Genetic Identification and Characterisation of Novel Genes Implicated in Cerebellar Ataxia

Roisin Hannah Sullivan

Thesis

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Queen Square Institute of Neurology

UCL

I, Roisin Sullivan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed,

Roisin Sullivan

Impact Statement

The research detailed within this thesis has the potential to impact the clinical and biological understanding of ataxic disorders, specifically inherited disorders that feature ataxia, including CANVAS, multiple systems atrophy (MSA) and late-onset cerebellar ataxia (LOCA), and autosomal recessive cerebellar ataxia.

Clinically, the identification of a biallelic repeat expansion in *RFC1* has uncovered a genetic basis of CANVAS and general late-onset ataxia cases and will aid in solving molecularly undiagnosed ataxic disorders. It also has the potential to help diagnostically distinguish between ataxic disorders that overlap in clinical presentation. It has identified cases of *RFC1*-associated disease previously diagnosed with MSA, and since publication, this implication has broadened to include additional neurodegenerative disorders.

This work also has the potential to impact the methods used to identify repeat expansions both using wet-laboratory techniques and bioinformatics. This work has optimised the wet-lab workflow involved in detecting the *RFC1* repeat expansion within the laboratory, refining the protocol for Southern blotting to ensure reliable and accurate repeat expansion sizing. Throughout this research, effective training of these techniques has been developed and delivered to many researchers across the globe, implementing the workflow into other laboratories with a standardised protocol.

Through a collaboration with Illumina, this work has also aided in the development and refinement of a bioinformatic tool that accurately detects individuals harbouring the recessive *RFC1* repeat expansion using WGS data. It has the potential to minimise the wet-lab standardised workflow, providing accurate and time-efficient molecular diagnosis. This bioinformatic tool is in discussions to be added to the 100,000 genomes project research environment and incorporated within the NHS neurogenetics diagnostic testing. This will aid in the diagnosis of suspected *RFC1*-associated disease. The RFC1 Caller method also has the potential to aid in the development of further tools able to detect complex repeat expansions, due to its ability to detect individual motif configurations.

This work has also identified a novel phenotype associated with ACBD5-deficient peroxisomal disorders, comprising pure ataxic features without retinal dystrophy. Through a series of collaborations, Patient 3 presented with autonomic dysfunction, cerebellar ataxia, the largest collection of ACBD5-deficient patients has been recruited to date, characterising the phenotypic profile associated with ACBD5-deficiency, and providing a valuable resource available to collaborators to investigate pathological mechanisms.

This research has led to a number of measurable outputs in scholarly journals, with findings still in the process of publication submission. These findings have also been presented at a number of national and international conferences, disseminating this research to both scientific and lay audiences in order to educate and circulate these findings.

Abstract

The inherited cerebellar ataxias are a group of heterogenous, neurogenerative disorders that affect the cerebellum and other connected regions. Four projects featuring cerebellar ataxia; Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome and autosomal recessive cerebellar ataxia, were explored from a genetic and functional prospective.

The first results chapter explores the ataxic disorder CANVAS and characterises its genetic cause; a recessive $AAGGG_{(n)}$ repeat expansion in *RFC1*. Using a workflow of flanking PCR, RPPCR and Southern blotting, the allelic frequency for the pathogenic $AAGGG_{(n)}$ repeat expansion was determined within the general population as 0.7%. The heterogenous region was characterised into additional non-pathogenic motifs and the allelic frequency was also identified in a general late-onset ataxia cohort as 3%. The Southern blotting technique was optimised, producing a more efficient diagnostic workflow. Additionally, a novel pathogenic motif conformation was identified in a Māori and Māori Cook Island population.

The second results chapter is a continuation of *RFC1* repeat expansion characterisation and its prevalence in other movement disorders. Due to the phenotypic overlap of CANVAS and multiple systems atrophy (MSA), two cohorts of MSA individuals, one pathologically confirmed, and one clinically diagnosed, were screened for the *RFC1* repeat expansion. The recessive repeat expansion was absent in the pathologically confirmed cohort, suggesting it is not involved in MSA pathophysiology. Three patients clinically diagnosed with 'possible' or 'probable' MSA from were positive for the recessive expansion. They had atypical presentations of both CANVAS and MSA, suggesting an MSA-mimic disorder. This highlights the expanding phenotypic spectrum associated with the repeat expansion and emphasises the importance of screening any cases within the clinic with atypical MSA symptoms.

The third results chapter details the development of a novel bioinformatic tool, the RFC1 caller, in collaboration with Illumina. The caller utilises ExpansionHunter output files, that are initially generated from WGS data. Using the 100,000 genomes project 62, 550 WGS samples with rare diseases were screened using the RFC1 caller, and 145 individuals were called as having the biallelic repeat expansion. These were confirmed using Southern blotting in cases where genomic DNA was available in the recruitment centre affiliated with the National Hospital for Neurology and Neurosurgery. The results highlighted further clinical phenotypes not yet associated with the repeat expansion, or possible incomplete penetrance of the disease. The RFC1 caller had a sensitivity and specificity of 96.2% and 100% respectively, using a cohort of samples independently selected for diagnostic screening based on their clinical presentation. This suggests that the RFC1 caller is an effective and accurate predictor of the biallelic repeat expansion in *RFC1* and could be beneficial for the molecular diagnosis of affected patients.

The fourth results chapter is a project involving mutations in *ACBD5* and its implication in autosomal recessive cerebellar ataxia in six patients. Clinically, three of the patients in the cohort presented with pure ataxic symptoms and the other three with ataxia and retinal dystrophy. All six patients showed a complete loss of function of ACBD5, a peroxisomal protein, and an increase in very-long chain fatty acids, suggesting an impairment in β-oxidation. Assays testing mitochondrial function suggested a subtle secondary impairment in complex I of the OXPHOS machinery, which appeared more severe in those patients without retinal dystrophy. All six mutations were predicted to undergo nonsense medicated decay using the bioinformatic tool Mutation Taster, western blotting, and cDNA sequencing. RNA-sequencing was performed to look for transcript differences between patients and identify other processes that may be involved in ACBD5-deficiency. No differences in ACBD5-transcript reads were identified between patients, suggesting the transcripts are degraded, and confirming loss of function. Four significantly downregulated genes were identified in patients compared to controls, however zero significantly differentially expressed genes were identified in patients with retinal dystrophy in comparison to those without. These results identify some potential novel candidate genes implicated in the pathogenic mechanism and enriched pathways that may indicate subtle differences in pathogenicity between phenotypic differences.

Acknowledgements

I would first and foremost like to thank all the patients and their families; without them this research would not be possible. Additionally, I would like to thank the Medical Research Council for funding this work, as well as the MSA Trust, Ataxia UK and the Wellcome Trust.

A huge thanks also to my supervisors at UCL Institute of Neurology, Professor Henry Houlden, Professor Mike Hanna, and Dr Jana Vandrovcova, for all their support over the past four years. I would especially like to thank Henry for the motivation, constant support, and guidance, and for always making time for me. I would also like to especially thank Jana for her guidance in bioinformatics and for her unwavering patience and kindness.

Thank you to my collaborators who have made this research possible. Thank you to Dr Andrea Cortese for all his guidance with *RFC1*. Thank you to Prof. Nancy Braverman, Dr Luisa Averdunk, Dr Irman Forghani, Prof Adeline Vanderver, Dr Sacha Ferdinandusse and Prof. Michael Schrader and his team for kindly providing me with patient fibroblast lines and carrying out assays which have contributed to this work. Special thanks to Dr Mike Eberle and Dr Sai Chen for all their help with the development of the RFC1 Caller. Thank you to Prof. Andrey Abramov and Dr Plamena Stroh who guided me in confocal microscopy and mitochondrial assay techniques. Thank you also to Dr Kerri Venner for teaching me electron microscopy.

Thank you to the members past and present of the Houlden lab: Dr Stephanie Efthymiou for always being there for me, Valentina Galassi Deforie and Natalia Dominik for getting me through hundreds of Southern blots and Dr Reza Maroofian, Dr Monika Madej, Dr Viorica Chelban, Dr Rauan Kaiyrzhanov, Dr Francesca Margrinelli, Dr Sarah Beecroft, Clarissa Rocca, Eva Galanaki, Alice Gennari, Huma Tariq, and Mohammed Ilyas. Thank you also to the members of the Neurogenetics Diagnostic Lab, particularly, Dr James Polke, Ambreen Tariq and Katrina Newland, and to all the security and staff at the Institute of Neurology, especially the front desk, who were always so kind and helpful.

I especially would like to thank all the members of office 708 who have constantly supported me over the years and made my time at the institute so fun: Dr Emer O'Connor, Dr Isabel Hostettler, Dr Benjamin O'Callaghan, Dr Wai Yan Yau, Dr Nourelhoda Haridy and Dr Thomas Bounarais. Special thanks to my partner in crime Emer for her kindness, positivity and sense of humour, I am so lucky to count her as one of one of my closest friends.

Thank you to my friends from home and further afield, especially to Helena O'Shea, Tania Nadarajah, Rio Cheung, Honor Morris, Dana Mokaddem, Sophie Norden, Agnes Rice, Adrianna Juraszek, Natalia Diaz-de-Cerio, Madison Pritchard, Faye Wilkes, and Sophie Stevens. They have all supported me through the highs and lows of the past four years; always championing me and insisting on celebrating successes, no matter how big or small.

I would also like to thank my family, especially my brothers Jack, Fionn and Sé, and my grandparents who are and were a constant source of encouragement. To my parents, who have always motivated me to achieve my best and given me unconditional support and love, thank you. They have made sacrifices to give me so many wonderful opportunities and all my successes are owed to them.

Finally, thank you to one of the most intelligent and inspirational men that I know, my partner Dr Benjamin Simpson, who has given me unrelenting support and encouragement. I will be forever thankful that I shared the experience of undertaking a PhD alongside my best friend.

Contents

List of Figures

List of tables

List of commonly used abbreviations

Abbreviation	Phrase
100kGP	100,000 genomes project
ABCD1	ATP binding cassette subfamily D member 1
ACBD5	Acyl-coA binding domain containing 5
ADCA	Autosomal dominant cerebellar ataxia
ALD	Adrenoleukodystrophy
ALS	Amyotrophic lateral sclerosis
AMD	Age-related macular degeneration
ARCA	Autosomal recessive cerebellar ataxia
ARSACS	Autosomal recessive spastic cerebellar ataxia of Charlevoix-Saguenay
AT	Ataxia telangiectasia
AUC	Area under the curve
AVED	Ataxia with vitamin E deficiency
BSA	Bovine Serum Albumin
BVF	Bilateral vestibular reflex
C9orf72	Chromosome 9 open reading frame 72
CAKUT	Congenital abnormalities of the kidneys and urinary tract
CANVAS	Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome
cDNA	Complimentary DNA
CMT	Charcot-Marie-Tooth
CNS	Central nervous system
Cyt c	Cytochrome c
DIG	Digoxigenin
DM	Myotonic dystrophies
DNA	Deoxyribonucleic acid
DNTM1	DNA methyltransferase 1
DRPLA	Dentatorubral-pallidoluysian atrophy
EJC	Exon junction complex
EPM1	Myoclonic epilepsy type 1
ER	Endoplasmic reticulum
FADH	Flavin adenine dinucleotide
FDR	False discovery rate
FFAT	Two phenylalanines (FF) in an acidic tract
FRDA	Friedreich’s ataxia
FSHD	Facioscapulohumeral muscular dystrophy
FTD	Frontotemporal dementia
FXN	Frataxin
GCI	Glial cytoplasmic inclusions
gDNA	Genomic DNA
GEL	Genomics England
GSH	Glutathione
GSSH	Oxidised glutathione

H_2O_2	Hydrogen peroxide
HA	Hereditary ataxia
HD	Huntington's disease
HLOD	Heterogeneity logarithm of the odds
hnRNPK	Heterogeneous nuclear ribonucleoprotein K
IGF	Insulin growth factor
ILOCA	Idiopathic late-onset cerebellar ataxia
IRD	Infantile Refum disease
IRR	In-Repeat-Reads
IUGR	Intrauterine growth restriction
kB	Kilobases
kDa	Kilodaltons
LOCA	Late-onset ataxia
MCB	Monochlorobimane
MHC	Major Histocompatibility Complex
miRNA	MicroRNA
MPTP	Mitochondrial permeability transition pore
mRNA	Messenger RNA
MSA	Multiple Systems Atrophy
MSA-C	MSA with predominantly cerebellar features
MSA-P	MSA with predominantly parkinsonian features
MSP	Major sperm protein
mtDNA	Mitochondrial DNA
NADH	Nicotinamide adenine dinucleotide
NALD	Neonatal adrenoleukodystrophy
NCS	Nerve conduction study
NGS	Next-generation-sequencing
NHNN	National Hospital for Neurology and Neurosurgery
NHS	National Health Service
NK	Natural killer
NMD	Nonsense-mediated decay
NO·	Nitric oxide
NOP56	Nuclear protein 56
OH	Orthostatic hypotension
OMIM	Online Mendelian Inheritance in Man
ONT	Oxford Nanopore Technology
OPC	Oligodendrocyte progenitor cells
ORA	Over-Representation Analysis
OXPHOS	Oxidative phosphorylation
PacBio	Pacific Biosciences
PBD	Peroxisomal biogenesis disorders
PCR	Polymerase-chain reaction
PD	Parkinson's disease
PED	Peroxisomal enzyme deficiencies
PKCδ	Mitochondrial protein kinase C δ

PNS	Peripheral nervous system
PolyQ	Polyglutamine
PSP	progressive supranuclear palsy
RAN	Repeat-associated non-ATG
REM	Rapid eye movement
RBP	Regulatory Binding Proteins
RCDP	Rhizomelic Chondrodysplasia Punctata
RD	Retinal dystrophy
RFC1	Replication Factor C Subunit 1
RNA	Ribonucleic acid
ROC	Receiver operating characteristic curve
ROS	Reactive oxygen species
RPPCR	Repeat-primed PCR
SCA	Spinocerebellar Ataxia
SINES	Short, interspersed elements
SMRT	Single-molecule real-time
SND	Striatonigral degeneration
SNP	Single nucleotide polymorphism
SRSF2	Rich Splicing Factor 2
STR	Short random repeats
TMRM	Tetramethylrhodamine
TPM	Transcripts per million
UCL	University College London
VAPA	VAMP Associated Protein A
VAPB	VAMP Associated Protein B
VLCFA	Very-long-chain-fatty-acids
VVOR	Visually enhanced vestibular-ocular reflex
WebGestalt	WEB-based Gene SeT AnaLysis Toolkit
WES	whole-exome-sequencing
WGS	whole-genome-sequencing
ZSD	Zellweger Spectrum Disorder

1. Introduction

1.1. Rare Mendelian diseases

Rare diseases that result from the altered function of single genes can have a debilitating effect on the quality of life of sufferers, and typically affect fewer than 1 in 2,000 people in Europe, collectively affecting millions worldwide (4, 5). Due to the rare nature of these diseases, effective therapies and treatments are also comparatively rare and many cases go molecularly unsolved. Estimates for the number of rare diseases are between 6,000 and 7000 from the Online Mendelian Inheritance in Man (OMIM) catalogue, with an estimate of between 7000 and 15,000 rare disease-causing genes reported (6, 7). However, greater than 3500 rare diseases are yet to be molecularly diagnosed, which is in part due to the unavailability of large numbers of affected individuals and relatives, locus heterogeneity and a lack of access to cutting-edge technologies (8).

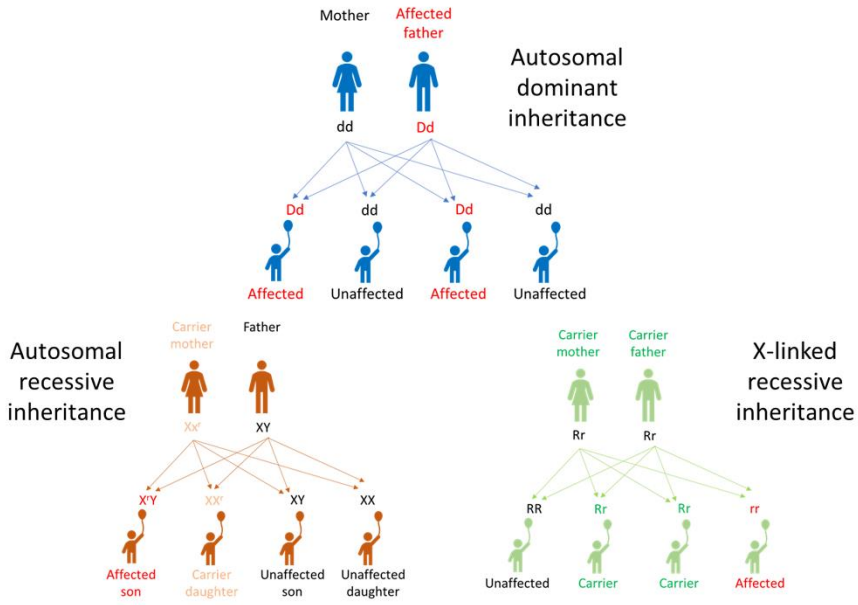


Figure 1.1 Patterns of Mendelian inheritance. Autosomal dominant inheritance which denotes a 50% risk of an affected child, recessive inheritance which denotes a 25% risk of an affected child and 25% risk of a carrier child and X-linked recessive inheritance which denotes a 50% risk of a carrier child and 25% risk of an affected child, following classical pattern of Mendelian inheritance.

1.1.1. Next-generation-sequencing in Mendelian disorders

The landscape of genetic research into rare diseases has changed and the rate of causal gene discovery accelerated by the advent of next-generation-sequencing (NGS) technologies. Whole-exome sequencing (WES) and whole-genome-sequencing (WGS) have paved the way for gene discovery and provide unbiased, powerful methods for determining genetic alterations with established and standardised pipelines. WGS allows for the generation of a map of genome variation, that is high-quality and depends on the depth of coverage obtained from the sequencing technology utilised. These technologies include short-read WGS that yields low error rates of 0.1-0.5% and paired-end 150 bp reads, long-read sequencing which yields 10-100kb reads, but with much higher error rates (10-15 %) and linked-read WGS which also provides short-reads (9). WES allows for the sequencing of all protein-coding regions of the human genome, with the exome thought to harbour approximately 85% of all disease-causing mutations (10), and a high proportion of all monogenic diseases (11).

