL	Hypomyelinating Leukodystrophies
HRP	Horseradish Peroxidase
IBD	Identical by Decent
IGV	Integrative Genomics Viewer
IP	Immunoprecipitation
JT	Jules Thorn ID number

LB	Luria Bertani
LD	Leukodystrophy
MAF	Minor Allele Frequency
MDCK	Madin Darby Canine Kidney
MIM	Mendelian Inhertiance in Man
MNP	Multi-Nucleotide Polymorphism
MRI	Magnetic resonance imaging
mRNA	messanger Ribonucleic acid
NAS	Nonsense-mediated Altered Splicing
ND	Neurodevelopmental family ID
NDS	Normal Donkey Serum
NGS	Next Generation Sequencing
NMD	Nonsense Mediated Decay
NRES	National Research Ethics Service
OCR	Oxygen Consumption Rate
OMIM	Online Mendelian Inheritance in Man
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
PBST	Phosphate Buffered Saline Tween-20
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PMD	Pelizaeus Merzbacher Disease
PNS	Peripheral Nervous System
PolyPhen-2	Polymorphism Phenotyping v2
PPI	Protein-Protein Interaction
PTC	Premature Termination Codon
PVDF	Polyvinylidene Difluoride
QC	Quality Control
RFLP	Restriction Fragment Length Polymorpshism
RIPA	Radio Immunoprecipitation Assay
RNA	Ribonucleic Acid
ROI	Region of Interest
RP	Retinitis Pigmentosa
RPE	Retinal Pigmented Epithelium
SDM	Side Directed Mutagenesis
SDS	Sodium Dodecyl Sulfate
SIFT	Sorting Introlerant From Tolerant
SNP	Single Nucleotide Polymorphism

TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TBS	Tris-Buffered Saline
TE	Tris-EDTA
TMHMM	Transmembrane Helix Prediction
UBD	Ubiquitin Binding Domains
UPR	Unfolded Protein Response
UPS	Ubiquitin Proteasome System
UV	Ultraviolet
VEP	Variant Effect Predictor
WCE	Whole Cell Extract
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
YFP	Yellow Fluorescent Protein

Appendix A

A.1 Inclusion Criteria

The inclusion criteria for family recruitment to this study is based on the phenotypic presence of neurodevelopmental defects. This may include global developmental defects, neurological defects, neuromuscular defects and brain abnormalities.

Most of the cases recruited were paediatric cases with autosomal recessive patterns of inheritance, and in a few instances some de novo cases too. In addition, the majority of the cases analysed in this study were consanguineous families, with the exception of a few non-consanguineous families.

All of the families recruited must meet the following criteria:

1) Presence of a rare paediatric neurodevelopmental phenotype.
2) No previous clinical diagnosis from NHS genetic testing.
3) Normal karyotype and array CGH in affected individuals.
4) Informed consent provided by a parent.
5) Patient under the care of Clinical Geneticists.

Appendix B

B.1 Participants information sheet

PARTICIPANT'S INFORMATION SHEET for parents

INVITATION TO TAKE PART IN RESEARCH TO STUDY GENES THAT CAUSE RECESSIVE INHERITED CONDITIONS:

"Molecular Genetic Investigations of Autosomal Recessive Conditions"

Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.

- **Part 1 tells you the purpose of this study and what will happen to you if you take part.**
- **Part 2 gives you more detailed information about the conduct of the study.**

Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

PART 1: to give you first thoughts about the project

WHAT DO WE WANT TO DO?

We are studying the inherited factors (genes) involved in recessive conditions. Recessive conditions are genetic conditions that a couple can pass on to their children, but do not affect the couple themselves. Some children might problems with their walking or learning, and might need operations because they have structural birth defects, such as like heart defects, kidney defects, or abnormalities picked up on a brain scan. In some cases the gene causing the condition isn't known and the aim of this study is to do research to find these faulty genes. This could lead to a better understanding of the condition.

WHY ARE WE ASKING FOR YOUR HELP?

You have been asked to take part in this research study because your child has a recessive condition but the genetic cause still isn't known. As most recessive conditions are very rare, we are seeking your help so that we can look at the genes of many families with exactly the same condition.

DO I HAVE TO TAKE PART?

No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. If at a later date you decide you do not want to take part in the study you can change your mind for any reason. If you do not wish to take part in the study this will not affect your or your family's medical care in any way.

WHAT WILL WE BE ASKING YOU TO DO?

A researcher will discuss the study with you and answer any questions you may have. When you are satisfied you have all the information you require and if you decide to take part, we will ask you to read and complete a consent form. The researcher will ask you to sign this form to give your consent. The form will ask your permission for an extra blood test on your child, yourself and your partner. We will try to only take blood from your child at a time when blood is being taken as part of the usual clinical care. If we already have a sample of your child's blood we may ask you if we can use that instead of taking a new sample. We may also ask for a saliva sample in addition or instead of a blood sample, and we may ask your permission to take samples from your other children if you have them, including those that are healthy. We will ask some questions about your medical and family history, and we may also ask your permission to look at your medical records. You will be given a copy of this information sheet and your signed consent form to keep.

WHAT INFORMATION WILL WE HOLD?

We will hold information that is routinely collected as part of your child's usual clinical care. We will also hold information that we gain from the blood samples that we have taken. In the future, we may need to contact you again for further medical information.

ARE THERE ANY POSSIBLE DISADVANTAGES TO TAKING PART IN THE STUDY?

We do not think that taking part in this study will harm you in any way. We will be gathering information which is already routinely collected by health professionals. If at any point you would like to discuss the study and your involvement in it, you will be able to speak to a researcher. Contact details are provided at the end of this information sheet.

WILL THIS PROJECT BENEFIT MY FAMILY?
There are no immediate short term advantages to taking part in this study. We cannot guarantee to discover anything that will directly benefit you or your family. However if we do find the faulty gene causing the condition in your family then tests may be available for your relatives to find out if they are at risk of having children with the condition. We also hope that studying the genetic causes of recessive conditions will help to discover the reasons why these conditions occur. By increasing understanding in this way we may, in the long term, be able to offer better help to children and their families

WHAT IF THERE IS A PROBLEM?
We will address any complaint about the way you have been dealt with during the study or any possible harm that you might suffer. The detailed information on this is given in Part 2.

THIS COMPLETES PART 1 OF THE INFORMATION SHEET

If the information in Part 1 has interested you and you are considering participation, please continue to read the additional information in Part 2 before making any decision.

PART 2: information you need to know if you still want to take part

WHAT IF NEW INFORMATION BECOMES AVAILABLE?
We hope that this study will uncover the genetic cause for your child's problems. If we do find the faulty gene, this information will be fed back to you in the course of the clinics run by the doctors caring for your child. We will also inform your local clinical genetics department so they can contact you. It may be possible to then offer tests to the wider family, if they wish to see if they carry the same gene or not. It may also be possible to offer tests to either you or your partner during a future pregnancy.

HOW DO WE ENSURE CONFIDENTIALITY?
All information recorded on paper and any biological samples stored will be kept under conditions of strict confidentiality as a legal requirement under the Data Protection and Human Tissue Acts. We will put information about your family onto a secure computer database. However, we will remove all personal details, such as names and addresses, so your family cannot be recognized from it. A study number, which can only be linked to you by the research team, will be the identifier of the information. This is necessary so that we can put together your information throughout the study. The results from any information or biological samples that we collect will only be used for research purposes and will not be available to anyone else.

WILL MY GENERAL PRACTITIONER/FAMILY DOCTOR KNOW THAT I AM PART OF THIS STUDY?
With your permission, we will contact your G.P. or other doctor involved in your family's care to let them know that your child is included in the study.

