Exome sequencing

in rare neurological disorders

Tetyana Smertenko

This Thesis submitted to the Newcastle University in candidature for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Institute of Genetic Medicine

January 2016

Abstract

Neurological disorders are complex traits, manifesting as a range of diverse phenotypes. The current diagnostic approach involves either stepwise testing, which is expensive and time consuming, or targeted next generation sequencing with a limited portfolio of genes. Both of these approaches have a lower diagnostic yield. Whole exome sequencing may be a more advantageous and faster method to discover disease causing gene mutations.

This study evaluates the use of whole exome sequencing for diagnostic purposes in neurological disorders. Whole exome sequencing was performed in a heterogeneous cohort of patients with suspected inherited ataxia as an example of a neurological disorder, with the aim to identify candidate gene mutations. The study cohort consisted of 35 affected individuals from 22 randomly selected families of white European descent with no known consanguinity. All common sporadic, inherited and metabolic causes were excluded on routine clinical investigations prior to inclusion in this study. Whole exome sequencing was performed on 30 affected individuals. In-house bioinformatic analysis was based on previously published tools. A variant filtering algorithm excluded synonymous variants and focused on protein altering variants, nonsense mutations, exonic insertions/deletions and splice site variants. Minor allele frequency (MAF) was set at 1% in dbSNP137, 1000 genomes (April 2012 data release) and NHLBI-ESP6500 databases as well as in 286 unrelated in-house controls. Selection of the remaining variants was based on mode of inheritance. The variants were prioritized for brain and nerve cell expression and defined using carefully selected criteria. Genetic analysis was supported further by molecular genetic approaches (Sanger sequencing, reverse transcription PCR, quantitative pyrosequencing, cloning for allelic *cis-trans* study) and proteomics (Western blotting, immunohistochemistry).

Confirmed pathogenic variants were found in 9/22 probands (41%) implicating 6 genes (*KCNC3, SPG7, TUBB4A, SLC1A3, SACS* and *NPC1*). Likely *de novo* dominant *TUBB4A* mutations were found in two families. In one family quantitative pyrosequencing revealed varying degrees of mosaicism in the mildly affected mother and heterozygosity in the severely affected offspring. *In silico* analysis further supported pathogenicity of the mutation and revealed that it could potentially disrupt

polymerizations of αβ-tubulin heterodimers.

Possible pathogenic variants were identified in 5/22 probands (23%) implicating 5 genes (*ZFYVE26, ZFYVE27, WFS1, WNK1* and *FASTKD2*). A predicted splice site mutation was detected in three members of an autosomal dominant pedigree in the previously described gene *ZFYVE27*. The ZFYVE27 protein (protrudin) levels were increased approximately 2.5 fold in the cerebellum but not in the frontal cortex of the affected individual. Protrudin is an endoplasmic reticulum (ER) protein and its anomalies have previously been shown to cause ER stress. In this study levels of the master regulator of ER stress, BiP/GRP78, were significantly increased in the patient's cerebellum, which may indicate the ER pathology. In one family with possible pathogenic compound heterozygous *FASTKD2* mutations, the *in silico* splice-site prediction was validated by sequencing analysis of cDNA clones. Likely *de novo* compound heterozygous mutations in *ZFYVE26* (SPG15) in one family was validated with sequencing of cloned alleles and the result confirmed occurrence of the mutations in *trans*, therefore supporting their autosomal recessive inheritance.

In conclusion, the likely molecular diagnosis in 14 out of 22 families (64%) was defined; a total of 11 genes were implicated. Disease genes previously described in isolated families were validated and the clinical phenotypes of known disease genes was broadened. This study has also demonstrated genetic heterogeneity of hereditary ataxias but shows the impact of exome sequencing in a group of patients difficult to diagnose genetically.

Acknowledgement

I would like to thank my supervisors Professor Patrick Chinnery, Professor Rita Horvath and Dr Angela Pyle for their support and guidance throughout my PhD. I am very grateful to Professor Patrick Chinnery for giving me this wonderful opportunity. My special thanks go to Dr Angela Pyle for encouragement, excellent knowledge and patience. I also thank Drs Marzena Kurzawa-Akanbi, Gavin Hudson and Florence Burte for knowledge and willingness to provide answers to my questions. I would like to thank Raf Hussain and Dr Jennifer Duff for excellent technical support. My gratitude goes to Gail Eglon for kindly providing data that accelerated this study. My gratitude is extended to all the patients and families who provided samples for this research. I would also like to thank the Newcastle Brain Tissue Resource and the Newcastle MRC Centre for Neuromuscular Disease Biobank staff for promptly providing valuable samples and data for this study. Most importantly, I would like to thank my family for their love and support. Special thanks go to I. and D. for their encouragement and understanding through these years. Finally, I thank A. for persuading me to embark upon this challenge and for his continued encouragement.

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Abbreviations

AD-HSP - autosomal dominant hereditary spastic paraplegia

ADCA - autosomal dominant cerebellar ataxias

Bp - base pairs

CCDS - consensus coding sequence

cDNA - copy DNA

CNS - central nervous system

ddNTP - dideoxynucleotides

DMSO - dimethylsulphoxide

dNTP - deoxyribonucleotide

DRPLA - dentatorubral-pallidoluysian atrophy

DYT4 - with dystonia type 4

EA - episodic ataxia

ER - endoplasmic reticulum

FC - fold change

GWAS - genome-wide association study

H-ABC - hypomyelination with atrophy of basal ganglia and cerebellum

IHC - immunohistochemistry

kDa - kilodalton

MAF - minor allele frequency

MRI - magnetic resonance imaging

NGS - next generation sequencing

PCR - polymerase chain reaction

RT-PCR - reverse transcription PCR

SBV - single base variants

SCA - spinocerebellar ataxias

SNP - single nucleotide polymorphisms

SNV - single nucleotide variants

SPAX - spastic ataxia

SPG - spastic paraplegia

UTR - untranslated region

WES - whole exome sequencing

WGS - whole genome sequencing

Publications

Keogh, M.J., Steele, H., Douroudis, K., Pyle, A., Duff, J., Hussain, R., **Smertenko, T**., Griffin, H., Santibanez-Koref, M., Horvath, R. and Chinnery, P.F. (2015) 'Frequency of rare recessive mutations in unexplained late onset cerebellar ataxia', *J Neurol*

Pyle, A., **Smertenko, T**., Bargiela, D., Griffin, H., Duff, J., Appleton, M., Douroudis, K., Pfeffer, G., Santibanez-Koref, M., Eglon, G., Yu-Wai-Man, P., Ramesh, V., Horvath, R. and Chinnery, P.F. (2014) 'Exome sequencing in undiagnosed inherited and sporadic ataxias', *Brain* 138(2)

Taylor, R.W., Pyle, A., Griffin, H., Blakely, E.L., Duff, J., He, L., **Smertenko, T**., Alston, C.L., Neeve, V.C., Best, A., Yarham, J.W., Kirschner, J., Schara, U., Talim, B., Topaloglu, H., Baric, I., Holinski-Feder, E., Abicht, A., Czermin, B., Kleinle, S., Morris, A.A., Vassallo, G., Gorman, G.S., Ramesh, V., Turnbull, D.M., Santibanez-Koref, M., McFarland, R., Horvath, R. and Chinnery, P.F. (2014) 'Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies', *JAMA*, 312(1), pp. 68-77

Pyle, A., Ramesh, V., Bartsakoulia, M., Boczonadi, V., Gomez-Duran, A., Herczegfalvi, A., Blakely, E.L., **Smertenko, T**., Duff, J., Eglon, G., Moore, D., Yu Wai Man, P., Douroudis, K., Santibanez-Koreff, M., Griffin, H., Lochmuller, H., Karcagi, V., Taylor, R.W., Patrick F. Chinnery, P.F., Rita Horvath, R. (2014) 'Behr's Syndrome is Typically Associated with Disturbed Mitochondrial Translation and Mutations in the C12orf65 Gene', *J of Neuromus Dis,* 1, pp.55–63,

Pyle, A., Griffin, H., Duff, J., Bennett, S., Zwolinski, S., **Smertenko, T**., Yu-Wai Man, P., Santibanez-Koref, M., Horvath, R. and Chinnery, P.F. (2013) 'Late-onset sacsinopathy diagnosed by exome sequencing and comparative genomic hybridization', *J Neurogenet*, 27(4), pp. 176-82

Author's declaration

This thesis is submitted for the degree of Doctor of Philosophy to Newcastle University. The research presented within was performed at the Wellcome Trust Centre for Mitochondrial Research within the Institute of Genetic Medicine and is my own work unless otherwise stated. The research was carried out under supervision of Professor Patrick Chinnery, Professor Rita Horvath and Dr Angela Pyle between October 2013 and October 2015.

I certify that none of the material offered in this thesis has been previously submitted by me for a degree or any other qualification at this or any other university.

Chapter 1

Introduction

Chapter 1. Introduction

1.1. Neurological diseases

Neurological diseases form a large group of disorders which exhibit a vast variation in age of onset, severity, clinical presentation, genetic background and response to therapeutics (Boone *et al.*, 2011). Neurological diseases caused by mutations are characterized by significant genetic and phenotypic heterogeneity and are challenging to diagnose and treat because of the highly divergent nature of each case. Analysis of the genetic cause of neurological diseases will lead to the characterization of novel molecular pathways and the identification of therapeutic targets. This in turn will aid the development of personalized medicines and genetic counselling. The hereditary nature of neurological diseases can be subdivided on the basis of their mode of inheritance into Mendelian, sporadic and mitochondrial.

1.1.1. Mendelian neurological diseases

Mendelian diseases are those, in which the phenotype is largely caused by mutation(s) in a single gene and inherited in an autosomal dominant, autosomal recessive or X-linked manner.

The simplest type of Mendelian neurological disease is a group of monogenic disorders, where the phenotype is caused by the presence of one (for a dominant trait) or two (for a recessive trait) mutant alleles. These disorders constitute only a small group of neurological diseases and are generally rare (Bras *et al.*, 2012). Nevertheless, understanding the mechanisms of diseases with simple inheritance patterns demonstrate how pathogenic mutations could affect gene functions and this knowledge could be subsequently applied to study more complex neurological diseases (Antonarakis and Beckmann, 2006).

1.1.2 Sporadic neurological diseases

Sporadic neurological diseases are those that occur in the absence of family history and a known genetic cause. *De novo* mutations are usually rare, often affect reproductive fitness and therefore are rarely transmitted to the next generation. *De novo* mutations

have not been subjected to purifying selection. Hence, they are usually more deleterious than inherited rare genetic variation and are the most likely cause of genetic diseases that occur sporadically (Foo *et al.*, 2013).

1.1.3. Mitochondrial diseases

Mitochondrial diseases are generally understood as diseases that arise as a result of dysfunction of the mitochondrial oxidative phosphorylation (OXPHOS) system, due to improper function of one or more of the mitochondrial respiratory chain complexes (Ylikallio and Suomalainen, 2012). They affect ~1 in 5000 live births (Schaefer *et al.*, 2004).

Mitochondrial diseases are a clinically heterogeneous group of inherited diseases, which can potentially affect any organ at any age and can be inherited in an autosomal dominant or recessive, X-linked, or maternal manner (Nunnari and Suomalainen, 2012).

1.2. Classical methods for identification of causative gene variants

1.2.1. Monogenic hereditary neurological disorders

Several major techniques were so far used for identification of the causative variants for monogenic hereditary neurological disorders.

Genetic linkage analysis

Genetic linkage analysis is a statistical method to identify regions on chromosomes that could contain the genes responsible for a particular phenotype. This method is based on the idea that the neighbouring genes on the chromosome are inherited together ("linked") during the meiosis. This method involves studying large pedigrees with multiple affected and unaffected individuals. DNAs from affected and unaffected individuals are genotyped for certain DNA sequence polymorphisms with known chromosomal locations (markers), which are scattered throughout the genome. If certain markers show segregation with the disease, then it could be assumed that these markers are located close on the chromosome to the disease gene. Linkage analysis makes use of linkage disequilibrium, which occurs when genetic markers are inherited together with a higher probability than would be expected if they were inherited by chance. This would

occur if markers are situated close to each other on the chromosome and thus are less likely to be separated during recombination in meiosis. Taking into consideration a recombination event, which takes place in meiosis, it is possible to determine the chromosomal region which segregates with the disease (Hossjer, 2005). The chromosomal region, which shows segregation with the disease, will probably contain the gene that harbours causative mutations. Subsequent Sanger sequencing is employed to pinpoint the mutation. An example of the successful application of genetic linkage analysis is the discovery that hereditary dysphasic disinhibition dementia, a frontotemporal dementia, is linked to 17q21-22 (Lendon *et al.*, 1998).

A significant drawback of linkage analysis is that it can only be performed in large families with multiple affected individuals. This limitation is particularly pronounced in families with a late disease onset, where affected individuals are already deceased and the young generation is yet to reach the disease onset (Bras *et al.*, 2012). In addition, the linkage analysis only identifies linkage regions and not a particular causative gene. Sometimes these regions are very large which makes it very difficult to prioritize the disease genes.

Homozygosity mapping

For consanguineous pedigrees, where an autosomal recessive mode of inheritance is hypothesized, homozygosity mapping has proved to be a powerful technique of mapping causative genes. Homozygosity mapping, a form of linkage analysis, defines the regions on chromosomes that are homozygous because they are inherited from a common ancestor and therefore could potentially contain a causative mutation (Seelow *et al.*, 2009). Homozygous regions are identified by genome-wide high-density single nucleotide polymorphism (SNP) genotyping arrays. The main principle of homozygosity mapping is that DNA markers, within close proximity to the disease gene, will be homozygous by descent due to being inherited from a common ancestor (Lander and Botstein, 1987). An example of the successful application of this technique is identifying a *COL18A1* mutation as a cause of an autosomal recessive neurological disorder in an Indian family (Paisan-Ruiz *et al.*, 2009).

Homozygosity mapping, however, does not always result in identifying a gene that harbours causative mutations. As in the case with linkage analysis, this technique would yield only the regions on chromosomes, which are predicted to be homozygous, rather

than a particular gene.

Array comparative genomic hybridization (aCGH)

DNA copy number variations (CNVs) were recently associated with common and rare diseases (Weischenfeldt *et al.*, 2013). CNV is a change in normal gene/loci dosage, which could occur due to multiplications or deletions of genomic regions. CNVs are commonly detected with array comparative genomic hybridization (aCGH). In this technique a reference and a patient's DNA samples are differently labelled with coloured fluorescent dyes, mixed and subjected to a competitive hybridization on a chip containing DNA probes (microarray). CNVs are called by comparison of fluorescence signal intensity of the two samples hybridized to each DNA probe.

However, the resolution of aCGH used for CNV calling is not high enough and is currently limited to deletions and duplications longer than 250 kb (Retterer *et al*., 2015). This could result in missing smaller CNVs, which potentially could be implicated in diseases.

Candidate gene association study

A candidate gene association study could be performed when genes known to carry disease causing mutations were already identified. In these studies different populations would be screened for the genes of interest and information such as prevalence and clinical significance would be obtained (Bras *et al.*, 2012).

The candidate gene association study approach would require prior knowledge about the genes to be studied. It is not always easy to obtain such information.

Case-control studies

In case-control studies the genes of interest would be investigated in a group of cases and a group of controls from the same population. Rare mutations are an evolutionary recent event that have occurred due to explosive accelerated population growth over the past 10000 years. It is anticipated now that most rare variants in the human genome are population-specific (Tennessen *et al.*, 2012). That is why case-control studies are of a particular importance.

1.2.2. Common disorders

Common, complex disorders, are those where phenotypes are caused by a combination of variants in several genes, none of which individually have an impact on the trait. These disorders form a group of complex diseases. There are two hypotheses as to which exact variants are contributing most to the development of the common complex disorders - "common disease, common variant" and "common disease, rare variant" (Schork *et al.*, 2009).

"Common disease, common variant" hypothesis

"Common disease, common variant" hypothesis postulates that common variants (i.e. variants with a Minor Allele Frequency (MAF) >1%), with relatively low penetrance, are major contributors to the risk of developing a common disease (Reich and Lander, 2001). By this hypothesis, not one of these variants individually is sufficient to cause the disease. This hypothesis was prevalent in the scientific community during the past decade.

Despite the progress in elucidating genes that harbour causative mutations for Mendelian diseases, common diseases have proved to be more difficult to dissect, mainly due to the locus heterogeneity. That is why a classical method for identification of causative mutations, such as linkage analysis, could not be successfully applied for common diseases (Altmuller *et al.*, 2001).

Genome-wide association study (GWAS) used the "common disease, common variant" hypothesis to investigate a large number of common single nucleotide polymorphisms (SNPs), in a large number of case and control samples, and to identify association between these common variants and the disease. GWAS was successfully applied to diseases such as Parkinson's disease (Pihlstrom *et al.*, 2012), Alzheimer's disease (Harold *et al.*, 2009) and multiple sclerosis (Matesanz *et al.*, 2012) yielding some common variants and loci which potentially modulate the risk of these common diseases. However, in each of these studies the identified loci explain only a small proportion of the heritable component of the disease. This means that inherited common variants alone cannot explain the majority of common diseases.

"Common disease, rare variant" hypothesis

The "common disease, rare variant" hypothesis argues that multiple rare variants (MAF<1%), each with relatively high penetrance, have a higher impact on the risk of developing common, complex diseases than common variants (Pritchard, 2001). Studies by several research groups subsequently demonstrated an effect of rare variants on common traits (Ji *et al.*, 2008; Nejentsev *et al.*, 2009). These high-risk variants do not cause disease directly, but rather have a higher impact on the risk of developing disease than common variants. Rare variants are hard to detect by GWAS because this technique employs common SNPs and thus would not account for rare variations. Rare variants, therefore, may explain "missing heritability" (Manolio *et al.*, 2009). The "common disease, rare variant" hypothesis is relatively recent and marks a major hypothesis shift in human genetics (McClellan and King, 2010). This hypothesis has several implications. For example, it has been proposed that in common diseases, which are highly heterogeneous, rare mutations in different genes involved in the same or related pathway could potentially cause the same disorder (McClellan and King, 2010). Indeed, recent studies in schizophrenia found that multiple genes harbouring multiple rare mutations in this disease are not a random group of genes. The majority of the mutated genes were, in fact, involved in signalling and critical neurodevelopmental pathways (Walsh *et al.*, 2008).

Employing GWAS for the identification of rare variants, that influence common disease, is not effective due to the low statistical power in detecting an association in these situations. Statistical power will decrease with a decreasing variant frequency, meaning that very large sample sizes are required in order to detect association of rare variants with common disease. (Raychaudhuri, 2011).

Considering that the traditional GWAS approach could not be successfully employed to reveal the association of rare variants with common disease, other methods are required. One of the approaches that could be used to identify rare variants involved in common disease is sequencing the extremes (Harismendy *et al.*, 2010).

To identify rare variants, involved in common disease, it has been suggested that sequencing candidate genes in individuals, at the high and low ends (extremes) of the distribution of a quantitative phenotype, should be performed (Figure 1.1).

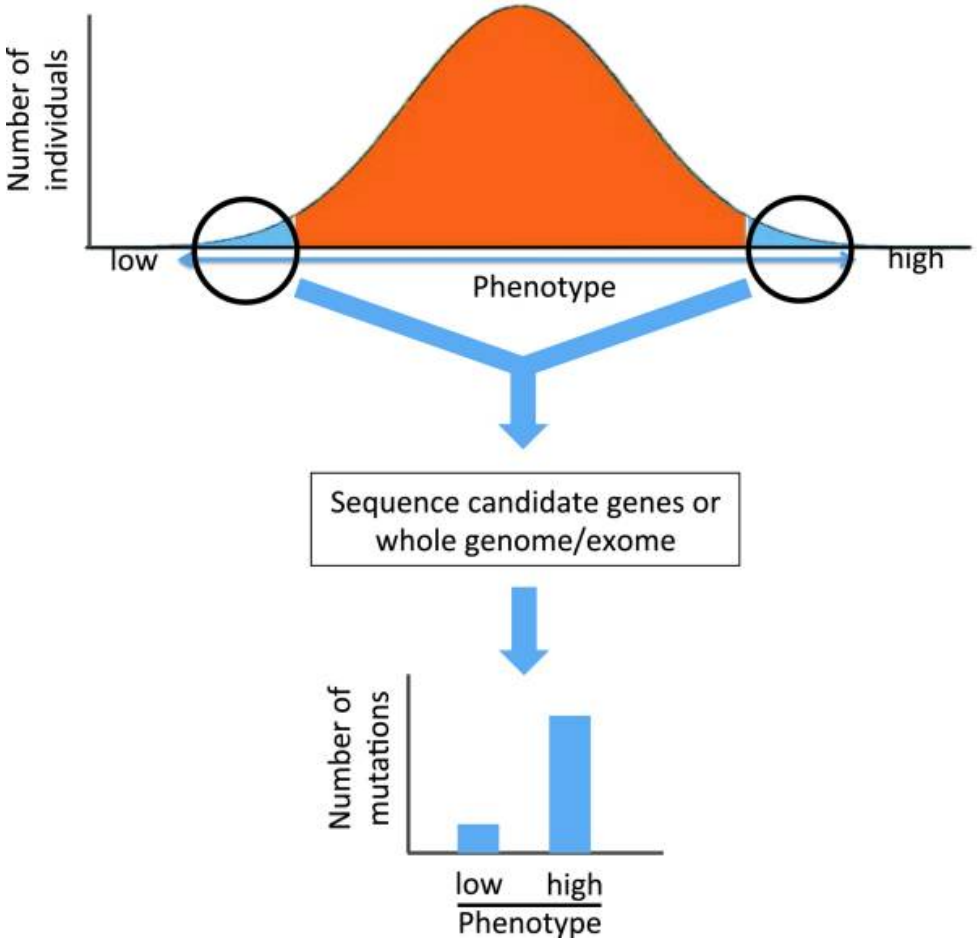


Figure 1.1: Sequencing of the extremes.
This figure shows the strategy of sequencing the extremes in order to identify rare variants involved in common disease (reproduced from Brunham and Hayden, 2013). The number of rare variants in each of the candidate genes relevant to the study is compared between the "high" and "low" groups. It is anticipated that an excess of rare variants in a given gene in the "high" group versus "low" provides evidence of involvement of this gene in the phenotype.

To study the impact of rare and private variants on the disease in a large number of individuals was very difficult until recent advances in technology. Next generation sequencing (NGS) technology is an emerging tool for identification of rare variants (Figure 1.2).

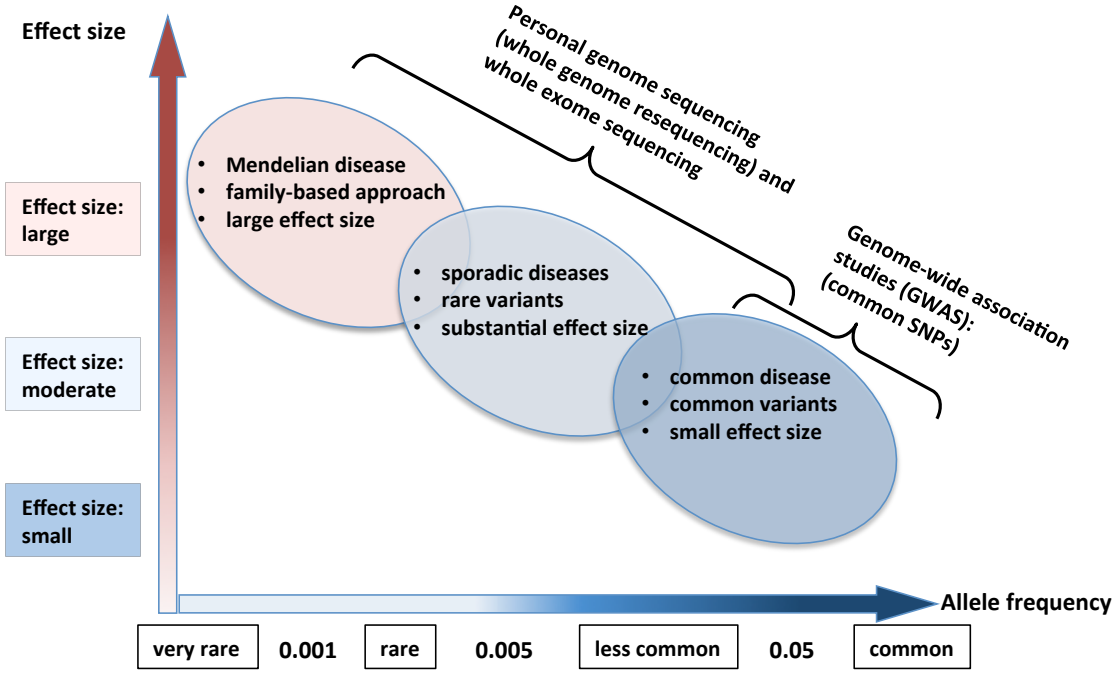


Figure 1.2: Research strategy for identification of disease-causing variants, based on the relationship between allele frequencies of variants in the population and the size of the effect caused by the variant.
For "common disease common variant" cases, where variants with high frequency and small effect size are thought to be major contributors to the risk of developing a common disease, the most effective study approach is genome-wide association studies (GWAS). These studies employ common single nucleotide polymorphisms (SNPs). Likewise, for "common disease rare variant" cases, where multiple variants with low frequency and substantial effect size are thought to contribute to developing the disease, next generation sequencing techniques such as whole genome and whole exome sequencing will be the most promising approach. For Mendelian diseases, which constitute a group of diseases caused by very rare variants with a large effect size, studying multiple or single genomes or exomes in the affected family is the most effective strategy (adapted from Tsuji, 2010).

1.3. Sequencing technologies

Automated Sanger sequencing is considered a "first generation" technology, and all newer sequencing technologies are referred to as "massively parallel sequencing" or "next generation sequencing" (NGS) (Rizzo and Buck, 2012).

1.3.1. Sanger sequencing as "first generation sequencing"

Sanger sequencing is a classical method of sequencing genes. This technology is based on the principle of chain termination and high-resolution electrophoresis. Fluorescently-labelled dideoxynucleotides (ddNTPs) are added to the reaction in addition to standard nucleotides (dNTPs). ddNTPs do not have a 3'-hydroxyl group, therefore incorporation of such nucleotides would result in termination of the DNA template elongation. This occurs by the DNA polymerase creating a pool of randomly terminated elongation products. Subsequently these terminated products are subjected to electrophoretic separation by size on a capillary-based polymer gel. Laser detection of fluorescent ddNTPs results in a fluorescence peak trace chromatogram.

1.3.2. Next Generation Sequencing

NGS refers to the technologies, which allow sequencing of massive amounts of DNA fragments in parallel. NGS allows, firstly, *de novo* sequencing, where the genome is assembled from short reads without prior knowledge of the reference genome sequence. This approach is suitable for *de novo* sequencing of smaller genomes, such as bacterial genomes (Chaisson and Pevzner, 2008). Secondly, larger genomes, like the human genome, are normally sequenced by alignment of the short reads to the reference sequence allowing mutation detection. NGS can also be applied in large-scale analysis of DNA methylation and transcriptome sequencing (Shendure and Ji, 2008).

NGS is not one single technology. There are numerous commercially available platforms. The most widely used platforms are Illumina HiSeq 2000 (Illumina, Inc.), Roche/454 (454 Life Sciences), SOLiD/ABI 5500 (Life Technologies Corporation), however, there are many others. They all differ in template preparation methods as well as the way the templates are sequenced and the data analysed (Metzker, 2010; Foo *et al.*, 2013).

Illumina HiSeq® Series (Illumina, Inc.)

The principal of the Illumina HiSeq® system is that the template is amplified on the solid surface of the flow cell using reversible dye terminator. The DNA sample is sheared into small fragments of about 200 bp and adapters are ligated to both ends of the fragments, creating a sequencing library. The library is then loaded on a flow cell and bound to its solid surface. Each fragment is amplified by bridge amplification polymerase chain reaction (PCR), which creates clusters of the fragments on the surface of the flow cell (Figure 1.3). The clusters are then sequenced through sequencing-by-synthesis.

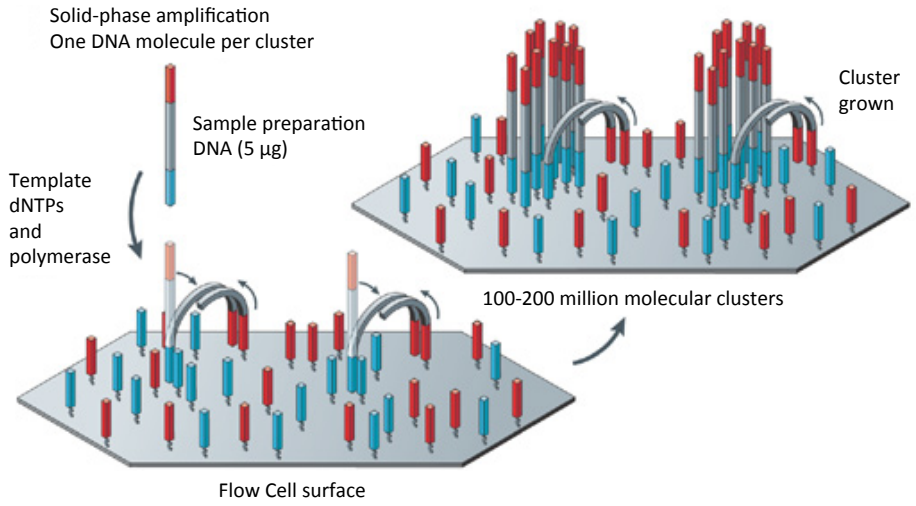


Figure 1.3: Illumina HiSeq® cluster amplification by bridge amplification PCR. For cluster generation, sample preparation DNA fragments, containing adapters ligated on both ends, are loaded on a flow cell where the fragments are captured on a flow cell surface. Each fragment is then amplified by bridge amplification PCR creating a clonal cluster (adapted from Metzker, 2010).

At the time of writing this Thesis, Illumina HiSeq® series included three platforms: HiSeq® 2500, HiSeq® 3000 and HiSeq® 4000, which differ in output range (from 10-300 Gb for HiSeq® 2500 to 125-1500 Gb for HiSeq®4000) and number of reads per flow cell (from 300 million for HiSeq® 2500 to 2.5 billion for HiSeq® 4000). The Illumina HiSeq® series allows sequencing of the whole genome, exome or transcriptome on a production scale.

Major applications of NGS

The two major applications of NGS, with regards to identification of disease causing gene variants, are whole genome sequencing and whole exome sequencing. Whole

genome sequencing (WGS) technology allows sequencing of a whole genome of a single individual, enabling identification of variants in coding and non-coding regions (Brunham and Hayden, 2013). WGS doesn't require a prior knowledge of target regions and covers the entire genome universally (Warr *et al.*, 2015). One of the main difficulties of WGS is interpretation of variations found in non-coding regions due to lack of knowledge about the importance of such regions. Hence, the consequences of variations in non-coding regions of the genome are less well understood than variations in coding regions (Ward and Kellis, 2012). WGS produces over 100 Gb of data and, despite being ultra-high-throughput technology, at present it is considered to be an expensive option in clinical practice. However, the cost of WGS is gradually going down and at present the whole genome of an individual could be sequenced in under 2 weeks for approximately US$ 4000 (Keogh and Chinnery, 2013). It is predicted that the cost of WGS will continue to fall and in the near future will drop below US$ 1000 (Foo *et al.*, 2012). Such a dramatic fall in cost of WGS, coupled with its high throughput abilities, can make WGS more accessible for clinicians in future (Keogh and Chinnery, 2013).

WGS has demonstrated its ability to successfully identify causative variants in neurological diseases. For example, in a recent study of a family with a recessive form of Charcot-Marie-Tooth disease, for which the genetic cause has not been identified, researchers sequenced the whole genome of the proband and identified all possible causative variants. Subsequent genotyping of all affected family members revealed segregation with the disease of a compound heterozygous mutation in the SH3 domain and tetratricopeptide repeats 2 (*SH3TC2*) gene (Lupski *et al.*, 2010).

Targeted sequencing of all known protein coding parts of the human genome (exome) is referred to as whole exome sequencing (WES). The other type of exome sequencing, which targets protein coding regions of only a set of chosen genes (panel) is known as targeted exome sequencing.

WES produces around 10 Gb of final data (Bras *et al.*, 2012), which is 10-fold less than WGS. WES and targeted exome sequencing are considerably less expensive than WGS (approximately 5-fold and 10-fold, respectively) (Kingsmore and Saunders, 2011). The current cost of targeted and whole exome sequencing is around US$ 620 (Nemeth *et al.*, 2013) and US$ 1000 (Foo *et al.*, 2012), respectively, which is considerably less than the

cost of WGS. The principles of exome sequencing and its applications, as well as the advantages and disadvantages of this technique, are discussed in detail in subsequent sections of this Thesis.

First proof of principle studies by Choi and colleagues and by Ng and co-workers demonstrated that exome sequencing could be successfully applied to identify causative mutations of Mendelian disorders and thus could be a potential diagnostic tool in the future (Choi *et al.*, 2009; Ng *et al.*, 2009).

1.4. Exome sequencing as a tool for dissecting neurological diseases

Neurological disorders manifest as a range of diverse phenotypes with varying age of onset, clinical presentation, severity and response to treatment (Foo *et al.*, 2013). For most neurological disorders there is no cure available at present (Foo *et al.*, 2013). Precise diagnosis is important because it could lead to the discovery of novel biochemical pathways involved in the pathogenesis of neurological disorders. This would aid in the discovery of new therapeutics and accurate genetic counselling (Biesecker *et al.*, 2012). The current diagnostic approach is stepwise testing, which is expensive, time consuming and has a low diagnostic yield. Exome sequencing is an emerging powerful technique for dissecting neurological disease (Foo *et al.*, 2013).

Exome sequencing is a hypothesis-free technique which allows an unbiased approach to mutation studies. The majority of allelic variants, known to be a cause of Mendelian disorders, disrupt protein-coding sequences, which account for approximately 1% of the human genome but harbour about 85% of disease mutations (Choi *et al.*, 2009; Stenson *et al.*, 2009). Also, a large fraction of rare non-synonymous variants in the human genome are predicted to have functional consequences and to be deleterious (Kryukov *et al.*, 2007).

1.4.1. Targeted next-generation sequencing panels

Disease specific sequencing panels allow interrogation of genes previously linked to a particular disease (Rehm *et al.*, 2013). Several custom-made panels, consisting of genes linked to neurological disorders, have been designed recently. Such panels enable targeted sequencing of genes of interest in patients with neurological disorders. For

example, the Nimblegen SeqCap EZ Neurology Panel targets 256 genes associated with 87 neurological disorders (Roche, Burgess Hill, UK) (available at http://www.nimblegen.com/products/seqcap/ez/designs/index.html). Another example of a targeted sequencing panel is the mitochondrial genome sequencing panel enabling sequencing of mitochondrial DNA and over 1000 nuclear genes encoding mitochondrial proteins (Calvo *et al.*, 2012). The latter panel could be used in patients with suspected mitochondrial disorders.

More recently several disease-specific sequencing panels were designed. Nemeth and co-workers used an ataxia panel consisting of 118 known and putative ataxia genes. They performed a pilot targeted exome sequencing study on a cohort of 50 highly heterogeneous ataxia patients (Nemeth *et al.*, 2013). The authors found 13 different mutations, with 9 being novel, in 8 different genes, namely: *PRKCG, TTBK2, SETX, SPTBN2, SACS, MRE11, KCNC3* and *DARS2*. The overall diagnostic yield was 18%, which supported an effectiveness of targeted exome sequencing in a clinical setting. Another example of a successful application of targeted exome sequencing is the identification of an inherited retinal degeneration (IRD) panel enabling the capture of exons from 73 IRD genes and 1 deep intronic splice site known to be associated with IRD (Shanks *et al.*, 2013). The authors performed targeted exome sequencing on 50 patients (36 "unknowns" and 14 positive controls) with IRD and found pathogenic mutations in 25% of patients in the "unknowns" group in a total of 6 genes, namely: *RHO, ABCA4, PDE6B, GUCY2D, CRX* and *CRB1T*.

Targeted exome sequencing has several advantages. It is argued that a well-defined number of genes being sequenced with a panel would reduce the final data output, enabling clinicians to reduce the time and cost of diagnosis (Rehm *et al.*, 2013). A pre-defined gene portfolio, hence, would reduce the amount of data input, enabling the use of less-time-consuming sequencing technology. This would therefore enable more patient samples to be analysed, compared to WGS and WES (Rehm *et al.*, 2013). The possibility to customise and update a panel is also beneficial. However, despite obvious advantages, targeted exome sequencing has several disadvantages over whole exome sequencing. Firstly, a panel has a set portfolio of genes; therefore identification of novel disease genes is not possible. Secondly, unusual clinical presentations might result in selecting the wrong panel. Currently we are witnessing a stream of reports describing novel presentations of established neurological diseases (Pfeffer *et al.*, 2015). Such

knowledge will eventually lead to updating existing panels. However, until then, the one should bear in mind the potential pitfalls of the current panels in use.

1.4.2. Whole exome sequencing

Whole exome sequencing targets the whole protein-coding part of the genome without prioritising any gene. WES does not rely on any postulation of involvement of particular genes in the given phenotype and therefore might be a less biased approach than targeted exome sequencing. In the past, numerous studies reported successful application of whole exome sequencing to various inheritable and complex neurological disorders enabling clinicians to establish their cause. Examples of neurological disorders and putative causative variants identified by whole exome sequencing are shown in Table 1.1.

Table 1.1: Neurological disorders and their putative causative variants, as identified by whole exome sequencing.
The table shows a selection of recent studies on various types of neurological disorders that used whole exome sequencing to identify putative causative variants. AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; EOAD, early-onset Alzheimer's disease; FALS, familial amyotrophic lateral sclerosis; LOAD, late-onset Alzheimer's disease; PD, Parkinson's disease; SCA, spinocerebellar ataxias; MS, multiple sclerosis (adapted and updated from Jiang *et al*., 2014)

Disease	Reference	Subtype	Gene
AD	(Guerreiro *et al.*, 2012)	EOAD	*NOTCH3*
	(Pottier *et al.*, 2012)	EOAD	*SORL1*
	(Pottier *et al.*, 2013)	EOAD	*TREM2*
	(Guerreiro *et al.*, 2013)	LOAD	*TREM2*
	(Sassi *et al.*, 2014)	EOAD	*Presenilin 1*
PD	(Koroglu *et al.*, 2013)	Juvenile parkinsonism	*DNAJC6*
	(Nuytemans *et al.*, 2013)		*VPS35*
	(Nuytemans *et al.*, 2013)		*EIF4G1*
	(Ruiz-Martinez *et al.*, 2015)	Late-onset Parkinson's disease with cognitive impairment.	*GIGYF2*
Epilepsy	(Heron *et al.*, 2012)	Autosomal dominant nocturnal frontal lobe epilepsy	*KCNT1*
	(Barcia *et al.*, 2012)	Malignant migrating partial seizures of infancy	*KCNT1*
	(Andrade *et al.*, 2012)	Teenage-onset progressive myoclonus epilepsy	*CLN6*
	(Zhou *et al.*, 2012)	Autosomal-recessive spinal muscular atrophy associated with progressive myoclonic epilepsy	*ASAH1*
	(Saitsu *et al.*, 2012b)	Ohtahara syndrome	*CASK*
	(Saitsu *et al.*, 2012a)	Ohtahara syndrome	*KCNQ2*
	(Berger *et al.*, 2012)	Intractable epilepsy of infancy	*EFHC1*
	(Lemke *et al.*, 2012)	Various types of epilepsy	*SCN1A; SCN2A; STXBP1; KCNJ10; KCTD7; KCNQ3; ARHGEF9; SMS; TPP1; MFSD8*
	(Karkheiran *et al.*, 2013)	Progressive myoclonus epilepsy syndrome	*COL6A2*
	(Hallmann *et al.*, 2014)	Progressive myoclonus epilepsy syndrome	*CARS2*
	(Blanchard *et al.*, 2015)	Early-infantile epileptic encephalopathy	*SCN8A*

Table 1.1: Continued

Disease	Reference	Subtype	Gene
MS	(Ramagopalan *et al.*, 2011)		*CYP27B1*
	(Dyment *et al.*, 2012)		*TYK2*
	(Granstrom *et al.*, 2014)		*EDNRB*
Stroke	(Cole *et al.*, 2012)	Ischemic stroke	*CSN3*
	(Kim *et al.*, 2014)	Ischemic stroke	*Paraoxonase-1*
ALS	(Wu *et al.*, 2012b)	FALS	*SOD1*
	(Williams *et al.*, 2012)	FALS	*UBQLN2*
	(Gonzalez-Perez *et al.*, 2012)	FALS	*VCP*
	(Wu *et al.*, 2012a)	FALS	*PFN1*
	(Daoud *et al.*, 2012)	Juvenile ALS	*SPG11*
	(Smith *et al.*, 2014)	FALS	*TUBA4A*
	(Johnson *et al.*, 2014)	FALS	*MATR3*
SCA	(Doi *et al.*, 2011)	Autosomal-recessive SCA	*STY14*
	(Sailer *et al.*, 2012)	Autosomal dominant SCA	*PRKCG*
	(Lee *et al.*, 2012)	Autosomal dominant SCA	*KCND3*
	(Duarri *et al.*, 2012)	Autosomal dominant SCA	*KCND3*
	(Huang *et al.*, 2012)	Autosomal dominant SCA	*ITPR1*
	(Dundar *et al.*, 2012)	Autosomal-recessive infantile onset SCA	*C10orf2*
	(Li *et al.*, 2013)	Autosomal dominant SCA	*TGM6*
	(Sun *et al.*, 2013)	Autosomal-recessive SCA	*TPP1*
	(Heimdal *et al.*, 2014)	Autosomal-recessive SCA	*STUB1*
	(Mallaret *et al.*, 2014)	Autosomal-recessive SCA	*WWOX*
	(Mancini *et al.*, 2015)	Autosomal-recessive SCA	*CLN5*

Recent data also suggest that whole exome sequencing could not only be a useful tool for the discovery of novel disease genes, but also enable the revision of diagnosis. For example, Dixon-Salazar at co-workers performed whole exome sequencing in 118 patients with paediatric-onset neurodevelopmental disorder and identified new disease genes in 19% of the cohort (Dixon-Salazar *et al.*, 2012). Notably, in 8% of the cohort putative causative mutations were discovered in genes known to cause diseases that were different from those previously diagnosed. This led to revision of the diagnosis and a change of management in these patients.

Applying whole exome sequencing to highly heterogeneous cohorts of patients, with neurological disorders, revealed that the diagnostic yield could be high. For example, Fogel and co-workers applied whole exome sequencing to a cohort of 76 patients who presented with predominantly adult- and sporadic-onset cerebellar ataxia (Fogel *et al.*, 2014). The researchers were able to find a definitive cause of disease in 21% of the patients and identified a further 40% of cases with potentially pathogenic mutations, bringing the overall diagnostic yield to 61%. In a similar study, whole exome sequencing was employed in order to establish a molecular cause of disease in a highly heterogeneous cohort of 35 patients with suspected inherited ataxia (Pyle *et al.*, 2014). The study reported identification of confirmed pathogenic variants in 41% of the patients and identified a possible cause of the disease in a further 23% of the patients, meaning that the overall likely molecular diagnosis was defined in 64% of the patients. The latter two studies highlighted the impact of whole exome sequencing on patients, which are difficult to diagnose by traditional means. As recent studies show, whole exome sequencing could also link known disease genes with unusual presentations broadening the phenotypic spectrum of the disease, for example *SPG7*, *SACS* (Pyle et al., 2014) and *ELOVL4* (Ozaki *et al.*, 2015).

1.4.3. Basic steps of exome sequencing

The basic steps of exome sequencing are: fragmentation of genomic DNA, ligation of adaptors to DNA fragments, capturing the fragments containing exons via hybridization, washing out the uncaptured DNA, elution of the captured DNA fragments, polymerase chain reaction (PCR) amplification and massively parallel sequencing followed by alignment and variant calling (Figure 1.4).

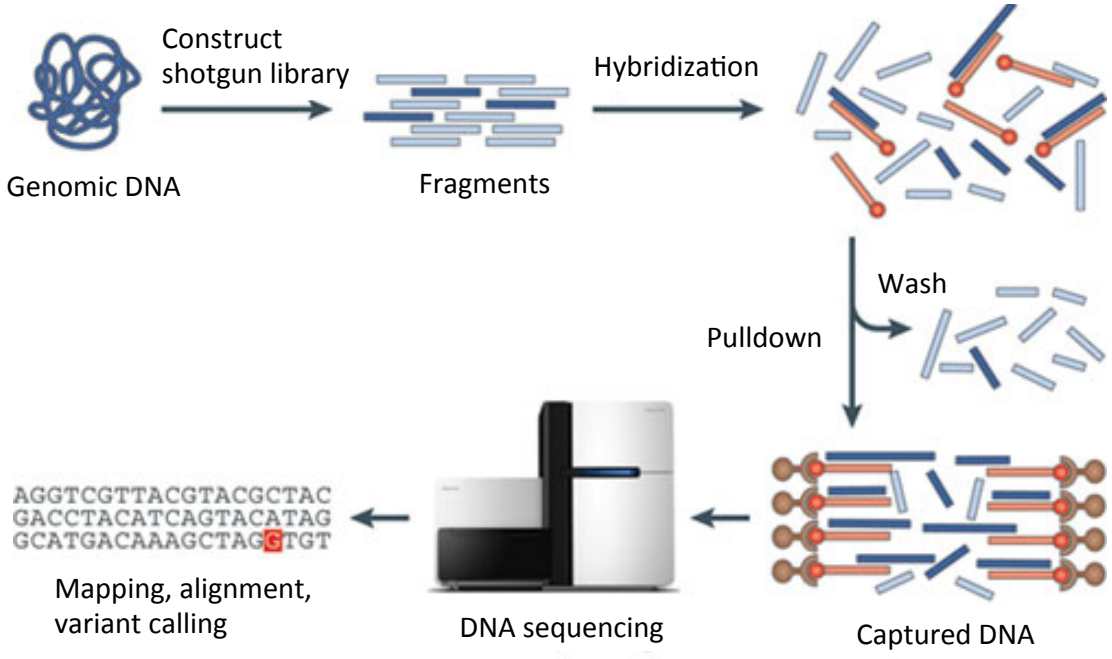


Figure 1.4: The basic steps required for exome sequencing.
The basic steps required for exome sequencing involve fragmentation of genomic DNA, constructing an *in vitro* shotgun library, flanking the library fragments by adaptors (not shown), enriching the library for sequences corresponding to exons (dark blue fragments) by hybridization of the fragments to biotinylated DNA or RNA baits (orange fragments), and recovery of the hybridized fragments by biotin–streptavidin-based pull-down. Finally, amplification and massively parallel sequencing of the enriched, amplified library is performed and mapping and calling of candidate causal variants follows (adapted from Bamshad *et al.,* 2011).

1.4.4. Prioritising of variants (filtering process)

Whole exome sequencing of an individual yields on average 20,000 variants (Bamshad *et al.*, 2011). In order to establish which of these variants account for the individual's phenotype, the variants are prioritized through a filtering process. In this Thesis the filtering process is described in detail in Chapter 3. Briefly, the variants are Compared to public databases such as 1000 genomes (Abecasis *et al.*, 2010), dbSNP (Sherry *et al.*, 2001) and Exome Variant Server (the latest release available at http://evs.gs.washington.edu/EVS/). This interrogation enables the filtering out of common variants with minor allele frequencies above 1% in these databases. This step considerably reduces the number of putative pathogenic variants, bringing it down to approximately 600 (Bamshad *et al.*, 2011). The variants are further filtered based on the mode of inheritance and involvement of the putative disease gene in the biological processes or known diseases, which would fit the phenotype. Of note, at the time this work commenced the Exome Aggregation Consortium (ExAC) data was not available and was not used. However, this has been subsequently reviewed and included in Chapter 4 of this Thesis. The latest release of ExAC data is available at http://exac.broadinstitute.org/ .

1.4.5. Establishing causality of filtered variants

After filtering of variants, as described in the previous section, the list of variants could still be long and contain dozen(s) of variants. Therefore each variant should be assessed manually and prioritized based on certain criteria, as recommended by the recent guidelines (MacArthur *et al.*, 2014; Richards *et al.*, 2015). There are numerous pathogenicity criteria that are considered before a variant could be called pathogenic, benign or classified as uncertain significance. These criteria are outlined in Table 1.2. The filtered variants are confirmed with Sanger sequencing. The pathogenicity of the variants is further investigated by a segregation study, where the available affected and unaffected relatives of the individual are investigated (segregation of the variant with the disease supports its pathogenicity). Following this, various pathogenicity prediction programs are employed in order to check how the putative causative mutation might impact on the function of the encoded protein (for example cause truncation or affect a functional domain in the protein) (Adzhubei *et al.*, 2010; Schwarz *et al.*, 2010). The variant of interest is then cross- examined in a cohort of unrelated controls of matched

ethnic background (pathogenicity of the variant is supported if the variant is not detected in the control population).

Table 1.2: Examples of evidence for variant pathogenicity.
This table shows examples of pathogenicity criteria at the gene and variant level including genetic, informatic and experimental data (adapted from MacArthur *et al.*, 2014)

Evidence level	Evidence class	Examples
Gene level	Genetic	Gene burden: the affected gene shows a statistical excess of rare (or *de novo*) probably damaging variants segregating in cases compared to control cohorts or null models.
	Experimental	Protein interactions: the gene product interacts with proteins previously implicated (genetically or biochemically) in the disease of interest. Biochemical function: the gene product performs a biochemical function shared with other known genes in the disease of interest, or is consistent with the phenotype. Expression: the gene is expressed in tissues relevant to the disease of interest and/or is altered in expression in patients who have the disease. Gene disruption: the gene and/or gene product function is demonstrably altered in patients carrying candidate mutations. Model systems: non-human animal or cell-culture models with a similarly disrupted copy of the affected gene show a phenotype consistent with human disease state. Rescue: the cellular phenotype in patient-derived cells or engineered equivalents can be rescued by addition of the wild-type gene product.

Table 1.2: Continued

Evidence level	Evidence class	Examples
Variant level	Genetic	Association: the variant is significantly enriched in cases compared to controls. Segregation: the variant is co-inherited with disease status within affected families and additional co-segregating pathogenic variants are unlikely or have been excluded. Population frequency: the variant is found at a low frequency, consistent with the proposed inheritance model and disease prevalence, in large population cohorts with similar ancestry to patients.
	Informatic	Conservation: the site of the variant displays evolutionary conservation consistent with deleterious effects of sequence changes at that location. Predicted effect on function: variant is found at the location within the protein predicted to cause functional disruption (for example, enzyme active site, protein-binding region).
	Experimental	Gene disruption: the variant significantly alters levels, splicing or normal biochemical function of the product of the affected gene. This is shown either in patient cells or a well-validated *in vitro* model system. Phenotype recapitulation: introduction of the variant, or an engineered gene product carrying the variant, into a cell line or animal model results in a phenotype that is consistent with the disease and that is unlikely to arise from disruption of genes selected at random. Rescue: the cellular phenotype in patient-derived cells, model organisms, or engineered equivalents can be rescued by addition of wild-type gene product or specific knockdown of the variant allele.

1.4.6. Strategies for finding disease-causing rare variants using exome sequencing

Sequencing of multiple unrelated affected individuals

If exomes of unrelated individuals with a similar phenotype have been sequenced, then their data could overlap. This would aid in reducing the number of candidate genes because more distantly related individuals would share less variants (Figure 1.5 a). Ng and colleagues used this approach to identify a nonsense mutation or frameshift insertions and deletions in myeloid/lymphoid or mixed-lineage leukemia 2 gene *MLL2* as a cause of Kabuki syndrome in ten unrelated individuals (Ng *et al.*, 2010a).

Sequencing the extremes

To identify rare variants involved in common disease the strategy of sequencing the extremes could be employed, as described in previous sections. This strategy is suitable for quantitative phenotypes. It involves sequencing of candidate genes relevant to the study, or whole exome sequencing, in the individuals from the both ends of the phenotype. Individuals with an excess of rare variants in the same gene are expected to fall into a "high end" of the phenotype and this particular gene is hypothesised to be a prime suspect for causing the observed phenotype (Figure 1.5 b). Using exome sequencing in this setting will accelerate the study considerably.

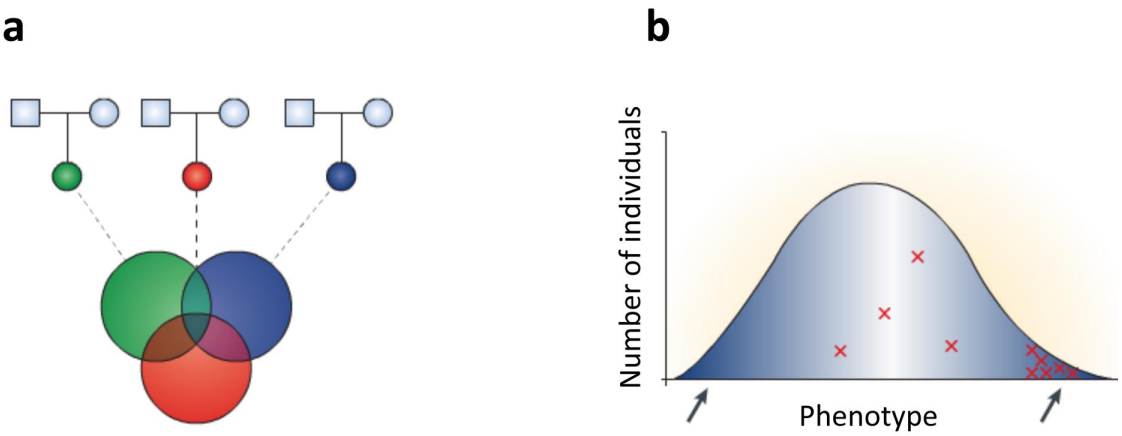


Figure 1.5: Several strategies where exome sequencing could be used for finding disease causing rare variants.
a – exome sequencing of multiple unrelated individuals exhibiting similar phenotypes is used to identify causative variants within the same gene, as outlined in the shaded region shared by all three individuals in this figure. **b** – for the quantitative phenotypes strategy "sequencing of the extremes" could be employed. It involves exome sequencing in individuals from the both ends of the phenotype. Individuals with an excess of rare variants in the same gene are expected to fall into the "high end" of the phenotype and this particular gene is hypothesised to be a prime suspect for causing the observed phenotype (adapted from Bamshad *et al.*, 2011).

Sequencing of multiple affected individuals within the same pedigree

It is important to decide which samples to sequence in an exome sequencing project because the cost of this technique is still prohibitive in comparison with other methods. The mode of inheritance of a monogenic disease will affect the study design and the way the analysis is undertaken (Bamshad *et al.*, 2011). Based on the original study's logic (Ng *et al.*, 2009), it is clear that analysing rare recessive diseases will be easier than dominant ones because there are far fewer genes in each exome that are homozygous or compound heterozygous for rare non-synonymous changes (i.e. genes with two rather than one rare protein altering allele per gene). That is why when analysing pedigrees with a recessive mode of inheritance of rare Mendelian traits, sequencing the exome of just one affected individual might be informative enough. If the family is consanguineous, then homozygosity mapping of this patient could be performed as well. The data are then combined in order to reduce the number of candidate genes.

For dominantly inherited traits it is necessary to exome sequence a minimum of two affected individuals from a studied pedigree and then to overlap the data in order to find the shared variants.

It is preferable that the exomes of most distantly related affected individuals are sequenced because in this case they would share fewer variants (Figure 1.6). Wang and co-workers sequenced exomes of four affected members in one four-generation Chinese family with autosomal dominant spinocerebellar ataxia and found that they all shared a missense mutation in a novel ataxia transglutaminase 6 gene *TGM6* (Wang *et al.*, 2010).

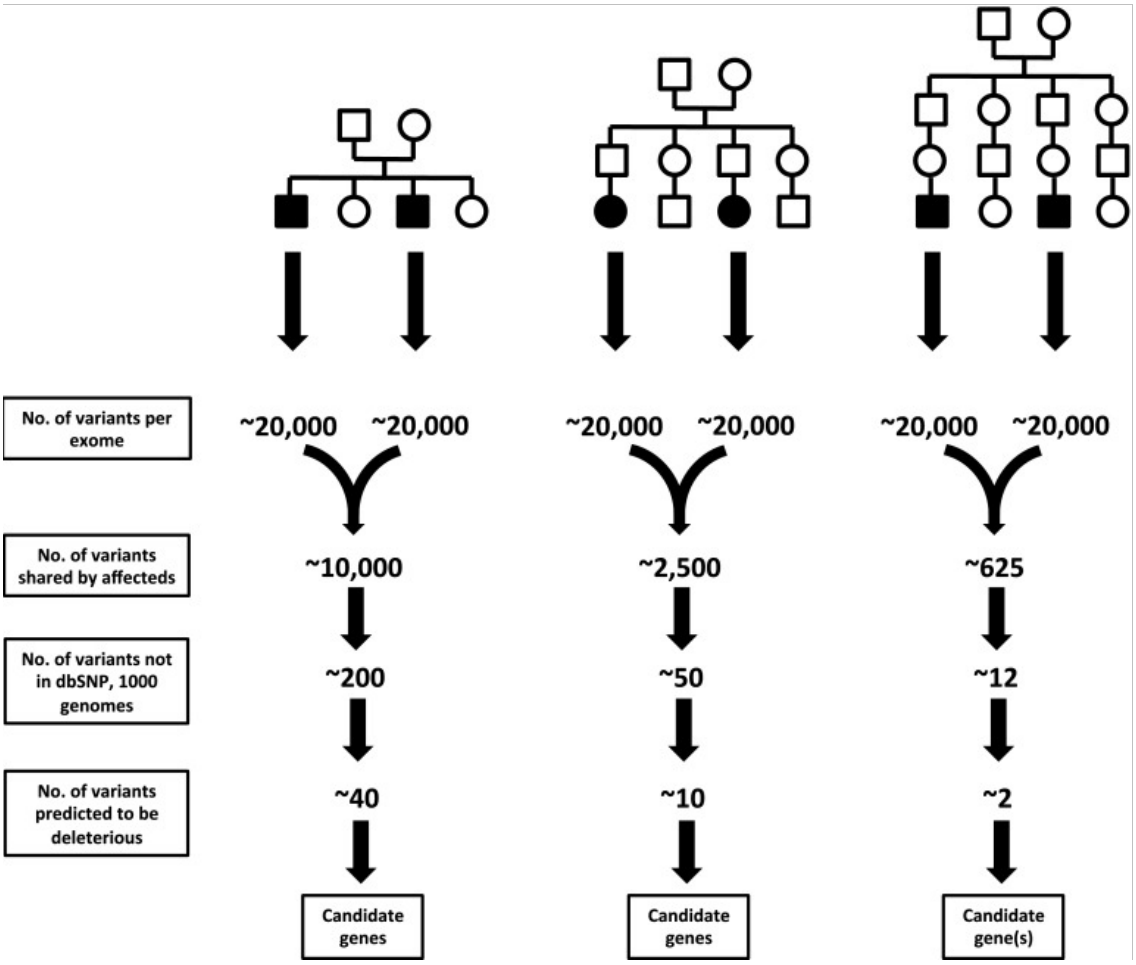


Figure 1.6: A theoretical example of how the degree to which two affected individuals are related could affect the outcome of exome sequencing data.
It is estimated, that, on the average, each sequenced exome will result in approximately 20,000 variants. When exomes of two affected siblings are intersected, the number of variants shared between them will be reduced to $1/2^{th}$ of the genome. Comparing the exomes of 1^{st} cousins and 2^{nd} cousins will reduce the number of candidate variants to roughly an $1/8^{th}$ and $1/32^{th}$ of the genome, respectively. This example assumes that all but 2% of variants will be identified in public databases such as dbSNP and 1000 genomes and the number of novel variants predicted to be deleterious will be approximately 20%. (reproduced from Brunham and Hayden, 2013).

Sequencing parents-child trios for de novo mutations

If a *de novo* mutation is suspected to be responsible for the phenotype, then exomes of parent-child trios should be sequenced. In this case only the child would be affected and excluding all variants shared between parents and their offspring would give a list of variants that are unique to the child and thus could be causative. This strategy was used in such complex traits as autism spectrum disorder (Neale *et al.*, 2012; O'Roak *et al.*, 2012), schizophrenia (Girard *et al.*, 2011), and mental retardation (Vissers *et al.*, 2010). Exome sequencing of parent-child trios was also proved to be effective in such common neurodegenerative disorders such as amyotrophic lateral sclerosis (Steinberg *et al.*, 2015) and multiple sclerosis (Dyment *et al.*, 2012).

1.4.7. Establishing number of cases needed for an exome sequencing study

There are several aspects to consider before undertaking an exome sequencing study. The most important step is to establish the number of cases which would be optimal for the particular exome sequencing study.

Mendelian diseases

It is anticipated that for Mendelian diseases, where disease-specific alleles have a large effect size, exome sequencing of as little as one individual will be needed for successful identification of causative variants (Figure 1.7).

Sporadic diseases

For sporadic and occasionally familial clustering cases only tens to hundreds of individuals might be enough to elucidate rare causative variants with moderate effect size (Tsuji, 2010).

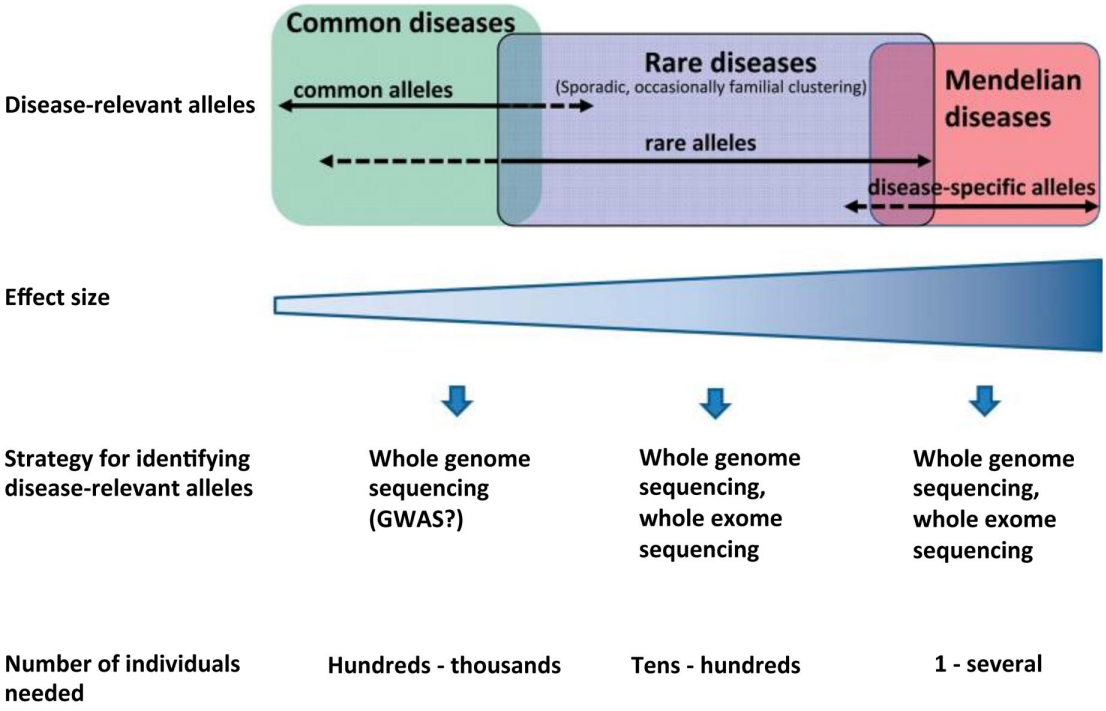


Figure 1.7: The estimated number of individuals needed for identification of disease causing variants in different groups of diseases employing whole genome sequencing and whole exome sequencing.
For Mendelian diseases, where disease-specific alleles have a large effect size, only several individuals are needed for successful identification of causative variants. For sporadic and occasionally familial clustering cases only tens to hundreds of individuals might be enough for elucidating rare causative variants with moderate effect size. For common diseases, where disease-causing common variants have small effect size, hundreds to thousands of individuals are needed (adapted from Tsuji, 2010).

Common diseases

To associate rare variants with common, complex diseases would require a large number of cases because for most human genes the power to detect associations is low (Tennessen *et al.*, 2012). It is estimated that for common diseases, where disease-causing common variants have a small effect size, hundreds to thousands of individuals are needed in order to identify causative variants (Tsuji, 2010).

1.4.8. Application of exome sequencing in different settings

Ng and colleagues were the first to use exome sequencing to find a cause of Miller syndrome, a rare recessive Mendelian disorder (Ng *et al.*, 2010b). This study pioneered the approach for identification of a cause of rare disorders in situations where large families are not available for linkage analysis and only a very small number of affected individuals are accessible for the study. Since then exome sequencing has proved to be useful in very different settings.

Combining exome sequencing with linkage analysis

Combining exome sequencing with linkage analysis, in multi-generation families with multiple affected and unaffected individuals, could aid in narrowing down a list of candidate genes. For example, in a study of spinocerebellar ataxia, combining exome sequencing and linkage analysis in three unrelated families of different ethnicity helped to identify the cause of the disease as mutations in the voltage-gated potassium channel Kv4.3-encoding gene *KCND3* (Lee *et al.*, 2012). In a more recent study of spinocerebellar ataxia type 21, in a large French family with 11 affected members, the addition of 12 family members enabled the authors to reconsider previous linkage analysis and map the disorder to chromosome 1 instead of chromosome 7, as previously proposed (Delplanque *et al.*, 2014). Furthermore, the subsequent whole exome sequencing of the affected members enabled the authors to identify the causative mutation in the transmembrane protein gene *TMEM240.* The combination of linkage analysis and whole exome sequencing enabled the authors to establish that a mutation in *KCTD17* was a cause of autosomal dominant myoclonus-dystonia in a British pedigree consisting of 7 affected subjects (Mencacci *et al.*, 2015). Finally, in a recent study of a consanguineous Palestinian Arab family, with recessive progressive myoclonus epilepsy with early ataxia, the combination of linkage analysis and WES led to the discovery of a novel homozygous missense mutation in *LMNB2* as a cause of the

observed phenotype (Damiano *et al.*, 2015).

Exome sequencing and homozygosity mapping

In situations where traditional approaches could be used (for example, homozygosity mapping), exome sequencing may simply accelerate the study. For example, Doi and co-workers, in their study of a Japanese family with autosomal recessive cerebellar ataxia, combined homozygosity mapping with exome sequencing and identified a homozygous missense mutation in synaptotagmin XIV (*SYT14*) as a cause of the disease (Doi *et al.*, 2011).

Exome sequencing in diseases with locus and phenotypic heterogeneity

When a conventional approach could not be successfully applied, exome sequencing becomes the method of first choice. For example, for diseases characterised by locus heterogeneity, it would be very laborious and expensive to screen all candidate genes for the casual mutation via PCR and Sanger sequencing. Exome sequencing could be a cost- and time-efficient approach in this setting. For example, in a recent study of three unrelated Tunisian families, diagnosed with autosomal recessive cerebellar ataxia (ARCA), exome sequencing was performed alongside homozygosity mapping and a new mutation in the *SACS* gene, as well as a known mutation in the *SPG11* gene, were identified. This study also identified a mutation in *APOB,* a gene previously associated with ataxia (Hammer *et al.*, 2012). Likewise, exome sequencing may prove very useful in dissecting diseases with phenotypic heterogeneity where only the most likely disease genes are routinely prioritised for screening (Ku *et al.*, 2012).

Exome sequencing in complex diseases

Complex neurological diseases are very difficult to dissect at present. Nevertheless, the application of exome sequencing for these scenarios was recently shown to be very promising. *De novo* mutations are the most deleterious form of rare genetic variations and are more damaging than inherited variants. This is because, being private, they have not been subjected to purifying selection (Brunham and Hayden, 2013). Until recently it was not possible to investigate the involvement and role of these mutations in the disease.

Exome sequencing of parent-child trios made it possible to realize that *de novo* mutations are the main contributors to complex forms of neurodevelopmental disorders

with severely reduced fitness, such as autism, intellectual disability and schizophrenia (Veltman and Brunner, 2012). Girard and colleagues, for example, sequenced exomes of 14 schizophrenia patients and their unaffected parents and demonstrated that *de novo* mutations occur at a much higher frequency than expected by chance (Girard *et al.*, 2011). Recent pilot studies of autistic-spectrum disorders (Neale *et al.*, 2012; O'Roak *et al.*, 2012; Sanders *et al.*, 2012) and intellectual disability (Vissers *et al.*, 2010) also suggested that *de novo* mutations are a major cause of sporadic cases of these diseases. This also highlighted the mechanism by which *de novo* mutations could contribute to these diseases being frequent in the human population. Interestingly, *de novo* mutations were not bound to a single gene, but rather were found in different genes with a role in neurodevelopment. It is anticipated that *de novo* mutations may also play a significant role in many other common diseases (Veltman and Brunner, 2012).

Complex neurological disorders, such as Parkinson's disease, Alzheimer's disease and amyotrophic lateral sclerosis are relatively common, with both common and rare variants contributing to their pathogenesis (Foo *et al.*, 2012). Over recent years, multiple studies reported successful application of exome sequencing to identify causative rare mutations in familial forms of complex neurological disorders.

In 2011 two research groups simultaneously published their studies identifying a mutation in *VPS35* as a cause of autosomal dominant late onset Parkinson's disease (PD) in several unrelated pedigrees. The first group (Vilarino-Guell *et al.*, 2011) applied whole exome sequencing to a Swiss kindred and three unrelated families who presented with late onset PD and a sporadic case. A mutation in *VPS35* (c.1858G>A; p.Asp620Asn) was found in all affected individuals but was absent in 3,309 controls. Similarly, the second group (Zimprich *et al.*, 2011) employed whole exome sequencing in 16 affected individuals from an autosomal dominant Austrian family who presented with late onset PD and found that a mutation in *VPS35* (c.1858G>A; p.Asp620Asn) segregated with the disease.

Whole exome sequencing was also successfully applied in recent studies on the pathogenesis of unexplained cases of Alzheimer's disease. (AD) For example Guerreiro and co-workers employed whole exome sequencing in order to find a genetic cause of AD in an early onset Turkish patient from a consanguineous family (Guerreiro *et al.*, 2012). The patient was found to have the same mutation in *NOTCH3* that was

previously linked to cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy.

Similar success of the application of whole exome sequencing in complex neurological disorders was reported recently in amyotrophic lateral sclerosis (ALS). Familial ALS cases constitute about 10% of all ALS cases (Dion *et al.*, 2009). Of all familial cases, about 50% still remain unsolved (Jiang *et al.*, 2014). For example, in one study whole exome sequencing was applied to a large Chinese family who presented with autosomal dominant ALS. A novel *SOD1* p.Cys146X mutation was found to segregate with the disease (Wu *et al.*, 2012b). In a more recent study, whole exome sequencing was employed to find a cause of ALS and dementia in a Caucasian family and found a *MATR3* mutation p.Phe115Cys in all affected individuals, but not in 27, 666 control chromosomes (Johnson *et al.*, 2014). Exome sequencing of an additional 108 cases with familial ALS led to the identification of a *MATR3* p.Thr622Ala mutation in a Sardinian family. These findings prompted re-sequencing of the entire *MATR3* gene in 96 British ALS cases and resulted in the identification of a *MATR3* mutation p.Pro154Ser in an individual who presented with sporadic ALS.

Studies in sporadic neurological diseases supported the causative role of *de novo* mutations in these diseases. For example, Rosewich and co-workers sequenced exomes of three parent-child trios and identified three heterozygous *de novo* mutations in the *ATP1A3* gene as a cause of alternating hemiplegia of childhood (Rosewich *et al.*, 2012). Interestingly, *ATP1A3* has been previously associated with rapid-onset dystonia. This and other studies highlighted the importance of the observation of *de novo* mutations in a gene previously reported to be involved in the phenotype.

Exome sequencing for mosaic variant discovery

Recent studies imply that exome sequencing could be applied in the identification of mosaic mutations. In order to find a cause of hypomyelination with atrophy of basal ganglia and cerebellum (H-ABC), Simons and co-workers performed exome sequencing in 10 families (5 trios, 2 quartets and 3 single probands) (Simons *et al.*, 2013). One of the quartets consisted of asymptomatic parents and severely affected siblings. Seemingly sporadic presentation and severity of H-ABC in an affected individual would normally indicate that the disease is caused by a *de novo* mutation. However, in the case of this quartet, exome sequencing showed that the father was homozygous wild-type

and the mother shared a *TUBB4A* c.745G>A; p.Asp249Asn mutation with both siblings. Detailed investigation of exome sequencing reads revealed that in the mother the ratio of wildtype to mutant allele was 3.6:1, which was considerably higher than in all other affected heterozygous individuals included in the study (1.0:1 – 1.4:1). This finding suggested mosaicism in the mother. Deep sequencing of DNA from blood, saliva and buccal cells confirmed the mosaicism in the asymptomatic mother. Overall, this study indicated that *de novo* mutations in asymptomatic mosaic individuals could cause disease in offspring, if inherited.

Mitochondrial exome sequencing

Calvo and colleagues used the principal of exome sequencing to identify the genetic cause of a mitochondrial disease in 42 unrelated infants with clinical and biochemical evidence of mitochondrial OXPHOS disease (Calvo *et al.*, 2012). Inherited defects in OXPHOS could be caused by mutations in either mitochondrial DNA (mtDNA) or nuclear DNA. Therefore, researchers performed "MitoExome" sequencing of the mtDNA as well as approximately 1000 nuclear genes encoding proteins targeted to mitochondria. They further prioritized rare variants disrupting protein function. Their choice of mitochondrial nuclear genes was based on the MitoCarta inventory (Pagliarini *et al.*, 2008). The approach by Calvo and co-workers enabled the identification of novel nuclear-encoded mitochondrial genes NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3 (*NDUFB3*) and acylglycerol kinase (*AGK*), which previously were not linked to disease.

1.5. Neurological disorders: the importance of a diagnosis

Neurological disorders represent a vast group of diseases which manifest as a range of diverse phenotypes. Most neurological disorders are highly heterogeneous and individually very rare. Furthermore, many neurological diseases exhibit phenotypic overlap. This substantial clinical spectrum, together with immense genetic heterogeneity, makes precise molecular diagnosis difficult. For patients with neurological phenotypes, establishing the diagnosis involves stepwise testing, which is expensive and time consuming (Warman Chardon *et al.*, 2015). Many patients, who present with a rare neurological phenotype, may never receive a molecular diagnosis (Shashi *et al.*, 2014). Accurate and timely diagnosis is crucial because for some neurological diseases treatment could be available (Toscano and Schoser, 2013).

Correct molecular diagnosis could also assist in patient management and limit unnecessary and invasive tests. For example, identifying mutations in the *SLC2A1* gene in patients presented with Glucose transporter type 1 deficiency syndrome enables implementation of the ketogenic diet. The latter treatment result in reduction of seizure frequency for up to 50% (Klepper, 2015; Schoeler *et al.*, 2015). Precise molecular diagnosis of rare neurological disorders is vital because it would eventually define the pathways involved in the disease. This would potentially result in identifying new therapeutic targets. Exome sequencing is gradually establishing itself as an excellent tool to elucidate causative mutations in various neurological diseases.

1.6. Research project structure

The research presented in this study aims to evaluate the application of whole-exome sequencing for the molecular diagnosis of unexplained neurological disorders. General materials and methods (Chapter 2) are followed by the investigation of approaches to filtering of exome sequencing data from large cohorts (Chapter 3). To investigate the use of whole exome sequencing for the molecular diagnosis of neurological disorders we performed whole exome sequencing in a heterogeneous cohort of patients with suspected inherited ataxia as an example of a neurological disorder (Chapter 4). We then move on to performing molecular genetic analysis to determine pathogenicity of a heterozygous splice site mutation in *ZFYVE27* in an autosomal dominant pedigree (Chapter 5) and *TUBB4A* mutations in an autosomal dominant pedigree and a sporadic case (Chapter 6). Finally, we evaluate the pathogenicity of mutations in *ZFYVE26*, *FASTKD2* and *KCNB2*.

1.6.1. Research Aims

- To assess the use of whole exome sequencing for diagnostic purposes in neurological disorders
- To identify novel ataxia genes
- To investigate the pathogenicity of mutations in novel ataxia genes

Chapter 2

Materials and Methods

Chapter 2. Materials and Methods

2.1. Ataxia patient cohort

We have collected 22 randomly selected families of white European origin with unexplained ataxia and no known consanguinity. All subjects studied gave written informed consent. The cohort consisted of 35 affected individuals identified through routine referrals to the regional neurogenetics service at Newcastle upon Tyne. 25/35 had a family history, suggestive of autosomal dominant inheritance (11/25) or autosomal recessive inheritance (14/25). All treatable and most common inherited causes of ataxia were excluded on routine clinical investigations. Clinical presentation and all routine clinical investigations performed on these patients were previously published and are summarised in Table 4.5 (Pyle *et al.*, 2014).