Mendelian diseases, defined by mutations in a single genetic locus, are inherited in either a homozygous, compound heterozygous or heterozygous pattern (figure 1.1) (12). The predicted mode of inheritance influences the approach used and selection and quantity of individuals analysed. Autosomal recessive inheritance predicts a 25% chance of having an affected child, 50% chance of carrier status and 25% chance of not being affected. One sibpair is often required to identify the disease-causing gene. X-linked inheritance can also be identified with NGS, with a 25% chance of being a carrier and 25% chance of being unaffected in females, and a 25% chance of being affected and 25% chance of being unaffected in males. Autosomal dominant inheritance dictates a 50% chance of being affected and 50% chance of being unaffected.

NGS, on top of providing a more low-cost and efficient workflow for genetic analysis, also pioneered reverse phenotyping. This is the process of retrospective clinical review of patients and their families, that often takes place when a pathogenic mutation in a gene is identified that was previously linked to a different phenotype (13). Around 25% of reported mutations in disease-causing genes have been reported to associate with a phenotype that, in retrospective studies, matched the clinical phenotype of the patient (8). This shows the heterogeneity in clinical presentation of patients and further defines the clinical spectrum of genetically caused disease.

1.2. Movement disorders

Movement disorders are defined as neurological syndromes that cause either a lack of or excess voluntary and autonomic movement that is unrelated to spasticity or weakness (14). Excess voluntary or autonomic movement is characterised by dyskinesias (unnatural movements), hyperkinesias (excessive movement) and abnormal involuntary movement. A lack of voluntary, or spontaneous,

involuntary movement that is independent of external stimuli, is characterised by bradykinesia (slow movement), hypokinesia (reduced amplitude of movement) and akinesia (the loss of movement) (15).

1.2.1. Neurodegenerative movement disorders

Neurodegenerative movement disorders are characterised by the degeneration of neurons, leading to neuronal loss and are generally age dependent (16). Neurodegenerative movement disorders typically lead to the accumulation of various protein aggregates within the brain, resulting in motor symptoms (17). Neurodegenerative movement disorders encompass a range of movement disorders including Parkinson's disease (PD), multiple systems atrophy (MSA), Huntington's disease (HD), and multiple hereditary ataxias including Friedreich's ataxia (FRDA). Mendelian disorders, where inheritance patterns show a single causal gene account for a number of neurodegenerative movement disorders including numerous forms of PD, HD, and several hereditary ataxias. However sporadic forms of neurodegenerative movement disorders also account for complex disorders, involve epigenetic, environmental, and genetic interaction that include progressive supranuclear palsy (PSP) and PD (18-20).

1.2.2. Ataxia

Ataxia is a physical finding, commonly involved in movement disorders, that involves impaired coordination of voluntary movement. It is usually a component of multiple presenting symptoms. The onset of ataxia can vary, with insidious onsets that have a slower, chronic progression (e.g. spinocerebellar ataxia), acute onsets that have a rapid progression, resulting from cerebellar haemorrhage, infarction or infection, and also subacute onset that results from immune disorders (21). Cerebellar dysfunction, along with impaired proprioceptive or vestibular afferent cerebellar input commonly underlies ataxia. Cerebellar lesions are often linked to the ataxic presentation. Diffuse cerebellar lesions cause generalised symmetric ataxic symptoms whereas lateralised lesions often cause ipsilateral ataxic symptoms (21). Appendicular (limb) ataxia is caused by lesions in the cerebellar hemisphere, whereas truncal and gait ataxia results from lesions of the vermis. Gait ataxia, vertigo and disequilibrium are caused by vestibulocerebellar lesions (21).

The cerebellum, cerebellar afferent and efferent pathways, the proprioceptive sensory pathway, and the vestibular system are all implicated in ataxia pathophysiology (figure 1.2). The midline cerebellum and cerebellar hemispheres separate the cerebellum, with lesions in each independent area resulting in varying presentations of ataxia. Gait and truncal ataxia is usually associated with midline cerebellar damage, whereas ipsilateral cerebellar ataxia is associated with unilateral cerebellar hemisphere damage (21).

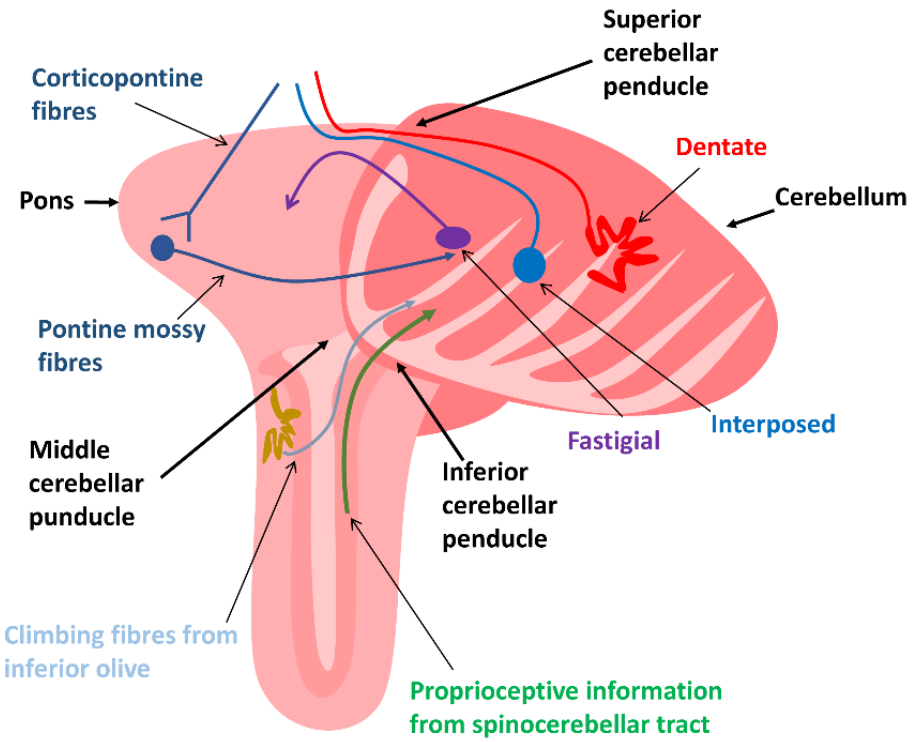


Figure 1.2: Neuroanatomy of the cerebellum showing afferent and efferent connections. Pontine mossy fibres through the middle cerebellar peduncles and climbing fibres from the inferior olives through the inferior cerebellar peduncles are the main afferent connections in the cerebellum. Efferent connections originate from the dendate nucleus and other deep cerebellar nuclei deep including the rostral fastigial nucleus that receives input from the vestibular nuclei and the interposed nucleus which is responsible for coordinating antagonist/agonist muscle pairs, that go through the superior cerebellar peduncles.

Clinically there are many terms used in describing ataxic presentation. Gait ataxia involves incoordination in the lower limbs and can be accompanied by an increase disturbance in the absence of visual cues such as the dark, which could suggest vestibular or sensory contribution. Truncal ataxia commonly occurs from midline cerebellar lesions and presents with truncal instability. Sensory ataxia is predominantly reflected in disturbed gait and can also be seen in a positive Romberg sign, high-stepping gait when walking due to motor weakness, feet-slapping gait and pseudo-athetosis, which describes random finger movements from outstretched hands, when sensory neuropathy also accompanies. Dysmetria describes the overshooting (hypermetria) or undershooting (hypometria) of targeted movement and is often found using the finger-chasing test. Ocular dysmetria describes when the eyes switch objects, hypometric saccades are when the eyes require a second move to capture the object and hypermetric saccades are when overshooting focus needs to be corrected when focusing on an object. Nystagmus describes when eyes make uncontrolled, repetitive movements, occurring up and

down, side to side, or in a circular pattern. Anterior vermis lesions usually cause upbeat nystagmus, foramen magnum lesions can cause downbeat nystagmus and lateral gaze-evoked nystagmus results in a slow drift towards midline vision, followed by fast saccades. Dysarthria defines irregular, slow speech, often with hesitation, and is prevalent in ataxic disorders. Additional ocular disturbances such as square-wave jerks, opsoclonus and ocular flutter can also occur in cerebellar disease (21).

1.3. Inherited ataxic disorders

Ataxic disorders, or diseases involving ataxia as a prominent feature are commonly inherited. The hereditary cerebellar ataxias encompass a range of highly heterogeneous neurodegenerative disorders that include Autosomal Dominant Cerebellar Ataxia (ADCA), also known as the spinocerebellar ataxias (SCAs), Autosomal Recessive Ataxias (ARCAs) and X-linked disorders. Other disorders including cerebellar ataxia, vestibular areflexia, neuropathy syndrome (CANVAS), Charcot-Marie-Tooth (CMT) disease and MSA, involve ataxia as a main feature.

The advancement of neurological genetics due to NGS has enhanced the understanding of the ataxias, specifically those that are inherited, such as the inherited cerebellar ataxias. However, there are sporadic ataxias that still remain undiagnosed.

1.3.1. Inherited ataxias

The inherited cerebellar ataxias are a group of disorders that share the clinical feature of ataxia. They encompass the hereditary ataxias, that can be transmitted in an autosomal dominant, autosomal recessive and X-linked fashion. They also span a range of inherited disorders that feature ataxia as a prominent feature including CANVAS.

1.3.1.1. Autosomal dominant cerebellar ataxia

ADCAs are a subset of hereditary ataxias that are inherited in an autosomal dominant fashion, they are often referred to as spinocerebellar ataxia (SCA). The SCAs have a global prevalence of 3 in 100,00 (22) and are caused commonly by CAG nucleotide expansion repeats, therefore involving the toxic polyglutamine (PolyQ) protein (23). SCA3 is the most prevalent globally, except for in Cuba where SCA2 is most frequent and Venezuela where SCA7 is most frequent due to a strong founder's effect (24-26).

Harding's classification of SCA details the symptomatic presentation and clinically classifies accordingly into ADCA type I, ADCA type II and ADCA type III (figure 1.3) (27). The physical presentation and age of onset of SCAs has overlap with ADCA type III presenting with 'pure cerebellar ataxia symptoms' that involves SCAs 5, 6, 11, 23, 26, 30, 37, 41 and 45 (28, 29). SCAs with additional signs including pyramidal, extrapyramidal, cognitive impairment and ophthalmoplegia are classified as

ADCA type I and includes SCA1-4, 8, 10, 12-14, 15, 17-22, 25, 27, 28, 31, 32, 34-37 38, 42-44, 46, 47, dentatorubral-pallidoluysian atrophy (DRPLA) and ataxia with DNA methyltransferase 1 (*DNTM1*) (28, 30-32). SCA 7 is the only disorder in ADCA type II which presents with pigmentary macular degeneration (28).

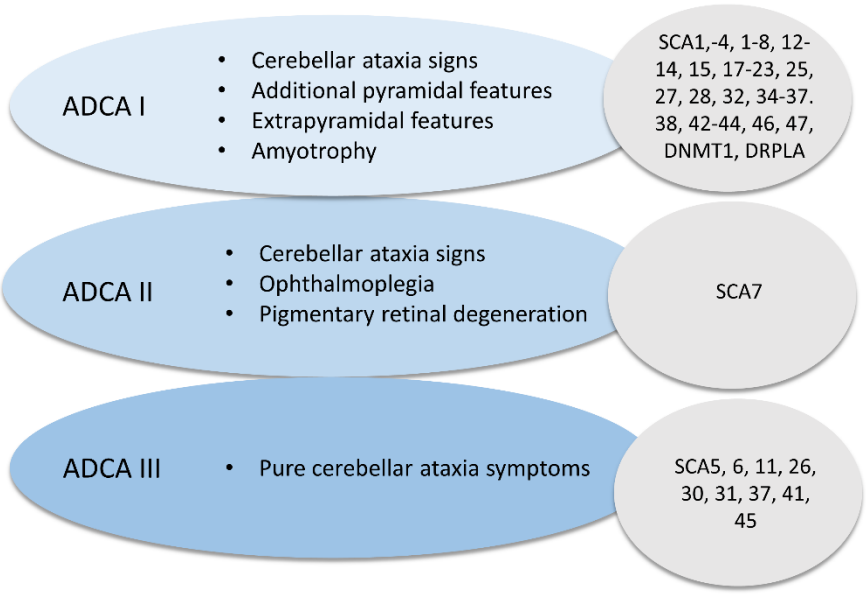


Figure 1.3: Harding's classification of spinocerebellar ataxia. SCAs are classified based on their symptom presentation into either ADCA I, ADCA II or ADCA III. *ADCA, autosomal dominant cerebellar ataxia. SCA, spinocerebellar ataxia.*

Multiple SCAs have additional characteristic features which aid in their clinical diagnosis. SCAs 17, 27 and DRPLA all present with chorea, SCAs 12, 15 and 27 all have upper limb postural tremor and SCA14 often presents with myoclonus. Table 1.1 details a non-exhaustive list of additional features seen within the SCAs. A new classification of SCA has been proposed by The International Parkinson and Movement Disorders Task Force where SCAs are categorised into pure or relatively pure ataxia, and complex ataxia (33).

Clinical Features	**Implicated SCA Subtype.**
Chorea	17, 27, DRPLA
Cognitive Impairment	2, 8, 13, 17, 19/22, 21, 36, 44, DRPLA
Dystonia	3, 14, 17, 20, 35
Face/Tongue fasciculation	36
Hearing Loss	31, 36, ataxia with DNTM1
Ichthyosiform plaques	34
Myoclonus	14
Narcolepsy	Ataxia with DNTM1
Ophthalmoplegia	2, 3, 28, 40
Parkinsonism	2, 3, 10, 14, 17, 19/22, 21
Psychiatric Symptoms	2, 17
Pyramidal signs	1, 3, 7, 8, 10, 14, 15, 17, 35, 40, 43
Seizures	10, 19/22, DRPLA
Tremor	12, 15, 27
Visual Impairment	7

Table 1.1. Clinical features of SCA, and associated subtypes (34). *DRPLA, dentatorubral-pallidoluysian atrophy; DNMT1, DNA methyltransferase 1, gene responsible for autosomal dominant cerebellar ataxia- deafness and narcolepsy*

Advances in NGS have enabled the molecular diagnosis of a large proportion of ADCAs, with exome sequencing having an average diagnostic rate of 36%, a significant improvement on the 17% yielded from targeted sequencing panels (35). Table 1.2 outlines the known genes implicated in the pathogenesis of specific SCAs.

Disease Subtype	***Gene/Locus***
SCA1	*ATXN1*
SCA2	*ATXN2*
SCA3	*ATXN3*
SCA4	*16q22.1*
SCA5	*SPTBN*
SCA6	*CACNA1A*
SCA7	*ATXN7*
SCA8	*ATXN8*
SCA9	Not assigned
SCA10	*ATXN10*
SCA11	*TTBK2*
SCA12	*PPP2R2B*
SCA13	*KCNC3*
SCA14	*PRKCG*
SCA15	*ITPR1*
SCA16	*ITPR1*
SCA17	*TBP*
SCA18	7q22-q32
SCA19/22	*KCND3*
SCA20	11q12
SCA21	*TMRM240*
SCA23	*PDYN*
SCA25	*SCA25*
SCA26	*EEF2*
SCA27	*FGF14*
SCA28	*AFG3L2*
SCA29	*ITPR1*
SCA30	4q34.3-q35.1
SCA31	*BEAN1*
SCA34	*ELOVL4*
SCA35	*TGM6*
SCA36	*NOP56*
SCA37	*1p32*
SCA38	*ELOVL5*
SCA40	*CCDC88C*
DRPLA	*ATN1*
SCA42	*CACNA1G*
ADCADN	*DNTM1*

Table 1.2: All currently known ADCA disease subtypes and corresponding genes. *ADCADN – Autosomal Dominant Cerebellar Ataxia Deafness and Narcolepsy.*

1.3.1.2. Autosomal recessive cerebellar ataxia

Cerebellar ataxias with a recessive mode of inheritance are classified as ARCAs (table 1.3). Much like the other cerebellar ataxia subtypes there is a vast heterogeneity with an estimated global prevalence of around 2.2–7 per 100,000 (36). However, unlike dominant disorders they commonly present sporadically (37). The most common ARCA is Friedreich's Ataxia (FRDA) which in the West has an estimated prevalence of between 1 in 20,000 and 1 in 725, 000 (38). FRDA is caused by a homozygous GAA trinucleotide repeat located in intron 1 of *FXN*, which codes for frataxin, on chromosome 9q13 (20). It can also be caused by a combination of GAA expansion on one allele and a heterozygous point mutation in the other in 2-5% of causes (39). The normal GAA repeat range is 14 -34, with pathogenicity occurring between 90-1700 repeats. However, repeats in the upper end of the normal range can expand in transmission, forming an intermediate range of up to 90 repeats or into disease causing alleles (40). FRDA arises from the reduced expression of frataxin, which is involved in cellular iron homeostasis. Downstream effects include respiratory chain complexes and Kreb cycle component disturbances due to a reduction in bound ferrous iron by frataxin; dysregulating synthesis of iron-containing molecules such as iron-sulphur clusters and haem by mitochondria, and ultimately causing cell death (41-43).