WILL ANYONE ELSE KNOW ABOUT MY TAKING PART?
The information collected about you during the course of the research project will be kept strictly confidential and you will not be identifiable from it. If any research results are published in medical articles as a result of this project, all personal details will be removed so that your family cannot be recognized from it.

WHAT WILL HAPPEN TO ANY SAMPLES THAT I GIVE?
We will request an extra sample bottle be filled at the time your child is having blood tests as part of their usual clinical care. We will also request a sample of blood from you and your partner. We may also ask to take samples from your other children if you have them, including those that are healthy. We could also ask you or your partner to give a saliva sample instead or in addition to a blood sample. The samples will be sent to our lab and DNA will be extracted. We will analyse the DNA to see if we can determine if there is a genetic cause for your child's problems. If you change your mind later and decide not to take part in the study then you should let us know what we should do with any DNA samples. We need to ask you because it can often take some years to find the faulty gene that causes the specific condition in your family. We can either keep the DNA for research in the future, or keep the DNA but remove any details that can identify it as yours, or destroy the sample.

WILL ANY GENETIC TESTS BE DONE?
Tests will be done to try to establish if your child's problems are caused by a faulty gene. We will not use the sample for any other genetic tests. We may use new technology (called "clonal" or next generation sequencing) to look at many genes at once rather than one after another. We want to use this new technology because we might find the faulty gene more quickly, but we will only use it for studies appropriate for your child's condition. It is possible that we may uncover other findings that are unconnected to your child's condition but might be important for your health.

With your permission, we will inform your local clinical genetics department so they can contact you about this information.

WHO IS ORGANISING AND FUNDING THE RESEARCH?
The study is organized by the University of Leeds. The study is funded by a medical research charity called the Sir Jules Thorn Charitable Trust. Neither you nor the researchers involved will benefit financially from this project.

WHO HAS REVIEWED THE STUDY?
All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given a favourable opinion by South Yorkshire Research Ethics Committee. The research has also been reviewed by the Sir Jules Thorn Charitable Trust, to ensure that the proper science is being done for this study.

WHAT IF I HAVE ANY CONCERNS?
If you have any concerns, other questions about this study or the way it has been carried out, you should contact the investigators in charge of running the study:

Prof. Colin A. Johnson

Section of Ophthalmology and Neurosciences
Wellcome Trust Brenner Building
Leeds Institute of Molecular Medicine
St James's University Hospital
Beckett Street
Leeds, LS9 7TF, U.K.

tel: (+44) 0113 343 8443

e-mail: c.johnson@leeds.ac.uk

Dr. Eamonn Sheridan:

Section of Genetics
Wellcome Trust Brenner Building
Leeds Institute of Molecular Medicine
St James's University Hospital
Beckett Street
Leeds, LS9 7TF, U.K.

tel: (+44) 0113 206 5927

e-mail: e.sheridan@leeds.ac.uk

COMPLAINTS
If you have a concern about any aspect of this study, you should ask to speak with the researchers who will do their best to answer your questions. Please contact Prof. Johnson on 0113 343 8443 or Dr. Sheridan on 0113 206 5927 in the first instance. If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital.

HARM
In the event that something does go wrong and you are harmed during the research study there are no special compensation arrangements. If you are harmed and this is due to someone's negligence then you may have grounds for a legal action for compensation against Leeds Teaching Hospitals NHS Trust but you may have to pay your legal costs. The normal National Health Service complaints procedure will still be available to you.

B.2 Consent form

UNIVERSITY OF LEEDS

The Leeds Teaching Hospitals NHS
NHS Trust

Title of Project: MOLECULAR GENETIC INVESTIGATIONS OF AUTOSOMAL RECESSIVE CONDITIONS

CONSENT FORM: for patients, or parents/guardians and children in the family
Participant to circle responses to all parts:

1. I confirm that I have read and understand the information sheet for the study. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily — **YES/NO**

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected — **YES/NO**

3. I understand that sections of any of my/my child's medical notes may be reviewed by the researchers and other responsible individuals from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in research. I give permission for these individuals to have access to my/our records

4. **i)** I agree to allow medical information about me/my family to be entered on a confidential computer database. I understand that personal details (including names and addresses) about myself and my family will be accessed **only** by researcher team leaders (Johnson, Sheridan & Bonthron). Information accessed by other researchers will have personal details removed, thereby maintaining our privacy — **YES/NO** **YES/NO**

ii) If further medical information is requested by researchers, I agree to be contacted again for this purpose. — **YES/NO**

5. **i)** I agree that a blood sample/saliva sample/stored DNA sample from me may be used for genetic research studies appropriate to my family's condition. I understand that any results arising from this research work will be kept **strictly confidential**. — **YES/NO**

ii) **OR** I agree that blood samples/saliva samples/stored DNA samples from the following **children** in the family can be used for the project: — **YES/NO**

list of **children** agreed to be part of the research:

1.	5.
2.	6.
3.	7.
4.	8.

6. **i)** If a genetic test, or other genetic finding that might have health implications, becomes available as a result of medical research on my own or my family's sample(s) I would like to have the opportunity to discuss the implications of these findings with appropriate medical experts, including my local clinical genetics department — **YES/NO**

ii) I understand that genetic testing may sometimes reveal information that might have health implications **unconnected** to my family's condition. I would like to have the opportunity to discuss the implications of any such findings with appropriate medical experts, including my local clinical genetics department — **YES/NO**

iii) I agree that my G.P, or other healthcare professional in charge of my or my family's care, is informed that I/we are taking part in this study and of the result of any genetic test — **YES/NO**

_____	_____	_____
Participant/Family Representative *on behalf of children in the family*	Date	Signature
_____	_____	_____
Name of Person taking consent	Date	Signature
_____	_____	_____
Researcher	Date	Signature

when completed <u>copies</u>: one for participant; one for researcher; signed original to be kept with hospital notes

Appendix C

C.1 Ethical approval

NHS
National Research Ethics Service

South Yorkshire Research Ethics Committee
Millside
Mill Pond Lane
Meanwood
Leeds
LS6 4RA

Telephone: 0113 305 0128

18 February 2011

Prof Colin A Johnson
Professor of Medical & Molecular Genetics
Section of Ophthalmology and Neuroscience
Leeds Inst of Molecular Medicine
St James University Hospital
Leeds LS9 7TF

Dear Prof Johnson

Study Title: **Molecular genetic investigations of autosomal recessive conditions**
REC reference number: **11/H1310/1**

Thank you for your letter of 8th February 2011, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where the only involvement of the NHS organisation is as a Participant Identification Centre (PIC), management permission for research is not required but the R&D office should be notified of the study and agree to the organisation's involvement. Guidance on procedures for PICs is available in IRAS. Further advice should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Investigator CV	1	14 December 2010
Covering Letter		08 February 2011
Letter from Sponsor		
REC application		14 December 2010
Response to Request for Further Information		
Participant Information Sheet: for children, aged 8 years or younger (Mirpuri Urdu translation)	1	14 December 2010
Evidence of insurance or indemnity		
Participant Information Sheet: Children aged 12 - 15 years	2	08 February 2011
Letter of invitation to participant	1	14 December 2010
Participant Information Sheet: for parents (Mirpuri Urdu translation)	1	14 December 2010
Participant Information Sheet: Relatives that are adults or young persons	2	08 February 2011
Participant Information Sheet: for children aged 8 year or younger	1	14 December 2010
Participant Information Sheet: Parents	2	08 February 2011
Participant Information Sheet: for children aged 8-12 years	1	14 December 2010
Participant Consent Form: Assent form for older children	1	14 December 2010
Protocol	1	14 December 2010

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

11/H1310/1	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Ms Jo Abbott
Chair

Email: Sinead.audsley@leedspft.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