2.1.2. Control group for case-control study

In order to exclude variants commonly found in specific ethnic groups this study collected DNA from 302 healthy individuals with the same ethnic background as the cohort presented here.

2.1.3. In-house control panel for exome sequencing

An in-house control panel of 286 exomes derived from unrelated individuals was used to filter for rare variants below a designated MAF (typically MAF<=0.01).

2.2. Patient DNA

Patients DNA was extracted from various tissues as described below. For some patients DNA samples were provided by Northern Genetics Service (The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne). All patient DNA samples were stored at -20^{o}C.

2.2.1. Genomic DNA extraction from blood

Genomic DNA was extracted from blood using a Nucleon® Genomic DNA Extraction Kit (Hologic Gen-Probe Incorporated, San Diego, USA) following the manufacturer's

protocol.

2.2.2. Genomic DNA extraction from Buccal swabs

Genomic DNA was extracted from Sterilin buccal swabs (Copan, Brescia, Italy) using QIAamp® DNA Mini Kit (QIAGEN, Manchester, UK) as per manufacturer's protocol.

2.2.3. Genomic DNA extraction from hair root cells

Genomic DNA was extracted from hair root cells using QIAamp® DNA Mini Kit (QIAGEN, Manchester, UK) following manufacturer's protocol.

2.2.4. Genomic DNA extraction from cultured fibroblasts

Genomic DNA was extracted from cultured fibroblasts using QIAamp® DNA Mini Kit (QIAGEN, Manchester, UK) as per manufacturer's protocol.

2.2.5. Genomic DNA amplification

In order to increase the amount of patient DNA available for sequencing studies, whole genome was amplified from patient DNA samples using a REPLI-g Mini Kit (QIAGEN, Manchester, UK) following the manufacturer's protocol.

2.3. Control DNA

Control DNA was extracted from anonymous blood samples from healthy individuals using a Nucleon® Genomic DNA Extraction Kit (Hologic Gen-Probe Incorporated, San Diego, USA) following the manufacturer's protocol.

2.4. Whole exome analysis

2.4.1. Whole exome sequencing

Whole-exome sequencing was performed on one or several individuals from each pedigree depending on the mode of inheritance of the disease. Whole-exome sequencing was outsourced to AROS (AROS Applied Biotechnology A/S, Aarhus,

Denmark). Genomic DNA was subjected to a library preparation using TruSeq™ DNA Sample Preparation Kit (Illumina Inc., San Diego, USA) and the targeted regions were captured using the TruSeq™ 62 Mb Exome Enrichment Kit (Illumina Inc., San Diego, USA). The captured fragments were sequenced on an Illumina HiSeq2000 platform (Illumina Inc., San Diego, USA) producing 100 bp paired-end reads.

2.4.2. Bioinformatics

Bioinformatic analysis was performed using an in-house algorithm incorporating the published tools. The following was performed by Dr. Helen Griffin (Newcastle University): the reads were aligned to the human reference genome (UCSChg19) using Burrows-Wheeler Aligner (Li and Durbin, 2010), PCR duplicates were removed with Picard v1.85 (available at http://broadinstitute.github.io/picard/), single base variants (SBV) and insertions/deletions (indels) were identified with Varscan v2.2 (Koboldt *et al.*, 2009) and Dindel v1.01 (Albers *et al.*, 2011) respectively..

2.5. Sanger sequencing validation of prospective variants

In order to verify that the gene variants identified by exome sequencing were not artefacts, Sanger sequencing was performed to confirm the mutations.

2.5.1. Primer design

Primer oligonucleotide sequences specific for the genes of interest were designed using Primer 3 (v.0.4.0) software (available at http://frodo.wi.mit.edu). Primer 3 specifies product size as well as melting temperature of the designed primers and their GC content (%). Target DNA sequences were uploaded into the online software and primer sequences selected to span the region of interest. The generated primer sequences were checked for specificity using the online program Basic Local Alignment Search Tool (BLAST) (available from http://blast.ncbi.nlm.nih.gov/Blast.cgi). This program is provided by the National Centre for Biotechnology Information (NCBI) (available from http://www.ncbi.nlm.nih.gov). BLAST identifies homologous regions in sequences by comparing nucleotide sequences to sequence databases and calculates statistical significance of matches. All primers used in this study to date are listed in Table 2.1.

Table 2.1: Primer sequences, annealing temperatures and amplicon sizes

Gene	Variant	Forward Primer Sequence 5'-3'	Reverse Primer Sequence 5'-3'	Anneal temp (˚C)	PCR product size (bp)
ABCB7		GCAGTGTGAAAGGCAG AGGAAAGT	TCTTTTGACCGAGATTT GCTGTTGTGT	65	400
AKAP4		GGGCTGCCCCATCGTTG CTA	GGAGGTGGCCAAAGTG CCAA	65	400
ATP1A3		CTGGCACCCCAGGCCTT CAC	CGCCACCTCGCAGGAC AAGA	65	394
ATXN7	1	TGGCTCACCATTCACCT GGGGG	CACACACACACACTCA TGCTCT	65	520
ATXN7	2	CCCTACCCCTCAACGGT AACA	CTTGTGCCCCTGGGAT GGTG	65	400
C12ORF65		GCATAATCTTGAGGGCA GATG	GGCCCAAGCCAGAAA AATA	65	486
CACNA1A		ACCACCCCCACTGCTTC CCA	ACCAATAGCAGGTACC CATTCCA	61, DMSO	346
CEP68		TCTCCCTTCCCCTGTTTT GTAGGA	GGCAGGCCACTGAGTG CAGG	65	376
CIT		CAGGCGGCTGTAGTCCC AG	GCAAACATCTCTGCCC TTGTGG	63	600
COL1CA1		ACAGTAAAAGCAGACA CAGGCATT	GGCCCAGCAGGAGCA AAGGG	65	395
CSTB		CTCTCACGGCCACAGCC CG	ACGATACCAGCCCCGG GAGG	65	499
DACH2		TTTTGAAATTAGAGGCA GCTGGAGAA	ACCCTGTCAAGGTACA GGAAATGACA	65	443
FASTKD2	1	GAAGCGCAGCTTCTCGG GGA	CGGCCGAACCCAGACC ATCA	65	373
FASTKD2	2	AGCCATTTGGAAGTGTT TCAGTGGAG	GTGCTTATGCCCTTTGT TTGAAAGCCA	65	274
FASTKD2	3, 4	CCCCTACAGGAAAACG ACAGCACG	AGGGCTGTTAGAGTGC TTATGCC	65	361
FASTKD2	5	TGAGCCTGGGAGGTTG AGGCT	TGGTACTGGAGGTCTT TGCAGGA	65	479
FASTKD2	6	ACCAACTGTCCAGAATG ATGTGTG	TCCACAGCAACACCTG CTCTGC	65	427
FASTKD2	7	ACATGAAGAATGCTTAC AAGCTGCATA	GGCCAGTTCCAGAGGC AATGCTG	65	338
FASTKD2	8	AGCTTTCTTAAACCCTA GCATGTAGTC	TGTGTGCTTTGCACATT TACAGCAG	65	296
HSPG2	1,2	CACCATCGGCTGAGTCG GGC	GTCCCCAGGATGTGGC ACAGG	65, DMSO	484
HSPG2	3	GCAGCGCCTGGAGACC TCTG	CCAGGGGTGGGGAGTA CGGG	65, Bet	470
HSPG2	4	GCCTTCGACCCACCCTC CCA	GGGAGACAGGTGCCCA CCCA	65 DMSO	458
HSPG2	5,6	CTGGGGCCAAAGGGGC AGAG	GCTTCCACCAGCCCTA CCAGC	65, DMSO	370
HSPG2	7	GCGCCACTTGTCTGTAC CCCC	GGCCACTGAGGAGGCA AGGC	65, Bet	295
HSPG2	8	GTGTGACGGGAAGGCG GTGG	GCTGGAGGGGCTGGAT GGGA	65, DMSO	366
HSPG2	9	TCGAGTGTTTGCCCTTC TGT	TAATATTGGCCCTGCA CCTG	65, DMSO	397
HSPG2	10	GGGCGTGAAGGGGGAG GGAA	TTGTGCCGAGCCCACA GCAG	65, DMSO	363

Table 2.1 continued

Gene	Variant	Forward Primer Sequence 5'-3'	Reverse Primer Sequence 5'-3'	Anneal temp (˚C)	PCR product size (bp)
HSPG2	11	CCCCACCCCTCTGCTCCCAG	TTCAGCCCCAGCCTGGCTCA	65, DMSO	333
IGDCC3	1	CCACATCTTTACACAGGAGGACCC	GTCCCCAGAGGCTGGGTGTT	65	364
IGDCC3	2	GGTGGGCAGGGGAGTGTGAA	GGAGTCCCAGTTCCACCACCA	65	579
JPH2		ACCCTCCACACTGTGCTCCA	CGATATCGACGCCACCACCA	58, DMSO	274
KCNA1		TTCTTCGCCTGCCCCAGCAA	ACACCACCGCCCACCAGAAA	65	388
KCNB2	1	TGCCTTCGCTCGAAGTATGGAACT	GGCCAGCTGCTCCTGTGGAC	65	394
KCNB2	2	CCAGCACAGCCAGGCCACTG	TCCTGCGGGGAGGAAGAGCC	65, DMSO	382
KCNC3		GGTCGTCATGGTGACCACAGCCCA	TGGAGGTGGAGACGGAGCCCT	65, DMSO	491
KCNH8		AGCACCTCTGTGTGTCCCTCC	GCTGAAATGGACTGCAGAGGCGT	65	350
KIAA0196		CTTAGGGTACTGACTCTTCCTTGGGA	AACATGGAGTGGGTATTGCCTTCT	58	444
LAMP3		GCAAGGCAACCTCCACTCAGACT	GGCCACACCCAACAACTCACACA	65, Bet	498
MICALCL	1	AACTGCTATTTTTGCCCATCCCA	TGGGGTGTGGGTCTCTGTGGT	65	299
MICALCL	2	TAGGGCAGCAGGAGAGCCCC	TGGGCAAACTGTGGGTGCGG	65	350
MLL2		CATGGTGAGCCCTGCCCTGC	CCAGCCTTGGAACCCAGTGCC	65	296
MYH14		CAGGCTCTGTGCTATGCGTGTG	TCCGCTCTCACCTACCTGGTCT	65	346
NEFH		AAATCCCCAGCCGAAGTC	TGGAGTCTTGGCTTCTGGAG	65	693
NPC1	1	GCTCTGACAAATGAAAGCTCAA	AGGACGAAGCAGCAAAACAT	65	494
NPC1	2	CCTGGTCTTAAGCAATTCTCTTG	TCTCTGTTCGTGCCCTTAGTTAG	65	447
OBFC2A		ATGGGACATAATCATTTGGGAAGTTTTTG	ACTCAGCACACCACAAAGTGCAA	65	447
SACS	1	ACTTGTTCAAGCTGTTCTTGTGGACC	CCATTCTTATGTCCTGAGCGGGC	65	247
SACS	2	GTGGTGAGCTGAGATTGTGCC	GTGTTCCGTGGAATATTCACTTCTCC	61, DMSO	390
SLC18A2		TGTGCGGTCTTGGACCCCCT	AACAGGCCAGCTCCGGGGAA	65	365
SLC1A3		TGCAGTTGGTAGATACGGCGAAC	ACATCTTGGTTTCACTGTCGATGGGT	65	328
SLC22A16		GTGTCCCAAGAAAACCAGATAGGAAA	AGCAGGAATAAGAGGGAGAACACA	65	300
SPG7		CGCGGATCACGCAGGCGCGG	TGCTACGCCGGAAACCGTTACTGGAT	61, DMSO	542
SPG7	1	CAAAGCCACCGTCACTTGT	GATCTTGGCTCACCACAACC	57, DMSO	483
SPG7	2	TGGTTTTTGTTTTGCTTTGG	CATCTCCCCATTTTCCAGAC	63	297

Table 2.1 continued

Gene	Variant	Forward Primer Sequence 5'-3'	Reverse Primer Sequence 5'-3'	Anneal temp (˚C)	PCR product size (bp)
SPG7	3	ACGAGGTAGACGGGCTCAG	CAGACGGGTTGGGAAAGTC	59	600
SRRM4	1	TCGGGAGGGATGGTGACCAGTG	ACCGCACCCAGGCTCTCCAG	65, DMSO	391
SRRM4	2	AGCCCCTCCTCATCCGGCAG	AAGAAGGCGGGGAGGGCAGG	67, Bet	348
TDP1		GCACAGCCAGCAGAATTATGGGC	GGGTATCATCTGCGGAATGCAGAGC	58	528
TNNT2		CACCCACAGCAGCTGGGAATCT	CAGCCCAGAATCAGGGTTTCCA	65	299
TPP1		TCCGACACAGCCTGCACCAG	ACAGACTGCCCCCTTCCCCA	65	385
TTBK2		ACAGGGACACACATGCATCAGGTT	TCGCAGGAGTCCCAGTGCCT	65, DMSO	363
TUBB4A	1	GATCCACTCCACGAAGTAGCTG	GAGGCACTCTACGACATCTGTTT	65	423
TUBB4A	2	GGTCATTCATGTTGCTCTCGGCCT	AACATGATGGCGGCGTGCGA	65	361
USH2A	1	TGGCTTGGATGGTGGTTGCC	AGGCCTCGTGAACTTCAATTGCT	65	388
USH2A	2	AGCCTTTCTGATGGGTTCTAAATGGA	ACGGGTTTTACCAGGAACAGTAGT	65	478
VARS2	1	CTCTTGGCAGGTTCCCCCTG	GCCCCCACTCCCCTCTCAC	66	300
VARS2	2	CACCTGTCAGCTGTTCTTCCTCAC	CCCAGAGACCAGGGCTGTCC	66	298
WNK1	1	TGCTGTTTCCCTTACTCCTCAGGTGG	GCAGACTATCCATTGGCTCTCATTGT	65	480
WNK1	2	ACTGGCCAATAGGAAGCCCAGA	CCCCAAAGATGGGGAAACTCTACTGA	65	400
ZFYVE26	3	TGCTCCCAGGTGGCTCTGGT	CCCAAAGAAGGTAGTTGGCACTTAGCA	65	448
ZFYVE26	4	TCCCTTCTCCCAGCTAAAATCTGGC	AGCCCACCCTTTGCCTTTGGT	65	398
ZFYVE27		CTTCAACCACTGTGCGATGT	GGCTCAACCAGAAACAGGAG	65	481
ZKSCAN4		GTCTTAAAATACAGCTCTCTGTGATGAGT	GGATCCGGGCCACATCTCCG	65	461

2.5.2. Standard polymerase chain reaction (PCR)

PCR was performed in a final reaction volume of 25 μL containing 1U Immolase™ DNA Polymerase (Bioline, London, UK), 0.2mM dNTPs (VHBio, Gateshead, UK), 1x ImmoBuffer (160mM $(NH_4)_2SO_4$, 1M Tris-HCl pH 8.3; Bioline, London, UK), 4mM $MgCl_2$ (Bioline, London, UK), 0.25μM of each forward and reverse primers and 1 μL of total genomic DNA extract containing on average 25ng DNA. Where GC-rich primers affected amplification efficiency, addition of 1M Betaine (Sigma-Aldrich, Dorset, UK) or 10% dimethylsulphoxide (DMSO) (Sigma-Aldrich, Dorset, UK) was required in order to reduce the formation of secondary structures in GC-rich regions (Sun *et al.*, 1993; Henke *et al.*, 1997). PCR reactions were performed in a Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK). PCR conditions involved a hot start at 95°C for 10 minutes, followed by 30 cycles of denaturation at 95°C for 1 minute, annealing at the relevant primer annealing temperature (Table 2.1) for 1 minute, extension at 72°C for 1 minute. The final extension step was performed at 72°C for 10 minutes, followed by cooling the reaction to 4°C. The PCR reactions were stored at -20°C.

2.5.3. Agarose gel electrophoresis

Agarose gel electrophoresis allows visualisation of PCR products and is used for determining the amount, purity and amplicon sizes of amplified DNA fragments. 2% agarose gel was prepared with molecular grade agarose (Helena Biosciences Europe, Gateshead, UK) and 1x TAE buffer (0.04M Tris acetate, 0.001M EDTA; Severn Biotech Ltd., Worcestershire, UK) and contained 0.4 μg/ml Ethidium Bromide (Sigma-Aldrich, Dorset, UK) to aid the visualisation of DNA. 5μL of each amplified PCR product was mixed with 5μL of Orange G loading buffer (50% glycerol (Sigma-Aldrich, Dorset, UK), 50% H_2O, Orange G (Sigma-Aldrich, Dorset, UK)), loaded onto the agarose gel, along with 5μL of Hyperladder IV (Bioline, London, UK) DNA marker. These were electrophoretically separated (Scie-Plas, Severn Biotech Ltd., Worcestershire, UK) in 1x TAE buffer at a constant voltage of 65 V. The PCR products were visualised by exposure to the UV light using GelDoc-It®2 Imager (UVP, LLC, Cambridge, UK)

2.5.4. Exonuclease I/FastAP treatment

The clean-up reaction removes unincorporated primers and degrades unincorporated nucleotides, which would otherwise interfere with the sequencing reaction. In order to clean-up the PCR product prior to sequencing, 10 U of Exonuclease I (Thermo Fisher Scientific Inc., Leicestershire, UK) and 1 U of Thermo Scientific FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific Inc., Leicestershire, UK) were mixed with 20 ng of the PCR amplicon in a 96-well plate (Greiner Bio-One Ltd., Gloucestershire, UK). The enzymatic reaction was performed in a Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK) and involved incubation at 37°C for 15 minutes, followed by enzyme inactivation at 85°C for 15 minutes.

2.5.5. BigDye® Terminator sequencing

Sequencing of the PCR amplicons was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies Ltd., Paisley, UK). This kit contains all reagents required for the sequencing reaction. Following the Exonuclease I/FastAP treatment, a mastermix of 1 μL of BigDye® Terminator v3.1 Ready Reaction Mix, 2 μL of BigDye® Terminator v1.1/v3.1 5X Sequencing Buffer and 0.5 μM of appropriate forward or reverse gene specific primer was added to the enzymatic reaction mix and made up to 20 μL with dH_2O. The BigDye® Terminator cycle sequencing was performed in a Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK). The reaction conditions involved an initial denaturing step at 96°C for 1 minute, followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes.

2.5.6. Ethanol/EDTA/sodium acetate precipitation

Unincorporated dye-labelled terminators in the sequencing reaction interfere with the downstream sequencing process and therefore should be completely removed prior to electrophoresis. In order to do this, ethanol-aided precipitation of the cycle sequencing product was performed. A mixture of 2 μL of 125mM EDTA (Sigma-Aldrich, Dorset, UK), 2 μL of 3M sodium acetate (Sigma-Aldrich, Dorset, UK), and 70 μL of 100% ethanol (VWR International Ltd., Leicestershire, UK) was added to each cycle sequencing reaction. After mixing and incubation for 15 minutes at room temperature,

the plate was centrifuged at 2000 g for 30 minutes. The supernatant was removed by inverting the plate on tissue and centrifuging briefly at 100 g. Following this, 70 µL of 70% ethanol (VWR International Ltd., Leicestershire, UK) was added and the sample centrifuged at 1650 g for 15 minutes. The supernatant was removed by inverting the plate on tissue and briefly centrifuging at 100 g. The sample was then air dried for 10 minutes in the dark to allow the evaporation of the remaining ethanol.

2.5.7. Capillary electrophoresis sequencing

The precipitated pellets were resuspended in 10 µL highly de-ionized formamide (Hi-Di™, Life Technologies Ltd., Paisley, UK) and incubated for 2 minutes at 95°C in a Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK). Following the incubation, the plate was briefly chilled on ice and then loaded into the ABI PRISM® 3100 Genetic Analyser (Life Technologies Ltd., Paisley, UK) to allow capillary electrophoresis sequencing.

2.5.8. Sequence analysis

The generated sequence data (electropherograms) were analysed using SeqScape® v2.6 Software (Life Technologies Ltd., Paisley, UK). This software allows mutation detection and analysis by comparing nucleotide sequences to manually uploaded chromosome reference sequences from Ensembl Genome Browser (available at http://www.ensembl.org/index.html).

2.6. Cloning from extracted PCR fragments and plasmid DNA sequencing

2.6.1. PCR and extraction of DNA fragments

PCR amplicons were generated as described in Section 2.5.2 and electrophoretically separated on 1-3% agarose gel as described in Section 2.5.3. The percentage of the agarose gel was dependent on the size of the amplicon of interest (i.e. 1% gels for 1kb amplicons) and the size difference between amplicons of interest (i.e. 3% gels for amplicons differing 15 bp in size). Following separation, the bands were excised from the gel and DNA was extracted using QIAquick® gel extraction Kit (QIAGEN, Manchester, UK) as per manufacturer's protocol. The elution of DNA from the column

was performed using 30μL of water.

2.6.2. Ligation

The pGEM®-T easy vector system (Promega, Southampton, UK) was used for ligation of the extracted DNA fragment. Ligation reaction was performed in a final reaction volume of 10μL containing 5ng of pGEM®-T easy vector, 3 Weiss units of T4 DNA ligase, approximately 30ng of DNA fragment preparation and 1x ligation buffer. The ligation was incubated at room temperature for at least 16 hours.

2.6.3. Transformation

5 μL of the ligation reaction was added to 50μL of JM109 competent cells (Promega, Southampton, UK) and incubated for 20 minutes on ice followed by heat shock at 42°C for 45 seconds. The cells were then put on ice and incubated for 2 minutes. 800 μL of sterile 2YT broth containing 16 g/L tryptone (Sigma-Aldrich, Dorset, UK), 10 g/L yeast extract (Sigma-Aldrich, Dorset, UK) and 5 g/L sodium chloride (Sigma-Aldrich, Dorset, UK) was added to the cells and incubated for 90 minutes at 37°C. The cells were plated out on 2YT agar plates containing 15g/L agar (Sigma-Aldrich, Dorset, UK) and supplemented with 100 μg/mL Ampicillin (Sigma-Aldrich, Dorset, UK), 0.5 mM IPTG (Promega, Southampton, UK) and 75 μg/mL X-GAL (Promega, Southampton, UK) to allow blue/white selection of recombinants. The plates were incubated overnight at 37°C. Successful incorporation of the DNA fragment interrupts the α-peptide coding region of β-galactosidase enabling identification of recombinants by white/blue colour screening on the plates.

2.6.4. Extraction of plasmid DNA (minipreps)

Single colonies were transferred from the agar plates into 15mL Falcon tubes (Greiner Bio-One, Stonehouse, UK) containing 5mL of sterile 2YT broth supplemented with 100 μg/mL Ampicillin (Sigma-Aldrich, Dorset, UK). The tubes were incubated overnight at 37°C in a rotary incubator. Following incubation the cultures were centrifuged at 3000 g for 10 minutes and the supernatant was discarded. Plasmid DNA was extracted using QIAprep® Spin Miniprep Kit (QIAGEN, Manchester, UK) as per manufacturer's protocol. DNA was eluted in 50μL of dH_2O. The plasmid DNA yield was determined using a NanoDrop 2000 Spectrophotometer (Fisher Scientific,

Leicestershire, UK).

2.6.5. Sequencing of plasmid DNA constructs

Sequencing of the plasmid DNA was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies Ltd., Paisley, UK). The cycle sequencing reaction was performed in a final reaction volume of 20 µL containing approximately 200 ng of plasmid DNA, 1 µL of BigDye® Terminator v3.1 Ready Reaction Mix, 2 µL of BigDye® Terminator v1.1/v3.1 5X Sequencing Buffer and 0.5µM of forward or reverse pUC/M13 sequencing primers (Promega, Southampton, UK) as outlined in Table 2.2. The BigDye® Terminator cycle sequencing was performed in a Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK) as described in Section 2.5.5.

Table 2.2: **pUC/M13 sequencing primers**

Forward primer	Reverse primer
5′-TCACACAGGAAACAGCTATGAC-3′	5′-CGCCAGGGTTTTCCCAGTCACGAC-3′

2.7. Segregation Analysis

If multiple affected and unaffected individuals' DNA was available from a pedigree, segregation analysis was performed. Each family member, where DNA was available, was analysed for the mutation of interest. The screening was performed using Sanger sequencing as described in Section 2.5. Segregation of the putative pathogenic variant with the disease supported pathogenicity of the mutation.

2.7.1. Pedigrees with family history

In the pedigrees with a dominant mode of inheritance we searched for genes with a single heterozygous variant which were shared between multiple affected and absent in unaffected individuals.

In the families with a suspected recessive mode of inheritance we searched for compound heterozygous variants shared between affected individuals. In the recessive families, where only one affected individual was available for the analysis, we sequenced DNA of unaffected parents, if available. In such cases we only considered

the compound heterozygous mutations in the affected individual if one of the variants was shared with the mother and the other one with father.

2.7.2. Pedigrees with no family history

For cases in which there was only one affected individual and no family history, we considered both single heterozygous variants (to fit a dominant inheritance model) and compound heterozygous variants (to fit a recessive mode of inheritance). Sequencing both unaffected parents, if available, enabled the discovery of *de novo* variants. In sporadic cases with a putative causative compound heterozygous mutation, where the parent's DNA was not available for screening, segregation of alleles was confirmed with sequencing of the cloned alleles. Compound heterozygous variants, which occurred *in trans,* supported a recessive mode of inheritance.

Cloning of genomic DNA for allelic cis-trans study

Cloning of genomic DNA from extracted PCR fragments was performed as described in Section 2.6.

Sanger sequencing validation of plasmid DNA clones for allelic cis-trans study

Plasmid DNA Sanger sequencing was performed as outlined in Section 2.6.

2.8. Cell culture

Control fibroblasts were obtained from the Newcastle BioBank (Newcastle University) from anonymous healthy individuals. Patient's fibroblasts were obtained from the Newcastle BioBank (Newcastle University). Informed consent was obtained from all probands before skin biopsies were performed.

The cells were incubated in in a humidified incubator at 37°C with an atmosphere of 5% CO_2, in minimum essential media (MEM) (Life Technologies Ltd., Paisley, UK) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich, Dorset, UK), 2mM L-glutamine (Life Technologies Ltd., Paisley, UK), 50 units/mL penicillin (Life Technologies Ltd., Paisley, UK), 50µg/mL streptomycin (Life Technologies Ltd., Paisley, UK), 110 mg/L sodium pyruvate (Sigma-Aldrich, Dorset, UK), Minimum Essential Medium vitamins at 1x (Life Technologies Ltd., Paisley, UK), non-essential amino acids at 1x (Life Technologies Ltd., Paisley, UK), and 50mg/L uridine (Life

Technologies Ltd., Paisley, UK).

2.8.1. Media change and passaging

The cells were initially grown on the surface of conventional tissue culture plastic 25cm^2 flasks (Greiner Bio-One, Stonehouse, UK) and then transferred into 75cm^2 flasks (Greiner Bio-One, Stonehouse, UK). Media was changed every 48 hours. Cells were passaged every 2-3 days when 80% confluent. In order to split cells, they were detached from the flask surface using trypsin-EDTA solution (Life Technologies Ltd., Paisley, UK) at 1x and incubated for 5 minutes at 37°C, followed by neutralization with the growth media. Cells were harvested by centrifugation at 200 x g for 5 minutes. The pellet was washed in Oxoid™ Phosphate Buffered Saline (PBS) (Thermo Fisher Scientific Inc., Leicestershire, UK), resuspended in growth media and seeded.

2.8.2. Freezing

In order to freeze cells, they were removed from the flask surface and pelleted, as described above, before resuspending in freezing media containing 90% FBS and 10% DMSO (Sigma-Aldrich, Dorset, UK). Cells were transferred to a -80°C freezer in an isopropanol filled container enabling slow cooling. After a minimum of 24 hours cells were transferred to liquid nitrogen for long-term storage.

2.8.3. Thawing

A cryovial, containing the relevant frozen cells, was removed from liquid nitrogen, thawed quickly in a 37°C water bath and transferred into a 75cm^2 flask with 11mL of supplemented growth media as described above. After pelleting at 200 x g for 5 minutes cells were washed in Oxoid™ Phosphate Buffered Saline (Thermo Fisher Scientific Inc., Leicestershire, UK), resuspended in growth media and seeded.

2.9. Sanger sequencing of complementary DNA (cDNA)

2.9.1. Total RNA extraction from cultured fibroblasts

Cultured fibroblast cells were harvested as described in Section 2.8 and the pelleted cells were washed in Oxoid™ PBS (Thermo Fisher Scientific Inc., Leicestershire, UK). Total RNA was extracted from the pelleted cultured cells using RNeasy® Mini Kit

(QIAGEN, Manchester, UK) as per manufacturer's protocol. The total RNA yield was determined using a NanoDrop 2000 Spectrophotometer (Fisher Scientific, Leicestershire, UK).

2.9.2. Generating first strand DNA (RT step)

Single-strand copy DNA (cDNA) was generated from total RNA extract using High Capacity cDNA Reverse Transcription (RT) Kit (Life Technologies Ltd., Paisley, UK) as per manufacturer's protocol. This Kit contains all reagents needed for reverse transcription of total RNA to single-stranded cDNA. The reverse transcription step was performed in a final reaction volume of 20µL containing 1x RT buffer, 4mM dNTPs, 50U MultiScribe™ Reverse transcriptase, 1x RT Random Primer and 10µL of total RNA extract containing approximately 1 µg of total RNA. RT reactions were carried out in Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK). Reverse transcription conditions involved annealing at 25^{o}C for 10 minutes, followed by cDNA synthesis at 37^{o}C for 120 minutes and termination of reaction at 85^{o}C for 5 minutes. The RT reactions were stored at -20^{o}C

2.9.3. Reverse transcription PCR (RT-PCR)

The first-strand cDNA obtained at the RT step was amplified directly with PCR. PCR was performed in a final reaction volume of 25µL containing 2.5µL of the first-strand reaction, 2.5U GoTaq® DNA Polymerase (Promega, Southampton, UK), 0.2mM dNTPs (VHBio, Gateshead, UK), 1 x GoTaq® Reaction Buffer containing $MgCl_2$ at a concentration of 7.5mM for a final concentration of 1.5mM in the 1x reaction; pH 8.5 (Promega, Southampton, UK) and 0.2µM of each forward and reverse primer. Where GC-rich primers affected amplification efficiency, addition of 10% DMSO (Sigma-Aldrich, Dorset, UK) was required in order to reduce the formation of secondary structure in GC-reach regions (Sun *et al.*, 1993). PCR reactions were performed in a Veriti® 96-well Thermal Cycler (Life Technologies Ltd., Paisley, UK). PCR conditions involved a hot start at 95^{o}C for 4 minutes, followed by 30-40 cycles of denaturation at 95^{o}C for 30 seconds, annealing at primer annealing temperature for 30 seconds, and extension at 72^{o}C. The final extension step was performed at 72^{o}C for 10 minutes, followed by cooling the sample to 4^{o}C. The PCR reactions were stored at -20^{o}C.

2.9.4. Cloning and sequencing of RT-PCR fragments

DNA fragments generated in RT-PCR step (Section 2.9.3) were extracted from an agarose gel as described in Section 2.6. DNA fragments were cloned and sequenced as described in Section 2.6.

2.10. Microscopy

2.10.1. Inverted microscopy for tissue culture samples

Images of cells were taken using DS-Fi2 camera mounted on inverted microscope Nikon Eclipse TS100 (Nikon UK Limited, Surrey, UK) and NIS-Elements D-3.2 microscope imaging software (Nikon UK Limited, Surrey, UK).

2.10.2. Upright microscopy for IHC

Images were captured using AxioCam HRc camera (Zeiss, Cambridge, UK) mounted on upright microscope Axio plan 2 (Zeiss, Cambridge, UK) and Axiovision microscope imaging software v.4.6.3 (Zeiss, Cambridge, UK).

Chapter 3

Approaches to filtering exome data from large cohorts

Chapter 3. Approaches to filtering exome data from large cohorts

3.1. Introduction

3.1.1. Principle of filtering

Whole exome sequencing produces about 10 GB of final data and it is a real challenge to distinguish novel variants for Mendelian or complex traits against a background of non-pathogenic polymorphisms and sequencing errors. Moreover, a recent study of 2,440 exomes of European and African ancestry demonstrated that out of 13,595 single nucleotide variants (SNVs), which each person carried on average, around 2.3% were predicted to alter protein function of approximately 313 genes per genome. Also, 95.7% of SNVs, predicted to be functionally significant, were rare (Tennessen *et al.*, 2012). In a similar independent study Nelson and co-workers sequenced 202 genes encoding drug targets in 14,002 individuals. They found that rare variants, many of which are deleterious, are abundant and geographically clustered (Nelson *et al.*, 2012). These studies hypothesised that an excess of rare deleterious variants is due to the recent explosion of population growth and weak purifying selection. These findings highlight potential difficulties researchers might confront when interpreting exome sequencing data. In order to prioritize candidate genes and to reduce their number, step-wise filtering is applied.

The main idea behind the filtering process is that in order to find rare or novel alleles, which are shared between affected individuals in one pedigree or several unrelated pedigrees, exome sequences of a small number of affected individuals are compared with a set of polymorphisms that are available in public databases (Ng *et al.*, 2010). The majority of exome sequencing studies use public databases such as dbSNP (Sherry *et al.*, 2001) (available at http://www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes Project (Abecasis *et al.*, 2010) (April 2012 release), and Exome Variant Server (May 2012 release) (the latest release available at http://evs.gs.washington.edu/EVS/). The novelty of the variant is based on the assumption that, being rare, this variant will be absent in public databases (Bamshad *et al.*, 2011). Rare variants might be present in those databases, albeit at a very low frequency (normally not more than 1% of the population).

3.1.2. Filtering steps

It is suggested that there is no optimal universal filtering pipeline and variant cut-off. This is because characteristics of functional variants will differ between genes, depending on the importance of the gene for the effective functioning of the human species and lack of purifying selection (Do *et al.*, 2012).

Functional annotation of all detected variants using ANNOVAR

In order to identify causal mutations, bioinformatic analysis, which utilises online tools, is performed. ANNOVAR is a tool which is used for functional annotation of identified novel variants (Wang *et al.*, 2010). ANNOVAR incorporates algorithms allowing the identification of protein-altering mutations. ANNOVAR also enables filtering out common variants that are present in dbSNP137 database (Sherry *et al.*, 2001), 1000 Genomes Project (Abecasis *et al.*, 2010) and Exome Variant Server (the latest release available at http://evs.gs.washington.edu/EVS/). Provided with the list of variants and their chromosomal positions, obtained from exome sequencing raw data, ANNOVAR is capable of creating a user-friendly output in the form of an Excel spread sheet. The output data created by ANNOVAR is arranged in columns, which contain information such as protein coding change, gene annotation and variant pathogenicity scores calculated by various pathogenicity prediction programs (Table 3.1).

Table 3.1: An example of ANNOVAR output for patients P23 and P24. This table depicts the output data created by ANNOVAR. In-house bioinformatic pipeline was modified in order to include an additional column containing MAF values for a control panel of 286 unrelated in-house exomes ("In-house MAF 286").

Chromosome	Position	Ref (R)	Var (V)	Gene	Associated GO Terms	OMIM	In House MAF 286	ESP6500	1000g 2012feb	dbSNP135	Func	Gene	Exonic Func	AA Change	PhyloP score	PhyloP Pred	SIFT score	SIFT Pred	Poly Phen2 score	Poly Phen2 Pred	LRT score	LRT Pred	Mutation Taster score	Mutation Taster Pred	GERP++	P23	P24
chr1	69428	T	G	OR4F5	integral to membrane, G-protein coupled receptor activity, plasma membrane, G-protein coupled receptor signaling pathway, olfactory receptor activity		0.04020979	0.030647		rs140739101	exonic	ENSG00000186092	nonsynonSNV	ENST00000335137:c.T338G:p.F113C	0.962332	C	1	D	0.98	D	0.994909	N	0.433294	N	0.499	R/V	V/V

Functional data for each variant are presented in columns “Function”, “Gene” and “Exonic variant function” (Table 3.1). Column “Function” contains values that specify whether the variant is positioned in the exon or intron or hits a region within a splicing junction etc. Possible values in column “Function” are summarized in Table 3.2. Column “Gene” provides the name of a gene where the variant resides. Column “Exonic variant function” shows functional consequences of the amino acid change as a result of the exonic variant. Only exonic and splice site (those situated ± 2bp from an exon boundary) variants are annotated in this column. All possible annotation values and their explanation are given in Table 3.3.

Table 3.2: Possible values in column "Function" and their explanations presented in ANNOVAR output (adapted from ANNOVAR website, available at http://annovar.openbioinformatics.org)

Value	Explanation
exonic	variant overlaps a coding region
splicing	variant is within 2-bp of a splicing junction
ncRNA	variant overlaps a transcript without coding annotation in the gene definition
UTR5	variant overlaps a 5' untranslated region
UTR3	variant overlaps a 3' untranslated region
intronic	variant overlaps an intron
upstream	variant overlaps 1-kb region upstream of transcription start site
downstream	variant overlaps 1-kb region downtream of transcription end site
intergenic	variant is in intergenic region

Table 3.3: Possible annotation values in column "Exonic variant function" and their explanations presented in ANNOVAR output (adapted from ANNOVAR website, available at http://annovar.openbioinformatics.org)

Annotation	Explanation
frameshift insertion	an insertion of one or more nucleotides that cause frameshift changes in protein coding sequence
frameshift deletion	a deletion of one or more nucleotides that cause frameshift changes in protein coding sequence
frameshift block substitution	a block substitution of one or more nucleotides that cause frameshift changes in protein coding sequence
stopgain	a nonsynonymous SNV, frameshift insertion/deletion, nonframeshift insertion/deletion or block substitution that lead to the immediate creation of stop codon at the variant site.
stoploss	a nonsynonymous SNV, frameshift insertion/deletion, nonframeshift insertion/deletion or block substitution that lead to the immediate elimination of stop codon at the variant site
nonframeshift insertion	an insertion of 3 or multiples of 3 nucleotides that do not cause frameshift changes in protein coding sequence
nonframeshift deletion	a deletion of 3 or multiples of 3 nucleotides that do not cause frameshift changes in protein coding sequence
nonframeshift block substitution	a block substitution of one or more nucleotides that do not cause frameshift changes in protein coding sequence
nonsynonymous SNV	a single nucleotide change that cause an amino acid change
synonymous SNV	a single nucleotide change that does not cause an amino acid change
unknown	unknown function (due to various errors in the gene structure definition in the database file)

ANNOVAR output also incorporates data from Online Mendelian Inheritance in Man (OMIM®, available at https://omim.org) and Gene Ontology (GO, available at http://www.geneontology.org), which provide data about gene associated diseases and gene product's biological properties, respectively. The latter two databases are linked by ANNOVAR to the identified variants enabling a quick genotype/phenotype interrogation.

Variant pathogenicity scores predicted by various pathogenicity prediction programs are presented in the ANNOVAR output in separate columns. Pathogenicity prediction programs that were incorporated in ANNOVAR at the time of writing this Thesis are outlined below.

- MutationTaster (Schwarz *et al.*, 2010) makes predictions on the basis of evolutionary conservation, loss of protein function and splice site changes (scores between 0 and 1 with the larger scores suggesting a higher probability to cause a disease).
- Sorting Intolerant from Tolerant (SIFT) algorithm (Kumar *et al.*, 2009) predicts the effect of the amino acid change (scores between 0 and 1 reflect the probability of being pathogenic: 0 = highest; 1 = lowest).
- Polymorphism Phenotyping 2 (PolyPhen2) (Adzhubei *et al.*, 2010) predicts the effect of the amino acid change (scores between 0 and 1 with the larger scores suggesting a higher probability to be damaging).
- Likelihood Ratio Test (LRT) method (Knudsen and Miyamoto, 2001) predicts pathogenicity of variants based on conservation of protein residues (scores 0 to 1 high score indicating a large deleterious effect).
- Phylogenetic p-values (PhyloP) (Pollard *et al.*, 2010) algorithm estimates the probability of observed variation under the hypothesis of neutral evolutionary rate and assigns conservation p-values (Conserved =>0.95, Non-Conserved <0.95).
- Genomic Evolutionary Rate Profiling (GERP) (Cooper *et al.*, 2005) measures base conservation via estimating probability of the observed substitution under

the hypothesis of neutral evolutionary rate (range from -12.3 to 6.17, with 6.17 being the most conserved).

Most of the programs outlined above use conservation of the affected residue and information about structure of the protein for prediction. Although using prediction programs may be successful in some cases (Brunham *et al.*, 2005), one should remember, that these programs might not be sensitive enough in other cases and thus should be used with great caution and only as a guide (Flanagan *et al.*, 2010; Ng *et al.*, 2010). Also, a poor agreement among different programs has been reported (Tennessen *et al.*, 2012).

Focusing on deleterious variants

The majority of exome sequencing studies only focus on deleterious variants such as exonic, protein altering and splice site alterations. The latter is defined as within 2 bp of the intron-exon boundary (Calvo *et al.*, 2010). Such approach is justified because the majority of protein altering variants are predicted to affect protein function and to be deleterious (Kryukov *et al.*, 2007). Also, most of the variants known to cause Mendelian disorders disrupt protein-coding regions of the genome (Stenson *et al.*, 2009).

The range of pathogenicity predicting programs is being used to predict the pathogenicity of variants, as outlined above.

Setting Minor Allele Frequency (MAF) restrictions

Minor allele frequency (MAF) is the frequency at which the less common allele occurs in a population. If, for example, the minor allele occurs in 5% of a given population, then this would mean that the MAF value for this allele is 0.05. Alleles with a MAF less than 0.01 are considered rare (Bamshad *et al.*, 2011). In order to filter out putative causative variants, MAF restrictions should be clearly set for the exome sequencing study.

Previously many groups ruled out any variant found in public databases such as dbSNP (Sherry *et al.*, 2001), 1000 Genomes Project (Abecasis *et al.*, 2010), and in-house exome databases. This is based on the assumption that, because causative variants are

very rare, any individual, who has a rare variant, will be affected. However, this is true only for fully penetrant variants (Brunham and Hayden, 2013). Also, as more whole exome and whole genome studies are undertaken, more data will be deposited in the public databases. This means that with time more rare variants will be accounted for and excluding them from sequencing studies could be very restrictive (Bamshad *et al.*, 2011). That is why many groups allow the minor allele frequency for variants found in public databases to be as high as 0.01 (McDonald *et al.*, 2012).

Consider mode of inheritance

A filtering workflow, used to prioritise putative causative gene variants in a pedigree, depends on mode of inheritance. For example, for dominant diseases only single heterozygous variants, which are shared between affected members, are considered. For recessive pedigrees, homozygous (for consanguineous families) or compound heterozygous (for non-consanguineous families) variants are considered. However, such approach should be used with a caution because an apparent homozygosity in a non-consanguineous recessive family could, for example, indicate a large genomic deletion on the other chromosome (Pyle *et al.*, 2013). For sporadic cases with no family history both single heterozygous and compound heterozygous mutations are considered in order to fit presumed dominant or recessive models, respectively. For suspected *de novo* mutations parent-child trios should be filtered for mutations absent in both parents (Ku *et al.*, 2012).

Select variants based on a gene or gene product function

Candidate gene variants can be prioritised based on a corresponding gene or protein function. This approach has been shown to be successful for some disorders. For example, in exome sequencing studies of mitochondrial disorders the genes could be stratified by their involvement with mitochondria (Calvo *et al.*, 2012). Recent studies suggest that using a list of key words may also be beneficial in prioritising putative causative mutations (Fogel *et al.*, 2014). Key words could be based on the best definition of the phenotype of the patient and should be used to search additional databases that are incorporated into ANNOVAR. For example, in a recent exome sequencing study of a highly heterogeneous cohort of ataxia patients the authors successfully used the key word “ataxia” to filter in variants in known ataxia genes

(Fogel *et al.*, 2014). The latest literature proposes that when prioritising the filtered variants, researchers should start from focusing on known or novel variants in known disease genes, followed by considering novel rare variants in "likely novel candidate genes" (Gonzaga-Jauregui *et al.*, 2015). It is also recommended that in order to link genomic data with a clinical phenotype, the list of filtered out genes should be reviewed by at least one neurologist. This would allow a review of the variants based on the latest literature and diagnostic criteria (Fogel *et al.*, 2014).

Average coverage of consensus coding sequence (CCDS) bases

The quality of exome sequencing data could be assessed by coverage of CCDS bases. In exome sequencing, sequence coverage (depth) is the average number of times the base is covered in a given experiment (Rizzo and Buck, 2012). Low coverage could result in misinterpreting the sequencing data and could result in potentially missing causative variants (Rehm *et al.*, 2013). Low coverage could also lead to a false-positive or false-negative variant call because variant calling algorithms take into consideration the number of times the base is seen as a reference as opposed to being seen as a variant. It is anticipated that it is impossible to impose a CCDS base coverage threshold, which would be universal for all laboratories. This is because the coverage would differ amongst different sequencing platforms, enrichment kits and analytical algorithms used by the laboratories for acquisition and interpretation of exome sequencing data (Rehm *et al.*, 2013).

Coverage analysis of genes previously linked to the disease

As recommended by the recent guidance for interrogation of putative causative variants in human disease (MacArthur *et al.*, 2014), the study on ascribing pathogenicity to the prospective variants should start from investigating genes, which were previously linked to that phenotype. Therefore it is necessary to assess the coverage of CCDS bases in those known disease genes along with the coverage of total CCDS bases. The list of genes for such study could be generated based on the latest published data. For example, gene panels designed for targeted exome sequencing in mitochondrial disease (Calvo *et al.*, 2012) and ataxia (Nemeth *et al.*, 2013) could be used in patients with these phenotypes.

3.2. Materials and methods

3.2.1. Functional annotation of variants using ANNOVAR

The study cohort consisted of 35 affected individuals (P1-P35) from 22 randomly selected families of white European descent with no known consanguinity. Whole exome sequencing was performed on 30 patients enrolled in this study In-house bioinformatic analysis was based on previously published tools and performed, as described in Section 2.4.2. All variants were functionally annotated by Dr. Helen Griffin (Newcastle University) using ANNOVAR (Wang *et al.*, 2010), which enabled identification of protein altering mutations along with their functional annotation. The ANNOVAR output file was presented as an Excel spread sheet containing a list of variants for each proband along with columns containing values for each variant. The values were as follows: chromosome position of the variants, reference and variant sequences, name of the gene in which the variant was recorded, OMIM® and Gene Ontology data linked to each variant, MAF value for each identified variant recorded in dbSNP137, 1000 genomes (April 2012 data release) and Exome Variant Server NHLBI-ESP6500 databases (May 2012 release). The in-house pipeline was modified in order to include MAF values for a control panel of 286 unrelated in-house exomes. The ANNOVAR output file also incorporated pathogenicity prediction programs presented in separate columns, which ascribed each variant a pathogenicity score (as described in Section 3.1.2). Finally, the ANNOVAR output file contained functional data for each variant presented in columns "Function" and "Gene" (as described in detail in Section 3.1.2).

3.2.2. Filtering for deleterious variants

In order to identify causal mutations, the list of variants annotated with ANNOVAR (as described in Section 3.2.1) was obtained from Dr. Helen Griffin (Newcastle University) and further manually filtered as outlined in Figure 3.1. This study only focused on deleterious variants (protein altering variants, nonsense mutations, exonic insertions/deletions (indels) and splice site variants) and excluded all synonymous variants. A range of bioinformatic programs for predicting pathogenicity of variants was

employed (MutationTaster (Schwarz *et al.*, 2010), SIFT (Kumar *et al.*, 2009), PolyPhen (Adzhubei *et al.*, 2010), PhyloP (Pollard *et al.*, 2010), LRT (Knudsen and Miyamoto, 2001), GERP (Cooper *et al.*, 2005)). However, due to the possible disagreement between these programs (Tennessen *et al.*, 2012) the current study used them with caution and only as a guide.

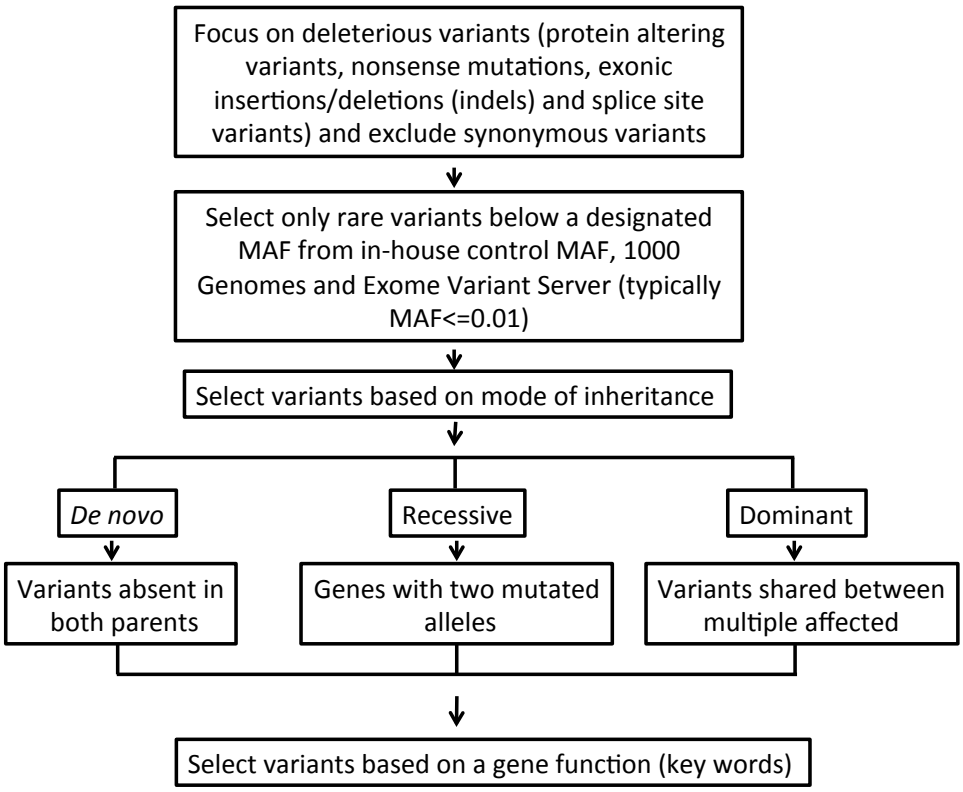


Figure 3.1: Overview of the filtering process to prioritise putative causative gene variants.

3.2.3. Setting MAF cut off

Minor allele frequency (MAF) cut off was set at 1% in dbSNP137, 1000 genomes (April 2012 data release) and Exome Variant Server NHLBI-ESP6500 databases (May 2012 release) as well as in 286 unrelated in-house controls (Figure 3.1). For families with a suspected recessive mode of inheritance or sporadic cases we set up a MAF cut off at <0.05 to accommodate for compound heterozygous mutations. Variants with MAF values above this cut off were excluded.

3.2.4. Selection of variants based on mode of inheritance

Selection of the remaining variants was based on mode of inheritance, as reflected in (Figure 3.1). Depending on the mode of inheritance we filtered for genes with compound heterozygous mutations (for suspected autosomal recessive and sporadic cases), genes with a single heterozygous mutation shared between multiple affected family members (for autosomal dominant traits) or variants absent in both parents of the affected individual (for suspected *de novo* cases).

3.2.5. Selection of variants based on gene function (using key words)

The variants were prioritized for brain and nerve cell expression and for connection with known neurological disease, which could fit the phenotype in the subjects being studied (Figure 3.1). Following the suggestions in the latest guidance for investigating putative causative variants in human disease (MacArthur *et al.*, 2014), this study started exploring variants in genes known to be associated with ataxia. For this key words were used (*cerebellum, ataxia*) and prioritized genes based on clinical information in OMIM® and Gene Ontology as described in Section 3.2.1. The filtered in variants were cross-referenced with the latest literature in order to identify known and novel variants in genes previously associated with ataxia or related neurological diseases. The variants were considered sequentially as follows:

- known variants in known ataxia and related neurological diseases genes;
- novel variants in known ataxia and related neurological diseases genes;

Following this, the current study investigated rare variants in genes linked to neurological disorders that could fit the phenotype in the affected individuals. Such genes are potentially novel genes associated with ataxia. This study further prioritised genes matching OMIM® and Gene Ontology terms *brain, neuro, nerve* (to aid in prioritizing genes expressed in the tissue relevant to the ataxia phenotype); *spastic*, *cephal, lysosome* (to prioritize genes linked to diseases with possible phenotype overlap). In order to determine the possible link between the variants and the patient's phenotype, two neurologists reviewed the list of putative causative mutations.

3.2.6. Average coverage of consensus coding sequence (CCDS) bases

Coverage was calculated for all Consensus Coding Sequence (CCDS) bases (31,935,069 bp), by Dr. Helen Griffin (Newcastle University), using custom scripts and following a protocol outlined below in Section 3.2.7. The percentage of bases that mapped to target regions (Illumina Truseq 62Mb targets +/-500bp) was also calculated (mean target base coverage).

3.2.7. Coverage analysis of ataxia genes from recent publications

In order to check if all known ataxia genes were covered in this study's patients by exome sequencing, a recently published list of genes used for a targeted capture in a similar ataxia cohort study was employed (Nemeth *et al.*, 2013). This list (Table 3.4) consisted of 118 genes and included known ataxia genes (42 genes, 786 exons), genes causing ataxia in mouse models (18 genes, 514 exons), transfer RNA synthetase genes (35 genes, 580 exons), genes associated with other neurological disorders with ataxia as a part of phenotype (6 genes, 79 exons), speculative genes based on function (7 genes, 167 exons) and genes associated with Joubert syndrome (10 genes, 243 exons). The latter is a specific ataxia disorder. Coverage was calculated for CCDS bases for genes studied by Nemeth and colleagues (386,444 bp). Coverage analysis was performed by Dr. Helen Griffin (Newcastle University) using custom scripts. For this the list of genes (Table 3.4) was uploaded by Dr. Helen Griffin (Newcastle University) into online formatting software SAMtools (available at http://samtools.sourceforge.net). SAMtools software allows alignment of sequence reads against large reference sequences, including the human genome, in a common alignment format (Li *et al.*, 2009). Per base files of coverage depth (pileup files) were created for each patient using files of aligned sequence reads and the 'mpileup' command in SAMtools formatting software. Custom Perl scripts were then used to calculate the mean per target base read depth and the number of target bases which were covered to a minimum depth of 20-, 10-, 5- and 1-fold. This was done for consensus coding sequence (CCDS) target bases and also specifically for all coding bases of genes selected for targeted capture by Nemeth and colleagues.

Table 3.4: List of genes selected for targeted capture by Nemeth and colleagues (taken from Nemeth *et al.,* 2013)

Autosomal Dominant	Autosomal Recessive	Ataxia in mouse models	tRNA synthetase genes	Other disorders associated with Ataxia	Speculative	Joubert
ATXN1	*APTX*	*AARS*	*AARS2*	*PRPS1*	*AIFM1*	*AHI1*
ATXN2	*SACS*	*CHD6*	*CARS*	*GAN*	*EN1*	*ARL13B*
ATXN3	*ABCB7*	*A2BP1*	*CARS2*	*SLC6A19*	*EN2*	*CC2D2A*
PLEKHG4	*ATM*	*ATP2B2*	*DARS*	*TINF2*	*KLHL1*	*CEP290*
SPTBN2	*MRE11A*	*AGTPBP1*	*DARS2*	*PSAP*	*RAB3A*	*INPP5E*
CACNA1A	*TTPA*	*DNAJC19*	*EARS2*	*SCN1A*	*SYNE2*	*NPHP1*
ATXN7	*ATCAY*	*DST*	*EPRS*		*EEF2*	*OFD1*
ATXN8OS	*FXN*	*GRID2*	*FARS2*			*RPGRIP1L*
ATXN10	*SIL1*	*HERC1*	*FARSA*			*TMEM216*
TTBK2	*TDP1*	*KCNJ6*	*FARSB*			*TMEM67*
KCNC3	*CABC1*	*MYO5A*	*GARS*			
PRKCG	*SETX*	*RELN*	*HARS*			
ITPR1	*SYNE1*	*RORA*	*HARS2*			
TBP	*MARS2*	*SLC12A6*	*IARS*			
FGF14	*GRM1*	*SNAP25*	*IARS2*			
AFG3L2	*KIAA0226*	*TRPC3*	*KARS*			
BEAN1	*DARS2*	*USP14*	*LARS*			
KCNA1	*KCNJ10*	*AFF1*	*LARS2*			
CACNB4	*AAAS*		*MARS*			
SLC1A3	*VLDLR*		*NARS*			
ATN1			*NARS2*			
PPP2R2B			*PARS2*			
			QARS			
			RARS			
			RARS2			
			SARS			
			SARS2			
			TARS			
			TARS2			
			VARS			
			VARS2			
			WARS			
			WARS2			
			YARS			
			YARS2			

3.2.8. Sanger sequencing validation and segregation study

In order to confirm putative pathogenic variants, Sanger sequencing was performed as described in Section 2.5. Where possible, multiple affected and unaffected family members were Sanger sequenced in order to segregate putative pathogenic mutations, as described in Section 2.7.

3.3. Results

3.3.1. Coverage and depth statistics

Coverage and depth statistics analysis for 30 patients was performed for CCDS bases, as described in Section 3.2.6, and the result was previously published (Pyle *et al.*, 2014). The mean target base coverage for all exomes was 72-fold. On average, 85% of all CCDS bases were covered 20-fold, 91% of all CCDS bases were covered 10-fold and 93% of all CCDS bases were covered 5-fold. Coverage and depth statistics for each patient is presented in Table 3.5.

Table 3.5: Coverage and depth statistics. Coverage calculated for Consensus Coding Sequence (CCDS) bases (31,935,069 bp). Target bases - bases 'on-target' to Illumina Truseq 62Mb targets +/-500bp. Adapted from Pyle *et al*., 2014.

Patient ID	Mean target base coverage	Number CCDS bases covered 20-fold	% CCDS bases covered 20-fold	Number CCDS bases covered 10-fold	% CCDS bases covered 10-fold	Number CCDS bases covered 5-fold	% CCDS bases covered 5-fold	Number CCDS bases covered 1-fold	% CCDS bases covered 1-fold
P1	51.2	19160285	60.0	22429550	70.2	24722918	77.4	27701324	86.7
P2	43.3	18655773	58.4	22188163	69.5	24570241	76.9	27574558	86.3
P3	48.2	19320591	60.5	22741113	71.2	25017636	78.3	27836153	87.2
P5	72.7	29223416	91.5	30558212	95.7	30976157	97.0	31336961	98.1
P7	40.5	24737579	77.5	29114027	91.2	30340208	95.0	31098022	97.4
P9	63.1	28163568	88.2	30066723	94.1	30729768	96.2	31281441	98.0
P10	83.7	29405477	92.1	30517261	95.6	30925225	96.8	31328606	98.1
P11	80.8	29275024	91.7	30463183	95.4	30869266	96.7	31280700	98.0
P12	80.9	29290200	91.7	30453008	95.4	30862332	96.6	31275473	97.9
P13	85.2	29419113	92.1	30519156	95.6	30890197	96.7	31289682	98.0
P14	65.0	28821117	90.2	30403109	95.2	30868327	96.7	31268983	97.9
P15	74.4	29233108	91.5	30482929	95.5	30895299	96.7	31285248	98.0
P16	111.8	29446426	92.2	30124845	94.3	30494261	95.5	31023782	97.1
P17	111.2	29568733	92.6	30318016	94.9	30687574	96.1	31190212	97.7

Table 3.5: continued

Patient ID	Mean target base coverage	Number CCDS bases covered 20-fold	% CCDS bases covered 20-fold	Number CCDS bases covered 10-fold	% CCDS bases covered 10-fold	Number CCDS bases covered 5-fold	% CCDS bases covered 5-fold	Number CCDS bases covered 1-fold	% CCDS bases covered 1-fold
P18	62.2	28634667	89.7	30350464	95.0	30866922	96.7	31283671	98.0
P20	77.3	29185621	91.4	30484459	95.5	30887362	96.7	31295467	98.0
P21	79.8	29221444	91.5	30463639	95.4	30862872	96.6	31268374	97.9
P22	102.7	29602038	92.7	30411380	95.2	30753878	96.3	31223181	97.8
P23	95.3	29824344	93.4	30657487	96.0	30976491	97.0	31356931	98.2
P25	75.9	29068110	91.0	30421050	95.3	30890606	96.7	31317631	98.1
P26	79.0	29206879	91.5	30478280	95.4	30915248	96.8	31356540	98.2
P27	77.6	29189639	91.4	30536770	95.6	30959369	96.9	31356138	98.2
P28	68.4	29038581	90.9	30465482	95.4	30900800	96.8	31287362	98.0
P29	79.8	29197944	91.4	30379082	95.1	30793888	96.4	31237188	97.8
P30	83.4	29410702	92.1	30581133	95.8	30986464	97.0	31375268	98.2
P31	39.9	25283305	79.2	29443210	92.2	30528049	95.6	31178979	97.6
P32	78.2	29134584	91.2	30484542	95.5	30953570	96.9	31378702	98.3
P33	38.6	18025493	56.4	21907641	68.6	24493070	76.7	27613377	86.5

Table 3.5: continued

Patient ID	Mean target base coverage	Number CCDS bases covered 20-fold	% CCDS bases covered 20-fold	Number CCDS bases covered 10-fold	% CCDS bases covered 10-fold	Number CCDS bases covered 5-fold	% CCDS bases covered 5-fold	Number CCDS bases covered 1-fold	% CCDS bases covered 1-fold
P34	53.6	19080357	59.7	22295709	69.8	24592972	77.0	27659325	86.6
P35	66.5	26849688	84.1	29588559	92.7	30568420	95.7	31305079	98.0

3.3.2. Coverage of ataxia genes and bases (based on the list studied by Nemeth et al., 2013)

Coverage of ataxia genes included in the recently published ataxia panel (Nemeth *et al.*, 2013) was calculated for the patients with ataxia studied here, as described in Section 3.2.7, and the result was previously published (Pyle *et al.*, 2014). Coverage calculated for total CCDS bases for the genes studied by Nemeth and colleagues. A total of 118 known and putative ataxia genes had 2,369 exons (386,444 bp). The mean target base coverage for all exomes was 64-fold. On average, 89% of all ataxia CCDS bases were covered 20-fold, 93% of CCDS bases were covered 10-fold and 95% of CCDS bases were covered 5-fold. These data are in line with findings by Nemeth and colleagues and the similar studies (Taylor *et al.*, 2014). Coverage and depth statistics for each patient is presented in Table 3.6.

Table 3.6: Coverage of ataxia genes studied by Nemeth et al., (2013) in the patients with ataxia studied here.
Coverage calculated for Total Consensus Coding Sequence (CCDS) bases for the genes studied by Nemeth *et al.,* 2013 (386,444 bp). Taken from Pyle *et al*., 2014.

Patient ID	Total CCDS bases (Nemeth panel - genes)	Mean per base Depth	Number bases covered 20-fold	% bases covered 20-fold	Number bases covered 10-fold	% bases covered 10-fold	Number bases covered 5-fold	% bases covered 5-fold	Number bases covered 1-fold	% bases covered 1-fold	Number CCDS exons (Nemeth panel - genes)	Number exons covered 20-fold	% exons covered 20-fold	Number exons covered 10-fold	% exons covered 10-fold	Number exons covered 5-fold	% exons covered 5-fold	Numbe r exons covered 1-fold	% exons covered 1-fold
P1	386444	54.8	259243	66.9	286151	73.6	303388	78.0	327346	83.8	2359	1674	71.0	1813	76.9	1889	80.1	2039	86.4
P2	386444	45.3	252131	64.7	283189	72.8	300767	77.6	324139	83.5	2359	1642	69.6	1800	76.3	1881	79.7	2021	85.7
P3	386444	51.2	259287	66.8	289107	74.4	304899	78.4	328304	84.3	2359	1680	71.2	1835	77.8	1889	80.1	2056	87.2
P5	386444	62.6	361107	93.8	377031	97.6	380905	98.5	383042	99.0	2359	2293	97.2	2322	98.4	2331	98.8	2337	99.1
P7	386444	61.7	358123	93.6	375557	97.6	379411	98.4	382156	98.9	2359	2296	97.3	2325	98.6	2330	98.8	2330	99.1
P9	386444	59.7	351611	92.4	373959	97.2	378680	98.2	382521	99.0	2359	2290	97.1	2320	98.3	2329	98.7	2339	99.2
P10	386444	72.3	364449	95.0	377463	97.9	380491	98.4	382702	99.0	2359	2306	97.8	2326	98.6	2330	98.8	2341	99.2
P11	386444	66.2	361529	94.3	376533	97.7	379961	98.4	382053	98.9	2359	2304	97.7	2323	98.5	2332	98.9	2339	99.2
P12	386444	67.1	362377	94.5	376551	97.7	379850	98.4	382505	98.9	2359	2305	97.7	2323	98.5	2331	98.8	2337	99.1
P13	386444	67.5	362923	94.5	377195	97.8	380278	98.4	382614	98.9	2359	2310	97.9	2324	98.5	2332	98.9	2338	99.1
P14	386444	55.2	354800	92.5	376076	97.6	380422	98.4	382635	98.9	2359	2288	97.0	2326	98.6	2332	98.9	2336	99.0
P15	386444	65.3	361752	94.2	376363	97.7	380008	98.4	382460	98.8	2359	2301	97.5	2326	98.6	2330	98.8	2336	99.0
P16	386444	96.6	370759	96.4	376417	97.8	378583	98.2	380873	98.8	2359	2309	97.9	2320	98.3	2332	98.9	2339	99.2
P17	386444	101.0	371604	96.9	376943	97.9	379170	98.3	381853	98.9	2359	2311	98.0	2323	98.5	2332	98.9	2340	99.2
P18	386444	55.2	356152	92.8	375686	97.5	379745	98.3	382734	99.0	2359	2285	96.9	2319	98.3	2330	98.8	2339	99.0
P20	386444	60.4	360037	93.6	376889	97.7	380350	98.4	382498	98.9	2359	2298	97.4	2324	98.5	2331	98.8	2338	99.1
P21	386444	62.7	360749	93.9	377245	97.8	380234	98.5	382437	98.8	2359	2306	97.8	2326	98.6	2331	98.8	2335	99.0

Table 3.6: continued

Patient ID	Total CCDS bases (Nemeth panel - genes)	Mean per base Depth	Number bases covered 20-fold	% bases covered 20-fold	Number bases covered 10-fold	% bases covered 10-fold	Number bases covered 5-fold	% bases covered 5-fold	Number bases covered 1-fold	% bases covered 1-fold	Number CCDS exons (Nemeth panel - genes)	Number exons covered 20-fold	% exons covered 20-fold	Number exons covered 10-fold	% exons covered 10-fold	Number exons covered 5-fold	% exons covered 5-fold	Number exons covered 1-fold	% exons covered 1-fold
P22	386444	83.5	369993	96.2	377391	97.9	379227	98.4	381909	98.9	2359	2310	97.9	2324	98.5	2331	98.8	2339	99.2
P23	386444	76.6	369406	96.0	378464	98.0	380778	98.5	382988	99.0	2359	2315	98.1	2329	98.7	2333	98.9	2339	99.2
P25	386444	65.4	359195	93.8	375917	97.5	380404	98.4	382775	98.9	2359	2299	97.4	2327	98.6	2330	98.8	2338	99.1
P26	386444	65.7	360687	94.2	376202	97.7	380103	98.5	383031	99.1	2359	2299	97.5	2326	98.6	2333	98.9	2343	99.3
P27	386444	61.9	360524	93.6	377193	97.7	380507	98.4	383269	99.1	2359	2295	97.3	2325	98.6	2330	98.8	2342	99.3
P28	386444	58.9	358872	93.4	375810	97.4	380212	98.4	383354	99.0	2359	2294	97.2	2321	98.4	2330	98.8	2339	99.2
P29	386444	64.4	362073	94.4	376806	97.8	379956	94.4	382415	98.9	2359	2301	97.5	2328	98.7	2331	98.8	2339	99.2
P30	386444	69.1	363523	94.5	377206	97.8	380357	98.4	382828	99.1	2359	2304	97.7	2326	98.6	2330	98.8	2342	99.3
P31	386444	35.7	306789	80.9	364983	95.1	377105	97.9	382264	98.9	2359	2184	92.6	2305	97.7	2326	98.6	2339	99.2
P32	386444	66.2	359471	93.9	375689	97.4	379807	98.3	382943	98.9	2359	2304	97.7	2321	98.4	2329	98.7	2341	99.2
P33	386444	40.8	246207	63.3	280984	72.1	301603	77.5	326255	83.7	2359	1612	68.3	1795	76.1	1883	79.8	2032	86.1
P34	386444	57.9	257857	66.5	284578	73.3	301361	77.5	327905	83.7	2359	1656	70.2	1798	76.2	1876	79.5	2032	86.1
P35	386444	59.8	341478	86.0	374078	96.1	382616	98.7	385785	99.8	2359	2208	93.6	2324	98.5	2346	99.4	2358	100.0

3.3.3. Filtering for putative causative variants

On average each individual had approximately 60,000 on target (within targets of Illumina Truseq 62Mb +/-500bp) single-nucleotide variants (SNV) and 6,000 on target indels (insertion/deletion). This study's filtering strategy yielded a list of prospective variants, all of which were verified with Sanger sequencing using custom designed primers (Chapter 2, Table 2.1). Population frequency data and pathogenicity scores for variants found in 27 patients from the ataxia cohort presented here were used in Pyle *et al.*, 2014 and are presented in Table 3.7.

Table 3.7: Population frequency data and pathogenicity scores for variants found in 27 patients with ataxia.
Key: P: Patient ID. In House MAF of 286 non-ataxia exomes; ESP6500 – MAF from 6500 exomes (NHLBI Exome Sequencing Project: http://evs.gs.washington.edu/EVS/); 1000g – MAF from 1000 genomes project, April 2012 (http://www.1000genomes.org/); SNP137 – SNP IDs from dbSNP137. Pathogenicity scores – PhyloP: higher scores are more deleterious, Conserved=>0.95, Non-Conserved=<0.95. SIFT: Sifting Intolerant from Tolerant, D: Deleterious (sift<=0.05); T: tolerated (sift>0.05). PolyPhen2: D: Probably damaging =>0.85, P:Possibly damaging=0.85-0.15, B:Benign=<0.15. LRT: Likelihood ratio test. High scores indicate a larger deleterious effect. D:Deleterious, N:Neutral, U: Unknown. Mutation Taster: The larger score suggest a higher probability to cause a human disease. A: Disease automatic, D: Disease causing, N:Polymorphism, P:Polymorphism automatic. GERP++: Higher scores are more deleterious. This table was used in Pyle *et al.,* 2014.

			Population Data				***Pathogenicity scores***										
P ID	***Gene***	***Variant***	***In House***	***ESP 6500***	***1000g***	***dbSNP137***	***PhyloP***		***SIFT***		***PolyPhen2***		***LRT***		***Mutation Taster***		***GERP***
							Score	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*
P1, P2	*SACS*	p.Thr692Thr fs*713; p.Gly1322Val fs*1343	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P3, P4	*SACS*	p.Glu4350*	0.0042	-	-	-	1	C	0.9	NA	0.74	NA	1	D	1	D	5.8
P5, P6	*SACS*	p.Ser527*	-	-	-		1	C	0.9	NA	0.74	NA	1	D	1	A	5.6
		p.Leu2261Ile	0.0084	0.0037	0.0041	rs146722795	1	C	1	D	0.16	P	1	D	0.75	D	5.4
P7, P8	*KCNC3*	p.Arg420His	-	-	-	rs104894699	0.99	C	1	D	1	D	1	U	0.98	D	2.46
P9, P10	*SPG7*	p.Ala510Val	0.007	0.0035	0.0014	rs61755320	1	C	1	D	0.98	D	1	D	1	D	4.56
		p.Ala572Val	-	-	-	rs72547551	1	C	1	D	0.11	B	1	D	1	D	5.3

Table 3.7: continued

			Population Data				Pathogenicity scores										
P ID	**Gene**	**Variant**	**In House**	**ESP 6500**	**1000g**	**dbSNP137**	**PhyloP**		**SIFT**		**PolyPhen2**		**LRT**		**Mutation Taster**		**GERP**
							Score	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*
P11, P12, P13	*TUBB4A*	p.Met300Ile	-	-	-	-	0.89	N	1	D	0.98	D	1	D	1	D	2.07
P14	*TUBB4A*	p.Ala364Asp	-	-	-	-	0.99	C	1	D	0.06	B	1	U	1	D	2.97
P15	*NPC1*	p.Met156Thr	0.0157	0.0115	0.01	rs147615070	1	C	0.86	T	0.27	P	1	D	1	D	5.12
P16		p.Pro237Sers	0.0052	0	-	rs80358251	1	C	0.6	T	0	B	1	D	0.97	D	5.4
P17,P18	*SLC1A3*	p.Arg454Gln	-	0.0003	-	rs138085358	0.99	C	0.81	T	0.36	P	1	N	1	D	4.13
P19	*ZFYVE26*	p.Leu817Cysfs*12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		p.Arg780*	-	-	-	-	0.99	C	0.9	NA	0.73	NA	0.99	N	1	A	5.58
P20	*WFS1*	p.Lys193Gln	0.0017	0.0022	0.0009	rs41264699	1	C	0.83	T	0.09	B	1	D	0.06	N	3.43
		p.Arg456His	0.0402	0.0467	0.06	rs1801208	1	C	0.78	T	0.99	D	1	D	0.69	P	4.42
P21	*FASTKD2*	c.-66A>G	0.0122	-	0.0046	-	-	-	-	-	-	-	-	-	-	-	-
		p.Lys50Arg	0.0052	0.0034	0.0009	rs141447598	0.88	N	0	T	0.031	B	0.43	N	0	N	0.74

Table 3.7: continued

			Population Data				Pathogenicity scores										
P ID	**Gene**	**Variant**	**In House**	**ESP 6500**	**1000g**	**dbSNP137**	**PhyloP**		**SIFT**		**PolyPhen2**		**LRT**		**Mutation Taster**		**GERP**
							Score	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*	*Pred*	*Score*
P22, P23, P24	*ZFYVE27*	c.805-2A>G	0	0.0001	-	-	-	-	-	-	-	-	-	-	-	-	-
P25	*WNK1*	p.Thr665Ile	0.0717	0.0554	0.07	rs2286007	1	C	1	D	0.06	B	1	D	0.49	P	5.36
		p.Thr1091Ile	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P26	*KCNB2*	p.Ser530Phe	0	0	-	-	1	C	1	D	0.98	D	0.1	N	0.07	N	5.12
		p.Leu784Pro	0	0	-	-	0.89	N	0.94	T	0	B	0	N	0.47	N	2.16
P27	*ABCB7*	p.Arg273Gln	0.0105	0.0027	0.0018	rs147584361											
	POLG	p.Trp748Ser	0	0.0004	-	rs113994097	1	C	0.67	T	1	D	1	D	1	D	5.22
		p.Glu1143Gly	0.0262	0.0275	0.02	rs2307441	0.79	N	0.96	D	0.74	P	1	D	0.95	D	2.02
	WFS1	p.Leu432Val	0.0017	0.0038	0.0027	rs35031397	1	C	1	D	0.98	D	1	D	0.87	D	4.38

3.4. Discussion

This study presents 35 individuals with unexplained ataxia from 22 randomly selected families of white European descent with no known consanguinity. Whole exome sequencing was performed on 30 patients enrolled in this study. In order to prioritize candidate genes, the raw data was carefully filtered. After filtering in prospective candidate genes, the putative causative variants were confirmed with Sanger sequencing in the patient and multiple affected and unaffected members of the family, if available.

As outlined in previous sections, many groups in earlier studies ruled out variants found in public databases such as dbSNP, the 1000 Genomes Project, and in in-house exome databases (Sherry *et al.*, 2001). Excluding rare variants, which were previously found in public databases, from the sequencing study could be very restrictive. This is because as more data are deposited in the public databases, more rare variants will be accounted for (Bamshad *et al.*, 2011). Also, rare pathogenic mutations could be deposited in the public databases because younger individuals with low penetrant neurological disorders have not yet presented with the phenotype (Keogh and Chinnery, 2013). In the more recent studies many groups would allow the minor allele frequency for variants found in public databases to be as high as 0.01 (McDonald *et al.*, 2012). The current study employed similar tactics. However, this study realize that a MAF cut off of 1% is arbitrary and should only be used as a guide. This is true for several reasons. Firstly, MAF values for rare variants could differ between public databases, as demonstrated in this study by an *NPC1* mutation c.709C>T:p.Pro237Ser in patients P15 and P16 (MAF=0.0153 in US-based ESP6500; MAF=0.0064 in UK-based 1000 Genomes project; MAF=0.0052 in an in-house panel of 286 unrelated exomes). The latter data was previously published (Pyle *et al.*, 2014; Pyle *et al.*, 2015) and presented here in Table 3.4. Secondly, MAF cut off values vary depending on different modes of inheritance and availability of the family history in the patients being studied. The current study filtered out all variants with MAF>=0.01 in presumed dominant pedigrees, but allowed MAF>0.01 for compound heterozygous variants in presumed recessive cases. The latter was considered only when the other allele was very rare. One example of such approach is compound heterozygous *FASTKD2* mutations, which was identified in patient P21 (Table 3.4). While mutation c.-66A>G has a MAF>0.01 in one of the databases, when taken together with the other mutation, it would cause disease with an incidence of approximately 1:17,000. This is well below recently suggested

incident rate of 1:10,000 for rare recessive disorders (Stranneheim and Wedell, 2015). Likewise, compound heterozygous *WFS1* mutations identified in patient P20 (Table 3.4) would be expected to cause a disease with an incidence rate of about 1:14,000. This is despite the p.Arg456His mutation having MAF=0.04 in ESP6500 database.