FRDA has a very similar clinical presentation to hereditary Ataxia with Vitamin E Deficiency (AVED) which presents primarily Vitamin E plasma concentrations of 3 µmol/l and below (normal range is 7-33 µmol/l) combined with abetalipoproteinemia and intestinal fat malabsorption (44). AVED is caused by mutations in the gene α-tocopherol transfer protein (*TTPA*) which aids vitamin E incorporation into circulating lipoproteins, therefore causing reduced vitamin E levels (45). Other ARCAs include autosomal recessive spastic cerebellar ataxia of Charlevoix-Saguenay (ARSACS) caused by the *SACS* gene, with over 70 mutations now implicated in pathogenesis, and Ataxia telangiectasia (AT). AT has an estimated prevalence of 1 in 300,000 and involves the gene AT mutated gene (*ATM*) which encodes a phophatidylinosytol-3 kinase protein that responds to DNA damage by phosphorylating vital substrates in DNA repair or cell control pathways (46-48).

Disorder	**Gene**	**Protein**
Friedreich Ataxia (FRDA)	*FXN*	Frataxin
Ataxia with vitamin E deficiency (AVED)	*TPPA*	A-tocopherol transfer protein
ARSACS	*SACS*	Sacsin
Ataxia telangiectasia-like disorder (ATLD)	*MRE11*	Meiotic recombination 11
Ataxia telangiectasia (AT)	*ATM*	Serine protein kinase
Ataxia with oculomotor apraxia type 1 (AOA1)	*APTX*	Aprataxin
Ataxia with oculomotor apraxia type 2 (AOA2)	*SETX*	Senataxin
Marinesco-Sjogren syndrome (MSS)	*SIL1*	Nucleotide exchange factor SIL1
Sensory ataxic neuropathy, dysarthria, and ophthalmoparesis/mitochondrial recessive ataxia syndrome) (SANDO / MIRAS)	*POLG1*	DNA Polymerase subunit γ-1
Autosomal recessive cerebellar ataxia type 1 (ARCA1)	*SYNE1*	Nesprin-1
Spastic paraplegia type 7 (HSP-SPG7)	*SPG7*	Paraplegin
Autosomal recessive cerebellar ataxia type 2 with coenxyme Q10 deficiency (ARCA2)	*CABC1/COQ8A*	Chaperone-activity of bc1 complex-like/Coenzyme Q8A
Autosomal recessive cerebellar ataxia type 3 caused by mutations in ANO10 (ARCA3)	*ANO10*	Anoctamin-10
Infantile onset spinocerebellar ataxia (IOSCA)	*C10orf2*	Twinkle
Spinocerebellar ataxia with axonal neuropathy (SCAN1)	TDP1	Tyrosyl-phosphodiesterase-1
Cataracts, intellectual disability, myopathy, short stature, childhood onset, cerebellar atrophy	SIL1	SIL1 Nucleotide Exchange Factor
Autosomal recessive spastic ataxia-4 (SPAX4)	MTPAP	Poly(A) polymerase-associated domain-containing protein 1
Salih ataxia	KIAA0226	Rundataxin
Dementia, pyramidal signs, epilepsy, tendon xanthomas, atherosclerosis, cataracts, diarrhea, elevated serum cholestanol, polyneuropathy, childhood to adult onset, variable cerebellar atrophy, cerebellar or cerebral white matter anomalies	CYP27A1	CYP27 Sterol 27-hydroxylase
Autosomal recessive spastic ataxia associated with frequent white matter changes (ARSAL)	2q33-q34	-
Spinocerebellar ataxia with blindness and deafness (SCABD)	6p21-p23	-
Childhood onset slowly progressive ataxia	11p15	-
Autosomal recessive spinocerebellar ataxia and epilepsy	17q21-q23	-
Autosomal recessive spastic ataxia (SAX2)	17p13	-
Early onset ataxia with developmental delay and failure to thrive	22q11	-

Table 1.3: Known causative genes and loci for primary most prevalent ARCA (49)

1.3.2. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome

CANVAS is a movement disorder that encompasses cerebellar ataxia as main feature. It is a slowly progressive, late-onset, and neurodegenerative disorder, that is multisystemic and variable in phenotype, with a pattern of spatial progression.

1.3.2.1. Clinical presentation

CANVAS presents as a predominantly late-onset, severe imbalance with absent visually enhanced vestibular-ocular reflex (VVOR), that suggests combined bilateral vestibular and cerebellar dysfunction. This diagnostic triad of symptoms is accompanied by non-length dependent sensory deficits, and often additional symptoms that can include autonomic failure. A chronic spasmodic cough is also often featured and is reported to start 15-29 years before onset of gait imbalance (50, 51). This diagnostic triad can remain incomplete due to phenotypic variability, with reports that the full range of symptoms is only present in approximately 2 out of 3 cases (2, 52). However, neuronopathy is a ubiquitous sign of CANVAS, as reported from a study involving a cohort of 100 genetically confirmed CANVAS patients (53). The onset range for the associated CANVAS features, excluding cough, is between 19 and 72 years of age, with a mean onset of 52 years (53).

Spatial progression from early sensory neuronal involvement to later cerebellar dysfunction and vestibular impairment has been suggested within CANVAS (53). Despite the proposed clinical criteria for CANVAS, diagnosis still proves a challenge, with a recent report of 66 confirmed cases only having 35% with an initial CANVAS clinical diagnosis (53). Postural instability is caused by each separate component of the CANVAS triad, but when combined together, they can result in severe imbalance from an impaired VVOR and this can be severely debilitating (54). Bilateral vestibular failure (BVF) and cerebellar dysfunction can be diagnosed using oculomotor testing. The combined failure of three compensatory eye movements, vestibulo-ocular reflex (VOR), optokinetic reflex, and smooth pursuit are reflective of VVOR, which yields a positive test result if compensatory eye movements are saccadic instead of smooth (55).

Neuropathy, presenting with non-length dependent sensory deficit in CANVAS, is commonly associated with hyperreflexia or normoreflexia (54). Normoreflexia is reported to result from the sparing of muscle spindle afferents (la fibres) (56, 57). Severe axonal sensory neuropathy with absent or reduced sensory nerve action potential (SNAP) in lower and upper limbs, reduced H-reflexes, and generally intact F-wave latencies have been revealed from electrophysiology studies on CANVAS patients, with electromyography commonly normal (53, 58, 59). Cerebellar ataxia often presents with ocular motor impairment encompassing downbeat nystagmus, gaze-evoked nystagmus, saccadic dysmetria and broken pursuit (52, 55, 60-63). Cerebellar atrophy has been reported from brain MRI data manifesting as predominant hemispheric atrophy in the horizontal and posterosuperior fissures and

also anterior and dorsal vermial atrophy (52, 55). Posterior cord atrophy and mild brain stem atrophy has also been reported (53). Neuropathological and otopathological studies have reported the hallmark feature of ganglionopathy (neuronopathy) that results in the vestibular areflexia, dysautonomia and peripheral sensory impairments seen in the disease (61, 64).

CANVAS can be difficult to diagnose when there are delays with the manifestation of one of more of the core triad of symptoms (54). Additionally, ataxia caused by impairments in vestibular, cerebellar and proprioceptive systems is not specific to just CANVAS, but multiple diseases that include MSA with predominantly cerebellar features (MSA-C), FRDA, and various SCAs such as SCA1, SCA3 and SCA6 (54). Some features of CANVAS, including bilateral vestibular failure and neuropathy require additional and deliberate specialised tests for confirmation, that are prompted by various clues from the clinical presentation (figure 1.4).

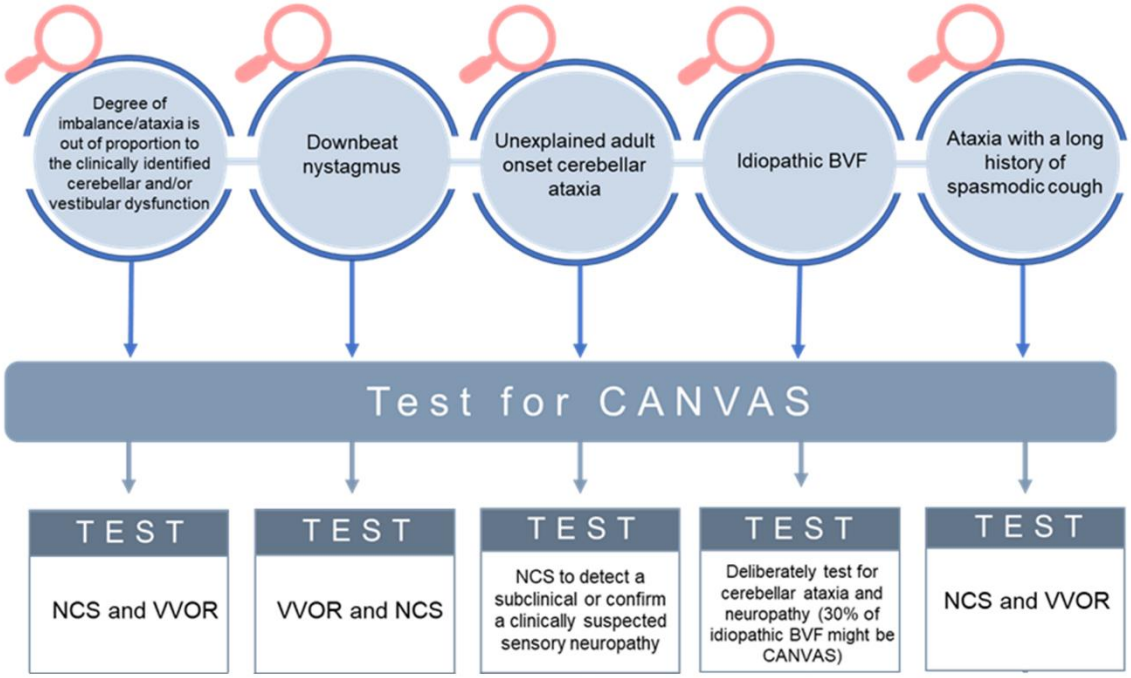


Figure 1.4: Signs and clues that lead to potential CANVAS diagnosis. NCS, nerve conduction study; *VVOR, visually-enhanced vestibulo-oculo reflex; BVF, bilateral vestibular failure*. Figure designed by R.S for publication in Sullivan et al., Curr. Opin. Neurol 2021 (54).

1.3.2.2. Treatment of CANVAS

CANVAS symptom management and symptom treatment is suggested to work best using a multidisciplinary approach. This includes managing the symptoms of autonomic dysfunction, whilst providing gait training, vestibular rehabilitation, speech and language therapy, pain management and providing devices to aid mobility issues (53, 54).

1.4. Repeat expansions in ataxic disorders

Repeat expansion mutations account for more than 40 diseases and are caused by expansions of sequence repeats within the human genome. They are implicated in HD, myotonic dystrophies DM1 and DM2, amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) caused by chromosome 9 open reading frame 72 (*C9ORF72*), FRDA and Fragile X syndrome, to name but a few. Repeat expansion disorders stem from polymorphic repeats within the population, with repeats at the higher end of the repeat range (mutable normal repeats) having the tendency to further expand and move into the pathogenic range. *de novo* repeat expansions have been shown to occur in unaffected families in HD, which has a frequency of 10% for *de novo* cases (65, 66).

Many repeat expansion disorders show a genotype-phenotype correlation, with longer repeat-length resulting in a more severe phenotype and commonly an earlier onset. This is seen in polyglutamine diseases, where CAG repeat expansions result in abnormally long regions of glutamine in disease proteins (figure 1.5). FRDA, DM1, Fragile X syndrome and many of the non-polyQ SCAs also show a genotype-phenotype correlation, with a more severe disease and earlier onset (67). The mode of inheritance is variable for repeat expansion disorders; however the majority are autosomal dominant. There are autosomal recessive repeat expansions, including Fragile X-syndrome and FRAXE mental retardation which are X-linked recessive, and FRDA and progressive myoclonic epilepsy type 1 (EPM1) which are autosomal recessive. Within the inherited ataxias, 45% of autosomal dominant ataxias, including multiple SCAs, are trinucleotide repeat expansion disorders, caused by expanded CAG. These include SCA1, SCA2, SCA3, SCA6, SCA7, SCA17 and dentatorubral-pallidoluysian atrophy (68, 69).

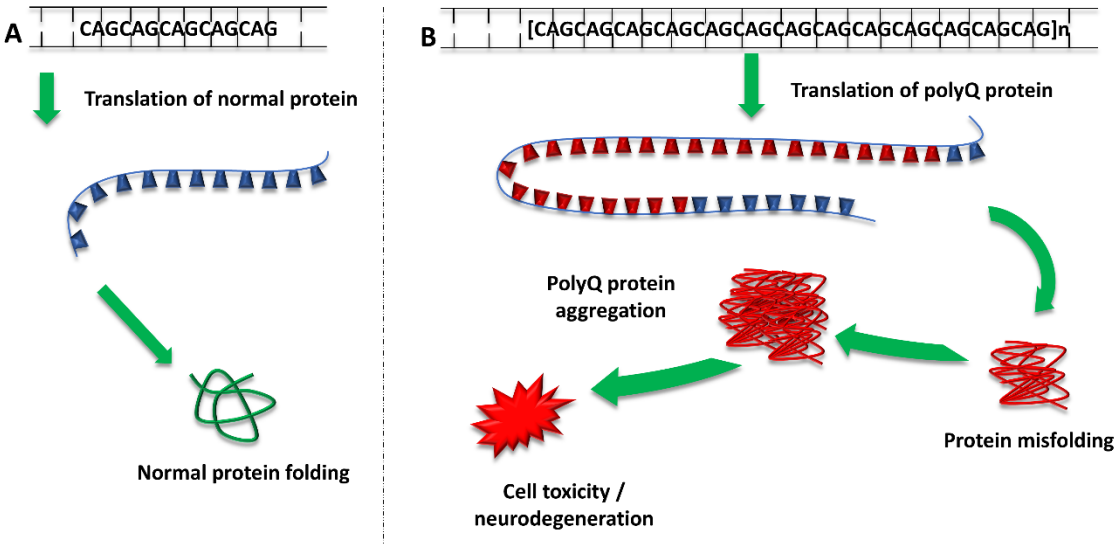


Figure 1.4: PolyQ-expansion repeat mechanism. A, normal translation of polyglutamine repeat within the normal range that results in a normal protein transcript and protein folding. B, A pathogenic polyglutamine expansion repeat length results in translation of an abnormal polyQ repeat. This leads to protein misfolding, with misfolded polyQ proteins aggregating and causing various downstream dysfunctions that result in cell toxicity and degeneration. *PolyQ, polyglutamine*. Figure made by RS for publication by Sullivan et al., Journal of Neurology, 2019 (3).

1.4.1.Non-coding repeat expansions and RNA foci formation

Multiple SCAS including SCA8, 10, 12, 31 and 37 are non-coding repeat expansion disorders. The hallmark of non-coding repeat expansions is a transcribed nuclear accumulation of repeat regulatory binding proteins (RBP) (figure 1.6). These accumulations are generated by expanded RNA aggregates, caused by an RNA toxic gain-of-function, forming nuclear ribonucleoprotein foci that sequester RBP and other cellular factors (70). There are several factors that influence the size and shape of the RNA foci and the colocalisation of proteins, and these include the tissue type, the type of proteins expressed and the expression level of the RNA transcript (71). This accumulation and formation of RNA foci leads to RNA toxicity. RNA foci are thought to be stuck in the nucleus due to the molecules of abnormal RNA produced that are covered in bound proteins (71). An intronic ATTCT repeat expansion in *ATXN10* has been implicated in SCA10, with cytoplasmic and nuclear foci forming in SCA10 cells and pathogenicity occurring in the range of 800–4500 repeating units (72). A SCA10 transgenic mouse model also shows aggregated AUUCU RNA that is resistant to degradation, formed from spliced ATXN10 pre-mRNA that lacks intron 9. This AUUCU RNA aggregate binds to splicing factor heterogeneous nuclear ribonucleoprotein K (hnRNPK), which results in its sequestration, subsequent loss of function, and accumulation of mitochondrial protein kinase C δ (PKCδ), ultimately resulting in

SCA10 cell apoptosis (73). SCA36 has been linked to an intronic GGCCTG hexanucleotide repeat expansion in nuclear protein 56 (*NOP56*), with RNA foci found in patient lymphoblastoid cells and a reduction in the microRNA (miRNA) MIR1292 (74). MiRNAs are small noncoding RNAs that inhibit translation of target mRNAs, regulating gene expression. The induction of RNA foci in *NOP56* caused sequestration of RBP Serine and Arginine Rich Splicing Factor 2 (SRSF2) and a toxic RNA gain-of-function mechanism, that resulted in SCA36 (74). Toxic gain of function promotes irregular interactions of the protein by increasing the interaction of the mutated protein with its usual binding partners and promoting protein misfolding or aggregation. SCA8 also has been found to have a toxic gain-of-function mechanism as part of its pathophysiology and is caused by a bidirectionally expressed CTG CAG repeat expansion, seen in a SCA8 mouse model. This results in an accumulation of [CUG]n transcript ribosomal inclusions that bind to RBP Mbn11 and cause downstream alternative splicing, leading to the phenotype (75). DM1 locus and HDL2 also are associated with the bidirectional expression of sense [CUG]n and antisense [CAG]n, which also result in ribonuclease inclusions (76, 77). SCA37 also has been linked to an intronic ATTTC repeat, which has been found to dysregulate reelin adaptor protein disabled coding (*DAB1*) gene expression, leading to altered splicing and an upregulation of cerebellar reelin-DAB1 signalling (78).