Copy to: *Mrs Rachel E de Souza, University of Leeds*
Mrs Anne Gowing, Leeds Teaching Hospitals NHS Trust

C.2 NHS permission for research

The Leeds Teaching Hospitals NHS
NHS Trust

Ref: Amanda Burd

29/04/2011

Research & Development
Leeds Teaching Hospitals NHS Trust
34 Hyde Terrace
Leeds
LS2 9LN

Tel: 0113 392 2878
Fax: 0113 392 6397

r&d@leedsth.nhs.uk
www.leedsth.nhs.uk

Dr Colin A. Johnson

Section of Ophthalmology and Neurosciences,
Wellcome Trust Brenner Bui
Leeds Institute of Molecular Medicine,
St. James's University Hospital
LS9 7TF

Dear Dr Johnson

Re: NHS Permission at LTHT for: Molecular genetic investigations of autosomal recessive conditions
LTHT R&D Number: CG11/9764 (53788/WY)
REC: 11/H1310/1

I confirm that *NHS Permission for research* has been granted for this project at The Leeds Teaching Hospitals NHS Trust (LTHT). NHS Permission is granted based on the information provided in the documents listed below. All amendments (including changes to the research team) must be submitted in accordance with guidance in IRAS. Any change to the status of the project must be notified to the R&D Department.

Permission is granted on the understanding that the study is conducted in accordance with the *Research Governance Framework for Health and Social Care*, ICH GCP (if applicable) and NHS Trust policies and procedures available at http://www.leedsth.nhs.uk/sites/research_and_development/.

This permission is granted only on the understanding that you comply with the requirements of the *Framework* as listed in the attached sheet "Conditions of Approval".

If you have any queries about this approval please do not hesitate to contact the R&D Department on telephone 0113 392 2878.

Indemnity Arrangements

The Leeds Teaching Hospitals NHS Trust participates in the NHS risk pooling scheme administered by the NHS Litigation Authority 'Clinical Negligence Scheme for NHS Trusts' for: (i) medical professional and/or medical malpractice liability; and (ii) general liability. NHS Indemnity for negligent harm is extended to researchers with an employment contract (substantive or honorary) with the Trust. The Trust only accepts liability for research activity that has been managerially approved by the R&D Department.

The Trust therefore accepts liability for the above research project and extends indemnity for negligent harm to cover you as investigator and the researchers listed on the Site Specific Information form. Should there be any changes to the research team please ensure that you inform the R&D Department and that s/he obtains an appropriate contract, or letter of access, with the Trust if required.

Yours sincerely

Dr D R Norfolk
Associate Director of R&D

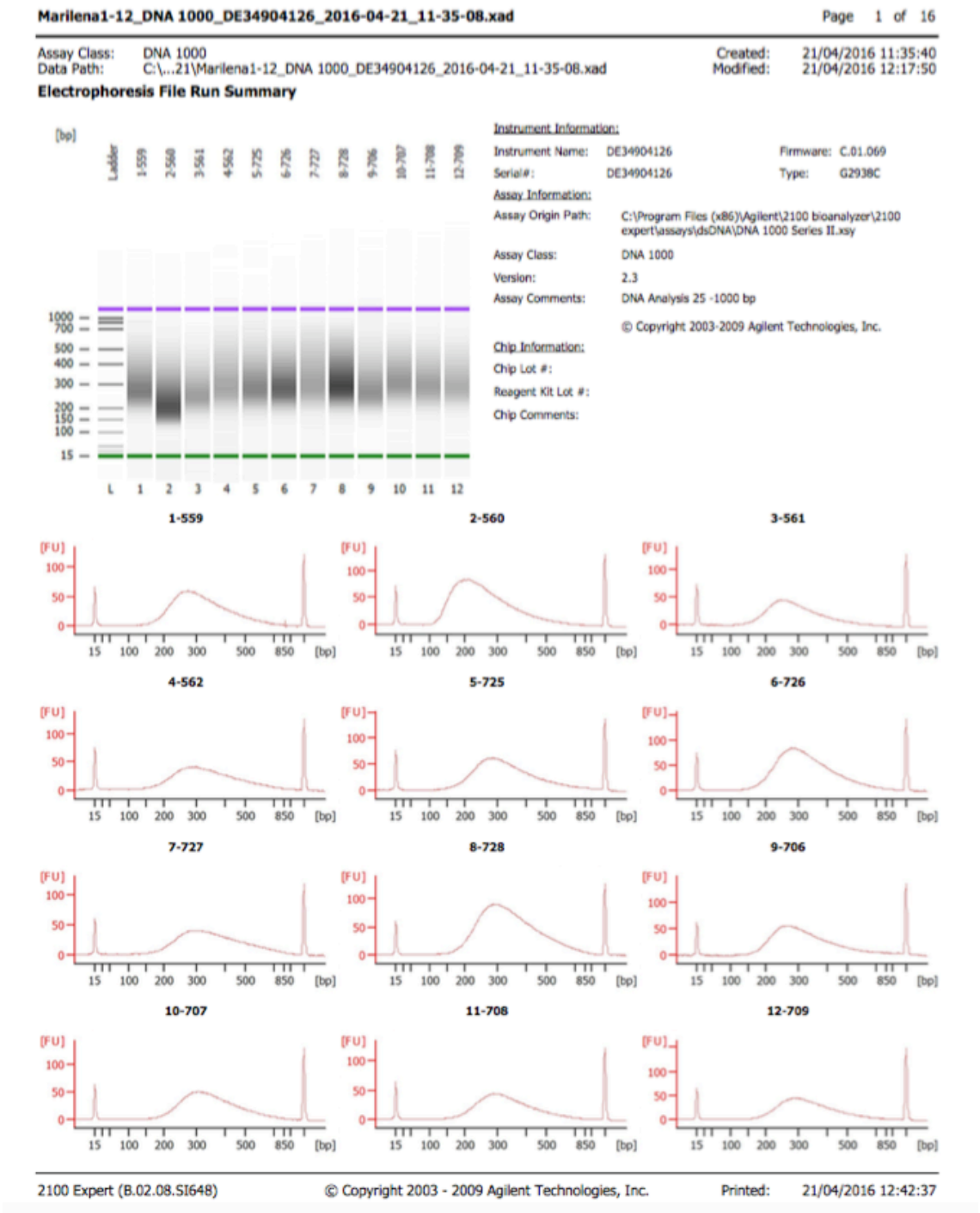
Approved documents
The documents reviewed and approved are listed as follows

Document	*Version*	*Date of document*
NHS R&D Form	3.1	02/03/2011
SSI Form	3.1	08/03/2011
Directorate Approval		08/03/2011
Radiology Approval		N/A
Pharmacy Approval		N/A
Protocol	1.0	14/12/2010
REC Letter confirming favourable opinion		18/02/2011
Evidence of Insurance		Not dated
Patient information sheet (REC Approved) Parent	2.0	08/02/2011
Patient information sheet (REC Approved) 8/younger URDU	1.0	14/12/2010
Patient information sheet (REC Approved) Parent URDU	1.0	14/12/2010
Patient information sheet (REC Approved) 12-15 yrs	2.0	08/02/2011
PIS (REC Approved) Adult relatives/young Person	2.0	08/02/2011
PIS (REC Approved) Children aged 8 or younger	1.0	14/12/2010
PIS (REC Approved) Children aged 8 – 12 years	1.0	14/12/2010
Informed Consent (REC Approved) Assent older children	1.0	14/12/2010
Letter of Invitation to participants (REC Approved)	1.0	14/12/2010