In order to prioritise putative causative variants, this study used keywords, as described in Section 3.2.5, to filter for variants in known ataxia and related neurological disorder genes. This enabled a significant reduction in the variant numbers (selected examples are presented in Figure 3.2). Three examples are highlighted here. Firstly, patient P7 had 1,649 rare/novel variants (Figure 3.2, a), out of which 396 variants were both heterozygous and protein altering. 43 of these variants contained the terms “nerv” or “neuro” in Gene Ontology and only one variant contained the term “ataxia” in OMIM database. Secondly, patients P9 and P10 had 1,419 and 1,609 rare/novel variants, respectively (Figure 3.2, b). Out of these, patient P9 had 522 and patient P10 had 510 rare/novel protein altering variants. The two patients shared 25 compound heterozygous protein altering variants, out of which only one variant had terms “nerv” or “neuro” or “brain” or “ataxia” in Gene Ontology or OMIM database. And finally, patient P17 had 1,421 rare/novel variants (Figure 3.2, c), out of which 624 were both heterozygous and protein altering. Three of these variants contained the terms “nerv” or “neuro” or “brain” or “ataxia” in Gene Ontology or OMIM database.

a

Patient ID/ inheritance	Variant type	On target (*)	Rare/novel variants (**)	Number of genes with rare variants: Heterozygous protein altering	Number of genes with rare variants: Contains "neuro" or "nerv" in Gene Onthology	Number of genes with rare variants: Contains "ataxia" in OMIM
P7/dominant	SNV	65,662	1,649	396	43	1
	Indel	7,538				

b

Patient ID/ inheritance	Variant type	On target (*)	Rare/novel variants (**)	Rare/novel protein altering variants	Number of shared genes with rare variants: Compound heterozygous protein altering	Number of shared genes with rare variants: Compound heterozygous protein altering and contains "neuro" or "nerv" or "brain" or "ataxia" in Gene Onthology or OMIM
P9/recessive	SNV	63,999	1,419	522	25	1
	Indel	7,508				
P10/recessive	SNV	67,390	1,609	510		
	Indel	7,902				

c

Patient ID/ inheritance	Variant type	On target (*)	Rare/novel variants (**)	Number of Genes with Rare Variants: Heterozygous protein altering	Number of Genes with Rare Variants: Single heterozygous protein altering and contains "neuro" or "nerv" or "brain" or "ataxia" in Gene Onthology or OMIM
P17/dominant	SNV	70,573	1,421	624	3
	Indel	7,004			

Figure 3.2: Genetic variant numbers identified in selected patients following exome sequencing and bioinformatic analysis pipeline. Variant numbers identified in patients P7 (a), P9 and P10 (b) and P17 (c). SNV=single-nucleotide variants; indel=insertion/ deletion variants. (*) Variants with position within targets (Illumina Truseq 62Mb) +/-500bp, observed on both (forward and reverse) strands and (SBVs only) variant allele frequency >24%. (**) Rare/novel variants (MAF<=0.01) with exclusion of common variants found to be shared with an in-house panel of 286 individuals, dbSNP137, 1000 genomes (April 2012 data release) and Exome Variant Server NHLBI-ESP6500 databases (May 2012 release).

After filtering with keywords and identifying putative candidate genes, an extensive literature search was performed. Following this, current variants associated with ataxia and related neurological disorders were identified. The study then cross-referenced the published variants with the variants filtered with the pipeline described here. The current study first concentrated on the known variants in the known ataxia and related neurological disorder genes. Following this the study searched for novel variants in the known ataxia and related neurological disorder genes. Finally, this study considered rare variants in potential novel ataxia genes, which would fit the phenotype in the patients from the cohort. Studies by other research groups, which were published after this study was performed, employed a similar approach (Fogel *et al.*, 2014; Gonzaga-Jauregui *et al.*, 2015). This study eventually was able to identify known pathogenic variants in genes associated with ataxia and related neurological disorders in two families. First, a dominant *KCNC3* c.1259G>A:p.Arg420His mutation in patient P7 (Table 3.4), which

was previously associated with ataxia (Waters *et al.*, 2006; Figueroa *et al.*, 2011). In the current study this mutation segregated with the disease (confirmed by a segregation study in patient P8 and unaffected members of this family). Second, compound heterozygous *SPG7* mutations in patients P9 and P10 (c.1529C>T:p.Ala510Val and c.1715C>T:p.Ala572Val) (Table 3.4), which were previously described in patients with ataxic features (Casari *et al.*, 1998; Pfeffer *et al.*, 2015). A novel mutation in a known ataxia gene *SLC1A3* (de Vries *et al.*, 2009) was identified in patient P17 (c.1361G>A:p.Arg45Gln) (Table 3.4). The latter mutation segregated with the disease in this family (it was also present in patient P18 and absent in the unaffected family members). This study also identified novel mutations in a known ataxia gene *SACS* in three unrelated families (Table 3.4).

In this study the variants have been chosen according to the inheritance mode in each family. For example, in families with a presumed dominant inheritance model this study primarily filtered for genes with a single heterozygous change (shared between multiple exomes, where available). Likewise, in families with a presumed recessive mode of inheritance the current study selected genes with at least two mutated alleles (for non-consanguineous families). None of the families, included in this study, were consanguineous, therefore the majority of the homozygous changes were excluded. However, one family (patients P3 and P4) presented with a seemingly homozygous *SACS* stop-codon mutation c.13048G>T:p.Glu4350*, which was present in the mother, but not in the father (Pyle *et al.*, 2013). The comparative genomic hybridisation study identified a 0.7 Mb deletion on chromosome 13q12.12, which includes the *SACS* gene, in both brothers. The deletion was heterozygous in the asymptomatic father. This example shows that in non-consanguineous families seemingly homozygous variants could, in fact, be hemizygous, indicating chromosomal rearrangements. The latter are increasingly recognised as a cause of late-onset neurological disease (Pyle *et al.*, 2013).

In conclusion, this study demonstrates that a stepwise filtering algorithm is an effective tool for interpretation of exome sequencing data. Filtering significantly reduces the number of putative causative variants from tens of thousands down to just about a dozen per individual. Such a tremendous reduction of data makes it easier to evaluate prospective variants. This aids in establishing a prompt molecular diagnosis in a highly heterogeneous group of patients with a neurological disorder.

Chapter 4

Whole exome sequencing for molecular diagnosis of neurological disorders using undiagnosed inherited and sporadic ataxias as a model

Chapter 4. Whole exome sequencing for molecular diagnosis of neurological disorders using undiagnosed inherited and sporadic ataxias as a model

4.1. Introduction

4.1.1. Hereditary ataxia overview

Clinical manifestation

Ataxia is defined as a disturbance of balance and movement coordination (Shakkottai and Fogel, 2013). Hereditary ataxias are a group of progressive neurological disorders, which are highly heterogeneous both clinically and genetically. Phenotypically, hereditary ataxias are characterized by incoordination of gait (ataxic gait) and disturbance of hand and eye movements and speech (Jayadev and Bird, 2013). In most cases, patients with ataxia are presented with atrophy of the cerebellum. Additionally, patients may present with neuropathy, spasticity, seizures and cognitive damage (Shakkottai and Fogel, 2013).

Ataxias can be subdivided on the basis of inheritance mode into autosomal dominant, autosomal recessive, X-linked, and mitochondrial (Bird, 1993 -). Ataxias are also categorized by a gene or chromosomal locus, which contain the causative mutation(s).

The full phenotypic spectrum of ataxias is not fully established; new types of ataxia and new ataxia genes are still being discovered. (Fogel *et al.*, 2014a). Therefore, the ataxia nomenclature is a work in progress. Also, there is a substantial overlap between ataxias and some other neurological disorders, which makes precise diagnosis difficult. It is estimated that, when taken together, prevalence of all ataxias could be 15-20:100,000 (Smeets and Verbeek, 2014).

Establishing the diagnosis

The first step in establishing the diagnosis is a clinical examination of the patient involving neurological tests, detailed family history and magnetic resonance imaging (MRI) of the brain (Fogel *et al.*, 2014a). The aim of the first step is to rule out common non-genetic (acquired) causes of ataxia. Acquired conditions causing ataxia are represented by an array of disorders including toxic, inflammatory, nutritional, neoplastic and infectious disorders (Shakkottai and Fogel, 2013). Hereditary nature of

ataxia in the patient is supported by positive family history, recognizing a phenotype previously linked to hereditary ataxia and identification of ataxia-causing mutation(s) (Jayadev and Bird, 2013). Distinguishing between hereditary and acquired ataxias is important because for some acquired ataxias a specific treatment may be available (Shakkottai and Fogel, 2013).

After ruling out acquired causes of ataxia, genetic investigations are concentrated on the most common ataxia causing genes (Fogel *et al.*, 2014a). Firstly, the patient is subjected to genetic tests for the repeat expansion disorders. Dominant cases are investigated for repeat expansions in genes *ATXN1, ATXN2, ATXN3, CACNA1A, ATXN7,* which cause spinocerebellar ataxias (SCA) SCA1, SCA2, SCA3, SCA6 and SCA7, respectively. Dominant cases also could be investigated for repeat expansions in genes *ATXN8* / *ATXN80S*, *ATXN10, PPP2R2B*, *TBP,* and *ATN1,* which cause SCA8, SCA10, SCA12, SCA17 and dentatorubral-pallidoluysian atrophy (DRPLA), respectively. It is estimated that up to 60% of hereditary dominant ataxias could be identified with testing for repeat expansion disorders (Jayadev and Bird, 2013). Patients with recessive family history are investigated for repeat expansions in *FXN,* which cause Friedreich ataxia. Male patients with suspected X-linked hereditary ataxia cases are investigated for repeat expansions in *FMR1*, which cause fragile X-associated tremor/ataxia syndrome. Secondly, the patient is investigated for the most common types of ataxia caused by non-repeat mutations (Fogel *et al.*, 2014b). Cases with dominant family history could be investigated for mutations in *SPTBN2*, *KCNC3, PRKCG, ITPR1, FGF14* and *AFG3L2,* which cause SCA5, SCA13, SCA14, SCA15, SCA27 and SCA28, respectively. Cases with recessive family history could be considered for genetic tests in *ATM*, *APTX*, *SETX* and *TTPA,* which cause ataxia-telangiectasia, ataxia with oculomotor apraxia type 1 (AOA1), ataxia with oculomotor apraxia type 2 (AOA2) and ataxia with vitamin E deficiency (AVED), respectively.

If the patient is tested negative for all the above investigations, then advanced diagnostic tests, such as clinical exome or genome sequencing, is recommended (Fogel *et al.*, 2014b).

Autosomal Dominant Cerebellar Ataxia

Prevalence of autosomal dominant cerebellar ataxias (ADCA) varies between countries, possibly due to a founder effect (Gasser *et al.*, 2010). In the Netherlands it is estimated

to be around 3:100,000 (van de Warrenburg *et al.*, 2002). More recent meta-analysis of the global ADCA prevalence was estimated to be 1-5:100,000 (Ruano *et al.*, 2014).

ADCA are further subdivided into spinocerebellar ataxias (SCA), episodic ataxias (EA), dentatorubral-pallidoluysian atrophy (DRPLA), autosomal dominant cerebellar ataxia with deafness and narcolepsy (ADCADN), hypomyelinating leukoencephalopathy and spastic ataxia (SPAX1) (Bird, 1993 -). SCAs represent the most numerous group of ADCAs. The most common cause of dominantly inherited ataxias is repeat expansion within the coding or noncoding regions of the corresponding genes (Duenas *et al.*, 2006). Out of these, CAG repeat expansions are the most common. (Durr, 2010). CAG codes for glutamine, therefore disorders caused by expansion of CAG repeat are referred to as polyglutamine disorders. Besides CAG repeat mutations, there are numerous conventional mutations in associated genes, which cause ADCA. In a recent study of French ataxias, the authors estimated that repeat expansions account for 45% of all dominant ataxias, whereas conventional mutations account for only 6% with the remaining 48% of all dominant ataxias remaining genetically undiagnosed (Durr, 2010). The nomenclature for the inheritable ataxias is still being updating due to phenotypic overlap between ataxia subtypes. The latest data for ADCA are presented in Table 4.1.

Table 4.1: Molecular genetics of autosomal dominant cerebellar ataxias.
This table depicts the molecular genetics of autosomal dominant cerebellar ataxias. Chromosomal locus is given only when the gene is unknown**.** For SCA9 there is no information published to date (adapted from (Bird, 1993 -)).

Disease Name	Gene Symbol or Chromosomal Locus	Type of Mutation	Reference
SCA1	*ATXN1*	CAG repeat	(Subramony and Ashizawa, 1993-)
SCA2	*ATXN2*	CAG repeat	(Pulst, 1993--a)
SCA3	*ATXN3*	CAG repeat	(Paulson, 2009)
SCA4	16q22.1	---	(Flanigan *et al.*, 1996; Hellenbroich *et al.*, 2003; Edener *et al.*, 2011)
SCA5	*SPTBN2*	Non-repeat mutations	(Ikeda *et al.*, 2006)
SCA6	*CACNA1A*	CAG repeat	(Gomez, 1993-)
SCA7	*ATXN7*	CAG repeat	(Garden, 1993-)
SCA8	*ATXN8 / ATXN80S*	CAG/CTG	(Ikeda *et al.*, 1993-)
SCA10	*ATXN10*	ATTCT repeat	(Matsuura and Ashizawa, 1993-)
SCA11	*TTBK2*	Non-repeat mutations	(Houlden, 1993-)
SCA12	*PPP2R2B*	CAG repeat	(Margolis *et al.*, 1993-)
SCA13	*KCNC3*	Non-repeat mutations	(Pulst, 1993--b)
SCA14	*PRKCG*	Non-repeat mutations	(Chen *et al.*, 1993-)
SCA15	*ITPR1*	Deletion of the 5' part of the gene	(Storey, 1993-)
SCA16	*CNTN4*	Non-repeat mutation	(Miura *et al.*, 2006)
SCA17	*TBP*	CAA/CAG repeat mutation	(Toyoshima *et al.*, 1993-)
SCA18	7q22-q32	---	(Brkanac *et al.*, 2002; Brkanac *et al.*, 2009)
SCA19/22	*KCND3*	Non-repeat mutations	(Verbeek *et al.*, 2002; Chung *et al.*, 2003; Schelhaas *et al.*, 2004; Duarri *et al.*, 2012; Lee *et al.*, 2012)
SCA20	11q12.2-11q12.3	260-kb duplication	(Storey, 1993-)
SCA21	*TMEM240*	Non-repeat mutations	(Delplanque *et al.*, 2014)
SCA23	*PDYN*	Non-repeat mutations	(Verbeek *et al.*, 2004; Bakalkin *et al.*, 2010)
SCA25	2p21-p13	---	(Stevanin *et al.*, 2003)

Table 4.1: continued

Disease Name	Gene Symbol or Chromosomal Locus	Type of Mutation	Reference
SCA26	*EEF2*	Non-repeat mutations	(Yu *et al.*, 2005; Hekman *et al.*, 2012)
SCA27	*FGF14*	Non-repeat mutations	(van Swieten *et al.*, 2003)
SCA28	*AFG3L2*	Non-repeat mutations	(Cagnoli *et al.*, 2006; Mariotti *et al.*, 2008)
SCA29	*ITPR1*	Non-repeat mutations	(Dudding *et al.*, 2004)
SCA30	4q34.3-q35.1	---	(Storey *et al.*, 2009)
SCA31	*BEAN1*	TGGAA Intronic repeat expansion	(Sato *et al.*, 2009; Sakai *et al.*, 2010; Edener *et al.*, 2011)
SCA34	*ELOVL4*	Non-repeat mutations	(Cadieux-Dion *et al.*, 2014)
SCA35	*TGM6*	Non-repeat mutations	(Wang *et al.*, 2010)
SCA36	*NOP56*	GGCCTG Intronic repeat expansion	(Kobayashi *et al.*, 2011)
SCA37	1p32	---	(Serrano-Munuera *et al.*, 2013)
SCA38	*ELOVL5*	Non-repeat mutations	(Di Gregorio *et al.*, 2014)
SCA40	*CCDC88C*	Non-repeat mutations	(Tsoi *et al.*, 2014)
DRPLA	*ATN1*	CAG repeat	(Tsuji, 1993-)
ADCADN	*DNMT1*	Non-repeat mutations	(Klein *et al.*, 2013)
Hypo-myelinating leukoen-cephalopathy	*TUBB4A*	Non-repeat mutations	(Hamilton *et al.*, 2014; Miyatake *et al.*, 2014)
EA1	*KCNA1*	Non-repeat mutation	(D'Adamo *et al.*, 1993-)
EA2	*CACNA1A*	Non-repeat mutations	(Spacey, 1993-)
EA3	1q42	---	(Cader *et al.*, 2005)
EA4	---	---	(Steckley *et al.*, 2001)
EA5	*CACNB4*	Non-repeat mutation	(Jen *et al.*, 2007)
EA6	*SLC1A3*	Non-repeat mutation	(Winter *et al.*, 2012)
EA7	19q13	---	(Kerber *et al.*, 2007)
SPAX1	*VAMP1*	Non-repeat mutations	(Meijer *et al.*, 2002; Bourassa *et al.*, 2012)

Autosomal Recessive Cerebellar Ataxia

Similarly to ADCAs, the prevalence of autosomal recessive cerebellar ataxias (ARCA) varies greatly between countries. The recent estimation of the global ARCA prevalence was 3:100,000 (Ruano *et al.*, 2014). The most common forms of ARCAs in Europe are Friedreich ataxia (FRDA), ataxia telangiectasia (AT), and ataxia with oculomotor apraxia type 1 and type 2 (AOA1 and AOA2, respectively). Prevalence of FRDA is estimated to be 1:50,000 in Europe (Gasser *et al.*, 2010). The most recent autosomal recessive hereditary ataxia nomenclature is outlined in Table 4.2. The table is divided into two parts with the first part listing more prevalent autosomal recessive ataxias (reported in more than 5 families worldwide) or treatable types. There are several types of recessive ataxia, for which a treatment is available. These include cerebrotendinous xanthomatosis (CTX) (treated with chenodeoxycholic acid), Refsum disease (treated with dietary phytanic acid), Brown-Vialetto-Van Laere syndrome 2 (treated with riboflavin), ataxia with vitamin E deficiency (AVED) (treated with vitamin E) and primary coenzyme Q10 deficiency-4 (COQ10D4), also known as autosomal recessive spinocerebellar ataxia-9 (SCAR9) (treated with coenzyme Q10 supplement). The first part of Table 4.2 also contains autosomal recessive ataxias more often found in a specific ethnic group, for example, autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), which is found in French-Canadians (Ogawa *et al.*, 2004). The second part of Table 4.2 lists less common autosomal recessive ataxias (reported in 1-5 families worldwide) (Musselman *et al.*, 2014).

Table 4.2: Molecular genetics of Autosomal Recessive Cerebellar Ataxias.
This table depicts the molecular genetics of Autosomal Recessive Cerebellar Ataxias (adapted from Bird, 1993 -). *More common - reported in >5 families; **treatable – types of ataxia with treatment available; ***less common – reported in 1-5 families

Gene / Locus	Disease	References
More common* and/or treatable**		
ANO10	Autosomal recessive spinocerebellar ataxia 10 (SCAR10)	(Vermeer *et al.*, 2010; Renaud *et al.*, 2014)
APTX	Ataxia with oculomotor apraxia type 1 (AOA1)	(Coutinho and Barbot, 1993-)
ATM	Ataxia-telangiectasia (AT)	(Gatti, 1993-)
C10orf2	Infantile-onset spinocerebellar ataxia (IOSCA)	(Nikali and Lonnqvist, 1993-)
CYP27A1	Cerebrotendinous xanthomatosis (CTX)	(Kim *et al.*, 1994)
FXN	Friedreich ataxia (FRDA)	(Bidichandani and Delatycki, 1993-)
PHYH PEX7	Refsum disease	(Wanders *et al.*, 1993-)
PNPLA6	Boucher-Neuhäuser syndrome	(Tarnutzer *et al.*, 2015)
SACS	Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)	(Vermeer *et al.*, 1993-)
SETX	Ataxia with oculomotor apraxia type 2 (AOA2)	(Moreira and Koenig, 1993-)
SIL1	Marinesco-Sjögren syndrome	(Anttonen and Lehesjoki, 1993-)
SLC52A2	Brown-Vialetto-Van Laere syndrome 2	(Fogel *et al.*, 2014a; Foley *et al.*, 2014)
TTPA	Ataxia with vitamin E deficiency (AVED)	(Schuelke, 1993-)
WFS1	Wolfram syndrome	(Chaussenot *et al.*, 2011; Fogel *et al.*, 2014a)
Less common***		
ABHD12	Polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract (PHARC)	(Fiskerstrand *et al.*, 2010; Chen *et al.*, 2013)
ACO2	Infantile cerebellar- retinal degeneration (ICRD)	(Spiegel *et al.*, 2012)
ADCK3 (CABC1)	Autosomal recessive spinocerebellar ataxia 9 (SCAR9)	(Lagier-Tourenne *et al.*, 2008; Mollet *et al.*, 2008)

Table 4.2: continued

Gene / Locus	Disease	References
Less common		
ATCAY	Cayman ataxia	(Bomar *et al.*, 2003)
ATP8A2	Cerebellar ataxia, mental retardation, and dysequilibrium syndrome 4 (CAMRQ4)	(Onat *et al.*, 2013)
STUB1 (CHIP)	Autosomal recessive spinocerebellarataxia16 (SCAR16)	(Shi *et al.*, 2013; Depondt *et al.*, 2014; Synofzik *et al.*, 2014)
CLCN2	Leukoencephalopathy with ataxia (LKPAT)	(Depienne *et al.*, 2013)
CLN5	Adult-onset autosomal recessive ataxiaassociated with neuronal ceriod-lipofuscinosis 5 (CLN5 disease)	(Mancini *et al.*, 2015)
CWF19L1	Autosomal recessive ataxia (Turkish)	(Burns *et al.*, 2014)
FLVCR1	Posterior column ataxia with rentinitis pigmentosa (AXPC1)	(Ishiura *et al.*, 2011)
GOSR2	Ramsay Hunt syndrome	(Corbett *et al.*, 2011)
GRID2	Autosomal recessive spinocerebellar ataxia 18 (SCAR18)	(Hills *et al.*, 2013; Van Schil *et al.*, 2014)
GRM1	Autosomal recessive spinocerebellar ataxia 13 (SCAR13)	(Guergueltcheva *et al.*, 2012)
KCNJ10	SeSAME syndrome	(Scholl *et al.*, 2009)
KIAA0226	Autosomal recessive spinocerebellar ataxia 15 (SCAR15)	(Assoum *et al.*, 2013)
LAMA1	Cerebellar dysplasia	(Aldinger *et al.*, 2014)
PNKP	Ataxia with oculomotor apraxia type 4 (AOA4)	(Bras *et al.*, 2015)
POLG	Mitochondrial recessive ataxiasyndrome (MIRAS)	(Mignarri *et al.*, 2015)
PTF1A	Pancreatic and cerebellar agenesis (PACA)	(Sellick *et al.*, 2004)
RNF216	Gordon Holmes syndrome	(Santens *et al.*, 2015)
SLC9A1	Lichtenstein-Knorr syndrome	(Guissart *et al.*, 2015)
SPTBN2	Autosomal recessive spinocerebellarataxia14 (SCAR14)	(Lise *et al.*, 2012; Elsayed *et al.*, 2014)
SYNE1	*SYNE1*-related autosomal recessive cerebellar ataxia	(Gros-Louis *et al.*, 2007)
SYT14	Autosomal recessive spinocerebellar ataxia 11 (SCAR11)	(Doi *et al.*, 2011)
TDP1	Spinocerebellar ataxia with axonal neuropathy (SCAN1)	(Takashima *et al.*, 2002)
TPP1	Autosomal recessive spinocerebellar ataxia 7 (SCAR7)	(Breedveld *et al.*, 2004; Sun *et al.*, 2013)

Table 4.2: continued

Gene / Locus	Disease	References
Less common		
VLDLR	*VLDLR*-associated cerebellar hypoplasia(CAMRQ1)	(Ali *et al.*, 2012)
WWOX	Autosomal recessive spinocerebellar ataxia 12 (SCAR12)	(Mallaret *et al.*, 2014)
ZNF592	Autosomal recessive spinocerebellar ataxia 5 (SCAR5)	(Mallaret *et al.*, 2014)
9q34-qter	Autosomal recessive spinocerebellar ataxia 2 (SCAR2)	(Delague *et al.*, 2001)

X-Linked Hereditary Ataxias

X-linked hereditary ataxias are uncommon (Bird, 1993 -), several types have been described to date with fragile X tremor ataxia syndrome (FXTAS) being the most common (Zanni and Bertini, 2011). FXTAS is caused by the expansion of a CGG-trinucleotide repeat in the fragile X mental retardation 1 gene (*FMR1)* (Leehey, 2009). The latest nomenclature for X-linked hereditary ataxias is presented in Table 4.3

Table 4.3: Molecular genetics of X-linked hereditary ataxias. This table depicts the molecular genetics of X-linked hereditary ataxias (adapted from Bird, 1993 -).

Gene / Locus	Disease	References
ABCB7	X-linked sideroblastic anemia and ataxia (XLSA/A)	(Allikmets *et al.*, 1999)
CASK	CASK-related disorders	(Moog *et al.*, 2011)
FMR1	fragile X-associated tremor/ataxia syndrome (FXTAS)	(Hagerman *et al.*, 2001)
OPHN1	X-linked mental retardation with cerebellar hypoplasia and distinctive facial appearance	(Zanni *et al.*, 2005)
SLC9A6	Syndromic X-linked mental retardation, Christianson type	(Garbern *et al.*, 2010)
Xq25-q27.1	X-linked spinocerebellar ataxia 5	(Zanni *et al.*, 2008)

Spastic ataxias

Spastic ataxias represent a group of disorders which combine spasticity with features of cerebellar ataxia (Bird, 1993 -). There are five disorders, which are designated spastic ataxia (SPAX1 through to SPAX5) with SPAX1 being inherited dominantly and the other four exhibiting an autosomal recessive inheritance. Spastic paraplegia type 7 (SPG7) is an autosomal recessive disorder caused by mutations in *SPG7*, which encodes paraplegin (Roxburgh *et al.*, 2013). Recently, mutations in *SPG7* were described as a major cause of unexplained ataxia with adult-onset presentation (Pfeffer *et al.*, 2015). Spastic ataxia types are presented in Table 4.4.

Table 4.4: Molecular genetics of disorders with spasticity and cerebellar ataxia. This table depicts the molecular genetics of spastic ataxias (adapted from Bird, 1993 -).

Disease	Gene / Locus	References
SPAX1	*VAMP1*	(Bourassa *et al.*, 2012)
SPAX2	*KIF1C*	(Dor *et al.*, 2014)
SPAX3	*MARS2*	(Bayat *et al.*, 2012)
SPAX4	*MTPAP*	(Crosby *et al.*, 2010)
SPAX5	*AFG3L2*	(Pierson *et al.*, 2011)
SPG7	*SPG7*	(Pfeffer *et al.*, 2015)

Mitochondrial Ataxias

Ataxia phenotype can associate with mutations in the mitochondrial DNA (mDNA). For example, mutations in ATP synthase 6 gene (*MTATP6)* were found to cause childhood and adult-onset cerebellar ataxia *(Pfeffer et al., 2012)*. Ataxia may also associate with mitochondrial disorders such as myoclonic epilepsy with ragged red fibres (MERRF), neuropathy, ataxia, and retinitis pigmentosa (NARP) and Kearns-Sayre syndrome (Rowland, 1983; Finsterer, 2009). Furthermore, mutations in nuclear-encoded genes, which regulate mitochondrial biogenesis, could result in hereditary ataxia (Table 4.2, *POLG1*).

4.1.2. Whole exome sequencing in ataxia

Precise diagnosis of hereditary ataxias is important because it leads to accurate genetic counselling and aids in diagnosis for families or patients with similar phenotypes. Correct diagnosis could also alter patient management as treatments are available for certain types of ataxia (Dixon-Salazar *et al.*, 2012). The current diagnostic laboratory approach for dissecting heterogeneous monogenic disorders, such as ataxia, is stepwise genetic testing, which is expensive and time consuming. After exclusion of common acquired causes of ataxia, the majority of inherited ataxias arise as a result of repeat expansions (Hersheson *et al.*, 2012). These mutations are usually detected on a routine diagnostic test, which provides a diagnosis for up to 60% of cases with a positive family history (Anheim *et al.*, 2010). The remaining 40% of inherited ataxias are individually rare (<1%). The diagnostic approach for the latter is screening for individual genes, which is a lengthy and expensive procedure. The current cost of the diagnosis of a neurological disorder could exceed £6,500 per individual

(Kingsmore and Saunders, 2011).

Targeted next generation sequencing was recently shown to dramatically improve diagnostic outcome in ataxia patients. Hoischen and co-workers performed a pilot study using autosomal recessive ataxia as a model neurological disorder to assess the application of next generation sequencing in diagnostic settings (Hoischen *et al.*, 2010). Coding and noncoding regions of seven known autosomal recessive ataxia genes were targeted in five patients with known mutations and two unaffected controls. On average 80% of the targeted region was covered 25-fold per base. Six out of seven known mutations were successfully identified. In another study, 58 known ataxia genes were captured in 50 highly heterogeneous ataxia patients without diagnosis (Nemeth *et al.*, 2013). The authors were able to establish diagnosis in only 18% of the patients, explaining this by the limited portfolio of genes included in the panel. An additional challenge in targeted next generation sequencing is unusual clinical presentation, which might lead a clinician to choose an incorrect gene panel for screening. Whole exome sequencing could overcome this challenge.

The aim of this study was to evaluate the use of whole exome sequencing for diagnostic purposes in neurological disorders. A heterogeneous cohort of patients with suspected inherited ataxia was used as an example of a neurological disorder and whole exome sequencing was applied with the aim to identify candidate gene mutations.

4.2 Materials and methods

4.2.1. Inclusion criteria

This study presents a cohort of 35 affected individuals with unexplained ataxia, as described in Section 2.1. The main inclusion criterion was ataxia as a predominant clinical feature. All common sporadic, inherited and metabolic causes were excluded on routine clinical investigations prior to inclusion in this study.

4.2.2. Molecular genetics and bioinformatics

All samples were prepared and sequenced, as described in Section 2.4.1. Bioinformatic analysis was performed as described in Section 2.4.2. Variant filtering was performed as described in Section 3.2. Coverage analysis of CCDS bases and known ataxia genes

was performed as described in Sections 3.2.6 and 3.2.7, respectively.

4.2.3. Variant definition

Variants were defined as previously described (Pyle *et al.*, 2014). Definition criteria were as follows:

1) Confirmed pathogenic variant – a previously described pathogenic variant or a variant in a known ataxia gene where the variant was predicted to affect protein structure and function, and segregated with at least one affected family member.
2) Possible pathogenic variant – a variant in a known disease gene where the variant was predicted to alter protein structure and function, but not fulfilled all criteria for confirmed pathogenic variants.
3) Variants of uncertain significance or no candidate variants found.

4.2.4. In silico mutation pathogenicity study (splice prediction using Alamut software)

Pathogenicity of the variants was assessed with mutation interpretation software Alamut (http://www.interactive-biosoftware.com/), which incorporates five splice prediction methods, namely: SpliceSiteFinder-like (Zhang, 1998), MaxEntScan (Yeo and Burge, 2004), NNSPLICE (Reese *et al.*, 1997), GeneSplicer (Pertea *et al.*, 2001) and Human Splicing Finder (Desmet *et al.*, 2009). Alamut software also has a capacity to predict exonic splicing enhancers (ESEs), which are short sequences within exons that enhance pre-mRNA splicing (Blencowe, 2000). Two programs are incorporated within Alamut in order to aid this, namely ESEfinder (Cartegni *et al.*, 2003) and RESCUE_ESE (Fairbrother *et al.*, 2004). An additional utility incorporated into Alamut in order to aid in determining which exonic variant has the highest chance to skip this exon is program EX-SKIP (available at http://ex-skip.img.cas.cz/).

4.3 Results

4.3.1. Clinical presentation, laboratory investigations and exome sequencing results for 35 patients with suspected inherited ataxia

Clinical presentation, comprehensive laboratory investigations and exome sequencing results for 35 patients with suspected inherited ataxia studied here were previously published (Pyle *et al.*, 2014) and are presented in Table 4.5. Additional clinical data and

segregation analysis results for cases with confirmed and possibly pathogenic mutations were previously published (Pyle *et al.*, 2014) and are presented in Appendix 1.

Table 4.5. Clinical presentation, laboratory investigations and exome sequencing results for 35 patients with suspected inherited ataxia
Abbreviations: M = male, F = female; Hom = homozygous; Het = heterozygous; '+' = present; '-' = absent, not applicable (test not carried out); sib = sibling; aff = affected; Normal = all results normal/negative; MRI = magnetic resonance imaging; [s]Segregation analysis performed in the family two have been reported previously (Pyle *et al*., 2012; Pyle *et al*., 2013)
This table included in Pyle *et al*., 2014.
Abbreviations: M = male, F = female; Hom = homozygous; Het = heterozygous; '+' = present; '-' = absent, not applicable (test not carried out); sib = sibling; aff = affected; Normal = all results normal/negative; MRI = magnetic resonance imaging; [s]Segregation analysis performed in the family two have been reported previously (Pyle *et al*., 2012; Pyle *et al*., 2013)
Adapted from Pyle *et al*., 2014.

Confirmed pathogenic

ID, Sex	***Age of onset (y)***	***Family History***	***Symptom at onset***	***Clinical signs***						***Investigations***					***Gene***	***Variants***
					Eyes					***Electrophysiology***			***MRI***			
				Cerebellar a ataxia	***Jerky ocular pursuit***	***Gaze evoked nystagmus***	***Hypometric saccades***	***Optic atrophy***	***Cerebellar dysarthria***	***Axonal sensorimotor neuropathy***	***Demyelinating sensorimotor neuropathy***	***Normal***	***Cerebellar atrophy***	***Generalised atrophy***		
P1, M	teens	aff sib (P2)	Learning difficulties	+	-	-	-	-	-	+	+	-	+	-	*SACS*	c.2076delG:p.Thr692Thrfs*713; c.3965_3966delAC:p.Gly1322Valfs*1343[s]

Table 4.5: continued
Confirmed pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variants*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar a ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P2, F	26	aff sib (P1)	Upper limb clumsiness	+	+	-	-	-	+	+	+	-	+	+		
P3, M	teens	aff sib (P4)	Upper limb clumsiness	+	+	+	+	+	+	-	-	+	-	-	*SACS*	hemizygous c.13048G>T: p.Glu4350*; 0.7Mb deletion on Chr13q12.12[s]
P4, M	Child hood	aff sib (P3)	Walking delay	+	+	+	-	+	+	-	-	+	+	+		
P5, M	40	aff sib (P6)	Gait disturbance	+	-	-	-	-	-	-	+	-	-	+	*SACS*	c.1580C>G:p.Ser527* c.6781C>A:p.Leu2261Ile[s]

Table 4.5: continued
Confirmed pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variants*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar a ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P6, F	40	aff sib (P5)	Gait disturbance	+	+	-	-	-	+	-	-	-	-	-		
P7, F	23	aff sib (P8)	Speech, Balance	+	+	-	-	+	-	-	-	-	+	-	*KCNC3*	het. c.1259G>A:p.Arg420His[s]
P8, F	57	aff sib (P7)	Upper limb clumsiness	+	+	-	-	+	-	-	-	-	+	-		
P9, M	30	aff sib (P10)	Upper limb clumsiness	+	+	+	-	+	+	-	-	+	-	+	*SPG7*	c.1529C>T:p.Ala510Val; c.1715C>T:p.Ala572Val

Table 4.5: continued
Confirmed pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variants*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar a ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P10, F	29	aff sib (P9)	Upper limb clumsiness	+	-	+	-	-	+	-	-	+	+	-		
P11, F	infancy	aff sib (P12), mother (P13)	Delayed motor and speech development	+	-	+	-	+	+	-	-	-	+	-		
P12, F	infancy	aff sib (P11), mother (P13)	Delayed motor and speech development	+	-	+	+	+	+	-	-	-	+	-	*TUBB4A*	het c.900C>A:p.Met300Ile[s]
P13, F	30	aff children (P11, P12)	Gait disturbance, slurred speech	+	-	+	-	+	+	-	-	+	+	-		mosaic c.900C>A:p.Met300Ile[s]

Table 4.5: continued
Confirmed pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variants*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar a ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P14, F	5	None	Learning difficulties, gait disturbance	+	-	-	+	-	+	-	-	+	-	-	*TUBB4A*	het c.1091C>A:p.Ala364Asp[s]
P15, F	21	aff sib (P16)	Gait disturbance	+	-	-	+	-	+	-	-	+	-	-	*NPC1*	c.467T>C: p.Met156Thr c.709C>T:p.Pro237Ser[s]
P16, F	21	aff sib (P15)	Upper limb clumsiness	+	+	-	+	-	+	-	-	+	-	-		

Table 4.5: continued
Confirmed pathogenic

ID, Sex	***Age of onset (y)***	***Family History***	***Symptom at onset***	***Clinical signs***						***Investigations***					***Gene***	***Variants***
					Eyes					***Electrophysiology***			***MRI***			
				Cerebellar a ataxia	***Jerky ocular pursuit***	***Gaze evoked nystagmus***	***Hypometric saccades***	***Optic atrophy***	***Cerebellar dysarthria***	***Axonal sensorimotor neuropathy***	***Demyelinating sensorimotor neuropathy***	***Normal***	***Cerebellar atrophy***	***Generalised atrophy***		
P17, M	30	aff mother, aunt, cousin (P18)	Speech disturbance Upper limb clumsiness	+	+	-	+	-	+	-	-	-	-	-	*SLC1A3*	het c.1361G>A:p.Arg454Gln[s]
P18, F	39	aff mother, cousin (P17)	Speech disturbance	+	+	-	-	-	+	-	-	-	-	-		

Table 4.5: continued
Possibly pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variant*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar Ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P19, M	32	None	Gait disturbance	+	-	-	-	-	-	+	-	-	-	+	*ZFYVE26*	c.2338C>T:p.Arg780* c.2450delT:p.Leu817Cysfs*12
P20, F	40	None	Gait disturbance	+	-	-	-	+	+	-	-	-	-	-	*WFS1*	c.577A>C:p.Lys193Gln; c.1367G>A:p.Arg456His
P21, F	infancy	None	Walking delay	+	-	+	-	+	+	-	-	-	-	-	*FASTKD2*	c.-66A>G; c.149A>G:p.Lys50Arg[s]

Table 4.5: continued
Possibly pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variant*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar Ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P22, M	37	aff son (P23)	Upper limb clumsiness	+	-	+	-	-	+	-	-	-	+	-	*ZFYVE27*	c.805-2A>G[s]
P23, M	35	aff father (P22)		+	-	-	-	-	+	-	-	+	+	-		
P24, M	25	aff father (P22) aff half-brother (P23)		+	+	+	+	-	+	-	-	-	+	-		

Table 4.5: continued
Possibly pathogenic

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variant*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar Ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P25, F	11	None	Upper limb clumsiness	+	-	+	-	-	+	-	-	+	-	-	*WNK1*	c.1994C>T:p.Thr665Ile[s] c.3272C>T:p.Thr1091Ile[s]

Table 4.5: continued
Uncertain significance or no candidate variants found

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variant*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar Ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P26, F	13	None	Upper limb clumsiness	+	+	+	+	-	+	-	-	-	-	-	*KCNB2*	c.1589C>T: p.Ser530Phe; c.2351T>C:p.Leu784Pro
P27, M	20	2 aff childre n	Gait disturbance	+	+	+	-	-	+	-	-	-	+	-	*ABCB7* *POLG* *WFS1*	hom (hemi) c.818G>A:p.Arg273Gln[s] c.2243G>C:p.Trp748Ser; c.3428A>G:p.Glu1143Gly het c.1294C>G:p.Leu432Val
P28, F	child hood	aff cousin	Gait disturbance	+	+	+	-	-	-	-	-	-	+	-	No candidate variants	
P29, F	40	aff child (P30)	Gait disturbance	+	+	+	-	-	-	-	-	-	-	+		

Table 4.5: continued
Uncertain significance or no candidate variants found

ID, Sex	*Age of onset (y)*	*Family History*	*Symptom at onset*	*Clinical signs*						*Investigations*					*Gene*	*Variant*
					Eyes					*Electrophysiology*			*MRI*			
				Cerebellar Ataxia	*Jerky ocular pursuit*	*Gaze evoked nystagmus*	*Hypometric saccades*	*Optic atrophy*	*Cerebellar dysarthria*	*Axonal sensorimotor neuropathy*	*Demyelinating sensorimotor neuropathy*	*Normal*	*Cerebellar atrophy*	*Generalised atrophy*		
P30, M	21	aff mother (P29)	Gait disturbance	+	+	-	-	-	-	+	-	-	+	-	No candidate variants	
P31, F	46	None	Gait disturbance	+	+	-	-	-	+	-	-	-	-	-		
P32, M	31	None	Upper limb clumsiness	+	+	+	-	-	+	-	-	-	+	+		
P33, M	teens	aff sib (P34)	Upper limb clumsiness	+	-	+	-	-	+	-	-	+	+	-		
P34, M	11	aff sib (P33)	Upper limb clumsiness	+	+	+	+	+	+	-	-	+	+	-		
P35, F	20	Mother has hypotonia	Gait disturbance	+	+	+	-	+	-	-	-	-	+	-		

4.3.2. Population frequency data and pathogenicity scores for variants found in 27 patients with ataxia

Population frequency data and pathogenicity scores for variants found in 27 patients with ataxia were previously published (Pyle *et al.*, 2014) and are presented in Chapter 3 of this Thesis (Table 3.4).

4.3.3. Confirmed pathogenic variants

Confirmed pathogenic variants were found in 9 out of 22 families (41% of families) (Pyle *et al.*, 2014) (presented in Table 4.5 and Appendix 1, and Chapter 3 (Table 3.4)).

SACS

Novel compound heterozygous *SACS* mutations were identified in three unrelated families (Appendices 1 and 2). Each family had a pair of affected siblings who shared mutations. Patients P1 and P2 were compound heterozygous for p.Thr692Thr fs*713 and p.Gly1322Val fs*1343 (Pyle *et al.*, 2012), patients P3 and P4 shared p.Glu4350* and a 0.7Mb deletion on Chr13q12.12, which included *SACS* (Pyle *et al.*, 2013) and finally, patients P5 and P6 shared mutations p.Ser527* and p.Leu2261Ile.

KCNC3

A known dominant *KCNC3* mutation p.Arg420His (Waters *et al.*, 2006; Figueroa *et al.*, 2010) was found to segregate with ataxia in four members of a three-generation autosomal dominant pedigree (patients P7 and P8), making this family the 5th worldwide to date with a p.Arg420His mutation.

SPG7

Previously reported compound heterozygous *SPG7* mutations (p.Ala510Val and p.Ala572Val) (Casari *et al.*, 1998) were found in two affected siblings from a family with no spasticity (patients P9 and P10).

TUBB4A

Likely *de novo* dominant *TUBB4A* mutations (Simons *et al.*, 2013) were found in two families. In one family, siblings P11 and P12 were heterozygous for p.Met300Ile, whereas their affected mother (P13) was mosaic for this mutation. In the other family,

the affected individual (P14) had a p.Ala364Asp mutation. Detailed investigations of pathogenicity of these mutations are presented in Chapter 6.

NPC1

Two siblings (P15 and P16) presenting with adult onset ataxia had compound heterozygous mutations in *NPC1* (p.Met156Thr and p.Pro237Ser). Subsequent oxysterol analysis confirmed pathogenicity of these mutations (Table 4.6; performed by Kim Barlett, Newcastle upon Tyne Foundation Hospitals NHS Trust) (Carstea *et al.*, 1997).

Table 4.6: Plasma oxysterol measurements in P15 and P16. Reproduced from Pyle *et al.*, 2014

Patient	*7 alpha (ng/ml)*	*7 beta (ng/ml)*	*7 keto (ng/ml)*
P15	21.3	12.9	33.9
P16	77.1	60.6	167.4
Normal Range	2.6-53.6	2.9-10.8	2.4-29.1

SLC1A3

A novel dominant p.Arg454Gln mutation, in a known ataxia gene *SLC1A3,* segregated with ataxia in two members of a family (P17 and P18) (de Vries *et al.*, 2009). Patient P17 had an affected mother.

4.3.4. Possible pathogenic variants

Possible pathogenic variants were found in 5 out of 22 families (23% of families) (Pyle *et al.*, 2014) and the result is presented in Appendices 1, 2 and Chapter 3 (Table 3.4).

ZFYVE26

One patient (P19) had a compound heterozygous mutation (p.Leu817Cysfs*12 and p.Arg780*) in *ZFYVE26* (SPG15). The p.Arg780* mutation occurs in a highly conserved residue and is predicted to be disease causing or damaging by three prediction programmes. The parental samples were unavailable for analysis. However, we were able to show, via cloning, that the variants were inherited *in trans*, supporting a recessive inheritance model (this investigation is presented in Chapter 7). This led the clinicians to re-assess the neuroimaging and confirm the thin corpus callosum, which is a characteristic of SPG15 (Goizet *et al.*, 2009).

WFS1

Previously described compound heterozygous *WFS1* mutations (p.Lys193Gln and p.Arg456His) were identified in one family (patient P20) (Cryns *et al.*, 2003). Both mutations occur at conserved residues and have been previously linked to Wolfram syndrome.

FASTKD2

One patient (P21) had compound heterozygous *FASTKD2* mutations (p.Lys50Arg, which is conserved among species, and c.-66A>G predicted to affect normal splicing of exon 1) (Ghezzi *et al.*, 2008). Sanger sequencing revealed segregation of the variants with the disease. Splice predictions for *FASTKD2* (NM_001136194.1):c.-66A>G - [c.-165 (Exon 1c) - c.-51+85 (Intron 1c)] show a gain of acceptor site at c.-65 by one program and loss of acceptor site at –c.64 by another program (Alamut v2.4, Interactive Biosoftware, Rouen, France) (Figure 4.1 A). The *in silico* splice-site prediction for the c.-66A>G mutation was validated with sequencing analysis of cDNA clones (this investigation is presented in Chapter 7).

.

Figure 4.1: Splice site prediction with Alamut software A: splice predictions for *FASTKD2* (NM_001136194.1):c.-66A>G - [c.-165 (Exon 1c) - c.-51+85 (Intron 1c)] show a gain of acceptor site at c.-65 by one program and loss of acceptor site at –c.64 by another program (Alamut v2.4, Interactive Biosoftware, Rouen, France). B: splice predictions for *ZFYVE27* c.805-2A>G show loss of acceptor site of exon 7 in five splice prediction programs (Alamut v2.4, Interactive Biosoftware, Rouen, France).

ZFYVE27

A predicted splice site mutation (c.805-2A>G) was detected in three members of an autosomal dominant pedigree (P22, P23 and P24) in the *ZFYVE27* gene, which was previously described in connection with spasticity, but not ataxia (Mannan *et al.*, 2006). Splice predictions for *ZFYVE27* c.805-2A>G show loss of acceptor site of exon 7 in five splice prediction programs (Alamut v2.4, Interactive Biosoftware, Rouen, France) (Figure 4.1 B). Detailed investigation of the possible pathogenicity of this mutation is presented in Chapter 5.

WNK1

Novel compound heterozygous mutations (p.Thr665Ile and p.Thr1091Ile) were found in *WNK1* in one patient (P25). The parents were both asymptomatic, each carrying an alternative or different allele, supporting the recessive inheritance model. The p.Thr665Ile mutation occurs in a highly conserved residue and is predicted to be pathogenic by four pathogenicity prediction programs. *WNK1* is a known disease gene, which has not been previously associated with ataxia. Mutations in *WNK1* were previously reported in cases with hereditary sensory and autonomic neuropathy type II (Lafreniere *et al.*, 2004; Shekarabi *et al.*, 2008).

4.3.5. Variants of uncertain significance or no candidate variants found

Variants of uncertain significance were found in 2 families (patients P26 and P27) and no candidate variant was found in 6 families (8 patients). This study, therefore, was not able to identify likely causative variants in 8 families (36% of families). The potential candidate variants are presented in Table 4.5 and Appendix 1. A list of variants in patients where we were not able to confidently identify likely candidates was previously published (Pyle *et al.*, 2014) and is presented in Appendix 2.

4.4. Discussion

4.4.1. Diagnostic yield

This exome sequencing study identified confirmed pathogenic variants in 9 out of 22 families (41%). Possible pathogenic variants were found in 5 out of 22 families (23%). Taken together, this accounts for a likely molecular diagnosis in 14 of 22 families (64%). These findings agree with whole exome sequencing studies in ataxia cohorts performed by other groups. Ohba and co-workers assessed patients with childhood-onset cerebellar ataxia and were able to define a molecular cause of the disease in 9 out of 23 families (39%) (Ohba *et al.*, 2013). Another study assessed the diagnostic yield of whole exome sequencing in 28 families with paediatric ataxia patients without a molecular diagnosis (Sawyer *et al.*, 2014). Molecular cause of the disease was established in 13 out of 28 families (46%). Of note, the lower diagnostic yield in the latter two studies might be due to lack of family history in some cases. Fogel and colleagues (Fogel *et al.*, 2014b) applied WES to a heterogeneous cohort of 76 patients with adult and sporadic onset cerebellar ataxia and were able to establish overall likely molecular diagnosis in 60% of the patients being studied. Interestingly, this study and the study by Fogel *et al*., 2014 shared only 3 known disease genes (*SPG7, WFS1* and *ZFYVE26*) as a likely cause of the disease.

This study, and the work by others mentioned above, contrasts with a targeted exome sequencing study in an ataxia cohort by Nemeth and colleagues (Nemeth *et al.*, 2013), which had an overall detection rate of 18% in a similar heterogeneous ataxia cohort. The possible explanation is greater genomic coverage and unbiased approach of the whole exome sequencing.

Importantly, this study is endorsed not only by exome sequencing studies in ataxia, but also by similar studies in other neurological disorders. Srivastava and co-workers assessed the diagnostic yield of whole exome sequencing in 78 patients with neurodevelopmental disorders and found an overall presumptive cause of the disease in 41% of the patients (Srivastava *et al.*, 2014). Similarly high diagnostic yield was documented in cohort studies with Charcot-Marie-Tooth disease (total yield = 53%) (Klein *et al.*, 2014) and hereditary spastic paraplegia (diagnostic yield =75%) (Novarino

et al., 2014). Of note, the higher diagnostic rate in the latter study might be due to exome sequencing parent-child trios in some cases. These samples were not always available in this study.

At the time this work commenced the Exome Aggregation Consortium (ExAC) data was not available and was not used. ExAC is a dataset which provides exome sequencing data for over 60,000 unrelated control individuals (available at http://exac.broadinstitute.org) The current study subsequently reviewed ExAC data and the outcome is presented in Table 4.7. The population frequency data generated using ExAC database are in line with other databases (Table 3.7) for all variants found in 27 patients in this study.

Table 4.7: Population frequency data generated from ExAC database for variants found in 27 patients with ataxia

			Population Data	
P ID	***Gene***	***Variant***	***Allele count***	***Allele frequency***
P1, P2	*SACS*	p.Thr692Thr fs*713	-	-
		p.Gly1322Val fs*1343	-	-
P3, P4	*SACS*	p.Glu4350*	-	-
P5, P6	*SACS*	p.Ser527*	-	-
		p.Leu2261Ile	584/120970	0.004828
P7, P8	*KCNC3*	p.Arg420His	-	-
P9, P10	*SPG7*	p.Ala510Val	306/121348	0.002522
		p.Ala572Val	1/121216	0.00000825
P11, P12, P13	*TUBB4A*	p.Met300Ile	-	-
P14	*TUBB4A*	p.Ala364Asp	-	-
P15	*NPC1*	p.Met156Thr	10/121362	0.0000824
P16		p.Pro237Sers	1262/120672	0.01046
P17,P18	*SLC1A3*	p.Arg454Gln	-	-
P19	*ZFYVE26*	p.Leu817Cys fs*12	-	-
		p.Arg780*	-	-
P20	*WFS1*	p.Lys193Gln	487/120064	0.004056
		p.Arg456His	6908/121250	0.05697
P21	*FASTKD2*	c.-66A>G	-	-
		p.Lys50Arg	304/120810	0.002516
P22, P23, P24	*ZFYVE27*	c.805-2A>G	3 / 121378	0.00002472
P25	*WNK1*	p.Thr665Ile	7755/121386	0.06389
		p.Thr1091Ile	-	-
P26	*KCNB2*	p.Ser530Phe	2/121214	0.0000165
		p.Leu784Pro	-	-
P27	*ABCB7*	p.Arg273Gln	-	-
	POLG	p.Trp748Ser	-	-
		p.Glu1143Gly	3410/121226	0.02813
	WFS1	p.Leu432Val	439/121400	0.003616

4.4.2. Heterogeneity of inherited ataxia

This study identifies 11 genes implicated in 14 families. No recurrent mutations in the same gene were observed (Pyle *et al.*, 2014). This result reflects the high genetic heterogeneity of hereditary ataxia.

4.4.3. Whole exome sequencing vs "ataxia multi-gene panel"

This study and the work by other groups suggest that one of the disadvantages of the targeted exome sequencing is a lower diagnostic yield. Diagnostic yield in this study (64%) and similar studies (39-60%) contrasted with a recent targeted exome sequencing study (diagnostic yield =18%) (Nemeth *et al.*, 2013). This could be due to a greater exome coverage, which allows identification of "new" ataxia genes, which were not previously considered as "known" ataxia genes and therefore were not included in the panel. Overlooking such genes is one of the disadvantages of targeted exome sequencing. An example of a gene, which was not included in the ataxia gene panel by Nemeth *et al*., 2013, but was found to be a definite cause of ataxia in one family in this study and in a similar work (Fogel *et al.*, 2014b), is *SPG7.* Moreover, both findings of this study and the work by Fogel *et al*., 2014, prompted a study in a cohort of 70 patients with adult onset unexplained ataxia (Pfeffer *et al.*, 2015), which identified mutations in *SPG7* as a cause of unexplained ataxia in 19% of the patients. This establishes *SPG7* mutations as a common cause of unexplained ataxia. The latest published ataxia gene panel (Nemeth *et al.*, 2013) included 118 known and speculative ataxia genes. In the current study an exome coverage analysis was performed for all these genes in all the patients and found that 89% of the coding regions of these genes were covered 20-fold (Chapter 3.3.2). This study, therefore, assumes that it would be able to detect the majority of the mutations identified by Nemeth *et al*., 2013. However, it is pointed out that only 29% of the disease genes identified in this study were included in the multi-gene panel of Nemeth *et al.,* 2014 (Pyle *et al.*, 2014).

The other disadvantage of targeted exome sequencing over whole exome sequencing is that different diagnostic laboratories might provide different gene panels with limited overlap. This could be due to the variations in the local demographics, which would affect prevalence of different types of ataxia in different populations. Therefore, various

laboratories might prefer to include different genes in their panels, reflecting the need of the local population. Also, the list of genes included in the panel is based on the current knowledge of ataxia genes and therefore is far from being complete. Considering the whole exome may help to remove these biases.

One of the advantages of whole exome sequencing is the identification of known disease genes in an unusual clinical setting, broadening the clinical phenotype of known disease genes. This study identified known pathogenic mutations in *SPG7* in two siblings (patients P9 and P10) from a recessive family, presented with ataxia but no spasticity. *SPG7* is a gene implicated in hereditary spastic paraplegia (Schule and Schols, 2011). Although *SPG7* previously has been linked to ataxia, it has never been described in cases without spasticity. In another family, two siblings (patients P15 and P16) were found to have compound heterozygous mutations in *NPC1*. Mutations in this gene cause Niemann Pick type C, a neurological disorder with characteristic eye movement. Patients presented in this study did not exhibit this feature. However, exome sequencing result of this study was supported by the biochemical finding (Table 4.5), which helped to establish the diagnosis. (Carstea *et al.*, 1997). In another patient (P19) this study identified compound heterozygous mutation in *ZFYVE26* known to cause spastic paraplegia type 15 (SPG15). The current study prompted a review of the neuroimaging in this patient and confirmed thin corpus callosum, which is a characteristic of SPG15.

Given that the cost of sequencing of a single candidate gene is comparable to whole exome sequencing (Pyle *et al.*, 2014) and to a targeted capture of >50 genes (Nemeth *et al.*, 2013), exome sequencing may be a faster and more cost effective method to discover disease causing gene mutations. Coupled with the continuing decrease in the cost of whole exome sequencing, this technique could be more advantageous in a clinical setting.

4.4.4. Possible explanations why did we not solve all of cases

This study was unable to identify the likely cause of ataxia in 36% of families. Given the relevant family history in most cases, this could be due to technical issues, for example, the capture probes being ineffective for GC-rich or repetitive regions, which

could result in incomplete coverage. Furthermore, large genomic rearrangements and trinucleotide repeat sequences are not reliably detected from exome-capture data. Finally, this study has not considered non-coding regions. It is likely that some causal variants will reside within non-coding regulatory regions. Some of these issues could be resolved by whole genome sequencing, albeit with additional cost. WGS produces over 100 Gb of data. The main difficulty of WGS is interpretation of variations found in non-coding regions due to lack of knowledge about the importance of such regions (Ward and Kellis, 2012). Therefore, the consequences of variations in non-coding regions of the genome are less well understood than variations in coding regions.

Chapter 5

Molecular genetic analysis and determining pathogenicity of a heterozygous splice site mutation in *ZFYVE27* in an autosomal dominant pedigree

Chapter 5. Molecular genetic analysis and determining pathogenicity of a heterozygous splice site mutation in *ZFYVE27* in an autosomal dominant pedigree

5.1. Introduction

Using whole exome sequencing we identified a predicted splice site mutation in *ZFYVE27* (zinc finger, FYVE domain containing 27) gene (c.805-2A>G), which segregated with the disease in three members of an autosomal dominant pedigree (patient P22, P23 and P24). Mutations in *ZFYVE27* have not been associated with ataxia before. Mannan and colleagues (Mannan *et al.*, 2006) identified a p.Gly191Val mutation in *ZFYVE27* (SPG33) and associated it with the clinical phenotype in a German family with autosomal dominant hereditary spastic paraplegia (AD-HSP), a neurological disease characterised by selective degeneration of axons. The most common form of AD-HSP is due to heterozygous mutations in the spastin (*SPG4*) gene (Hazan *et al.*, 1999). Mannan and co-workers reported ZFYVE27 as a specific spastin-binding protein and suggested that ZFYVE27 binds spastin through its C-terminus via a short motif within the FYVE domain. Intriguingly, using a co-immunoprecipitation experiment, the authors showed reduced binding of the p.Gly191Val mutant ZFYVE27 to spastin. The p.Gly191Val mutation was found to segregate with the disease in the family and was not present in 210 control chromosomes. The mutation falls within the third transmembrane domain of the protein. It was suggested that the mutation shifts the transmembrane domain motif forward and thus alters the conformation of ZFYVE27 in the region of the FYVE domain. They thus proposed that the failure of ZFYVE27 to interact with spastin was potentially a cause of the disease. However, it should be pointed out that the current study subsequently investigated the frequency of the p.Gly191Val mutation on ExAC database (available at http://exac.broadinstitute.org) and found that the p.Gly191Val is a common variant (allele count 1329/121408; MAF=0.01095). No other *ZFYVE27* pathogenic mutations were reported in other families so far.

Shirane and Nakayama (Shirane and Nakayama, 2006) identified ZFYVE27 as a mammalian protein that promotes neurite formation through the interaction with a

guanosine diphosphate (GDP)-bound form of Rab11. Rab11 takes part in the regulation of membrane trafficking during neurite formation and has a specific role in membrane recycling. The authors showed that ZFYVE27 became phosphorylated by extracellular signal-regulated kinase (ERK) in response to nerve growth factor (NGF) and this leads to interaction between ZFYVE27 and Rab11. Down-regulation of ZFYVE27 inhibited neurite formation.

Martignoni and colleagues (Martignoni *et al.*, 2008) assessed the ability of wildtype and p.Gly191Val-mutant ZFYVE27 to extend neurites in immortalised murine motor neuronal cell lines and the ability to elongate HeLa cells, but found no difference between the wildtype and the mutant. The p.Gly191Val mutant of ZFYVE27 was also found to retain the ability to bind to Rab11-GDP and spastin indicating that the p.Gly191Val mutation does not result in the loss of function. These data question the contribution of the p.Gly191Val mutation to the disturbance of the interaction between ZFYVE27 and spastin, and therefore, the involvement of ZFYVE27 mutation in the disease phenotype, presented by Mannan et al., is still inconclusive (Mannan *et al.*, 2006).

Saita and colleagues (Saita *et al.*, 2009) found that ZFYVE27 associates and co-localizes with a vesicle-associated membrane protein isoform A (VAP-A), which is expressed in the endoplasmic reticulum (ER). The VAP-A-binding domain of ZFYVE27 was found to be the prerequisite for the ZFYVE27-dependent formation of membrane protrusions in HeLa cells. Furthermore, depletion of VAP-A by RNA interference lead to the inhibition of nerve growth factor (NGF)-induced neurite outgrowth in the rat pheochromocytoma PC12 cells. Subsequent forced expression of human wildtype VAP-A, but not the VAP-A mutants that lost their ability to bind ZFYVE27, corrected the defect. Altogether, these results suggest that the interaction of ZFYVE27 with VAP-A is, firstly, an important regulator of subcellular localisation of ZFYVE27 and secondly, is essential for the ability of ZFYVE27 to stimulate neurite formation.

A study by Zhang and co-workers (Zhang *et al.*, 2012) supported the work by Martignoni and colleagues, in finding no difference in the ability of wildtype ZFYVE27

and the p.Gly191Val mutant to extend neurites in PC12 cells or to elongate the motor neuron axon of zebrafish embryos. They also showed that spastin promotes ZFYVE27-dependent neurite outgrowth in PC12 cells and spinal cord motor neuron axon outgrowth in zebrafish and proposed that spastin and ZFYVE27 may work together in promoting motor axon outgrowth. They also suggested that the N-terminal region of ZFYVE27, and not the FYVE domain, mediated the interaction with spastin. Chang and co-workers proposed a membrane topology model for the ZFYVE27 protein, protrudin (Figure 5.1) (Chang *et al.*, 2013).

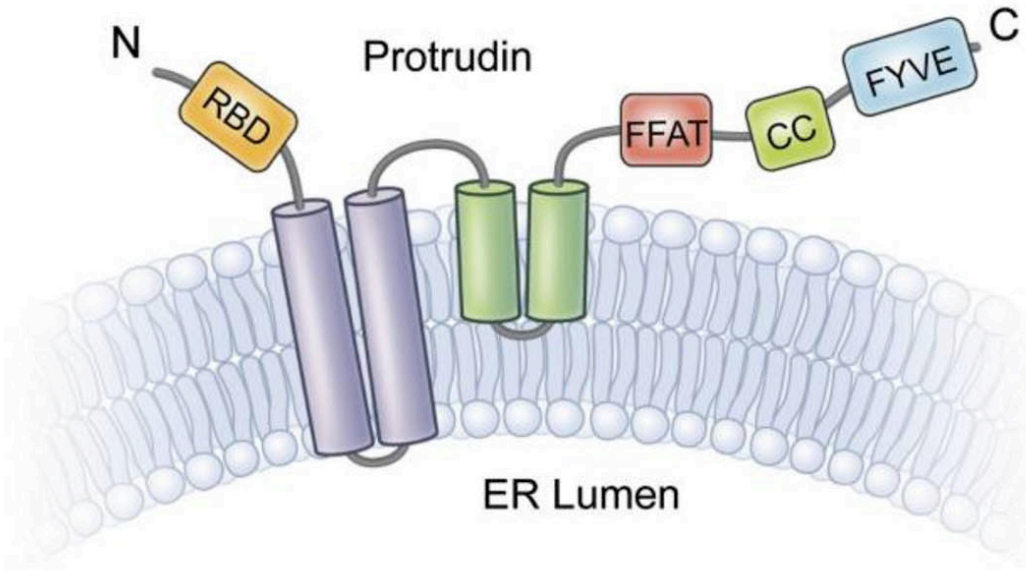


Figure 5.1: Membrane topology model for ZFYVE27. ZFYVE27 contains a Rab11-binding domain (RBD). Three trans-membrane domains are followed by FFAT (two phenylalanines in an acidic tract) domain, an endoplasmic reticulum targeting signal. On its C-terminus protrudin contains a coiled coil (CC) domain followed by the FYVE (Fab1, YOTB, Vac1, and EEA1) domain (adapted from Chang *et al*., 2013).

The RBD domain is necessary for interaction with RAB11A, which promotes neurite outgrowth (Shirane and Nakayama, 2006). The FFAT domain is necessary for interaction with VAP-A (vesicle-associated membrane protein-associated protein isoform A) (Saita *et al.*, 2009). VAP-A is expressed in the ER and is responsible for subcellular localisation of ZFYVE27. The FFAT domain also mediates the interaction of protrudin with microtubule-associated molecular motor KIF5 (Matsuzaki *et al.*,

2011). Mutated *KIF5A*, a neuron-specific isoform of *KIF5*, was identified as a cause of autosomal dominant hereditary spastic paraplegia type 10 (SPG10) (Reid *et al.*, 2002). Interaction of ZFYVE27 with KIF5A plays an important role in vesicular trafficking in neurons (Matsuzaki *et al.*, 2011). The FYVE domain is suggested to be responsible for bringing proteins to the endosome (Stenmark *et al.*, 2002).

In a recent paper, investigating expression and function of *ZFYVE27*, the authors found that in mice the region corresponding to human exon 8 (21bp) is present only in a transcript expressed in neuronal tissue (cerebrum and spinal cord) (Ohnishi *et al.*, 2014). The authors named the transcripts, including and lacking this 21 bp fragment, as Protrudin-L and Protrudin-S, respectively. Quantitative RT-PCR analysis of various mouse tissues revealed that Protrudin-L mRNA was expressed in the cerebrum and spinal cord but it was absent in the liver, heart, kidney, lung, thymus, spleen, stomach and the intestine. In brain the highest expression of Protrudin-L mRNA was in the cerebellum followed by the midbrain and the medulla. The authors concluded that Protrudin-L is a neuron-specific isoform. Importantly, this 21bp fragment encodes a VAP-A binding domain. VAP-A is expressed in ER and is responsible for subcellular localisation of ZFYVE27. Knock-in and knock-down experiments in PC12 cells highlighted the ability of VAP-A to stimulate neurite outgrowth (Saita *et al.*, 2009).

The mechanism by which ZFYVE27 regulates polarised endosome translocation and neurite outgrowths has recently been proposed (Raiborg *et al.*, 2015).

There are seven protein coding transcripts of the *ZFYVE27* gene. (Table 5.1). As mentioned previously, the encoded protein is called protrudin. Protrudin belongs to the FYVE-finger family, with the majority of its members serving as regulators of the endocytic membrane trafficking (Stenmark *et al.*, 2002).

Table 5.1: *ZFYVE27* RNA transcripts and corresponding protein isoforms. The table depicts *ZFYVE27* RNA transcript variants with corresponding NCBI reference sequences for mRNA and protein isoforms.

Transcript variant	NCBI Reference Sequence (mRNA)	Length (bp)	NCBI Reference Sequence (protein)	Length (aa)
1	NM_001002261	2858	NP_001002261	416
2	NM_144588	3047	NP_653189	411
3	NM_001002262	2997	NP_001002262	404
4	NM_001174119	2927	NP_001167590	372
5	NM_001174120	2765	NP_001167591	318
6	NM_001174121	2769	NP_001167592	311
7	NM_001174122	2669	NP_001167593	286

This study aims to elucidate pathogenicity of the *ZFYVE27* c.805-2A>G mutation and establish its causative role in a dominant family presenting with cerebellar ataxia. In order to do this, we will perform transcription and protein expression studies.

5.2. Materials and Methods

5.2.1. In silico splice-site prediction with Alamut software

In silico prediction of pathogenicity of the putative causative variant was performed using Alamut software as described in Section 4.2.4.

5.2.2. Sequencing of fibroblast cDNA (total PCR product)

Reverse transcription PCR (RT-PCR) was performed as described in Section 2.9 using *ZFYVE27* cDNA specific primers (Table 5.2).

Table 5.2: **Primers for RT-PCR**

Primer	Sequence
Forward	5′- TGTGGAGTTCTTCCGAGTTGTGT-3′
Reverse R1	5′- AGCCCGTGCAGTTCCCGAAG-3′
Reverse R2	5′- CAGCACCACCAAGTGGGTCTC-3′

PCR conditions involved a hot start at 95°C for 4 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 53°C for 30 seconds, and extension at 72°C for 1 minute. A final extension step was performed at 72°C for 10 minutes, followed by cooling and storage of the sample at 4°C. Sanger sequencing of total PCR product was performed using *ZFYVE27* cDNA specific primers (Table 5.2), as described in Section 2.5.

5.2.3. Emetine inhibition of nonsense-mediated decay in fibroblasts from controls and patient P23

The *ZFYVE27* c.805-2A>G mutation was predicted to affect mRNA splicing by causing downstream exon skipping. This, in turn, could introduce a frame shift and result in the introduction of a premature stop codon. Premature termination codons, which account for approximately one third of mutations causing human disorders, have been shown to trigger degradation of mutant transcripts by the nonsense-mediated decay pathway (Frischmeyer and Dietz, 1999). Treatment of a patient's fibroblasts with 100 µg/ml emetine dihydrochloride hydrate has been shown to successfully inhibit the nonsense-mediated decay pathway (Noensie and Dietz, 2001). Inhibition of the nonsense-mediated decay pathway would allow read-through of premature stop codons and expression of the transcripts, that otherwise would be eliminated by the nonsense-mediated decay pathway. In order to elucidate whether the c.805-2A>G mutation leads to nonsense-mediated decay, control and patient fibroblasts were treated with emetine dihydrochloride hydrate in order to inhibit nonsense-mediated decay. cDNA (total PCR product) was sequenced with the aim to study the effect of the candidate mutation on the possible exon skipping and activation of the nonsense mediated decay pathway. The outline of this experiment is presented in Figure 5.2. This study hypothesized that by inhibiting the nonsense mediated decay pathway and allowing read-through of a premature stop codon in the patient, it would be possible to amplify an additional *ZFYVE27* RNA transcript by means of RT-PCR. Detecting an additional (aberrant) transcript in the patient's sample would indicate that the *ZFYVE27* c.805-2A>G mutation leads to nonsense-mediated decay

In order to elucidate whether the c.805-2A>G mutation leads to nonsense-mediated decay, this study aimed to inhibit nonsense-mediated decay in fibroblasts from controls and the patient by treating them with emetine. Untreated cells were cultured in parallel. The control and patient's primary fibroblast cell lines were cultured as described in Section 2.8. Cells were grown in the standard fibroblast medium (Section 2.8) with exclusion of uridine, penicillin and streptomycin. After reaching 60-80% confluency, half of the cells were supplemented with 100 µg/mL emetine (Sigma-Aldrich, Dorset, UK) and the other half was left intact. After 10 hours incubation, all cell lines were harvested, total RNA extracted (Section 2.9.1), cDNA generated (Section 2.9.2) and amplified using primers specific for *ZFYVE27* cDNA (Section 2.9.3 and Section 5.2.2). Sanger sequencing of the total PCR product was performed as described in Section 5.2.2. The results for treated and untreated cells were compared.

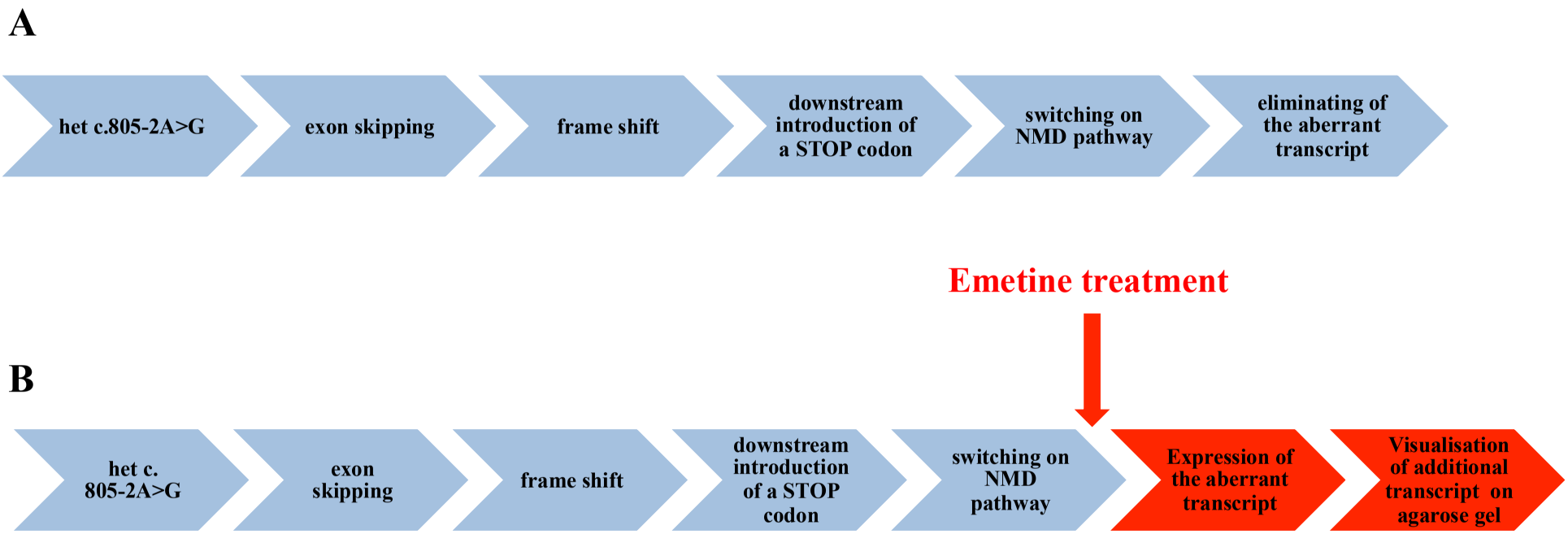


Figure 5.2: Functional study of the effect of the *ZFYVE27* c.805-2A>G mutation on mRNA splicing in fibroblasts. This figure depicts the possible consequence of the *ZFYVE27* c.805-2A>G mutation (A) and the experimental design for elucidating the role of the c.805-2A>G mutation in the nonsense-mediated decay via emetine treatment (B).

5.2.4. Human tissue samples

Control muscle tissue samples were obtained from the Mitochondrial Control Tissue Bank (Newcastle University) from anonymous healthy individuals. Informed consent was obtained from all individuals before skin biopsies were performed. Frozen muscle tissue sections were kindly provided by Dr Florence Burte (Newcastle University).

Post-mortem brain tissue from patient P22 and three age-matched controls were obtained from the Newcastle Brain Tissue Resource (Newcastle University) in accordance with Newcastle University Ethics board. For each individual a sample was obtained from the frontal cortex and the cerebellum. The clinical and neuropathological data were obtained from the Newcastle Brain Tissue Resource. The quality of the post-mortem tissue was assessed by means of pH readings of the starting material (Stan *et al.*, 2006) (pH reading obtained from Newcastle Brain Tissue Resource). Frozen brain tissue sections for the protein expression study were obtained from the Newcastle Brain Tissue Resource.

5.2.5. RNA extraction from human brain tissue

A section of frozen brain tissue (not more than 0.5 cm in the smallest dimension) was placed into a sterile RNase-free 15ml tube (Corning) containing at least 10 volumes of ice cold RNA*later*®-ICE Tissue Transition Solution (Life Technologies Ltd., Paisley, UK) and incubated at -20^{o}C for at least 16 hours. RNA*later*®-ICE Tissue Transition Solution is specifically designed for transition of tissue from frozen to pliable state at -20^{o}C without compromising the RNA integrity. RNA was extracted using RiboPure™ Kit (Life Technologies Ltd., Paisley, UK) according to the manufacturer's protocol. The extracted RNA samples were treated with DNase I in order to remove DNA contamination. This was performed using TURBO DNA-free™ Kit (Life Technologies Ltd., Paisley, UK) as per the manufacturer's manual.