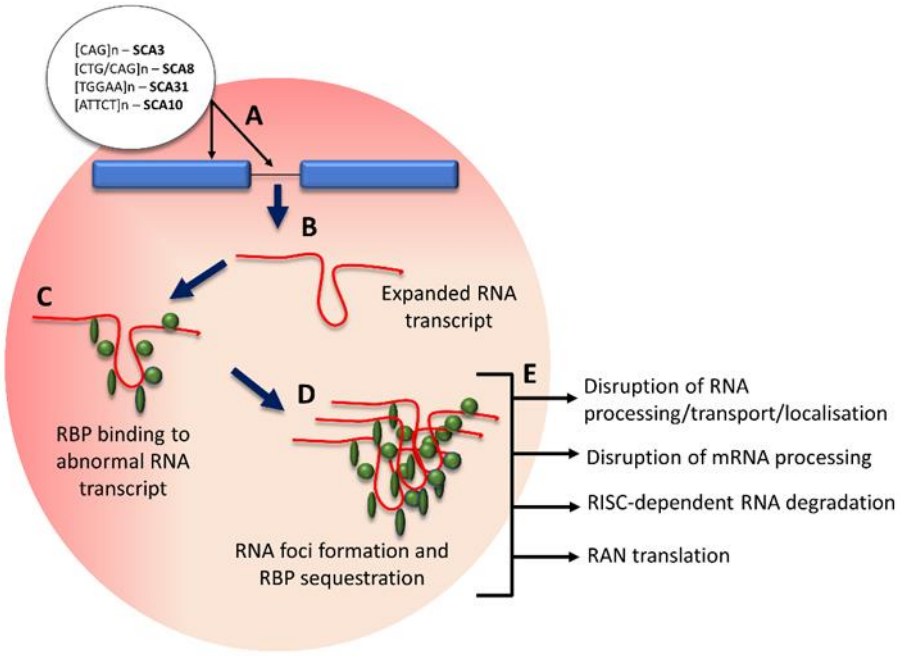


Figure 1.5: RNA foci formation. A, intronic and exonic pathogenic repeat expansions implicated in SCAs. B, expanded repeat is transcribed into expanded mRNA/pre-mRNA. C, regulatory binding proteins attach to abnormal mRNA transcript. D, Abnormal transcripts aggregate and RBPs are sequestered. E, downstream effects of RBP sequestration on various processes. *RBP, regulatory binding protein, RAN, repeat-associated non-ATG, SCA, spinocerebellar ataxia.* Figure made by R.S for publication by Sullivan., et al., Journal of Neurology, 2019 (3).

1.4.2. Molecular diagnosis of repeat expansion disorders

WES and WGS has revealed the vast genetic heterogeneity of the inherited ataxias, identifying conventional disease-causing mutations including deletions, insertions, and single nucleotide changes. High-throughput short-read sequencing technologies are currently limited by the length of read, approximately 150 base pairs (bp), which makes these technologies inefficient to pick up pathogenic repeat expansions that can be thousands of repeating units in length.

The standard workflow for detecting repeat expansions includes repeat-primed PCR (RPPCR), followed by fragment analysis, which is used diagnostically to detect the *C9ORF72* repeat expansion causative of Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). These techniques are frequently misinterpreted due to a high rate of false negatives and positives due to both indels in the repeat flanking regions and varying protocols between diagnostic laboratories (3, 79). Repeat length size cannot be estimated using RPPCR or fragment analysis and Southern blotting is often employed as a secondary test which is considered the gold standard for repeat length size estimation. Southern blotting also has its limitations, including the amount of genomic DNA required, which ranges from 5 to 10 µg. Its accuracy of size estimation is often affected by somatic heterogeneity, which it is susceptible to (80). Tethering PCR has been suggested to negate the need to perform Southern blotting, and has been tested in SCA1, 2, 3, 6 and 7 (81).

Oxford Nanopore sequencing and PacBio single-molecule real-time (SMRT) are two forms of long-read sequencing which have been proposed as the new cutting-edge technology for detecting repeat expansions. They both enable sequencing of DNA more than 10,000 bp in length. PacBio uses real-time imaging of nucleotides, that are fluorescently tagged, during DNA synthesis that uses DNA polymerase as a reaction facilitator (82). Oxford Nanopore identifies unlabelled nucleotide 5'monophosphate molecules by using a protein nanopore covalently attached to an adaptor molecule (83). It has an accuracy of 99.8%, whilst PacBio reads have an average higher error rate of 15% (83, 84). Oxford Nanopore are limited with their MiniION sequencer, which has a reported base-calling error rate of 38.2% (85). ExpansionHunter is a bioinformatic package developed by Illumina that utilises WGS data. It has the capability to detect the size of repeats of varying size from PCR-free data, which include very large pathogenic expansions that are longer than the read length (86). It has the potential to revolutionise the workflow for detecting repeat expansions by reducing the length of time required and reducing the errors and inaccuracy of existing methods of detection.

1.5. Mitochondrial dysfunction in the ataxias

Mitochondria have been demonstrated as a crucial player in the pathogenesis and progression of neurodegenerative diseases including Alzheimer's disease, PD, HD, and neurodegenerative ataxic disorders including CMT disease and FRDA (87). Mitochondria are key to critical cell functions

including synthesising adenosine triphosphate (ATP), apoptosis modulation, inflammasome activation, ROS production and ferroptosis. They therefore have a rigid response system to stressors that could affect cell-viability (87).

1.5.1. Mitochondrial biogenesis

The process of mitochondrial biogenesis details existing mitochondria growth and division to form new mitochondria, which leads to increased cellular mitochondrial mass. Environmental stresses including low temperature, caloric restriction and physical exercise influences mitochondrial biogenesis, as well as cell division, differentiation, and renewal.

Mitochondrial biogenesis involves replication of mitochondrial DNA (mtDNA) and the synthesis of various proteins that are encoded by both nuclear and mitochondrial genomes. mtDNA is a double stranded molecule that contains 37 genes encoding 13 subunits of complexes I, III, IV and V of the electron transport chain (88). Approximately 1000–1500 proteins are required for correct mitochondrial biogenesis to occur. A large proportion of these proteins are encoded in the nucleus, and therefore, a mechanism exists for the correct import of these proteins and assembly to ensure mitochondrial shape and function is correct. Cytosolic mRNAs are translated to precursor proteins which are formed with a positively charged N-terminal that can form a basic, amphiphilic α-helix. They cross the inner and outer mitochondrial membranes via protein translocases. Translocation is driven by matrix heat-shock protein 70 (Hsp70) and the mitochondrial membrane potential. A matrix protease then cleaves the presequence and the important protein is folded. There are four major membrane protein translocase complexes, the translocase of the outer membrane (TOM), translocase of the inner membrane (TIM), presequence translocase-associated motor (PAM) and sorting and assembly machinery (SAM) (88).

Mitochondrial biogenesis is regulated by peroxisome proliferator activated receptor-gamma (PPARγ) and transcriptional coactivator-1alpha (PGC-1α). PGC-1α, activates various transcription factors, such as nuclear respiratory factors 1 and 2 (NRF-1 and NRF-2), which promote the expression of mitochondrial factor A (Tfam), which drives the transcription and replication of mtDNA (89). Various studies have confirmed the contribution of PGC-1α to mitochondrial biogenesis regulation, which have found that it also induces uncoupling protein (UPC) 2. It also been found to bind to NRF-1, co-activating NRF-1's transcriptional function on the promotor for Tfam (90). PGC-1α expression has also been found to increase in studies linking exercise to an increase in mitochondrial biogenesis (91). PGC-1β also has been implicated in mitochondrial biogenesis, however it has not been found to be upregulated in either brown adipose tissue, or in muscles in response to cold and exercise stimuli, respectively (92).

Signalling molecules are also crucial to the activation of the transcriptional cascade that leads to mitochondrial biogenesis. AMP-activated protein kinase (AMPK) responds to acute energy crises, regulating intracellular energy metabolism (93). AMPK activity has been reported to decline during aging which could contribute to impaired intracellular lipid metabolism and mitochondrial dysfunction

(94). Nitric oxide (NO) is involved in PGC-1α gene transcription and therefore mitochondrial biogenesis (95). Calcium/calmodulin-dependent protein kinase IV (CaMKIV) has a key role in mitochondrial biogenesis regulation, however the role in its response to exercise is not yet fully understood (88). Several other signalling molecules activate PGC-1α, including silent mating type information regulation 2 homologue-1 (SIRT1) and transducer of regulated CREB (cAMP-response-element-binding protein)-binding protein (TORC) 1. Resveratrol, a phytoalexin, activates SIRT1 and has been reported to trigger mitochondrial biogenesis in skeletal muscle (96). TORC1, along with TORC2 and TORC3, strongly activate PGC-1α transcription, inducing the endogenous mRNA of PGC-1α (97). Receptor-interacting protein 140 (RIP140) is a transcriptional co-repressor, which has also been implicated in mitochondrial biogenesis regulation. It is abundantly expressed in skeletal muscle with fast-twitch, type II and mixed muscle fibres, and much more lowly expressed in slow-twitch type I muscles (98). RIP140 deficiency has been shown to result in an increased mitochondrial oxidative capacity and gene expression, whereas transgenic expression of RIP140 in skeletal muscle result in a decreased mitochondrial gene expression and an increase in oxidative muscle fibres (99). This suggests that RIP140 is similar to adipocytes in its regulation of mitochondrial oxidative metabolism. p38 Mitogen-activated protein kinase (MAPK) has been suggested to have a key role in skeletal muscle mitochondrial biogenesis in response to exercise, by stimulating PGC-1α expression (100).

1.5.2. Mitochondrial biogenesis dysfunction

Deficits in mitochondrial biogenesis have been shown to contribute to mitochondrial dysfunction in several neurodegenerative diseases including Alzheimer's disease, PD, CMT and FRDA. A major player in neurodegenerative disease is oxidative stress from reactive oxygen species (ROS) production, of which mitochondria are a major source and target of elevated ROS exposure. ROS exposure increases when mitochondrial biogenesis decreases, due to an increase in ROS production and a reduction in ROS clearance rate (101-103). ROS production can induce the intrinsic apoptotic cascade which is caused by the reduced functionality and integrity of mitochondria, triggering mitochondrial outer membrane permeabilization (MOMP) (104). The cascade involves the release of cell death modulators including apoptosis-inducing factor, endonuclease G and cytochrome c, as well as the inhibition of respiratory chain function and termination of ATP production (105). MOMP is activated by the altering of proteins in the mitochondrial permeability transition pore (MPTP) complex, resulting in an increase in MPTP permeability and cytochrome c release (106-108). Cytochrome c is involved in the initiation of a proteolytic cascade dependent on caspase 3 and induced by caspase 9, whilst AIF and endonuclease G mediate DNA fragmentation (105). ROS can also activate c-Jun N-terminal kinase (JNK), inducing a DNA damage response that involves pro-apoptotic signalling activation and subsequent cell death (109). It can also induce cytochrome C release via altering the mitochondria inner membrane phospholipid cardiolipin, which results in the binding of cytochrome C to the inner membrane, disrupting interactions between cardiolipin and cytochrome C (110).

FRDA pathophysiology involves mitochondrial biogenesis deficits due to the transcriptional silencing and reduced expression of frataxin, caused by the expanded GAA repeat in the *FXN* gene. Frataxin is a mitochondrial protein that is involved in iron-sulphur cluster biogenesis. The activity of iron-sulphur cluster containing enzymes such as the respiratory chain complexes is reduced in frataxin-deficiency (57). In model systems this results in iron accumulation, oxidative stress, and a reduction in ATP production (87). Fibroblast cells show a decrease in PGC-1α in the absence of frataxin, and subsequent reductions of NRF1, Tfam and mitochondrial copy number (111). FRDA knockout models also show mitochondrial biogenesis deficits in muscle and brain cells, as well as the RNA of peripheral blood mononuclear cells (PBMCs) (111, 112). Frataxin overexpression has been shown to cause an increase in mitochondrial biogenesis (111). Markers of oxidative stress are shown in multiple FRDA in vitro models, but the contribution of oxidative stress to FRDA pathophysiology is still not fully understood (113-116). The antioxidant response in FRDA in patient cells is perturbed when PGC-1α expression is reduced, which mirrors the reduced expression of SOD2, another antioxidant gene (117). SOD2 response to hydrogen peroxide (H_2O_2) has been shown to be restored when PGC-1α expression is triggered by AMPK and PPARγ (101).

1.5.3. Mitochondrial dynamics

Mitochondria are dynamic organelles and undergo mitochondrial dynamics as a response to environmental stimuli that cause mitochondrial fragmentation and fusing, which allows mitochondria to keep up with the cell's energetic demands (118). Disruptions to the process of mitochondrial dynamics have been shown to contribute to disease, via an imbalance between mitochondrial fusion and fission.

Mitochondrial fission is regulated by dynamin-related protein 1 (DRP1), a GTPase that is regulated by post-translational modifications that include phosphorylation at its Ser616 and Ser637 residues (119-121). Mitochondrial fission activated via DRP1 has been reported to be triggered by cyclin-dependent kinase 1 (CDK1) phosphorylating DRP1 at its Ser616 residue, which causes a translocation of Ser616-phospho-Drp1 to the outer mitochondrial membrane (122). Ser616-phospho-Drp1 then binds to mitochondrial fission protein 1 (FIS1). Mitochondrial dynamics proteins MiD49 and MiD51 also have the capability of promoting mitochondrial fission by binding to Drp1 on the outer membrane. Ser616-phospho-Drp1 binds to other phospho-proteins on the outer mitochondrial membrane, forming a ring around the mitochondria. This ring then severs the mitochondria, employing GTPase activity (87). The inhibition of fission activity via Drp1 has been shown to occur through phosphorylation of DRP1's Ser637 residue by AMPK (123).

Mitochondrial fusion is regulated by a number of proteins including mitofusins MFN1 and MFN2 which tether to the endoplasmic reticulum and promote mitochondrial fusion (83). DRP1's GTPase activity is inhibited by phosphatidic acid and cardiolipin, inhibiting mitochondrial fission, whilst promoting

fusion. Optic atrophy protein 1 (OPA1) mediates the fusion of the mitochondrial inner membrane, which is key for mitochondrial cristae formation and the regulation of mitochondrial bioenergetics (124, 125).

Disruption to these processes can lead to abnormal mitochondrial morphology, affecting function, and may be a secondary effect of impairments to other mitochondrial pathways and organelles that interact with the mitochondria. The homeostatic state of a cell affects mitochondrial dynamic processes, with elongated mitochondrial networks permitting efficient function of the complexes involved in the respiratory chain, and quick membrane potential transmission (126). Conversely, mitochondrial fragmentation allows for the transfer of mitochondria to cellular compartments with a high bioenergetic demand (127). Efficient mitochondrial fission also allows for mitophagy and the clearance of mitochondria, with efficient cell death even supported by mitochondrial fragmentation which may reduce energy production during necrosis (128).

Neurodegeneration can be caused by alterations to mitochondrial dynamics. Mutations in Mfn2 cause CMT type IIA, by resulting in impaired mitochondrial distribution and the failure of mitochondria to be transferred to nerve terminals, instead causing their accumulation in small caliber axons (126, 129, 130). Mutations in ganglioside-induced differentiation associated protein 1 (GDAP1) also cause CMT type IVA, by impairing mitochondrial fission by causing an increase in mitochondrial fragmentation. PD with underlying genetic mutations, including alpha-synuclein (α-syn), leucine-rich repeat kinase 2 (*LRRK2)*, parkin (*PRKN*) and PTEN-induced kinase 1 (*PINK2)* also involve impaired mitochondrial dynamics. Drp1 is phosphorylated by LRRK2, which has been reported to cause an increased Drp1 and Ser616-phospo-DRP1 activity in PD patient PBMCs (129, 131). Additionally knockdown of Pink1 and Parkin in neuroblastoma has been shown to lead to mitochondrial fragmentation (87). FRDA pathophysiology has been reported to include increased mitochondrial fragmentation caused by increased oxidative stress and overactivation of Ser616-phospho-Drp1, shown in frataxin-deficient yeast and fibroblast cells, and patient fibroblast cells (132).

1.5.4. Mitochondrial ferroptosis

Ferroptosis is a non-apoptotic cell death pathway that relies on iron that leads to the rupture of the plasma membrane and reduces mitochondrial size (133). It is caused by the accumulation of lipid peroxides and ROS and is combated by glutathione (GSH), an antioxidant enzyme. GSH protects against peroxidation and binds to glutathione peroxidase 4 (GPX4), and its reduction results in ferroptosis via GPX4 inhibition (87). Ferroptosis is implicated in multiple neurodegenerative diseases including FRDA which results in iron overload and subsequent ROS overproduction. In AD, ferroptosis is linked to amyloid protein precursor (APP) dysfunction and its regulation of ferritin levels (133). Increased levels of transferrin, necessary for neuronal iron uptake, are involved in PD, resulting in neuronal iron accumulation (134).

1.5.5. Mitochondrial oxidative phosphorylation

Mitochondria are the site of oxidative phosphorylation (OXPHOS) which is a chain of sequential redox reactions, facilitated by the reduction of equivalents resulting from carbon substrate oxidative degradation. This chain comprises respiratory complexes, containing mtDNA encoded subunits that shuttle electrons from complex to complex, with the help of coenzyme Q (ubiquinone, CoQ) and cytochrome c (cyt c), Figure 1.7 details the pathways involved in the OXPHOS pathway.

Complex 1 (NADH:ubiquinine oxidoreductase) is the primary site of the respiratory chain. Nicotinamide adenine dinucleotide (NADH) is reduced to NAD + H^+, generating ubiquinol ($CoQH_2$), pumping four protons into the intermembrane space and transferring two electrons (e^-) to Complex III (135). Complex II (succinate:ubiquinone oxidoreductase) runs in parallel to complex I and facilitates the transfer of electrons from flavin adenine dinucleotide ($FADH_2$) to CoQ, which are fed into complex III. It releases less energy than complex I, so no protons are pumped into the intermembrane space. CoQ also receives electrons from glycerol 3-phosphate dehydrogenase, generated from ubiquinol. Complex III then transfers two electrons from $CoQH_2$ to *cyt c*, whilst pumping four protons into the intermembrane space. *Cyt c* then shuttles the electrons onto complex IV (ferrycytochrome:oxygen oxidoreductase, *cytochrome c oxidase*, COX), the final component of the respiratory chain. It transfers electrons to the final acceptor, oxygen, which produces water.