Appendix D

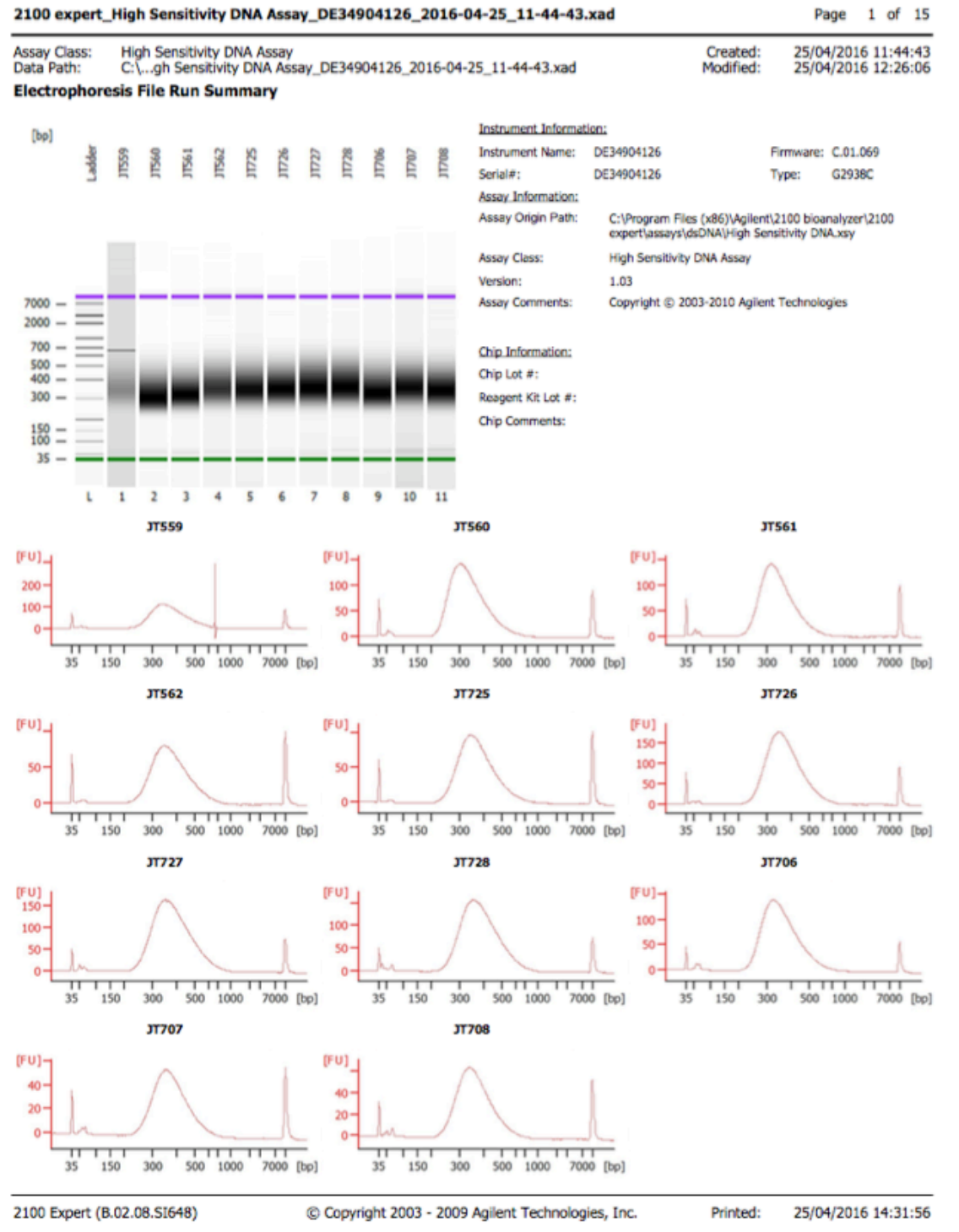
D.1 Agilent 2100 Bioanalyzer DNA 1000 Assay

The following traces outline a typical Bioanalyzer DNA 1000 assay trace, evaluating the DNA fragments after tagmentation reaction in WES.



D.2 Agilent 2100 Bioanalyzer High Sensitivity Assay

The following traces outline a typical Bioanalyzer High Sensitivity assay trace, evaluating the quantity and quality of the indexed libraries prior to pooling.



Appendix E

E.1 Linux command-line for WES data analysis

The following commands were used in the standard data analysis pipeline for filtering variants from WES data.

1. **Alignment to GRCh37 using BWA**

bwa mem -t 12 -M /home/ref/b37/human_g1k_v37.fasta /data/medmelp/JTXXX_S1_L001_R1_001.fastq.gz /data/medmelp/JTXXX_S1_L001_R2_001.fastq.gz -v 1 -R '@RG\tID:JTXXX\tSM:JTXXX\tPL:Illumina\tPU:HiSeq3000\tLB:JTXXX_WES_exome' -M |samtools view -Sb - > /data/medmelp/JTXXX.bwamem.bam

2. **Sort bam file**

java -Xmx4g -jar /home/picard/picard-tools-2.5.0/picard.jar SortSam I=/data/medmelp/JTXXX.bwamem.bam O=/data/medmelp/JTXXX.bwamem.sort.bam SO=coordinate CREATE_INDEX=TRUE

3. **Mark duplicates**

java -Xmx4g -jar /home/picard/picard-tools-2.5.0/picard.jar MarkDuplicates I=/data/medmelp/JTXXX.bwamem.sort.bam O=/data/medmelp/JTXXX.bwamem.sort.dedup.bam M=/data/medmelp/JT7XXX.bwamem.sort.dedup.metrics CREATE_INDEX=TRUE

4. **Create indel realigner targets**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T RealignerTargetCreator -R /home/ref/b37/human_g1k_v37.fasta -known /home/ref/b37/1000G_phase1.indels.b37.vcf -known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I /data/medmelp/JTXXX.bwamem.sort.dedup.bam -o /data/medmelp/JTXXX.bwamem.sort.dedup.intervals

5. **Perform indel realignment**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T IndelRealigner -R /home/ref/b37/human_g1k_v37.fasta -known /home/ref/b37/1000G_phase1.indels.b37.vcf -known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I /data/medmelp/JTXXX.bwamem.sort.dedup.bam -targetIntervals /data/medmelp/JTXXX.bwamem.sort.dedup.intervals -o /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.bam

6. **Get recalibration model**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T BaseRecalibrator -R /home/ref/b37/human_g1k_v37.fasta -knownSites /home/ref/b37/1000G_phase1.indels.b37.vcf -knownSites /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -knownSites /home/ref/b37/dbSnp146.b37.vcf.gz -I /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.bam -o /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.grp -nct 6

7. **Apply recalibration**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T PrintReads -R /home/ref/b37/human_g1k_v37.fasta -I /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.bam -BQSR /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.grp -o /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.bam

8. **Haplotype Caller for variant calling**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T HaplotypeCaller --emitRefConfidence GVCF --variant_index_type LINEAR --variant_index_parameter 128000 -R /home/ref/b37/human_g1k_v37.fasta -D /home/ref/b37/dbSnp146.b37.vcf.gz -stand_call_conf 30 -stand_emit_conf 10 -I /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.bam -o /data/medmelp/JTXXX.g.vcf

9. **Genotype gVCF and merge family vcfs together**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T GenotypeGVCFs -R /home/ref/b37/human_g1k_v37.fasta -D /home/ref/b37/dbSnp146.b37.vcf.gz -stand_call_conf 30 -stand_emit_conf 10 -V /data/medmelp/JTXXX.g.vcf -V /data/medmelp/JTXXX.g.vcf -V /data/medmelp/JTXXX.g.vcf -V /data/medmelp/JTXXX.g.vcf -o /data/medmelp/JTXXX-XXX.raw.vcf –showFullBamList