5.2.6. Sequencing of cDNA from the human brain tissue (total PCR product and plasmid DNA cloned from extracted PCR fragments)

Reverse transcription PCR (RT-PCR) was performed as described in Section 2.9 using *ZFYVE27* specific primers (Table 5.2) and PCR conditions outlined in Section 5.2.2.

The efficiency of the cDNA amplification from each tissue specimen was assessed by a control RT-PCR using primers specific to a 133 bp fragment of the housekeeping gene GAPDH (forward: 5'-CTGACTTCAACAGCGACACC; reverse: 5'-ATGAGGTCCACCACCCTGT). RT-PCR was performed as described in Section 2.9.3. PCR conditions were as follows: a hot start at 95^{o}C for 3 minutes followed by 30 cycles of denaturation at 95^{o}C for 30 seconds, annealing at 55^{o}C for 30 seconds and extension at 72^{o}C for 45 seconds. The reaction was finished by a final extension step performed at 72^{o}C for 10 minutes, followed by cooling the sample to 4^{o}C.

Sanger sequencing of the total PCR product was performed as described in Section 5.2.2. Sanger sequencing of plasmid DNA cloned from the extracted PCR fragments was performed as described in Section 2.6.

5.2.7. Protein extraction from the human brain tissue

Post-mortem human brain tissue samples were rapidly thawed and approximately 500mg of tissue was homogenised on ice using a rotor stator homogenizer Polytron® (Thermo Fisher Scientific Inc., Leicestershire, UK) in 2 ml of lysis buffer containing 0.2 M Triethylammonium bicarbonate buffer (TEAB) (Sigma-Aldrich, Dorset, UK) and cOmplete Protease Inhibitor Cocktail (Roche, Indianapolis, USA). Following this, Sodium Dodecyl Sulphate (SDS) (Sigma-Aldrich, Dorset, UK) was added to the homogenate to a final concentration of 0.2% and the sample was gently sonicated for 40 minutes on ice in a sonicating bath. Finally, the homogenate was sonicated for 10 seconds using ultrasonic processor Sonics® (Sonics & Materials, Newtown, USA). Protein concentration was determined using a Bradford assay (Bradford, 1976).

5.2.8. Western blotting

Homogenate samples (Section 5.2.7) containing 20 µg of protein and an equal volume of loading dye (250 mM Tris-HCl, pH 6.8; 20% glycerol; 4% SDS; 0.1% Bromophenol Blue, and 10mM Dithiothreitol (DTT), all Sigma-Aldrich, Dorset, UK) were heated at 95^{o}C for 5 minutes and separated on precast 4-15% Mini-PROTEAN®TGX™ sodium dodecyl sulphate-polyacrylamide gels (SDS-PAGE) (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) with 1 x Tris-glycine running buffer (25mM Tris base; 190 mM glycine; 0.1% SDS, all Sigma-Aldrich, Dorset, UK). The protein gels were transferred to polyvinylidene fluoride (PVDF) membranes (iBlot®2 PVDF stacks; Life

Technologies Ltd., Paisley, UK) using the iBlot®2 Dry Blotting system (Life Technologies Ltd., Paisley, UK). The membranes were blocked in 5% milk in 1 x TBST (Tris buffered saline (Sigma-Aldrich, Dorset, UK), containing 0.2% Tween 20 (Sigma-Aldrich, Dorset, UK)), for one hour at room temperature. Following the blocking step, membranes were incubated with primary antibody (Table 5.3) diluted in 5% milk in 1 x TBST overnight at 4°C. After three washes in 1 x TBST, membranes were incubated with secondary antibody (Table 5.3) diluted in 5% milk in 1 x TBST for one hour at room temperature. Finally, membranes were washed three times in 1 x TBST and developed using ECL blotting substrate (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) following the manufacturer's protocol. The signal was detected using a BioSpectrum®500 Imaging System (UVP, LLC, Cambridge, UK). Protein molecular weight was determined by comparison with a SeeBlue® Plus2 Pre-Stained Standard (Life Technologies Ltd., Paisley, UK) and Biotinylated Protein Ladder (Cell Signaling Technology, Inc., Danvers, USA).

Table 5.3: Antibodies used for the detection of GAPDH, ZFYVE27 and HSPA5 (BiP/GRP78) proteins in human frontal cortex and cerebellum

Antibody	Source	Catalogue number	Dilution
GAPDH (FL-335) rabbit polyclonal (primary)	Santa Cruz	sc-25788 HRP	1/500
ZFYVE27 rabbit polyclonal (primary)	Abnova	PAB23125	1/250
HSPA5 (BiP/GRP78) rabbit polyclonal (primary)	Abnova	PAB2462	1/1000
Polyclonal Swine anti-rabbit Horseradish peroxidase (HRP)-conjugated (secondary)	Dako	p0399	1/2 000

Statistical analysis

Densitometric analysis was performed using ImageJ software (available at http://imagej.nih.gov/ij/). GAPDH was used to normalise the result. Ratios of the protein of interest/GAPDH, means and standard deviations were calculated for all samples. Data represent the mean of three independent replicates. Fold change (FC) was calculated as the ratio of average of ratios (protein of interest/GAPDH) for the patient / average of means of ratios (protein of interest/GAPDH) for the controls.

5.2.9. Immunohistochemistry (IHC) on frozen muscle and human brain tissue sections

Frozen tissue sections were fixed in 4% paraformaldehyde (Sigma-Aldrich, Dorset, UK), permeabilised in 0.5% Triton-X100 (Sigma-Aldrich, Dorset, UK) in phosphate-buffered saline (PBS) (Sigma-Aldrich, Dorset, UK), blocked in 10% goat normal serum (Sigma-Aldrich, Dorset, UK) in PBS (Sigma-Aldrich, Dorset, UK), for 30 minutes at room temperature and incubated with primary ZFYVE27 antibody diluted in 10% goat normal serum (Sigma-Aldrich, Dorset, UK) in PBS (Sigma-Aldrich, Dorset, UK) (Table 5.4) overnight at 4^{o}C.

Table 5.4: Antibodies used for the detection of ZFYVE27 protein in human frontal cortex and cerebellum frozen sections

Antibody	Source	Catalogue number	Dilution
ZFYVE27 rabbit polyclonal (primary)	Sigma	HPA037523	1/200
Biotinylated goat anti-rabbit IgG (secondary)	Vector Laboratories	BA-1000	1/200

After two washes in PBS (Sigma-Aldrich, Dorset, UK), the tissue sections were treated with 0.3% hydrogen peroxide (Sigma-Aldrich, Dorset, UK) in water for 5 minutes in order to quench endogenous peroxidase activity, blocked in 10% goat normal serum (Sigma-Aldrich, Dorset, UK) in PBS (Sigma-Aldrich, Dorset, UK) for 10 minutes and incubated with biotinylated goat anti-rabbit IgG secondary antibody (Vector Laboratories, Peterborough, UK). The target antigen signal was amplified with Vectastain® ABC detection kit (Vector Laboratories, Peterborough, UK). Immunoreactive product was visualised with diaminobenzidine (DAB) substrate for peroxidase (ImmPACT™ DAB Substrate Kit; Vector Laboratories, Peterborough, UK). The IHC protocol procedure was first tested on human control muscle tissue (positive control). Primary antibody was omitted from the procedure to act as a negative control, specifically testing for non-specific binding of the secondary antibody. Images were captured as described in Section 2.10.2.

5.3. Results

5.3.1. In silico mutation pathogenicity study with Alamut software

Splice predictions for *ZFYVE27* c.805-2A>G show loss of acceptor site of exon 7 in five splice prediction programs (Alamut v2.4, Interactive Biosoftware, Rouen, France) (Figure 4.1).

5.3.2. Functional studies of ZFYVE27 *c.805-2A>G mutation in fibroblasts*

Transcriptional studies

Sequencing of ZFYVE27 cDNA (total PCR product)

The schematic representation of *ZFYVE27* RNA transcripts is outlined in Figure 5.3.

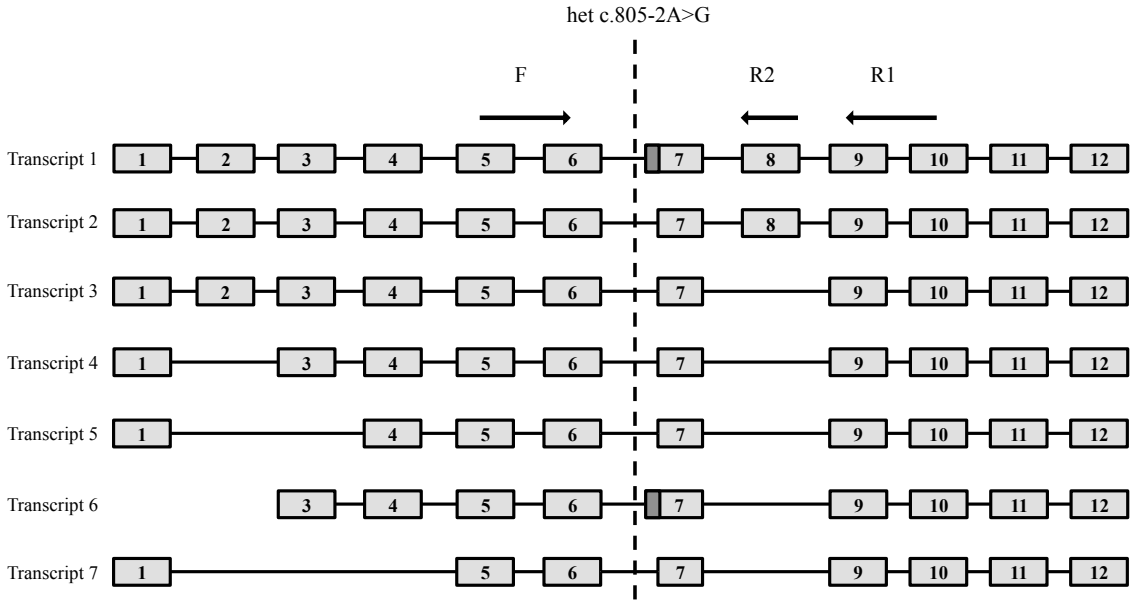


Figure 5.3: *ZFYVE27* RNA transcripts. *ZFYVE27* gene has seven protein coding RNA transcripts as a result of alternative splicing. Numbered boxes represent exons. The dotted line indicates the position of the c.805-2A>G mutation. Exons are aligned for convenience. The forward and reverse primer positions are represented by arrows. Primer R2 spans the entire exon 8 region. Primers F and R1 could potentially amplify PCR products of 425 bp, 410 bp, 404 bp and 389 bp; primers F and R2 could potentially amplify products of 265 bp and 250 bp.

Patient and control cDNA, derived from RNA extracted from fibroblasts, was amplified using F and R1 primers which were specific for *ZFYVE27* cDNA (figure 5.3). Agarose gel electrophoresis of the resulting PCR product revealed the amplification of two products (404 bp and 389 bp) (Figure 5.4).

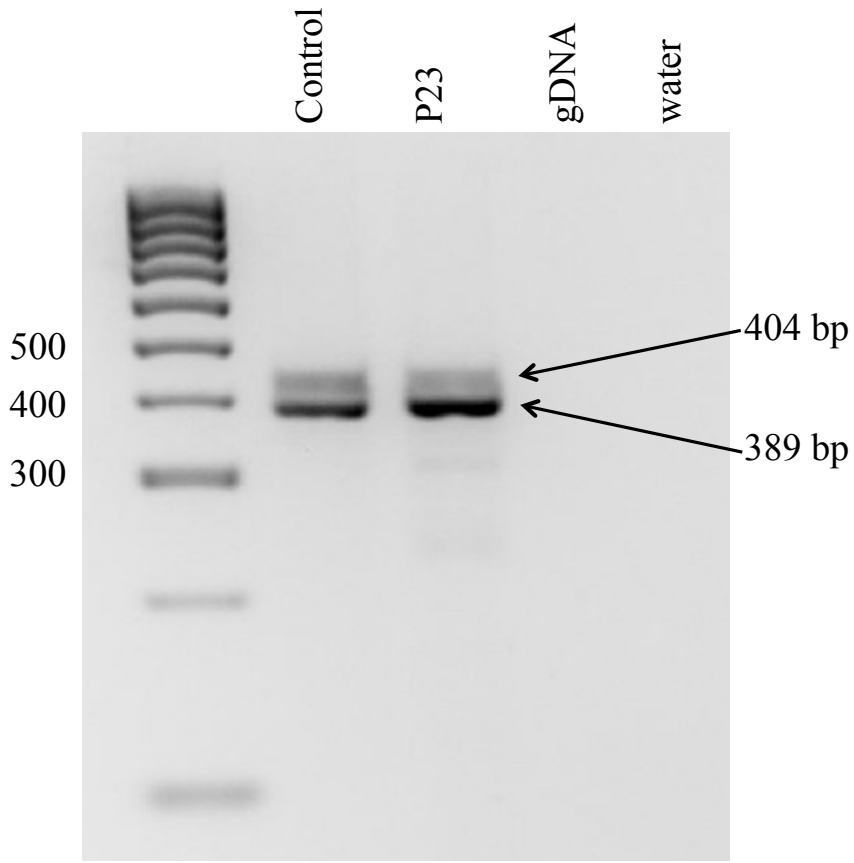


Figure 5.4: Agarose gel (3.5%) electrophoresis of the PCR product amplified from *ZFYVE27* cDNA using primers F and R1 (RNA was derived from fibroblasts of a healthy control and the patient P23)

Sequencing of cDNA (total PCR product) revealed no differences between the control and patient. However, the sequence chromatograms did show overlayed peaks in the 5' region of exon 7 and a homozygous deletion of the entire exon 8 (Figure 5.5). This study therefore concludes that both transcripts 1 and 2 (Figure 5.3) are absent in the fibroblasts of this particular control and patient. The sequencing results confirm that the band of 404 bp corresponds to transcript 6, whereas the band of 389 bp represents transcripts 3, 4, 5 and 7 (these could be present in any combination) (Figure 5.3). Research published after this study have been undertaken (Ohnishi *et al.*, 2014) suggests that in mice the region corresponding to human exon 8 (21bp) is present only in transcripts expressed in neuronal tissue (cerebrum and spinal cord). The findings of this study are in line with these previous observations and provide evidence that human fibroblasts do not express the neuron-specific transcripts 1 and2 of *ZFYVE27* (Figure 5.3).

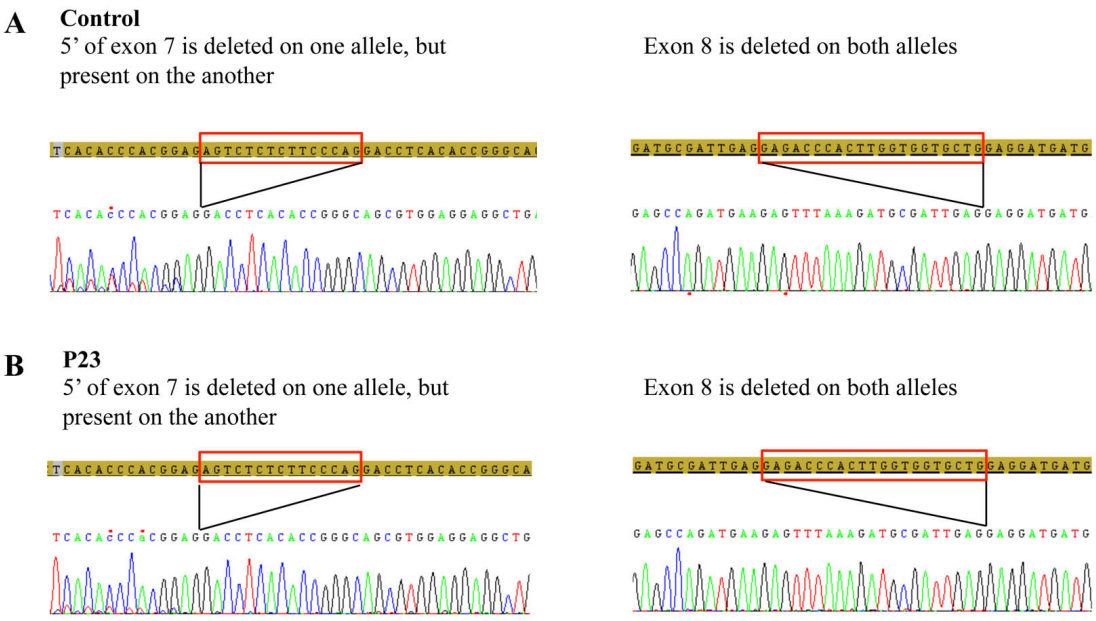


Figure 5.5: Sanger sequencing chromatograms for *ZFYVE27* cDNA generated from the fibroblast RNA of a healthy control (A) and patient P23 (B) using primers F and R1. Red rectangles represent deleted sequences.

Inhibition of nonsense-mediated decay in fibroblasts of controls and patient P23 with emetine compound (sequencing of total PCR product)

Treated and untreated fibroblast cells from controls and patient P23 exhibited different morphology (Figure 5.6). Not treated fibroblasts from controls and the patient were characterised by normal morphology. Treated fibroblasts from controls and the patient exhibited accumulation of intracellular vesicles. This may represent the response to stress (Li *et al.*, 2013) due to the emetine treatment.

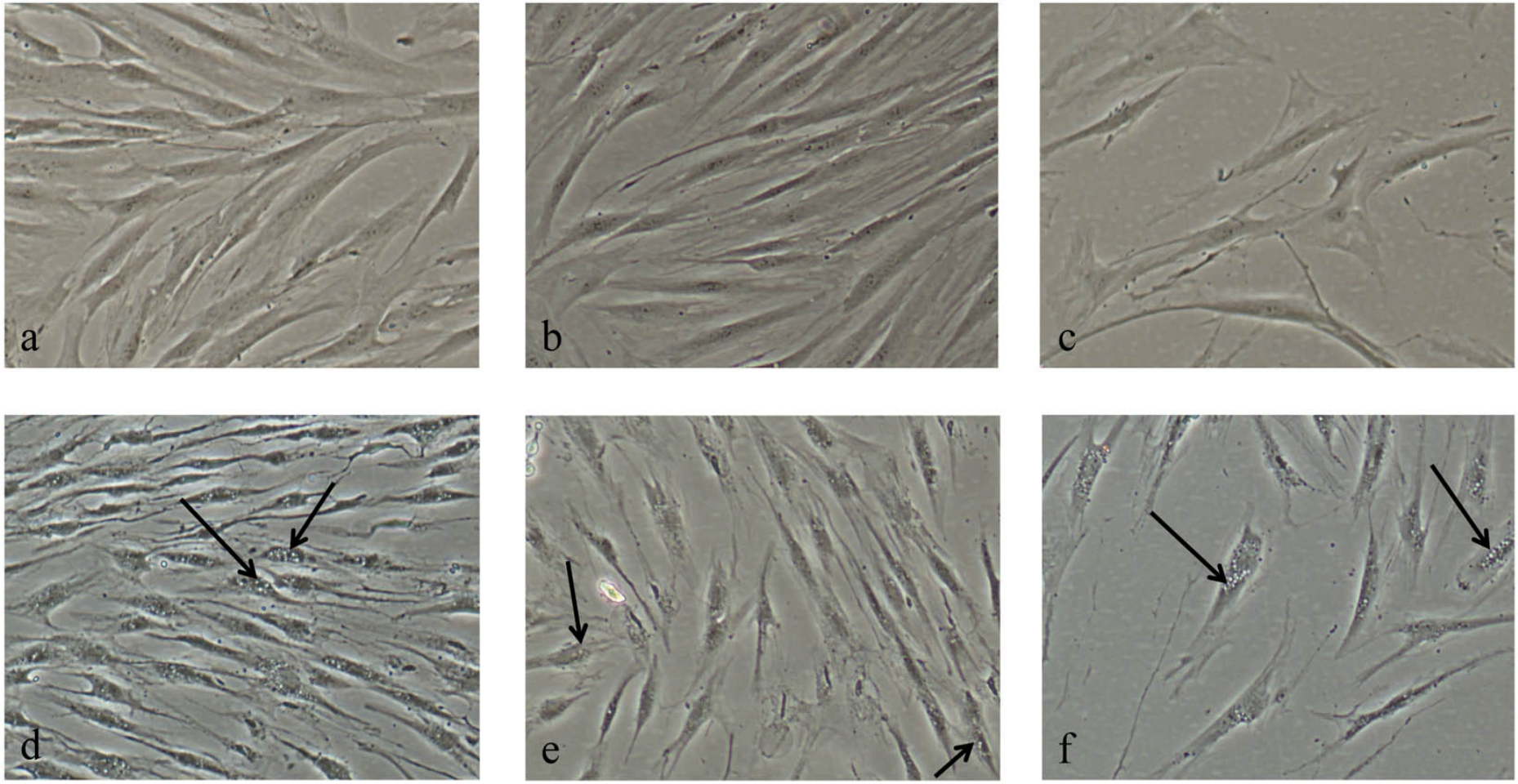


Figure 5.6: Fibroblast cells from emetine treatment experiment. Not treated fibroblast cells from controls (a and b) and patient P23 (c) exhibit normal morphology, albeit the patient's cells grow at a slower rate, possibly due to the mutation. Fibroblast cells from controls (d and f) and patient P23 (f) treated with emetine exhibit smaller cells and an accumulation of intracellular vesicles (arrows). The latter is present in virtually every cell.

Agarose gel electrophoresis of RT-PCR product generated using F and R1 primers which were specific for *ZFYVE27* cDNA (RNA was derived from fibroblasts of healthy controls and the patient P23, which were treated with emetine) revealed amplification of two PCR products (404 bp and 389 bp) (Figure 5.7). The same result was observed for not treated cells. Notably, when compared with the treated control cells, the treated cells of the patient did not produce any additional band, which would be indicative of the effect of the mutation on RNA splicing.

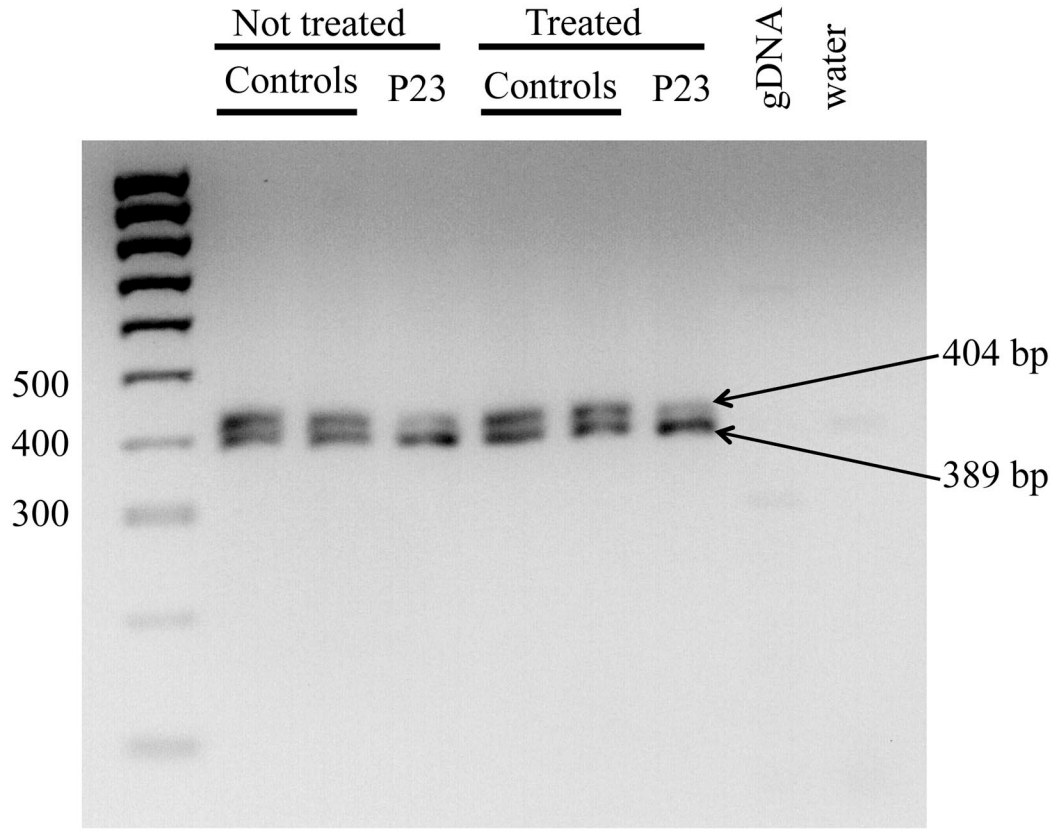


Figure 5.7: Agarose gel (3.5%) electrophoresis of the RT-PCR product generated using primers F and R1 which were specific for *ZFYVE27* cDNA (RNA was derived from fibroblasts of healthy controls and the patient P23 without [not treated] and with [treated] inhibition of nonsense-mediated decay by emetine)

Sequencing of ZFYVE27 cDNA (total PCR product) did not reveal any difference between treated and not treated fibroblast cells in controls and the patient. Sequence chromatograms showed mixed peaks in the 5' region of exon 7 and the homozygous deletion of the entire exon 8 in all samples (Figure 5.8). Based on the sequencing result, this study concludes that the band of 404 bp corresponds to the transcript 6, whereas the

band at 389 bp represents transcripts 3, 4, 5 and 7, which could be present in any combination (Figure 5.3).

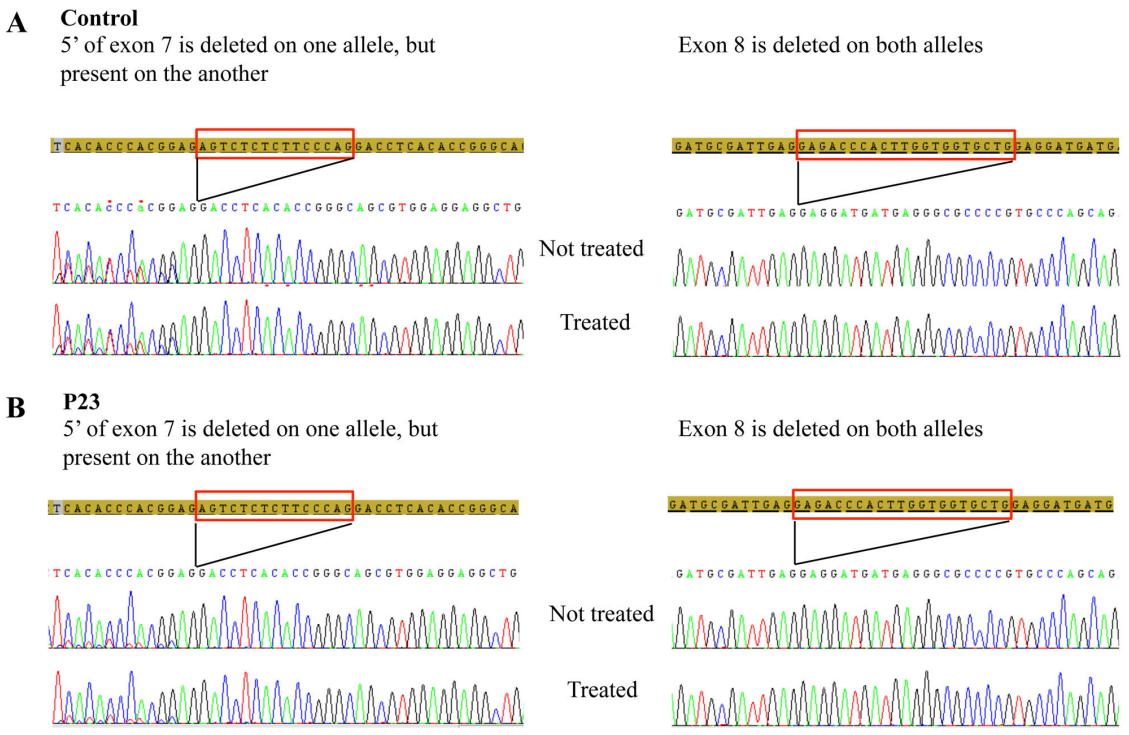


Figure 5.8: Sanger sequencing chromatograms for *ZFYVE27* cDNA obtained from the emetine treated fibroblasts of a healthy control (A) and patient P23 (B) using primers F and R1 which were specific for *ZFYVE27* cDNA. Red rectangles represent deleted sequences.

5.3.3. Functional studies of the ZFYVE27 *candidate mutation in the human frontal cortex and cerebellum*

Transcriptional study

Prior to the transcriptional study, the efficiency of the cDNA amplification was assessed by performing a control RT-PCR using primers specific to the housekeeping gene GAPDH, as described in Section 5.2.6. Agarose gel electrophoresis of RT-PCR product revealed amplification of GAPDH-specific product of 133 bp in all samples, confirming the satisfactory quality of the initial total RNA preparation (Figure 5.9).

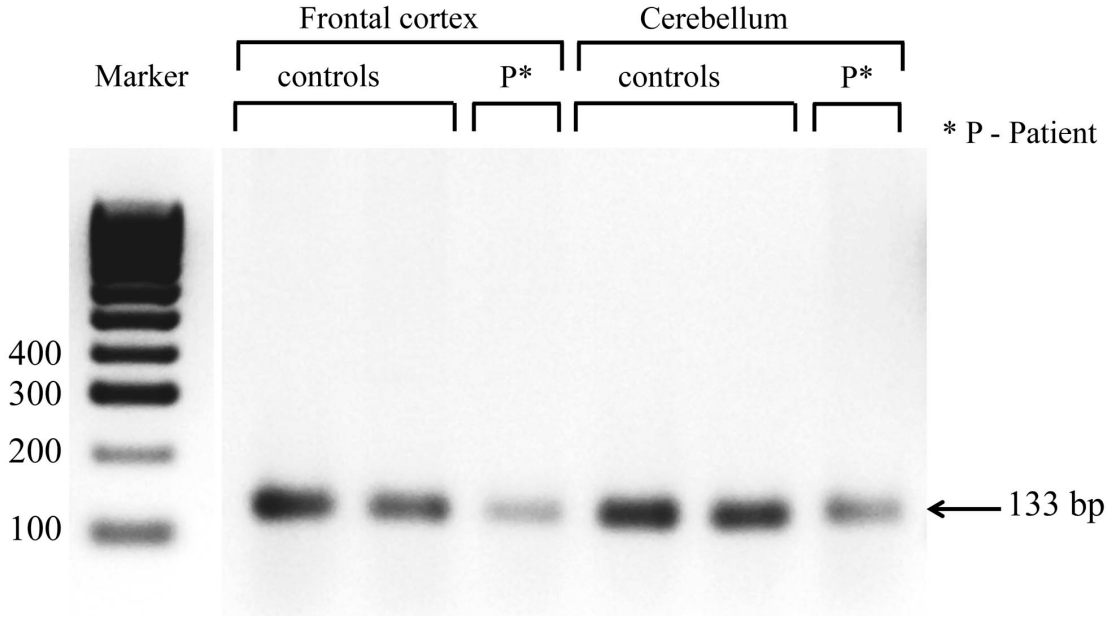


Figure 5.9: Agarose gel electrophoresis of *GAPDH* RT-PCR product derived from the frontal cortex and cerebellum of controls and the patient

Sequencing of ZFYVE27 cDNA (total PCR product and cloned cDNA fragments)

This study used one forward (F) and two reverse (R1 and R2) primers to generate a range of amplicons (Figure 5.3). Primer R2 covers the entire neuron-specific exon 8, hence amplicons generated with primers F and R2 would represent only neuron-specific transcripts (transcript 1 and transcript 2). R1 spans across a boundary between exons 9 and 10 and is pan-specific for all transcripts. Therefore, amplicons generated with primers F and R1 would represent all the variety of transcripts present in a sample.

This study initially used three age-matched controls alongside the patient's sample. Agarose gel electrophoresis of RT-PCR product for amplicon FR2 revealed expression of two bands of 265 bp and 250 bp (representing transcript 1 and transcript 2, respectively) in all samples but the patient's cerebellum, which expressed only one band of 250 bp (Figure 5.10). In one control there was a striking difference in amounts of amplified PCR product (frontal cortex of control 3) when compared with the other samples. A subsequent enquiry into this control's medical history revealed that the individual underwent a pituitary radiotherapy treatment, which could underlie the abnormal observations. This led to exclusion of this control from further investigations. Direct sequencing of RT-PCR reactions for amplicon FR2 revealed mixed peaks on the

electropherograms (which indicates the presence of both transcripts 1 and 2) for all samples but the patient's cerebellum, where this study confirmed the presence of only transcript 2.

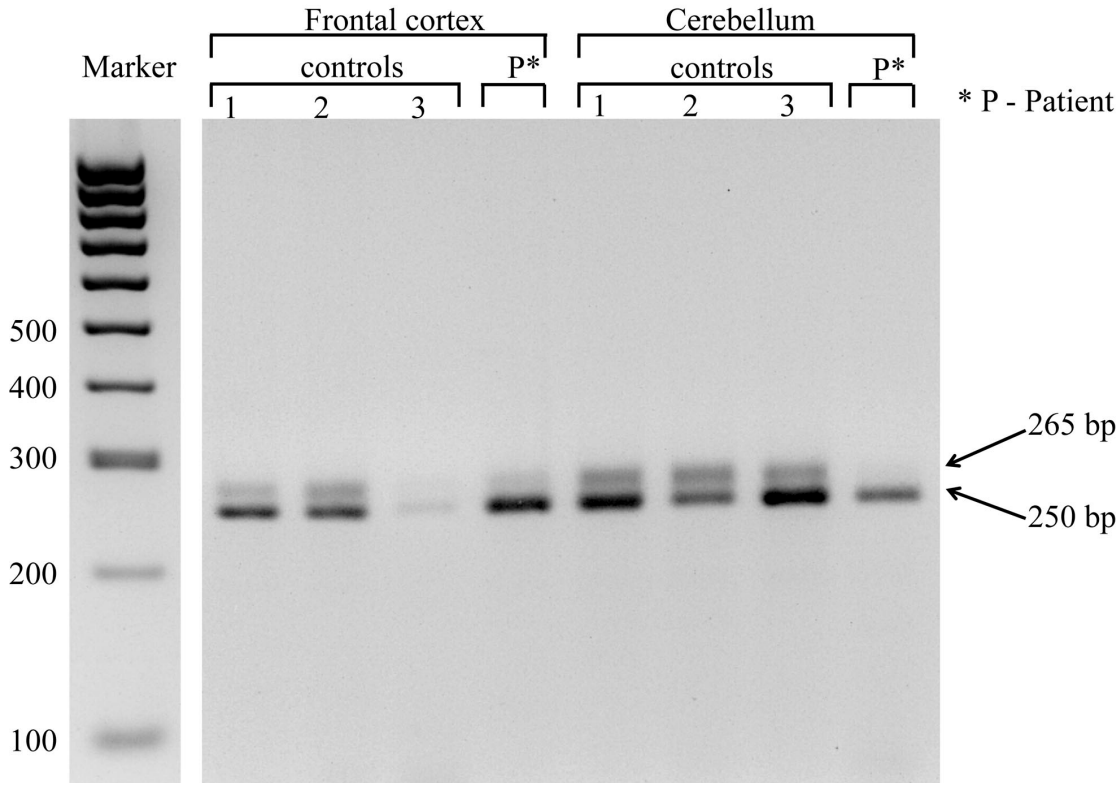


Figure 5.10: 3.5% Agarose gel electrophoresis of *ZFYVE27* RT-PCR product (amplicon FR2) derived from the frontal cortex and cerebellum of controls and the patient

Agarose gel electrophoresis of RT-PCR product for amplicon FR1 showed amplification of multiple bands in all samples, with the patient's cerebellum producing the least number of bands (Figure 5.11). Similar to the results for amplicon FR2 above, in the frontal cortex of control 3 there was a striking difference in quantities of amplified PCR product when compared with the other samples. As previously mentioned, this study excluded this control from further investigations. Direct sequencing of the total RT-PCR product for amplicon FR1 produced inconclusive results for all samples, which was manifested by mixed peaks on the sequencing traces in all samples.

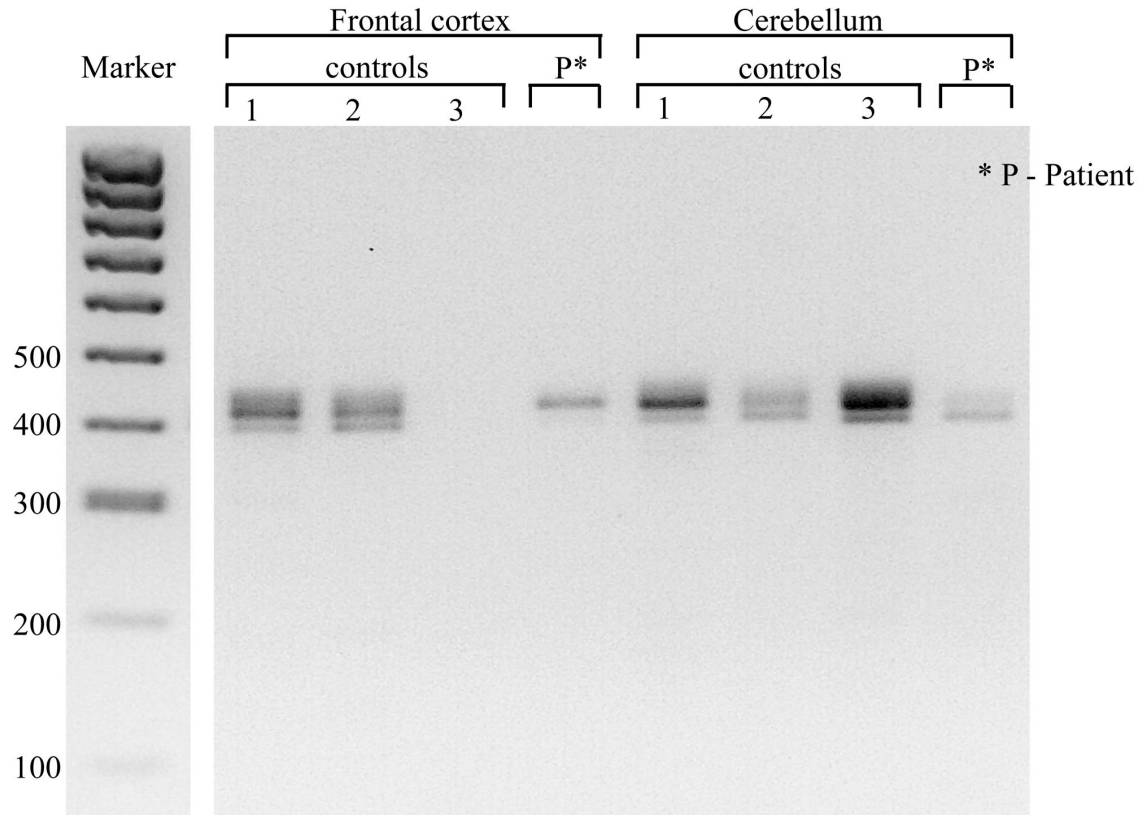


Figure 5.11: 3.5% Agarose gel electrophoresis of *ZFYVE27* RT-PCR product (amplicon FR1) derived from the frontal cortex and cerebellum of controls and the patient

The current study concluded that the direct sequencing of total RT-PCR product was inconclusive and therefore the study moved on to cloning the cDNA fragments from each sample and subsequent sequencing of the clones. This investigation was performed only for amplicon FR1.

Sequencing analysis of cDNA clones for amplicon FR1 revealed that whilst control frontal cortex and cerebellum express potentially all seven *ZFYVE27* RNA transcripts, the patient's frontal cortex does not express transcript 6. More strikingly, the patient's cerebellum expressed only one transcript (the neuron-specific transcript 2).

Protein expression study

Western blot analysis of ZFYVE27

Western blotting of ZFYVE27 was performed as described in Section 5.2.8. The commercial anti-ZFYVE27 antibody was generated against the C-terminus region of the

protein. The ZFYVE27 protein (protrudin) levels were increased approximately 2.5 fold in the cerebellum but not in the frontal cortex of the affected individual, when compared to controls (Figure 5.12).

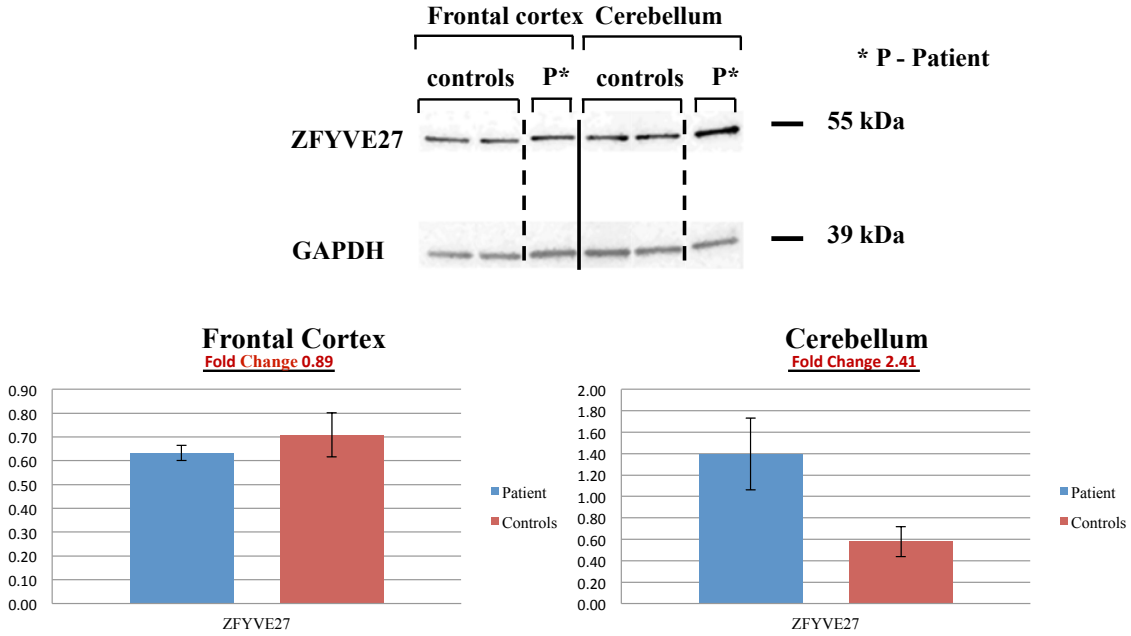


Figure 5.12: ZFYVE27 and GAPDH protein levels in the frontal cortex and the cerebellum in controls and the patient

Western blot analysis of ER stress and UPR markers

Western blotting of the master regulator of ER stress, BiP/GRP78, was performed as described in Section 5.2.8. The levels of BiP/GRP78 were significantly increased (approximately 2 fold) in the patient's cerebellum but not in the frontal cortex, when compared to controls (Figure 5.13). The large standard deviation values for the controls' frontal cortex may reflect the biological variation in the general population (Stan *et al.*, 2006).

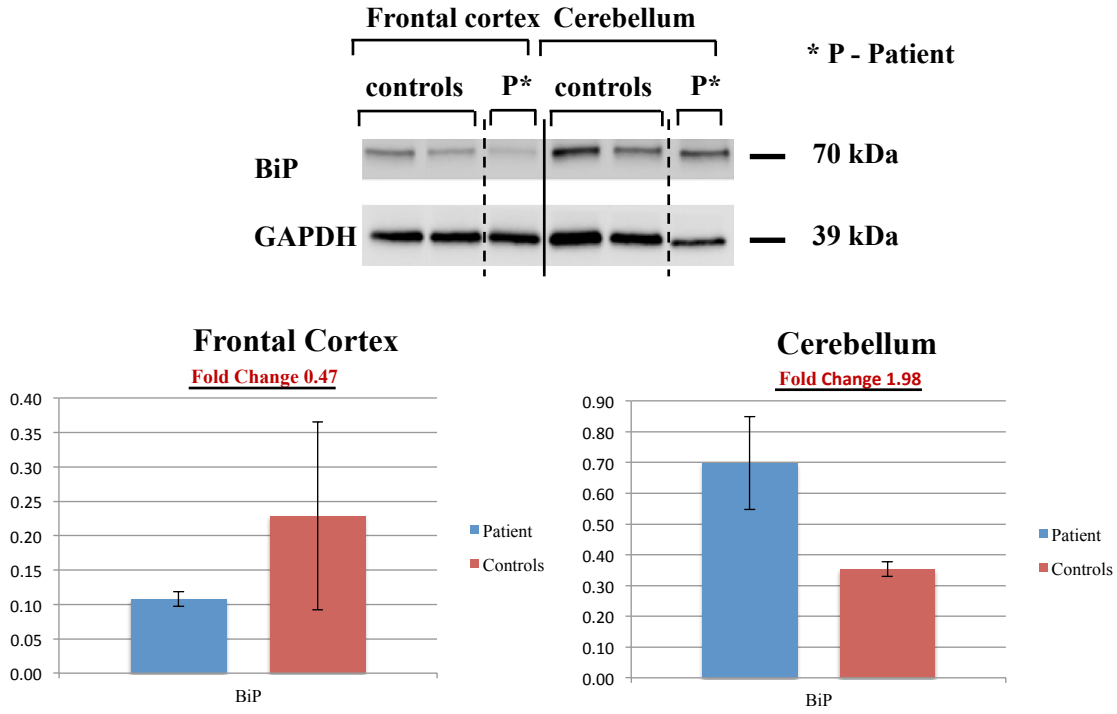


Figure 5.13: BiP/GRP78 and GAPDH protein levels in the frontal cortex and the cerebellum in controls and the patient. The levels of BiP/GRP78 are significantly increased (approximately 2 fold) in the patient's cerebellum but not in the frontal cortex, when compared to controls. The large standard deviation values for the controls' frontal cortex may reflect the biological variation in the general population.

IHC of ZFYVE27

The commercially available ZFYVE27 antibody, which was used in this study, was generated against C-terminus region of the protein.

- *Optimization of the protocol using human control muscle*

Human post-mortem brain is a precious sample and not always easily available. Therefore, prior to performing IHC for ZFYVE27 on the human brain tissue, the protocol was optimized on frozen human control muscle sections (Figure 5.14). This study observed some tissue preparation and handling artefacts (cell bubbles and separation of cells). Positive staining was observed in the cytoplasm of the muscle cells (Figure 5.14, B). No staining was observed in the negative control (Figure 5.14, A), however some background staining was observed.

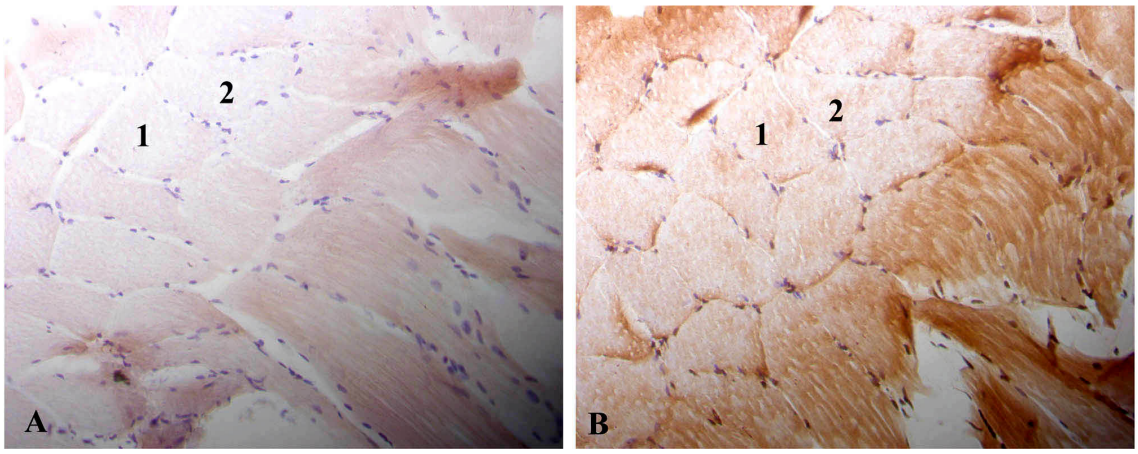


Figure 5.14: IHC staining of frozen human control muscle tissue sections with anti-ZFYVE27 antibody. Magnification x20. Negative control (no primary antibody) (A) and 1:100 dilution of anti-ZFYVE27 antibody (B). Sections are sequential. Numbers represent the same cell.

- *ZFYVE27 detection in human frontal cortex and cerebellum*

This study detected expression of ZFYVE27 in the frontal cortex and the cerebellum of controls and the patient, which supported the Western blot findings. The latter revealed that the ZFYVE27 protein (protrudin) levels were increased approximately 2.5 fold in the cerebellum but not in the frontal cortex of the affected individual, when compared to controls (Figure 5.12). In the frontal cortex ZFYVE27 staining was observed in neurons (Figure 5.15, arrows) and glial cells (Figure 5.15, arrow heads). No difference in staining intensity or localisation between controls and the patient was observed. In the cerebellum this study detected immunostaining of Purkinje cells in controls (Figure 5.16 A, arrows). In the patient this study did not observe Purkinje cells immunostaining, probably due to their loss (Figure 5.16 E, arrows). This study observed some tissue damage in the patient's cerebellum and some gaps in the Purkinje cell layer, which would indicate a loss of Purkinje cells. The current study observed occasional immunostaining in cells suggestive of astrocytes in the cerebellum of the controls (Figure 5.16 B, arrow heads). In the patient's cerebellum this study detected pronounced immunostaining in cells resembling astrocytes in the molecular layer (Figure 5.16 F, G, H, arrow heads). The current study also noticed immunostaining in parallel fibres of the patient cerebellum (Figure 5.16 G, arrow heads).

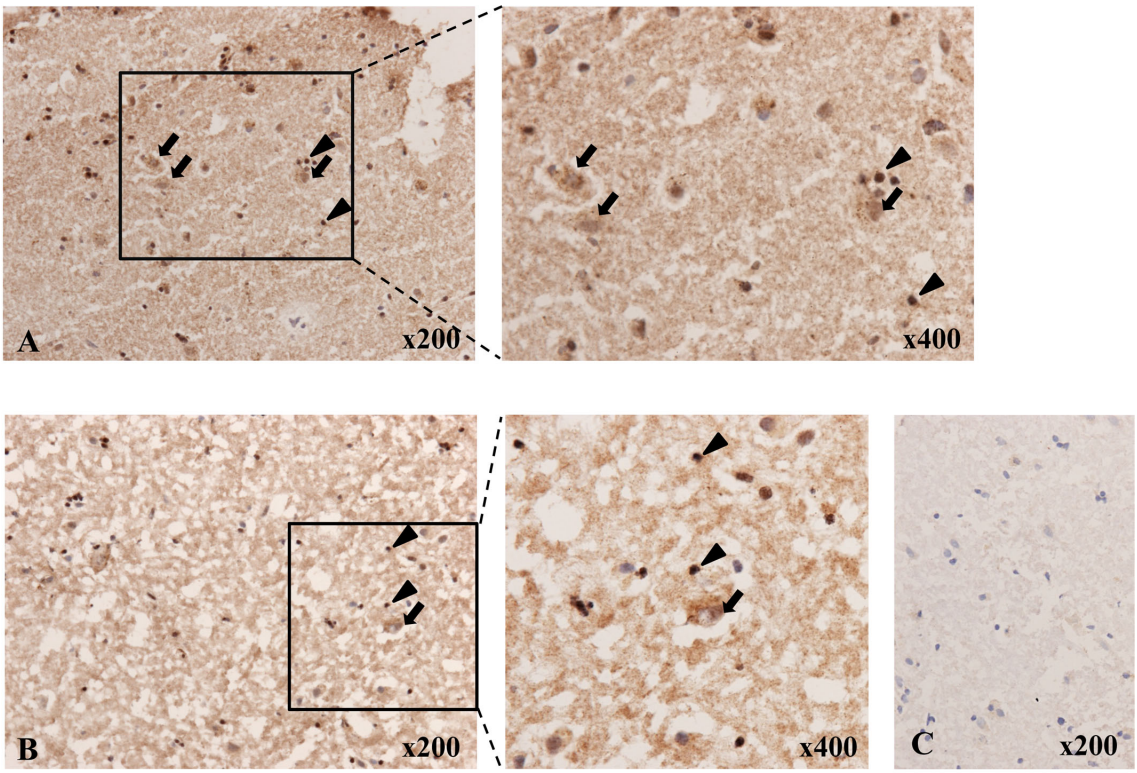


Figure 5.15: IHC staining of frozen frontal cortex sections in controls and the patient with anti-ZFYVE27 antibody. Staining in patient (A) and control (B) frontal cortex with anti-ZFYVE27 antibody diluted 1:200. Negative control (C): patient frontal cortex with no primary antibody. Arrows represent neurons, arrow heads represent glial cells. Magnification x200.

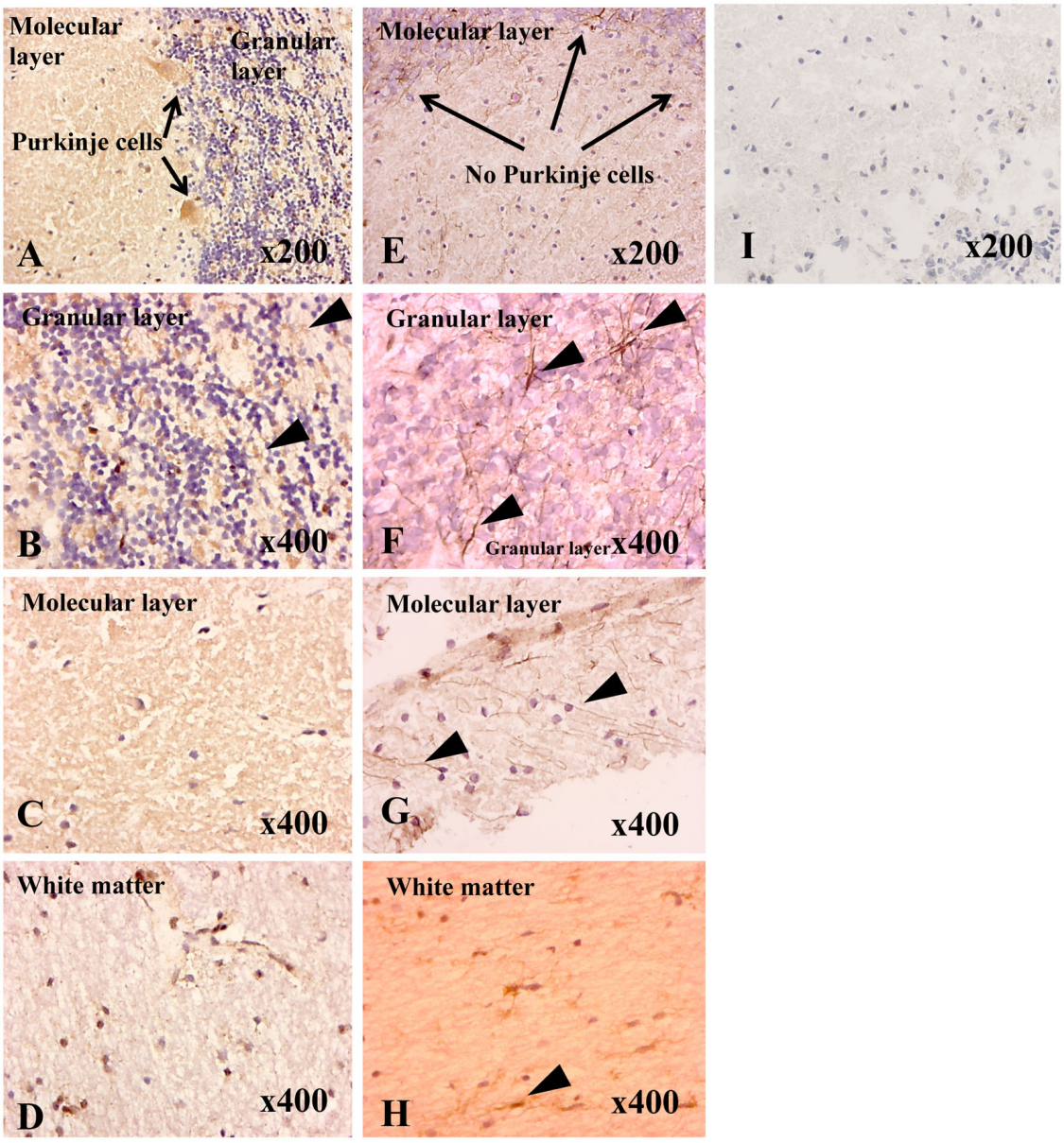


Figure 5.16: IHC staining of frozen cerebellum sections in controls and the patient with anti-ZFYVE27 antibody. Staining in a control (A, B, C, D) and the patient (E, F, G, H). Negative control: no primary antibody in the patient's cerebellum (I). Arrowheads represent cells resembling astrocytes. Magnification x200 and x400.

5.4. Discussion

5.4.1. Transcriptional studies in fibroblasts

Using an unbiased whole exome sequencing approach, in three members of an autosomal dominant pedigree (patient P22, P23 and P24), a predicted splice site mutation (c.805-2A>G) was detected in the *ZFYVE27* gene, previously associated with spasticity, but not ataxia (Mannan *et al.*, 2006). It is documented that approximately 15% of mutations linked to human inheritable diseases cause splicing defects in mRNA (Blencowe, 2000). Mutations affecting splice sites are therefore of a particular interest. The routine method of studying the effect of a genomic DNA mutation on mRNA splicing involves sequencing of RNA from patients and controls. The total RT-PCR product is sequenced and the sequence chromatograms are compared between patients and controls. Any potential aberrant splicing event is detected when the sequence chromatogram of mutant and wildtype alleles show differences, which is manifested by mixed peaks on the chromatogram. In this study *ZFYVE27* cDNA (total RT-PCR product) were initially sequenced in fibroblasts of the patient (P23) and two unrelated controls, however any abnormal splicing in the patient cells in comparison to the controls were not detected. Moreover, sequencing of cDNA from fibroblasts of the patient P23 and controls showed that exon 8 was absent in both the patient and controls. Therefore, it was concluded that fibroblasts do not express RNA transcripts 1 and 2, both of which contain exon 8. Furthermore, it was reasoned that fibroblasts are not the tissue of interest in a family with ataxia. The most relevant tissue for this study would be brain tissue, but it is precious and has not been available at that time.

The mutation (c.805-2A>G) in *ZFYVE27* was predicted to result in a loss of acceptor site of the coding exon 7 in five splice prediction programs [Alamut version 2.4 (Interactive Biosoftware, Rouen, France)] (Figure 4.1). This could hypothetically lead to skipping of the downstream exon and result in a frameshift, which potentially could introduce a downstream premature stop codon. Literature suggests that premature stop codons, which account for approximately one third of mutations causing human disorders, could trigger degradation of mutant transcripts by nonsense mediated RNA decay pathway (Frischmeyer and Dietz, 1999). Nonsense-mediated decay is a quality control mechanism found in eukaryotic cells (Conti and Izaurralde, 2005). The primary role of nonsense-mediated decay is to degrade mRNA harbouring premature stop

codons, in order to prevent expression of truncated proteins with altered functions. If expressed, such proteins would be either partially functional, non-functional, or even deleterious to the cell (Wengrod *et al.*, 2013). Elimination of the premature stop codon-containing mRNA transcripts by nonsense-mediated decay creates a major problem in detecting the effect of mutations on mRNA splicing. Inhibition of the nonsense-mediated decay pathway would allow read-through of premature stop codons and expression of the transcripts, that otherwise would be eliminated by the nonsense-mediated decay pathway. Treatment of the patient's fibroblasts with 100 µg/ml emetine dihydrochloride hydrate has been shown to successfully inhibit the nonsense mediated decay pathway (Noensie and Dietz, 2001). Based on the above observations, the identification of nonsense-mediated decay as a molecular mechanism of a disease could potentially lead to a therapy employing inhibition of this pathway. Recent clinical trials have indicated that treatment with gentamicin, an inhibitor of nonsense-mediated decay, could be effective in fibrosis and muscular dystrophy - diseases caused by nonsense-mediated decay triggering mutations (Welch *et al.*, 2007). This study was interested, therefore, to elucidate whether the c.805-2A>G mutation leads to nonsense-mediated decay. In this study control and patient fibroblasts were treated with emetine dihydrochloride hydrate in order to inhibit nonsense-mediated decay. cDNA (total PCR product) was sequenced with the aim to study the effect of the candidate mutation on the possible exon skipping and activation of the nonsense mediated decay pathway. The outline of this experiment is presented in Figure 5.2. This study hypothesized that by inhibiting the nonsense mediated decay pathway and allowing read-through of a premature stop codon in the patient, it would be possible to amplify an additional *ZFYVE27* RNA transcript by means of RT-PCR. Detecting an additional (aberrant) transcript in the patient's sample would indicate that the *ZFYVE27* c.805-2A>G mutation leads to nonsense-mediated decay.

Sequencing of cDNA (total RT-PCR product) from emetine treated and non-treated fibroblasts from both controls and the patient did not show any difference, however. The latter experiment showed also that exon 8 was absent in both the patient's and controls' fibroblasts. Since these experiments were carried out, new research has been published describing the first *ZFYVE27* transcriptional study in mice (Ohnishi *et al.*, 2014). The authors found that in mice the region corresponding to human exon 8 (21bp) is present only in a transcript expressed in the neuronal tissue (cerebrum and spinal cord). This finding explained why this study was not able to detect exon 8 in control

and the patient's fibroblasts. More importantly, this study showed that fibroblast cell lines are not ideal to study neurological diseases.

5.4.2. Transcriptional study of ZFYVE 27 in the post-mortem frontal cortex and cerebellum

Brain tissue is precious and has not been available at the beginning of this study. Later, regrettably a member of the family died and the brain tissue became available. Hence, a transcriptional study was carried out in the frontal cortex and cerebellum of post-mortem brains from the affected individual (patient P22) and three age-matched controls. The yield and quality of RNA extracted from post-mortem brain tissue could be affected by many factors, which cannot be controlled. Therefore it is important to determine the quality of the starting material prior to research. It is established that pH is a good indicator of post-mortem brain tissue quality (Stan *et al.*, 2006). Tissue samples used in this study were characterised by pH>6.4, which is considered good quality (Stan *et al.*, 2006). The quality of the total RNA extracted from the post-mortem tissue was assessed by means of RT-PCR using primers specific to a housekeeping gene *GAPDH*. High expression levels of *GAPDH* were detected in all samples which is an indicator of good quality total RNA (Figure 5.9). Agarose gel electrophoresis of the RT-PCR product for amplicon FR2 of *ZFYVE27* revealed the amplification of two products, representing transcript 1 and transcript 2, in all samples apart from the patient's cerebellum, which expressed only the lower band of 250 bp corresponding to RNA transcript 2 (Figure 5.10). This result may indicate that the patient's cerebellum and Purkinje cells are more severely affected by the mutation than other neuronal populations. It has been previously documented that Purkinje cells, one of the largest neurons, are uniquely vulnerable and more susceptible to damage in response to genetic changes than other neuronal populations (Hekman and Gomez, 2015). Elsewhere, it has been documented that in mice ER stress leads to degeneration of Purkinje cells, which result in cerebellar degeneration and ataxia (Kyuhou *et al*., 2006). Also, it has been documented that defects in ER structure and function in dendrites and axons could result in neurological phenotypes (Ramirez and Couve, 2011). The aforementioned studies suggest that Purkinje cells might be more susceptible to damage due to ER stress, than other neuronal cell populations. Agarose gel electrophoresis of the RT-PCR product for amplicon FR1 showed amplification of multiple bands in all samples (Figure 5.11). Sequencing analysis of total RT-PCR products and cDNA clones

revealed that there is a difference in the expression of RNA transcripts in the brain of the patient, when compared to the controls. This study found that in control individuals frontal cortex and cerebellum potentially express all seven *ZFYVE27* RNA transcripts (Figure 5.3). Interestingly, the current study did not detect expression of transcript 6 in the patient's frontal cortex, however all other transcripts were evident. The patient's cerebellum expressed only one transcript (neuron-specific transcript 2). The *ZFYVE27* gene has seven protein coding RNA transcripts as a result of alternative splicing (Figure 5.3). Out of seven splice transcripts, only transcripts 1 and 2 contain exon 8. A recent study suggested that in mice, expression of *ZFYVE27* transcripts, containing the 21bp region corresponding to human exon 8, is restricted to cerebrum and spinal cord (Ohnishi *et al.*, 2014). The latter study also found that *ZFYVE27* transcripts, which did not contain the 21bp region, were expressed in all tissues tested. The current study provides the evidence that human fibroblasts do not express the neuron-specific *ZFYVE27* transcripts 1 and 2 (Section 5.4.1). The current study also suggests that the human brain potentially expresses all known *ZFYVE27* transcripts. The results of this study showed a difference in the expression of *ZFYVE27* RNA transcripts in the brain of the patient, when compared to the controls. For some genes, the ratio of alternatively spliced transcripts is strictly regulated. Splice site mutations might disrupt this ratio and result in a disease (Ward and Cooper, 2010). A recent study has shown that splice site mutations in the microtubule associated protein tau (*MAPT*) gene cause frontotemporal dementia and Parkinsonism linked to chromosome 17 (FDTP-17) (Liu and Gong, 2008). Moreover, the latter study found that splice site mutations in *MAPT* disrupt alternative splicing of exon 10. This resulted in altering the ratio of alternatively spliced *MAPT* transcripts encoding isoforms containing either three (3R) or four (4R) microtubule-binding repeats. The authors suggested that deregulation of the 3R/4R ratio contributes to FDTP-17, however, the molecular mechanism of this deregulation is currently not understood. All the aforementioned studies suggest, that the splice site mutation in *ZFYVE27* could result in deregulation of the ratio of splice variants in the brain of the patient. The deregulation of the ratio might change cellular conditions and lead to ER stress. The latter might lead to degeneration of Purkinje cells, which result in cerebellar degeneration and ataxia in the patient. It has to be highlighted that this study analysed only two regions of the brain (the frontal cortex and the cerebellum) for expression of *ZFYVE27* gene transcripts in the subjects involved in the study. It would be beneficial to analyze other brain regions from the patient and control individuals in order to gain more insight into the expression profiles of ZFYVE27 under normal and

pathological conditions.

The *ZFYVE27* mutation (c.805-2A>G) is predicted to cause a loss of acceptor site of the coding exon 7. The putative role of the c.805-2A>G mutation in the hypothetical nonsense-mediated decay, which might have contributed to the patient's clinical phenotype, has been explored in the patient's fibroblasts. This study did not detect any alterations in expression of RNA transcripts in the patient cells compared to controls, as described in Section 5.4.1. Analysis of post-mortem brain tissue however is more relevant as *ZFYVE27* RNA transcripts 1 and 2 are neuron specific and not expressed in fibroblasts (Figure 5.3) (Ohnishi *et al.*, 2014). Nonsense-mediated decay has been established as a mechanism to eliminate altered mRNA species. Recently it has been shown however that nonsense-mediated decay also targets non-mutated transcripts, including transcripts which play an important part in cellular stress response, more specifically, ER stress response (Mendell *et al.*, 2004). It is likely therefore that the c.805-2A>G mutation in *ZFYVE27* leads to nonsense-mediated decay in the patient, which is supported by different expression of RNA transcripts in the patient as opposed to controls. This study speculates that nonsense-mediated decay, triggered by the mutation, could eliminate functional, not mutated *ZFYVE27* transcripts and stabilize mutated transcripts (Gardner, 2010). Recently it has been documented that nonsense-mediated decay efficiency varies between different tissues (Linde *et al.*, 2007). This could explain the differences in transcript expression between the frontal cortex and the cerebellum observed in the patient. To support the hypothesis that the c.805-2A>G mutation could cause nonsense-mediated decay in the patient, and to clarify the molecular nature of the splicing, it would be useful to study the effect of the mutation in primary neuronal cell cultures. The experiment could involve transfection with constructs containing mutated or wildtype *ZFYVE27* gene fragments. Subsequent treatment with emetine in order to inhibit nonsense-mediated decay, as described earlier in this Thesis, would help to elucidate the mechanisms of transcriptional regulation occurring in neuronal cells.

5.4.3. Analysis of ZFYVE27 protein expression in the frontal cortex and cerebellum

This study carried out a protein expression study using post-mortem frontal cortex and the cerebellum of patient P22 and age matched controls. Firstly, Western blot analysis was employed to study ZFYVE27 expression in the frontal cortex and the cerebellum. This study found that the ZFYVE27 levels are increased approximately 2.5 fold in the

cerebellum but not in the frontal cortex of the affected individual, when compared to controls. Many splice site mutations result in exon skipping, which could lead to the production of inactive or unstable protein. The latter is manifested in the decreased level of protein expression (Cooper *et al.*, 2009). However, in some cases, a splice site mutation could result in the aggregation of protein (Liu and Gong, 2008). ZFYVE27 is expressed in the endoplasmic reticulum (ER) and is thought to be involved in shaping the ER network (Chang *et al.*, 2013). ER is a key site for protein synthesis and quality control. Accumulation of unfolded proteins leads to ER stress and activation of unfolded protein response (UPR) (Wang *et al.*, 2010). ZFYVE27 anomalies have previously been shown to cause ER stress (Hashimoto *et al.*, 2014) and result in a neurological phenotype SPG33 (Mannan *et al.*, 2006). This study therefore used Western blotting to determine levels of the master regulator of ER stress, BiP/GRP78 (Wang *et al.*, 2010). The results of this study show that levels of BiP/GRP78 were significantly increased in the patient's cerebellum, which may indicate ER pathology. It would be interesting to use another ER stress marker (for example, 70-kDa heat shock proteins (HSP70)) to confirm the result with BiP/GRP78. It should be mentioned that in this family mitochondrial disease was not considered because standard mitochondrial tests were negative. Therefore, investigating mitochondrial parameters in the patient's fibroblasts was not considered. It should also be highlighted that although overexpression of BiP/GRP78 suggests ER pathology on the one hand, and overexpression of ZFYVE27 is indicative of the consequence of the mutation on the other hand, this study only tested two regions of the brain (cerebellum and the frontal cortex) in the subjects involved in the study. Analysis of other brain regions from the patient and neurologically normal individuals would be beneficial in order to gain more insight into the expression profiles of ZFYVE27 under normal and pathological conditions. IHC with anti-ZFYVE27 antibody revealed staining in both frontal cortex (Figure 5.15) and the cerebellum (Figure 5.16) of controls and the patient, which supported this study's Western blot findings. In the frontal cortex of controls and the patient ZFYVE27 staining was detected in neurons and glial cells. In the control cerebellum immunostaining was detected of Purkinje cells. In the patient immunostaining of Purkinje cells was not detected, probably due to their loss. This observation is supported by the pathological report of post-mortem tissue examination of the patient's brain (provided by Newcastle Brain Tissue Resource, Newcastle University), which states depletion of the Purkinje cell population and cerebellar atrophy in the patient. There were detected gaps in the Purkinje cells' layer of the

patient's cerebellum. Based on the pathological report mentioned above, this finding suggests death of Purkinje cells, possibly due to apoptosis, in the patient (Dusart *et al*., 2006). To elucidate the mechanism of Purkinje cells loss, screening of proapoptotic markers would be beneficial. Purkinje cells are the largest neurons in the brain, which together with other neurons and neuronal fibres form the cerebellar neuronal circuitry (Figure 5.17). A recent study in neurological disorders, featuring structural changes in the cerebellum, highlights the role of Purkinje cell degeneration in the development of the clinical phenotype. For example, post-mortem studies of essential tremor identified structural changes in the cerebellum, including Purkinje cell loss (Louis *et al.*, 2013), which indicates that Purkinje cell degeneration may be central to the pathogenesis of essential tremor. In another study on essential tremor, authors found that abnormalities in climbing fibre - Purkinje cell connections could lead to dysfunctional motor control and possibly clinical tremor and/or ataxia (Lin *et al.*, 2014). In mouse models of spinocerebellar ataxias, expression of mutant proteins in Purkinje cells causes Purkinje cell degeneration and abnormal climbing fibre-Purkinje cell connections (Furrer *et al.*, 2013). All the aforementioned studies suggest, that proper connections between Purkinje cells and other neuronal structures are of paramount importance for the normal function of the cerebellum.

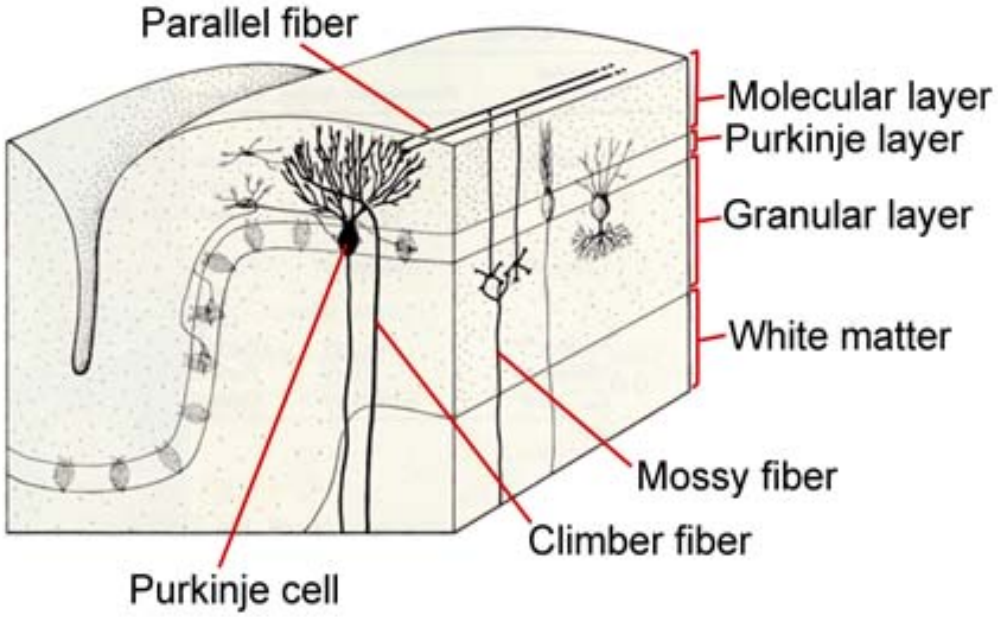


Figure 5.17: Schematic representation of a cerebellar cross-section. Figure depicts basic structure of the cerebellar cortex and shows Purkinje cells in relation to

cerebellum layers and other neurons and neuronal fibres including parallel, mossy and climber fibres (taken from De Schutter and Bower, 1994)

Interestingly, current IHC studies detected immunostaining in cells suggestive of astrocytes in both – the control individual and the patient. These cells were not tested for expression of glial or astrocyte markers, therefore the term "astrocytes" is used to describe findings presented here. The control sample showed only an occasional staining of "astrocytes", whereas the patient presented much more pronounced staining suggestive of proliferation of "astrocytes" and reactive astrogliosis. Astrocytes are the most abundant cells in the central nervous system (CNS), which traditionally were thought to play an important role in supporting neurons both structurally and metabolically (Hol and Pekny, 2015). Recent research has shown that astrocytes play a crucial role in synaptogenesis (Allen *et al.*, 2012). Proliferation of astrocytes (reactive astrogliosis) is a response of the CNS to neuronal damage (Hol and Pekny, 2015). Astrogliosis plays an important role in health, stimulating recovery processes (Burda and Sofroniew, 2014) and disease progression (Barres, 2008). In this study the putative reactive astrogliosis in the patient is supported by the Western blot analysis, which showed almost a 2.5 fold increase in expression of ZFYVE27 in the patient's cerebellum, when compared to controls. IHC analysis also revealed ZFYVE27 staining of parallel "fibres", resembling astrocyte processes, in the patient cerebellum. As staining for astrocyte or glial markers was not performed, this study is not able to draw conclusions regarding the expression of ZFYVE27 in astrocytes in the cerebellum. Therefore, future work should focus on performing a colocalisation study which would involve immunofluorescent staining of ZFYVE27 along with a glia marker (for example, glial fibrillary acidic protein (GFAP)) (Hol and Pekny, 2015) and an astrocyte marker (for example, aldehyde dehydrogenase 1 family, member L1 (ALDH1L)) (Tong *et al.*, 2014). More control subjects and more brain regions should also be analysed for expression of ZFYVE27. It would also be interesting to look into the expression of ZFYVE27 in other types of ataxia, to support current studies and confirm whether the identified genetic defect in *ZFYVE 27* is a cause of disease in the case investigated in this Thesis.

Recently ZFYVE27 has been found to serve as an adaptor molecule that connects the motor protein KIF5A and its' cargoes in vesicular transport during process formation (Matsuzaki *et al.*, 2011). KIF5A is a cause of human hereditary spastic paraplegia (SPG10) (Reid *et al.*, 2002). The role of the interactions between ZFYVE27 and

KIF5A, which lead to connecting the ER with late endosomes and translocation of the latter during neurite outgrowths, has recently been elucidated (Raiborg *et al.*, 2015). ZFYVE27 binds the C-terminus of the KIF5A protein and the binding site could potentially be disrupted by *ZFYVE27* c.805-2A>G mutation. Given that both *ZFYVE27* and *KIF5A* are known disease genes in hereditary spastic paraplegia, it would be interesting to explore the effect of the *ZFYVE27* mutation on the protein interactions by means of protein expression studies.