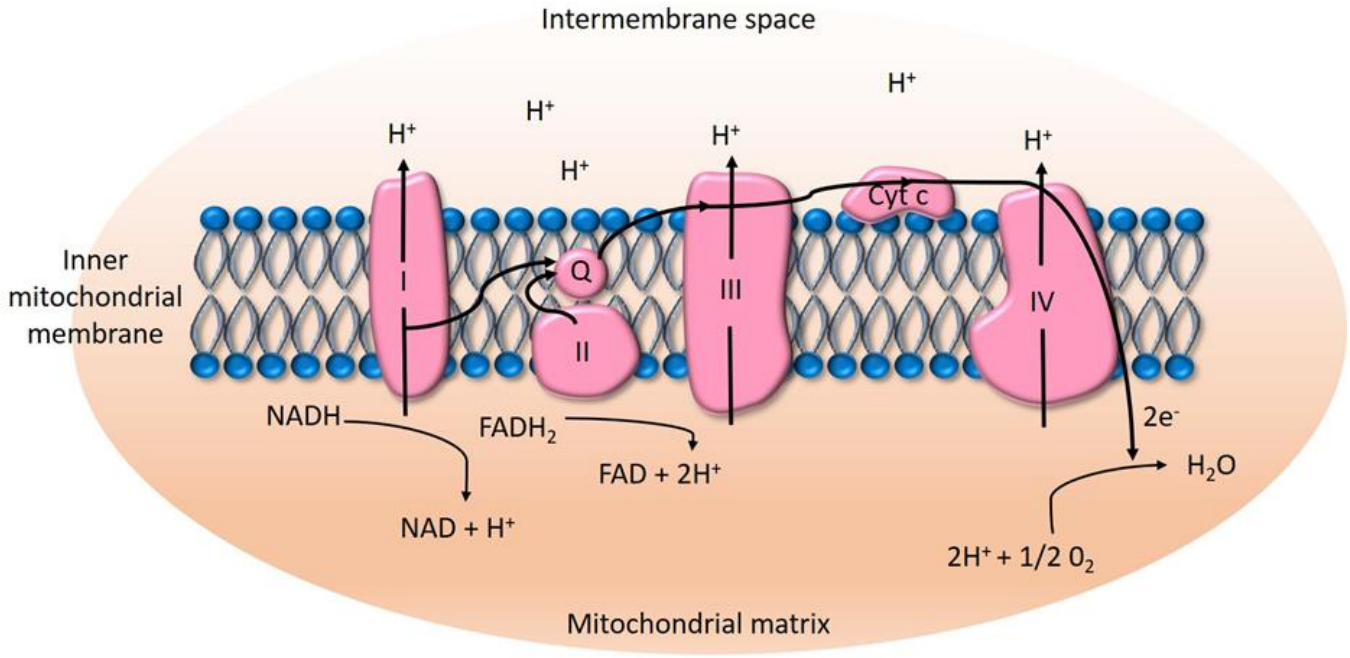


Figure 1.6: A schematic of oxidative phosphorylation (OXPHOS). Complexes I to IV of the electron transport chain (ETC) are seen in the above schematic. NADH is oxidised to NAD + H^+ by complex I, with four protons pumped into the intermembrane space and donation of two electrons. Complex II oxidises succinate to fumarate, but releases less energy than complex I, so no protons are transported to the intermembrane space, but electrons are passed through the ETC. Electron transfer flavoprotein-ubiquinone oxidoreductase (seen as 'Q' in the schematic) accepts electrons but does not cross the lipid bilayer. It is important in the β-oxidation of fatty acids and catabolism of amino acids and choline. Complex III pumps four protons into the intermembrane space and involves the electron transfer to cytochrome c acceptor. Complex IV is the final complex in the ETC and mediates the transfer of electrons to oxygen, whilst pumping protons across the membrane. *NADH, nicotinamide adenine dinucleotide; FADH, flavin adenine dinucleotide*

All these redox reactions, electron transfers and proton translocations across the inner mitochondrial membrane result in an electrochemical gradient (ΔΨ). The gradient powers the flow of protons back to complex V (ATP synthase, F_1F_0-ATPase), which catalyses ATP synthesis and subsequent phosphorylation of ADP to ATP (136). ATP generated is transported back across the inner mitochondrial membrane via the adenine nucleotide translocator (ANT), which facilitates an exchange with cytosolic ADP which is imported.

1.5.6. Mitochondrial reactive oxygen species

The mitochondrial respiratory chain produces ROS generated by a leak of electrons from the electron transport chain forming superoxide anion ($O_2^{\cdot -}$) via a reaction with oxygen. $O_2^{\cdot -}$ is then

converted into hydrogen peroxide (H_2O_2) by manganese superoxide dismutase (MnSOD, SOD2), which is further metabolised into H_2O by glutathione peroxidase (GPx). H_2O can also generate the hydroxyl radical (OH·), producing peroxynitrate (ONOO·). Antioxidants maintain the concentration of H_2O_2 in the mitochondrion and cytosol in order to limit the formation of the membrane permeant hydroxyl radical (137, 138).

The balance between ROS generation and ROS scavenging is tightly controlled, and depending on physiological conditions, oxidative stress caused by ROS overproduction has been found to initiate several cellular responses that range from activation of mitochondrial fission and autophagy to help clear damaged mitochondria and cells, protecting the spread of damage, to triggering signalling pathways involved in protecting the cell (139-141). These responses work to re-establish the 'redox homeostasis'. However, oxidative stress, when unregulated and caused by ROS generation without compensatory levels of scavenging, results in severe damage to the cell and ultimately cell death. Mitochondrial dysfunction and ROS have been suggested to be implicated in SCA2 pathophysiology, with patient fibroblast cells exhibiting increased mitochondrial oxidative stress, as well as altered mitochondrial respiratory chain enzymes and mitochondrial morphology (142). It has been reported that mitochondrial dysfunction may precede disease onset by a fibroblast line from a patient with no clinical symptoms, but with an expanded CAG repeat, showing exacerbated mitochondrial dysfunction (142).

The intermembrane space within mitochondria, under physiological conditions, maintains a redox state influenced by the reactions occurring in the respiratory chain. The more reduced the interior of the mitochondria is, the more likely primary ROS formation is to occur, for example, ROS formation is proportional to the level of reduction of NAD that occurs during the redox reaction of NAD to $NADH^+$ (143). Under pathological conditions, this steady state is perturbed by an imbalance in the reactions controlling this redox state. Issues with complex I, for example, could alter the ratio of $NAD/NADH^+$, causing a reduction of ROS-producing sites in complex I. Complex I deficits are associated with a range of pathologies, for example PD and the production of toxic oxidants. In PD dopamine undergoes non-enzymatic oxidation by superoxide which forms dopamine quinone. This can be reduced by complex I which then generates semiquinone and ultimately superoxide, which begins the cycle again (144).

1.6. Aims and hypotheses

The overall aims of this research are to further the genetic understanding of specific cerebellar ataxic diseases. Characterising genetic causes of rare, neurological diseases is pivotal for ensuring that timely molecular diagnoses are made in the clinic and appropriate patient care results. This work also aims to

optimise and further develop laboratory and bioinformatic diagnostic tools that aid these timely diagnoses and advance the technologies used within the detection of inherited rare neurological diseases.

1.6.1. Chapter 3 aims

The primary aim of this chapter was to identify the common genetic cause of late-onset ataxia CANVAS. Using a combination of flanking PCR, RPPCR and Southern blotting, an intronic, pentanucleotide, recessive repeat expansion in Replication Factor C subunit 1 (*RFC1*) was identified in 29 families with CANVAS. It was hypothesised that the general population would not harbour recessive pathogenic expansions, thus confirming the implication with late-onset ataxia and CANVAS. It was also hypothesised that the region harbouring the repeat expansion would be highly polymorphic within the general population. Using the same techniques, the allelic frequencies within the general population were confirmed, as well as the heterogeneity within the expanded region. As part of this work the Southern blotting method used to estimate expansion size was optimised, to provide a more reliable, efficient, and standardised protocol to size these large repeat expansions.

1.6.2. Chapter 4 aims

The primary aim of this chapter was to evaluate the frequency of the repeat expansion in cohorts of pathologically confirmed, and clinically diagnosed multiple systems atrophy (MSA) patients. MSA is a movement disorder that shares phenotypic features with CANVAS, including autonomic dysfunction and cerebellar ataxia. It is also a challenging disease to diagnose and is in fact a common differential diagnosis for CANVAS. It was hypothesised that a small proportion of clinically diagnosed MSA patients may harbour the recessive pathogenic repeat expansion in *RFC1*, and present with an atypical MSA mimic. To investigate this, a cohort of clinically diagnosed MSA patients were screened for the expansion using flanking PCR, RPPCR and Southern blotting. It was also hypothesised that the *RFC1* repeat expansion would not be involved in MSA pathology, and this was evaluated by screening a cohort of pathologically confirmed MSA patients.

1.6.3. Chapter 5 aims

The primary aim of chapter 5 was to aid in the development of a bioinformatic tool used to detect the *RFC1* repeat expansion using output data from ExpansionHunter. The project was a collaboration with Illumina, who developed the ExpansionHunter software package. Using the 100,000 genomes project (100k GP) from Genomics England (GEL) research environment, a large cohort of 62, 225 rare disease probands and their relatives were screened using the RFC1 caller developed by Dr Michael Eberle and Dr Sai Chen. It was hypothesised that the allelic frequency would be similar to that found in the general population. It was also hypothesised that some recessive cases of the pathogenic expansion would be identified in diseases not associated with the classical CANVAS phenotype, based on its prevalence in

other diseases such as parkinsonism-related disorders and MSA mimics. By looking at segregation within trios, it was hypothesised that a Mendelian recessive pattern of inheritance would be identified in those harbouring the biallelic pathogenic repeat expansion. It was hypothesised that this tool could be successfully used diagnostically to identify patients carrying the recessive expansion if it presented with a high enough specificity and sensitivity.

1.6.4. Chapter 6 aims

The principal aim of chapter 6 was to functionally characterise a mutation found in Acyl-CoA Binding Domain Containing 5 (*ACBD5*) found in two brothers with ARCA. It was hypothesised that this mutation would cause loss-of-function of the peroxisomal membrane protein, with the implication of nonsense-mediated decay (NMD). This was investigated using western blotting, immunofluorescence, and complimentary DNA (cDNA) sanger-sequencing. Due to peroxisomal function involving metabolic processes such as very-long-chain fatty acid β-oxidation, it was hypothesised that this loss of ACBD5 would impair VLC β-oxidation, and this was tested using metabolic screening carried out by a collaborator, Dr Sacha Ferdinandusse. Peroxisomes interact with mitochondria, and so it was hypothesised that the loss of ACBD5 could impact mitochondrial function, which would contribute to the pathophysiology seen, either as a primary or secondary cause. To investigate this, mitochondrial assays including tetramethylrhodamine, methyl ester (TMRM) assay, NAD^+/NADH autofluorescence, ROS production, lipid peroxidation and a glutathione concentration assay, were carried out under the guidance of the Abramov research group at UCL, to evaluate mitochondrial function. During the course of this work, further ACBD5 cases came to light through Genematcher and collaborations between a research group led by Professor Nancy Braverman from the McGill University in Quebec, Canada, and with Dr Luisa Averdunk from Heinrich-Heine-University, Düsseldorf, Germany were formed. A further four patients were recruited to the functional work, revealing a phenotypic split between patients, with three presenting with retinal dystrophy and three without. It was hypothesised that there could be mechanistic differences contributing to these phenotypic differences, potentially by differences in compensatory pathway regulation. To investigate this further, those patients without retinal dystrophy were sent to a collaborator in Exeter, Professor Michael Schrader, to undergo peroxisomal motility testing, and all samples were sent for RNA sequencing to reveal any differentially expressed genes involved in ACBD5-deficiency and phenotypic group.

1.7. Introduction associated publications

Spinocerebellar ataxia: an update.
Sullivan R, Yau WY, O'Connor E, Houlden H.J Neurol. 2019 Feb;266(2):533-544. doi: 10.1007/s00415-018-9076-4. Epub 2018 Oct 3.

Cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS): Genetic and Clinical Insights.
Sullivan R, Kaiyrzhanov R and Houlden H. Current Opinion in Neurology August 2021 34(4):556-564

DNA repair in trinucleotide repeat ataxias.
Yau WY, O'Connor E, **Sullivan R**, Akijian L, Wood NW.FEBS J. 2018 Oct;285(19):3669-3682. doi: 10.1111/febs.14644. Epub 2018 Sep 10.

2. Chapter 2: Methods

2.1. Ethics

The work presented within this thesis was investigated under approval of the joint ethics committee of the UCL Institute of Neurology and the National Hospital for Neurology and Neurosurgery (NHNN), London, UK (UCLH: 04/N034).

2.2. Patient and sample cohorts

2.2.1. Biobank DNA samples

DNA samples were procured from the NHNN biobank under NHNN ethics and written permission from Prof. Henry Houlden and Prof. Michael Hanna. Samples were obtained from -80 ºC stores by Mr Mark Gaskin, and samples were aliquoted under his supervision. The Biobank was used for RFC1 sample investigations using gDNA from patients with CANVAS, or generalised late-onset ataxia, clinically diagnosed MSA cases and all control DNA.

2.2.2. MSA pathologically confirmed cases

gDNA from cerebellar brain tissue from 336 pathologically confirmed MSA cases, provided by Queen Square Brain Bank, Cambridge Brain Bank, Newcastle Brain Tissue Resource, Oxford Brain Bank, Edinburgh Brain Bank, Southwest Dementia Brain Bank, London Degenerative Disease Brain Bank, Sheffield Brain Tissue Bank, Manchester Brain Bank, Maryland Brain and Tissue Bank, and the Harvard Brain Tissue Resource Center.

2.2.3. RNA sequencing control cohort

Samples used as controls for RNA sequencing were recruited as cluster headache patients under the NHNN (n=33), under ethics code 04/N034. Informed consent was given, and skin biopsies were collected for all patients.

2.2.4. Fibroblast samples

For fibroblast work pertaining to the research detailed in chapter 6 fibroblast samples from patients 1 and 2 were obtained from NHNN patients under the care of Prof. Henry Houlden and Dr Emer O'Connor. Three family members from family 1 (mother, father, and unaffected sister) also gave consent for fibroblast samples to be taken. The fibroblast sample from patient 3 was sent by Dr Sacha Ferdinandusse from Amsterdam UMC, by permission of Dr Adeline Vanderver, from the Children's Hospital of Philadelphia (145) . Fibroblast samples from patient 4 were approved for use by Dr Forghani from the John T. Macdonald Foundation Department of Human Genetics at the University of Miami Miller School of Medicine, Miami, Florida, USA (146). Fibroblast samples from patient 5 were sent by

Prof. Nancy Braverman from the McGill University, Quebec, Canada. Patient 6 fibroblast sample was sent by Dr Luisa Averdunk from the Department of General Pediatrics, Heinrich-Heine-University, Düsseldorf, Germany. All patients (or parents, on behalf of patient) gave informed consent.

Control fibroblasts lines for research detailed in chapter 6 include an adrenoleukodystrophy (ALD) patient obtained and kindly given by Dr Eavan McGovern from the Salpetriere Hospital, Paris, and a fibroblast sample from a Zellweger Spectrum Disorder (ZSD) patient kindly sent by Prof. Nancy Braverman. Informed consent was obtained from all patients.

Healthy control fibroblast lines, and control fibroblast lines of cluster headache patients for RNA sequencing detailed in chapter 6, were obtained from the NHNN fibroblast biobank at the UCL, Queen Square Institute of Neurology.

2.2.5. Genematcher

Two patients were recruited to the project detailed in chapter 6 via the online portal Genematcher, a freely accessible web site that enables connections between researchers with interests in the same genes. Users are matched based on their submitted gene of interest (147).

2.2.6. Patient clinical examination

Patients involved in the research project documented in chapter 4 were clinically examined by Dr Wai Yan Yau, a neurologist from the NHNN. The patients gave consent for clinical examination videos to be recorded and for these to be published for research purposes.

Patients 1 and 2 involved in the research project discussed in chapter 6 were clinically examined by Dr Emer O'Connor, a neurologist from the NHNN and Prof. Henry Houlden, consultant neurologist at the NHNN. These patients gave informed consent for clinical examination videos to be recorded and published for research purposes.

2.2.7. Genomics England samples

Whole genome data for patients recruited into the 100,000 Genomes Project was analysed under research registry ID RR31. The 100,000 Genomes Project is a programme set up to assess the genomes of patients with molecularly unsolved rare diseases and cancer. These patients were recruited under health-care professionals and researchers from 13 centres across the UK. Full informed and written consent for data to be used for research purposes was given, and family members were also recruited where possible. WGS was carried out by Genomics England (Hinxton, UK), and genomic data analysed on the GEL research platform. Genomes were aligned to the hg38 reference genome (GRCh38 Decoy assembly), using the illumina Isaac pipeline v03.16.02.1964.