10. **Separate SNPs**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T SelectVariants -R /home/ref/b37/human_g1k_v37.fasta --variant /data/medmelp/JTXXX-XXX.raw.vcf -selectType SNP -o /data/medmelp/JTXXX-XXX.snps.vcf

11. **Separate INDELS**

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T SelectVariants -R /home/ref/b37/human_g1k_v37.fasta --variant /data/medmelp/JTXXX-XXX.raw.vcf -selectType INDEL -selectType MNP -o /data/medmelp/JTXXX-XXX.indels.vcf

12. Variant hard filtering for SNPs

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T VariantFiltration -R /home/ref/b37/human_g1k_v37.fasta -V /data/medmelp/JTXXX-XXX.snps.vcf --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MappingQualityRankSum < -12.5" --filterName "snp_hard_filter" -o /data/medmelp/JTXXX-XXX.fltd-snps.vcf

13. Variant hard filtering for INDELS

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T VariantFiltration -R /home/ref/b37/human_g1k_v37.fasta -V /data/medmelp/JTXXX-XXX.indels.vcf --filterExpression "QD < 2.0 || FS > 200.0" --filterName "indel_hard_filter" -o /data/medmelp/JTXXX-XXX.fltd-indels.vcf

14. Combine variants into a merged vcf

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T CombineVariants -R /home/ref/b37/human_g1k_v37.fasta --variant /data/medmelp/JTXXX-XXX.fltd-snps.vcf --variant /data/medmelp/JTXXX-XXX.fltd-indels.vcf -o /data/medmelp/JTXXX-XXX.fltd-combinedvars.vcf --genotypemergeoption UNSORTED

15. Filter on dbSNP with <1%MAF

perl /home/vcfhacks-v0.2.0/annotateSnps.pl -d /home/ref/b37/dbSnp146.b37.vcf.gz /home/ref/b37/clinvar_20160531.vcf.gz -f 1 -pathogenic -i /data/medmelp/JTXXX-XXX.fltd-combinedvars.vcf -o /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.vcf

16. Filter on EVS with <1% MAF

perl /home/vcfhacks-v0.2.0/filterOnEvsMaf.pl -d /home/ref/evs/ -f 1 --progress -i /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.vcf -o /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.vcf

17. Filter on ExAC with <1% MAF

perl /home/vcfhacks-v0.2.0/filterVcfOnVcf.pl -f /home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz -w -y 0.01 -i /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.vcf -o /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.exac.vcf -progress -fork 12

18. Apply variant effect predictor (VEP)

perl /home/variant_effect_predictor/variant_effect_predictor.pl --offline --vcf --everything --dir /home/variant_effect_predictor/vep_cache --plugin

Condel,/home/variant_effect_predictor/vep_cache/Plugins/config/Condel/config/ --plugin SpliceConsensus --assembly GRCh37 -input /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.exac.vcf -o /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.exac.vep.vcf

19. **Find biallelic variants (based on pedigree file)**

perl /home/vcfhacks-v0.2.0/findBiallelic.pl -i /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.exac.vep.vcf -o /data/medmelp/JTXXX-XXX.fltd-biallelic.vep.vcf -f /data/medmelp/JTXXX-XXX.ped

Or filter for de novo variants (for some cases)

perl /home/vcfhacks-v0.2.0/filterOnSample.pl --input /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.vep.vcf --samples [samples to keep variants if present in all] --reject [samples to reject variants from if present in any]

20. **Get functional variants (optional)**

perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl -i /data/medmelp/JTXXX-XXX.vep.vcf --damaging all --keep_any_damaging --pass --progress -o /data/JTXXX.functvyarvep.vcf

21. **Rank on CADD score**

perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl -c /data/shared/cadd/v1.3/*.gz -i /data/medmelp/JTXXX-XXX.fltd-biallelic.vep.vcf -o /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.vep.cadd_ranked.vcf -n cadd_not_found.tsv --progress

22. **Gene annotator**

perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb -m vep -i /data/medmelp/JTXXX-XXX.fltd-combinedvars.1pcdbsnp.1pcEVS.vep.cadd_ranked.vcf -o /data/medmelp/JTXXX-XXX.fltd-BiallelicVars.gene_anno

23. **Convert to excel file**

perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i /data/medmelp/JTXXX-XXX.fltd-BiallelicVars.gene_anno --vep --gene_anno -o /data/medmelp/JTXXX-XXX.fltd-BiallelicVars.vep.cadd_ranked.filtered.gene_anno.xlsx -f

Appendix F

F.1 Quality control using Picard tools

The following command line was used for quality control of the .bam files generated for each sample sequenced. The tool is known as Picard tools CollectMultipleMetrics (https://broadinstitute.github.io/picard/command-line-overview.html#CollectMultipleMetrics).

Clean Bam:

java -Xmx4g -jar /home/picard/picard-tools-1.141/picard.jar CleanSam I=/data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.bam O=/data/medmelp/XXX.bwamem.sort.dedup.indelrealn.recal.cln.bam CREATE_INDEX=TRUE

Collect Multiple Metrics – Graphical representation

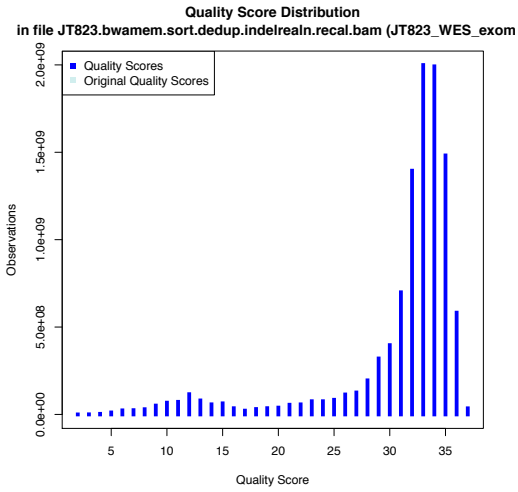
java -Xmx4g -jar /home/picard/picard-tools-1.141/picard.jar CollectMultipleMetrics I=/data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.cln.bam O=/data/medmelp/JTXXX/alignment_metrics/JTXXX.bwamem.sort.dedup.indelrealn.recal.cln R=/home/ref/b37/human_g1k_v37.fasta

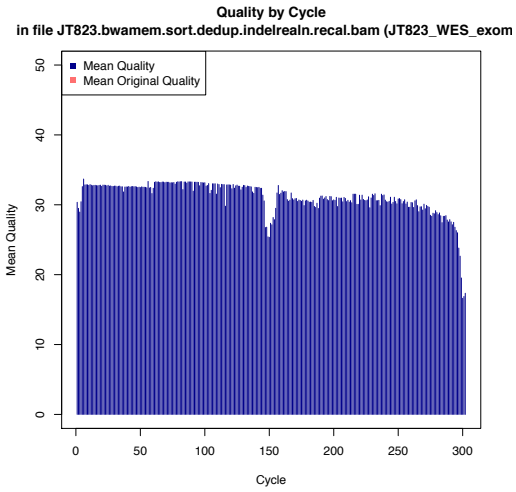
Checking % Duplicates

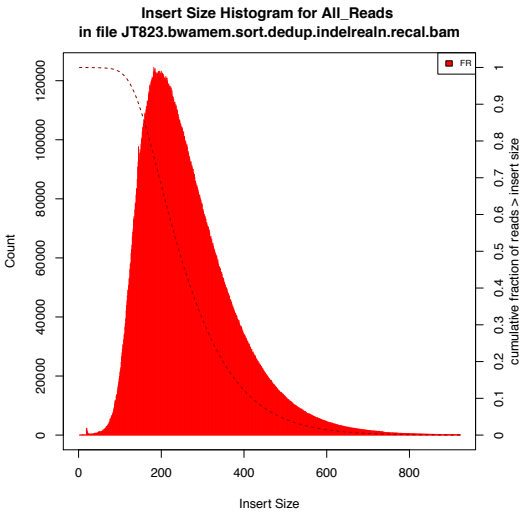
samtools flagstat /data/medmelp/JTXXX.bwamem.sort.dedup.indelrealn.recal.cln.bam > /data/medmelp/JTXXX/alignment_metrics/JTXXX.flagstat.metrics.txt