5.4.4. Proposed model of functional consequences of the c.805-2A>G mutation

Based on the findings presented here, this study proposes a model of functional consequences of the c.805-2A>G mutation in *ZFYVE27* (Figure 5.18). Mutant protrudin may lead to ER stress, as evidenced in this study, however the mechanisms have not been identified as yet (Gardner, 2010). The effects of this mutation in *ZFYVE27* may also disrupt the normal interaction of protrudin with KIF5A, which may further lead to the failure to recruit late endosomes and eventually result in failure of neurite outgrowths (Raiborg *et al.*, 2015).

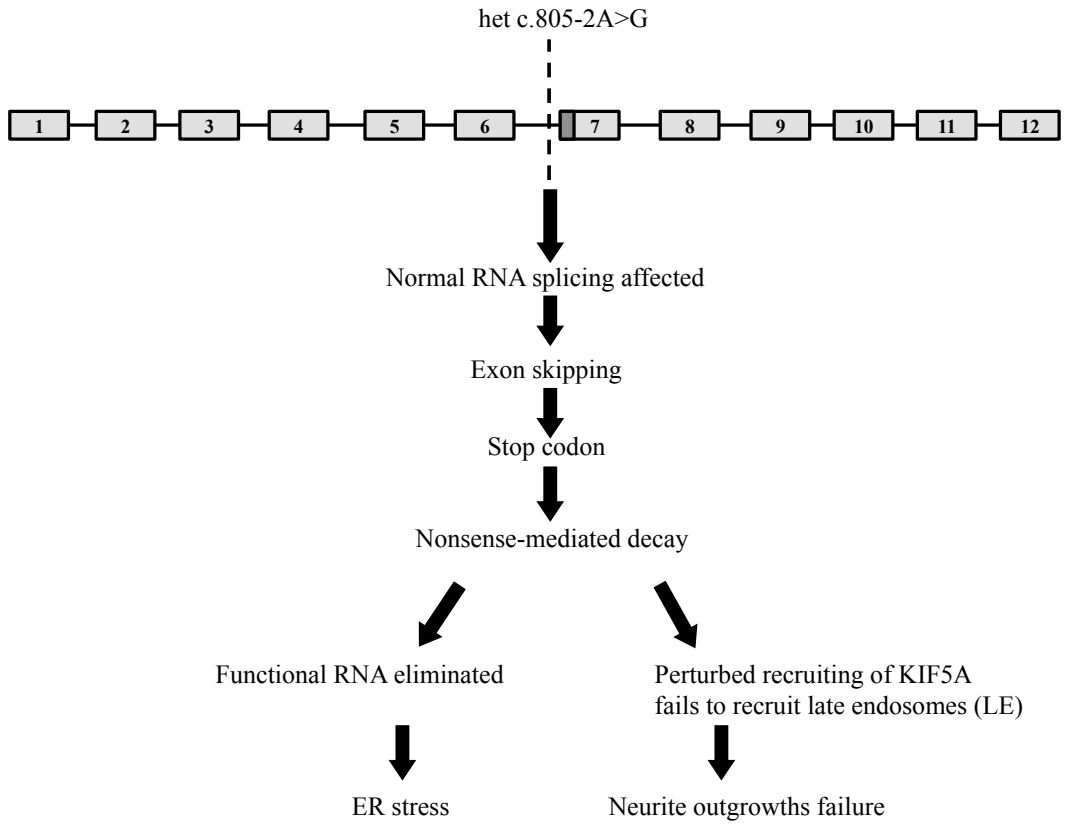


Figure 5.18: Proposed model of functional consequences of the ZFYVE27 c.805-2A>G mutation.

In summary, this study presents the evidence supporting a causative role of *ZFYVE27* gene mutation (c.805-2A>G) in a dominant family with cerebellar ataxia. The transcriptional studies, presented in this Thesis, show that the c.805-2A>G mutation could affect normal RNA splicing and cause nonsense-mediated decay of *ZFYVE27* transcripts. The protein expression studies show that the c.805-2A>G mutation could cause ER stress and lead to neuronal damage. This study has an important implication because it could potentially lead to therapy employing inhibition of nonsense-mediated decay with gentamicin (Welch *et al*., 2007). It should be highlighted that the current study subsequently investigated the frequency of the *ZFYVE27* mutation c.805-2A>G on ExAC database (available at http://exac.broadinstitute.org) and found that it is a rare mutation (allele count 3/121378; MAF=0.00002472). However, with no indisputable *ZFYVE27* pathogenic mutations identified to date, it is difficult to draw conclusions as to the pathogenicity of the c.805-2A>G mutation. Therefore, at the time of writing this Thesis, these data could not be used in the clinical setting. It should be mentioned that although Patient 23's muscle biopsy tests showed multiple mitochondrial DNA deletions and a small frequency of COX negative fibres, the appearances of the biopsy and the multiple mitochondrial DNA deletions were consistent with age-related changes, and thus not considered to be diagnostically relevant. However, it would be interesting to investigate mitochondrial dynamics in the patient's fibroblasts. In order to elucidate the cause of disease in this family, it would be beneficial to perform whole genome sequencing with haplotype reconstruction. Haplotype reconstruction employs computational methods, which enables the identification of alleles that are located on the same chromosome. When performed on related affected individuals, haplotype reconstruction enables the identification of identical-by-decent haplotypes shared between those individuals. It is anticipated that haplotypes, being more specific markers than single nucleotide variants, will uncover novel haplotype-disease associations (Glusman *et al*., 2014). The latter will play an increasingly important role in personalised medicine.

.

Chapter 6

Molecular genetic analysis and determining pathogenicity of *TUBB4A* gene mutations in an autosomal dominant pedigree and a sporadic case

Chapter 6. Molecular genetic analysis and determining pathogenicity of *TUBB4A* gene mutations in an autosomal dominant pedigree and a sporadic case

6.1. Introduction

Tubulins are building blocks of microtubules, which are the major components of the cytoskeleton in eukaryotic cells (Oakley, 2000). Tubulins are organised into αβ-tubulin heterodimers which bind head to tail to form a protofilament (Figure 6.1). About 13 protofilaments form a tubular structure (microtubule) with well-defined polarity (Akhmanova and Steinmetz, 2008).

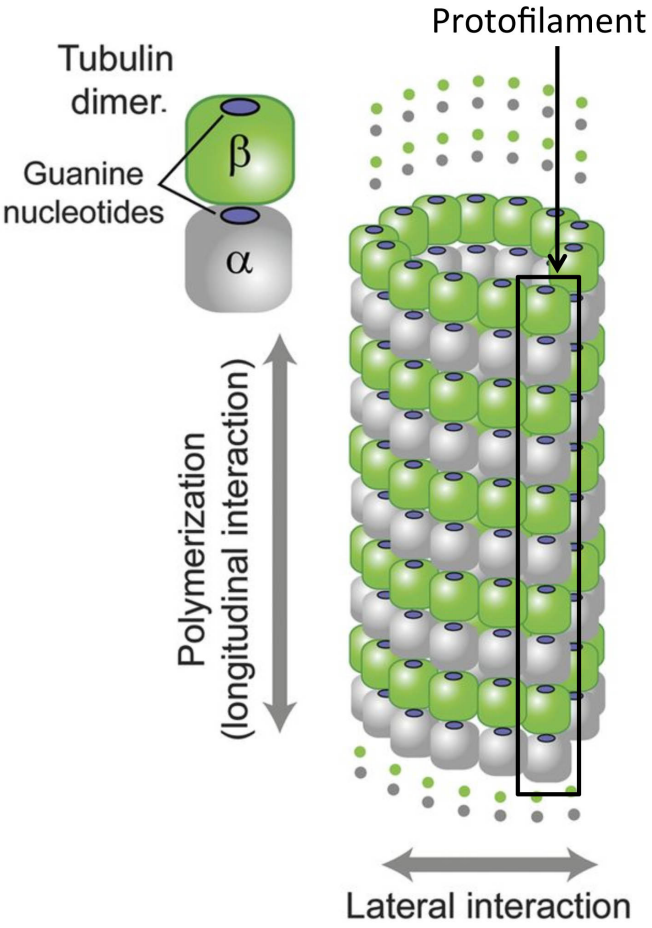


Figure 6.1: Tubulin heterodimer structure. αβ-Tubulin heterodimers polymerize head-to-tail along their longitudinal direction to form protofilaments. Protofilaments interact with each other laterally in order to organize into a tubular structure (microtubule) (adapted from Miyatake *et al*., 2014).

Microtubules are highly dynamic structures, which undergo assembly and disassembly (Kirschner and Mitchison, 1986). This process depends on guanine nucleotide binding and hydrolysis. Microtubules are involved in cell division including the formation of mitotic spindles, intracellular transport and cell motility. Intriguingly, β-tubulin contains a motif, which serves as a binding site for a pharmacological compound taxol (Turner and Margolis, 1984). Binding of taxol to β-tubulin leads to suppression of disassembly of microtubules and results in stabilization of the latter. This phenomenon made taxol and its derivatives as a drug of preference in cancer therapy because stabilized microtubules would supress cell division (Galassi, 1992). Strikingly, another group of drugs, which affect microtubule dynamics, was discovered recently. Compounds such as nocodazole and vinblastine were found to promote microtubule depolymerization (Jordan and Wilson, 2004).

The role of microtubules in the formation and function of the nervous system was extensively studied in the past. It is well established that the neuronal microtubules are required for the division and migration of neurons, neuronal differentiation and circuit formation (Kuijpers and Hoogenraad, 2011). Depletion of microtubules in axons and dendrites is associated with neurodegenerative diseases (Sudo and Baas, 2011). Therefore, microtubule-stabilizing drugs are a promising target for therapy of injury and diseases of the nervous system, resulting in the prevention of microtubules from degrading (Baas and Ahmad, 2013).

The pivotal role of microtubules in CNS development and maintenance was recently underscored by numerous reports implicating tubulins in a spectrum of neurological disorders. More recently several genes encoding α- and β-tubulin isotypes, crucial for microtubule assembly in neurons, have been linked to neurological phenotypes characterised by impaired neuronal migration, differentiation, axon guidance and maintenance. Missense mutations in *TUBA1A* (Tischfield *et al.*, 2011), *TUBB2B* (Jaglin *et al.*, 2009), *TUBB3* (Tischfield *et al.*, 2010) (Poirier *et al.*, 2010), *TUBB5* (Breuss *et al.*, 2012), *TUBA8* (Abdollahi *et al.*, 2009) and *TUBG1* (Poirier *et al.*, 2013) were all implicated.

More recently, *TUBB4A* was linked to several neurological phenotypes. *TUBB4A* is located on the chromosome 19 and encodes a member of β-tubulin family comprising 444 amino acids. TUBB4A is a major isotype in brain representing 46% of all beta tubulins. The highest expression in the brain is in the cerebellum, followed by putamen

and white matter (Leandro-Garcia *et al.*, 2010; Lohmann *et al.*, 2012).

In 2012 two groups independently studied an extended Australian family with dystonia type 4 (DYT4) (Hersheson *et al.*, 2012; Lohmann *et al.*, 2012). Both groups performed genetic linkage and subsequent exome sequencing and identified a missense *TUBB4A* p.Arg2Gly mutation as a cause of the disease in the family. Lohmann and colleagues studied *TUBB4A* mRNA expression level in different tissues in one affected person. Compared with controls, *TUBB4* RNA levels were decreased in all tissues of the affected person. The authors concluded that the disease mechanism is through reduced levels of *TUBB4* RNA rather than affecting its functionality. This group also screened 394 unrelated individuals with dystonia and identified one individual with a causative *TUBB4A* p.Ala271Thr mutation. Interestingly, shortly after these two studies were published, another neurological phenotype was linked to *TUBB4A*. Simons and colleagues employed whole exome sequencing in five family trios, two family quartets and three single probands who presented with hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC syndrome). They subsequently identified a *de novo* p.Asp249Asn mutation in *TUBB4A* as a cause of disease in all 11 affected individuals (Simons *et al.*, 2013). H-ABC is a rare neurological disease diagnosed on the basis of MRI findings such as hypomyelination and cerebellar atrophy. The mutation was predicted to affect longitudinal interaction between tubulin subunits. Strikingly, in one family, with unaffected parents and two affected siblings, detailed investigations revealed maternal mosaicism for the mutation. The authors reasoned that the phenotypically neutral *de novo* mosaic mutation in the mother caused the disease in the subsequent generation. Numerous reports on a causative role of *TUBB4A* in H-ABC soon followed the findings by Simons and colleagues. Presently, 19 additional *TUBB4A* mutations have been reported in H-ABC (Ferreira *et al.*, 2014; Hamilton *et al.*, 2014; Miyatake *et al.*, 2014; Purnell *et al.*, 2014)(Blumkin *et al.*, 2014)(Carvalho *et al.*, 2015; Erro *et al.*, 2015). Mutations in *TUBB4A* were recently reported in other neuroimaging phenotypes. Pizzino and colleagues performed whole exome sequencing on 5 patients from 4 families with hypomyelinating leukodystrophy with or without cerebellar atrophy but lacking other characteristic features of H-ABC such as basal ganglia abnormalities (Pizzino *et al.*, 2014). Four novel *de novo* mutations in *TUBB4A* were identified in all 5 affected individuals (p.Val255Ile, p.Arg282Pro, p.Gln292Lys, p.Arg391His). All these mutations were predicted to affect longitudinal interactions between tubulin subunits, similar to the p.Asp249Asn mutation published by Simons

and colleagues. In a more recent publication, Kancheva and colleagues applied whole exome sequencing to a highly consanguineous family with five children who presented with complicated hereditary spastic paraplegia (HSP), exhibiting regional hypomyelination, mild cerebellar atrophy, and no dystonia, basal ganglia atrophy, or cognitive dysfunction (Kancheva *et al.*, 2015). The authors identified a p.His190Tyr mutation in *TUBB4A* in all affected children. Noticeably, the mutation was also present in the asymptomatic mother suggesting mosaicism. Quantitative PCR in the mother's blood showed that the mutant allele was present at 10%, therefore supporting mosaicism

The precise molecular mechanism underlying *TUBB4A*-associated phenotypes remains to be elucidated.

This study aims to evaluate pathogenicity of two *TUBB4A* mutations, c.900G>T and c.1091C>A, and establish their causative role in two families who presented with ataxia and hypomyelinating leukodystrophy.

6.2. Materials and Methods

6.2.1. Patients

This study involved two families. Family 1 included three affected individuals (the mother P13 and daughters P11 and P12), who were subjected to whole exome sequencing. Both siblings presented during infancy with a severe disease phenotype. The mother presented in her late 30s with milder disease features (Appendices 1 and 2). Family 2 was represented by a female patient P14 with no family history (Appendices 1 and 2).

6.2.2. MRI scan

Magnetic resonance imaging (MRI) was performed at the Royal Victoria Infirmary (Newcastle upon Tyne Foundation Hospitals NHS Trust, UK) following the routine referrals of the patients to the regional neurogenetic service at Newcastle upon Tyne, UK.

6.2.3. Exome read depth at the position of the TUBB4A mutation c.900G>T

Bioinformatic analysis was performed as described in Section 2.4.2. Coverage analysis

of CCDS bases and known ataxia genes was performed as described in Sections 3.2.6 and 3.2.7, respectively. Analysis of the variant file in the variant caller Varscan, the raw Varscan output file, variant frequency and exome read depth at the position of the *TUBB4A* mutation were performed by Dr Helen Griffin (Newcastle University). The readings for the reference and variant alleles were obtained from Dr Helen Griffin (Newcastle University) and the ratios of reference to variant reads were calculated for patients P11, P12 and P13.

6.2.4. Conservation study

Multiple sequence alignment of TUBB4A, in human and animal orthologs, was performed using Clustal Omega software (available on http://www.ebi.ac.uk/Tools/msa/clustalo/).

6.2.5. Pyrosequencing

Pyrosequencing is a DNA sequencing technology, which is based on the sequencing-by-synthesis principle. The technique employs four enzymes, which work in a cascade reaction. Each event of incorporation of a nucleotide results in a release of pyrophosphate (PPi) in an equimolar quantity to the amount of nucleotide incorporated. PPi is then quantitatively converted to adenosine triphosphate (ATP). The end of the reaction is manifested by emission of light in an amount that is proportional to the amount of ATP. The light is detected as a peak on the pyrogram. The proportion of mutated alleles is calculated by the software by comparing wildtype and mutant peak heights.

The proportion of mutated alleles in various tissues was determined using a quantitative pyrosequencing assay according to the manufacturer's protocol. Pyromark Assay Design software v.2.0 (QIAGEN, Manchester, UK) was used to design gene-specific PCR and pyrosequencing primers (Table 6.1). Pyrosequencing was performed on the Pyromark Q24 platform (QIAGEN, Manchester, UK).

The assay was validated with a series of controls using a reference DNA sample. The controls were as follows:

1) PCR without template DNA. This control is used to check unspecific

amplification during PCR.

2) PCR with template DNA but with no sequencing primer. This control will ensure that the template is not forming a loop, which could result in a background signal in pyrosequencing reactions.

3) Sequencing primer without any PCR product. This control will check whether the sequencing primer forms duplexes or hairpins, which could result in unspecific DNA amplification during PCR.

4) Biotinylated primer without any PCR product. This control will check whether biotinylated primer forms duplexes or hairpins, which could result in a background signal in pyrosequencing reactions.

5) Sequencing primer and biotinylated primer together without PCR product. This control will ensure that the sequencing primer and the biotinylated primer don't form duplexes, which could result in a background signal in pyrosequencing reactions.

In order to pass the quality control, pyrograms from all the above controls should not show any significant peak after any nucleotide addition.

Table 6.1: Primer sequences used for mutation load analysis by pyrosequencing. The mutation is described along with the Gene Name and ID. Forward and reverse primers were designed for *TUBB4A* amplification prior to pyrosequencing analysis. One primer in each pyrosequencing primer pair had a 5' biotinylation modification (Bio-5'). An additional sequencing primer was required for pyrosequencing analysis which was used to assess mutational load. Pyrosequencing assays were designed using PyroMark assay design v.2.0 (QIAGEN, Manchester, UK) (this table was used in Pyle *et al*., 2014).

Gene Name and Mutation	Forward Amplification primer sequence	Reverse Amplification primer sequence	Pyrosequencing primer sequence
TUBB4A, c.900G>T	5' TAGCTGCTGTTCTTGCTCTGC 3'	Bio-5' CTGGCCGTCAACATGGTTC 3'	5' CGGGTCGCACGCCGC 3'

6.2.6. 3D structure modelling

Refined structure of bovine αβ-tubulin heterodimer, published by Lowe and colleagues (Lowe *et al.*, 2001), was obtained from Protein Database (PDB number 1JFF) and the stereo plot was obtained using the online program Cn3D (available at http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml). The mutated amino acids were mapped to the 3D structure of αβ-tubulin heterodimer. The assessment of the possible effect of the mutations was based on a linear structure of β-tubulin published by Wade and colleagues (Wade *et al.*, 2009).

6.3. Results

6.3.1. MRI scan

The images were kindly provided by Newcastle upon Tyne Foundation Hospitals NHS Trust, UK. In family 1, MRI for the mother (patient P13, Figure 6.2 A and B) and daughter (patient P12, Figure 6.2 C and D) revealed leukodystrophy and cerebellar atrophy in both patients (Figure 6.2). T2 and T1 images (Figure 6.2 A and B) revealed generalized atrophy and periventricular high signal, indicative of an abnormality within the white matter. T2 images (Figure 6.2 C and D) showed marked cerebellar atrophy and diffuse hypomyelination. MRI for patient P13 revealed similar findings to patient P12 (image is not provided). In family 2 (patient P14) MRI revealed diffuse hypomyelination including brainstem and diffuse white matter changes (image is not provided).

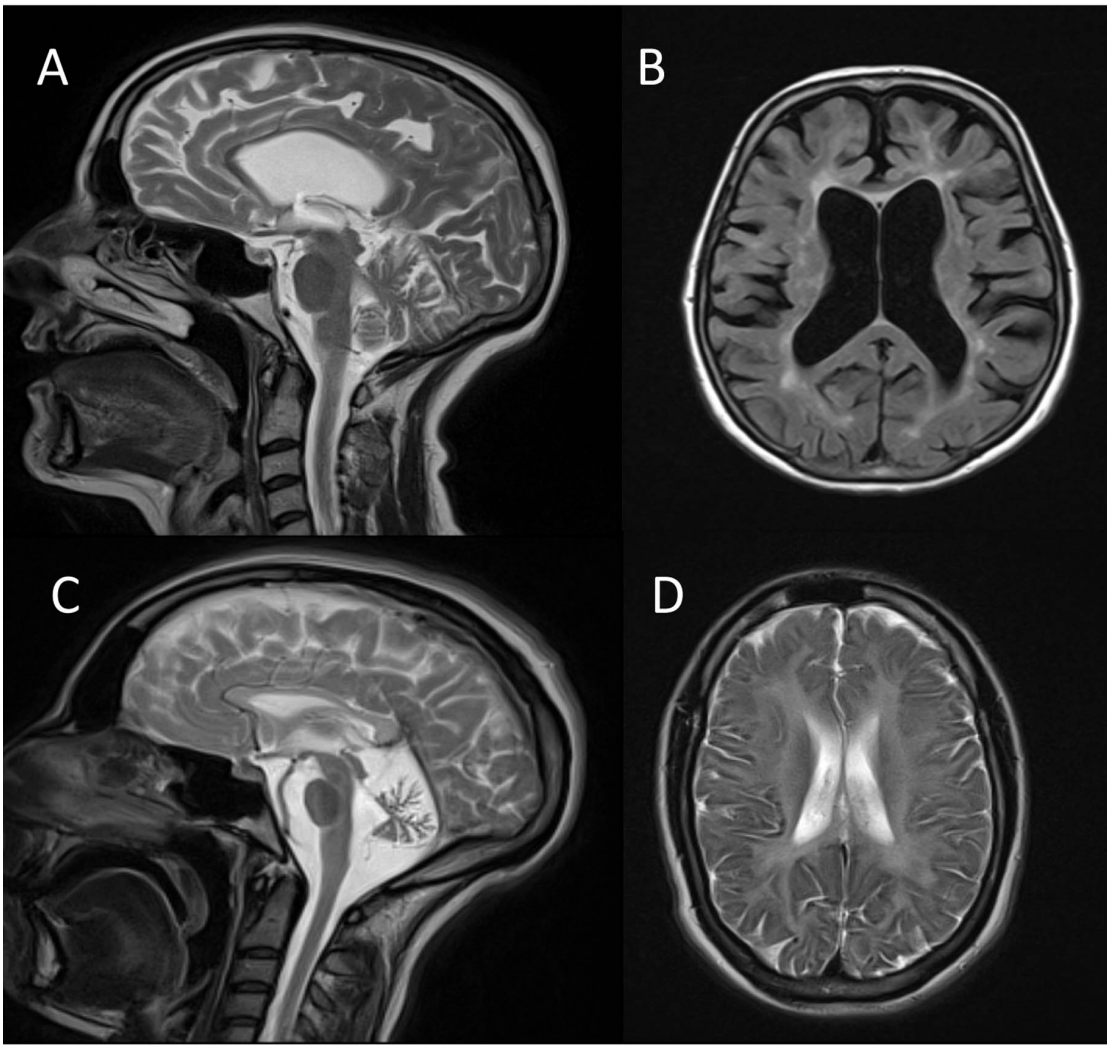


Figure 6.2: MRI from the mother P13 and daughter P12. Brain MRI for patient P13, the mother (A and B) and patient P12, the daughter (C and D) revealed leukodystrophy and cerebellar atrophy in both patients. T2 and T1 images (A and B) revealed generalized atrophy and periventricular high signal, indicative of an abnormality within the white matter. T2 images (C and D) showed marked cerebellar atrophy and diffuse hypomyelination. Adapted from Pyle *et al*., 2014

6.3.2. Analysis of the exome read depth at the position of the p.Met300Ile mutation in family 1

Exome read depth at the position of the TUBB4A mutation revealed that both offspring had a normal ratio of reference to variant reads (approximately 1:1) reflecting heterozygosity. The mother had an abnormal ratio (13:1) suggesting mosaicism (Table

6.2). With no history of the disorder in previous generations of this family, it was suspected that the mother had a *de novo* mutation.

Table 6.2: Exome read depth at the position of the p.Met300Ile mutation in patients P11, P12 and P13.

Patient	Target	Chr	Variant Position	Coverage	Ref-base	Var-base	Ref-reads	Var-reads	% Var-freq	Ratio Ref:Var
P11	Chr19:6494330-6495958:*TUBB4A*	19	6495610	400	C	A	162	146	47.4	1.1:1
P12				477			145	148	50.51	0.98:1
P13				422			359	27	6.99	13.3:1

6.3.3. Conservation of residues p.Met300Ile and p.Ala364Asp

Both the p.Met300Ile and p.Ala364Asp mutations were found to be evolutionary conserved (Figures 6.3 and 6.4, respectively).

Figure 6.3: Multiple sequence alignment of a section of human TUBB4A and animal orthologs. Residue Met300 is boxed in red.

Figure 6.4: Multiple sequence alignment of a section of human TUBB4A and animal orthologs. Residue Ala364 is boxed in red.

6.3.4. Quantification of the mutated allele p.Met300Ile with pyrosequencing (family 1)

Validation of the assay revealed that pyrograms from all the validation controls didn't show any significant peaks after any nucleotide addition (Figure 6.5).

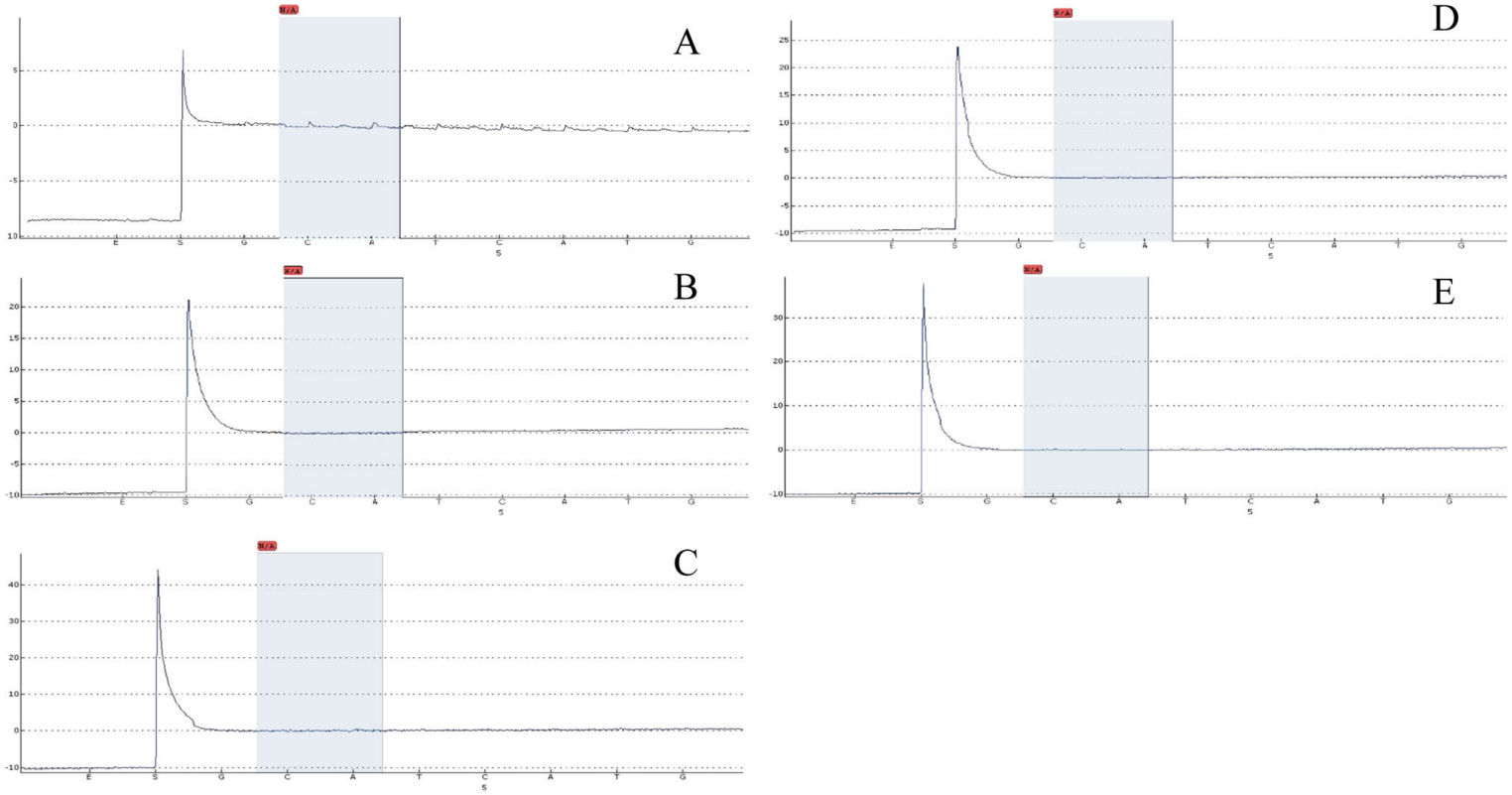


Figure 6.5: Pyrosequencing assay validation. This figure depicts pyrograms for control pyrosequencing runs: PCR without template DNA (A); PCR with template DNA but with no sequencing primer (B); Sequencing primer without any PCR product (C); Biotinylated primer without any PCR product (D); Sequencing primer and biotinylated primer together without PCR product E).

Proportion of the mutated allele was determined by pyrosequencing. This study tested blood, hair root cells, buccal epithelial cells and fibroblasts in the mother and blood cells in both offspring. The result supported mosaicism in the mother (Figure 6.6, Figure 6.7, Table 6.3).

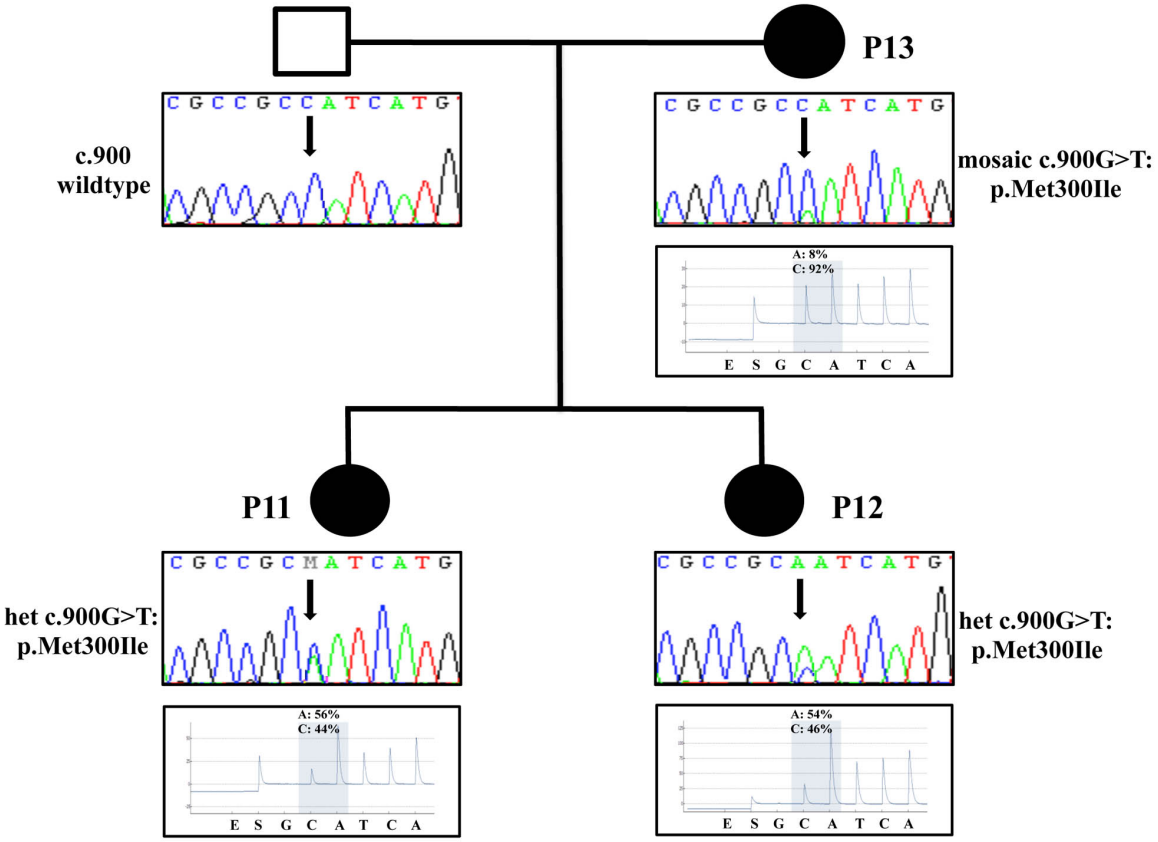


Figure 6.6: Confirmatory Sanger sequencing and pyrosequencing in blood DNA samples from the mother (P13) and her daughters P11 and P12.

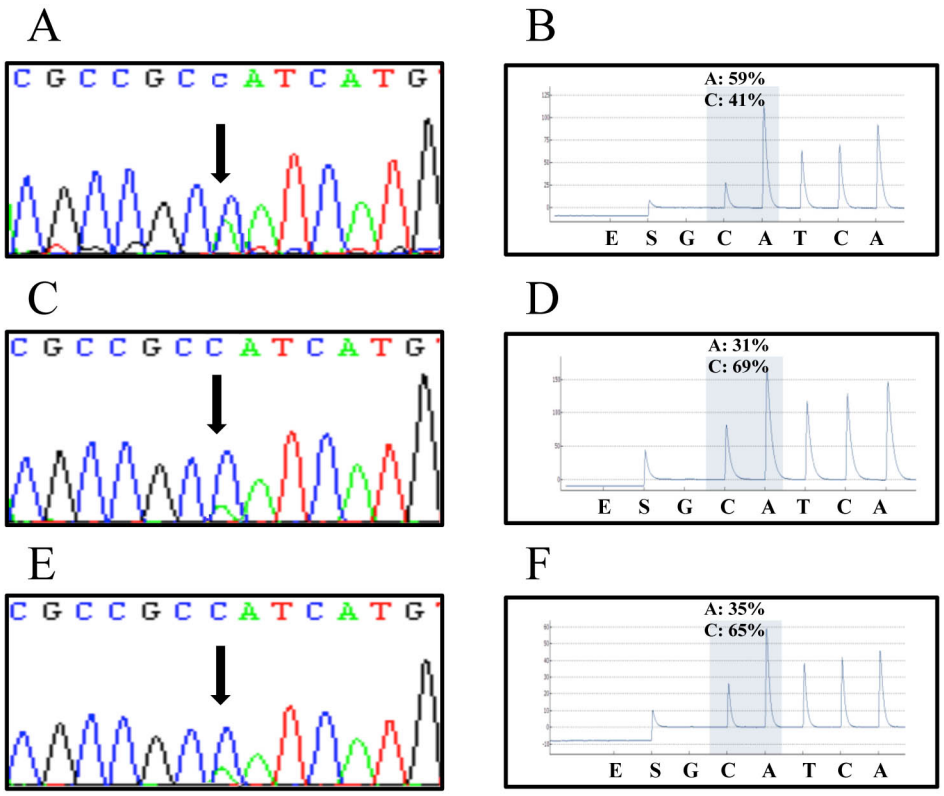


Figure 6.7: Confirmatory Sanger sequencing (left) and pyrosequencing (right) in hair root cells (A, B), buccal epithelial cells (C, D), and fibroblasts in the mother (P13).

Table 6.3: Pyrosequencing result for allele quantification of the p.Met300Ile mutation in different tissue in the mother (P13) and her daughters (P11 and P12). This table was used in Pyle *et al*., 2014

Patient	***Tissue***	***Pyrosequencing***	
		% variant	***% reference***
P13	Blood	8	92
	Hair root cells	59	41
	Buccal epithelial cells	31	69
	Fibroblasts	35	65
P11	Blood	56	44
P12	Blood	54	46

6.3.5. Investigation of putative pathogenicity of TUBB4A mutations p.Met300Ile and p.Ala364Asp via 3D modelling

3D structure of αβ-tubulin heterodimer (Lowe *et al*., 2001), with mapped mutated residues, is presented in Figure 6.8. The assessment of the possible effect of the mutations was based on a linear structure of β-tubulin published by Wade and colleagues (Wade *et al.*, 2009). In family 1 (patients P11, P12 and P13) the mutation p.Met300Ile might affect longitudinal interactions between tubulin monomers, possibly affecting polymerizations of αβ-tubulin heterodimers

In family 2 (patient P14) the p.Ala364Asp mutation affects a residue which is directly involved in the taxol binding. Therefore, the mutation might affect taxol binding and therefore prevent stabilizing of microtubules.

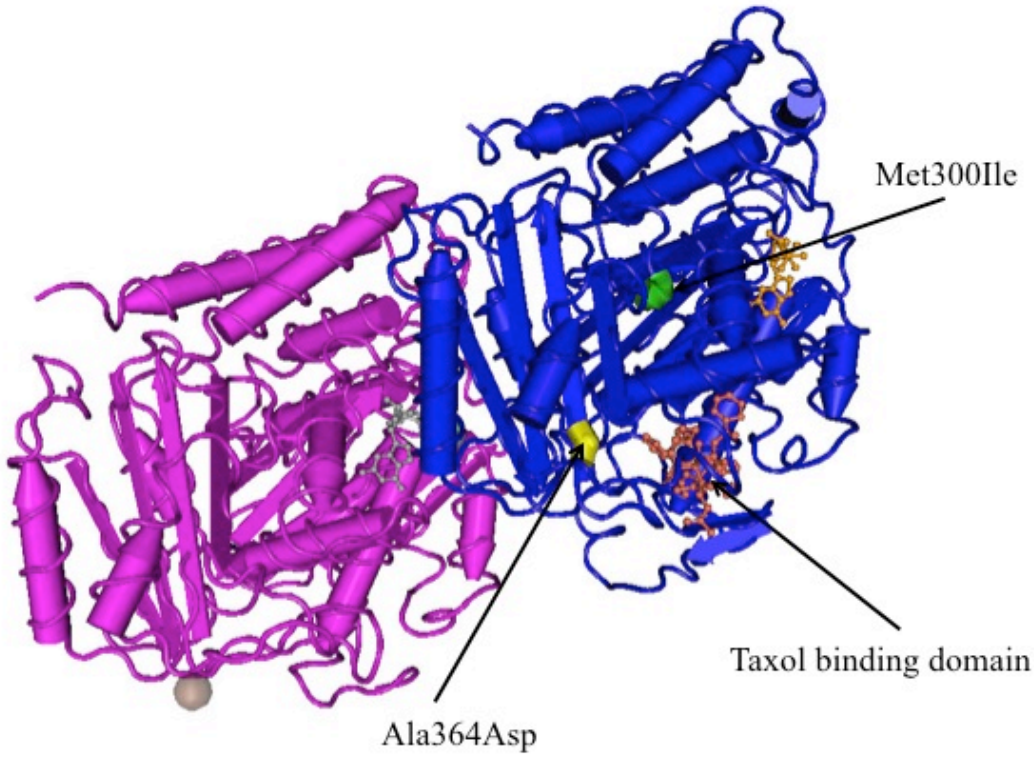


Figure 6.8: 3D structure of αβ-tubulin heterodimer (Lowe *et al.*, 2001). α-tubulin subunit is represented in magenta and β-tubulin in blue. Helices, β-sheets and loops are represented by crayons, flat arrows and threads, respectively. Taxol binding motif is shown in orange. The green and yellow tubes represent Met300Ile and Ala364Asp mutated residues, respectively

6.4. Discussion

Using whole exome sequencing, likely *de novo* dominant *TUBB4A* mutations were found in two families who presented with cerebellar ataxia. In family 1 (patients P11, P12 and P13) MRI additionally revealed leukodystrophy in the affected individuals. This studye found that the *TUBB4A* p.Met300Ile mutation segregated with the disease in this family. In family 2 the affected individual (patient P14) exhibited diffuse hypomyelination including brainstem (revealed by MRI). The *TUBB4A* p.Ala364Asp mutation segregated with the disease in the latter family. Both these mutations are evolutionary conserved (Figure 6.3 and Figure 6.4).

TUBB4A seemed to fit the phenotype in the both families (Simons *et al.*, 2013). However, in family 1 exome sequencing revealed that the variant p.Met300Ile was in the heterozygous state in both offspring, but wildtype in the mother. The mother presented with a similar, but much milder phenotype and later onset, which led this study to suspect mosaicism. Sanger sequencing of blood DNA samples from all three patients confirmed the mutation in the both offspring, but, strikingly, revealed heterozygozity in the mother as well. This finding pointed towards mosaicism and prompted revisiting the exome sequencing output data. The current study first analyzed the variant file in the variant caller Varscan and found that the mother (P13) didn't carry the variant allele for the p.Met300Ile mutation. This finding prompted a detailed investigation of the raw Varscan output file. This study found that the mother had a predicted variant frequency of 6.99% (Varscan detected 359 reference allele readings and 27 variant allele readings at the chromosome position Chr19:6495610). It was reasoned that these readings allowed the variant to pass the Varscan base quality filter. However, the variant didn't get reported in the 'on-target' variant file because this file is filtered for variants with frequencies greater than 25%. The ratios of reference to variant reads were calculated and found that both offspring had a normal ratio of reference to variant reads (approximately 1:1) reflecting heterozygosity, the mother had an unexpected ratio (13:1), suggesting mosaicism. Exome sequence read depth also revealed an even coverage of the mutated base at approximately 400-fold in all three patients. Similar findings were recently reported in the literature (Simons *et al.*, 2013). In order to establish a molecular cause of the disease, Simons and co-workers (2013)

performed exome sequencing in five family trios, two family quartets, and three single probands with a total of 11 individuals, who presented with hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC). The authors found that all 11 individuals carried the same *de novo* p.Asp249Asn mutation in *TUBB4A*. One quartet with two affected children and seemingly unaffected parents was investigated further. Exome sequencing showed that the siblings and the mother were heterozygous for the mutation, while the father was wildtype. More detailed investigation of the exome readings revealed that in the heterozygous mother the ratio of wildtype to alternative reads is much higher than in all other heterozygous patients from the study. Taking into account that the mutation was covered at least 50-fold in all samples, the authors concluded that the mother is mosaic for that variant. Deep sequencing (MiSEQ platform, Illumina) of DNA samples from blood, saliva and buccal cells, from all individuals of the quartet, revealed that in the asymptomatic mother the alternative allele was present in 25% of reads from blood, 29% of reads from buccal cells and 26% of reads from saliva. The result confirmed mosaicism in the mother and showed that a rare *de novo* mutation, which is phenotypically neutral in a mosaic individual, could be transmitted in a dominant manner and result in a disease phenotype in the subsequent generation.

In the family 1, in order to determine the proportion of alternative alleles in different tissues in the mother (P13), this study performed quantitative pyrosequencing in blood, hair root cells, buccal epithelial cells and fibroblasts. To allow comparison, this study also performed quantitative pyrosequencing in blood cells in both offspring. Confirmatory Sanger sequencing was also performed in all these tissues. The p.Met300Ile mutation showed varying degrees of tissue mosaicism in the mildly affected mother and heterozygosity in the two severely affected offspring (Figures 6.6 and 6.7 and Table 6.3). Exome reads support this finding (Table 6.2).

The current study also investigated how the two *TUBB4A* mutations, mentioned above, might affect the microtubule structure. The mutations were mapped to the refined structure of bovine αβ-tubulin heterodimer (Lowe *et al.*, 2001) and the data was visualised by Cn3D software (Figure 6.8). Based on the TUBB4A structure, it was predicted that the p.Ala364Asp mutation could disrupt the taxol binding domain (Wade

et al., 2009), possibly affecting stabilization of microtubules. The p.Met300Ile mutation could affect longitudinal interactions between tubulin monomers (Wade *et al.*, 2009), possibly disrupting polymerizations of αβ-tubulin heterodimers

The precise molecular mechanism of *TUBB4A* mutations, discovered to this day, is not known. However, the proposed working hypothesis is alteration of microtubule dynamic and stability (Hamilton *et al.*, 2014). It is also proposed that the defective microtubule system would impair axonal transport, which would lead to axonal loss and subsequently to hampering of the myelination process. Disrupted axonal transport could also account for the loss of neurons in basal nuclei and cerebellum. Indeed, impaired microtubule dynamic and stability was recently found to be a common functional consequence of *TUBB4A* mutations (Table 6.4).

Table 6.4: Functional consequences of the reported *TUBB4A* mutations. This table depicts all reported mutations in *TUBB4A* and the phenotypes associated with these mutations, along with the effect of the mutations (modified from Carvalho *et al.,* 2015).

Nucleotide change	Amino acid change	Number of patients	References (number of cases per article)	Phenotype	Effect of the mutation (interaction affected)
c.4C>G	p.Arg2Gly	9	(Hersheson *et al.*, 2012)	DYT4	Auto-regulation of mRNA stability
c.4C>T	p.Arg2Trp	1	(Hamilton *et al.*, 2014)	H-ABC	Auto-regulation of mRNA stability
c.5G>A	p.Arg2Gln	2	(Miyatake *et al.*, 2014) (1) (Hamilton *et al.*, 2014) (1)	H-ABC	Auto-regulation of mRNA stability
c.467G>T	p.Arg156Leu	1	(Purnell *et al.*, 2014)	H-ABC	Lateral
c.533C>G	pThr178Arg	1	(Miyatake *et al.*, 2014)	H-ABC	Lateral
c.568C>T	p.His190Tyr		(Kancheva *et al.*, 2015)	HSP	Lateral
c.716G>T	p.Cys239Phe	1	(Hamilton *et al.*, 2014) (Ferreira *et al.*, 2014)	H-ABC	Taxol binding
c.730G>A	p.Gly244Ser	4	(Hamilton *et al.*, 2014) (3) (Carvalho *et al.*, 2015) (1)	H-ABC	Longitudinal
c.731G>T	p.Gly244Val	1	(Hamilton *et al.*, 2014)	H-ABC	Longitudinal
c.745G>A	p.Asp249Asn	30	(Miyatake *et al.*, 2014) (2) (Hamilton *et al.*, 2014) (16) (Ferreira *et al.*, 2014) (1) (Simons *et al.*, 2013) (11)	H-ABC	Longitudinal

Table 6.4: continued

Nucleotide change	Amino acid change	Number of patients	References (number of cases per article)	Phenotype	Effect of the mutation (interaction affected)
c.763G>A	p.Val255Ile	1	(Pizzino *et al.*, 2014)	Isolated hypomyelination	Longitudinal
c.785G>A	p.Arg262His	2	(Miyatake *et al.*, 2014) (1) (Ferreira *et al.*, 2014) (1)	H-ABC	Kinesin interaction
c.845G>C	p.Arg282Pro	2	(Pizzino *et al.*, 2014)	Isolated hypomyelination	Lateral
c.874C>A	p.Gln292Lys	1	(Pizzino *et al.*, 2014)	Isolated hypomyelination	Lateral
c.900G>T	p.Met300Ile	3	This study (Pyle *et al.*, 2014)	leukodystrophy	Longitudinal
c.941C>T	p.Ala314Val	1	(Erro *et al.*, 2015)	H-ABC	Lateral
c.968T>G	p.Met323Arg	1	(Hamilton *et al.*, 2014)	H-ABC	Longitudinal
c.1045G>A	p.Ala352Thr	1	(Hamilton *et al.*, 2014)	H-ABC	Taxol binding
c.1061G>A	p.Cys354Tyr	1	(Hamilton *et al.*, 2014)	H-ABC	Taxol binding
c.1091C>A	p.Ala364Asp		This study (Pyle *et al.*, 2014)	hypomyelination	Taxol binding
c.1099T>A	p.Phe367Ile	1	(Hamilton *et al.*, 2014)	H-ABC	Taxol binding
c.1099T>C	p.Phe367Leu	1	(Hamilton *et al.*, 2014)	H-ABC	Taxol binding
c.1162A>G	p.Met388Val	2	(Miyatake *et al.*, 2014) (1) (Hamilton *et al.*, 2014) (1)	H-ABC	Longitudinal
c.1163T>C	p.Met388Thr	1	(Hamilton *et al.*, 2014)	H-ABC	Taxol binding

Table 6.4: continued

Nucleotide change	Amino acid change	Number of patients	References (number of cases per article)	Phenotype	Effect of the mutation (interaction affected)
c.1164G>A	p.Met388Ile	1	(Hamilton *et al.*, 2014)	H-ABC	Taxol binding
c.1172G>A	p.Arg391His	1	(Pizzino *et al.*, 2014)	Isolated hypomyelination	Longitudinal
c.1181T>G	p.Phe394Cys	1	(Carvalho *et al.*, 2015)	H-ABC	Lateral
c.1228G>A	p.Glu410Lys	3	(Miyatake *et al.*, 2014) (2) (Blumkin *et al.*, 2014) (1)	H-ABC	Lateral, kinesin interaction

Kancheva and co-workers mapped known *TUBB4A* mutations to a 3D model in order to establish genotype-phenotype correlation (Figure 6.9). The most common mutation causing H-ABC (p.Asp249Asn) was found to cause the most benign H-ABC, in comparison with other *TUBB4A* mutations causing H-ABC. Interestingly, there are several mutations of Arg2. A heterozygous change of Arg2 to glycine caused a milder phenotype (dystonia, DYT4), whereas an Arg2 to tryptophan or glutamine causes a more severe phenotype H-ABC (Table 6.4). The possible explanation for this could be that a change to the chemically inert glycine would have a very mild effect on the function of the resulting TUBB4A. On the other hand, an Arg2 change to the polar glutamine or hydrophobic tryptophan would probably lead to a drastic change of TUBB4A function (Hamilton *et al.*, 2014).

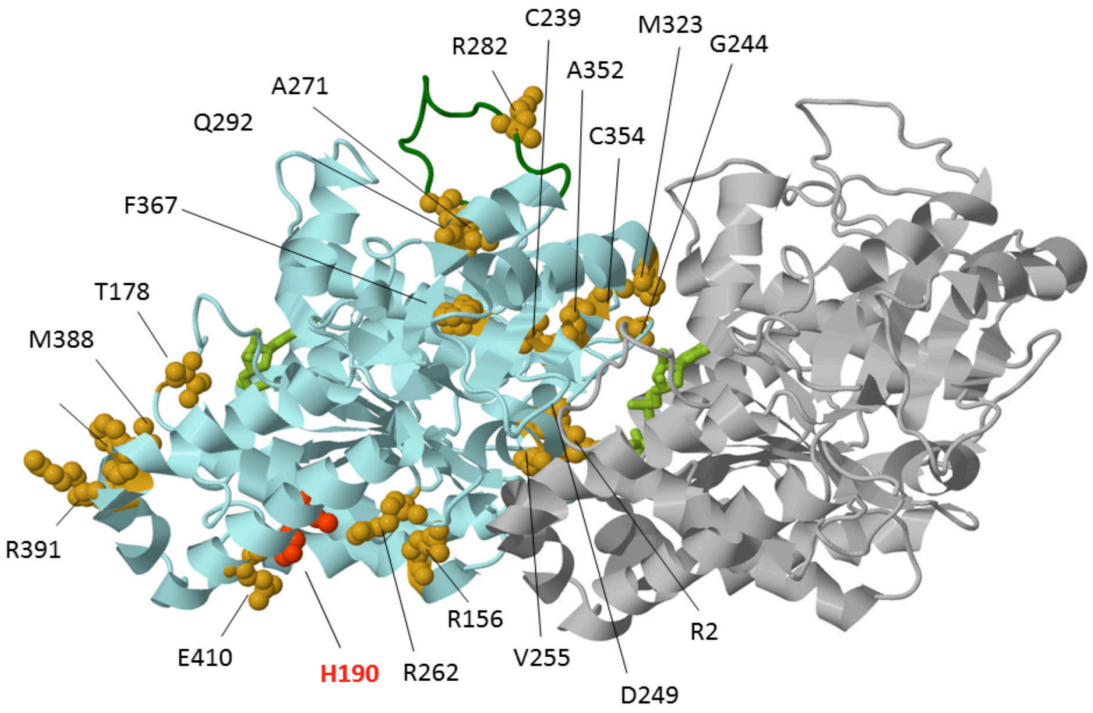


Figure 6.9: Published TUBB4A mutations mapped to a 3D structural model of TUBB4A. This picture depicts the 3D structure of TUBB4A with the β tubulin subunit represented in light blue and the α subunit in grey. Alpha helixes are depicted as ribbons and beta strands are shown as arrows. Loops are presented as threads. The m-loop is shown as a green thread. The GDP and GTP molecules are depicted in light green. The mutated residues are presented as yellow spheres. Adapted from Kancheva *et al.*, 2015.

To elucidate the precise molecular mechanism of the *TUBB4A* mutations and to

understand how these mutations result in a clinical phenotype and imaging abnormalities, functional studies are warranted. For the c.900G>T mutation, which occurred in the family with three affected individuals, it would be worthy to undertake a transcriptional study. Different tissues of the mother could be employed in order to determine the expression levels of *TUBB4A* mRNA when compared to controls. In order to investigate the effect of both mutations (c.900G>T and c.1091C>A) on microtubule dynamics *in vitro*, it would be interesting to transfect wildtype and mutant FLAG-tagged human TUBB4A expression constructs into commercial cells (for example HEK293). To observe incorporation of wildtype and mutant constructs into the microtubule polymer, co-localisation for transgenic and endogenous tubulins should be performed. For this the cells are stained sequentially with anti-FLAG (to detect transgenic TUBB4A) and anti- α-tubulin (to detect endogenous tubulin) and the result is visualised by means of confocal microscopy.

It is of particular interest that mutations in *TUBB4A* were recently associated with several neuroimaging phenotypes, namely dystonia type 4 (DYT4) (Hersheson *et al.*, 2012; Lohmann *et al.*, 2012), hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC) (Simons *et al.*, 2013), isolated hypomyelination (Pizzino *et al.*, 2014) and hereditary spastic paraplegia (HSP) with no basal ganglia involvement or cognitive impairment (Kancheva *et al.*, 2015). This makes *TUBB4A* a primary candidate for genetic screening, following MRI, meaning that the overall cost of a diagnosis could be greatly reduced. It is clear, that to this day the full spectrum of neurological phenotypes associated with *TUBB4A* mutations is not exhausted. Therefore, it is anticipated that whole exome sequencing will be instrumental in discovering new *TUBB4A* mutations.

Chapter 7

Molecular genetic analysis and determining pathogenicity of mutations in *ZFYVE26*, *FASTKD2* and *KCNB2*

Chapter 7. Molecular genetic analysis and determining pathogenicity of mutations in *ZFYVE26*, *FASTKD2* and *KCNB2*

7.1. Introduction

Delineating the molecular cause of clinically and genetically heterogeneous ataxias is important because treatment is available for some ataxias (Jayadev and Bird, 2013). Whole exome sequencing is rapidly becoming instrumental in establishing a genetic diagnosis in ataxia. New disease genes are being discovered, new mutations in previously implicated genes are being identified and the phenotypic spectrum of ataxia is being broadened. However, the interpretation of exome sequencing data is not always straightforward. Identification of a new disease gene or an unexpected gene would need further confirmation by functional studies and sometimes re-assessment of previous clinical tests such as neuroimaging.

In this chapter investigation of the potential pathogenicity of the variants in the following three genes is described: *ZFYVE26, FASTKD2* and *KCNB2*. These genes were identified following exome sequencing approach in 35 affected individuals with suspected inherited ataxia.

ZFYVE26

Zinc Finger, FYVE Domain Containing 26 (*ZFYVE26*) gene encodes a protein of 2539 amino acids, which is localized to the cytoskeleton and the lysosome. The ortholog of spastizin in zebrafish is required for spinal motor neuron axon outgrowth (Martin *et al.*, 2012). Recently spastizin was shown to mediate the formation of new lysosomes and therefore, to be a key player in lysosomal biogenesis (Chang *et al.*, 2014). Mutations in *ZFYVE26* were identified as the cause of hereditary spastic paraplegia type 15 (SPG15) (Hanein *et al.*, 2008). Subsequently, thin corpus callosum was identified as an MRI hallmark of SPG15 (Goizet *et al.*, 2009). Interestingly, mutations in *ZFYVE26* were recently identified in a patient with cerebellar ataxia (Fogel *et al.*, 2014).

FASTKD2

A fas activated serine-threonine domain 2 (*FASTKD2*) gene encodes a protein, which is localized in the mitochondrial inner compartment. FASTKD2 was predicted to be involved in regulating apoptosis in breast cancer (Yeung *et al.*, 2011). A nonsense mutation in this gene was found in two siblings with infantile mitochondrial

encephalopathy associated with a deficiency of cytochrome c oxidative activity, however the molecular mechanism of the disease was not elucidated (Ghezzi *et al.*, 2008). Recently, decreased expression of *FASTKD2* in astrocytes was associated with Alzheimer's disease, which suggested that FASTKD2 might be involved in apoptosis regulation in astrocytes (Sekar *et al.*, 2015).

KCNB2

A potassium channel, voltage gated subfamily B, member 2 (*KCNB2*) gene encodes a member of a diverse voltage-gated potassium (Kv) channel family. Kv channels are pivotal mediators of neuronal excitability and key participants in the signalling pathway that regulates neuronal apoptosis (Guan *et al.*, 2007). Voltage-gated potassium channels form protein pores, which open and close in response to the voltage difference across the membrane, allowing potassium ion transport in excitable membranes (Jensen *et al.*, 2012). The precise function of KCNB2 is not known. However, in a recent study Guan and co-workers showed that *KCNB2* mRNA and protein are expressed in the dendrites of pyramidal neurons in rat cortex (Guan *et al.*, 2007). The authors therefore speculated that KCNB2 might play a role in regulating distal dendritic excitability. Ion channels are of a particular interest due to the recent link between mutations in these genes and several neurological phenotypes, including ataxia (Kullmann and Waxman, 2010). Intriguingly, voltage-gated potassium channels are predicted to become the latest therapeutic targets in the treatment of a wide range of neurological disorders (Shah and Aizenman, 2014).

The aim of this study is to elucidate pathogenicity of mutations in *ZFYVE26* (c.2450delT:p.Leu817Cysfs*12 and c.2338C>T:p.Arg780*; patient P19), *FASTKD2* (c.-66A>G and c.149A>G:p.Lys50Arg; patient P21) and *KCNB2* (p.Ser530Phe and p.Leu784Pro; patient P26) by means of functional studies. For this, transcription (*FASTKD2*) and segregation studies (*ZFYVE26* and *KCNB2*) will be performed.

7.2. Materials and Methods

7.2.1. In silico mutation analysis (splice-site prediction with Alamut software)

In silico prediction of pathogenicity of the putative causative variants was assessed with Alamut software, as described in Section 4.2.4.

7.2.2. Allelic cis-trans study via cloning of genomic DNA (sequencing of cloned alleles)

PCR amplicons were generated as described in Section 2.5.2 using gene-specific primers (Table 7.1) and electrophoretically separated on a 1.5% agarose gel as described in Section 2.5.3. Following separation, the bands were excised from the gel, DNA was extracted, cloned and sequenced, as described in Section 2.6.

Table 7.1: Primers used for generating PCR amplicons for *KCNB2* and *ZFYVE26*

Gene	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'	Anneal temp. (˚C)	PCR product size (bp)
KCNB2	GAAACAAGCTCCA ACAAGTC	CCTGGGAGCTCTA AGAAGTCT	58	959
ZFYVE26	TGCTCCCAGGTGG CTCTGGT	AGCCCACCCTTTG CCTTTGGT	58	904

7.2.3. Sequencing of fibroblast cDNA (total PCR product and cDNA cloned from extracted PCR fragments)

Reverse transcription PCR (RT-PCR) was performed as described in Section 2.9 using *FASTKD2* specific primers (Table 7.2).

Table 7.2: ***FASTKD2* primers used for RT-PCR**

***FASTKD2* primer**	**Sequence**
Forward	5′- GAAGCGCAGCTTCTCGGGGA -3′
Reverse	5′- CATTGATACGTTCCTGGGTCAC -3′

PCR conditions included a hot start at 95^{o}C for 3 minutes, followed by 35 cycles of denaturation at 95^{o}C for 1 minute, annealing at 53^{o}C for 1 minute, and extension at 72^{o}C for 2 minutes. The final extension step was performed at 72^{o}C for 10 minutes, followed by cooling the sample to 4^{o}C. Sanger sequencing of total PCR product was performed using *FASTKD2* specific primers (Table 7.2), as described in Section 2.5.

The efficiency of the cDNA amplification from fibroblasts of controls and the patient

was assessed by conducting a control RT-PCR test using primers specific for a 353 bp fragment of the housekeeping gene beta-actin (Table 7.3)

Table 7.3: Beta-actin primers for mRNA quality control RT-PCR

Beta actin primer	Sequence
Forward	5′- GCTCGTCGTCGACAACGGCTC -3′
Reverse	5′- CAAACATGATCTGGGTCATCTTCTC -3′

The quality control RT-PCR was performed as described in Section 2.9.3. PCR conditions were as follows: a hot start at 95^{o}C for 2 minutes followed by 30 cycles of denaturation at 95^{o}C for 15 seconds, annealing at 58^{o}C for 30 seconds and extension at 72^{o}C for 1 minute. The reaction was finished by a final extension step performed at 72^{o}C for 10 minutes, followed by cooling the sample to 4^{o}C.

Sanger sequencing of plasmid DNA cloned from the extracted PCR fragments was performed as described in Section 2.6.

7.3. Results

7.3.1. ZFYVE26 (patient 19)

This male patient presented at age 32 with cerebellar ataxia. There was no family history. Whole exome sequencing identified one frameshift mutation (c.2450delT:p.Leu817Cysfs*12) and one stop gain mutation at a highly conserved residue (c.2338C>T:p.Arg780*) in *ZFYVE26* in patient P19. The p.Arg780* is predicted to be disease causing or damaging by three prediction programmes (Table 3.4). Detailed clinical investigations are presented in Table 4.5. Additional clinical data are presented in Appendix 1.

Allelic cis-trans study via cloning of genomic DNA

The parental samples were unavailable for analysis. Segregation of alleles was confirmed with sequencing of the cloned alleles.

PCR amplicon was generated from the patient's genomic DNA using *ZFYVE26* specific primers spanning both mutations. Agarose gel electrophoresis of the PCR amplicon

revealed amplification of a 904 bp PCR product in patient P19 (not shown). The band was excised from the gel and the DNA fragment was ligated into pGEM®-T easy vector. Competent cells JM109 were then transformed with the ligation reaction, and the cells were plated out on agar plates supplemented with ampicillin, IPTG and X-Gal to allow blue/white colonies selection. Plasmid DNA was extracted from the single white colonies and Sanger sequenced.

31 clones were sequenced. Sanger sequencing of the cloned alleles revealed that the mutations occur in *trans* (Figure 7.1). There were two types of alleles each containing only one homozygous mutation. Allele A contained a homozygous mutation c.2338C>T:p.Arg780* and a wildtype c.2450T (Figure 7.1, A). Allele B contained a homozygous mutation c.2450_2450delT:Leu817Cysfs*12 and wildtype c.2338C (Figure 7.1, B). Out of 31 clones, 17 were type A and 14 represented type B allele.

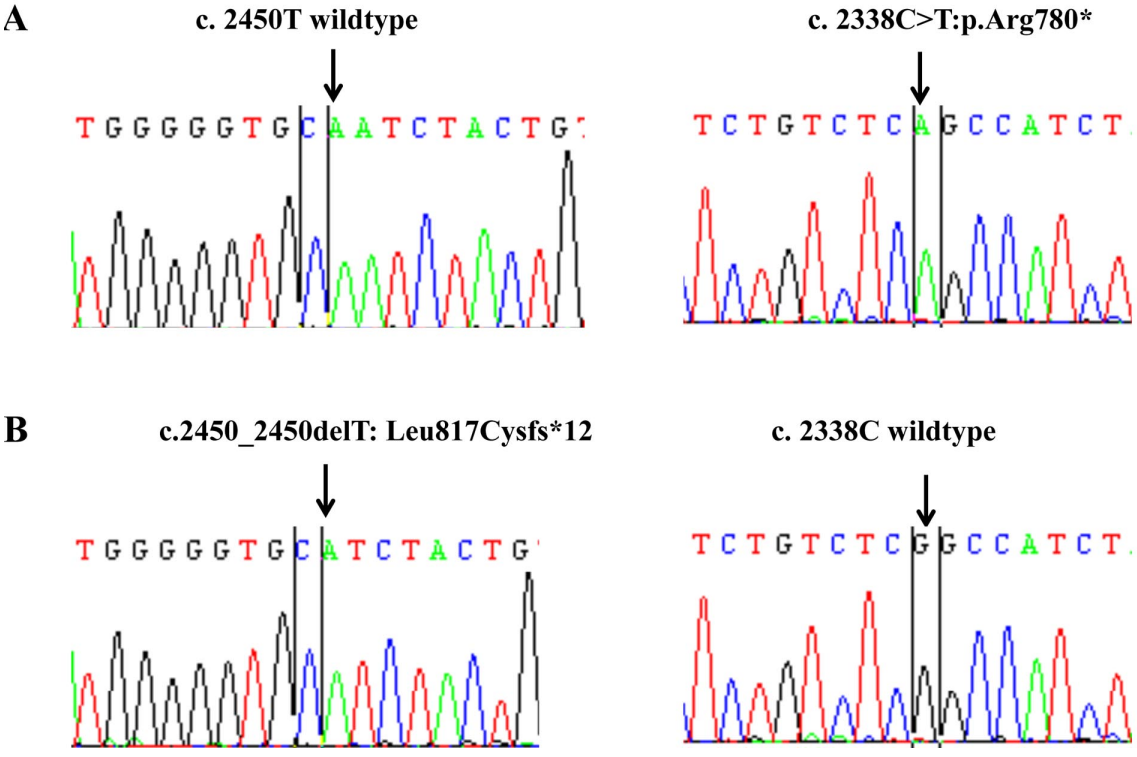


Figure 7.1: Sanger sequencing chromatograms for cloned alleles types A and B. This figure depicts two types of alleles that were cloned. Allele A containes a homozygous mutation c.2338C>T:p.Arg780* and a wildtype c.2450T (A). Allele B contained a homozygous mutation c.2450_2450delT:Leu817Cysfs*12 and wildtype c.2338C (B). The mutations are inherited in *trans*.

7.3.2. FASTKD2 (patient P21)

This female patient presented in infancy with a walking delay at 23 months, wobbly

walking and frequent falls. There was no evidence of family history. Whole exome sequencing identified putative pathogenic compound heterozygous *FASTKD2* mutations (c.-66A>G and c.149A>G:p.Lys50Arg) in this patient with ataxia (Table 4.5). Sanger sequencing confirmed segregation of these variants with disease in the family (Appendix 1). The p.Lys50Arg mutation is rare in the population and conserved among species (Table 3.4).

In silico mutation pathogenicity study (splice-site prediction with Alamut software)

Splice predictions for *FASTKD2* (NM_001136194.1) c.-66A>G show a gain of acceptor site at c.-65 by one program and loss of acceptor site at –c.64 by another program (Alamut v2.4, Interactive Biosoftware, Rouen, France) (Figure 4.1).

Sequencing of FASTKD2 fibroblast cDNA (total PCR product and plasmid DNA cloned from extracted PCR fragments)

Prior to the transcription study, the efficiency of the cDNA amplification was assessed by means of RT-PCR using primers specific for the housekeeping gene beta-actin. Agarose gel electrophoresis of the RT-PCR product revealed amplification of a beta-actin-specific product of 353 bp in all samples, confirming the satisfactory quality of the initial total RNA preparation (Figure 7.2).

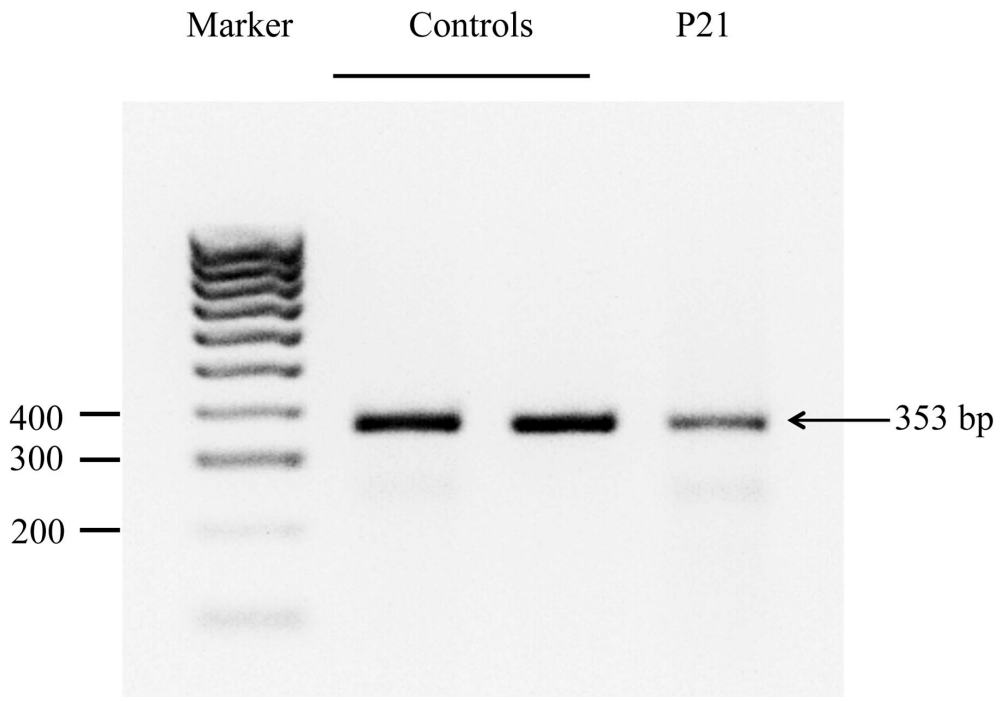


Figure 7.2: Agarose gel electrophoresis of beta-actin RT-PCR product generated from controls' and patient's fibroblasts

Agarose gel electrophoresis of the *FASTKD2* RT-PCR product, generated from RNA derived from fibroblasts of healthy controls and patient P21 using primers specific to *FASTKD2* cDNA, revealed amplification of a strong band of 907 bp and a weak band of approximately 1000 bp in controls. The patient sample produced two strong bands of 1000 bp and 907 bp (Figure 7.3)

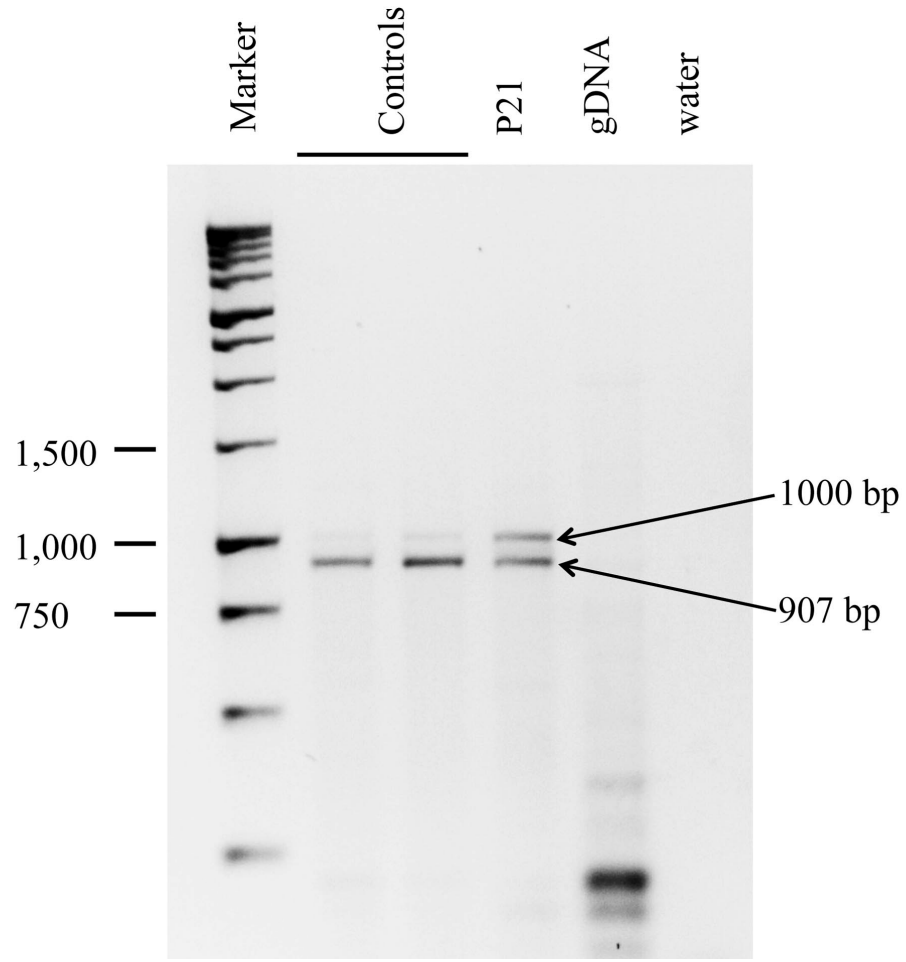


Figure 7.3: Agarose gel (1.5%) electrophoresis of *FASTKD2* PCR product, amplified from cDNA (RNA was derived from fibroblasts of healthy controls and the patient P21)

Sequencing of *FASTKD2* RT-PCR product generated from fibroblasts (total PCR product) was inconclusive in the patient. Sequence chromatograms obtained from the patient showed mixed peaks indicating the presence of several alleles. This study concluded that direct sequencing of total RT-PCR product was inconclusive in patient P21 and therefore the study moved on to cloning the cDNA fragments in the patient and subsequently sequencing the cloned DNA. For controls, sequence chromatograms showed a deletion of 108 bp from the 5' untranslated region (UTR) of exon 1 on both

alleles. It was therefore concluded that the normal splicing event involves a deletion of 108 bp. Interestingly, the c.-66A mutation falls within the 108 bp deletion. Based on the clear Sanger sequencing result, cDNA cloning for controls was not performed. Also, the clear Sanger sequencing result for controls, which reveals the presence of only one allele, suggests that the weak band at approximately 1000 bp in both controls represents a non-specific PCR product (Figure 7.3).

In order to extract the patient's DNA from the gel for subsequent cloning, both bands were excised in a single preparation. This fragment was cloned into pGEM®-T Easy vector, 22 white colonies were selected and sequenced. Sequencing in 4 clones failed, possibly due to the primers' failure to anneal. From the remaining clones, 3 represented the aberrant allele and 15 represented the normal allele (Figure 7.4).

A normal allele (Figure 7.4, A and B) has a deletion of 108 bp on the 5'UTR of exon 1, including the c.-66A mutation; the second c.149A>G:p.Lys50Arg mutation is present on this allele confirming occurrence of the mutations *in trans.* An aberrant allele (Figure 7.4, C and D) has a deletion of 15 bp from the 5'UTR of exon 1, including the c.-66A mutation; the second c.149A>G:p.Lys50Arg mutation is shown to be present on the opposite allele confirming occurrence of the mutations *in trans.*

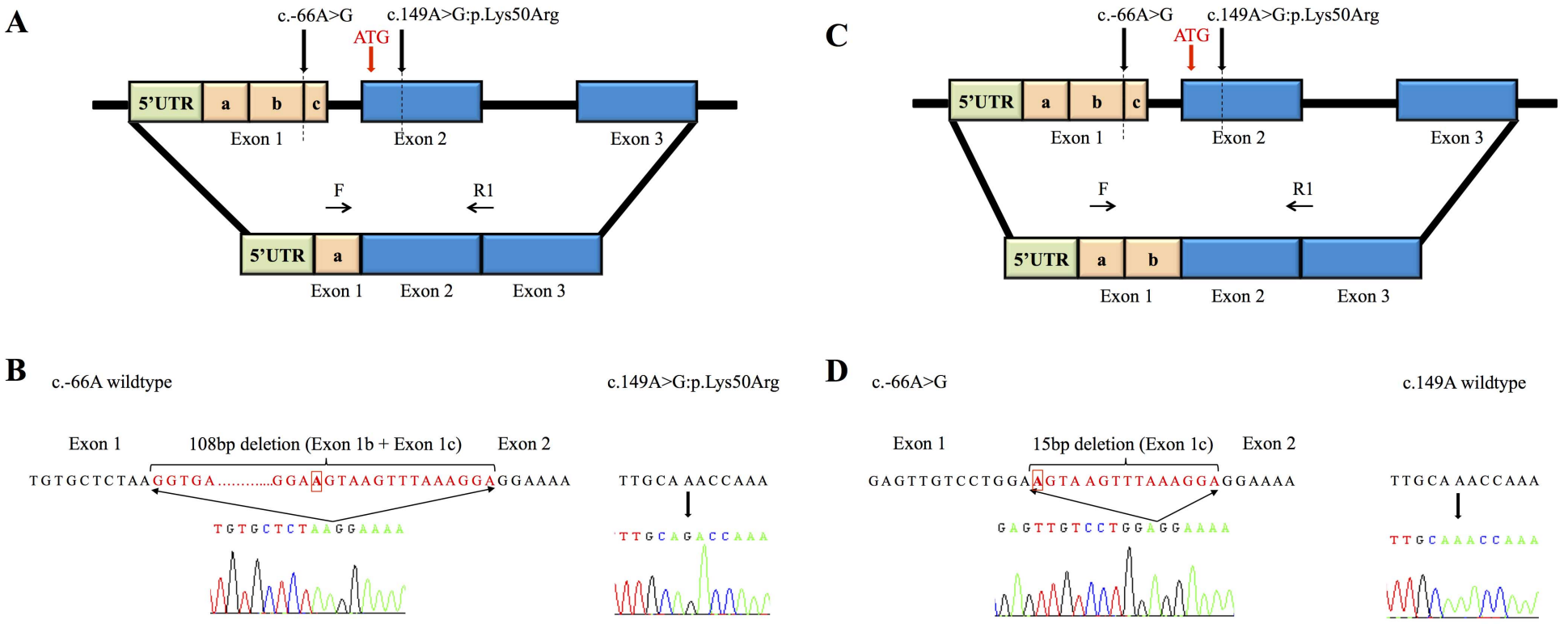


Figure 7.4: Sanger sequencing analysis of *FASTKD2* cDNA clones for amplicon FR1. This figure depicts normal (A, B) and aberrant (C, D) splicing in the patient. The c.-66A mutation is boxed in red. A: schematic representation of a normal splicing event. B: sequence chromatograms of the patient's cDNA clone representing a normal allele (left: deletion of 108 bp from the 5'UTR of exon 1, including the c.-66A mutation, represents a normal splicing event; right: the second c.149A>G:p.Lys50Arg mutation is present on the same allele confirming occurrence of the mutations *in trans*. C: schematic representation of an aberrant splicing event. D: sequence chromatograms of the patient's cDNA clone representing an aberrant allele (left: deletion of 15 bp from the 5'UTR of exon 1, including the c.-66A mutation, represents an aberrant splicing event; right: the second c.149A>G:p.Lys50Arg mutation is shown to be present on the opposite allele confirming occurrence of the mutations *in trans*.