2.3. Genomic techniques

2.3.1. Primer sequences

Primer	Sequence
Flanking PCR	
Pathogenic AAGGG	Fw: TCAAGTGATACTCCAGCTACACCGTTGC Rv: GTGGGAGACAGGCCAATCACTTCAG
RPPCR	
Fam-labelled Forward	FAM-TCAAGTGATACTCCAGCTACACCGT
Anchor	CAGGAAACAGCTATGACC
Pathogenic AAGGG	*Rv1:* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAAAGGGAA GG GAAGGGAAGGGAA *Rv2* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAAGGGAAG GG AAGGGAAGGGAA *Rv3* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAGGGA GG GAAGGGAAGGGAA
Expanded AAAGG	*Rv1:* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAAAGGA AAGGAAAGGAAAGGAAA *Rv2* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAGGAAAG GAAAGGAAAGGAAA *Rv3:* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAGGAAAG GAAAGGAAAGGAAA
Expanded AAAAG	*Rv1:* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAAAGAAA AGAAAAGAAAAGAAAA *Rv2:* CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGTTTCAAAAAGAAAA GAAAAGAAAAGAAAA *Rv3:* CAGGAAACAGCTATGACCAACAGAGCAAGACTC TGTTTCAAAAGAAAAG AAAAGAAAAGAAAA
Southern Blotting	
DIG Probe	*Fw:* ATTAGGTGTCTGGTGAGGGC *Rv:* GAAGAATGGCCCCAAAAGCA
Sanger sequencing	
ACBD5 DNA primers	*Fw:* GCACGCCTTCTCTTTGATCC *Rvs:* TAGGAAAGAGGCTGGACACA
ACBD5 cDNA primers	*Fw:* TACTCCAAACGCCAAAACCG *Rvs:* ATGTTGCCATTTCCAGGAGG
DIG probe synthesis	
Forward	*Fw*: ATTAGGTGTCTGGTGAGGGC
Reverse	*Rv*: GAAGAATGGCCCCAAAAGCA

Table 2.1: Primer sequences

2.3.2. Flanking Polymerase Chain-Reaction (PCR)

Flanking PCR was used to differentiate samples with normal sized DNA fragments, from those with expansions. 1µl DNA added to 7.5 µl FastStart™ PCR Master, 5 µl H_2O and 0.75 µl of forward and reverse primer (table 2.1). Reaction run on following thermocycler programme: 95 ºC for 4 minutes, [95 ºC for 30 seconds, 63 ºC for 30 seconds, 72 ºC for 1 minute] for 35 cycles, 72 ºC for 5 minutes and 4 ºC hold. 5 µl of PCR reaction volume mixed with 5 µl DNA loading dye and run on 1% agarose gel (100 ml of 1x Tris-borate-EDTA (TBE), 1 g of agarose and 3 µl of ethidium bromide (EtBr), alongside 3 µl of GelPilot® wide range ladder. Gel run at 90 V for 30 minutes and then visualised in Ultraviolet (UV) transilluminator.

2.3.3. Repeat-primed PCR

Repeat-primed PCR (RPPCR) was used to identify the presence of an expanded AAGGG repeat, expanded reference AAAAG allele or the expanded variant AAAGG allele, using primers targeted to each allelic conformation (table 2.1). The design of the reverse primers by Dr Andrea Cortese produced multiple incrementally sized amplicons by binding at different points within the expansion repeat. To give preferential amplification of the larger PCR product 25 – 27 nucleotides flanked the repeat which also increased binding affinity to the polymorphic (A/AA/-)3' end of the microsatellite and flanking region. Reactions consisted of 1 µl of DNA, 7.5 µl Phusion® High-Fidelity PCR Master mix (ThermoScientific®), 5 µl H_2O, 0.75 µl fluorescently labelled forward, 0.75 µl anchor primer and 0.075 µl reverse primer which was exhausted early on during the reaction allowing the anchor primer to take over amplification. Each specific primer set was run on a specific thermocycler programme (table 2.2). 1µl of each reaction was mixed with 9.2 µl HiDi Formamide (Applied Biosystems®) and 0.3 µl GeneScan™ 500 LIZ® Size Standard (Applied Biosystems®). Fragment analysis was performed on AB1 3730xl genetic analyser (Applied Biosystems®, Foster City, CA, USA). Fragment analysis via capillary electrophoresis was carried out to identify amplified fluorescent DNA fragments (a product of the FAM-labelled forward primer). The GeneScan™ 500 LIZ® Size Standard allows for a spectral calibration, to allow for accurate detection of the dye-labelled primer. During capillary electrophoresis, the PCR products are injected electrokinetically into polymer-filled capillaries and a high voltage is applied. The fluorescent DNA fragments are separated by size, which are then detected by a laser system. Data was analysed using GeneMapper V5 software, which precisely measures the size of fragments. Samples with a 'saw-toothed' pattern associated with the presence of a repeat expansion were put forward for sizing on a Southern Blot if enough DNA was available.

Primer set	Thermocycler programme
Pathogenic AAGGG	98 ºC – 3 minutes **35 CYCLES:** 98 ºC – 10 seconds 70 ºC – 15 seconds 72 ºC – 2 minutes 72 ºC – 1 minute 4 ºC – Hold
Expanded AAAAG	98 ºC – 3 minutes **35 CYCLES:** 98 ºC – 10 seconds 55 ºC – 15 seconds 72 ºC – 2 minutes 72 ºc – 1 minute 4 ºC – Hold
Expanded AAAGG	98 ºC – 3 minutes **35 CYCLES:** 98 ºC – 10 seconds 65 ºC – 15 seconds 72 ºC – 2 minutes 72 ºc – 1 minute 4 ºC – Hold

Table 2.2: Thermocycler conditions

2.3.4. Southern Blot

2.3.4.1. DNA Digestion and Agarose Gel Loading

To confirm the presence of expanded alleles, Southern blotting was performed. 5 µg of sample gDNA was diluted using autoclaved H_2O into a total volume of 33 µl. If the DNA concentration was below 151.5 ng/µl, 33 µl of neat gDNA was used. This was digested using 4 µl of 10X CutSmart Buffer, 2 µl of Spermidine and 1 µl of EcoR1 and incubated at 37 ºC for 1 hour, after which 1 µl of EcoR1 was added and then a further incubation at 37 ºC for 2 hours. The digested sample was mixed with 8.6 µl of Promega Blue/Orange 6x loading dye and 5 µl was transferred to a separate PCR plate to perform a checker gel. The mixed samples were loaded onto an agarose gel prepared using 480 ml 1X TBE and 3.84 g of agarose. Two Roche Dig-labelled size markers were used alongside the samples, Molecular Weight Marker II and III. Both markers prepared using 4 µl of autoclaved H_2O, 1 µl of marker and 1 µl of Promega Orange/Blue 6x loading dye. 1x TBE was used as a buffer in the electrophoresis tank, which was run at 50 V for 15 hours.

2.3.4.2. DNA transfer onto Membrane

The gel briefly washed in H_2O and then washed in 500 ml of 0.25 M Hydrochloric acid (HCL), Depurination solution (475 ml H_2O and 25 ml of conc. HCL), then 500 ml of Denaturing solution (10g

sodium hydroxide (NaOH) and 29.2g sodium chloride (NaCl) dissolved in 500 ml of H_2O) and then 500 ml of Neutralising solution (Alfa Aesar) for 45 minutes each on the slow shaker. The blot was then assembled using 4 pieces of 3MM Whatmann (Roche) soaked in 10X Saline Sodium Citrate (SSC), a 20 cm x 20 cm piece of Nylon Membrane, Positively Charged (Roche) of which the edges of the gel were masked using saran wrap, a piece of 20 cm x 20 cm piece of 1 CHR Whatmann paper (Roche) soaked in 2X (SSC) buffer and then 4 additional 3MM Whatmann papers followed by 3 even stacks of paper towels, a glass plate and a 1 L bottle of water used as a weight. This was left for 4 hours, with periodic checks to ensure the glass plate was level. Once finished, the membraned was placed DNA side down onto a 3MM sheet of Whatmann paper was soaked in 2x SSC and submitted to UV light for 3 minutes then left to dry overnight.

2.3.4.3. Membrane Hybridisation

The membrane was hybridised for 3 hours with 40 ml Roche DIG Easy-Hyb Solution (Roche) in a pre-warmed 49 ºC shaking incubator at 70 rpm with a saran wrap covering to prevent evaporation. The membrane was drained and probe mixture, consisting of denatured 60 µl of 1kb probe added to 30 ml of DIG Easy Hyb solution (Roche), was added to the membrane, again covered in Saran wrap. This was incubated at 49 ºC overnight at 70 rpm.

2.3.4.4. Washing and detection

The membrane was then drained and washed 2X with 500 ml 0.5 X SSC /0.1% Sodium dodecyl Sulphate (SDS) solution for 15 minutes, at 65 ºC at 40 rpm, followed by a 1X wash with 200 ml Washing Buffer (Roche DIG Wash and Block Buffer Set) and a 1-hour incubation with Blocking Solution (Roche DIG Wash and Block Buffer Set). The membrane was drained and placed in 40 ml of Block Solution containing 4 µl of freshly centrifuged 'Anti-DIG AP' antibody (Roche) for 30 minutes at room temperature on the slow shaker. The membrane was washed twice with 200 ml of Washing Buffer for 15 mins (Roche DIG Wash and Block Buffer Set) and then 40 ml of Detection Buffer for 5 mins (Roche DIG Wash and Block Buffer Set). The membrane was then transferred into a clean box and 6 ml of CDP-Star® ready-to-use (Sigma-Aldrich) was applied and incubated for 1 minute. The membrane was drained as much as possible, wrapped in saran wrap and autoclave taped into a film cassette, DNA side up. Films were developed for 5 minutes before visualisation using Super RX-N Fuji Medical X-Ray Film (FujiFilm).

2.3.4.5. Membrane stripping and reprobing

The membrane was washed with 500 ml of 0.5x SSC and 1% SDS solution for 1 hour at 65 ºC at 40 rpm, followed by 3x 5 min washes of 2x SSC were performed, followed by membrane hybridisation, and washing and detection.

2.3.5. DNA Extraction

gDNA extracted from fibroblast cells using the Wizard Genomic DNA Purification Kit (Promega), following the manufacturers protocol.

2.3.6. RNA Extraction

RNA was extracted using the Quick-RNA Miniprep Kit (Zymo) following the manufacturers protocol and including the on-column DNAse treatment for removing contaminated gDNA. RNA concentration was quantified on a Nanodrop spectrophotometer, measuring the 260 nm absorbance.

2.3.7. Reverse Transcription

A reverse transcription reaction using 500 ng of RNA was used to create a complementary DNA (cDNA) library using the SuperScript II (SSII) Reverse Transcriptase. RNA was denatured in 10 µl $_{dd}$H2O using random primers (Promega) at 65 ºC for 5 minutes, and then snap cooled on ice. This mix was added to 1 x first synthesis buffer, 10 nM, , 0.5 mM dNTPs (C, G, A and T), Dithiothreitol (DTT), 1 U/µl RNAseOUT and 1.25 U/µl SSII reverse transcriptase (Thermo Scientific), made up to 20 µl. The reverse transcription reaction was catalysed by a 1-hour incubation at 42 ºC, followed by a 15-minute denaturation at 70 ºC. A RT^- control reaction lacking SSII reverse transcription enzyme was performed for each RNA sample, to control for gDNA contamination.

2.3.8. Sanger Sequencing

gDNA was extracted from blood and polymerase-chain reaction (PCR) performed using primer sequences (table 2.1) and FastStart™ master mix (Sigma). The PCR product was cleaned up using Exo-Fast containing 50 µl Exo I (Thermo Scientific), 200 µl Fast-Ap (Thermo Scientific) and 750 µl H_2O, to remove single strand DNA including unused deoxynucleotide triphosphates (dNTPs) and unused primers. Excess fluorescently labelled di-dNTPs (ddNTPs) were expelled via filtration through G50 Sephadex columns (Sigma) using 0.66 mm glass fibre FiltrEX™ plates (Corning). The product was loaded onto AB1 3730xl genetic analyser (Applied Biosystems®, Foster City, CA, USA). Data analysed using Sequencher® software.

To identify nonsense-mediated decay (NMD), sanger sequencing was carried out on cDNA of patient and parents. Sanger sequencing was performed on cDNA using primers (table 2.1). Data analysed using Sequencher® software.

Additional Sanger sequencing obtained by sending clean PCR product of samples to Source Bioscience, along with necessary primers.

2.3.9. SNP Haplotype Screening

Samples were genotyped using primers targeting two disease-associated haplotype markers rs11096992 (AA) and rs2066790 (AA) on all healthy control screen samples and late-onset ataxia samples.

2.3.10. RNA-sequencing

RNA sequencing performed on six ACBD5-deficient patients and 33 controls (methods section 3.1.2). RNA was extracted as previously described on fibroblast cell lines maintained as described in Methods section 3.3.1. Each sample was sequenced with a read depth of 40million. RNA sequencing pipeline performed by UCL Genomics facility; in brief samples (100ng RNA per sample) underwent library prep using Kapa mRNA Hyper Prep (Roche), quality control using a Bioanalyzer and sequencing via NovaSeq S1 v1.5 (100 cycles) (Illumina). Please see Methods section 3.4.3 for analysis methods.

2.3.11. Whole-exome-sequencing

Patients 1 and 2 detailed in chapter 3.4 underwent whole-exome sequencing at GeneDX, where gDNA was available. Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA) was used to capture the exonic regions and flanking splice junctions. Massively parallel (NextGen) sequencing was done on an Illumina system with 100 bp or greater paired end reads. Reads were aligned to human genome build GRCh37/UCSC hg19 and analysed for sequence variants using a custom-developed analysis tool by Dr David Lynch.

2.4. Molecular biology techniques

2.4.1. Cell culturing

Primary skin fibroblasts were cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% fetal bovine serum (Sigma). Cells were passaged at 70-90% confluency using 0.25% w/v Trypsin-EDTA solution (Gibco) and Dulbeco's Phosphate Buffered Saline (Sigma), in a 1:2 split. Cells were kept in a humidified incubator, maintained at 37 ºC, 95% air/5% CO2 gas mixture. Cells underwent a medium change every 4-7 days.

2.4.2. Mycoplasma testing

Mycoplasma contamination was routinely tested in actively cultured cell lines, every 2 months, using the biolyminescence based MycoAlert™ Mycoplasma Detection Kit (Lonza), following the manufacturers protocol. 500 µl of cell medium was transferred to a sterile Eppendorf and centrifuged for 10 minutes at 500g to pellet cells and cell debris. 50 µl of supernatant was transferred to a 96-well plate, that also contained 50 µl of MycoAlert™ Buffer (used as negative control) or 50 µl of MycoAlert™ positive control. 50 µl of MycoAlert™ reagent was added to each well and the luminescence of each well was read by the Fluostar Omega Microplate reader (BMG Labtech). 50ul of

MycoAlert™ substrate was then added to each well, and luminescence readings repetitively performed across a 1-hour period. Cell lines without mycoplasma contamination showed no luminescence, compared to increasing luminescence in the positive control. No mycoplasma contamination was found in any cell lines.

2.4.3. Metabolic Screening

Metabolic screening was carried out by Dr. Sacha Ferdinandusse at the Laboratory of Genetic and Metabolic Disease, Academic Medical Centre, Amsterdam. Frozen cryovials of primary cell fibroblasts for patients 1, 2, 4 and 6 were sent to Amsterdam on dry ice. Patients 3 and 5 had previously both undergone the same metabolic screening at the Laboratory of Genetic and Metabolic Disease. C26:0 and C22:0 were measured in cell pellets. For the D3-C22:0 loading test, D3_C22:0 was added to the medium and fatty acid analysis was performed 3 days later.

2.4.4. Whole fibroblast cell protein lysate generation

Centrifuged fibroblast cells were washed with ice-cold Dulbecco's Phosphate Buffered Saline (D-PBS, Sigma) and placed at -80 ºC. Alternatively, washed cells were physically detached from the cell culture dish using a cell scraper (VWR) whilst on ice and pelleted in 1 ml DPBS, these were then placed at -80 ºC before cell lysing step. Cell pellet was resuspended in ice-cold radioiummoprecipitation assay (RIPA) lysis buffer (R0278, Sigma), containing protease inhibitor cocktail (11836170001, Roche). Cells were lysed on ice for 1 hour, with intermittent vortexing every 10 minutes. The cell lysate was then centrifuged at 20,000g for 10 minutes at 4 ºC, to pellet cell debris and the supernatant containing the protein lysate was transferred to a fresh, pre-chilled Eppendorf tube.

2.4.5. Bradford Assay

Cell lysate diluted 1:10 in autoclaved H_2O. 5µg/ml of Bovine Serum Albumin (BSA, Thermofisher) stock diluted as follows into separate aliquots (table 2.3):

Vol. of BSA for 4 replicates (µl)	Vol. of RIPA buffer (µl)	Vol. of H_2O (µl)
0	4	36
1.2	4	34.8
1.6	4	34.4
2.0	4	34
2.4	4	33.6
2.8	4	33.2
3.2	4	32.8
3.6	4	32.4

Table 2.3: BSA stock dilutions

BSA dilutions split into three 10µl replicates per BSA dilution and one triplicate of 10µl diluted cell lysate prepared, into a 96-well plate. 200 µl of Bradford Reagent added into every well. Absorbance values measured at 590 nm read on Fluorspar Omega software and Bradford Assay converted into BSA

Standard Curve. Protein concentration for samples calculated with reference to the standard curve. Protein concentration was calculated by comparison of absorbance values of protein samples to known BSA standard curve.

2.4.6. BCA Assay protein quantification

Protein concentration also quantified using the Pierce bicinchoninic acid (BCA) assay (Thermofisher), using the manufacturers protocol. Duplicates of known BSA (Thermofisher) concentration were added into a 96 well plate to create standard curve. Duplicate wells of 1 µl of lysis buffer used as blanks, and 200 µl of Pierce BCA assay reagent added to each sample. Absorbance was measured at 562nm using Fluorspar Omega software. Protein concentration was calculated by comparison of absorbance values of protein samples to known BSA standard curve.