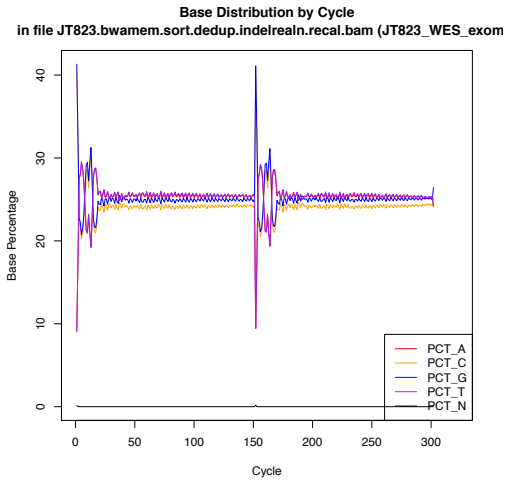
F.2 Exemplar of CollectMultipleMetrics output

The following graphical representations outline the output obtained for each sample sequenced using CollectMultipleMetrics from Picard tools.









Appendix G

G.1 Depth of coverage command

$ java –Xmx10g –jar GATK.jar –T DepthOfCoverage –R /home/ref/b37/human_g1k_v37.fasta -i sample_1.bam -i sample_2.bam -I sample_3.bam –o output.coveragedepth.txt –L /home/ref/SureSelectAllExonV5/S04380110_Regions_b37.bed –ct 5 –ct 10 –ct 20

G.2 Exemplar output of depth of coverage

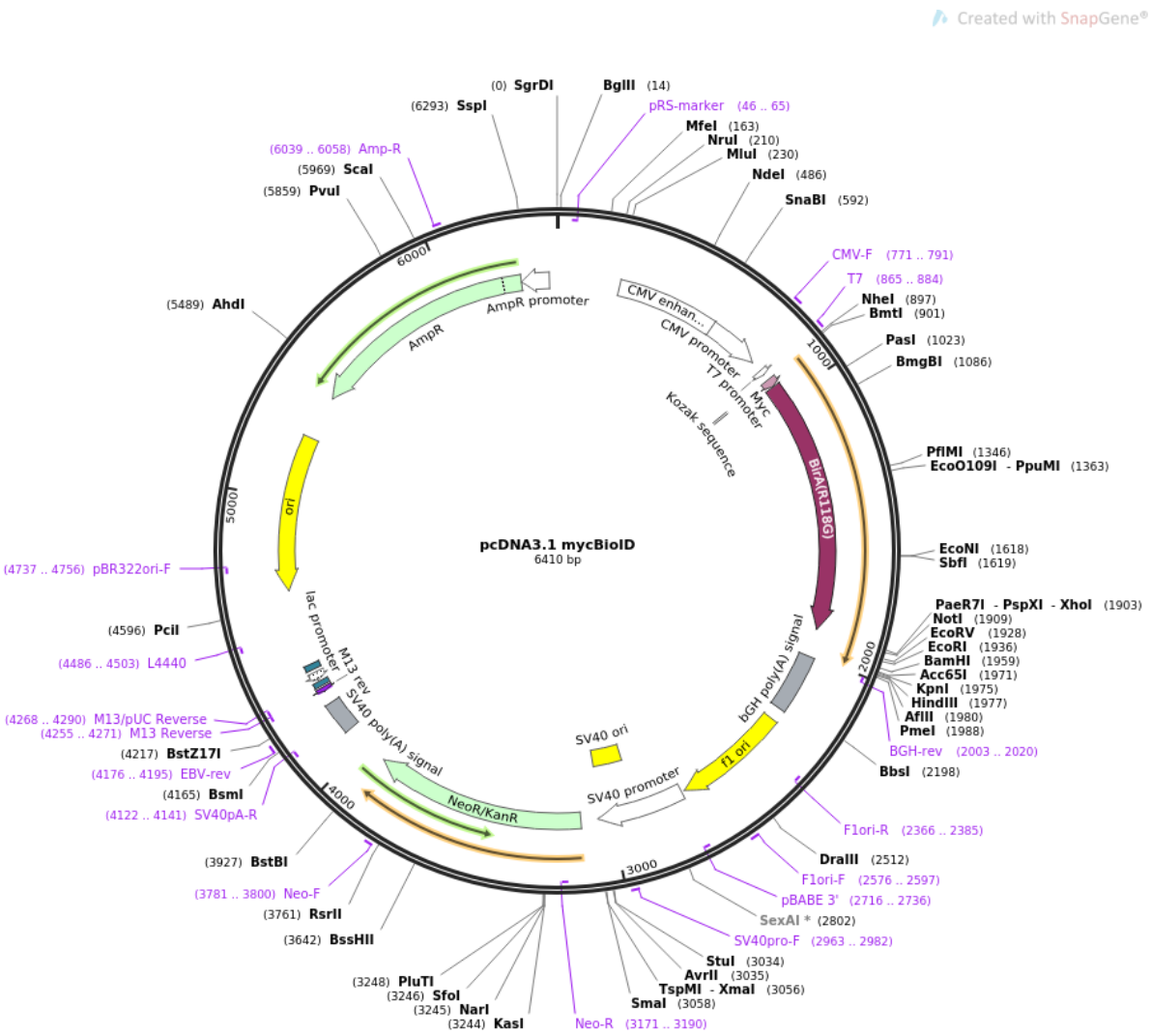
Sample ID	total	mean	third quartile	median	first quartile
JT725	8354771371	138.28	181	128	85
JT726	7582420392	125.49	166	113	72
JT727	4459553701	73.81	98	66	41
JT728	9370917192	155.09	205	139	88
Total	29767662656	492.67			

Sample ID	%bases_ above5	%bases_ above10	%bases_ above15	%bases_ above20
JT725	99.5	99.2	98.8	98.2
JT726	99.6	99.1	98.4	97.5
JT727	99.3	98.1	96	93.2
JT728	99.7	99.4	98.9	98.3

Appendix H

H.1 Plasmid used for BioID cloning

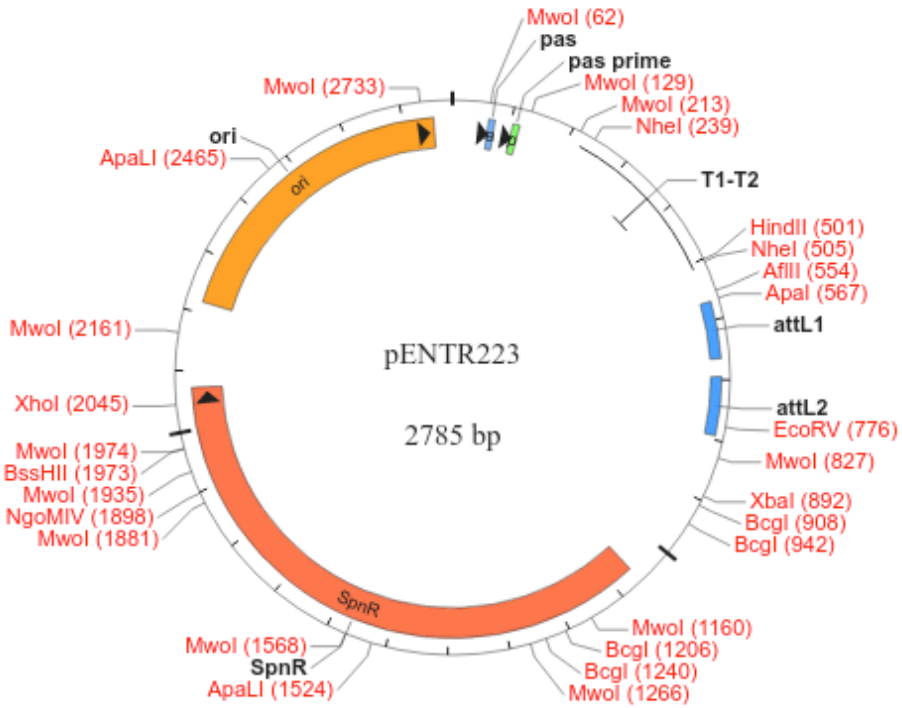
The following plasmid, pcDNA3.1_mycBioID_N-term, was obtained from Addgene (www.addgene.org) for BioID cloning purposes.