7.3.3. KCNB2 (patient P26)

Allelic cis-trans study via cloning of genomic DNA

Patient P26 (female) presented with cerebellar ataxia in her teens. There was no family history. Whole exome sequencing identified putative pathogenic compound heterozygous *KCNB2* mutations in this patient (p.Ser530Phe and p.Leu784Pro). Both mutations were not found in the public databases, such as NHLBI Exome Sequencing Project, 1000 genomes and dbSNP137, or the in-house control panel of 286 unrelated exomes (Table 3.4). Mutation p.Ser530Phe is conserved across species and p.Leu784Pro is conserved across the primates (Table 3.4). Detailed clinical investigations are presented in Table 4.5. Segregation study employing Sanger sequencing confirmed both mutations in the patient, but revealed that the mother was wildtype for both variants. This suggested that the patient inherited the variants in *cis*. However, with no father's DNA available, it was not possible to affirm this with confidence due to a chance, albeit very little, for these mutations being *de novo* compound heterozygous. Therefore the *KCNB2* amplicon, which was amplified from the patient's genomic DNA, was cloned and sequenced. PCR amplicon was generated from the patient's genomic DNA using *KCNB2* specific primers spanning both mutations. Agarose gel electrophoresis of the PCR amplicon revealed amplification of a 959 bp PCR product in patient P26 (Figure 7.5). This band was excised from the gel and cloned.

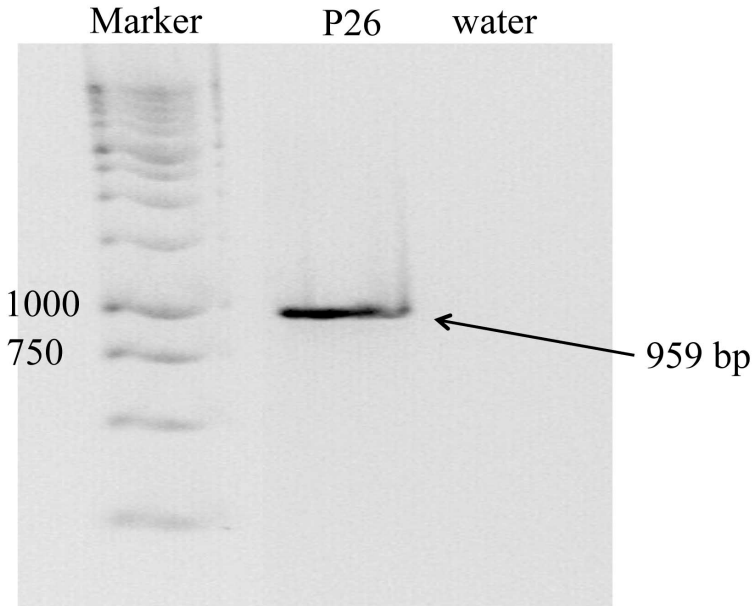


Figure 7.5: Agarose gel (1.5%) electrophoresis of a PCR amplicon generated from the patient's genomic DNA using *KCNB2* specific primers spanning both *KCNB2* mutations c.1589C>T:p.Ser530Phe and c.2351T>C:p.Leu784Pro

24 clones were sequenced. Sanger sequencing of the cloned alleles revealed that there were two type of alleles present (Figure 7.6). Allele A represented a wildtype allele and had both bases c.1589C and c.2351T present in a homozygous wildtype state (Figure 7.6, A). Allele B had both *KCNB2* mutations c.1589C>T:p.Ser530Phe and c.2351T>C:p.Leu784Pro present in a homozygous state. Out of 24 clones, there were 9 clones representing the wildtype allele A and 15 clones representing allele B containing both mutations.

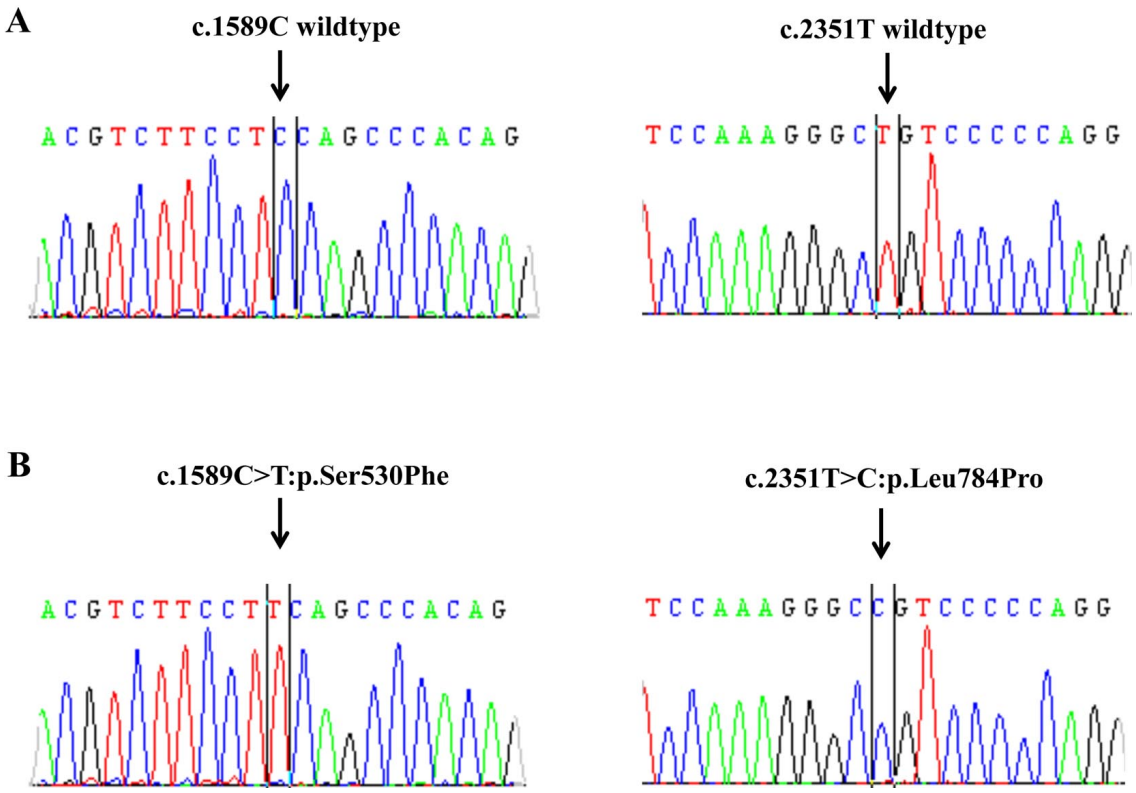


Figure 7.6: Sanger sequencing chromatograms for cloned *KCNB2* alleles type A and B. This figure depicts two types of alleles that were cloned. Type A represents a wild type allele, type B contains both *KCNB2* mutations **c.1589C>T:p.Ser530Phe and c.2351T>C:p.Leu784Pro**. The mutations are inherited in *cis*.

Segregation study

Sanger sequencing of blood DNA of an unaffected brother confirmed that the variants are inherited in *cis* (Figure 7.7)

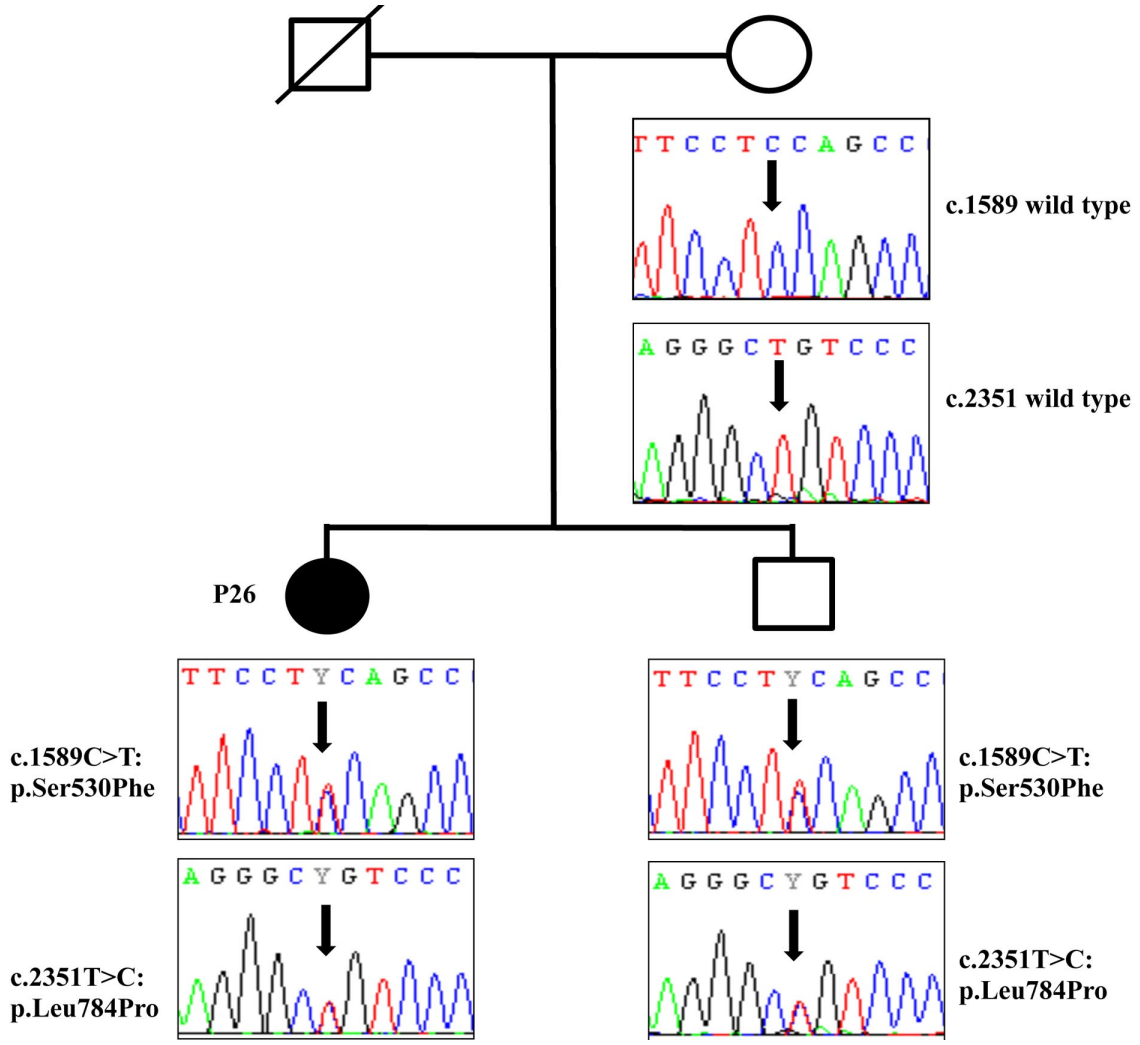


Figure 7.7: Segregation analysis of *KCNB2* mutations in the family of patient P26. This figure shows that the patient P26 and her unaffected brother inherit the *KCNB2* heterozygous mutations in *cis*, probably from the father.

7.4. Discussion

7.4.1. ZFYVE26 (patient 19)

Whole exome sequencing identified *ZFYE26* (SPG15) compound heterozygous mutations (p.Leu817Cysfs*12 and p.Arg780*) in patient P19. Detailed clinical investigations are presented in Table 4.5. Additional clinical data are presented in Appendix 1. The p.Arg780* was predicted to be disease causing or damaging by three pathogenicity prediction programs (Table 3.4). The parental samples were unavailable for analysis. However, this study shows, via sequencing of cloned alleles from the patient, that the variants were inherited in *trans*, supporting a recessive inheritance

model. This led the clinicians to re-assess the neuroimaging and confirm the thin corpus callosum, which is a characteristic of SPG15 (Figure 7.8).

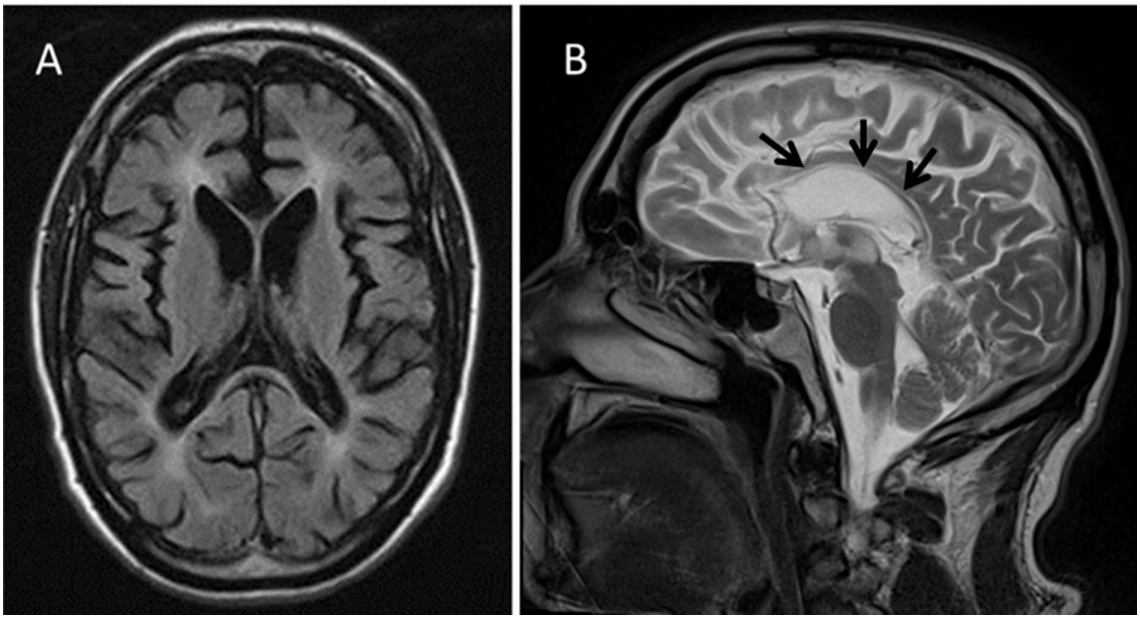


Figure 7.8: T1 Axial (A) and T2 sagittal (B) MRI in P19. The MRI shows global cerebral atrophy with marked atrophy of the corpus callosum (arrows) (adapted from Pyle *et al*., 2014).

SPG15 is a neurological disorder with a recessive mode of inheritance, which is characterised by slowly progressive spastic paraparesis, cognitive impairment, ophthalmologic abnormalities and often ataxia. The main MRI characteristic of SPG15 is thin corpus callosum (Goizet *et al.*, 2009). SPG15 is caused by mutations in *ZFYVE26,* which encodes spastizin (Hanein *et al.*, 2008).

This study underscores the advantages of whole exome sequencing in a clinical setting. The finding of compound heterozygous mutations in *ZFYVE26* prompted re-assessing the neuroimaging of the patient's brain and subsequently establishing the diagnosis in a patient with previously undiagnosed ataxia.

7.4.2. FASTKD2 (patient P21)

Using unbiased whole exome sequencing this study identified putative pathogenic compound heterozygous *FASTKD2* mutations (c.-66A>G and c.149A>G:p.Lys50Arg) in one patient (P21) with ataxia (Table 4.5). Sanger sequencing confirmed segregation of these variants with disease in the family (Appendix 1). The p.Lys50Arg mutation is

rare in the population and conserved among species (Table 3.4). The c.-66A>G mutation, which is situated in the untranslated exon 1 (5'UTR) was predicted to affect normal splicing of exon 1 (Figure 4.1).

Genomic variations causing splicing abnormalities account for approximately 15% of mutations linked to human inheritable diseases (Blencowe, 2000). 5'- and 3' UTRs of the mRNA contain regulatory elements, which mediate protein synthesis (Chatterjee and Pal, 2009). Recently it has been shown that mutations in the 5'- and 3' UTRs are causing human diseases (Scheper *et al.*, 2007). In particular, mutations in 5'UTR were shown to cause such neurological phenotypes as Alzheimer's disease (Mihailovich *et al.*, 2007) and Charcot-Marie-Tooth disease (Hudder and Werner, 2000). This study shows, via sequencing analysis of cDNA clones, that the c.-66A>G mutation, which was predicted to affect splicing, is indeed causing an aberrant splicing event in the patient. This study therefore hypothesises that the aberrant splicing could cause a neurological phenotype in the patient presented here. Previously, a nonsense *FASTKD2* mutation was found in two siblings with infantile mitochondrial encephalopathy (Ghezzi *et al.*, 2008). FASTKD2 is located to the mitochondria and was shown to be involved in regulating apoptosis in breast cancer (Yeung *et al.*, 2011). Recently it was shown that in the astrocytes of patients with Alzheimer's disease *FASTKD2* expression level is decreased. This suggested that FASTKD2 might be involved in the regulation of apoptosis in astrocytes (Sekar *et al.*, 2015). The latter finding is of a particular interest because recently astrocytes emerged as the cells playing a crucial role in synaptogenesis (Allen *et al.*, 2012). This study therefore hypothesises that the *FASTKD2* mutation c-66A>G, which resides in the 5'UTR, could, together with the second mutation p.Lys50Arg, cause dis-regulation of apoptosis in the astrocytes and therefore affect synaptogenesis in the patient. The functional studies of these and other *FASTKD2* mutations are warranted in order to dissect the molecular mechanism of the associated phenotypes. This could potentially lead to therapy and counselling of families with similar disorders.

7.4.3. KCNB2 (patient P26)

Using whole exome sequencing this study identified putative pathogenic compound heterozygous *KCNB2* mutations in patient P26 (p.Ser530Phe and p.Leu784Pro) (Appenxix 1). Both mutations were not found in the public databases, such as NHLBI

Exome Sequencing Project, 1000 genomes and dbSNP137, or the in-house control panel of 286 unrelated exomes (Table 3.4). Mutation p.Ser530Phe is conserved across species and p.Leu784Pro is conserved across the primates. Segregation study employing Sanger sequencing confirmed both mutations in the patient, but revealed that the mother was wildtype for both variants. This suggested that the patient inherited the variants in *cis*. However, with no father's DNA available, this study was unable to affirm this with confidence due to a chance, albeit very little, for these mutations being *de novo* compound heterozygous. Therefore the *KCNB2* amplicon, which was amplified from the patient's genomic DNA, was cloned and sequenced. The latter confirmed inheritance of the variants in *cis*. Sanger sequencing of the blood genomic DNA of an unaffected brother revealed the presence of both variants. This finally confirmed the inheritance of the both variants in *cis* by the patient P26 and her unaffected brother. This study therefore excluded *KCNB2.* The current study was unable to solve this case possibly due to the technical issues such as ineffectiveness of the exome sequencing probes for GC-reach or repetitive issues, which could result in incomplete coverage. The causative variants could also reside within non-coding regulatory regions, large genomic rearrangements and trinucleotide repeat sequences, which are not reliably detected from exome-sequencing data. Therefore, in order to solve this case the patient's DNA should be subjected to whole genome sequencing.

Chapter 8

Concluding Discussion

8. Concluding Discussion

8.1. Overview

Neurological disorders are complex traits, manifesting as a range of diverse phenotypes. Most neurological disorders are highly heterogeneous and individually very rare. This clinical spectrum, coupled with vast genetic heterogeneity, makes genetic diagnosis particularly difficult. Moreover, many neurological diseases exhibit phenotypic overlap, which further complicates establishing the diagnosis. Accurate and timely diagnosis is crucial because for some neurological diseases treatment could be available (Jayadev and Bird, 2013; Toscano and Schoser, 2013). Precise molecular diagnosis could also assist patient management (Dixon-Salazar *et al.*, 2012). Application of next generation sequencing tools promises to speed up the discovery of causative variants and subsequently aid in elucidating the molecular mechanisms of different neurological diseases.

The most comprehensive type of NGS is whole genome sequencing, which was shown to be effective in identifying mutations causing neurological phenotypes (Lupski *et al.*, 2010). However, the drawbacks include the large data output, difficulties in interpreting non-coding variations and ultimately a higher cost (Warman Chardon *et al.*, 2015). The solution could be targeted next generation sequencing, which involves panels of known disease genes. Targeted next generation sequencing is considerably less expensive and more rapid than whole genome sequencing. This approach has proved to be successful in establishing a diagnosis for neurological patients (Calvo and Mootha, 2010; Nemeth *et al.*, 2013; Rehm *et al.*, 2013; Shanks *et al.*, 2013). Although effective, this approach remains a biased technique, which requires a subjective decision as to which exact genes to include in the panel. This could lead to a limited portfolio of genes in the panel. Also, as new disease genes are being discovered and unusual clinical presentations are being described, such panels may quickly become outdated and the correct diagnosis could be overlooked. Unlike the targeted sequencing panels, whole exome sequencing allows discovery of rare mutations in novel disease genes, which were not linked to a disease before. After the proof of principle for using whole exome sequencing for dissecting rare Mendelian disorders (Ng *et al.*, 2009; Ng *et al.*, 2010b), numerous exome sequencing studies have followed. In the early exome sequencing studies,

several groups simultaneously reported an unexpected excess of rare variations, many of which are deleterious, in populations of various ancestry (Nelson *et al.*, 2012; Tennessen *et al.*, 2012; Fu *et al.*, 2013). The authors hypothesised that the excess of rare deleterious variants is due to the recent explosion of the population growth and weak purifying selection. This new paradigm highlighted potential difficulties researchers could confront when interpreting exome sequencing data. An additional obstacle is phenotypic overlap, which is not uncommon among neurological disorders. As outlined above, defining the genetic cause of rare neurological disorders is crucial because it would ultimately define the pathways involved in the disease. This would potentially result in identifying new therapeutic targets. Exome sequencing is gradually establishing itself as an excellent tool to elucidate causative variants in various neurological diseases.

In order to prioritize candidate genes and to reduce their number, step-wise filtering is applied (Ng *et al.*, 2010a). The filtered variants are subjected to verifications by means of Sanger sequencing. The confirmed mutations are additionally verified in multiple affected and unaffected family members via Sanger sequencing (segregation study). If the variant is found to segregate with the disease, then this would point towards pathogenicity of the variant and support candidate gene status. Also, if the variant segregates with the disease in other unrelated cases or pedigrees, then this would strengthen the causality of the gene. *De novo* mutations in genes, previously reported to be involved with a similar phenotype, are regarded as being highly informative about their functionality. *De novo* mutations in previously unreported genes require further investigation (Sunyaev, 2012). Case-control studies should also be performed. It has been documented that many rare variants will be restricted to particular geographical regions (Nelson *et al.*, 2012). Therefore it is particularly important to carefully match case and control samples. In a case-control study, allele frequencies of genetic variants in patients are compared with a group of controls from the same population.

The fundamental question of every exome sequencing study is how do we prove that the particular mutation is a cause of the disease? Literature suggests that evidence of causality of a particular mutation could be finding different mutations in the same gene in other unrelated individuals with a similar phenotype (Ng *et al.*, 2010a; Ng *et al.*, 2010b). For rare phenotypes, which segregate in a family, the accepted proof of pathogenicity of a mutation in a given gene is establishing involvement of this gene

with the same phenotype in three unrelated families (Brunham and Hayden, 2013). Mutations in the another gene from the same biological pathway, found in unrelated individuals with the same phenotype, would also suggest involvement of that gene in this biological event (McClellan and King, 2010). The ultimate goal of a genetic study is to understand the mechanisms behind the disease onset and progression. That is why it is important to know the functional significance of identified novel rare variants. Functional studies depend on the type of the putative causative variant and could include, but will not be exhausted by, transcriptional and translational studies, as well as studies in live cells and animal models (MacArthur *et al.*, 2014).

After the first exome sequencing studies were carried out, researchers began to verify this technique for a clinical setting. Yang and colleagues sequenced exomes of 250 unselected patients with suspected Mendelian disorders, 80% of which were children with neurodevelopmental disorders (Yang *et al.*, 2013). The authors reported the overall diagnostic yield of 25%. Whole exome sequencing of parent-child trios was shown to be very useful in dissecting neurological disorders, such as intellectual disability and developmental disorders. For example, de Ligt and co-workers performed whole exome sequencing on 100 patients with unexplained severe intellectual disability and their unaffected parents (de Ligt *et al.*, 2012). Having evaluated *de novo*, autosomal recessive and X-linked mutations, the authors established the molecular cause in 16% of the patients tested. Interestingly, the majority of identified putative causative mutations were *de novo.* In another study, exomes of 51 patients with severe sporadic intellectual disability and their unaffected parents were sequenced (Rauch *et al.*, 2012). The authors found that *de novo* variants (both point mutations and small indels) were present in 50% of patients in the cohort.

The research presented in this study aimed to evaluate the use of whole exome sequencing for diagnostic purposes in neurological disorders. Whole exome sequencing was performed in a heterogeneous cohort of patients with suspected inherited ataxia as an example of a neurological disorder, with the aim to identify candidate gene mutations. 35 affected individuals from 22 randomly selected families of white European descent with no known consanguinity were studied. Prior to inclusion in this study, all participating individuals had undertaken routine clinical investigations in order to exclude common sporadic, inherited and metabolic causes of ataxia. 30 affected individuals were subjected to whole exome sequencing. The interpretation of the raw

whole exome sequencing data utilised an in-house bioinformatic pipeline. A subsequent variant filtering algorithm excluded synonymous variants and focused on protein altering variants, nonsense mutations, exonic insertions/deletions and splice site variants. All variants with a minor allele frequency (MAF) greater than 1%, in the public databases (dbSNP137, 1000 genomes (April 2012 data release) and NHLBI-ESP6500 databases) and in 286 unrelated in-house controls, were excluded. The remaining variants were selected based on mode of inheritance. Priority was given to variants in genes that were expressed in brain or nerve cells, and variants in genes previously linked to ataxia or related neurological disorders. Subsequently, an extensive literature search was performed in order to identify current genes associated with ataxia and related neurological disorders. The published variants were cross-referenced with the filtered variants. All variants were defined using carefully selected criteria. Further molecular genetic techniques (Sanger sequencing, reverse transcription PCR, quantitative pyrosequencing, cloning for allelic *cis-trans* study) and proteomics (Western blotting, immunohistochemistry) were employed in order to support the exome sequencing finding.

8.2. Main findings

In this study, whole exome sequencing was performed on 35 individuals with suspected inherited ataxia, which were selected randomly from 22 families with no known consanguinity. This work made a new and certain molecular diagnosis in several families. Confirmed pathogenic variants were found in 9/22 probands (41%) implicating 6 genes. Three families, each presenting with a pair of affected siblings, had novel compound *SACS* mutations (Pyle *et al.*, 2012; Pyle *et al.*, 2013). Interestingly, one of these families exhibited an adult onset disorder considered to present in childhood. A known dominant *KCNC3* mutation was found to segregate with ataxia in four members of a three-generation autosomal dominant pedigree (Waters *et al.*, 2006). Previously reported compound heterozygous *SPG7* mutations were found in two affected siblings from a family with no spasticity (Casari *et al.*, 1998). *SPG7* is a gene implicated in hereditary spastic paraplegia. Although *SPG7* has previously been linked to ataxia, it has yet to be described in cases without spasticity. Therefore this study allowed to broaden the phenotype implicated with *SPG7* mutations. Likely *de novo* dominant *TUBB4A* mutations were found in two families (Simons *et al.*, 2013). Both

mutations are evolutionary conserved. In one family, quantitative pyrosequencing showed varying degrees of mosaicism in the mildly affected mother and heterozygosity in the severely affected offspring. *In silico* analysis further supported pathogenicity of the mutation and revealed that it could potentially disrupt polymerizations of αβ-tubulin heterodimers. Two siblings who presented with adult onset ataxia had compound heterozygous mutations in *NPC1,* confirmed by subsequent oxysterol analysis (Carstea *et al.*, 1997). These two patients were lacking the characteristic clinical features linked to Niemann Pick type C, which is diagnosed in childhood. This finding allowed us to consider a known disease gene in an unusual clinical presentation, hence broadening the phenotype. A novel dominant mutation in a known ataxia gene *SLC1A3* segregated with ataxia in three members of one family (de Vries *et al.*, 2009).

Possible pathogenic variants were identified in 5/22 probands (23%) implicating 5 genes. This study provides information on inheritance patterns and potential prognosis. Likely *de novo* compound heterozygous mutations in *ZFYVE26* (SPG15) in one family was validated with sequencing of cloned alleles and the result confirmed the occurrence of the mutations in *trans*, therefore supporting their autosomal recessive inheritance (Hanein *et al.*, 2008). This finding led the clinicians to re-assess the neuroimaging and confirmed thinning of the corpus callosum, which is a characteristic of SPG15. Previously described compound heterozygous *WFS1* mutations were identified in one family (Cryns *et al.*, 2002; Cryns *et al.*, 2003). In one family with possible pathogenic compound heterozygous *FASTKD2* mutations, the *in silico* splice-site prediction was validated by sequencing analysis of cDNA clones. A predicted splice site mutation was detected in three members of an autosomal dominant pedigree in the previously described spastic paraplegia gene *ZFYVE27* (SPG33) (Mannan *et al.*, 2006). This study showed that the ZFYVE27 protein (protrudin) levels were increased approximately 2.5 fold in the cerebellum but not in the frontal cortex of the affected individual. Protrudin is an endoplasmic reticulum (ER) protein and its anomalies have previously been shown to cause ER stress. In this study levels of the master regulator of ER stress, BiP/GRP78, were significantly increased in the patient's cerebellum, which may indicate the ER pathology. The preliminary data presented in this study also suggested that protrudin levels were increased in the astrocytes in the cerebellum but not in the frontal cortex of the affected individual. Finally, novel compound heterozygous mutations, which segregated in the family, were found in *WNK1* in one patient (Lafreniere *et al.*, 2004).

WNK1 is a known disease gene, which was previously described in families with hereditary sensory and autonomic neuropathy type II (HSAN2), but has yet to be associated specifically with ataxia.

Taken together, this study defined the likely molecular diagnosis in 14 of 22 families (64%) with 11 genes implicated. With no recurrent mutation in the same gene observed, this study reiterates the known genetic heterogeneity of inherited ataxias. Importantly, the diagnostic yield in this study is similar to studies in other neurological disorders (Klein *et al.*, 2014; Novarino *et al.*, 2014; Gonzaga-Jauregui *et al.*, 2015) and ataxia (Fogel *et al.*, 2014; Sawyer *et al.*, 2014). Together, these studies highlight the potential clinical significance of whole exome sequencing.

The diagnostic yield in this study contrasted with a targeted next generation sequencing approach of a similar sized UK cohort of inherited ataxia patients (18% of patients) (Nemeth *et al*., 2013). This could be due to greater exome coverage in this study. The unbiased approach of the current study also means that genes, which are linked to ataxia, but are not considered as "ataxia genes", were not excluded from this study. In fact, this study identified confirmed pathogenic mutations in *SPG7,* which is known to cause spastic paraplegia type 7 (SPG7), in one family with ataxia but no spasticity. This surprising finding prompted the prospective exome sequencing study of 70 patients with unexplained ataxia by Pfeffer and the colleagues (Pfeffer *et al.*, 2015). The latter study identified mutations in *SPG7* as a likely cause of disease in 19% of patients with unexplained ataxia. This suggested that the occurrence of *SPG7*-related diseases could be comparable with that of dominant spinocerebellar ataxia (SCA) and Friedrich ataxia. The implication of this finding is significant, because it could lead to a change in a diagnostic workflow making *SPG7* a top candidate for initial genetic tests. This could eventually decrease the cost of a diagnostic test.

The current study reveals the importance of whole exome sequencing in identifying pathogenic variants in known disease genes when the clinical presentation is less specific, broadening the phenotype. This study identified compound heterozygous *WFS1* mutations in one family (patient P20). The patient didn't present with early diabetes mellitus, which is a characteristic of classic Wolfram syndrome or other *WFS1*-related disorders (Fogel *et al*., 2014). Based on the phenotype, *WSF1* would not have been considered for routine clinical investigations in this patient. Similarly, this study

identified confirmed pathogenic mutations in *SPG7,* which is known to cause spastic paraplegia type 7 (SPG7), in one family with ataxia but no spasticity (patients P9 and P10). The absence of spasticity is uncharacteristic of SPG7, therefore *SPG7* would not have been prioritised for routine gene test.

The current study identified likely *de novo* dominant *TUBB4A* mutations in two families presented with cerebellar ataxia and leukodystrophy. In one family quantitative pyrosequencing revealed varying degrees of mosaicism in the mildly affected mother and heterozygosity in the severely affected offspring. *In silico* analysis further supported pathogenicity of the mutation and revealed that it could potentially disrupt polymerizations of αβ-tubulin heterodimers. Mutations in *TUBB4A* were recently associated with several neuroimaging phenotypes, namely dystonia type 4 (DYT4) (Hersheson *et al.*, 2012; Lohmann *et al.*, 2012), hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC) (Simons *et al.*, 2013), isolated hypomyelination (Pizzino *et al.*, 2014) and hereditary spastic paraplegia (HSP) with no basal ganglia involvement or cognitive impairment (Kancheva *et al.*, 2015). This makes *TUBB4A* a primary candidate for genetic screening, following MRI, meaning that the overall cost of a diagnosis could be greatly reduced.

This study has demonstrated the impact of exome sequencing in a group of patients difficult to diagnose genetically. This study indicates that exome sequencing is an efficient tool, which can help with rapid, cost effective identification of genetic mutations causing neurological diseases. The approach presented in this study could be applied to other neurological disorders, which are also challenging to solve.

8.3. Study limitations and future directions

As demonstrated by recent studies, exome sequencing is an excellent tool for discovering rare variants as a cause of classical rare Mendelian (Ng *et al.*, 2010a) and more common sporadic (O'Roak *et al.*, 2011) neurological disorders. In addition to studying individual pedigrees, exome sequencing, when applied to large cohorts of cases and controls, could reveal if any single gene is responsible for the phenotype (Bras and Singleton, 2011).

However, despite the clear potential of whole exome sequencing in solving rare inherited neurological disorders, there are several drawbacks. Firstly, an exome sequencing study could uncover unexpected findings, such as rare mutations predisposing to some types of cancer or mutations causing familial neurodegenerative disorders. This ethical issue should be considered when obtaining informed consent for an exome sequencing study (Bamshad *et al.*, 2011). Secondly, exome sequencing studies of different populations may elucidate different causative alleles and disease genes. This is because different populations might exhibit local genetic variations, which could be due, for example, to a founder effect. Also, different populations are pre-disposed to different types of neurological disorders (including ataxia), which would affect the outcome of the exome sequencing study. The cohort presented in this study consisted of individuals from an outbred population of white European ancestry. It is anticipated, that similar exome sequencing studies of cohorts from other geographic regions may result in an outcome, which may differ from the one described here. Thirdly, an exome sequencing study potentially faces some technical issues, for example, incomplete coverage of exons. Typical coverage of exons in exome sequencing is about 90-95% (Bamshad *et al*., 2011). Poor coverage could be due to the capture probes being ineffective for GC-rich or repetitive regions (Sulonen *et al.*, 2011). Also, because the probes for exome capture are designed based on known sequences found in the public databases, all yet-to-be-found exons cannot be captured. That is why exome sequencing cannot be considered exhaustive (Ku *et al.*, 2011). Another technical issue for an exome sequencing study is detecting copy number variants (CNVs), which were recently associated with common and rare diseases (Weischenfeldt *et al.*, 2013). Owing to the fact that exome sequencing employs PCR-based sample preparation techniques, it will be difficult to detect copy number variation in a sample where all templates are amplified to a similar concentration. The exome sequencing study described here did not consider non-coding regions or synonymous variants. It is still not clear to what extent these types of variants can be classed as disease causing (Cooper *et al.*, 2010). It could be, however, foreseen that some of the causal variants may reside within these regions. It is anticipated that the technical issues outlined above prevented this study from confidently identifying likely causative variants in 8/22 probands (36%). It is assumed that some of these technical issues could be addressed by whole genome sequencing, albeit with additional cost. In fact, 8/22 probands where diagnosis has not been established, have been sent for whole genome sequencing.

Another limitation of this research needs to be acknowledged. One of the aims of the current study was to elucidate pathogenicity of the rare *ZFYVE27* c.805-2A>G mutation (allele count 3/121378; MAF=0.00002472 in ExAC database) and establish its causative role in a dominant family presenting with cerebellar ataxia. This study resulted in several interesting findings. The ZFYVE27 protein (protrudin) levels were increased in the cerebellum but not in the frontal cortex of the affected individual, when compared with controls. This study also founds that levels of the master regulator of ER stress, BiP/GRP78, were significantly increased in the patient's cerebellum, which may indicate the ER pathology. It would be beneficial to use another ER stress marker to confirm the result with BiP/GRP78. It has to be highlighted that this study analysed only two regions of the brain (the frontal cortex and the cerebellum). It would be valuable to analyze other brain regions from the patient and control individuals in order to gain more insight into the expression profiles of ZFYVE27 under normal and pathological conditions. Despite the interesting findings mentioned above, with no other indisputable *ZFYVE27* pathogenic mutations identified to this date, it is difficult to draw conclusions as to the pathogenicity of the c.805-2A>G mutation. Therefore, at the time of writing this Thesis, these data could not be used in the clinical setting. In order to elucidate the cause of disease in this family, it would be beneficial to perform whole genome sequencing with haplotype reconstruction.

It is predicted that the cost of NGS will continue to fall (Boycott *et al.*, 2013), which will increase the speed of the discovery of rare variations. Many of the mutations will be extremely rare, if not private, therefore ascribing the pathogenicity might become problematic for simplex cases. Studies like this not only verify whole exome sequencing as an excellent tool in delineating the cause of rare neurological disorders. This study, together with similar work by other groups, which report unusual clinical presentations and rare variations with uncertain significance, will aid in creating public databases. The latter will reinforce collaboration and data sharing, which in turn will aid in defining pathways implicated in neurological disorders. Therefore, there is no doubt that whole exome sequencing will aid in establishing a clear diagnosis and finding new therapies in a complex field of neurological disorders.

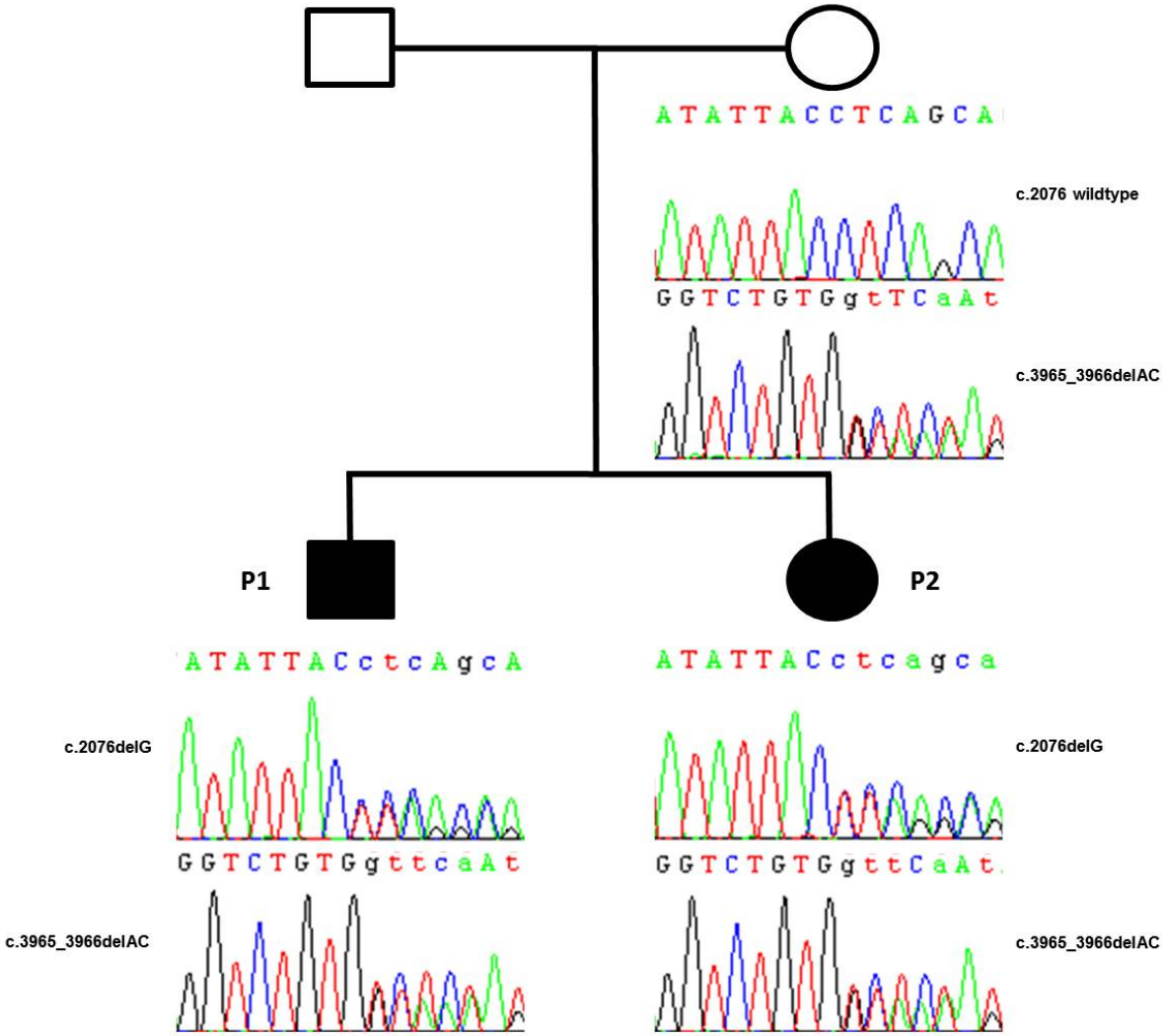
Appendices

Appendix 1. Clinical data and segregation analysis

The data presented here are included in Pyle *et al.* 2014.

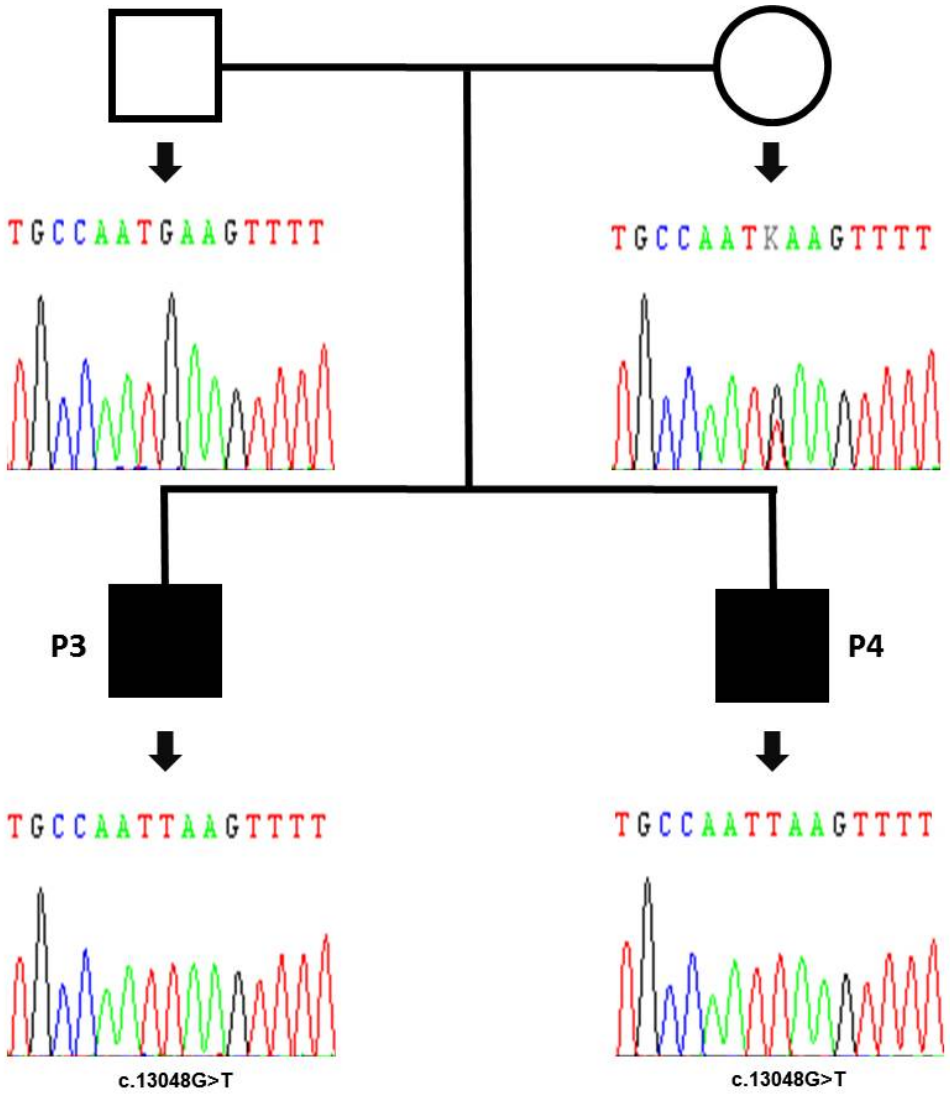
Confirmed pathogenic mutations

Patient 1 and 2: (*SACS*, c.2076delG:p.Thr692Thr fs*713 and c.3965_3966delAC:p. Gly1322Valfs*1343). Patient 1 (male) presented with learning difficulties and pyramidal signs in his teens. He now has a gait disturbance and is wheelchair bound. MRI revealed cerebellar atrophy. Peripheral electrophysiology showed a large-fibre sensorimotor axonal-demyelinating neuropathy. Muscle biopsy showed normal mitochondrial histochemistry and respiratory complex activity with no evidence of mitochondrial DNA deletions. His affected sister (P2) presented at age 26 with poor coordination and urinary urgency. She also developed a gait disturbance and is wheelchair bound. The patient developed a jerky ocular pursuit, and had reduced visual acuity secondary to bilateral epiretinal membrane formation within the macula. She progressed to develop dysarthria and difficulty swallowing. MRI revealed generalised atrophy, and marked cerebellar atrophy. Peripheral electrophysiology showed mixed demyelinating axonal neuropathy. Type I fibre clustering was found in her muscle biopsy, however respiratory complex activity was normal and there was no evidence of mitochondrial DNA deletions. Exome sequencing was performed on both these patients. DNA was available from the mother and Sanger sequencing confirmed that the mutations segregated with the disease phenotype in the family.


ATATTACCTCAGCA
c.2076 wildtype
GGTCTGTGgtTCaAt
c.3965_3966delAC
P1
P2
ATATTACctcAgcA
ATATTACctcagca
c.2076delG
c.2076delG
GGTCTGTGgttcaAt
GGTCTGTGgttCaAt
c.3965_3966delAC
c.3965_3966delAC

Patient 3 and 4: (*SACS,* hemizygous c.13048G>T: p.Glu4350*; 0.7Mb deletion on Chr13q12.12). Patient 3 (male) presented with clumsiness and poor sporting performance in his teens. He went on to develop a spastic ataxic gait, bilateral optic atrophy, jerky ocular pursuits, hypometric saccades and gaze-evoked nystagmus. His speech became slurred with a brisk jaw jerk. His affected brother walked late at 18 months and was diagnosed with a mild cerebral palsy. He progressed with a spastic ataxic gait, bilateral optic atrophy, jerky ocular pursuits and gaze evoked nystagmus in all directions. Patient 4 also developed cerebellar dysarthria and brisk jaw jerk. MRI revealed marked cerebellar vermian atrophy, moderate diffuse cortical atrophy and reduced spinal cord volume. Neurogenic features were present on muscle biopsy examination including angulated fibres, small group atrophy, type II fibre predominance and fibre type grouping. Exome sequencing was performed on patient P3.

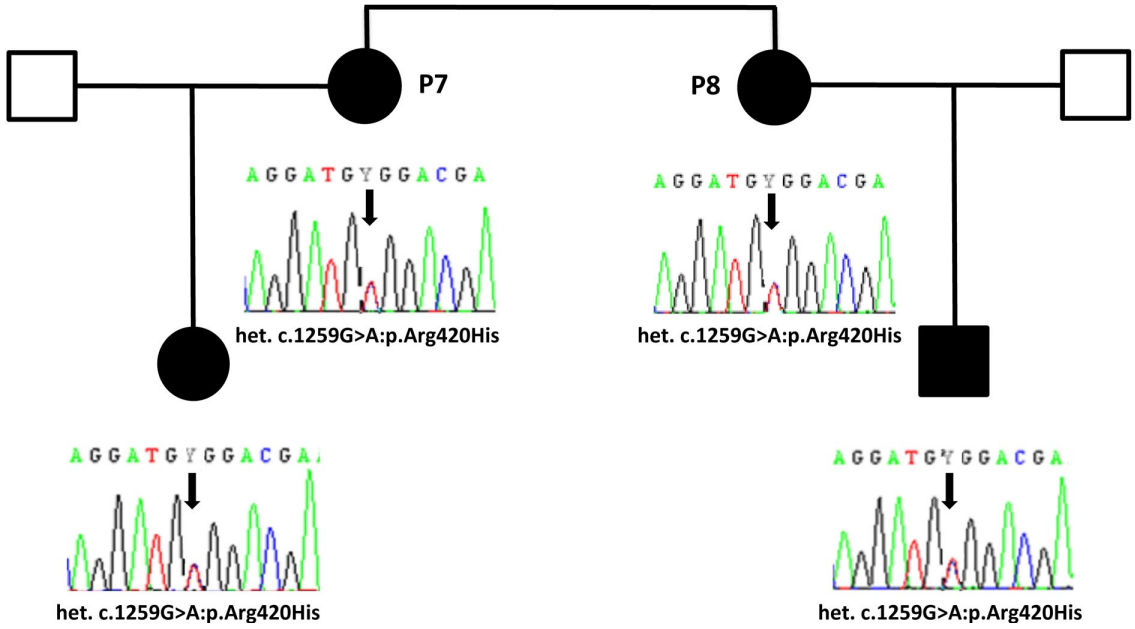
Sanger sequencing of the mutation in all family members confirmed segregation with the disease phenotype.


TGCCAATGAAGTTTT
TGCCAATKAAGTTTT
P3
P4
TGCCAATTAAGTTTT
TGCCAATTAAGTTTT
c.13048G>T
c.13048G>T

Patient 5 and 6: (*SACS*, c.1580C>G:p.Ser527* c.6781C>A:p.Leu2261Ile). Patient 5 (male) presented with difficulty walking and increasing leg stiffness at age 40. He went on to develop spastic paraplegia and is now wheelchair bound. MRI revealed generalised atrophy with foci of subcortical periventricular white matter signal change, consistent with early small cerebrovascular disease. Electrophysiology tests highlighted demyelinating sensory-motor neuropathy. His affected sister (Patient 6) presented during her 40s with difficulty walking and balance problems. She went on to develop an ataxic gait with a positive Romberg test. Her speech developed a subtle slurring and eye examination revealed jerky ocular pursuits with a restricted upward gaze (20 degrees). Mixed sensory-motor neuropathy was identified on peripheral electrophysiology testing. This patient died at age 62. Exome sequencing was performed on patient P5. Segregation analysis in the daughter of patient 5 confirmed that the *SACS* mutations were on complementary chromosomes.

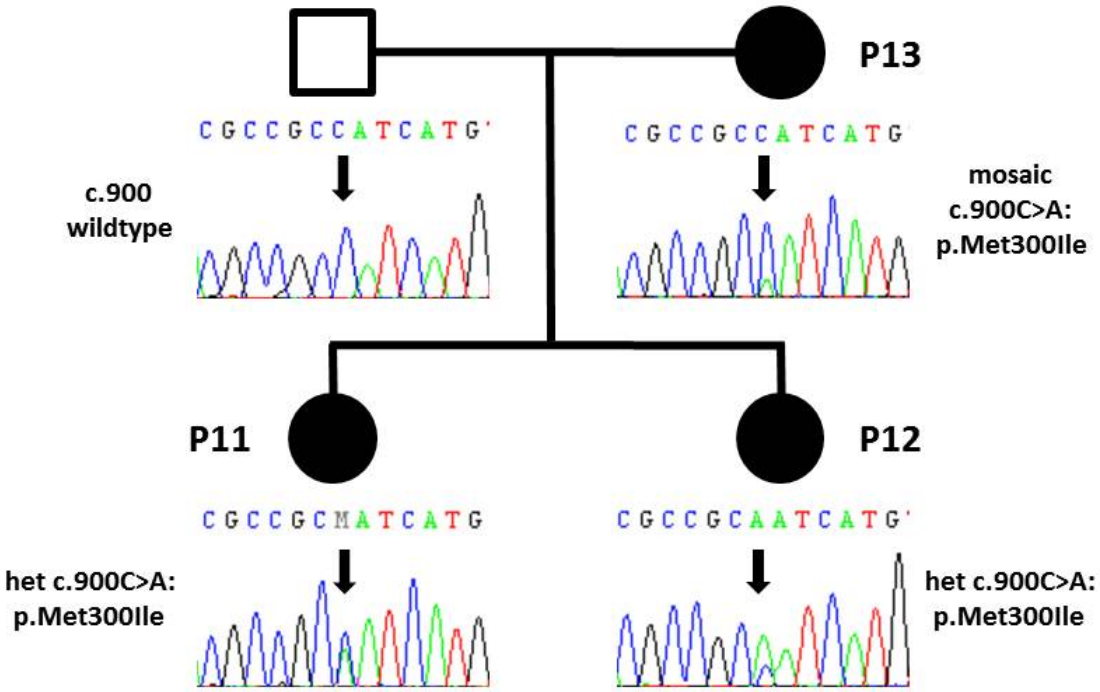
Patient 7 and 8: (*KCNC3*, het. c.1259G>A:p.Arg420His). Patient 7 presented at age 23 with speech and balance problems. She went on to develop an appendicular and ataxic gait with a positive Romberg test. Eye examination revealed bilateral disc pallor, an afferent pupillary defect, square wave jerks with a jerky pursuit. Additional symptoms included dysarthria and hearing impairment. MRI results showed cerebellar atrophy most marked in vermis and superior folia. Myopathic changes and a single cytochrome c oxidase (COX) negative fibre were found on muscle biopsy examination. Her affected sister (Patient 8) presented at age 57 with upper limb clumsiness followed by gait ataxia, jerky ocular pursuits, optic atrophy and dysarthria. MRI examination showed cerebellar atrophy and incidental degenerative disc disease in cervical spine. Exome sequencing was performed on patient P7.

The mutation segregated with the phenotype in the next generation.



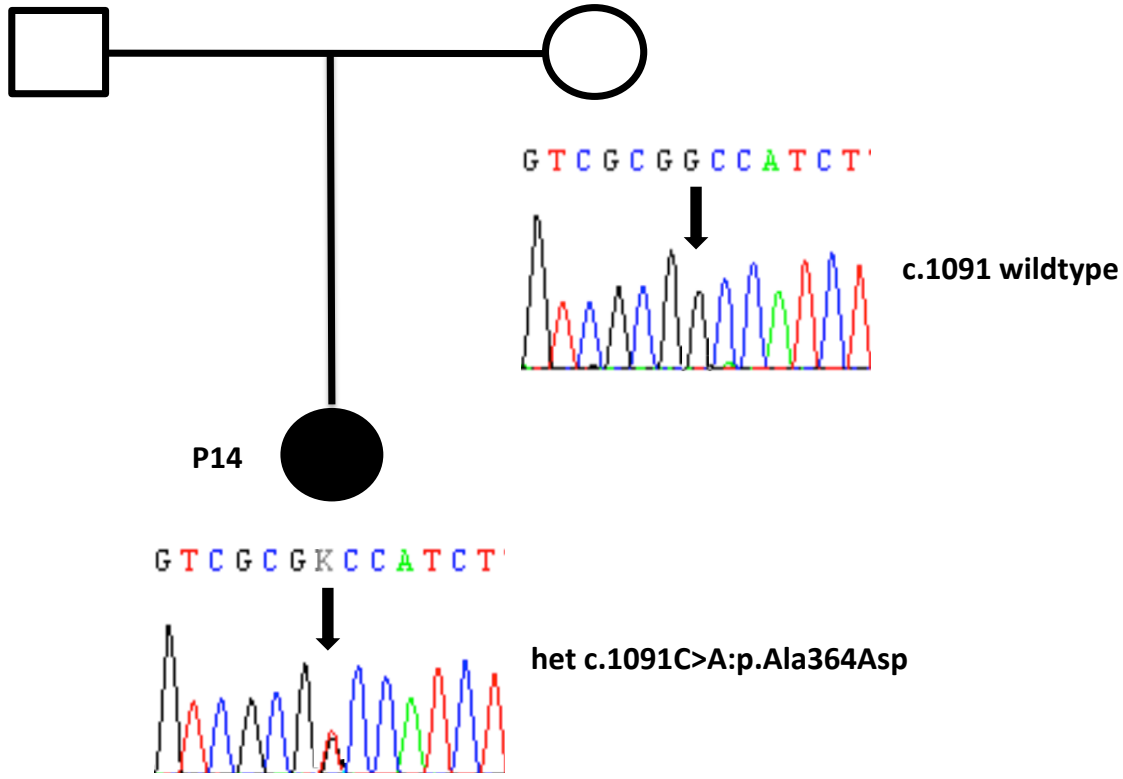
Patient 9 and 10: (*SPG7*, c.1529C>T:p.Ala510Val; c.1715C>T:p.Ala572Val). Patient 9 (male) presented at age 30 with clumsiness. He went on to develop an ataxic gait, dystarthria, a convergent squint, bilateral optic atrophy, jerky ocular pursuit movements which were brisk but with sustained horizontal nystagmus. MRI revealed cerebral atrophy and his muscle biopsy showed a single COX deficient fibre. Routine haematological and biochemical blood tests were normal apart from a mild elevation of aminotransferase (ALT) 50 IU/L. Patient 10 is his affected sister who presented at age 29 with clumsiness and a tendency to fall. She has a short stature. She subsequently developed gait and appendicular ataxia, nystagmus, night blindness, slurred speech, wasting and fasciculations of the right side of tongue. MRI revealed cerebellar atrophy affecting vermis and hemispheres. Muscle biopsy results showed mild or moderate fibre size variation with 1% COX negative fibres. Routine haematological and biochemical blood tests were normal apart from a mild elevation of the ALT 44 and alkaline phosphatase (ALP) 124 IU/L. Exome sequencing was performed on both these patients. Both siblings had the same genotype on Sanger sequencing. The mother was hemizygous, confirming that Patient 9 and 10 were compound heterozygotes.

Patient 11 and 12: (*TUBB4A*, het c.900C>A:p.Met300Ile). **Patient 13**: (*TUBB4A*, mosaic c.900C>A:p.Met300Ile). Patient 11 (female) presented in infancy with delayed motor and speech developments. She developed dysarthria, slow speech, titubation and is now wheelchair bound. Eye examination revealed nystagmus on lateral gaze, slow saccades in the horizontal plane and bilateral optic disc pallor. MRI results showed hypomyelination, leukodystrophy and cerebellar atrophy. Her affected sister (Patient 12) also presented in infancy with delayed motor and speech developments. She went on to develop, dysarthria, truncal ataxia and is wheelchair bound. Slow saccades were found on eye examination. MRI revealed hypomyelination, leukodystrophy and cerebellar atrophy. The mother of these affected sibs (Patient 13) presented at age 30 with balance problems and slurred speech. She went on to develop dystonic posturing, bilateral optic atrophy, dysarthria, echolalia and some swallowing difficulty. MRI revealed hypomyelination, leukodystrophy and cerebellar atrophy (Figure 6.2). Muscle biopsy results showed an age-related multiple mitochondrial DNA deletions. Exome sequencing was performed on all three patients. Sanger sequencing of the candidate variant confirmed segregation with the disease phenotype in the family, and the proportion of the mutated allele was determined by pyrosequencing (Table 6.3).

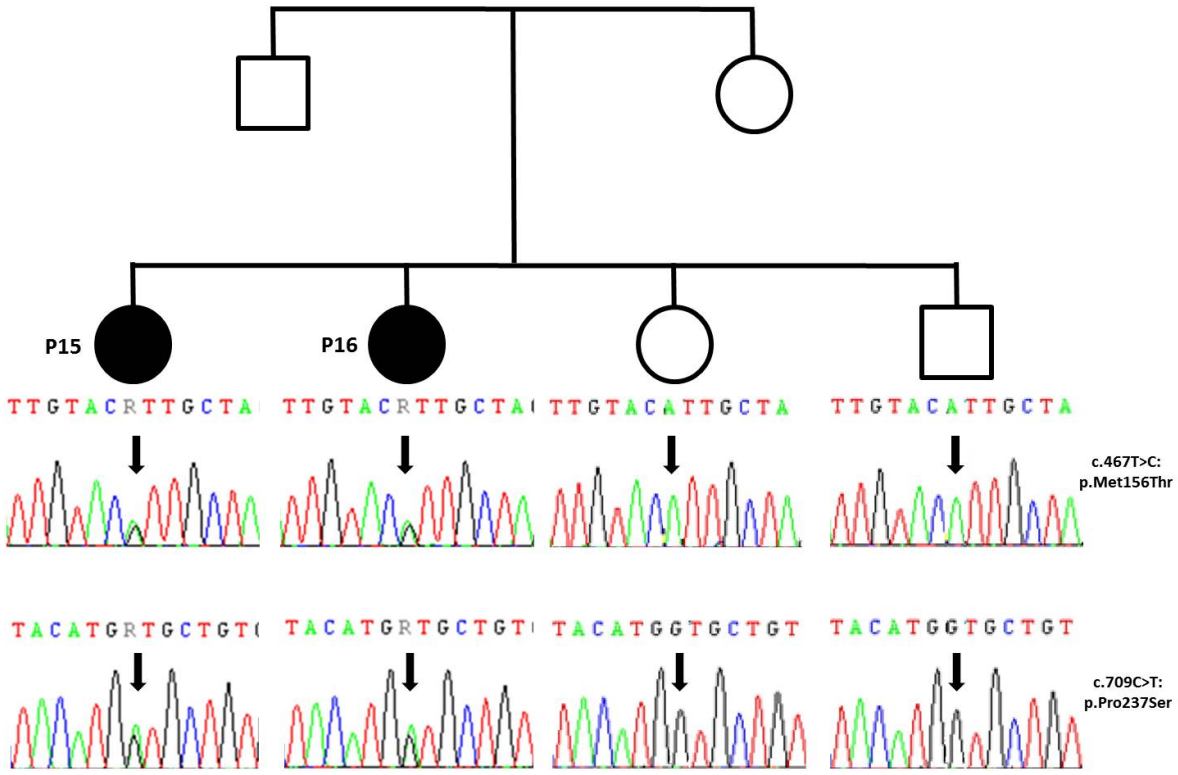


Patient 14: (*TUBB4A*, het c.1091C>A:p.Ala364Asp). This female patient present at age 5 with learning difficulties, ataxia and falls. There was no relevant family history. Developing symptoms included spastic, scissoring gait with wheelchair dependence and use of a walking frame. Eye examination revealed horizontal ocular dyspraxia which was poor in vertical directions, with blinking to relocate saccades. Additional symptoms were a quiet, spastic dysarthria, slow tongue movements, dribbling tendency, brisk jaw jerks and positive glabellar tap. MRI revealed diffuse hypomyelination including brainstem and diffuse white matter changes with T2 hyperintensity. This patient was subjected to exome sequencing.

DNA was only available from the patient and mother, who was consequently wildtype for the variant.

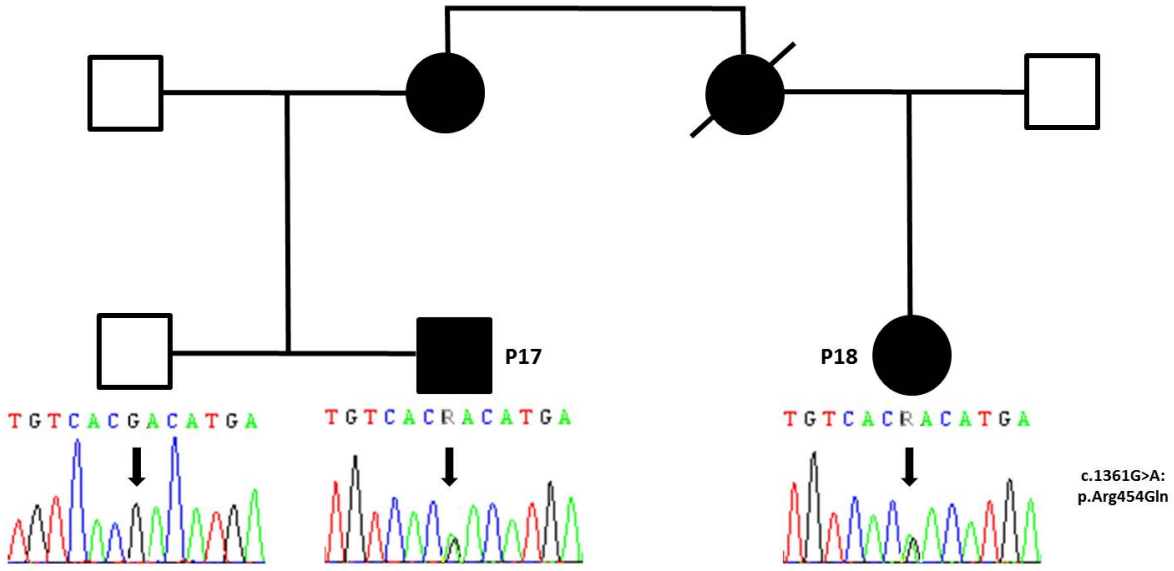


Patient 15 and 16: (*NPC1*, c.467T>C: p.Met156Thr c.709C>T:p.Pro237Ser). These two affected sisters presented at age 21. Patient 15 presented with unsteady walking which resulted in frequent falls. She went on to develop a spastic gait, hypermetric saccades and subtle cerebellar dysarthria. MRI revealed bilateral symmetrical abnormal high T2 signal involving corticospinal tracts. Patient 16 presented with clumsy hands, pins and needles and slurred speech. She went on to develop ataxia, a high steppage gait, and was Romberg test positive. Additional symptoms included jerky ocular pursuits, slow saccades, myoclonic jerks and cerebellar dysarthria. Dysphasia and hearing impairment developed late in the progression of the disease. Exome sequencing was performed on both these patients. Sanger sequencing of the variant confirmed segregation of the mutation with the disease phenotype in the family. Biochemical analysis of plasma oxysterols confirmed the functional consequences of the *NPC1* mutation (Table 4.5).



Patient 17 and 18: (*SLC1A3*, het c.1361G>A:p.Arg454Gln). Patient 17 (male) presented at age 30 with a speech disturbance, dexterity problems and a deteriorating gait. He went on to develop ataxia, jerky ocular pursuits and hypometric saccades. Additional symptoms included cerebellar dysarthria, dysphagia and slow repeating tongue movements. His mother is also affected. Patient 18 is a female cousin who presented at age 39 with a speech disturbance. She went on to develop cerebellar ataxia, jerky ocular pursuits, dysarthria, urinary urgency and frequency. Exome sequencing was performed on patient P17 and his unaffected brother.

Sanger sequencing of the mutation confirmed segregation with the disease in the family.



Possibly pathogenic mutations

Patient 19: (*ZFYVE26,* c.2338C>T:p.Arg780*; c.2450delT:p.Leu817Cysfs*12). This male patient developed muscle weakness and frequent falls aged 32. He was wheelchair bound. There was no family history. The patient developed bilateral facial weakness with dystonia of the facial muscles. MRI investigations revealed generalised atrophy of both the white and grey matter with hazy diffuse gliotic white matter change in both corona radiata. The corpus callosum was severely atrophic (Figure 7.8) showing the characteristic thin corpus callosum frequently associated with mutations in *ZFYVE26*/SPG15 (Goizet *et al.*, 2009). There was also some secondary midbrain involution. Electrophysiology showed large fibre axonal sensory motor neuropathy. This patient was subjected to exome sequencing.

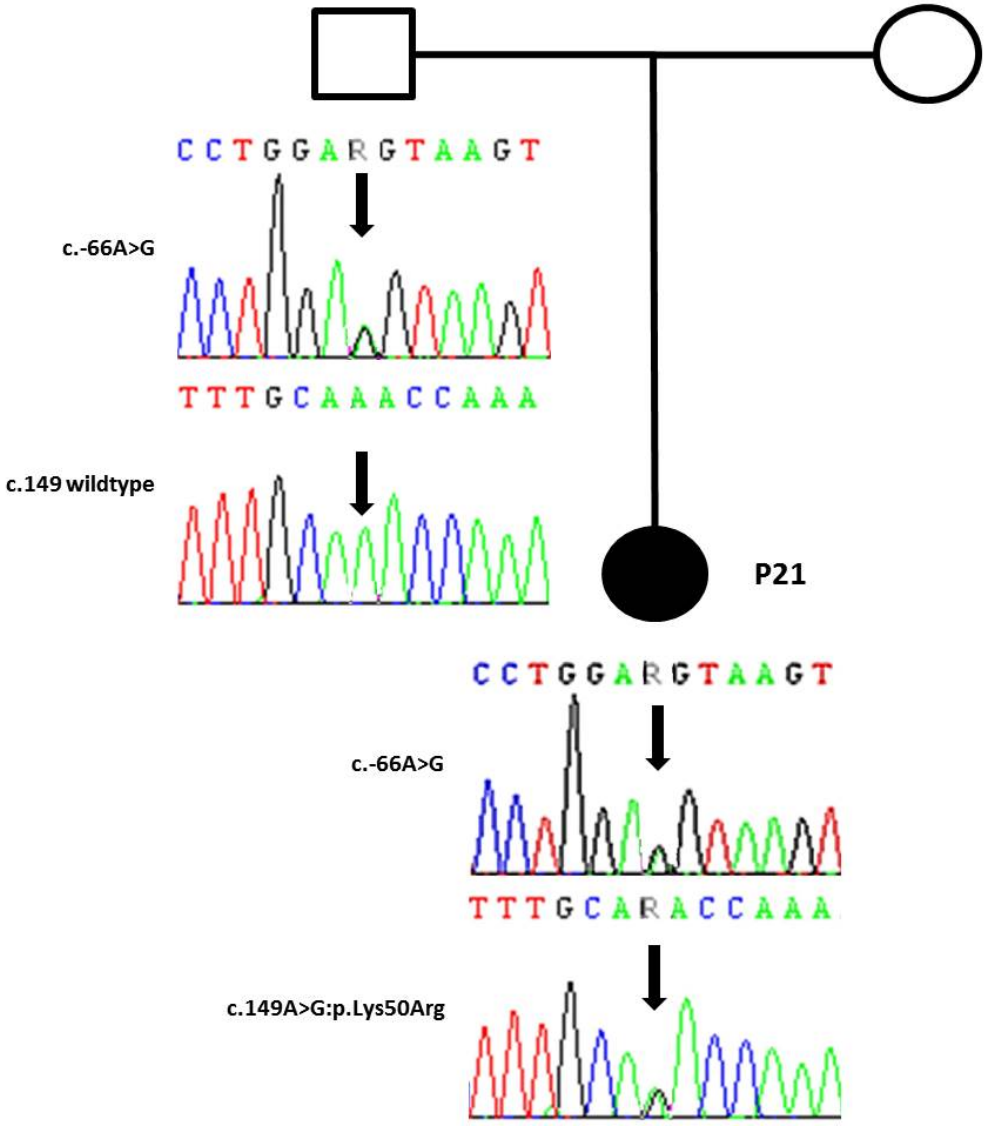
This patient has one frameshift mutation (c.2450delT:p.Leu817Cysfs*12) and one stop gain mutation at a highly conserved residue (c.2338C>T:p.Arg780*) in *ZFYVE26*. The p.Arg780* is predicted to be disease causing or damaging by three prediction programmes. The majority of reported *ZFYVE26* mutations are a loss of function type (33 of 38 mutations described are nonsense and frameshifts). *ZFYVE26* encodes the protein spastizin, in zebrafish this protein is essential for the establishment of motor neuron axonal development (Martin *et al.*, 2012). Spastizin is known to interact with Beclin 1 which is a key regulator of autophagy (Sagona *et al.*, 2011). A recent study by Vantaggiato and co-workers investigated four patient cell lines, two with truncating mutations and two with missense mutations in *ZFYVE26*. The truncating mutations led to a loss of spastizin protein. They also observed a lack of spastizin interaction with Beclin 1 and therefore an impairment of autophagosome maturation. This led to an accumulation of immature autophagosomes. The authors hypothesised that this is likely to occur in neuronal cells linking defective autophagy in the neurodegenerative process underlying the SPG15, although they did not directly test this (Vantaggiato *et al.*, 2013). In addition *ZFYVE26* knockout mice develop late-onset spastic paraplegia with cerebellar ataxia giving further confirmation that SPG15 is caused by spastizin deficiency (Khundadze *et al.*, 2013).

Patient 20: (*WFS1*, c.577A>C:p.Lys193Gln; c.1367G>A:p.Arg456His). This female presented aged 40 with ataxia, dysarthria and vertigo. She went on to develop a worsening visual acuity, optic atrophy, dysarthria and dysphasia. Initial hearing problems improved with surgery. Facial weakness was apparent during stroke-like episodes. Muscle biopsy showed 5% COX negative ragged-red fibres and multiple mitochondrial DNA deletions in muscle DNA. Detailed neuroendocrine testing was not possible because the patient died before the molecular diagnosis was reached. This patient was subjected to exome sequencing. Sanger sequencing showed that only one of the *WFS1* variants was present in her unaffected sibling, confirming that the mutations were compound heterozygous.

Mutations in *WFS1* cause a severe early onset neurodegenerative disorder, which includes diabetes and optic atrophy, of which ataxia is also one symptom. Dominant milder combinations of optic atrophy and deafness (but not ataxia) have also been described with *WFS1* missense mutations (Hogewind *et al.*, 2010). The mutations p.Lys193Gln and p.Arg456His have been previously described in patients with Wolfram syndrome (Cryns *et al.*, 2003). Both mutations occur at conserved residues with two and three prediction tools supporting a pathogenic role.

Patient 21: (*FASTKD2*, c.-66A>G; c.149A>G:p.Lys50Arg). This female patient presented in infancy with a walking delay at 23 months, wobbly walking and frequent falls. There was no evidence of family history. She went on to develop moderate learning difficulties, dysarthria, gait or truncal ataxia and was wheelchair bound with spastic quadraparesis. Eye examination revealed poor bilateral visual acuity, spectacle corrected hypermetropia, pendular nystagmus, ocular motor apraxia and bilateral optic atrophy. Electrophysiology revealed electroencephalography and non-specific mild cortical disturbance. Muscle biopsy results showed low coenzyme Q10 in muscle. Exome sequencing was performed on this patient.