2.4.7. Polyacrylamide Gel Electrophoresis and Western Blotting

1x Lithium Dodecyl Sulphate buffer (LDS) (Thermofisher) and 10 mM Dithiothreitol (DTT) added to equal amounts of proteins (15-30 µg) and samples were denatured at 70 ºC for 10 minutes and then cooled at 4 ºC. Samples were separated by polyacrylamide gel electrophoresis (PAGE). 4-12% Bis-Tris gel (Invitrogen) used in Bolt Mini gel tank (Thermofisher) containing 1x MES running buffer. 10 µl SeeBlue ladder (Life Technologies) loaded into first well, followed by sample into remaining wells. Gel ran at 120 V for 10 minutes and then 180 V for an hour. Separated proteins were electrophoretically wet transferred to 0.22 µm nitrocellulose membrane (BioRad) at 100 V for 60 minutes at 4 ºC, using 1x tris/glycine buffer with 20% v/v methanol (Sigma). Following transfer, membrane was washed in phosphate-buffered saline (PBS, GIBCO) with 0.1% Tween-20 (PBST, Sigma) and membrane stained with 50% v/v Ponceau S Solution (Sigma), diluted in PBST, in order to gauge transfer efficiency and as a guide for cutting the membrane for simultaneous staining of antibodies for proteins with different molecular weights. Membrane was washed a further three times with PBST for 10 minutes at room temperature. Membrane was blocked with 10 ml of 1 g of Blocking Agent (GE Healthcare) in 20 ml of PBST for 1 hour at room temperature, shaking Membrane was incubated with primary antibody Rabbit Anti-ACBD5 (HPA012145, Sigma) (table 2.4) added in a 1:500 dilution with 1% Blocking Agent, and Anti-βactin (A5441, Sigma) in 1:1000 dilution, and left overnight shaking at room temperature. Three washes in PBST were carried out, followed by application of horseradish peroxidase conjugated secondary antibodies, goat anti-rabbit (31466, Invitrogen) and goat anti-mouse (31431, Invitrogen) in a 1:2000 dilution, which were incubated for 1 hour at room temperature, shaking. Membranes were visualised using the Amersham ECL Prime Western Blotting Detection Reagent (GE Life Sciences). They were imaged using a ChemiDoc Touch imaging system (BioRad).

2.4.8. Antibody Immunogen sequences

Western Blotting Antibodies	
Anti-ACBD5	AQEEVKGAEQSDNDKKMMKKSADHKNNGYDKDGFVQDI QNDIHASSSLNGRSTEEKPIDENLGQTGKSAVCIHQDINDDH VEDVTGEVYCDSMEQFGQEESLDSFTSNNGPFQYYLGGHSSQPM
Anti- βactin	Slightly modified β-cytoplasmic actin N-terminal peptide, Ac-Asp-Asp-Asp-Ile-Ala-Ala-Leu-Val-Ile-Asp-Asn-Gly-Ser-Gly-Lys, conjugated to KLH.

Table 2.4: Immunogen sequences for primary antibodies

2.4.9. Immunofluorescence microscopy

Primary skin fibroblasts cultured on cover slips of approximately 70% confluency. Cells were fixed and permeabilised in D-PBS (Sigma) containing 4% paraformaldehyde (PFA, Sigma). Cells were blocked in a blocking and permeabilization solution containing D-PBS, 10% donkey serum and 0.1% w/v Triton X-100 for 1hr, covered in foil. The rabbit polyclonal primary antibody (HPA012145, Sigma) was diluted 1:200 in Dako antibody diluent (Agilent) and added to cells for an overnight incubation at 4 ºC. The cells were allowed to come to room temperature and washed three times in DPBS. A secondary antibody (donkey anti-rabbit, Alexa Fluoro 488 (ThermoFisher), was diluted in 1:1000 and added to cells and incubated for 1 hr at room temperature. Cells were mounted to glass slides using Dako fluorescence mounting medium (Agilent).

Images were acquired using a Zeiss 710 VIS CLMS confocal microscopy and an x40 oil immersion objective (Zeiss, Oberkochen, Germany). Images were processed and analysed using ImageJ FIJI software.

2.4.10. Measurement of mitochondrial membrane potential by TMRM

Primary skin fibroblasts grown on cover slips were incubated for 40 minutes with 25 nM 25 nM tetramethylrhodamine methyl ester (TMRM, ThermoFisher). TMRM is positively charged and accumulates in the negatively charged mitochondria matrix. TMRM fluorescence is proportional to the mitochondrial membrane potential (ψ_m), therefore a reduction in TMRM intensity corresponds to depolarisation of the ψ_m.

Images were acquired using a Zeiss 710 VIS CLMS confocal microscope with a META detection system and an x40 oil immersion objective (Zeiss, Oberkochen, Germany). The 560 nm laser line was used to excite TMRM, and emitted fluorescence was measured above 580 nm for TMRM, keeping the laser at a minimum (<0.3%) to avoid phototoxicity. Z stacks were acquired for the basal ψ_m, at maximum intensity projection. 5-10 Z stacks were taken per sample. Images were analysed using ImageJ, with a minimum pixel intensity threshold set to exclude non-mitochondrial fluorescence from analysis.

To assess the function of the OXPHOS complexes and their contribution to Ψ_m maintenance, cells were incubated with 25 nM TMRM and TMRM intensity was measured over time, with toxins added sequentially. A single 2 µm focal plane was used to focus on mitochondria and images were taken every 5-10 seconds. Baseline was established and measured for >4 timepoints with stable TMRM intensity, before 10 µg/ml of Oligomycin (Sigma) was added, in order to inhibit complex V. A slight increase in fluorescence is expected in healthy cells following oligomycin addition, as CV utilises the Ψ_m to synthesise ATP. Once fluorescence had stabilised, 5 µM rotenone (Sigma) was added to inhibit complex I. Once fluorescence had stabilised, 1 µM Carbonyl cynanide-p- trifluoromethoxyphenylhydrazone (FCCP) protonphore was added to induce complete Ψ_m depolarisation, by dissipating the mitochondrial H^+ gradient. ImageJ FIJI software was used to analyse TMRM intensity over time, with a minimum pixel intensity threshold set to exclude non-mitochondrial fluorescence from analysis. TMRM intensity values were normalised to basal levels of TMRM which were taken as 100% and TMRM intensity in the mitochondria after complete depolarization from the addition of FCCP was taken as 0%.

2.4.11. Measurement of NADH redox state

NADH autofluorescence was measured using an epifluorescence-inverted microscope with a x40 oil objective. A Xenon arc lamp passed through a monochromator provided excitation light at 360 nm. A cooled CCD camera picked up the emitted light reflected through a 455 nm long-pass filter. Images were acquired using Andor Software and analysed using ImageJ. Basal autofluorescence was read and then 1 µM FCCP was added to the cells, causing complete mitochondrial membrane depolarization and oxidation of the mitochondrial pool of NADH to NAD^+ which lacks fluorescence. This was taken as 0%. 1 mM of nitrogen cyanide (NaCN) was added, which inhibited respiration and allowed the regeneration of the NADH mitochondrial pool, which was taken as 100%. The NADH pool was calculated from the difference between the minimum and maximum values of autofluorescence. The NADH redox index was calculated as the percentage represented by basal levels when extrapolating its value in the 0-100% range produced by FCCP and NaCN.

2.4.12. Measurement of ROS production via Glutathione level measurement

To analyse glutathione (GSH) levels, fibroblast cells were incubated with 50 µM monochlorobimane (MCB) in HBSS at room temperature for 30 minutes. Monochlorobimane (MCB) is a non-fluorescent bimane which is freely permeable across the cell membrane and forms a fluorescent adduct when combined with reduced glutathione (GSH) in a reaction catalyzed by glutathione s-transferase. Cells were then washed with HBSS. Live cell images were acquired using a Zeiss 710 LSM confocal microscope with a META detection system, using a x40 oil immersion objective. MCB was excited at 405 nm, with emission light detected between 430 and 525 nm. Z-stacks were acquired, and fluorescence intensity was analysis using ImageJ FIJI software. Mean pixel intensity was measured,

using a minimum threshold. Experiments were repeated three times using independent passages of cell culture, with three technical replicates per cell line.

2.4.13. Measurement of rate of lipid peroxidation

The rate of lipid peroxidation was measured using Zeiss 710 LSM confocal microscope. BODIPY (655/675, 2 µM, Molecular Probes) was excited by 563 and 630 nm lasers and measured from 580 to 610 and above 650 nm, using an oil-immersion 40x objection. Illumination intensity was kept to a minimum (at 0.1 - 0.2% of laser output) to avoid phototoxicity and the pinhole set giving an optical slice of 2 µm. Data were measured using ZEN2009 software, with mean pixel intensity measured over time.

2.4.14. Peroxisomal number and biogenesis analysis

The Schrader research group at Exeter University carried out investigations into peroxisomal number, with the protocol being previously described (148).The work was carried out by Mrs Tina A. Schrader and Miss Rechal Kumar, and data was kindly provided by Prof. Michael Schrader. Peroxisomal biogenesis and organelle number were analysed in fibroblast cells from patient 1, patient 6, and two age-matched controls. Fibroblast cells were processed or immunofluorescence using antibodies against the peroxisomal membrane marker PEX14. Peroxisomal number in fibroblasts were calculated using an in-house ImageJ macro, utilising the Analyze Particles Function, by the Schrader research group from Exeter University (149).

2.4.15. Transmission electron microscopy and quantitation of organelle-organelle contacts

Peroxisomal-endoplasmic reticulum contacts were assessed in patients 1, patient 6 and two aged-matched healthy controls by Professor Michael Schrader at Exeter University, UK, using a published method (150). Fibroblast cell cultures were maintained by Mrs Tina A Schrader from Exeter University and data was generated by Dr Christian Hacker from the Exeter Bioimaging Centre and was kindly sent to me by Professor Michael Schrader.

In brief, for ultrastructural analysis, monolayers of ACBD5 patient cell lines as well as age-matched control cell lines, were fixed and processed for transmission electron microscopy as previously described. Cells were fixed in 0.5% glutaraldehyde in 0.2M PIPES, pH 7.2, post-fixed in 1% osmium tetroxide (reduced with 1.5% w/v potassium ferrocyanide) in 0.1M sodium cacodylate, pH 7.2. Cells were then dehydrated through a graded ethanol series (30%-100% ethanol) and embedded in Durcupan resin (Sigma Aldrich, Merck, Gillingham, UK). 60nm ultrathin sections, contrasted with lead citrate, were imaged with a JEOL JEM 1400 transmission electron microscope operated at 120 kV and a digital camera (Gatan, EW1000S, Abingdon, UK). For quantitation of organelle contacts between PO's and ER, cells were sampled and assessed as previously described in Costello et al., 2018. Briefly, cell profiles were scanned applying a systematic uniform random (SUR) approach to ensure unbiasedness

of the organelle selection. Peroxisomes were identified as single membrane-bound organelles of round to oval shape, displaying a fine granular homogenous matrix and matching the size range previously published (Costello et al., 2018). Positively identified peroxisomes were imaged at a nominal magnification of 80k X to include 35-68 PO profiles per experiment (mean of 42). Images were then opened in Fiji (Schindelin et al., 2012) and overlaid with a random square lattice grid (grid spacing of 0.01 μ^2). Grid lines intersecting with PO membranes were categorized as being an organelle contact site (distance between PO and ER membrane ≤ 15 nm) or non-associated (yielding 232 ± 34 [STDEV] intersections in total per experiment). Data was statistically analysed in GraphPad Prism 5 by either a one-way analysis of variance followed by Tukey's multiple comparison test or unpaired t-test with Welch's correction.

2.5. Bioinformatic and statistical analysis

2.5.1. Predicting mutational effects

For predicting mutation effects, the online software tool Mutation Taster was used, which has been previously published (151). Mutations were also run through the online software tool Expasy Translate, to predict protein transcripts, which has also been previously published (152).

2.5.2. Pathway analysis using STRING

Gene pathways and protein-protein networks were analysed using the online STRING database, with results download from the STRING website (153).

2.5.3. Gene Set Enrichment and Over-Representation Analysis

Gene Set Enrichment and Over-Representation Analysis (ORA) were performed using WEB-based Gene SeT AnaLysis Toolkit (WebGestalt) (154). The top 200 genes ranked firstly by the adjusted P value (q-value), and then by P value. The gene symbols for these top 200 genes were uploaded to WebGestalt for ORA. For Gene Set Enrichment the -LOG(p value)*LOG(foldchange) was calculated as a gene ranking, and this ranking was uploaded, alongside the corresponding gene symbol. Tables of the top 200 ranked genes used for Gene Set Enrichment and ORA can be found in tables 8.3.1 and 8.3.2 in the Appendix. Full results of the Gene Set Enrichment and ORA results can be found in the appendix, tables 8.3.3-8.3.5.

2.5.4.ExpansionHunter and RFC1 Caller

All scripts used for initial ExpansionHunter run were downloaded from https://github.com/Illumina/ExpansionHunter, which is publicly available, and used within the GEL research environment. Initial ExpansionHunter processing run as previously documented (155).

RFC1 caller algorithms and script were developed by Dr Sai Chen and Dr Michael Eberle from Illumina. RFC1 caller script was privately provided by Dr Sai Chen. All subsequent scripts were written and run by Roisin Sullivan within the GEL research environment, using the RStudio statistical environment, with guidance from Dr Jana Vandrovcova.

ExpansionHunter was first run on all samples from 100kGP rare diseases' cohort samples (n = 66,250). ExpansionHunter repeat calls were then run on RFC1 caller using BAM and VCF output files. In brief the caller first identified whether an individual has one or two expansions using an aggressive definition of an expansion (125bp or 150bp). Based on this samples were classified as 'normal' (no expansions), 'potential carriers' (one expansion) or 'potentially pathogenic' (two expansions). For samples called as having at least one expansion, the RFC1 caller then identified the repeated motif within the expanded allele(s). Heuristic rules were applied to define the likely *RFC1* genotype. Samples were considered 'normal' unless any of the following were observed: (1) it has two expansions and >80% of the counted repeat motifs are one of the known pathogenic motifs (currently only AAGGG) and is therefore called as pathogenic, (2) it has two expansions and has more than 20% reads with >80% of their counted repeat motifs as AAGGG and is therefore called as a carrier, (3) the sample has two expansions and 30-80% of the counted repeat motifs are AAGGG, and is therefore a carrier, (4) the sample has one expansion and >80% of the counted repeat motifs are AAGGG, and is therefore called as a carrier. Caller outputs define the allelic distribution of the motifs, and combinations of motifs. Cases where pathogenic AAGGG has been identified on both alleles is called as 'Patient: two expanded pathogenic alleles', alleles that contain expanded but non-pathogenic motifs are called as 'expanded benign', and alleles that contain the reference motif AAAAG within the normal range are called as 'short reference allele'.

Analyses of caller results and allelic distribution were performed on the rare diseases cohort a as whole. Segregation analyses were conducted on available trios (n =9657). Motif distribution was calculated using a subset of alleles from the entire cohort (n=53,792)

2.5.5.RNA sequencing Analysis

RNA sequencing data first processed using Kallisto software by Dr Jana Vandrovcova to quantify the abundances of transcripts (156). Kallisto output transcript abundancies were further analysed by myself using Sleuth software within the RStudio Statistical Environment (157).

2.5.6. Statistical Analysis

All statistical analysis carried out in either RStudio statistical environment or GraphPad Version 6. All variables were first assessed for normal distribution by plotting histograms, Q-Q plots and tested using a Shapiro-Wilke test. When variables were normally distributed parametric statistical tests were used, including T-test for two group means comparison and One-Way Anova for multiple groups mean comparisons, followed by a Tukey HSD post-hoc test. In the case of non-normally distributed data, non-parametric statistical methods were used which included a Mann Whitney U test for two group means comparison and a One-Way Kruskal-Wallace test, followed by pairwise duns for post-hoc comparisons. All RNA-sequencing results were corrected for multiple testing using the Benjamini-Hochberg correction. For all comparisons an alpha value equal to or less than 0.05 was considered statistically significant.

3. Results

3.1. Repeat expansions in Replication Factor C Subunit 1 (*RFC1*)

3.1.1. Introduction

Late-onset ataxia (LOCA) is a common, neurological condition that can be classified into hereditary, non-hereditary and acquired degenerative ataxias, with around 60% of familial cases and 19% sporadic having a genetic basis (158-160). LOCA is caused by the failure of the systems controlling motor coordination, with 50% of cases remaining idiopathic, classified as idiopathic late onset ataxia (ILOCA). One geographical study of LOCA within South Wales identified 86% of LOCA patients as having ILOCA, giving an estimated prevalence of 8.4 per 100,000 within the South Wales population (161).

Pure idiopathic late-onset cerebellar ataxia (ILOCA) has a range of clinical signs that include gait and stance ataxia, commonly leading to falls, daily life limitations and neurological referral. Studies have shown that factors such as polyneuropathy (PNP) and/or peripheral vestibular deficits, as well as cerebellar disease can be combined as contributory causes, such as bilateral vestibulopathy and additional cerebellar ataxia (162). The various combination of symptoms coined the acronym CANVAS which is defined as a progressive ataxic disorder with, as described, a combination of cerebellar and somatosensory and vestibular impairment (64).

Clinically the onset of CANVAS symptoms varies in timing, however the typical bilateral vestibulopathy, somatosensory deficit and cerebellar impairment triad of symptoms allows for a clear diagnosis once other causes such as FRDA or SCA3 are ruled out (table 3.3.1) (52). However, CANVAS can be difficult to diagnose when there is a delay in symptoms as there are a spectrum of clinical signs including reports of autonomic dysfunction and chronic cough also associated with the syndrome (62). Patients with CANVAS present with an abnormal visually enhanced vestibulo-ocular reflex (VVOR), which is a result of a composite deficit of the three oculomotor reflexes: the optokinetic reflex, vestibulo-ocular reflex and smooth pursuit (60). This results in saccadic compensatory eye movements, instead of smooth (52).

CANVAS triad symptom	Impairment
Bilateral vestibulopathy	Decreased VOR Irregular VVOR
Cerebellar Impairment	Cellular Atrophy Cerebellar dysarthria Dysphagia of cerebellar basis Cerebellar oculomotor abnormalities Appendicular ataxia
Somatosensory impairment	Evidence of neuronopathy
Additional genetic testing	Negative diagnosis of genetic ataxias, including SCA3 and Friedreich's ataxia

Table 3.1.1: Core features of CANVAS. *VOR – vestibulo-ocular reflex, VVOR – visually enhanced vestibulo-ocular reflex.*

The peripheral sensory deficits associated with CANVAS have been linked to marked dorsal root ganglion (DRG) atrophy, with DRG neuronal loss causing axonal degeneration and secondary demyelination in the posterior columns. Specifically, a loss of neurons from the dorsal root and V, VII and VIII cranial nerve ganglia, therefore suggesting that sensory neuronopathy is involved in a subset of CANVAS patients (52).