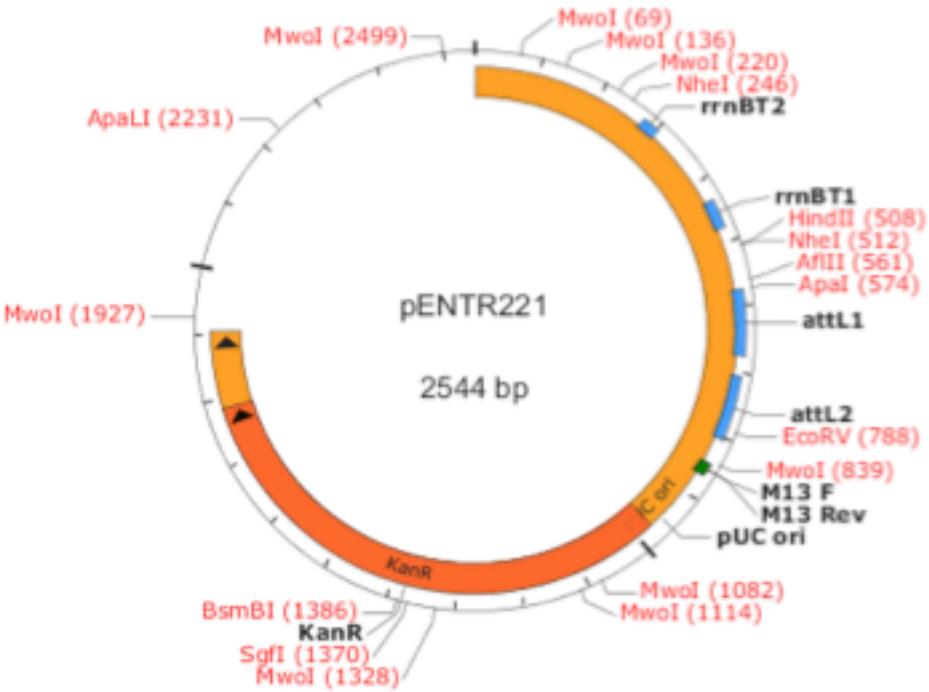


Appendix I

I.1 Expression vectors

The following expression vectors for *MAL* (pENTR223) and *KLHL7* (pENTR221) were used for cloning purposes.





Appendix J

J.1 Raw data from KLHL7 BioID experiment

T: Majority protein IDs	T: Gene names	KLHL7_B_x/y_ KLHL7	KLHL7_B_x/y_ Control_B	KLHL7_x/y_ Control	C: Student's T-test Significant KLHL7_B_Control_B	C: Student's T-test Significant KLHL7_Control	C: Student's T-test significant	C: KLHL7_B_x/y_KLHL7 A significant	C: KLHL7_B_x/y_Control_B A significant	C: KLHL7_x/y_Control A significant	C: KB_x/y_K A significant	C: KB_x/y_CB A significant	N: Peptides
Q9NVP2	ASF1B	104017000	5.27536	1.32E-08		+	KLHL7_Control	+	+				4
P46060	RANGAP1	48477000	48477000	1				+	+				8
O15111	CHUK	4.55756	3.60065	10176500				+		+			6
P60604	UBE2G2	3.80422	3.6424	41460500				+		+		+	3
Q86UE8	TLK2	3.76747	3.24132	17671000				+		+			4
Q8IXQ5	KLHL7	3.44239	4197300000	1.219E+09	+	+	KLHL7_B_Control_B;KLHL7_Control	+	+	+		+	27
Q6NYC1	JMJD6	3.41327	71545500	20961000				+	+	+			4
Q14142	TRIM14	3.32829	62878000	18892000				+	+	+	+	+	4
Q8WUF5	PPP1R13L	3.14766	2.57776	11814500				+		+			8
Q9NRW7	VPS45	3.14639	0.897755	22710500				+		+			5
P19474	TRIM21	3.00211	90378600	30105000				+	+	+		+	4
Q8IWT6	LRRC8A	2.95473	0.966026	27036500				+		+			7
Q6P3W7	SCYL2	2.79667	0.893638	22758500				+		+			7
Q9HAV4	XPO5	2.75177	1.12651	44331500				+		+			11
P36405	ARL3	2.40501	87917500	36556000					+	+	+	+	3
Q9H9Y6	POLR1B	2.24282	103045000	45944500		+	KLHL7_Control		+	+	+	+	7
P51148	RAB5C	2.20573	167845000	76095000	+		KLHL7_B_Control_B		+	+			6
Q8TEM1	NUP210	2.15364	84028500	39017000		+	KLHL7_Control		+	+			4
Q99685	MGLL	1.7238	114308000	66311500		+	KLHL7_Control		+	+		+	4
Q96GD4	AURKB	1.589	69528400	43756000					+	+			6
Q8IY21	DDX60	1.54659	18.0078	45.1951	+		KLHL7_B_Control_B		+	+		+	26
Q8TCS8	PNPT1	1.52308	146520000	96199600	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+			13
Q9NUQ7	UFSP2	1.50545	78386000	52068000		+	KLHL7_Control		+	+		+	3

T: Majority protein IDs	T: Gene names	KLHL7_B_x/y_ KLHL7	KLHL7_B_x/y_ Control_B	KLHL7_x/y_ Control	C: Student's T-test Significant KLHL7_B_Control_B	C: Student's T-test Significant KLHL7_Control	C: Student's T-test significant	C: KLHL7_B_x/y_KLHL7 A significant	C: KLHL7_B_x/y_Control_B A significant	C: KLHL7_x/y_Control A significant	C: KB_x/y_K A significant	C: KB_x/y_CB A significant	N: Peptides
Q8IVG5	SAMD9L	1.38192	16.2139	32.4615	+		KLHL7_B_Control_B		+	+		+	29
Q13287	NMI	1.34436	146985000	109334000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	8
P09914	IFIT1	1.31651	43.6681	70.7665	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	39
Q8IXQ6	PARP9	1.28534	12.6096	484540000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	19
P20591	MX1	1.24914	149340000	119555000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	9
Q8N6T3	ARFGAP1	1.22028	90674000	74306000					+	+	+	+	7
P29728	OAS2	1.17395	841460000	716779000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	19
Q8IVU3	HERC6	1.16779	539885000	462315000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	25
Q96C10	DHX58	1.16014	136925000	118025000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	8
P20339	RAB5A	1.15216	97575000	84689100	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+			7
Q8TDB6	DTX3L	1.15047	191070000	166080000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	13