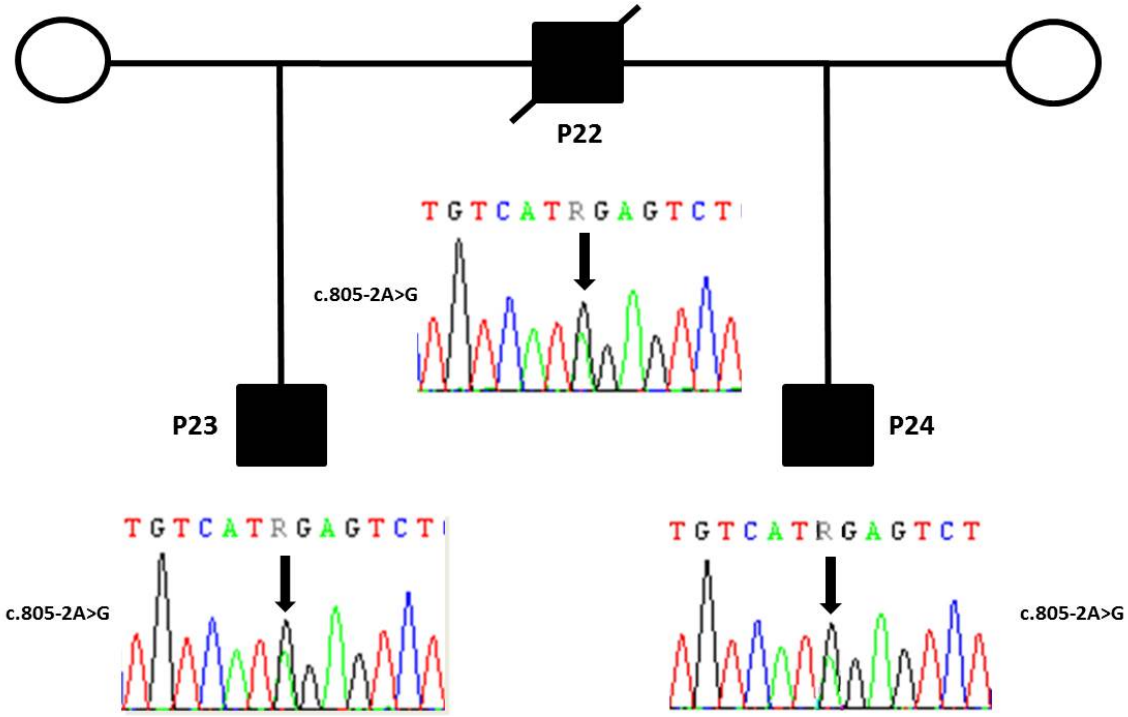
Sanger sequencing of these variants showed the segregation within the family.



The c.-66A>G *FASTKD2*mutation is predicted to affected the splicing of exon 1 in *FASTKD2* (Figure 4.1, A). The p.Lys50Arg mutation is rare in the population. A missense mutation in *FASTKD2* was reported in a patient with infantile mitochondrial encephalopathy. The patient also had developmental delay, asymmetrical brain atrophy and low cytochrome c oxidase activity in skeletal muscle (Ghezzi *et al.*, 2008).

Patient 22, 23 and 24: (*ZFYVE27*, c.805-2A>G). Patient 22, the affected father of Patient 23, presented at age 37 with clumsiness. He went on to develop ataxia with decreased mobility, dysarthria, bilateral horizontal nystagmus on lateral gaze, left cataract and dysarthria. MRI revealed cerebellar and pontine atrophy. His affected son (Patient 23) presented at age 35 with clumsiness. He went on to develop ataxia, dysarthria, diplopia and dysphagia. MRI revealed cerebellar atrophy. Muscle biopsy tests showed multiple mitochondrial DNA deletions and a small frequency of COX negative fibres. Patient 24 (affected half-brother of P22 and son of P23) presented at age 25 with clumsiness. He went on to develop ataxia, dysarthria, jerky ocular pursuits, gaze-evoked nystagmus and hypometric saccades. MRI revealed cerebellar ataxia. Exome sequencing was performed on these three patients.

Sanger sequencing of this variant showed the segregation of the mutations with the disease phenotype in the family.

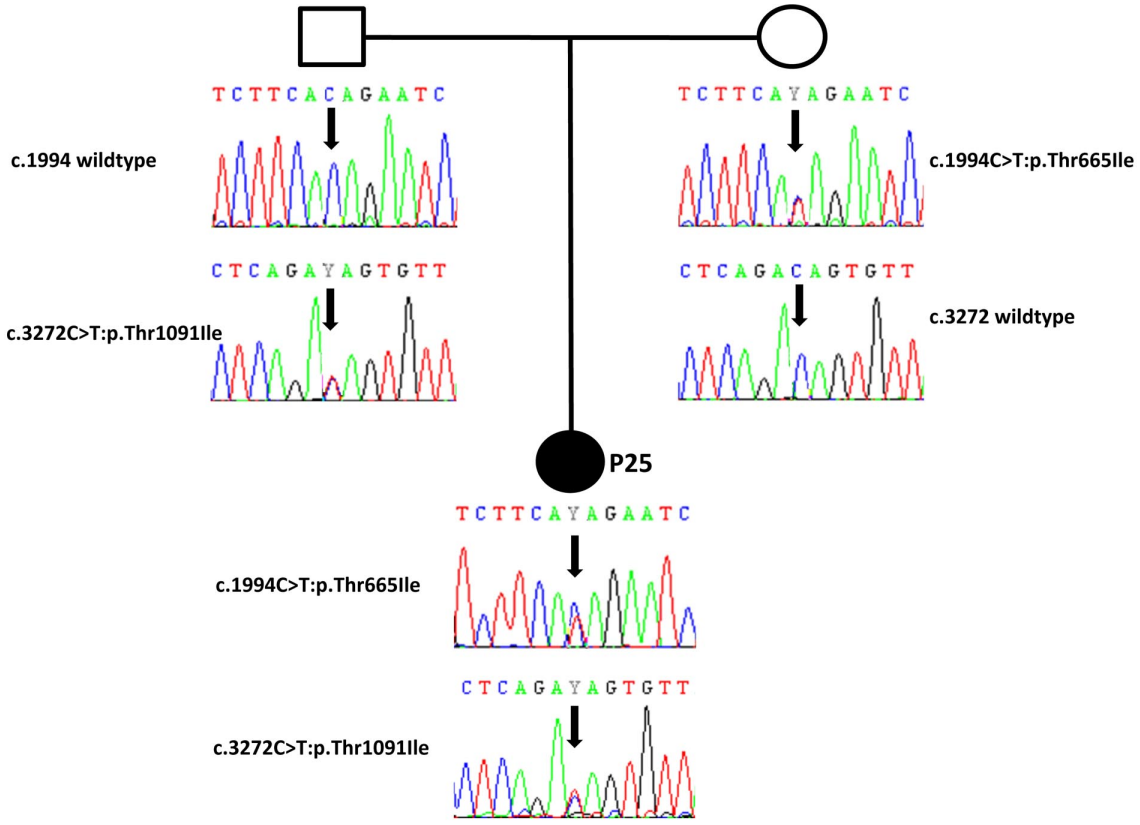


ZFYVE27, c.805-2A>G is very rare in the population and predicted to cause the loss of acceptor site of exon 7 (Figure 4.1, B). Sanger sequencing of this variant showed the segregation of the mutations with the disease phenotype in the family. *ZFYVE27*

encodes Protrudin, a novel member of the FYVE family, which play a role in the formation of neurite extensions by promoting directional membrane trafficking in neurons. Protrudin is a specific spastin-binding protein (Stenmark *et al.*, 2002). In autosomal dominant hereditary spastic paraplegia, spastin is the most commonly mutated protein. Although mutations in *ZFYVE27* have not previously been associated with ataxia, p.G191V in *ZFYVE27* has been reported in one family with hereditary spastic paraplegia (SPG33) (Mannan *et al.*, 2006). Although the functional consequences of p.G191V have been disputed (Martignoni *et al.*, 2008). subsequent studies showed that the interaction with spastin is mediated by the *N*-terminal domain of protruding, and that p.G191V altered the subcellular distribution of protrudin and impaired the yolk sac extension in zebrafish (Zhang *et al.*, 2012).

Patient 25: (*WNK1*, c.1994C>T:p.Thr665Ile; c.3272C>T:p.Thr1091Ile). This female patient presented at age 11 with unsteadiness and was unable to tie her shoe laces. Her symptoms progressed with ataxia, dysarthria, gaze-evoked nystagmus. MRI revealed cavum velum interpositum variant. This patient was subjected to exome sequencing.

Sanger sequencing of these variants showed the segregation of the mutations with the disease phenotype in the family.



WNK1 p.Thr665Ile occurs at a highly conserved residue and is predicted to be pathogenic by four different prediction tools. *WNK1* is a known disease gene but has not been previously associated specifically with ataxia. Mutations in *WNK1* (*HSN2*) have previously been reported in families with hereditary sensory and autonomic neuropathy type II (HSAN2) (Lafreniere *et al.*, 2004; Shekarabi *et al.*, 2008).

Appendix 2. List of variants in patients where we were unable to confidently identify likely candidates

Key: AFP – alpha-fetoprotein
This table is included in Pyle *et al*., 2014.

Patient ID	*Gene*	*Mutation*	*Reason for exclusion*	*Sanger sequenced*
P28	*CYP27A1*	het c.1183C>T:p.Arg395Cys	Cholestanol results for P28 were normal	no
	KCNA1	het c.912G>C:p.Lys304Asn	Variant was wildtype on Sanger sequencing	yes
	AKAP1	het c.98T>G:p.Val33Gly		no
	ASPM	het c.8558G>A:p.Arg2853Gln		no
	NAB2	het c.689C>G:p.Thr230Ser		no
	SLC35C1	het c.1055C>A:p.Pro352His		no
	GARS	het c.112G>A:p.Ala38Thr		no
	NDE1	het c.921C>A:p.Ser307Arg		no
	GSC	het c.488A>G:p.His163Arg		no
	CELSR2	c.2008G>C:p.Ala670Pro; c.5099A>C:p.His1700Pro; c.6044G>A:p.Arg2015Lys; c.7649G>A:p.Arg2550Gln		no
	AHNAK	het c.12113A>G:p.Lys4038Arg		no
	PDE1B	het c.307T>C:p.Ser103Pro		no
	LRP1	het c.10295G>A:p.Arg3432His		no
	PKD1	c.11815T>G:p.Trp3939Gly ; c.7913A>G:p.His2638Arg; c.7065+2T>G; c.4195T>C:p.Trp1399Arg		no
	SNTG2	c.149A>C:p.Glu50Ala; c.1018G>C:p.Asp340His		no
	PKDREJ	het c.4498C>T:p.Pro1500Ser		no
	CHL1	het c.671T>C:p.Val224Ala		no
	BSN	c.10318A>G:p.Ser3440Gly ; c.10331G>A:p.Arg3444Gln		no

	HTT	c.107A>C:p.Gln36Pro; c.110A>C:p.Gln37Pro; c.113A>C:p.Gln38Pro		no
	REST	c.2013G>T:p.Glu671Asp; c.2206C>T:p.Pro736Ser		no
	UTRN	c.404_405insG: p.Val136Glyfs*18; c.7984C>T:p.Arg2662Cys		no
	PLXNA4	c.5608C>T:p.His1870Tyr; c.1249G>A:p.Asp417Asn		no
	SFTPD	het c.883G>A:p.Ala295Thr		no
	APPL2	c.209A>G:p.Lys70Arg		no
	APOBR	c.1083G>C:p.Glu361Asp; c.1090G>T:p.Gly364Trp		no
	DOK2	het c.821C>T:p.Pro274Leu		no
	RASGRP 2	c.1528G>A:p.Val510Ile; c.1466C>G:p.Ser489Cys		no
P29 and P30	*TTBK2*	c.3691A>C:p.Thr1231Pro	Variant was wildtype in both patients on Sanger sequencing	yes
	ATXN2	het c.454T>C:p.Cys152Arg	Genetic test for *ATXN2* trinucleotide repeat expansions was negative	no
	ATP6V1B 1	het c.1493A>G:p.Tyr498Cys		no
	ASPM	het c.8128A>G:p.Thr2710Ala		no
	TYRO3	het c.1246C>T:p.Arg416Cys		no
	SLC18A2	het c.607+2T>G		no
	SRRM4	het c.1091C>T:p.Pro364Leu		no
	CSPG4	het c.4397T>C:p.Val1466Ala		no
	ABCC1	het c.3725A>G:p.Asn1242Ser		no
	CDH13	het c.2069T>G:p.Leu690Arg		no
	FEZ2	het c.1025C>G:p.Pro342Arg		no
	GLI1	het c.651C>G:p. Ser217Arg		no
	PLEKHA 6	het c.3082A>C:p.Met1028Leu		no

	EXOC8	het c.485T>C:p.Leu162Pro		no
	DDHD1	het c.455C>T:p.Ala152Val		no
	SREBF1	het c.422C>T:p.Pro141Leu		no
P31	*MT-ATP6*	m. 9006_A>G	Not pathogenic	no
	MT-ND5	m.14002_A>G	Not pathogenic	no
	MT-ND6	m.14040_G>A	Not pathogenic	no
	IKBKAP	c.934G>A:p.Glu312Lys; c.751A>G:p.Ser251Gly		no
	HSPG2	c.9422G>A:p.Arg3141His; c.G1238A:p.Arg413Gln		no
	NOTCH1	het c.4111A>C:p.Thr1371Pro		no
	CELSR2	c.2008G>C:p.Ala670Pro; c.3602C>T:p.Pro1201Leu; c.5099A>C:p.His1700Pro; c.5699A>G:p.Asn1900Ser		no
	MEF2D	het c.1330T>C:p.Ser444Pro		no
	NCSTN	het c.1229C>T:p.Ala410Val		no
	CACNA1S	het c.2099C>T:p.Thr700Met		no
	ITGA8	het c.851T>C:p.Leu284Ser		no
	NELL1	het c.578G>A:p.Arg193His		no
	MLST8	het c.175G>C:p.Ala59Pro		no
	TSHZ3	het c.3082A>G:p.Met1028Val		no
	SNTG2	het c.149A>C:p.Glu50Ala		no
	JAG1	het c.800C>G:p.Pro267Arg		no
	PLCG1	het c.1073G>A:p.Arg358Gln		no
	PIGT	het c.109A>C:p.Thr37Pro		no
	EOMES	het c.200A>G:p.Glu67Gly		no
	SRRT	het c.1520A>G:p.Asn507Ser		no
	ACHE	het c.1673A>C:p.Gln558Pro		no
	LAMB1	het c.1046C>G:p.Ala349Gly		no
	CSGALNACT1	het c.16C>T:p.Arg6Trp		no
	EXOC8	het c.395A>C:p.His132Pro		no
	ARHGAP21	het c.1718G>A:p.Ser573Asn		no
	ABCA7	het c.717C>A:p.Asn239Lys		no
	AP3D1	het c.3418C>T:p.Arg1140Trp		no

	UGT1A4	het c.310_310delG: p.Glu104Asnfs*4		no
	NCOA3	het c.3374G>A:p.Gly1125Asp		no
	SLCO2A1	het c.166C>T:p.Leu56Phe		no
	CLVS2	het c.922G>T:p.Val308Leu		no
	NEU1	het c.428T>A:p.Val143Glu		no
	CADPS2	c.1151G>A:p.Arg384Gln		no
	PCLO	het c.2290A>G:p.Thr764Ala		no
	PREX2	het c.2971A>G:p.Thr991Ala		no
	MELK	c.77A>G:p.Lys26Arg; c.1481_1484delTGTT: p.Phe495Glyfs*30		no
	UROS	het c.689G>C:p.Arg230Pro		no
	OGDHL	het c.2869C>T:p.Pro957Ser		no
	APEX1	c.416_419delCACT: p.Leu140Lysfs*6		no
	TBC1D21	het c.902A>G:p.Asn301Ser		no
	TSHZ3	het c.3082A>G:p.Met1028Val		no
	TMEM147	het c.488C>T:p.Thr163Ile		no
	PTPRH	het c.2425G>A:p.Val809Ile		no
	CYP2D6	het c.360_361insGTT: p.Phe120delinsLeuPhe		no
	LRRN1	c.460C>A:p.Gln154Lys; c.1676C>A:p.Ala559Asp		no
	SCAF8	het c.80A>G:p.Glu27Gly		no
	VARS2	c.423G>A:p.Met141Ile; c.1735+1G>T	variants inherited in *cis*	yes
	GNMT	het c.113C>G:p.Thr38Ser		no
	ATM	het c.146C>G:p.Ser49Cys	only one variant (ATM is recessive); AFP test neg	no
P32	*PPP2R2B* (SCA12)	het c.752A>G:p.Tyr251Cys	SCA12 trinucleotide repeat expansion test neg	no
	MFN2	het c.1988G>T:p.Arg663Leu	*MFN2* sanger sequencing neg	no
	TPP1	het c.101G>A:p.Gly34Asp	*TPP1* biochemical assay normal	yes
	POLG	c.2492A>G:p.Tyr831Cys; c.1550G>T:p.Gly517Val	Not pathogenic	no

	MYH14	het c.5504G>A:p.Arg1835His	Variant does not segregate with disease phenotype in the family	yes
	KIAA0196 (SPG8)	het c.925G>A:p.Ala309Thr	Variant does not segregate with disease phenotype in the family	yes
	CEP152	het c.1507A>G:p.Ile503Val		no
	PNKP	het c.1360C>A:p.Leu454Met		no
	INPPL1	het c.1706C>T:p.Thr569Met		no
	BRCA1	het c.1487G>A:p.Arg496His		no
	BARHL2	het c.371C>A:p.Pro124His		no
	AGRN	het c.3353C>A:p.Thr1118Lys		no
	P2RY2	het c.515T>C:p.Val172Ala		no
	GAS6	het c.1690C>A:p.Leu564Ile		no
	SYNM	het c.1050G>T:p.Arg350Ser		no
	IRX3	het c.1016C>T:p.Ala339Val		no
	PRKCD	het c.1213G>A:p.Ala405Thr		no
	PCDHA11	het c.1930C>G:p.His644Asp		no
	PRKAR1B	het c.695G>A:p.Arg232Gln		no
	PRKACG	het c.558C>G:p.Phe186Leu		no
	ACAP3	het c.2336C>G:p.Ala779Gly		no
	ATP8B2	het c.2321A>G:p.Lys774Arg		no
	FERMT3	het c.119C>T:p.Ser40Leu		no
	PLCB3	het c.1492T>C:p.Ser498Pro		no
	CDC42BPG	het c.4272-2A>G		no
	APOBR	c.1083G>C:p.Glu361Asp; c.1090G>T:p.Gly364Trp		no
	ATP8B3	c.2950G>A:p.Val984Met; c.637G>A:p.Ala213Thr		no
	LIPE	het c.2827_2828insC:p.Glu943fs		no
	ROCK2	het c.1802A>T:p.Asp601Val		no

	LSS	het c.1564G>A:p.Gly522Arg		no
P33 and P34	*IGDCC3*	c.1984G>A:p.Gly662Ser; c.1207G>A:p.Glu403Lys	Variant does not segregate with disease phenotype in the family	yes
	HIF1a	c.217C>A:p.Leu73Met; c.517G>T:p.Asp173Tyr	Variant does not segregate with disease phenotype in the family	yes
	MT-CO1	m.6869_C>T	polymorphism	no
	ZEB2	c.34T>G:p.Cys12Gly; c.49C>A:p.Gln17Lys		no
	IRS1	c.2482A>C:p.Ser828arg; c.1159A>C:p.Thr387Pro		no
	CHRND	c.158A>C:p.His53Pro; c.280C>G:p.Gln94Glu		no
	PCDHB15	c.1490A>C:p.His497Pro; c.1717A>C:p.Thr573Pro		no
	TNC	c.5731C>T:p.Arg1911Trp; c.5341G>A:p.Ala1781Thr; c.2374G>T:p.Gly792Cys		no
	PLXNA2	c.4606G>A:p.Val1536Met; c.2161A>C:p.Thr721Pro		no
	ACAD8	c.118A>C:p.Thr40Pro; c.124A>C:p.Thr42Pro		no
	SYT7	c.950T>G:p.Val317Gly; c.607A>C:p.Thr203Pro		no
P35	*TIMMDC1*	c.176G>A:p.Arg59Gln; c.285G>T:p.Arg95Ser	inconclusive (mother is heterozygous for one variant; father's DNA sample unavailable)	yes
	SACS	Het c.11032C>G:p.P3678A	only one variant found	yes
	MTATP6	m.9052 A>G	Not pathogenic	yes
	ASCL1	c.150_152delGCA:p.50_51del; c.152_153insGCA:p.Q51delinsQQ		no
	ABCB6	c.1028G>A:p.Arg343Gln; c.490G>A:p.Ala164Thr		no
	HTT	c.107A>C:p.Gln36Pro; c.110A>C:p.Gln37Pro; c.113A>C:p.Gln38Pro		no
	BHLHE22	c.673G>A:p.Gly225Ser; c.676A>G:p.Ser226Gly		no

	UGT1A4	c.119C>A:p.Pro40His; c.310_310delG:p.Glu104Asnfs*4		no
	TNXB	c.12520G>A:p.Asp4174Asn; c.10673G>A:p.Arg3558His; c.9631C>G:p.Arg3211Gly; c.7235C>T:p.Pro2412Leu		no

Bibliography

Abdollahi, M.R., Morrison, E., Sirey, T., Molnar, Z., Hayward, B.E., Carr, I.M., Springell, K., Woods, C.G., Ahmed, M., Hattingh, L., Corry, P., Pilz, D.T., Stoodley, N., Crow, Y., Taylor, G.R., Bonthron, D.T. and Sheridan, E. (2009) 'Mutation of the variant alpha-tubulin TUBA8 results in polymicrogyria with optic nerve hypoplasia', *Am J Hum Genet*, 85(5), pp. 737-44.

Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) 'A map of human genome variation from population-scale sequencing', *Nature*, 467(7319), pp. 1061-73.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) 'A method and server for predicting damaging missense mutations', *Nat Methods*, 7(4), pp. 248-9.

Akhmanova, A. and Steinmetz, M.O. (2008) 'Tracking the ends: a dynamic protein network controls the fate of microtubule tips', *Nat Rev Mol Cell Biol*, 9(4), pp. 309-22.

Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H. and Durbin, R. (2011) 'Dindel: accurate indel calls from short-read data', *Genome Res*, 21(6), pp. 961-73.

Aldinger, K.A., Mosca, S.J., Tetreault, M., Dempsey, J.C., Ishak, G.E., Hartley, T., Phelps, I.G., Lamont, R.E., O'Day, D.R., Basel, D., Gripp, K.W., Baker, L., Stephan, M.J., Bernier, F.P., Boycott, K.M., Majewski, J., Parboosingh, J.S., Innes, A.M. and Doherty, D. (2014) 'Mutations in LAMA1 cause cerebellar dysplasia and cysts with and without retinal dystrophy', *Am J Hum Genet*, 95(2), pp. 227-34.

Ali, B.R., Silhavy, J.L., Gleeson, M.J., Gleeson, J.G. and Al-Gazali, L. (2012) 'A missense founder mutation in VLDLR is associated with Dysequilibrium Syndrome without quadrupedal locomotion', *BMC Med Genet*, 13, p. 80.

Allen, N.J., Bennett, M.L., Foo, L.C., Wang, G.X., Chakraborty, C., Smith, S.J. and Barres, B.A. (2012) 'Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors', *Nature*, 486(7403), pp. 410-4.

Allikmets, R., Raskind, W.H., Hutchinson, A., Schueck, N.D., Dean, M. and Koeller, D.M. (1999) 'Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A)', *Hum Mol Genet*, 8(5), pp. 743-9.

Altmuller, J., Palmer, L.J., Fischer, G., Scherb, H. and Wjst, M. (2001) 'Genomewide scans of complex human diseases: true linkage is hard to find', *Am J Hum Genet*, 69(5), pp. 936-50.

Andrade, D.M., Paton, T., Turnbull, J., Marshall, C.R., Scherer, S.W. and Minassian, B.A. (2012) 'Mutation of the CLN6 gene in teenage-onset progressive myoclonus epilepsy', *Pediatr Neurol*, 47(3), pp. 205-8.

Anheim, M., Fleury, M., Monga, B., Laugel, V., Chaigne, D., Rodier, G., Ginglinger, E., Boulay, C., Courtois, S., Drouot, N., Fritsch, M., Delaunoy, J.P., Stoppa-Lyonnet, D., Tranchant, C. and Koenig, M. (2010) 'Epidemiological, clinical, paraclinical and molecular study of a cohort of 102 patients affected with autosomal recessive progressive cerebellar ataxia from Alsace, Eastern France: implications for clinical management', *Neurogenetics*, 11(1), pp. 1-12.

Antonarakis, S.E. and Beckmann, J.S. (2006) 'Mendelian disorders deserve more attention', *Nat Rev Genet*, 7(4), pp. 277-82.

Anttonen, A.K. and Lehesjoki, A.E. (1993-) 'Marinesco-Sjogren Syndrome', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301371.

Assoum, M., Salih, M.A., Drouot, N., Hnia, K., Martelli, A. and Koenig, M. (2013) 'The Salih ataxia mutation impairs Rubicon endosomal localization', *Cerebellum*, 12(6), pp. 835-40.

Baas, P.W. and Ahmad, F.J. (2013) 'Beyond taxol: microtubule-based treatment of disease and injury of the nervous system', *Brain*, 136(Pt 10), pp. 2937-51.

Bakalkin, G., Watanabe, H., Jezierska, J., Depoorter, C., Verschuuren-Bemelmans, C., Bazov, I., Artemenko, K.A., Yakovleva, T., Dooijes, D., Van de Warrenburg, B.P., Zubarev, R.A., Kremer, B., Knapp, P.E., Hauser, K.F., Wijmenga, C., Nyberg, F., Sinke, R.J. and Verbeek, D.S. (2010) 'Prodynorphin mutations cause the neurodegenerative disorder spinocerebellar ataxia type 23', *Am J Hum Genet*, 87(5), pp. 593-603.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A. and Shendure, J. (2011) 'Exome sequencing as a tool for Mendelian disease gene discovery', *Nat Rev Genet*, 12(11), pp. 745-55.

Barcia, G., Fleming, M.R., Deligniere, A., Gazula, V.R., Brown, M.R., Langouet, M., Chen, H., Kronengold, J., Abhyankar, A., Cilio, R., Nitschke, P., Kaminska, A., Boddaert, N., Casanova, J.L., Desguerre, I., Munnich, A., Dulac, O., Kaczmarek, L.K., Colleaux, L. and Nabbout, R. (2012) 'De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy', *Nat Genet*, 44(11), pp. 1255-9.

Barres, B.A. (2008) 'The mystery and magic of glia: a perspective on their roles in health and disease', *Neuron*, 60(3), pp. 430-40.

Bayat, V., Thiffault, I., Jaiswal, M., Tetreault, M., Donti, T., Sasarman, F., Bernard, G., Demers-Lamarche, J., Dicaire, M.J., Mathieu, J., Vanasse, M., Bouchard, J.P., Rioux, M.F., Lourenco, C.M., Li, Z., Haueter, C., Shoubridge, E.A., Graham, B.H., Brais, B. and Bellen, H.J. (2012) 'Mutations in the mitochondrial methionyl-tRNA synthetase cause a neurodegenerative phenotype in flies and a recessive ataxia (ARSAL) in humans', *PLoS Biol*, 10(3), p. e1001288.

Berger, I., Dor, T., Halvardson, J., Edvardson, S., Shaag, A., Feuk, L. and Elpeleg, O. (2012) 'Intractable epilepsy of infancy due to homozygous mutation in the EFHC1 gene', *Epilepsia*, 53(8), pp. 1436-40.

Bidichandani, S.I. and Delatycki, M.B. (1993-) 'Friedreich Ataxia', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301458.

Biesecker, L.G., Burke, W., Kohane, I., Plon, S.E. and Zimmern, R. (2012) 'Next-generation sequencing in the clinic: are we ready?', *Nat Rev Genet*, 13(11), pp. 818-24.

Bird, T.D. (1993 -) 'Hereditary Ataxia Overview', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301317.

Blanchard, M.G., Willemsen, M.H., Walker, J.B., Dib-Hajj, S.D., Waxman, S.G., Jongmans, M.C., Kleefstra, T., van de Warrenburg, B.P., Praamstra, P., Nicolai, J., Yntema, H.G., Bindels, R.J., Meisler, M.H. and Kamsteeg, E.J. (2015) 'De novo gain-of-function and loss-of-function mutations of SCN8A in patients with intellectual disabilities and epilepsy', *J Med Genet*, 52(5), pp. 330-7.

Blencowe, B.J. (2000) 'Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases', *Trends Biochem Sci*, 25(3), pp. 106-10.

Blumkin, L., Halevy, A., Ben-Ami-Raichman, D., Dahari, D., Haviv, A., Sarit, C., Lev, D., van der Knaap, M.S., Lerman-Sagie, T. and Leshinsky-Silver, E. (2014) 'Expansion of the spectrum of TUBB4A-related disorders: a new phenotype associated with a novel mutation in the TUBB4A gene', *Neurogenetics*, 15(2), pp. 107-13.

Bomar, J.M., Benke, P.J., Slattery, E.L., Puttagunta, R., Taylor, L.P., Seong, E., Nystuen, A., Chen, W., Albin, R.L., Patel, P.D., Kittles, R.A., Sheffield, V.C. and Burmeister, M. (2003) 'Mutations in a novel gene encoding a CRAL-TRIO domain cause human Cayman ataxia and ataxia/dystonia in the jittery mouse', *Nat Genet*, 35(3), pp. 264-9.

Boone, P.M., Wiszniewski, W. and Lupski, J.R. (2011) 'Genomic medicine and neurological disease', *Hum Genet*, 130(1), pp. 103-21.

Bourassa, C.V., Meijer, I.A., Merner, N.D., Grewal, K.K., Stefanelli, M.G., Hodgkinson, K., Ives, E.J., Pryse-Phillips, W., Jog, M., Boycott, K., Grimes, D.A., Goobie, S., Leckey, R., Dion, P.A. and Rouleau, G.A. (2012) 'VAMP1 mutation causes dominant hereditary spastic ataxia in Newfoundland families', *Am J Hum Genet*, 91(3), pp. 548-52.

Boycott, K.M., Vanstone, M.R., Bulman, D.E. and MacKenzie, A.E. (2013) 'Rare-disease genetics in the era of next-generation sequencing: discovery to translation', *Nat Rev Genet*, 14(10), pp. 681-91.

Bradford, M.M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal Biochem*, 72, pp. 248-54.

Bras, J., Alonso, I., Barbot, C., Costa, M.M., Darwent, L., Orme, T., Sequeiros, J., Hardy, J., Coutinho, P. and Guerreiro, R. (2015) 'Mutations in PNKP cause recessive ataxia with oculomotor apraxia type 4', *Am J Hum Genet*, 96(3), pp. 474-9.

Bras, J., Guerreiro, R. and Hardy, J. (2012) 'Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease', *Nat Rev Neurosci*, 13(7), pp. 453-64.

Bras, J.M. and Singleton, A.B. (2011) 'Exome sequencing in Parkinson's disease', *Clin Genet*, 80(2), pp. 104-9.

Breedveld, G.J., van Wetten, B., te Raa, G.D., Brusse, E., van Swieten, J.C., Oostra, B.A. and Maat-Kievit, J.A. (2004) 'A new locus for a childhood onset, slowly progressive autosomal recessive spinocerebellar ataxia maps to chromosome 11p15', *J Med Genet*, 41(11), pp. 858-66.

Breuss, M., Heng, J.I., Poirier, K., Tian, G., Jaglin, X.H., Qu, Z., Braun, A., Gstrein, T., Ngo, L., Haas, M., Bahi-Buisson, N., Moutard, M.L., Passemard, S., Verloes, A., Gressens, P., Xie, Y., Robson, K.J., Rani, D.S., Thangaraj, K., Clausen, T., Chelly, J., Cowan, N.J. and Keays, D.A. (2012) 'Mutations in the beta-tubulin

gene TUBB5 cause microcephaly with structural brain abnormalities', *Cell Rep*, 2(6), pp. 1554-62.

Brkanac, Z., Fernandez, M., Matsushita, M., Lipe, H., Wolff, J., Bird, T.D. and Raskind, W.H. (2002) 'Autosomal dominant sensory/motor neuropathy with Ataxia (SMNA): Linkage to chromosome 7q22-q32', *Am J Med Genet*, 114(4), pp. 450-7.

Brkanac, Z., Spencer, D., Shendure, J., Robertson, P.D., Matsushita, M., Vu, T., Bird, T.D., Olson, M.V. and Raskind, W.H. (2009) 'IFRD1 is a candidate gene for SMNA on chromosome 7q22-q23', *Am J Hum Genet*, 84(5), pp. 692-7.

Brunham, L.R. and Hayden, M.R. (2013) 'Hunting human disease genes: lessons from the past, challenges for the future', *Hum Genet*, 132(6), pp. 603-17.

Brunham, L.R., Singaraja, R.R., Pape, T.D., Kejariwal, A., Thomas, P.D. and Hayden, M.R. (2005) 'Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene', *PLoS Genet*, 1(6), p. e83.

Burda, J.E. and Sofroniew, M.V. (2014) 'Reactive gliosis and the multicellular response to CNS damage and disease', *Neuron*, 81(2), pp. 229-48.

Burns, R., Majczenko, K., Xu, J., Peng, W., Yapici, Z., Dowling, J.J., Li, J.Z. and Burmeister, M. (2014) 'Homozygous splice mutation in CWF19L1 in a Turkish family with recessive ataxia syndrome', *Neurology*, 83(23), pp. 2175-82.

Cader, M.Z., Steckley, J.L., Dyment, D.A., McLachlan, R.S. and Ebers, G.C. (2005) 'A genome-wide screen and linkage mapping for a large pedigree with episodic ataxia', *Neurology*, 65(1), pp. 156-8.

Cadieux-Dion, M., Turcotte-Gauthier, M., Noreau, A., Martin, C., Meloche, C., Gravel, M., Drouin, C.A., Rouleau, G.A., Nguyen, D.K. and Cossette, P. (2014) 'Expanding the clinical phenotype associated with ELOVL4 mutation: study of a large French-Canadian family with autosomal dominant spinocerebellar ataxia and erythrokeratodermia', *JAMA Neurol*, 71(4), pp. 470-5.

Cagnoli, C., Mariotti, C., Taroni, F., Seri, M., Brussino, A., Michielotto, C., Grisoli, M., Di Bella, D., Migone, N., Gellera, C., Di Donato, S. and Brusco, A. (2006) 'SCA28, a novel form of autosomal dominant cerebellar ataxia on chromosome 18p11.22-q11.2', *Brain*, 129(Pt 1), pp. 235-42.

Calvo, S.E., Compton, A.G., Hershman, S.G., Lim, S.C., Lieber, D.S., Tucker, E.J., Laskowski, A., Garone, C., Liu, S., Jaffe, D.B., Christodoulou, J., Fletcher, J.M., Bruno, D.L., Goldblatt, J., Dimauro, S., Thorburn, D.R. and Mootha, V.K. (2012) 'Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing', *Sci Transl Med*, 4(118), p. 118ra10.

Calvo, S.E. and Mootha, V.K. (2010) 'The mitochondrial proteome and human disease', *Annu Rev Genomics Hum Genet*, 11, pp. 25-44.

Calvo, S.E., Tucker, E.J., Compton, A.G., Kirby, D.M., Crawford, G., Burtt, N.P., Rivas, M., Guiducci, C., Bruno, D.L., Goldberger, O.A., Redman, M.C., Wiltshire, E., Wilson, C.J., Altshuler, D., Gabriel, S.B., Daly, M.J., Thorburn, D.R. and Mootha, V.K. (2010) 'High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency', *Nat Genet*, 42(10), pp. 851-8.

Carstea, E.D., Morris, J.A., Coleman, K.G., Loftus, S.K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M.A., Pavan, W.J., Krizman, D.B., Nagle, J., Polymeropoulos, M.H., Sturley, S.L., Ioannou, Y.A., Higgins, M.E., Comly, M., Cooney, A., Brown, A., Kaneski, C.R., Blanchette-Mackie, E.J., Dwyer, N.K., Neufeld, E.B., Chang, T.Y., Liscum, L., Strauss, J.F., 3rd, Ohno, K., Zeigler, M., Carmi, R., Sokol, J., Markie, D., O'Neill, R.R., van Diggelen, O.P., Elleder, M., Patterson, M.C., Brady, R.O., Vanier, M.T., Pentchev, P.G. and Tagle, D.A. (1997) 'Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis', *Science*, 277(5323), pp. 228-31.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (2003) 'ESEfinder: A web resource to identify exonic splicing enhancers', *Nucleic Acids Res*, 31(13), pp. 3568-71.

Carvalho, D., Santos, S., Martins, B. and Pinto Marques, F. (2015) 'TUBB4A novel mutation reinforces the genotype-phenotype correlation of hypomyelination with atrophy of the basal ganglia and cerebellum', *Brain*, 138(Pt 2), p. e327.

Casari, G., De Fusco, M., Ciarmatori, S., Zeviani, M., Mora, M., Fernandez, P., De Michele, G., Filla, A., Cocozza, S., Marconi, R., Durr, A., Fontaine, B. and Ballabio, A. (1998) 'Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease', *Cell*, 93(6), pp. 973-83.

Chaisson, M.J. and Pevzner, P.A. (2008) 'Short read fragment assembly of bacterial genomes', *Genome Res*, 18(2), pp. 324-30.

Chang, J., Lee, S. and Blackstone, C. (2013) 'Protrudin binds atlastins and endoplasmic reticulum-shaping proteins and regulates network formation', *Proc Natl Acad Sci U S A*, 110(37), pp. 14954-9.

Chang, J., Lee, S. and Blackstone, C. (2014) 'Spastic paraplegia proteins spastizin and spatacsin mediate autophagic lysosome reformation', *J Clin Invest*, 124(12), pp. 5249-62.

Chatterjee, S. and Pal, J.K. (2009) 'Role of 5'- and 3'-untranslated regions of mRNAs in human diseases', *Biol Cell*, 101(5), pp. 251-62.

Chaussenot, A., Bannwarth, S., Rouzier, C., Vialettes, B., Mkadem, S.A., Chabrol, B., Cano, A., Labauge, P. and Paquis-Flucklinger, V. (2011) 'Neurologic features and genotype-phenotype correlation in Wolfram syndrome', *Ann Neurol*, 69(3), pp. 501-8.

Chen, D.H., Bird, T.D. and Raskind, W.H. (1993-) 'Spinocerebellar Ataxia Type 14', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301573.

Chen, D.H., Naydenov, A., Blankman, J.L., Mefford, H.C., Davis, M., Sul, Y., Barloon, A.S., Bonkowski, E., Wolff, J., Matsushita, M., Smith, C., Cravatt, B.F., Mackie, K., Raskind, W.H., Stella, N. and Bird, T.D. (2013) 'Two novel

mutations in ABHD12: expansion of the mutation spectrum in PHARC and assessment of their functional effects', *Hum Mutat*, 34(12), pp. 1672-8.

Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S. and Lifton, R.P. (2009) 'Genetic diagnosis by whole exome capture and massively parallel DNA sequencing', *Proc Natl Acad Sci U S A*, 106(45), pp. 19096-101.

Chung, M.Y., Lu, Y.C., Cheng, N.C. and Soong, B.W. (2003) 'A novel autosomal dominant spinocerebellar ataxia (SCA22) linked to chromosome 1p21-q23', *Brain*, 126(Pt 6), pp. 1293-9.

Cole, J.W., Stine, O.C., Liu, X., Pratap, A., Cheng, Y., Tallon, L.J., Sadzewicz, L.K., Dueker, N., Wozniak, M.A., Stern, B.J., Meschia, J.F., Mitchell, B.D., Kittner, S.J. and O'Connell, J.R. (2012) 'Rare variants in ischemic stroke: an exome pilot study', *PLoS One*, 7(4), p. e35591.

Conti, E. and Izaurralde, E. (2005) 'Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species', *Curr Opin Cell Biol*, 17(3), pp. 316-25.

Cooper, D.N., Chen, J.M., Ball, E.V., Howells, K., Mort, M., Phillips, A.D., Chuzhanova, N., Krawczak, M., Kehrer-Sawatzki, H. and Stenson, P.D. (2010) 'Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics', *Hum Mutat*, 31(6), pp. 631-55.

Cooper, G.M., Stone, E.A., Asimenos, G., Program, N.C.S., Green, E.D., Batzoglou, S. and Sidow, A. (2005) 'Distribution and intensity of constraint in mammalian genomic sequence', *Genome Res*, 15(7), pp. 901-13.

Cooper, T.A., Wan, L. and Dreyfuss, G. (2009) 'RNA and disease', *Cell*, 136(4), pp. 777-93.

Corbett, M.A., Schwake, M., Bahlo, M., Dibbens, L.M., Lin, M., Gandolfo, L.C., Vears, D.F., O'Sullivan, J.D., Robertson, T., Bayly, M.A., Gardner, A.E., Vlaar, A.M.,

Korenke, G.C., Bloem, B.R., de Coo, I.F., Verhagen, J.M., Lehesjoki, A.E., Gecz, J. and Berkovic, S.F. (2011) 'A mutation in the Golgi Qb-SNARE gene GOSR2 causes progressive myoclonus epilepsy with early ataxia', *Am J Hum Genet*, 88(5), pp. 657-63.

Coutinho, P. and Barbot, C. (1993-) 'Ataxia with Oculomotor Apraxia Type 1', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301629.

Crosby, A.H., Patel, H., Chioza, B.A., Proukakis, C., Gurtz, K., Patton, M.A., Sharifi, R., Harlalka, G., Simpson, M.A., Dick, K., Reed, J.A., Al-Memar, A., Chrzanowska-Lightowlers, Z.M., Cross, H.E. and Lightowlers, R.N. (2010) 'Defective mitochondrial mRNA maturation is associated with spastic ataxia', *Am J Hum Genet*, 87(5), pp. 655-60.

Cryns, K., Pfister, M., Pennings, R.J., Bom, S.J., Flothmann, K., Caethoven, G., Kremer, H., Schatteman, I., Koln, K.A., Toth, T., Kupka, S., Blin, N., Nurnberg, P., Thiele, H., van de Heyning, P.H., Reardon, W., Stephens, D., Cremers, C.W., Smith, R.J. and Van Camp, G. (2002) 'Mutations in the WFS1 gene that cause low-frequency sensorineural hearing loss are small non-inactivating mutations', *Hum Genet*, 110(5), pp. 389-94.

Cryns, K., Sivakumaran, T.A., Van den Ouweland, J.M., Pennings, R.J., Cremers, C.W., Flothmann, K., Young, T.L., Smith, R.J., Lesperance, M.M. and Van Camp, G. (2003) 'Mutational spectrum of the WFS1 gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease', *Hum Mutat*, 22(4), pp. 275-87.

D'Adamo, M.C., Hanna, M.G., Di Giovanni, G. and Pessia, M. (1993-) 'Episodic Ataxia Type 1', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301785.

Damiano, J.A., Afawi, Z., Bahlo, M., Mauermann, M., Misk, A., Arsov, T., Oliver, K.L., Dahl, H.M., Shearer, A.E., Smith, R.J., Hall, N.E., Mahmood, K., Leventer, R.J., Scheffer, I.E., Muona, M., Lehesjoki, A.E., Korczyn, A.D.,

Herrmann, H., Berkovic, S.F. and Hildebrand, M.S. (2015) 'Mutation of the nuclear lamin gene LMNB2 in progressive myoclonus epilepsy with early ataxia', *Hum Mol Genet*.

Daoud, H., Zhou, S., Noreau, A., Sabbagh, M., Belzil, V., Dionne-Laporte, A., Tranchant, C., Dion, P. and Rouleau, G.A. (2012) 'Exome sequencing reveals SPG11 mutations causing juvenile ALS', *Neurobiol Aging*, 33(4), pp. 839 e5-9.

de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B.B., Brunner, H.G., Veltman, J.A. and Vissers, L.E. (2012) 'Diagnostic exome sequencing in persons with severe intellectual disability', *N Engl J Med*, 367(20), pp. 1921-9.

De Schutter, E. and Bower, J.M. (1994) 'An active membrane model of the cerebellar Purkinje cell. I. Simulation of current clamps in slice', *J Neurophysiol*, 71(1), pp. 375-400.

de Vries, B., Mamsa, H., Stam, A.H., Wan, J., Bakker, S.L., Vanmolkot, K.R., Haan, J., Terwindt, G.M., Boon, E.M., Howard, B.D., Frants, R.R., Baloh, R.W., Ferrari, M.D., Jen, J.C. and van den Maagdenberg, A.M. (2009) 'Episodic ataxia associated with EAAT1 mutation C186S affecting glutamate reuptake', *Arch Neurol*, 66(1), pp. 97-101.

Delague, V., Bareil, C., Bouvagnet, P., Salem, N., Chouery, E., Loiselet, J., Megarbane, A. and Claustres, M. (2001) 'Nonprogressive autosomal recessive ataxia maps to chromosome 9q34-9qter in a large consanguineous Lebanese family', *Ann Neurol*, 50(2), pp. 250-3.

Delplanque, J., Devos, D., Huin, V., Genet, A., Sand, O., Moreau, C., Goizet, C., Charles, P., Anheim, M., Monin, M.L., Buee, L., Destee, A., Grolez, G., Delmaire, C., Dujardin, K., Dellacherie, D., Brice, A., Stevanin, G., Strubi-Vuillaume, I., Durr, A. and Sablonniere, B. (2014) 'TMEM240 mutations cause spinocerebellar ataxia 21 with mental retardation and severe cognitive impairment', *Brain*, 137(Pt 10), pp. 2657-63.

Depienne, C., Bugiani, M., Dupuits, C., Galanaud, D., Touitou, V., Postma, N., van Berkel, C., Polder, E., Tollard, E., Darios, F., Brice, A., de Die-Smulders, C.E., Vles, J.S., Vanderver, A., Uziel, G., Yalcinkaya, C., Frints, S.G., Kalscheuer, V.M., Klooster, J., Kamermans, M., Abbink, T.E., Wolf, N.I., Sedel, F. and van der Knaap, M.S. (2013) 'Brain white matter oedema due to ClC-2 chloride channel deficiency: an observational analytical study', *Lancet Neurol*, 12(7), pp. 659-68.

Depondt, C., Donatello, S., Simonis, N., Rai, M., van Heurck, R., Abramowicz, M., D'Hooghe, M. and Pandolfo, M. (2014) 'Autosomal recessive cerebellar ataxia of adult onset due to STUB1 mutations', *Neurology*, 82(19), pp. 1749-50.

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) 'Human Splicing Finder: an online bioinformatics tool to predict splicing signals', *Nucleic Acids Res*, 37(9), p. e67.

Di Gregorio, E., Borroni, B., Giorgio, E., Lacerenza, D., Ferrero, M., Lo Buono, N., Ragusa, N., Mancini, C., Gaussen, M., Calcia, A., Mitro, N., Hoxha, E., Mura, I., Coviello, D.A., Moon, Y.A., Tesson, C., Vaula, G., Couarch, P., Orsi, L., Duregon, E., Papotti, M.G., Deleuze, J.F., Imbert, J., Costanzi, C., Padovani, A., Giunti, P., Maillet-Vioud, M., Durr, A., Brice, A., Tempia, F., Funaro, A., Boccone, L., Caruso, D., Stevanin, G. and Brusco, A. (2014) 'ELOVL5 mutations cause spinocerebellar ataxia 38', *Am J Hum Genet*, 95(2), pp. 209-17.

Dion, P.A., Daoud, H. and Rouleau, G.A. (2009) 'Genetics of motor neuron disorders: new insights into pathogenic mechanisms', *Nat Rev Genet*, 10(11), pp. 769-82.

Dixon-Salazar, T.J., Silhavy, J.L., Udpa, N., Schroth, J., Bielas, S., Schaffer, A.E., Olvera, J., Bafna, V., Zaki, M.S., Abdel-Salam, G.H., Mansour, L.A., Selim, L., Abdel-Hadi, S., Marzouki, N., Ben-Omran, T., Al-Saana, N.A., Sonmez, F.M., Celep, F., Azam, M., Hill, K.J., Collazo, A., Fenstermaker, A.G., Novarino, G., Akizu, N., Garimella, K.V., Sougnez, C., Russ, C., Gabriel, S.B. and Gleeson, J.G. (2012) 'Exome sequencing can improve diagnosis and alter patient management', *Sci Transl Med*, 4(138), p. 138ra78.

Do, R., Kathiresan, S. and Abecasis, G.R. (2012) 'Exome sequencing and complex disease: practical aspects of rare variant association studies', *Hum Mol Genet*, 21(R1), pp. R1-9.

Doi, H., Yoshida, K., Yasuda, T., Fukuda, M., Fukuda, Y., Morita, H., Ikeda, S., Kato, R., Tsurusaki, Y., Miyake, N., Saitsu, H., Sakai, H., Miyatake, S., Shiina, M., Nukina, N., Koyano, S., Tsuji, S., Kuroiwa, Y. and Matsumoto, N. (2011) 'Exome sequencing reveals a homozygous SYT14 mutation in adult-onset, autosomal-recessive spinocerebellar ataxia with psychomotor retardation', *Am J Hum Genet*, 89(2), pp. 320-7.

Dor, T., Cinnamon, Y., Raymond, L., Shaag, A., Bouslam, N., Bouhouche, A., Gaussen, M., Meyer, V., Durr, A., Brice, A., Benomar, A., Stevanin, G., Schuelke, M. and Edvardson, S. (2014) 'KIF1C mutations in two families with hereditary spastic paraparesis and cerebellar dysfunction', *J Med Genet*, 51(2), pp. 137-42.

Duarri, A., Jezierska, J., Fokkens, M., Meijer, M., Schelhaas, H.J., den Dunnen, W.F., van Dijk, F., Verschuuren-Bemelmans, C., Hageman, G., van de Vlies, P., Kusters, B., van de Warrenburg, B.P., Kremer, B., Wijmenga, C., Sinke, R.J., Swertz, M.A., Kampinga, H.H., Boddeke, E. and Verbeek, D.S. (2012) 'Mutations in potassium channel kcnd3 cause spinocerebellar ataxia type 19', *Ann Neurol*, 72(6), pp. 870-80.

Dudding, T.E., Friend, K., Schofield, P.W., Lee, S., Wilkinson, I.A. and Richards, R.I. (2004) 'Autosomal dominant congenital non-progressive ataxia overlaps with the SCA15 locus', *Neurology*, 63(12), pp. 2288-92.

Duenas, A.M., Goold, R. and Giunti, P. (2006) 'Molecular pathogenesis of spinocerebellar ataxias', *Brain*, 129(Pt 6), pp. 1357-70.

Dundar, H., Ozgul, R.K., Yalnizoglu, D., Erdem, S., Oguz, K.K., Tuncel, D., Temucin, C.M. and Dursun, A. (2012) 'Identification of a novel Twinkle mutation in a family with infantile onset spinocerebellar ataxia by whole exome sequencing', *Pediatr Neurol*, 46(3), pp. 172-7.

Durr, A. (2010) 'Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond', *Lancet Neurol*, 9(9), pp. 885-94.

Dusart, I., Guenet, J.L. and Sotelo, C. (2006) 'Purkinje cell death: differences between developmental cell death and neurodegenerative death in mutant mice', *Cerebellum*, 5(2), pp. 163-73.

Dyment, D.A., Cader, M.Z., Chao, M.J., Lincoln, M.R., Morrison, K.M., Disanto, G., Morahan, J.M., De Luca, G.C., Sadovnick, A.D., Lepage, P., Montpetit, A., Ebers, G.C. and Ramagopalan, S.V. (2012) 'Exome sequencing identifies a novel multiple sclerosis susceptibility variant in the TYK2 gene', *Neurology*, 79(5), pp. 406-11.

Edener, U., Bernard, V., Hellenbroich, Y., Gillessen-Kaesbach, G. and Zuhlke, C. (2011) 'Two dominantly inherited ataxias linked to chromosome 16q22.1: SCA4 and SCA31 are not allelic', *J Neurol*, 258(7), pp. 1223-7.

Elsayed, S.M., Heller, R., Thoenes, M., Zaki, M.S., Swan, D., Elsobky, E., Zuhlke, C., Ebermann, I., Nurnberg, G., Nurnberg, P. and Bolz, H.J. (2014) 'Autosomal dominant SCA5 and autosomal recessive infantile SCA are allelic conditions resulting from SPTBN2 mutations', *Eur J Hum Genet*, 22(2), pp. 286-8.

Erro, R., Hersheson, J., Houlden, H. and Bhatia, K.P. (2015) 'A novel TUBB4A mutation suggests that genotype-phenotype correlation of H-ABC syndrome needs to be revisited', *Brain*.

Fairbrother, W.G., Yeo, G.W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.A. and Burge, C.B. (2004) 'RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons', *Nucleic Acids Res*, 32(Web Server issue), pp. W187-90.

Ferreira, C., Poretti, A., Cohen, J., Hamosh, A. and Naidu, S. (2014) 'Novel TUBB4A mutations and expansion of the neuroimaging phenotype of hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC)', *Am J Med Genet A*.

Figueroa, K.P., Minassian, N.A., Stevanin, G., Waters, M., Garibyan, V., Forlani, S., Strzelczyk, A., Burk, K., Brice, A., Durr, A., Papazian, D.M. and Pulst, S.M. (2010) 'KCNC3: phenotype, mutations, channel biophysics-a study of 260 familial ataxia patients', *Hum Mutat*, 31(2), pp. 191-6.

Figueroa, K.P., Waters, M.F., Garibyan, V., Bird, T.D., Gomez, C.M., Ranum, L.P., Minassian, N.A., Papazian, D.M. and Pulst, S.M. (2011) 'Frequency of KCNC3 DNA variants as causes of spinocerebellar ataxia 13 (SCA13)', *PLoS One*, 6(3), p. e17811.

Finsterer, J. (2009) 'Mitochondrial ataxias', *Can J Neurol Sci*, 36(5), pp. 543-53.

Fiskerstrand, T., H'Mida-Ben Brahim, D., Johansson, S., M'Zahem, A., Haukanes, B.I., Drouot, N., Zimmermann, J., Cole, A.J., Vedeler, C., Bredrup, C., Assoum, M., Tazir, M., Klockgether, T., Hamri, A., Steen, V.M., Boman, H., Bindoff, L.A., Koenig, M. and Knappskog, P.M. (2010) 'Mutations in ABHD12 cause the neurodegenerative disease PHARC: An inborn error of endocannabinoid metabolism', *Am J Hum Genet*, 87(3), pp. 410-7.

Flanagan, S.E., Patch, A.M. and Ellard, S. (2010) 'Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations', *Genet Test Mol Biomarkers*, 14(4), pp. 533-7.

Flanigan, K., Gardner, K., Alderson, K., Galster, B., Otterud, B., Leppert, M.F., Kaplan, C. and Ptacek, L.J. (1996) 'Autosomal dominant spinocerebellar ataxia with sensory axonal neuropathy (SCA4): clinical description and genetic localization to chromosome 16q22.1', *Am J Hum Genet*, 59(2), pp. 392-9.

Fogel, B.L., Lee, H., Deignan, J.L., Strom, S.P., Kantarci, S., Wang, X., Quintero-Rivera, F., Vilain, E., Grody, W.W., Perlman, S., Geschwind, D.H. and Nelson, S.F. (2014a) 'Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia', *JAMA Neurol*, 71(10), pp. 1237-46.

Fogel, B.L., Lee, H., Deignan, J.L., Strom, S.P., Kantarci, S., Wang, X., Quintero-Rivera, F., Vilain, E., Grody, W.W., Perlman, S., Geschwind, D.H. and Nelson,

S.F. (2014b) 'Exome Sequencing in the Clinical Diagnosis of Sporadic or Familial Cerebellar Ataxia', *JAMA Neurol*.

Foley, A.R., Menezes, M.P., Pandraud, A., Gonzalez, M.A., Al-Odaib, A., Abrams, A.J., Sugano, K., Yonezawa, A., Manzur, A.Y., Burns, J., Hughes, I., McCullagh, B.G., Jungbluth, H., Lim, M.J., Lin, J.P., Megarbane, A., Urtizberea, J.A., Shah, A.H., Antony, J., Webster, R., Broomfield, A., Ng, J., Mathew, A.A., O'Byrne, J.J., Forman, E., Scoto, M., Prasad, M., O'Brien, K., Olpin, S., Oppenheim, M., Hargreaves, I., Land, J.M., Wang, M.X., Carpenter, K., Horvath, R., Straub, V., Lek, M., Gold, W., Farrell, M.O., Brandner, S., Phadke, R., Matsubara, K., McGarvey, M.L., Scherer, S.S., Baxter, P.S., King, M.D., Clayton, P., Rahman, S., Reilly, M.M., Ouvrier, R.A., Christodoulou, J., Zuchner, S., Muntoni, F. and Houlden, H. (2014) 'Treatable childhood neuronopathy caused by mutations in riboflavin transporter RFVT2', *Brain*, 137(Pt 1), pp. 44-56.

Foo, J.N., Liu, J. and Tan, E.K. (2013) 'Next-generation sequencing diagnostics for neurological diseases/disorders: from a clinical perspective', *Hum Genet*, 132(7), pp. 721-34.

Foo, J.N., Liu, J.J. and Tan, E.K. (2012) 'Whole-genome and whole-exome sequencing in neurological diseases', *Nat Rev Neurol*, 8(9), pp. 508-17.

Frischmeyer, P.A. and Dietz, H.C. (1999) 'Nonsense-mediated mRNA decay in health and disease', *Hum Mol Genet*, 8(10), pp. 1893-900.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., Nickerson, D.A., Bamshad, M.J. and Akey, J.M. (2013) 'Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants', *Nature*, 493(7431), pp. 216-20.

Furrer, S.A., Waldherr, S.M., Mohanachandran, M.S., Baughn, T.D., Nguyen, K.T., Sopher, B.L., Damian, V.A., Garden, G.A. and La Spada, A.R. (2013) 'Reduction of mutant ataxin-7 expression restores motor function and prevents cerebellar synaptic reorganization in a conditional mouse model of SCA7', *Human Molecular Genetics*, 22(5), pp. 890-903.

Galassi, A. (1992) 'The next generation: new chemotherapy agents for the 1990s', *Semin Oncol Nurs*, 8(2), pp. 83-94.

Garbern, J.Y., Neumann, M., Trojanowski, J.Q., Lee, V.M., Feldman, G., Norris, J.W., Friez, M.J., Schwartz, C.E., Stevenson, R. and Sima, A.A. (2010) 'A mutation affecting the sodium/proton exchanger, SLC9A6, causes mental retardation with tau deposition', *Brain*, 133(Pt 5), pp. 1391-402.

Garden, G. (1993-) 'Spinocerebellar Ataxia Type 7', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301433.

Gardner, L.B. (2010) 'Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis', *Mol Cancer Res*, 8(3), pp. 295-308.

Gasser, T., Finsterer, J., Baets, J., Van Broeckhoven, C., Di Donato, S., Fontaine, B., De Jonghe, P., Lossos, A., Lynch, T., Mariotti, C., Schols, L., Spinazzola, A., Szolnoki, Z., Tabrizi, S.J., Tallaksen, C.M., Zeviani, M., Burgunder, J.M. and Harbo, H.F. (2010) 'EFNS guidelines on the molecular diagnosis of ataxias and spastic paraplegias', *Eur J Neurol*, 17(2), pp. 179-88.

Gatti, R. (1993-) 'Ataxia-Telangiectasia', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301790.

Ghezzi, D., Saada, A., D'Adamo, P., Fernandez-Vizarra, E., Gasparini, P., Tiranti, V., Elpeleg, O. and Zeviani, M. (2008) 'FASTKD2 nonsense mutation in an infantile mitochondrial encephalomyopathy associated with cytochrome c oxidase deficiency', *Am J Hum Genet*, 83(3), pp. 415-23.

Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., Thibodeau, P., Bachand, I., Bao, J.Y., Tong, A.H., Lin, C.H., Millet, B., Jaafari, N., Joober, R., Dion, P.A., Lok, S., Krebs, M.O. and Rouleau, G.A. (2011) 'Increased exonic de novo mutation rate in individuals with schizophrenia', *Nat Genet*, 43(9), pp. 860-3.

Glusman, G., Cox, H.C. and Roach, J.C. (2014) 'Whole-genome haplotyping approaches and genomic medicine', *Genome Med*, 6(9), p. 73.

Goizet, C., Boukhris, A., Maltete, D., Guyant-Marechal, L., Truchetto, J., Mundwiller, E., Hanein, S., Jonveaux, P., Roelens, F., Loureiro, J., Godet, E., Forlani, S., Melki, J., Auer-Grumbach, M., Fernandez, J.C., Martin-Hardy, P., Sibon, I., Sole, G., Orignac, I., Mhiri, C., Coutinho, P., Durr, A., Brice, A. and Stevanin, G. (2009) 'SPG15 is the second most common cause of hereditary spastic paraplegia with thin corpus callosum', *Neurology*, 73(14), pp. 1111-9.

Gomez, C.M. (1993-) 'Spinocerebellar Ataxia Type 6', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301319.

Gonzaga-Jauregui, C., Harel, T., Gambin, T., Kousi, M., Griffin, L.B., Francescatto, L., Ozes, B., Karaca, E., Jhangiani, S.N., Bainbridge, M.N., Lawson, K.S., Pehlivan, D., Okamoto, Y., Withers, M., Mancias, P., Slavotinek, A., Reitnauer, P.J., Goksungur, M.T., Shy, M., Crawford, T.O., Koenig, M., Willer, J., Flores, B.N., Pediaditrakis, I., Us, O., Wiszniewski, W., Parman, Y., Antonellis, A., Muzny, D.M., Baylor-Hopkins Center for Mendelian, G., Katsanis, N., Battaloglu, E., Boerwinkle, E., Gibbs, R.A. and Lupski, J.R. (2015) 'Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy', *Cell Rep*.

Gonzalez-Perez, P., Cirulli, E.T., Drory, V.E., Dabby, R., Nisipeanu, P., Carasso, R.L., Sadeh, M., Fox, A., Festoff, B.W., Sapp, P.C., McKenna-Yasek, D., Goldstein, D.B., Brown, R.H., Jr. and Blumen, S.C. (2012) 'Novel mutation in VCP gene causes atypical amyotrophic lateral sclerosis', *Neurology*, 79(22), pp. 2201-8.

Granstrom, A.L., Markljung, E., Fink, K., Nordenskjold, E., Nilsson, D., Wester, T. and Nordenskjold, A. (2014) 'A novel stop mutation in the EDNRB gene in a family with Hirschsprung's disease associated with multiple sclerosis', *J Pediatr Surg*, 49(4), pp. 622-5.

Gros-Louis, F., Dupre, N., Dion, P., Fox, M.A., Laurent, S., Verreault, S., Sanes, J.R., Bouchard, J.P. and Rouleau, G.A. (2007) 'Mutations in SYNE1 lead to a newly

discovered form of autosomal recessive cerebellar ataxia', *Nat Genet*, 39(1), pp. 80-5.

Guan, D., Tkatch, T., Surmeier, D.J., Armstrong, W.E. and Foehring, R.C. (2007) 'Kv2 subunits underlie slowly inactivating potassium current in rat neocortical pyramidal neurons', *J Physiol*, 581(Pt 3), pp. 941-60.

Guergueltcheva, V., Azmanov, D.N., Angelicheva, D., Smith, K.R., Chamova, T., Florez, L., Bynevelt, M., Nguyen, T., Cherninkova, S., Bojinova, V., Kaprelyan, A., Angelova, L., Morar, B., Chandler, D., Kaneva, R., Bahlo, M., Tournev, I. and Kalaydjieva, L. (2012) 'Autosomal-recessive congenital cerebellar ataxia is caused by mutations in metabotropic glutamate receptor 1', *Am J Hum Genet*, 91(3), pp. 553-64.

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S., Younkin, S., Hazrati, L., Collinge, J., Pocock, J., Lashley, T., Williams, J., Lambert, J.C., Amouyel, P., Goate, A., Rademakers, R., Morgan, K., Powell, J., St George-Hyslop, P., Singleton, A., Hardy, J. and Alzheimer Genetic Analysis, G. (2013) 'TREM2 variants in Alzheimer's disease', *N Engl J Med*, 368(2), pp. 117-27.

Guerreiro, R.J., Lohmann, E., Kinsella, E., Bras, J.M., Luu, N., Gurunlian, N., Dursun, B., Bilgic, B., Santana, I., Hanagasi, H., Gurvit, H., Gibbs, J.R., Oliveira, C., Emre, M. and Singleton, A. (2012) 'Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease', *Neurobiol Aging*, 33(5), pp. 1008 e17-23.

Guissart, C., Li, X., Leheup, B., Drouot, N., Montaut-Verient, B., Raffo, E., Jonveaux, P., Roux, A.F., Claustres, M., Fliegel, L. and Koenig, M. (2015) 'Mutation of SLC9A1, encoding the major Na(+)/H(+) exchanger, causes ataxia-deafness Lichtenstein-Knorr syndrome', *Hum Mol Genet*, 24(2), pp. 463-70.

Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B. and Hagerman, P.J. (2001) 'Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X', *Neurology*, 57(1), pp. 127-30.

Hallmann, K., Zsurka, G., Moskau-Hartmann, S., Kirschner, J., Korinthenberg, R., Ruppert, A.K., Ozdemir, O., Weber, Y., Becker, F., Lerche, H., Elger, C.E., Thiele, H., Nurnberg, P., Sander, T. and Kunz, W.S. (2014) 'A homozygous splice-site mutation in CARS2 is associated with progressive myoclonic epilepsy', *Neurology*, 83(23), pp. 2183-7.

Hamilton, E.M., Polder, E., Vanderver, A., Naidu, S., Schiffmann, R., Fisher, K., Raguz, A.B., Blumkin, L., van Berkel, C.G., Waisfisz, Q., Simons, C., Taft, R.J., Abbink, T.E., Wolf, N.I. and van der Knaap, M.S. (2014) 'Hypomyelination with atrophy of the basal ganglia and cerebellum: further delineation of the phenotype and genotype-phenotype correlation', *Brain*, 137(Pt 7), pp. 1921-30.

Hammer, M.B., Eleuch-Fayache, G., Gibbs, J.R., Arepalli, S.K., Chong, S.B., Sassi, C., Bouhlal, Y., Hentati, F., Amouri, R. and Singleton, A.B. (2012) 'Exome sequencing: an efficient diagnostic tool for complex neurodegenerative disorders', *Eur J Neurol*.

Hanein, S., Martin, E., Boukhris, A., Byrne, P., Goizet, C., Hamri, A., Benomar, A., Lossos, A., Denora, P., Fernandez, J., Elleuch, N., Forlani, S., Durr, A., Feki, I., Hutchinson, M., Santorelli, F.M., Mhiri, C., Brice, A. and Stevanin, G. (2008) 'Identification of the SPG15 gene, encoding spastizin, as a frequent cause of complicated autosomal-recessive spastic paraplegia, including Kjellin syndrome', *Am J Hum Genet*, 82(4), pp. 992-1002.

Harismendy, O., Bansal, V., Bhatia, G., Nakano, M., Scott, M., Wang, X., Dib, C., Turlotte, E., Sipe, J.C., Murray, S.S., Deleuze, J.F., Bafna, V., Topol, E.J. and Frazer, K.A. (2010) 'Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level', *Genome Biol*, 11(11), p. R118.

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A.R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Morgan, K.,

Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Love, S., Kehoe, P.G., Hardy, J., Mead, S., Fox, N., Rossor, M., Collinge, J., Maier, W., Jessen, F., Schurmann, B., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frolich, L., Hampel, H., Hull, M., Rujescu, D., Goate, A.M., Kauwe, J.S., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Muhleisen, T.W., Nothen, M.M., Moebus, S., Jockel, K.H., Klopp, N., Wichmann, H.E., Carrasquillo, M.M., Pankratz, V.S., Younkin, S.G., Holmans, P.A., O'Donovan, M., Owen, M.J. and Williams, J. (2009) 'Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease', *Nat Genet*, 41(10), pp. 1088-93.

Hashimoto, Y., Shirane, M., Matsuzaki, F., Saita, S., Ohnishi, T. and Nakayama, K.I. (2014) 'Protrudin regulates endoplasmic reticulum morphology and function associated with the pathogenesis of hereditary spastic paraplegia', *J Biol Chem*, 289(19), pp. 12946-61.

Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P., Brottier, P., Cattolico, L., Barbe, V., Burgunder, J.M., Prud'homme, J.F., Brice, A., Fontaine, B., Heilig, B. and Weissenbach, J. (1999) 'Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia', *Nat Genet*, 23(3), pp. 296-303.

Heimdal, K., Sanchez-Guixe, M., Aukrust, I., Bollerslev, J., Bruland, O., Jablonski, G.E., Erichsen, A.K., Gude, E., Koht, J.A., Erdal, S., Fiskerstrand, T., Haukanes, B.I., Boman, H., Bjorkhaug, L., Tallaksen, C.M., Knappskog, P.M. and Johansson, S. (2014) 'STUB1 mutations in autosomal recessive ataxias - evidence for mutation-specific clinical heterogeneity', *Orphanet J Rare Dis*, 9, p. 146.

Hekman, K.E. and Gomez, C.M. (2015) 'The autosomal dominant spinocerebellar ataxias: emerging mechanistic themes suggest pervasive Purkinje cell vulnerability', *J Neurol Neurosurg Psychiatry*, 86(5), pp. 554-61.

Hekman, K.E., Yu, G.Y., Brown, C.D., Zhu, H., Du, X., Gervin, K., Undlien, D.E., Peterson, A., Stevanin, G., Clark, H.B., Pulst, S.M., Bird, T.D., White, K.P. and Gomez, C.M. (2012) 'A conserved eEF2 coding variant in SCA26 leads to loss of translational fidelity and increased susceptibility to proteostatic insult', *Hum Mol Genet*, 21(26), pp. 5472-83.

Hellenbroich, Y., Bubel, S., Pawlack, H., Opitz, S., Vieregge, P., Schwinger, E. and Zuhlke, C. (2003) 'Refinement of the spinocerebellar ataxia type 4 locus in a large German family and exclusion of CAG repeat expansions in this region', *J Neurol*, 250(6), pp. 668-71.

Henke, W., Herdel, K., Jung, K., Schnorr, D. and Loening, S.A. (1997) 'Betaine improves the PCR amplification of GC-rich DNA sequences', *Nucleic Acids Res*, 25(19), pp. 3957-8.

Heron, S.E., Smith, K.R., Bahlo, M., Nobili, L., Kahana, E., Licchetta, L., Oliver, K.L., Mazarib, A., Afawi, Z., Korczyn, A., Plazzi, G., Petrou, S., Berkovic, S.F., Scheffer, I.E. and Dibbens, L.M. (2012) 'Missense mutations in the sodium-gated potassium channel gene KCNT1 cause severe autosomal dominant nocturnal frontal lobe epilepsy', *Nat Genet*, 44(11), pp. 1188-90.

Hersheson, J., Haworth, A. and Houlden, H. (2012a) 'The inherited ataxias: genetic heterogeneity, mutation databases, and future directions in research and clinical diagnostics', *Hum Mutat*, 33(9), pp. 1324-32.

Hersheson, J., Mencacci, N.E., Davis, M., Macdonald, N., Trabzuni, D., Ryten, M., Pittman, A., Paudel, R., Kara, E., Fawcett, K., Plagnol, V., Bhatia, K.P., Medlar, A.J., Stanescu, H.C., Hardy, J., Kleta, R., Wood, N.W. and Houlden, H. (2012b) 'Mutations in the autoregulatory domain of beta-tubulin 4a cause hereditary dystonia', *Ann Neurol*.

Hills, L.B., Masri, A., Konno, K., Kakegawa, W., Lam, A.T., Lim-Melia, E., Chandy, N., Hill, R.S., Partlow, J.N., Al-Saffar, M., Nasir, R., Stoler, J.M., Barkovich, A.J., Watanabe, M., Yuzaki, M. and Mochida, G.H. (2013) 'Deletions in GRID2 lead to a recessive syndrome of cerebellar ataxia and tonic upgaze in humans', *Neurology*, 81(16), pp. 1378-86.

Hogewind, B.F., Pennings, R.J., Hol, F.A., Kunst, H.P., Hoefsloot, E.H., Cruysberg, J.R. and Cremers, C.W. (2010) 'Autosomal dominant optic neuropathy and sensorineual hearing loss associated with a novel mutation of WFS1', *Mol Vis*, 16, pp. 26-35.

Hoischen, A., Gilissen, C., Arts, P., Wieskamp, N., van der Vliet, W., Vermeer, S., Steehouwer, M., de Vries, P., Meijer, R., Seiqueros, J., Knoers, N.V., Buckley, M.F., Scheffer, H. and Veltman, J.A. (2010) 'Massively parallel sequencing of ataxia genes after array-based enrichment', *Hum Mutat*, 31(4), pp. 494-9.

Hol, E.M. and Pekny, M. (2015) 'Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system', *Curr Opin Cell Biol*, 32, pp. 121-30.

Hossjer, O. (2005) 'Information and effective number of meioses in linkage analysis', *J Math Biol*, 50(2), pp. 208-32.

Houlden, H. (1993-) 'Spinocerebellar Ataxia Type 11', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301723.

Huang, L., Chardon, J.W., Carter, M.T., Friend, K.L., Dudding, T.E., Schwartzentruber, J., Zou, R., Schofield, P.W., Douglas, S., Bulman, D.E. and Boycott, K.M. (2012) 'Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia', *Orphanet J Rare Dis*, 7, p. 67.

Hudder, A. and Werner, R. (2000) 'Analysis of a Charcot-Marie-Tooth disease mutation reveals an essential internal ribosome entry site element in the connexin-32 gene', *J Biol Chem*, 275(44), pp. 34586-91.

Ikeda, Y., Dalton, J.C., Day, J.W. and Ranum, L.P.W. (1993-) 'Spinocerebellar Ataxia Type 8', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301445.

Ikeda, Y., Dick, K.A., Weatherspoon, M.R., Gincel, D., Armbrust, K.R., Dalton, J.C., Stevanin, G., Durr, A., Zuhlke, C., Burk, K., Clark, H.B., Brice, A., Rothstein, J.D., Schut, L.J., Day, J.W. and Ranum, L.P. (2006) 'Spectrin mutations cause spinocerebellar ataxia type 5', *Nat Genet*, 38(2), pp. 184-90.

Ishiura, H., Fukuda, Y., Mitsui, J., Nakahara, Y., Ahsan, B., Takahashi, Y., Ichikawa, Y., Goto, J., Sakai, T. and Tsuji, S. (2011) 'Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1', *Neurogenetics*, 12(2), pp. 117-21.

Jaglin, X.H., Poirier, K., Saillour, Y., Buhler, E., Tian, G., Bahi-Buisson, N., Fallet-Bianco, C., Phan-Dinh-Tuy, F., Kong, X.P., Bomont, P., Castelnau-Ptakhine, L., Odent, S., Loget, P., Kossorotoff, M., Snoeck, I., Plessis, G., Parent, P., Beldjord, C., Cardoso, C., Represa, A., Flint, J., Keays, D.A., Cowan, N.J. and Chelly, J. (2009) 'Mutations in the beta-tubulin gene TUBB2B result in asymmetrical polymicrogyria', *Nat Genet*, 41(6), pp. 746-52.

Jayadev, S. and Bird, T.D. (2013) 'Hereditary ataxias: overview', *Genet Med*, 15(9), pp. 673-83.

Jen, J.C., Graves, T.D., Hess, E.J., Hanna, M.G., Griggs, R.C. and Baloh, R.W. (2007) 'Primary episodic ataxias: diagnosis, pathogenesis and treatment', *Brain*, 130(Pt 10), pp. 2484-93.

Jensen, M.O., Jogini, V., Borhani, D.W., Leffler, A.E., Dror, R.O. and Shaw, D.E. (2012) 'Mechanism of voltage gating in potassium channels', *Science*, 336(6078), pp. 229-33.

Ji, W., Foo, J.N., O'Roak, B.J., Zhao, H., Larson, M.G., Simon, D.B., Newton-Cheh, C., State, M.W., Levy, D. and Lifton, R.P. (2008) 'Rare independent mutations in

renal salt handling genes contribute to blood pressure variation', *Nat Genet*, 40(5), pp. 592-9.

Jiang, T., Tan, M.S., Tan, L. and Yu, J.T. (2014) 'Application of next-generation sequencing technologies in Neurology', *Ann Transl Med*, 2(12), p. 125.

Johnson, J.O., Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., Gibbs, J.R., Nalls, M.A., Morgan, S., Shoai, M., Hardy, J., Pittman, A., Orrell, R.W., Malaspina, A., Sidle, K.C., Fratta, P., Harms, M.B., Baloh, R.H., Pestronk, A., Weihl, C.C., Rogaeva, E., Zinman, L., Drory, V.E., Borghero, G., Mora, G., Calvo, A., Rothstein, J.D., Consortium, I., Drepper, C., Sendtner, M., Singleton, A.B., Taylor, J.P., Cookson, M.R., Restagno, G., Sabatelli, M., Bowser, R., Chio, A. and Traynor, B.J. (2014) 'Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis', *Nat Neurosci*, 17(5), pp. 664-6.

Jordan, M.A. and Wilson, L. (2004) 'Microtubules as a target for anticancer drugs', *Nat Rev Cancer*, 4(4), pp. 253-65.

Kancheva, D., Chamova, T., Guergueltcheva, V., Mitev, V., Azmanov, D.N., Kalaydjieva, L., Tournev, I. and Jordanova, A. (2015) 'Mosaic dominant TUBB4A mutation in an inbred family with complicated hereditary spastic paraplegia', *Mov Disord*.

Karkheiran, S., Krebs, C.E., Makarov, V., Nilipour, Y., Hubert, B., Darvish, H., Frucht, S., Shahidi, G.A., Buxbaum, J.D. and Paisan-Ruiz, C. (2013) 'Identification of COL6A2 mutations in progressive myoclonus epilepsy syndrome', *Hum Genet*, 132(3), pp. 275-83.

Keogh, M.J. and Chinnery, P.F. (2013) 'Next generation sequencing for neurological diseases: new hope or new hype?', *Clin Neurol Neurosurg*, 115(7), pp. 948-53.

Kerber, K.A., Jen, J.C., Lee, H., Nelson, S.F. and Baloh, R.W. (2007) 'A new episodic ataxia syndrome with linkage to chromosome 19q13', *Arch Neurol*, 64(5), pp. 749-52.

Khundadze, M., Kollmann, K., Koch, N., Biskup, C., Nietzsche, S., Zimmer, G., Hennings, J.C., Huebner, A.K., Symmank, J., Jahic, A., Ilina, E.I., Karle, K., Schols, L., Kessels, M., Braulke, T., Qualmann, B., Kurth, I., Beetz, C. and Hubner, C.A. (2013) 'A hereditary spastic paraplegia mouse model supports a role of ZFYVE26/SPASTIZIN for the endolysosomal system', *PLoS Genet*, 9(12), p. e1003988.

Kim, D.S., Crosslin, D.R., Auer, P.L., Suzuki, S.M., Marsillach, J., Burt, A.A., Gordon, A.S., Meschia, J.F., Nalls, M.A., Worrall, B.B., Longstreth, W.T., Jr., Gottesman, R.F., Furlong, C.E., Peters, U., Rich, S.S., Nickerson, D.A., Jarvik, G.P. and on behalf of the, N.E.S.P. (2014) 'Rare coding variation in paraoxonase-1 is associated with ischemic stroke in the NHLBI Exome Sequencing Project', *J Lipid Res*, 55(6), pp. 1173-1178.

Kim, K.S., Kubota, S., Kuriyama, M., Fujiyama, J., Bjorkhem, I., Eggertsen, G. and Seyama, Y. (1994) 'Identification of new mutations in sterol 27-hydroxylase gene in Japanese patients with cerebrotendinous xanthomatosis (CTX)', *J Lipid Res*, 35(6), pp. 1031-9.

Kingsmore, S.F. and Saunders, C.J. (2011) 'Deep sequencing of patient genomes for disease diagnosis: when will it become routine?', *Sci Transl Med*, 3(87), p. 87ps23.

Kirschner, M. and Mitchison, T. (1986) 'Beyond self-assembly: from microtubules to morphogenesis', *Cell*, 45(3), pp. 329-42.

Klein, C.J., Bird, T., Ertekin-Taner, N., Lincoln, S., Hjorth, R., Wu, Y., Kwok, J., Mer, G., Dyck, P.J. and Nicholson, G.A. (2013) 'DNMT1 mutation hot spot causes varied phenotypes of HSAN1 with dementia and hearing loss', *Neurology*, 80(9), pp. 824-8.

Klein, C.J., Middha, S., Duan, X., Wu, Y., Litchy, W.J., Gu, W., Dyck, P.J., Gavrilova, R.H., Smith, D.I. and Kocher, J.P. (2014) 'Application of whole exome sequencing in undiagnosed inherited polyneuropathies', *J Neurol Neurosurg Psychiatry*.

Klepper, J. (2015) 'GLUT1 deficiency syndrome and ketogenic diet therapies: missing rare but treatable diseases?', *Dev Med Child Neurol*, 57(10), pp. 896-7.

Knudsen, B. and Miyamoto, M.M. (2001) 'A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins', *Proc Natl Acad Sci U S A*, 98(25), pp. 14512-7.

Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., Habu, T., Liu, W., Okuda, H. and Koizumi, A. (2011) 'Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement', *Am J Hum Genet*, 89(1), pp. 121-30.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K. and Ding, L. (2009) 'VarScan: variant detection in massively parallel sequencing of individual and pooled samples', *Bioinformatics*, 25(17), pp. 2283-5.

Koroglu, C., Baysal, L., Cetinkaya, M., Karasoy, H. and Tolun, A. (2013) 'DNAJC6 is responsible for juvenile parkinsonism with phenotypic variability', *Parkinsonism Relat Disord*, 19(3), pp. 320-4.

Kryukov, G.V., Pennacchio, L.A. and Sunyaev, S.R. (2007) 'Most rare missense alleles are deleterious in humans: implications for complex disease and association studies', *Am J Hum Genet*, 80(4), pp. 727-39.

Ku, C.S., Cooper, D.N., Polychronakos, C., Naidoo, N., Wu, M. and Soong, R. (2012) 'Exome sequencing: dual role as a discovery and diagnostic tool', *Ann Neurol*, 71(1), pp. 5-14.

Ku, C.S., Naidoo, N. and Pawitan, Y. (2011) 'Revisiting Mendelian disorders through exome sequencing', *Hum Genet*, 129(4), pp. 351-70.

Kuijpers, M. and Hoogenraad, C.C. (2011) 'Centrosomes, microtubules and neuronal development', *Mol Cell Neurosci*, 48(4), pp. 349-58.

Kullmann, D.M. and Waxman, S.G. (2010) 'Neurological channelopathies: new insights into disease mechanisms and ion channel function', *J Physiol*, 588(Pt 11), pp. 1823-7.

Kumar, P., Henikoff, S. and Ng, P.C. (2009) 'Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm', *Nat Protoc*, 4(7), pp. 1073-81.

Kyuhou, S., Kato, N. and Gemba, H. (2006) 'Emergence of endoplasmic reticulum stress and activated microglia in Purkinje cell degeneration mice', *Neurosci Lett*, 396(2), pp. 91-6.

Lafreniere, R.G., MacDonald, M.L., Dube, M.P., MacFarlane, J., O'Driscoll, M., Brais, B., Meilleur, S., Brinkman, R.R., Dadivas, O., Pape, T., Platon, C., Radomski, C., Risler, J., Thompson, J., Guerra-Escobio, A.M., Davar, G., Breakefield, X.O., Pimstone, S.N., Green, R., Pryse-Phillips, W., Goldberg, Y.P., Younghusband, H.B., Hayden, M.R., Sherrington, R., Rouleau, G.A. and Samuels, M.E. (2004) 'Identification of a novel gene (HSN2) causing hereditary sensory and autonomic neuropathy type II through the Study of Canadian Genetic Isolates', *Am J Hum Genet*, 74(5), pp. 1064-73.

Lagier-Tourenne, C., Tazir, M., Lopez, L.C., Quinzii, C.M., Assoum, M., Drouot, N., Busso, C., Makri, S., Ali-Pacha, L., Benhassine, T., Anheim, M., Lynch, D.R., Thibault, C., Plewniak, F., Bianchetti, L., Tranchant, C., Poch, O., DiMauro, S., Mandel, J.L., Barros, M.H., Hirano, M. and Koenig, M. (2008) 'ADCK3, an ancestral kinase, is mutated in a form of recessive ataxia associated with coenzyme Q10 deficiency', *Am J Hum Genet*, 82(3), pp. 661-72.

Lander, E.S. and Botstein, D. (1987) 'Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children', *Science*, 236(4808), pp. 1567-70.

Leandro-Garcia, L.J., Leskela, S., Landa, I., Montero-Conde, C., Lopez-Jimenez, E., Leton, R., Cascon, A., Robledo, M. and Rodriguez-Antona, C. (2010) 'Tumoral and tissue-specific expression of the major human beta-tubulin isotypes', *Cytoskeleton (Hoboken)*, 67(4), pp. 214-23.

Lee, Y.C., Durr, A., Majczenko, K., Huang, Y.H., Liu, Y.C., Lien, C.C., Tsai, P.C., Ichikawa, Y., Goto, J., Monin, M.L., Li, J.Z., Chung, M.Y., Mundwiller, E., Shakkottai, V., Liu, T.T., Tesson, C., Lu, Y.C., Brice, A., Tsuji, S., Burmeister, M., Stevanin, G. and Soong, B.W. (2012) 'Mutations in KCND3 cause spinocerebellar ataxia type 22', *Ann Neurol*, 72(6), pp. 859-69.

Leehey, M.A. (2009) 'Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment', *J Investig Med*, 57(8), pp. 830-6.

Lemke, J.R., Riesch, E., Scheurenbrand, T., Schubach, M., Wilhelm, C., Steiner, I., Hansen, J., Courage, C., Gallati, S., Burki, S., Strozzi, S., Simonetti, B.G., Grunt, S., Steinlin, M., Alber, M., Wolff, M., Klopstock, T., Prott, E.C., Lorenz, R., Spaich, C., Rona, S., Lakshminarasimhan, M., Kroll, J., Dorn, T., Kramer, G., Synofzik, M., Becker, F., Weber, Y.G., Lerche, H., Bohm, D. and Biskup, S. (2012) 'Targeted next generation sequencing as a diagnostic tool in epileptic disorders', *Epilepsia*, 53(8), pp. 1387-98.

Lendon, C.L., Lynch, T., Norton, J., McKeel, D.W., Jr., Busfield, F., Craddock, N., Chakraverty, S., Gopalakrishnan, G., Shears, S.D., Grimmett, W., Wilhelmsen, K.C., Hansen, L., Morris, J.C. and Goate, A.M. (1998) 'Hereditary dysphasic disinhibition dementia: a frontotemporal dementia linked to 17q21-22', *Neurology*, 50(6), pp. 1546-55.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*, 26(5), pp. 589-95.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078-9.

Li, M., Pang, S.Y., Song, Y., Kung, M.H., Ho, S.L. and Sham, P.C. (2013a) 'Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family', *Clin Genet*, 83(3), pp. 269-73.

Li, Z., Fang, F. and Xu, F. (2013b) 'Effects of different states of oxidative stress on fetal rat alveolar type II epithelial cells in vitro and ROSinduced changes in Wnt signaling pathway expression', *Mol Med Rep*, 7(5), pp. 1528-32.

Lin, C.Y., Louis, E.D., Faust, P.L., Koeppen, A.H., Vonsattel, J.P. and Kuo, S.H. (2014) 'Abnormal climbing fibre-Purkinje cell synaptic connections in the essential tremor cerebellum', *Brain*, 137(Pt 12), pp. 3149-59.

Linde, L., Boelz, S., Neu-Yilik, G., Kulozik, A.E. and Kerem, B. (2007) 'The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells', *Eur J Hum Genet*, 15(11), pp. 1156-62.

Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Schnekenberg, R.P., Suminaite, D., Hope, J., Baker, I., Gregory, L., Green, A., Allan, C., Lamble, S., Jayawant, S., Quaghebeur, G., Cader, M.Z., Hughes, S., Armstrong, R.J., Kanapin, A., Rimmer, A., Lunter, G., Mathieson, I., Cazier, J.B., Buck, D., Taylor, J.C., Bentley, D., McVean, G., Donnelly, P., Knight, S.J., Jackson, M., Ragoussis, J. and Nemeth, A.H. (2012) 'Recessive mutations in SPTBN2 implicate beta-III spectrin in both cognitive and motor development', *PLoS Genet*, 8(12), p. e1003074.

Liu, F. and Gong, C.X. (2008) 'Tau exon 10 alternative splicing and tauopathies', *Mol Neurodegener*, 3, p. 8.

Lohmann, K., Wilcox, R.A., Winkler, S., Ramirez, A., Rakovic, A., Park, J.S., Arns, B., Lohnau, T., Groen, J., Kasten, M., Bruggemann, N., Hagenah, J., Schmidt, A., Kaiser, F.J., Kumar, K.R., Zschiedrich, K., Alvarez-Fischer, D., Altenmuller, E., Ferbert, A., Lang, A.E., Munchau, A., Kostic, V., Simonyan, K., Agzarian, M., Ozelius, L.J., Langeveld, A.P., Sue, C.M., Tijssen, M.A. and Klein, C. (2012) 'Whispering dysphonia (DYT4 dystonia) is caused by a mutation in the TUBB4 gene', *Ann Neurol*.

Louis, E.D., Babij, R., Lee, M., Cortes, E. and Vonsattel, J.P. (2013) 'Quantification of cerebellar hemispheric purkinje cell linear density: 32 ET cases versus 16 controls', *Mov Disord*, 28(13), pp. 1854-9.

Lowe, J., Li, H., Downing, K.H. and Nogales, E. (2001) 'Refined structure of alpha beta-tubulin at 3.5 A resolution', *J Mol Biol*, 313(5), pp. 1045-57.

Lupski, J.R., Reid, J.G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D.C., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D.A., McGuire, A.L., Zhang, F., Stankiewicz, P., Halperin, J.J., Yang, C., Gehman, C., Guo, D., Irikat, R.K., Tom, W., Fantin, N.J., Muzny, D.M. and Gibbs, R.A. (2010) 'Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy', *N Engl J Med*, 362(13), pp. 1181-91.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., Barrett, J.C., Biesecker, L.G., Conrad, D.F., Cooper, G.M., Cox, N.J., Daly, M.J., Gerstein, M.B., Goldstein, D.B., Hirschhorn, J.N., Leal, S.M., Pennacchio, L.A., Stamatoyannopoulos, J.A., Sunyaev, S.R., Valle, D., Voight, B.F., Winckler, W. and Gunter, C. (2014) 'Guidelines for investigating causality of sequence variants in human disease', *Nature*, 508(7497), pp. 469-76.

Mallaret, M., Synofzik, M., Lee, J., Sagum, C.A., Mahajnah, M., Sharkia, R., Drouot, N., Renaud, M., Klein, F.A., Anheim, M., Tranchant, C., Mignot, C., Mandel, J.L., Bedford, M., Bauer, P., Salih, M.A., Schule, R., Schols, L., Aldaz, C.M. and Koenig, M. (2014) 'The tumour suppressor gene WWOX is mutated in autosomal recessive cerebellar ataxia with epilepsy and mental retardation', *Brain*, 137(Pt 2), pp. 411-9.

Mancini, C., Nassani, S., Guo, Y., Chen, Y., Giorgio, E., Brussino, A., Di Gregorio, E., Cavalieri, S., Lo Buono, N., Funaro, A., Pizio, N.R., Nmezi, B., Kyttala, A., Santorelli, F.M., Padiath, Q.S., Hakonarson, H., Zhang, H. and Brusco, A. (2015) 'Adult-onset autosomal recessive ataxia associated with neuronal ceroid lipofuscinosis type 5 gene (CLN5) mutations', *J Neurol*, 262(1), pp. 173-8.

Mannan, A.U., Krawen, P., Sauter, S.M., Boehm, J., Chronowska, A., Paulus, W., Neesen, J. and Engel, W. (2006) 'ZFYVE27 (SPG33), a novel spastin-binding protein, is mutated in hereditary spastic paraplegia', *Am J Hum Genet*, 79(2), pp. 351-7.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A. and Visscher, P.M. (2009) 'Finding the missing heritability of complex diseases', *Nature*, 461(7265), pp. 747-53.

Margolis, R.L., O'Hearn, E., Holmes, S.E., Srivastava, A.K., Mukherji, M. and Sinha, K.K. (1993-) 'Spinocerebellar Ataxia Type 12', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301381.

Mariotti, C., Brusco, A., Di Bella, D., Cagnoli, C., Seri, M., Gellera, C., Di Donato, S. and Taroni, F. (2008) 'Spinocerebellar ataxia type 28: a novel autosomal dominant cerebellar ataxia characterized by slow progression and ophthalmoparesis', *Cerebellum*, 7(2), pp. 184-8.

Martignoni, M., Riano, E. and Rugarli, E.I. (2008) 'The role of ZFYVE27/protrudin in hereditary spastic paraplegia', *Am J Hum Genet*, 83(1), pp. 127-8; author reply 128-30.

Martin, E., Yanicostas, C., Rastetter, A., Naini, S.M., Maouedj, A., Kabashi, E., Rivaud-Pechoux, S., Brice, A., Stevanin, G. and Soussi-Yanicostas, N. (2012) 'Spatacsin and spastizin act in the same pathway required for proper spinal motor neuron axon outgrowth in zebrafish', *Neurobiol Dis*, 48(3), pp. 299-308.

Matesanz, F., Gonzalez-Perez, A., Lucas, M., Sanna, S., Gayan, J., Urcelay, E., Zara, I., Pitzalis, M., Cavanillas, M.L., Arroyo, R., Zoledziewska, M., Marrosu, M., Fernandez, O., Leyva, L., Alcina, A., Fedetz, M., Moreno-Rey, C., Velasco, J., Real, L.M., Ruiz-Pena, J.L., Cucca, F., Ruiz, A. and Izquierdo, G. (2012) 'Genome-wide association study of multiple sclerosis confirms a novel locus at 5p13.1', *PLoS One*, 7(5), p. e36140.

Matsuura, T. and Ashizawa, T. (1993-) 'Spinocerebellar Ataxia Type 10', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301354.

Matsuzaki, F., Shirane, M., Matsumoto, M. and Nakayama, K.I. (2011) 'Protrudin serves as an adaptor molecule that connects KIF5 and its cargoes in vesicular transport during process formation', *Mol Biol Cell*, 22(23), pp. 4602-20.

McClellan, J. and King, M.C. (2010) 'Genetic heterogeneity in human disease', *Cell*, 141(2), pp. 210-7.

McDonald, K.K., Stajich, J., Blach, C., Ashley-Koch, A.E. and Hauser, M.A. (2012) 'Exome analysis of two limb-girdle muscular dystrophy families: mutations identified and challenges encountered', *PLoS One*, 7(11), p. e48864.

Meijer, I.A., Hand, C.K., Grewal, K.K., Stefanelli, M.G., Ives, E.J. and Rouleau, G.A. (2002) 'A locus for autosomal dominant hereditary spastic ataxia, SAX1, maps to chromosome 12p13', *Am J Hum Genet*, 70(3), pp. 763-9.

Mencacci, N.E., Rubio-Agusti, I., Zdebik, A., Asmus, F., Ludtmann, M.H., Ryten, M., Plagnol, V., Hauser, A.K., Bandres-Ciga, S., Bettencourt, C., Forabosco, P., Hughes, D., Soutar, M.M., Peall, K., Morris, H.R., Trabzuni, D., Tekman, M., Stanescu, H.C., Kleta, R., Carecchio, M., Zorzi, G., Nardocci, N., Garavaglia, B., Lohmann, E., Weissbach, A., Klein, C., Hardy, J., Pittman, A.M., Foltynie, T., Abramov, A.Y., Gasser, T., Bhatia, K.P. and Wood, N.W. (2015) 'A Missense Mutation in KCTD17 Causes Autosomal Dominant Myoclonus-Dystonia', *Am J Hum Genet*, 96(6), pp. 938-47.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F. and Dietz, H.C. (2004) 'Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise', *Nat Genet*, 36(10), pp. 1073-8.

Metzker, M.L. (2010) 'Sequencing technologies - the next generation', *Nat Rev Genet*, 11(1), pp. 31-46.

Mignarri, A., Cenciarelli, S., Da Pozzo, P., Cardaioli, E., Malandrini, A., Federico, A. and Dotti, M.T. (2015) 'Mitochondrial recessive ataxia syndrome: A neurological rarity not to be missed', *J Neurol Sci*, 349(1-2), pp. 254-5.

Mihailovich, M., Thermann, R., Grohovaz, F., Hentze, M.W. and Zacchetti, D. (2007) 'Complex translational regulation of BACE1 involves upstream AUGs and stimulatory elements within the 5' untranslated region', *Nucleic Acids Res*, 35(9), pp. 2975-85.

Miura, S., Shibata, H., Furuya, H., Ohyagi, Y., Osoegawa, M., Miyoshi, Y., Matsunaga, H., Shibata, A., Matsumoto, N., Iwaki, A., Taniwaki, T., Kikuchi, H., Kira, J. and Fukumaki, Y. (2006) 'The contactin 4 gene locus at 3p26 is a candidate gene of SCA16', *Neurology*, 67(7), pp. 1236-41.

Miyatake, S., Osaka, H., Shiina, M., Sasaki, M., Takanashi, J.I., Haginoya, K., Wada, T., Morimoto, M., Ando, N., Ikuta, Y., Nakashima, M., Tsurusaki, Y., Miyake, N., Ogata, K., Matsumoto, N. and Saitsu, H. (2014) 'Expanding the phenotypic spectrum of TUBB4A-associated hypomyelinating leukoencephalopathies', *Neurology*.

Mollet, J., Delahodde, A., Serre, V., Chretien, D., Schlemmer, D., Lombes, A., Boddaert, N., Desguerre, I., de Lonlay, P., de Baulny, H.O., Munnich, A. and Rotig, A. (2008) 'CABC1 gene mutations cause ubiquinone deficiency with cerebellar ataxia and seizures', *Am J Hum Genet*, 82(3), pp. 623-30.

Moog, U., Kutsche, K., Kortum, F., Chilian, B., Bierhals, T., Apeshiotis, N., Balg, S., Chassaing, N., Coubes, C., Das, S., Engels, H., Van Esch, H., Grasshoff, U., Heise, M., Isidor, B., Jarvis, J., Koehler, U., Martin, T., Oehl-Jaschkowitz, B., Ortibus, E., Pilz, D.T., Prabhakar, P., Rappold, G., Rau, I., Rettenberger, G., Schluter, G., Scott, R.H., Shoukier, M., Wohlleber, E., Zirn, B., Dobyns, W.B. and Uyanik, G. (2011) 'Phenotypic spectrum associated with CASK loss-of-function mutations', *J Med Genet*, 48(11), pp. 741-51.

Moreira, M.C. and Koenig, M. (1993-) 'Ataxia with Oculomotor Apraxia Type 2', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301333.

Musselman, K.E., Stoyanov, C.T., Marasigan, R., Jenkins, M.E., Konczak, J., Morton, S.M. and Bastian, A.J. (2014) 'Prevalence of ataxia in children: a systematic review', *Neurology*, 82(1), pp. 80-9.

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E.L., Campbell, N.G., Geller, E.T., Valladares, O., Schafer, C., Liu, H., Zhao, T., Cai, G., Lihm, J., Dannenfelser, R., Jabado, O., Peralta, Z., Nagaswamy, U., Muzny, D., Reid, J.G., Newsham, I., Wu, Y., Lewis, L., Han, Y., Voight, B.F., Lim, E., Rossin, E., Kirby, A., Flannick, J., Fromer, M., Shakir, K., Fennell, T., Garimella, K., Banks, E., Poplin, R., Gabriel, S., DePristo, M., Wimbish, J.R., Boone, B.E., Levy, S.E., Betancur, C., Sunyaev, S., Boerwinkle, E., Buxbaum, J.D., Cook, E.H., Jr., Devlin, B., Gibbs, R.A., Roeder, K., Schellenberg, G.D., Sutcliffe, J.S. and Daly, M.J. (2012) 'Patterns and rates of exonic de novo mutations in autism spectrum disorders', *Nature*, 485(7397), pp. 242-5.

Nejentsev, S., Walker, N., Riches, D., Egholm, M. and Todd, J.A. (2009) 'Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes', *Science*, 324(5925), pp. 387-9.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M.D., Nangle, K., Wang, J., Abecasis, G., Cardon, L.R., Zollner, S., Whittaker, J.C., Chissoe, S.L., Novembre, J. and Mooser, V. (2012) 'An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people', *Science*, 337(6090), pp. 100-4.

Nemeth, A.H., Kwasniewska, A.C., Lise, S., Parolin Schnekenberg, R., Becker, E.B., Bera, K.D., Shanks, M.E., Gregory, L., Buck, D., Zameel Cader, M., Talbot, K., de Silva, R., Fletcher, N., Hastings, R., Jayawant, S., Morrison, P.J., Worth, P., Taylor, M., Tolmie, J., O'Regan, M., Valentine, R., Packham, E., Evans, J., Seller, A. and Ragoussis, J. (2013) 'Next generation sequencing for molecular

diagnosis of neurological disorders using ataxias as a model', *Brain*, 136(Pt 10), pp. 3106-18.

Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., Lee, C., Turner, E.H., Smith, J.D., Rieder, M.J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D.A., Bamshad, M.J. and Shendure, J. (2010a) 'Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome', *Nat Genet*, 42(9), pp. 790-3.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., Shendure, J. and Bamshad, M.J. (2010b) 'Exome sequencing identifies the cause of a mendelian disorder', *Nat Genet*, 42(1), pp. 30-5.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., Bamshad, M., Nickerson, D.A. and Shendure, J. (2009) 'Targeted capture and massively parallel sequencing of 12 human exomes', *Nature*, 461(7261), pp. 272-6.

Nikali, K. and Lonnqvist, T. (1993-) 'Infantile-Onset Spinocerebellar Ataxia', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301746.

Noensie, E.N. and Dietz, H.C. (2001) 'A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition', *Nat Biotechnol*, 19(5), pp. 434-9.

Novarino, G., Fenstermaker, A.G., Zaki, M.S., Hofree, M., Silhavy, J.L., Heiberg, A.D., Abdellateef, M., Rosti, B., Scott, E., Mansour, L., Masri, A., Kayserili, H., Al-Aama, J.Y., Abdel-Salam, G.M., Karminejad, A., Kara, M., Kara, B., Bozorgmehri, B., Ben-Omran, T., Mojahedi, F., Mahmoud, I.G., Bouslam, N., Bouhouche, A., Benomar, A., Hanein, S., Raymond, L., Forlani, S., Mascaro, M., Selim, L., Shehata, N., Al-Allawi, N., Bindu, P.S., Azam, M., Gunel, M., Caglayan, A., Bilguvar, K., Tolun, A., Issa, M.Y., Schroth, J., Spencer, E.G., Rosti, R.O., Akizu, N., Vaux, K.K., Johansen, A., Koh, A.A., Megahed, H., Durr, A., Brice, A., Stevanin, G., Gabriel, S.B., Ideker, T. and Gleeson, J.G.

(2014) 'Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders', *Science*, 343(6170), pp. 506-11.

Nunnari, J. and Suomalainen, A. (2012) 'Mitochondria: in sickness and in health', *Cell*, 148(6), pp. 1145-59.

Nuytemans, K., Bademci, G., Inchausti, V., Dressen, A., Kinnamon, D.D., Mehta, A., Wang, L., Zuchner, S., Beecham, G.W., Martin, E.R., Scott, W.K. and Vance, J.M. (2013) 'Whole exome sequencing of rare variants in EIF4G1 and VPS35 in Parkinson disease', *Neurology*, 80(11), pp. 982-9.

O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., Rieder, M.J., Nickerson, D.A., Bernier, R., Fisher, S.E., Shendure, J. and Eichler, E.E. (2011) 'Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations', *Nat Genet*, 43(6), pp. 585-9.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., Turner, E.H., Stanaway, I.B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J.M., Borenstein, E., Rieder, M.J., Nickerson, D.A., Bernier, R., Shendure, J. and Eichler, E.E. (2012) 'Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations', *Nature*, 485(7397), pp. 246-50.

Oakley, B.R. (2000) 'An abundance of tubulins', *Trends Cell Biol*, 10(12), pp. 537-42.

Ogawa, T., Takiyama, Y., Sakoe, K., Mori, K., Namekawa, M., Shimazaki, H., Nakano, I. and Nishizawa, M. (2004) 'Identification of a SACS gene missense mutation in ARSACS', *Neurology*, 62(1), pp. 107-9.

Ohba, C., Osaka, H., Iai, M., Yamashita, S., Suzuki, Y., Aida, N., Shimozawa, N., Takamura, A., Doi, H., Tomita-Katsumoto, A., Nishiyama, K., Tsurusaki, Y., Nakashima, M., Miyake, N., Eto, Y., Tanaka, F., Matsumoto, N. and Saitsu, H. (2013) 'Diagnostic utility of whole exome sequencing in patients showing cerebellar and/or vermis atrophy in childhood', *Neurogenetics*, 14(3-4), pp. 225-32.

Ohnishi, T., Shirane, M., Hashimoto, Y., Saita, S. and Nakayama, K.I. (2014) 'Identification and characterization of a neuron-specific isoform of protrudin', *Genes Cells*, 19(2), pp. 97-111.

Onat, O.E., Gulsuner, S., Bilguvar, K., Nazli Basak, A., Topaloglu, H., Tan, M., Tan, U., Gunel, M. and Ozcelik, T. (2013) 'Missense mutation in the ATPase, aminophospholipid transporter protein ATP8A2 is associated with cerebellar atrophy and quadrupedal locomotion', *Eur J Hum Genet*, 21(3), pp. 281-5.

Ozaki, K., Doi, H., Mitsui, J., Sato, N., Iikuni, Y., Majima, T., Yamane, K., Irioka, T., Ishiura, H., Doi, K., Morishita, S., Higashi, M., Sekiguchi, T., Koyama, K., Ueda, N., Miura, Y., Miyatake, S., Matsumoto, N., Yokota, T., Tanaka, F., Tsuji, S., Mizusawa, H. and Ishikawa, K. (2015) 'A Novel Mutation in ELOVL4 Leading to Spinocerebellar Ataxia (SCA) With the Hot Cross Bun Sign but Lacking Erythrokeratodermia: A Broadened Spectrum of SCA34', *JAMA Neurol*.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., Hill, D.E., Vidal, M., Evans, J.G., Thorburn, D.R., Carr, S.A. and Mootha, V.K. (2008) 'A mitochondrial protein compendium elucidates complex I disease biology', *Cell*, 134(1), pp. 112-23.

Paisan-Ruiz, C., Scopes, G., Lee, P. and Houlden, H. (2009) 'Homozygosity mapping through whole genome analysis identifies a COL18A1 mutation in an Indian family presenting with an autosomal recessive neurological disorder', *Am J Med Genet B Neuropsychiatr Genet*, 150B(7), pp. 993-7.

Paulson, H.L. (2009) 'The spinocerebellar ataxias', *J Neuroophthalmol*, 29(3), pp. 227-37.

Pertea, M., Lin, X. and Salzberg, S.L. (2001) 'GeneSplicer: a new computational method for splice site prediction', *Nucleic Acids Res*, 29(5), pp. 1185-90.

Pfeffer, G., Blakely, E.L., Alston, C.L., Hassani, A., Boggild, M., Horvath, R., Samuels, D.C., Taylor, R.W. and Chinnery, P.F. (2012) 'Adult-onset spinocerebellar

ataxia syndromes due to MTATP6 mutations', *J Neurol Neurosurg Psychiatry*, 83(9), pp. 883-6.

Pfeffer, G., Pyle, A., Griffin, H., Miller, J., Wilson, V., Turnbull, L., Fawcett, K., Sims, D., Eglon, G., Hadjivassiliou, M., Horvath, R., Nemeth, A. and Chinnery, P.F. (2015) 'SPG7 mutations are a common cause of undiagnosed ataxia', *Neurology*, 84(11), pp. 1174-6.

Pierson, T.M., Adams, D., Bonn, F., Martinelli, P., Cherukuri, P.F., Teer, J.K., Hansen, N.F., Cruz, P., Mullikin For The Nisc Comparative Sequencing Program, J.C., Blakesley, R.W., Golas, G., Kwan, J., Sandler, A., Fuentes Fajardo, K., Markello, T., Tifft, C., Blackstone, C., Rugarli, E.I., Langer, T., Gahl, W.A. and Toro, C. (2011) 'Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases', *PLoS Genet*, 7(10), p. e1002325.

Pihlstrom, L., Axelsson, G., Bjornara, K.A., Dizdar, N., Fardell, C., Forsgren, L., Holmberg, B., Larsen, J.P., Linder, J., Nissbrandt, H., Tysnes, O.B., Ohman, E., Dietrichs, E. and Toft, M. (2012) 'Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease', *Neurobiol Aging*.

Pizzino, A., Pierson, T.M., Guo, Y., Helman, G., Fortini, S., Guerrero, K., Saitta, S., Murphy, J.L., Padiath, Q., Xie, Y., Hakonarson, H., Xu, X., Funari, T., Fox, M., Taft, R.J., van der Knaap, M.S., Bernard, G., Schiffmann, R., Simons, C. and Vanderver, A. (2014) 'TUBB4A de novo mutations cause isolated hypomyelination', *Neurology*, 83(10), pp. 898-902.

Poirier, K., Lebrun, N., Broix, L., Tian, G., Saillour, Y., Boscheron, C., Parrini, E., Valence, S., Pierre, B.S., Oger, M., Lacombe, D., Genevieve, D., Fontana, E., Darra, F., Cances, C., Barth, M., Bonneau, D., Bernadina, B.D., N'Guyen, S., Gitiaux, C., Parent, P., des Portes, V., Pedespan, J.M., Legrez, V., Castelnau-Ptakine, L., Nitschke, P., Hieu, T., Masson, C., Zelenika, D., Andrieux, A., Francis, F., Guerrini, R., Cowan, N.J., Bahi-Buisson, N. and Chelly, J. (2013) 'Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly', *Nat Genet*, 45(6), pp. 639-47.

Poirier, K., Saillour, Y., Bahi-Buisson, N., Jaglin, X.H., Fallet-Bianco, C., Nabbout, R., Castelnau-Ptakhine, L., Roubertie, A., Attie-Bitach, T., Desguerre, I., Genevieve, D., Barnerias, C., Keren, B., Lebrun, N., Boddaert, N., Encha-Razavi, F. and Chelly, J. (2010) 'Mutations in the neuronal ss-tubulin subunit TUBB3 result in malformation of cortical development and neuronal migration defects', *Hum Mol Genet*, 19(22), pp. 4462-73.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. (2010) 'Detection of nonneutral substitution rates on mammalian phylogenies', *Genome Res*, 20(1), pp. 110-21.

Pottier, C., Hannequin, D., Coutant, S., Rovelet-Lecrux, A., Wallon, D., Rousseau, S., Legallic, S., Paquet, C., Bombois, S., Pariente, J., Thomas-Anterion, C., Michon, A., Croisile, B., Etcharry-Bouyx, F., Berr, C., Dartigues, J.F., Amouyel, P., Dauchel, H., Boutoleau-Bretonniere, C., Thauvin, C., Frebourg, T., Lambert, J.C., Campion, D. and Collaborators, P.G. (2012) 'High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease', *Mol Psychiatry*, 17(9), pp. 875-9.

Pottier, C., Wallon, D., Rousseau, S., Rovelet-Lecrux, A., Richard, A.C., Rollin-Sillaire, A., Frebourg, T., Campion, D. and Hannequin, D. (2013) 'TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease', *J Alzheimers Dis*, 35(1), pp. 45-9.

Pritchard, J.K. (2001) 'Are rare variants responsible for susceptibility to complex diseases?', *Am J Hum Genet*, 69(1), pp. 124-37.

Pulst, S.M. (1993--a) 'Spinocerebellar Ataxia Type 2', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301452.

Pulst, S.M. (1993--b) 'Spinocerebellar Ataxia Type 13', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301404.

Purnell, S.M., Bleyl, S.B. and Bonkowsky, J.L. (2014) 'Clinical Exome Sequencing Identifies a Novel TUBB4A Mutation in a Child With Static Hypomyelinating Leukodystrophy', *Pediatr Neurol*, 50(6), pp. 608-11.

Pyle, A., Griffin, H., Duff, J., Bennett, S., Zwolinski, S., Smertenko, T., Yu-Wai Man, P., Santibanez-Koref, M., Horvath, R. and Chinnery, P.F. (2013) 'Late-onset sacsinopathy diagnosed by exome sequencing and comparative genomic hybridization', *J Neurogenet*, 27(4), pp. 176-82.

Pyle, A., Griffin, H., Keogh, M.J., Horvath, R. and Chinnery, P.F. (2015) 'Reply: Evaluation of exome sequencing variation in undiagnosed ataxias', *Brain*.

Pyle, A., Griffin, H., Yu-Wai-Man, P., Duff, J., Eglon, G., Pickering-Brown, S., Santibanez-Korev, M., Horvath, R. and Chinnery, P.F. (2012) 'Prominent sensorimotor neuropathy due to SACS mutations revealed by whole-exome sequencing', *Arch Neurol*, 69(10), pp. 1351-4.

Pyle, A., Smertenko, T., Bargiela, D., Griffin, H., Duff, J., Appleton, M., Douroudis, K., Pfeffer, G., Santibanez-Koref, M., Eglon, G., Yu-Wai-Man, P., Ramesh, V., Horvath, R. and Chinnery, P.F. (2014) 'Exome sequencing in undiagnosed inherited and sporadic ataxias', *Brain*.

Raiborg, C., Wenzel, E.M., Pedersen, N.M., Olsvik, H., Schink, K.O., Schultz, S.W., Vietri, M., Nisi, V., Bucci, C., Brech, A., Johansen, T. and Stenmark, H. (2015) 'Repeated ER-endosome contacts promote endosome translocation and neurite outgrowth', *Nature*, 520(7546), pp. 234-8.

Ramagopalan, S.V., Dyment, D.A., Cader, M.Z., Morrison, K.M., Disanto, G., Morahan, J.M., Berlanga-Taylor, A.J., Handel, A., De Luca, G.C., Sadovnick, A.D., Lepage, P., Montpetit, A. and Ebers, G.C. (2011) 'Rare variants in the CYP27B1 gene are associated with multiple sclerosis', *Ann Neurol*, 70(6), pp. 881-6.

Ramirez, O.A. and Couve, A. (2011) 'The endoplasmic reticulum and protein trafficking in dendrites and axons', *Trends Cell Biol*, 21(4), pp. 219-27.

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., Joset, P., Ropke, A., Moog, U., Riess, A., Thiel, C.T., Tzschach, A., Wiesener, A., Wohlleber, E., Zweier, C., Ekici, A.B., Zink, A.M., Rump, A., Meisinger, C., Grallert, H., Sticht, H., Schenck, A., Engels, H., Rappold, G., Schrock, E., Wieacker, P., Riess, O., Meitinger, T., Reis, A. and Strom, T.M. (2012) 'Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study', *Lancet*, 380(9854), pp. 1674-82.

Raychaudhuri, S. (2011) 'Mapping rare and common causal alleles for complex human diseases', *Cell*, 147(1), pp. 57-69.

Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. (1997) 'Improved splice site detection in Genie', *J Comput Biol*, 4(3), pp. 311-23.

Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R. and Lyon, E. (2013) 'ACMG clinical laboratory standards for next-generation sequencing', *Genet Med*, 15(9), pp. 733-47.

Reich, D.E. and Lander, E.S. (2001) 'On the allelic spectrum of human disease', *Trends Genet*, 17(9), pp. 502-10.

Reid, E., Kloos, M., Ashley-Koch, A., Hughes, L., Bevan, S., Svenson, I.K., Graham, F.L., Gaskell, P.C., Dearlove, A., Pericak-Vance, M.A., Rubinsztein, D.C. and Marchuk, D.A. (2002) 'A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10)', *Am J Hum Genet*, 71(5), pp. 1189-94.

Renaud, M., Anheim, M., Kamsteeg, E.J., Mallaret, M., Mochel, F., Vermeer, S., Drouot, N., Pouget, J., Redin, C., Salort-Campana, E., Kremer, H.P., Verschuuren-Bemelmans, C.C., Muller, J., Scheffer, H., Durr, A., Tranchant, C. and Koenig, M. (2014) 'Autosomal recessive cerebellar ataxia type 3 due to ANO10 mutations: delineation and genotype-phenotype correlation study', *JAMA Neurol*, 71(10), pp. 1305-10.

Retterer, K., Scuffins, J., Schmidt, D., Lewis, R., Pineda-Alvarez, D., Stafford, A., Schmidt, L., Warren, S., Gibellini, F., Kondakova, A., Blair, A., Bale, S., Matyakhina, L., Meck, J., Aradhya, S., Haverfield, E. (2015) 'Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort', 17(8), pp. 623-9.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K. and Rehm, H.L. (2015) 'Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology', *Genet Med*.

Rizzo, J.M. and Buck, M.J. (2012) 'Key principles and clinical applications of "next-generation" DNA sequencing', *Cancer Prev Res (Phila)*, 5(7), pp. 887-900.

Rosewich, H., Thiele, H., Ohlenbusch, A., Maschke, U., Altmuller, J., Frommolt, P., Zirn, B., Ebinger, F., Siemes, H., Nurnberg, P., Brockmann, K. and Gartner, J. (2012) 'Heterozygous de-novo mutations in ATP1A3 in patients with alternating hemiplegia of childhood: a whole-exome sequencing gene-identification study', *Lancet Neurol*, 11(9), pp. 764-73.

Rowland, L.P. (1983) 'Molecular genetics, pseudogenetics, and clinical neurology. The Robert Wartenberg Lecture', *Neurology*, 33(9), pp. 1179-95.

Roxburgh, R.H., Marquis-Nicholson, R., Ashton, F., George, A.M., Lea, R.A., Eccles, D., Mossman, S., Bird, T., van Gassen, K.L., Kamsteeg, E.J. and Love, D.R. (2013) 'The p.Ala510Val mutation in the SPG7 (paraplegin) gene is the most common mutation causing adult onset neurogenetic disease in patients of British ancestry', *J Neurol*, 260(5), pp. 1286-94.

Ruano, L., Melo, C., Silva, M.C. and Coutinho, P. (2014) 'The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies', *Neuroepidemiology*, 42(3), pp. 174-83.

Ruiz-Martinez, J., Krebs, C.E., Makarov, V., Gorostidi, A., Marti-Masso, J.F. and Paisan-Ruiz, C. (2015) 'GIGYF2 mutation in late-onset Parkinson's disease with cognitive impairment', *J Hum Genet*.

Sagona, A.P., Nezis, I.P., Bache, K.G., Haglund, K., Bakken, A.C., Skotheim, R.I. and Stenmark, H. (2011) 'A tumor-associated mutation of FYVE-CENT prevents its interaction with Beclin 1 and interferes with cytokinesis', *PLoS One*, 6(3), p. e17086.

Sailer, A., Scholz, S.W., Gibbs, J.R., Tucci, A., Johnson, J.O., Wood, N.W., Plagnol, V., Hummerich, H., Ding, J., Hernandez, D., Hardy, J., Federoff, H.J., Traynor, B.J., Singleton, A.B. and Houlden, H. (2012) 'Exome sequencing in an SCA14 family demonstrates its utility in diagnosing heterogeneous diseases', *Neurology*, 79(2), pp. 127-31.

Saita, S., Shirane, M., Natume, T., Iemura, S. and Nakayama, K.I. (2009) 'Promotion of neurite extension by protrudin requires its interaction with vesicle-associated membrane protein-associated protein', *J Biol Chem*, 284(20), pp. 13766-77.

Saitsu, H., Kato, M., Koide, A., Goto, T., Fujita, T., Nishiyama, K., Tsurusaki, Y., Doi, H., Miyake, N., Hayasaka, K. and Matsumoto, N. (2012a) 'Whole exome sequencing identifies KCNQ2 mutations in Ohtahara syndrome', *Ann Neurol*, 72(2), pp. 298-300.

Saitsu, H., Kato, M., Osaka, H., Moriyama, N., Horita, H., Nishiyama, K., Yoneda, Y., Kondo, Y., Tsurusaki, Y., Doi, H., Miyake, N., Hayasaka, K. and Matsumoto, N. (2012b) 'CASK aberrations in male patients with Ohtahara syndrome and cerebellar hypoplasia', *Epilepsia*, 53(8), pp. 1441-9.

Sakai, H., Yoshida, K., Shimizu, Y., Morita, H., Ikeda, S. and Matsumoto, N. (2010) 'Analysis of an insertion mutation in a cohort of 94 patients with spinocerebellar ataxia type 31 from Nagano, Japan', *Neurogenetics*, 11(4), pp. 409-15.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A., Song, Y., El-Fishawy, P., Murtha, R.C., Choi,

M., Overton, J.D., Bjornson, R.D., Carriero, N.J., Meyer, K.A., Bilguvar, K., Mane, S.M., Sestan, N., Lifton, R.P., Gunel, M., Roeder, K., Geschwind, D.H., Devlin, B. and State, M.W. (2012) 'De novo mutations revealed by whole-exome sequencing are strongly associated with autism', *Nature*, 485(7397), pp. 237-41.

Santens, P., Van Damme, T., Steyaert, W., Willaert, A., Sablonniere, B., De Paepe, A., Coucke, P.J. and Dermaut, B. (2015) 'RNF216 mutations as a novel cause of autosomal recessive Huntington-like disorder', *Neurology*, 84(17), pp. 1760-6.

Sassi, C., Guerreiro, R., Gibbs, R., Ding, J., Lupton, M.K., Troakes, C., Lunnon, K., Al-Sarraj, S., Brown, K.S., Medway, C., Lord, J., Turton, J., Mann, D., Snowden, J., Neary, D., Harris, J., Bras, J., Consortium, A., Morgan, K., Powell, J.F., Singleton, A. and Hardy, J. (2014) 'Exome sequencing identifies 2 novel presenilin 1 mutations (p.L166V and p.S230R) in British early-onset Alzheimer's disease', *Neurobiol Aging*, 35(10), pp. 2422 e13-6.

Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., Takahashi, M., Matsuura, T., Flanigan, K.M., Iwasaki, S., Ishino, F., Saito, Y., Murayama, S., Yoshida, M., Hashizume, Y., Takahashi, Y., Tsuji, S., Shimizu, N., Toda, T., Ishikawa, K. and Mizusawa, H. (2009) 'Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n', *Am J Hum Genet*, 85(5), pp. 544-57.

Sawyer, S.L., Schwartzentruber, J., Beaulieu, C.L., Dyment, D., Smith, A., Warman Chardon, J., Yoon, G., Rouleau, G.A., Suchowersky, O., Siu, V., Murphy, L., Hegele, R.A., Marshall, C.R., Bulman, D.E., Majewski, J., Tarnopolsky, M. and Boycott, K.M. (2014) 'Exome sequencing as a diagnostic tool for pediatric-onset ataxia', *Hum Mutat*, 35(1), pp. 45-9.

Schaefer, A.M., Taylor, R.W., Turnbull, D.M. and Chinnery, P.F. (2004) 'The epidemiology of mitochondrial disorders--past, present and future', *Biochim Biophys Acta*, 1659(2-3), pp. 115-20.

Schelhaas, H.J., Verbeek, D.S., Van de Warrenburg, B.P. and Sinke, R.J. (2004) 'SCA19 and SCA22: evidence for one locus with a worldwide distribution', *Brain*, 127(Pt 1), p. E6; author reply E7.

Scheper, G.C., van der Knaap, M.S. and Proud, C.G. (2007) 'Translation matters: protein synthesis defects in inherited disease', *Nat Rev Genet*, 8(9), pp. 711-23.

Schoeler, N.E., Cross, J.H., Drury, S., Lench, N., McMahon, J.M., MacKay, M.T., Scheffer, I.E., Sander, J.W. and Sisodiya, S.M. (2015) 'Favourable response to ketogenic dietary therapies: undiagnosed glucose 1 transporter deficiency syndrome is only one factor', *Dev Med Child Neurol*, 57(10), pp. 969-976.

Scholl, U.I., Choi, M., Liu, T., Ramaekers, V.T., Hausler, M.G., Grimmer, J., Tobe, S.W., Farhi, A., Nelson-Williams, C. and Lifton, R.P. (2009) 'Seizures, sensorineural deafness, ataxia, mental retardation, and electrolyte imbalance (SeSAME syndrome) caused by mutations in KCNJ10', *Proc Natl Acad Sci U S A*, 106(14), pp. 5842-7.

Schork, N.J., Murray, S.S., Frazer, K.A. and Topol, E.J. (2009) 'Common vs. rare allele hypotheses for complex diseases', *Curr Opin Genet Dev*, 19(3), pp. 212-9.

Schuelke, M. (1993-) 'Ataxia with Vitamin E Deficiency', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301419.

Schule, R. and Schols, L. (2011) 'Genetics of hereditary spastic paraplegias', *Semin Neurol*, 31(5), pp. 484-93.

Schwarz, J.M., Rodelsperger, C., Schuelke, M. and Seelow, D. (2010) 'MutationTaster evaluates disease-causing potential of sequence alterations', *Nat Methods*, 7(8), pp. 575-6.

Seelow, D., Schuelke, M., Hildebrandt, F. and Nurnberg, P. (2009) 'HomozygosityMapper--an interactive approach to homozygosity mapping', *Nucleic Acids Res*, 37(Web Server issue), pp. W593-9.

Sekar, S., McDonald, J., Cuyugan, L., Aldrich, J., Kurdoglu, A., Adkins, J., Serrano, G., Beach, T.G., Craig, D.W., Valla, J., Reiman, E.M. and Liang, W.S. (2015) 'Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes', *Neurobiol Aging*, 36(2), pp. 583-91.

Sellick, G.S., Barker, K.T., Stolte-Dijkstra, I., Fleischmann, C., Coleman, R.J., Garrett, C., Gloyn, A.L., Edghill, E.L., Hattersley, A.T., Wellauer, P.K., Goodwin, G. and Houlston, R.S. (2004) 'Mutations in PTF1A cause pancreatic and cerebellar agenesis', *Nat Genet*, 36(12), pp. 1301-5.

Serrano-Munuera, C., Corral-Juan, M., Stevanin, G., San Nicolas, H., Roig, C., Corral, J., Campos, B., de Jorge, L., Morcillo-Suarez, C., Navarro, A., Forlani, S., Durr, A., Kulisevsky, J., Brice, A., Sanchez, I., Volpini, V. and Matilla-Duenas, A. (2013) 'New subtype of spinocerebellar ataxia with altered vertical eye movements mapping to chromosome 1p32', *JAMA Neurol*, 70(6), pp. 764-71.

Shah, N.H. and Aizenman, E. (2014) 'Voltage-gated potassium channels at the crossroads of neuronal function, ischemic tolerance, and neurodegeneration', *Transl Stroke Res*, 5(1), pp. 38-58.

Shakkottai, V.G. and Fogel, B.L. (2013) 'Clinical neurogenetics: autosomal dominant spinocerebellar ataxia', *Neurol Clin*, 31(4), pp. 987-1007.

Shanks, M.E., Downes, S.M., Copley, R.R., Lise, S., Broxholme, J., Hudspith, K.A., Kwasniewska, A., Davies, W.I., Hankins, M.W., Packham, E.R., Clouston, P., Seller, A., Wilkie, A.O., Taylor, J.C., Ragoussis, J. and Nemeth, A.H. (2013) 'Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease', *Eur J Hum Genet*, 21(3), pp. 274-80.

Shashi, V., McConkie-Rosell, A., Rosell, B., Schoch, K., Vellore, K., McDonald, M., Jiang, Y.H., Xie, P., Need, A. and Goldstein, D.B. (2014) 'The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders', *Genet Med*, 16(2), pp. 176-82.

Shekarabi, M., Girard, N., Riviere, J.B., Dion, P., Houle, M., Toulouse, A., Lafreniere, R.G., Vercauteren, F., Hince, P., Laganiere, J., Rochefort, D., Faivre, L., Samuels, M. and Rouleau, G.A. (2008) 'Mutations in the nervous system--specific HSN2 exon of WNK1 cause hereditary sensory neuropathy type II', *J Clin Invest*, 118(7), pp. 2496-505.

Shendure, J. and Ji, H. (2008) 'Next-generation DNA sequencing', *Nat Biotechnol*, 26(10), pp. 1135-45.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) 'dbSNP: the NCBI database of genetic variation', *Nucleic Acids Res*, 29(1), pp. 308-11.

Shi, Y., Wang, J., Li, J.D., Ren, H., Guan, W., He, M., Yan, W., Zhou, Y., Hu, Z., Zhang, J., Xiao, J., Su, Z., Dai, M., Jiang, H., Guo, J., Zhang, F., Li, N., Du, J., Xu, Q., Hu, Y., Pan, Q., Shen, L., Wang, G., Xia, K., Zhang, Z. and Tang, B. (2013) 'Identification of CHIP as a novel causative gene for autosomal recessive cerebellar ataxia', *PLoS One*, 8(12), p. e81884.

Shirane, M. and Nakayama, K.I. (2006) 'Protrudin induces neurite formation by directional membrane trafficking', *Science*, 314(5800), pp. 818-21.

Simons, C., Wolf, N.I., McNeil, N., Caldovic, L., Devaney, J.M., Takanohashi, A., Crawford, J., Ru, K., Grimmond, S.M., Miller, D., Tonduti, D., Schmidt, J.L., Chudnow, R.S., van Coster, R., Lagae, L., Kisler, J., Sperner, J., van der Knaap, M.S., Schiffmann, R., Taft, R.J. and Vanderver, A. (2013) 'A de novo mutation in the beta-tubulin gene TUBB4A results in the leukoencephalopathy hypomyelination with atrophy of the basal ganglia and cerebellum', *Am J Hum Genet*, 92(5), pp. 767-73.

Smeets, C.J. and Verbeek, D.S. (2014) 'Cerebellar ataxia and functional genomics: Identifying the routes to cerebellar neurodegeneration', *Biochim Biophys Acta*.

Smith, B.N., Ticozzi, N., Fallini, C., Gkazi, A.S., Topp, S., Kenna, K.P., Scotter, E.L., Kost, J., Keagle, P., Miller, J.W., Calini, D., Vance, C., Danielson, E.W., Troakes, C., Tiloca, C., Al-Sarraj, S., Lewis, E.A., King, A., Colombrita, C.,

Pensato, V., Castellotti, B., de Belleroche, J., Baas, F., ten Asbroek, A.L., Sapp, P.C., McKenna-Yasek, D., McLaughlin, R.L., Polak, M., Asress, S., Esteban-Perez, J., Munoz-Blanco, J.L., Simpson, M., Consortium, S., van Rheenen, W., Diekstra, F.P., Lauria, G., Duga, S., Corti, S., Cereda, C., Corrado, L., Soraru, G., Morrison, K.E., Williams, K.L., Nicholson, G.A., Blair, I.P., Dion, P.A., Leblond, C.S., Rouleau, G.A., Hardiman, O., Veldink, J.H., van den Berg, L.H., Al-Chalabi, A., Pall, H., Shaw, P.J., Turner, M.R., Talbot, K., Taroni, F., Garcia-Redondo, A., Wu, Z., Glass, J.D., Gellera, C., Ratti, A., Brown, R.H., Jr., Silani, V., Shaw, C.E. and Landers, J.E. (2014) 'Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS', *Neuron*, 84(2), pp. 324-31.

Spacey, S. (1993-) 'Episodic Ataxia Type 2', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301674.

Spiegel, R., Pines, O., Ta-Shma, A., Burak, E., Shaag, A., Halvardson, J., Edvardson, S., Mahajna, M., Zenvirt, S., Saada, A., Shalev, S., Feuk, L. and Elpeleg, O. (2012) 'Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2', *Am J Hum Genet*, 90(3), pp. 518-23.

Srivastava, S., Cohen, J.S., Vernon, H., Baranano, K., McClellan, R., Jamal, L., Naidu, S. and Fatemi, A. (2014) 'Clinical whole exome sequencing in child neurology practice', *Ann Neurol*, 76(4), pp. 473-83.

Stan, A.D., Ghose, S., Gao, X.M., Roberts, R.C., Lewis-Amezcua, K., Hatanpaa, K.J. and Tamminga, C.A. (2006) 'Human postmortem tissue: what quality markers matter?', *Brain Res*, 1123(1), pp. 1-11.

Steckley, J.L., Ebers, G.C., Cader, M.Z. and McLachlan, R.S. (2001) 'An autosomal dominant disorder with episodic ataxia, vertigo, and tinnitus', *Neurology*, 57(8), pp. 1499-502.

Steinberg, K.M., Yu, B., Koboldt, D.C., Mardis, E.R. and Pamphlett, R. (2015) 'Exome sequencing of case-unaffected-parents trios reveals recessive and de novo genetic variants in sporadic ALS', *Sci Rep*, 5, p. 9124.

Stenmark, H., Aasland, R. and Driscoll, P.C. (2002) 'The phosphatidylinositol 3-phosphate-binding FYVE finger', *FEBS Lett*, 513(1), pp. 77-84.

Stenson, P.D., Mort, M., Ball, E.V., Howells, K., Phillips, A.D., Thomas, N.S. and Cooper, D.N. (2009) 'The Human Gene Mutation Database: 2008 update', *Genome Med*, 1(1), p. 13.

Stevanin, G., Bouslam, N., Broussolle, E., Ravaux, L., Boland, A., Durr, A. and Brice, A. (2003) 'Autosomal dominant cerebellar ataxia with sensory neuropathy maps to the spinocerebellar ataxia 25 (SCA25) locus on chromosome 2p15-p21.', *American Journal of Human Genetics*, 73(5), pp. 550-550.

Storey, E. (1993-) 'Spinocerebellar Ataxia Type 15', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301536.

Storey, E., Bahlo, M., Fahey, M., Sisson, O., Lueck, C.J. and Gardner, R.J. (2009) 'A new dominantly inherited pure cerebellar ataxia, SCA 30', *J Neurol Neurosurg Psychiatry*, 80(4), pp. 408-11.

Stranneheim, H. and Wedell, A. (2015) 'Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders', *J Intern Med*.

Subramony, S.H. and Ashizawa, T. (1993-) 'Spinocerebellar Ataxia Type 1', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301363.

Sudo, H. and Baas, P.W. (2011) 'Strategies for diminishing katanin-based loss of microtubules in tauopathic neurodegenerative diseases', *Hum Mol Genet*, 20(4), pp. 763-78.

Sulonen, A.M., Ellonen, P., Almusa, H., Lepisto, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., Joensuu, H., Raivio, T., Suomalainen, A. and Saarela, J. (2011) 'Comparison of solution-based exome capture methods for next generation sequencing', *Genome Biol*, 12(9), p. R94.

Sun, Y., Almomani, R., Breedveld, G.J., Santen, G.W., Aten, E., Lefeber, D.J., Hoff, J.I., Brusse, E., Verheijen, F.W., Verdijk, R.M., Kriek, M., Oostra, B., Breuning, M.H., Losekoot, M., den Dunnen, J.T., van de Warrenburg, B.P. and Maat-Kievit, A.J. (2013) 'Autosomal recessive spinocerebellar ataxia 7 (SCAR7) is caused by variants in TPP1, the gene involved in classic late-infantile neuronal ceroid lipofuscinosis 2 disease (CLN2 disease)', *Hum Mutat*, 34(5), pp. 706-13.

Sun, Y., Hegamyer, G. and Colburn, N.H. (1993) 'PCR-direct sequencing of a GC-rich region by inclusion of 10% DMSO: application to mouse c-jun', *Biotechniques*, 15(3), pp. 372-4.

Sunyaev, S.R. (2012) 'Inferring causality and functional significance of human coding DNA variants', *Hum Mol Genet*, 21(R1), pp. R10-7.

Synofzik, M., Schule, R., Schulze, M., Gburek-Augustat, J., Schweizer, R., Schirmacher, A., Krageloh-Mann, I., Gonzalez, M., Young, P., Zuchner, S., Schols, L. and Bauer, P. (2014) 'Phenotype and frequency of STUB1 mutations: next-generation screenings in Caucasian ataxia and spastic paraplegia cohorts', *Orphanet J Rare Dis*, 9, p. 57.

Takashima, H., Boerkoel, C.F., John, J., Saifi, G.M., Salih, M.A., Armstrong, D., Mao, Y., Quiocho, F.A., Roa, B.B., Nakagawa, M., Stockton, D.W. and Lupski, J.R. (2002) 'Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy', *Nat Genet*, 32(2), pp. 267-72.

Tarnutzer, A.A., Gerth-Kahlert, C., Timmann, D., Chang, D.I., Harmuth, F., Bauer, P., Straumann, D. and Synofzik, M. (2015) 'Boucher-Neuhauser syndrome: cerebellar degeneration, chorioretinal dystrophy and hypogonadotropic hypogonadism: two novel cases and a review of 40 cases from the literature', *J Neurol*, 262(1), pp. 194-202.

Taylor, R.W., Pyle, A., Griffin, H., Blakely, E.L., Duff, J., He, L., Smertenko, T., Alston, C.L., Neeve, V.C., Best, A., Yarham, J.W., Kirschner, J., Schara, U., Talim, B., Topaloglu, H., Baric, I., Holinski-Feder, E., Abicht, A., Czermin, B., Kleinle, S., Morris, A.A., Vassallo, G., Gorman, G.S., Ramesh, V., Turnbull,

D.M., Santibanez-Koref, M., McFarland, R., Horvath, R. and Chinnery, P.F. (2014) 'Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies', *JAMA*, 312(1), pp. 68-77.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H.M., Jordan, D., Leal, S.M., Gabriel, S., Rieder, M.J., Abecasis, G., Altshuler, D., Nickerson, D.A., Boerwinkle, E., Sunyaev, S., Bustamante, C.D., Bamshad, M.J. and Akey, J.M. (2012) 'Evolution and functional impact of rare coding variation from deep sequencing of human exomes', *Science*, 337(6090), pp. 64-9.

Tischfield, M.A., Baris, H.N., Wu, C., Rudolph, G., Van Maldergem, L., He, W., Chan, W.M., Andrews, C., Demer, J.L., Robertson, R.L., Mackey, D.A., Ruddle, J.B., Bird, T.D., Gottlob, I., Pieh, C., Traboulsi, E.I., Pomeroy, S.L., Hunter, D.G., Soul, J.S., Newlin, A., Sabol, L.J., Doherty, E.J., de Uzcategui, C.E., de Uzcategui, N., Collins, M.L., Sener, E.C., Wabbels, B., Hellebrand, H., Meitinger, T., de Berardinis, T., Magli, A., Schiavi, C., Pastore-Trossello, M., Koc, F., Wong, A.M., Levin, A.V., Geraghty, M.T., Descartes, M., Flaherty, M., Jamieson, R.V., Moller, H.U., Meuthen, I., Callen, D.F., Kerwin, J., Lindsay, S., Meindl, A., Gupta, M.L., Jr., Pellman, D. and Engle, E.C. (2010) 'Human TUBB3 mutations perturb microtubule dynamics, kinesin interactions, and axon guidance', *Cell*, 140(1), pp. 74-87.

Tischfield, M.A., Cederquist, G.Y., Gupta, M.L., Jr. and Engle, E.C. (2011) 'Phenotypic spectrum of the tubulin-related disorders and functional implications of disease-causing mutations', *Curr Opin Genet Dev*, 21(3), pp. 286-94.

Tong, X., Ao, Y., Faas, G.C., Nwaobi, S.E., Xu, J., Haustein, M.D., Anderson, M.A., Mody, I., Olsen, M.L., Sofroniew, M.V. and Khakh, B.S. (2014) 'Astrocyte Kir4.1 ion channel deficits contribute to neuronal dysfunction in Huntington's disease model mice', *Nat Neurosci*, 17(5), pp. 694-703.

Toscano, A. and Schoser, B. (2013) 'Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review', *J Neurol*, 260(4), pp. 951-9.

Toyoshima, Y., Onodera, O., Yamada, M., Tsuji, S. and Takahashi, H. (1993-) 'Spinocerebellar Ataxia Type 17', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301611.

Tsoi, H., Yu, A.C., Chen, Z.S., Ng, N.K., Chan, A.Y., Yuen, L.Y., Abrigo, J.M., Tsang, S.Y., Tsui, S.K., Tong, T.M., Lo, I.F., Lam, S.T., Mok, V.C., Wong, L.K., Ngo, J.C., Lau, K.F., Chan, T.F. and Chan, H.Y. (2014) 'A novel missense mutation in CCDC88C activates the JNK pathway and causes a dominant form of spinocerebellar ataxia', *J Med Genet*, 51(9), pp. 590-5.

Tsuji, S. (1993-) *Drpla*. GeneReviews.

Tsuji, S. (2010) 'Genetics of neurodegenerative diseases: insights from high-throughput resequencing', *Hum Mol Genet*, 19(R1), pp. R65-70.

Turner, P.F. and Margolis, R.L. (1984) 'Taxol-induced bundling of brain-derived microtubules', *J Cell Biol*, 99(3), pp. 940-6.

van de Warrenburg, B.P., Sinke, R.J., Verschuuren-Bemelmans, C.C., Scheffer, H., Brunt, E.R., Ippel, P.F., Maat-Kievit, J.A., Dooijes, D., Notermans, N.C., Lindhout, D., Knoers, N.V. and Kremer, H.P. (2002) 'Spinocerebellar ataxias in the Netherlands: prevalence and age at onset variance analysis', *Neurology*, 58(5), pp. 702-8.

Van Schil, K., Meire, F., Karlstetter, M., Bauwens, M., Verdin, H., Coppieters, F., Scheiffert, E., Van Nechel, C., Langmann, T., Deconinck, N. and De Baere, E. (2014) 'Early-onset autosomal recessive cerebellar ataxia associated with retinal dystrophy: new human hotfoot phenotype caused by homozygous GRID2 deletion', *Genet Med*.

van Swieten, J.C., Brusse, E., de Graaf, B.M., Krieger, E., van de Graaf, R., de Koning, I., Maat-Kievit, A., Leegwater, P., Dooijes, D., Oostra, B.A. and Heutink, P. (2003) 'A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]', *Am J Hum Genet*, 72(1), pp. 191-9.

Vantaggiato, C., Crimella, C., Airoldi, G., Polishchuk, R., Bonato, S., Brighina, E., Scarlato, M., Musumeci, O., Toscano, A., Martinuzzi, A., Santorelli, F.M., Ballabio, A., Bresolin, N., Clementi, E. and Bassi, M.T. (2013) 'Defective autophagy in spastizin mutated patients with hereditary spastic paraparesis type 15', *Brain*, 136(Pt 10), pp. 3119-39.

Veltman, J.A. and Brunner, H.G. (2012) 'De novo mutations in human genetic disease', *Nat Rev Genet*, 13(8), pp. 565-75.

Verbeek, D.S., Schelhaas, J.H., Ippel, E.F., Beemer, F.A., Pearson, P.L. and Sinke, R.J. (2002) 'Identification of a novel SCA locus (SCA19) in a Dutch autosomal dominant cerebellar ataxia family on chromosome region 1p21-q21', *Hum Genet*, 111(4-5), pp. 388-93.

Verbeek, D.S., van de Warrenburg, B.P., Wesseling, P., Pearson, P.L., Kremer, H.P. and Sinke, R.J. (2004) 'Mapping of the SCA23 locus involved in autosomal dominant cerebellar ataxia to chromosome region 20p13-12.3', *Brain*, 127(Pt 11), pp. 2551-7.

Vermeer, S., Hoischen, A., Meijer, R.P., Gilissen, C., Neveling, K., Wieskamp, N., de Brouwer, A., Koenig, M., Anheim, M., Assoum, M., Drouot, N., Todorovic, S., Milic-Rasic, V., Lochmuller, H., Stevanin, G., Goizet, C., David, A., Durr, A., Brice, A., Kremer, B., van de Warrenburg, B.P., Schijvenaars, M.M., Heister, A., Kwint, M., Arts, P., van der Wijst, J., Veltman, J., Kamsteeg, E.J., Scheffer, H. and Knoers, N. (2010) 'Targeted next-generation sequencing of a 12.5 Mb homozygous region reveals ANO10 mutations in patients with autosomal-recessive cerebellar ataxia', *Am J Hum Genet*, 87(6), pp. 813-9.

Vermeer, S., van de Warrenburg, B.P. and Kamsteeg, E.J. (1993-) 'Arsacs', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301432.

Vilarino-Guell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., Behrouz, B., Melrose, H.L., Hentati, E., Puschmann, A., Evans, D.M., Conibear, E., Wasserman, W.W., Aasly, J.O., Burkhard, P.R., Djaldetti, R., Ghika, J., Hentati,

F., Krygowska-Wajs, A., Lynch, T., Melamed, E., Rajput, A., Rajput, A.H., Solida, A., Wu, R.M., Uitti, R.J., Wszolek, Z.K., Vingerhoets, F. and Farrer, M.J. (2011) 'VPS35 mutations in Parkinson disease', *Am J Hum Genet*, 89(1), pp. 162-7.

Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., van Bon, B.W., Hoischen, A., de Vries, B.B., Brunner, H.G. and Veltman, J.A. (2010) 'A de novo paradigm for mental retardation', *Nat Genet*, 42(12), pp. 1109-12.

Wade, R.H., Garcia-Saez, I. and Kozielski, F. (2009) 'Structural variations in protein superfamilies: actin and tubulin', *Mol Biotechnol*, 42(1), pp. 49-60.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S.M., Rippey, C.F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R.L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., Butler, P., Eckstrand, K., Noory, L., Gochman, P., Long, R., Chen, Z., Davis, S., Baker, C., Eichler, E.E., Meltzer, P.S., Nelson, S.F., Singleton, A.B., Lee, M.K., Rapoport, J.L., King, M.C. and Sebat, J. (2008) 'Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia', *Science*, 320(5875), pp. 539-43.

Wanders, R.J.A., Waterham, H.R. and Leroy, B.P. (1993-) 'Refsum Disease', in Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds.) *GeneReviews*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301527.

Wang, J.L., Yang, X., Xia, K., Hu, Z.M., Weng, L., Jin, X., Jiang, H., Zhang, P., Shen, L., Guo, J.F., Li, N., Li, Y.R., Lei, L.F., Zhou, J., Du, J., Zhou, Y.F., Pan, Q., Wang, J., Li, R.Q. and Tang, B.S. (2010a) 'TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing', *Brain*, 133(Pt 12), pp. 3510-8.

Wang, K., Li, M. and Hakonarson, H. (2010b) 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Res*, 38(16), p. e164.

Wang, M., Ye, R., Barron, E., Baumeister, P., Mao, C., Luo, S., Fu, Y., Luo, B., Dubeau, L., Hinton, D.R. and Lee, A.S. (2010c) 'Essential role of the unfolded protein response regulator GRP78/BiP in protection from neuronal apoptosis', *Cell Death Differ*, 17(3), pp. 488-98.

Ward, A.J. and Cooper, T.A. (2010) 'The pathobiology of splicing', *J Pathol*, 220(2), pp. 152-63.

Ward, L.D. and Kellis, M. (2012) 'Interpreting noncoding genetic variation in complex traits and human disease', *Nat Biotechnol*, 30(11), pp. 1095-106.

Warman Chardon, J., Beaulieu, C., Hartley, T., Boycott, K.M. and Dyment, D.A. (2015) 'Axons to Exons: the Molecular Diagnosis of Rare Neurological Diseases by Next-Generation Sequencing', *Curr Neurol Neurosci Rep*, 15(9), p. 584.

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N. and Watson, M. (2015) 'Exome Sequencing: Current and Future Perspectives', *G3 (Bethesda)*.

Waters, M.F., Minassian, N.A., Stevanin, G., Figueroa, K.P., Bannister, J.P., Nolte, D., Mock, A.F., Evidente, V.G., Fee, D.B., Muller, U., Durr, A., Brice, A., Papazian, D.M. and Pulst, S.M. (2006) 'Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes', *Nat Genet*, 38(4), pp. 447-51.

Weischenfeldt, J., Symmons, O., Spitz, F. and Korbel, J.O. (2013) 'Phenotypic impact of genomic structural variation: insights from and for human disease', *Nat Rev Genet*, 14(2), pp. 125-38.

Welch, E.M., Barton, E.R., Zhuo, J., Tomizawa, Y., Friesen, W.J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C.R., Hwang, S., Wilde, R.G., Karp, G., Takasugi, J., Chen, G., Jones, S., Ren, H., Moon, Y.C., Corson, D., Turpoff, A.A., Campbell, J.A., Conn, M.M., Khan, A., Almstead, N.G., Hedrick, J., Mollin, A., Risher, N., Weetall, M., Yeh, S., Branstrom, A.A., Colacino, J.M., Babiak, J., Ju, W.D., Hirawat, S., Northcutt, V.J., Miller, L.L., Spatrick, P., He, F., Kawana, M., Feng, H., Jacobson, A., Peltz, S.W. and Sweeney, H.L. (2007)

'PTC124 targets genetic disorders caused by nonsense mutations', *Nature*, 447(7140), pp. 87-91.

Wengrod, J., Martin, L., Wang, D., Frischmeyer-Guerrerio, P., Dietz, H.C. and Gardner, L.B. (2013) 'Inhibition of nonsense-mediated RNA decay activates autophagy', *Mol Cell Biol*, 33(11), pp. 2128-35.

Williams, K.L., Warraich, S.T., Yang, S., Solski, J.A., Fernando, R., Rouleau, G.A., Nicholson, G.A. and Blair, I.P. (2012) 'UBQLN2/ubiquilin 2 mutation and pathology in familial amyotrophic lateral sclerosis', *Neurobiol Aging*, 33(10), pp. 2527 e3-10.

Winter, N., Kovermann, P. and Fahlke, C. (2012) 'A point mutation associated with episodic ataxia 6 increases glutamate transporter anion currents', *Brain*, 135(Pt 11), pp. 3416-25.

Wu, C.H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D.M., Kost, J.E., Gonzalez-Perez, P., Fox, A.D., Adams, J., Taroni, F., Tiloca, C., Leclerc, A.L., Chafe, S.C., Mangroo, D., Moore, M.J., Zitzewitz, J.A., Xu, Z.S., van den Berg, L.H., Glass, J.D., Siciliano, G., Cirulli, E.T., Goldstein, D.B., Salachas, F., Meininger, V., Rossoll, W., Ratti, A., Gellera, C., Bosco, D.A., Bassell, G.J., Silani, V., Drory, V.E., Brown, R.H., Jr. and Landers, J.E. (2012a) 'Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis', *Nature*, 488(7412), pp. 499-503.

Wu, J., Shen, E., Shi, D., Sun, Z. and Cai, T. (2012b) 'Identification of a novel Cys146X mutation of SOD1 in familial amyotrophic lateral sclerosis by whole-exome sequencing', *Genet Med*, 14(9), pp. 823-6.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M.R., Leduc, M.S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S.E., Lupski, J.R., Beaudet, A.L., Gibbs, R.A. and Eng, C.M. (2013) 'Clinical whole-exome sequencing for the diagnosis of mendelian disorders', *N Engl J Med*, 369(16), pp. 1502-11.

Yeo, G. and Burge, C.B. (2004) 'Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals', *J Comput Biol*, 11(2-3), pp. 377-94.

Yeung, K.T., Das, S., Zhang, J., Lomniczi, A., Ojeda, S.R., Xu, C.F., Neubert, T.A. and Samuels, H.H. (2011) 'A novel transcription complex that selectively modulates apoptosis of breast cancer cells through regulation of FASTKD2', *Mol Cell Biol*, 31(11), pp. 2287-98.

Ylikallio, E. and Suomalainen, A. (2012) 'Mechanisms of mitochondrial diseases', *Ann Med*, 44(1), pp. 41-59.

Yu, G.Y., Howell, M.J., Roller, M.J., Xie, T.D. and Gomez, C.M. (2005) 'Spinocerebellar ataxia type 26 maps to chromosome 19p13.3 adjacent to SCA6', *Ann Neurol*, 57(3), pp. 349-54.

Zanni, G., Bertini, E., Bellcross, C., Nedelec, B., Froyen, G., Neuhauser, G., Opitz, J.M. and Chelly, J. (2008) 'X-linked congenital ataxia: a new locus maps to Xq25-q27.1', *Am J Med Genet A*, 146A(5), pp. 593-600.

Zanni, G. and Bertini, E.S. (2011) 'X-linked disorders with cerebellar dysgenesis', *Orphanet J Rare Dis*, 6, p. 24.

Zanni, G., Saillour, Y., Nagara, M., Billuart, P., Castelnau, L., Moraine, C., Faivre, L., Bertini, E., Durr, A., Guichet, A., Rodriguez, D., des Portes, V., Beldjord, C. and Chelly, J. (2005) 'Oligophrenin 1 mutations frequently cause X-linked mental retardation with cerebellar hypoplasia', *Neurology*, 65(9), pp. 1364-9.

Zhang, C., Li, D., Ma, Y., Yan, J., Yang, B., Li, P., Yu, A., Lu, C. and Ma, X. (2012) 'Role of spastin and protrudin in neurite outgrowth', *J Cell Biochem*, 113(7), pp. 2296-307.

Zhang, M.Q. (1998) 'Statistical features of human exons and their flanking regions', *Hum Mol Genet*, 7(5), pp. 919-32.

Zhou, J., Tawk, M., Tiziano, F.D., Veillet, J., Bayes, M., Nolent, F., Garcia, V., Servidei, S., Bertini, E., Castro-Giner, F., Renda, Y., Carpentier, S., Andrieu-Abadie, N., Gut, I., Levade, T., Topaloglu, H. and Melki, J. (2012) 'Spinal

muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1', *Am J Hum Genet*, 91(1), pp. 5-14.

Zimprich, A., Benet-Pages, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., Rossle, S.C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., Hotzy, C., Reinthaler, E., Harutyunyan, A., Kralovics, R., Peters, A., Zimprich, F., Brucke, T., Poewe, W., Auff, E., Trenkwalder, C., Rost, B., Ransmayr, G., Winkelmann, J., Meitinger, T. and Strom, T.M. (2011) 'A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease', *Am J Hum Genet*, 89(1), pp. 168-75.