T: Majority protein IDs	T: Gene names	KLHL7_B_x/y_KLHL7	KLHL7_B_x/y_Control_B	KLHL7_x/y_Control	C: Student's T-test Significant KLHL7_B_Control_B	C: Student's T-test Significant KLHL7_Control	C: Student's T-test significant	C: KLHL7_B_x/y_KLHL7 A significant	C: KLHL7_B_x/y_Control_B A significant	C: KLHL7_x/y_Control A significant	C: KB_x/y_K A significant	C: KB_x/y_CB A significant	N: Peptides
Q53EU6	AGPAT9	1.14034	126235000	110699000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	6
Q9H930	SP140L	1.12216	90198900	80380000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	11
P20592	MX2	1.11896	582190000	520295000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	17
Q9UP95	SLC12A4	1.10931	73467000	66227500		+	KLHL7_Control		+	+		+	10
P09913	IFIT2	1.10717	1803850000	1.629E+09	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	27
O14879	IFIT3	1.08895	260580000	239295000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	15
O14933	UBE2L6	1.05953	13.0554	182665000		+	KLHL7_Control		+	+		+	7
Q9BYX4	IFIH1	1.0513	332785000	316545000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	18
Q9UII4	HERC5	1.0283	868250000	844355000	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	18
Q9Y6K5	OAS3	1.02347	26.2604	28.0346	+	+	KLHL7_B_Control_B;KLHL7_Control		+	+		+	45
O95786	DDX58	1.01778	97.8502	121.96	+		KLHL7_B_Control_B		+	+		+	42
Q03518	TAP1	1.01075	63.1628	2.243E+09		+	KLHL7_Control		+	+		+	16
Q03519	TAP2	1.00764	14.6324	958286000		+	KLHL7_Control		+	+		+	13

Appendix K

K.1 Raw data from MAL BioID2 experiment (wildtype)

T: Majority protein IDs	T: Gene names	WB_x/y_CB	WB_x/y_W	MB_x/y_CB	MB_x/y_M	MB_x/y_WB	C: WB_x/y_CBA significant	C: MB_x/y_CBA significant	C: MB_x/y_WBA significant
O14828	SCAMP3	113270000	113270000	1	1	8.83E-09	+		+
O15294	OGT	205540000	205540000	1	1	4.87E-09	+		+
O43747	AP1G1	183570000	183570000	1	5.55E-09	5.45E-09	+		+
O43865;Q96HN2	SEPT3.	127180000	127180000	1	1	7.86E-09	+		+
O95292	VAPB	243170000	243170000	1	1	4.11E-09	+		+
P02511	CRYAB	424730000	424730000	1	1	2.35E-09	+		+
P09874	PARP1	235580000	235580000	1	1	4.24E-09	+		+
P12081	HARS	175210000	175210000	1	5.21E-09	5.71E-09	+		+
P13861	PRKAR2A	93972000	93972000	1	9.70E-09	1.06E-08	+		+
P14314	PRKCSH	292710000	292710000	1	1	3.42E-09	+		+
P27105	STOM	128690000	128690000	1	1	7.77E-09	+		+
P30876	POLR2B	302040000	302040000	1	1	3.31E-09	+		+
P42704	LRPPRC	211640000	211640000	1	4.36E-09	4.73E-09	+		+
P49591	SARS	176420000	0.973084	1	1	5.67E-09	+		+
P52788	SMS	137330000	137330000	1	1	7.28E-09	+		+
P84095	RHOG	140110000	140110000	1	1	7.14E-09	+		+
Q13557	CAMK2D	309760000	309760000	1	1	3.23E-09	+		+
Q13573	SNW1	149920000	149920000	1	1	6.67E-09	+		+
Q14008	CKAP5	153390000	153390000	1	1	6.52E-09	+		+
Q14254	FLOT2	80492000	80492000	1	1	1.24E-08	+		+
Q15836;P63027	VAMP3;VAMP2	143230000	143230000	1	7.40E-09	6.98E-09	+		+
Q6P1M3	LLGL2	125650000	125650000	1	1	7.96E-09	+		+
Q7Z2W4	ZC3HAV1	136670000	136670000	1	1	7.32E-09	+		+
Q92900	UPF1	236770000	236770000	1	2.93E-09	4.22E-09	+		+
Q99653	CHP1	105170000	105170000	1	1	9.51E-09	+		+
Q9BS26	ERP44	109800000	109800000	1	1	9.11E-09	+		+
Q9NP72	RAB18	107010000	107010000	1	1	9.34E-09	+		+
Q9NT62	ATG3	201520000	201520000	1	1	4.96E-09	+		+
Q9Y224	C14orf166	649450000	649450000	1	1	1.54E-09	+		+
Q9Y5B9	SUPT16H	3.70569	6.75382	1.9331	1.14503	0.521657	+		

K.2 Raw data from MAL BioID2 experiment (mutant)

T: Majority protein IDs	T: Gene names	WB_x/y_CB	WB_x/y_W	MB_x/y_CB	MB_x/y_M	MB_x/y_WB	C: WB_x/y_CBA significant	C: MB_x/y_CBA significant	C: MB_x/y_WBA significant
O14776	TCERG1	1	1	101620000	101620000	101620000		+	+
O15126	SCAMP1	1	1	188730000	188730000	188730000		+	+
O15372	EIF3H	1	1	196220000	0.896268	196220000		+	+
O60664	PLIN3	1	1	123240000	123240000	123240000		+	+
O75348;O95670	ATP6V1G1;ATP6V1G2	1	1	122070000	0.733242	122070000		+	+
O76094	SRP72	1	1	84873000	84873000	84873000		+	+
P08574	CYC1	1	1	219840000	219840000	219840000		+	+
P08754	GNAI3	1	1	164840000	0.928415	164840000		+	+
P15374	UCHL3	1	1.17E-08	90696000	0.94543	90696000		+	+
P30038	ALDH4A1	1	1	125740000	0.946125	125740000		+	+
P30040	ERP29	1	1	247120000	247120000	247120000		+	+
P31949	S100A11	1	1	1405700000	0.850393	1405700000		+	+
P39748	FEN1	1	1	174610000	174610000	174610000		+	+
P46087	NOP2	1	1	265550000	265550000	265550000		+	+
P49770	EIF2B2	1	1	31099000	0.969844	31099000		+	+
P54105	CLNS1A	1	1	203530000	0.745887	203530000		+	+
P54577	YARS	1	1	385610000	385610000	385610000		+	+
P55884	EIF3B	1	1	143150000	143150000	143150000		+	+
P63010	AP2B1	1	1	376530000	0.911077	376530000		+	+
Q00688	FKBP3	1	1	195280000	195280000	195280000		+	+
Q02543	RPL18A	2.2794	1.42122	9.39644	2.06444	4.12233		+	+
Q13310	PABPC4	1	1	460830000	460830000	460830000		+	+
Q13586	STIM1	1	1	159700000	159700000	159700000		+	+
Q14683	SMC1A	1	1	151560000	0.720205	151560000		+	+
Q9H8S9;Q7L9L4	MOB1A;MOB1B	1	1	154550000	154550000	154550000		+	+
Q8WVJ2	NUDCD2	1	1	125930000	1.25959	125930000		+	+
Q8WWI1	LMO7	1	1	342840000	342840000	342840000		+	+
Q92599	SEPT8	1	1	250070000	0.848817	250070000		+	+
Q92804	TAF15	1	1	314490000	314490000	314490000		+	+
Q99757	TXN2	1	1	212780000	212780000	212780000		+	+
Q9NZL4	HSPBP1	1	1	61683000	61683000	61683000		+	+
Q9P2R3	ANKFY1	1	1	144060000	144060000	144060000		+	+
Q9UI12	ATP6V1H	1	1	235350000	235350000	235350000		+	+
Q9UKV3	ACIN1	1	1	336970000	336970000	336970000		+	+
Q9UKY7	CDV3	1	1	356720000	356720000	356720000		+	+
Q9Y2B0	CNPY2	1	1	261770000	261770000	261770000		+	+