There is suggestion of a possible late-onset recessive transmission from reports of CANVAS presenting in siblings (61). However, despite advances in NGS, recent attempts in identifying a genetic basis for CANVAS have been unsuccessful.

3.1.1.1. Recessive expansion in RFC1 is a cause of CANVAS

The work detailed in this study has contributed in part to a project led by Dr. Andrea Cortese which used parametric linkage analysis in siblings to find a causative genetic component implicated in CANVAS. Subsequent significant genetic linkage has been identified on chromosome 4 overlapping a recessive intronic AAGGG repeat expansion in replication factor C subunit 1 (*RFC1*), which has been replicated in sporadic, LOCA cases.

Genotyping was carried out by Dr Cortese and Dr Vandrovcova on 29 individuals from 11 families, comprising 23 affected and 6 unaffected, of which the majority consisted of sibships, with two additional affected first-degree cousins from non-consanguineous families. Importantly, there was no evidence of vertical transmission in any of the families, with all parents remaining unaffected despite living until old age.

Linkage analysis was carried out, assuming a recessive mode of inheritance and a single peak at position 4q14 was identified, which had a cumulative maximum heterogeneity logarithm of the odds (HLOD) score of 5.8. A 1.7MB region between markers rs6814637 and rs10008483 was defined by haplotype analysis, mapping to chr4:38977921-40712231 which contains the same maternal and paternal alleles

within single family affected siblings, compared to unaffected siblings. This region contains 21 known genes, however 4 single nucleotide polymorphisms (SNPs) (rs2066790, rs11096992, rs17584703 and rs6844176) map to a region encompassing all exons of *RFC1*, and the last exon of WD repeat domain 19 (WDR19) were present in all affected individuals across all families, barring one, suggesting a potential founder haplotype.

RFC1 is the large subunit of the RFC complex, which also comprises, RFC2, RFC3, RFC4 and RFC5. RFC is also known as the DNA clamp loader and is key in replication where it binds to the 3' end of the primed nascent DNA strand, loading DNA polymerase γ or ε and proliferating cell nuclear antigen (PCNA) onto the replication forks. Therefore, any dysfunction in RFC1 could impair replication machinery assembly, lead to replication fork collapse and formation of double strand breaks (DSBs) (163). WGS performed by Dr Cortese and Dr Jana Vandrovcova revealed a reduced read depth in a region within the 1.7 Mb region that encompassed a simple tandem $AAAAG_{(11)}$ repeat, located at position chr4:39350045-39350103. Results showed that in affected patients, this $AAAAG_{(11)}$ repeat was replaced with a $AAGGG_{(n)}$ conformation of variable repeat size. This $AAGGG_{(n)}$ repeat appeared to be a biallelic expansion as none of the read spanned the entire microsatellite region. Indeed, results from an unaffected sibling showed an equal distribution of interrupted reads containing both the mutated AAGGG repeat and AAAAG conformation. RPPCR and long-range PCR confirmed the presence of the AAGGG expansion in 63% of 33 patients with sporadic LOCA. However, size estimation of the expansion could not be confirmed by either repeat due to the high error rate associated with PCR as well as the previously demonstrated repeat region contraction during PCR cycling (164).

3.1.2. Aims of study

The primary purpose of this contribution to the discovery of a repeat expansion in *RFC1* as a genetic cause of CANVAS was to further explore the mutated $AAGGG_{(n)}$ repeat expansion identified by Cortese et al., in both affected and normal populations. The goal was to further extend and support the existing data, as well as characterise the expansion repeat. Using a combination of RPPCR, Southern blotting, Sanger sequencing and SNP screening, the pathogenic expansion was identified in CANVAS patients, along with the estimation of repeat size. The frequency and size of the mutated repeat expansion and polymorphic conformations within the normal population was also identified, along with LOCA populations.

During the course of this project, I aimed to optimise the techniques used within the study, specifically Southern blotting, and the workflow of flanking PCR and RPPCR. I had the opportunity to collaborate with the Laing research group from the Harry Perkins Institute of Medical Research in Perth, Australia, working with Dr Sarah Beecroft to confirm *RFC1* repeat expansions in a cohort of New Zealand Māori and Cook Island Māori population reported to have CANVAS (62).

3.1.3. Summary of methods

Study aim	Methods used	Contributors	Methods section
Identification of repeat expansions in CANVAS patients	Flanking PCR RPPCR Southern blotting	Dr Andrea Cortese, Roisin Sullivan Dr Andrea Cortese, Roisin Sullivan Dr Andrea Cortese, Roisin Sullivan	2.2.2, 2.2.3 2.2.4
Identification of repeat expansions in healthy control population	Flanking PCR RPPCR Southern blotting	Roisin Sullivan, Dr Andrea Cortese Roisin Sullivan Roisin Sullivan, Dr Andrea Cortese	2.2.2 2.2.3 2.2.4
Identification of repeat expansions in late-onset ataxia cohort	Flanking PCR RPPCR Southern blotting	Roisin Sullivan Roisin Sullivan Roisin Sullivan	2.2.2 2.2.3 2.2.4
Southern blotting optimisation	Southern blotting	Roisin Sullivan, Dr Wai Yan Yau	2.2.4
SNP haplotype sequencing	Flanking PCR Sanger sequencing	Dr Andrea Cortese, Roisin Sullivan Dr Andrea Cortese, Roisin Sullivan	2.2.2 2.2.8
Māori and Cook Island population investigations	Flanking PCR RPPCR Southern blotting	Roisin Sullivan, Dr Sarah Beecroft Roisin Sullivan, Dr Sarah Beecroft Roisin Sullivan, Dr Wai Yan Yau	2.2.2 2.2.3 2.2.4

3.1.4. Results

3.1.4.1. Biallelic pathogenic repeat expansion in RFC1 identified in all affected CANVAS patients from 11 families, and 16 sporadic cases, confirmed through Southern blotting

Southern blotting revealed that all affected patients (n = 18) from 11 families, and 16 sporadic cases, showed either two discrete bands or one large overlapping band suggestive of the recessive pathogenic AAGGG repeat expansion (figure 3.1.1). Control samples all showed either one band associated with the short reference allele $AAAAG_{(11)}$, approximately 5kb, and/or a slightly raised band corresponding to a benign expanded motif. Expansion repeat size ranged from 400 to 2000 repeating units, however most cases showed expansions of approximately 1,000 repeating units. Throughout families, repeat size appears relatively stable between siblings. Expansion sizes were estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used.

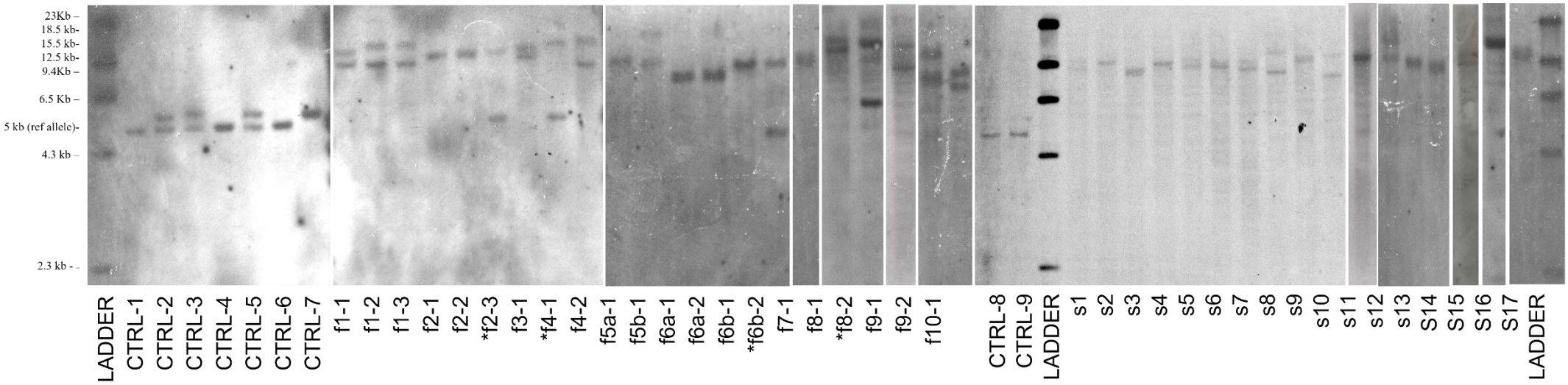


Figure 3.1.1: Southern blot of 18 patients from 11 families and 16 sporadic cases. Two discrete or one large overlapping band can be seen for patients that span between 7 and 15 kb. " * " denotes unaffected siblings, where one expanded allele can be seen due to carrier status, with one additional band within the normal range. CTRL indicates control samples that correspond with the 5 kb band that is the estimated size for the short reference $AAAAG_{(11)}$ allele, however larger bands can also be seen, indicated benign expansions. Expansion sizes were estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used.

Ladder II refers to DIG-labelled DNA Molecular Weight Marker II (Roche) that contain 8 fragments of base pair lengths 125, 564, 2,027, 2,322, 4,361, 6,557, 9,416, 23,130 bp. Ladder III refers to DIG-labelled DNA Molecular Weight Marker III (Roche) containing 13 fragments of base pair lengths 564, 831, 947, 1,375, 1,584, 1,904, 2,027, 3,530, 4,268, 4,973, 5,148, and 21,226 bp.

This work was done in collaboration with Dr Andrea Cortese.

Work published in Cortese et al., 2019 (2).

3.1.4.2. Allelic and polymorphic conformations of the repeat expansion in the normal population

To gauge the frequency of the repeat expansion within the normal population, targeted flanking PCR and RPPCR was performed on 304 healthy controls (figure 3.1.2). The chr4:39350045-39350103 locus was shown to be highly polymorphic within the normal population and three other conformations were observed; $AAAAG_{(11)}$, $AAAAG_{(n)}$ and $AAAGG_{(n)}$ (table 3.1.2 and table 3.1.3). Three targeted primer sets were used for flanking PCR and RPPCR, as three polymorphic conformations; wildtype sequence AAAAG, variant AAAGG and pathogenic AAGGG were screened. The pathogenic recessive AAGGG expansion, defined by a combination of positive RPPCR for the AAGGG repeat conformation and lack of PCR amplifiable product on flanking PCR, was not observed in 304 healthy controls. 0.7% (4 out of 608 chromosomes screened) of the normal population were shown to have AAGGG expansion repeat in the heterozygous state, as defined by the absence of a PCR amplifiable product on flanking PCR, a positive RPPCR for the AAGGG expansion and another non-pathogenic conformation for the second allele. The allelic distribution for the $AAAGG_{(n)}$ conformation was 7.9% (n = 48), 75.5% (n = 459) for the $AAAAG_{(11)}$ allele and 13% (n = 79) for the $AAAAG_{(n)}$ allele.

To estimate mean expansion repeat size, a selection of samples that showed no amplifiable product on flanking PCR and a positive RPPCR for any of the three primer sets were run on a Southern blot (figure 3.1.3 A-E, figure 3.1.4). To aid size estimate of the AAAGG and AAAAG expansion conformations a clamp was included in these targeted primers, attaching to the start of the sequence adjacent to the expansion repeat region and picked up on RPPCR. The Southern blots revealed an estimated repeat size for the $AAGGG_{(n)}$ expansion ranging from 200 to 900 repeats (mean 740 ± 304), 15 to 200 repeats (mean 72 ± 43) for the $AAAAG_{(n)}$ expansion and 40 to 1000 repeats (mean 173 ± 239) for the $AAAGG_{(n)}$ expansion) (figure 3.1.3). In addition to enabling repeat size estimation, the Southern blots highlighted that some cases likely had two expansions of the same repeating unit, with two distinct bands on the Southern blot, a lack of amplifiable PCR product on flanking PCR and a positive RPPCR result for either the AAAGG or AAAAG primer set. 22 cases likely had two $AAAAG_{(11)}$ expansion repeats and nine samples likely had two $AAAGG_{(11)}$ expansions. However, it cannot be ruled out that another polymorphic conformation could be present on either of the two alleles which was not targeted in the present study. Indeed 9 additional subjects had no amplifiable product on flanking PCR and a negative RPPCR result for all three primer sets, suggesting that they potentially contain another allelic polymorphic conformation, which would account for 3% (n=18) of screened chromosomes. The limitation of insufficient DNA did not allow Southern blots in these cases.

The RPPCR results targeting AAGGG and AAAAG showed several cases that contained interruptions in repeat expansion, particularly for $AAAGG_{(n)}$, (figure 3.1.4), suggesting that there are other polymorphic conformations interrupting the targeted repeat. These included bursts of targeted repeat at the start, followed by interruptions in the middle of the fragment analysis. Targeted sanger sequencing

picked up some interrupting polymorphic repeats including AAAGGG and AAAAGG interspersed in some samples analysed with interruptions on RPPCR.

It was shown that a proportion of the healthy screen cohort had biallelic repeat expansions of a distinct repeated conformation; one subject had $AAAAG_{(n)}$ and $AAGGG_{(n)}$, six cases had $AAAAG_{(n)}/AAAGG_{(n)}$ and one case of $AAAGG_{(n)}$ and $AAGGG_{(n)}$ (table 3.1.3).

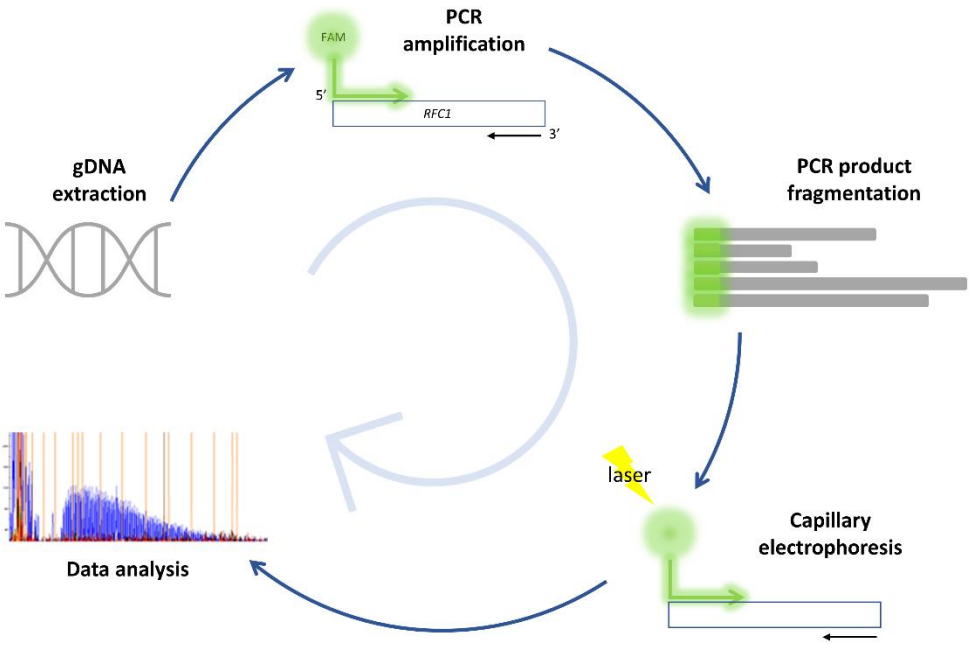


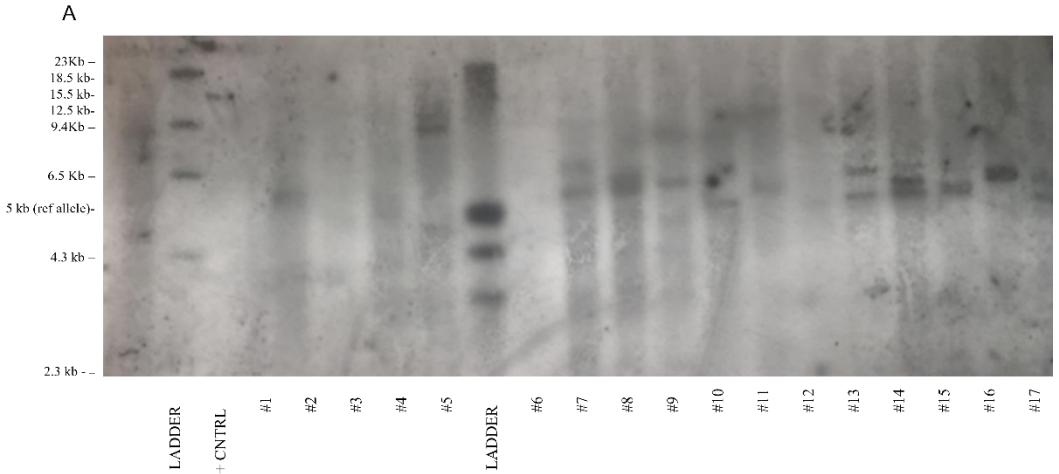
Figure 3.1.2: Schematic of repeat-primed PCR (RPPCR) method. gDNA is extracted from patient blood and undergoes a PCR reaction using fluorescently labelled FAM primers that fluorescently tag the 5' end of the gene of interest. The amplified PCR product undergoes fragmentation, resulting in fragments of different lengths, using HiDi formamide. A size standard (LIZ-500) is also added at this stage to allow sizing of sample peaks and allow for spectral calibration for injection variations in the next step. This fragmented and denatured PCR product then undergoes capillary electrophoresis in a 3730 DNA analyzer, which detects the FAM dye tagged onto the 5' end of the DNA fragments. During capillary electrophoresis, the PCR products are injected electrokinetically into polymer-filled capillaries and a high voltage is applied. The fluorescent DNA fragments are separated by size, which are then detected by a laser system. Data was analysed using GeneMapper V5 software, which precisely measures the size of fragments. Data analysis then occurs using the electropherograms produced which are a plot of DNA fragment sizes A positive trace contains a large area of fragmented peaks, that correspond to the dye, measured in RFU, that are sized along the x-axis and commonly appear in a sawtooth pattern of decremental peaks. *gDNA, genomic DNA; PCR, polymerase chain reaction; RFC1, replication factor C subunit 1; FAM, 5' 6-FAM (Fluorescein), RFU, relative fluorescence units*

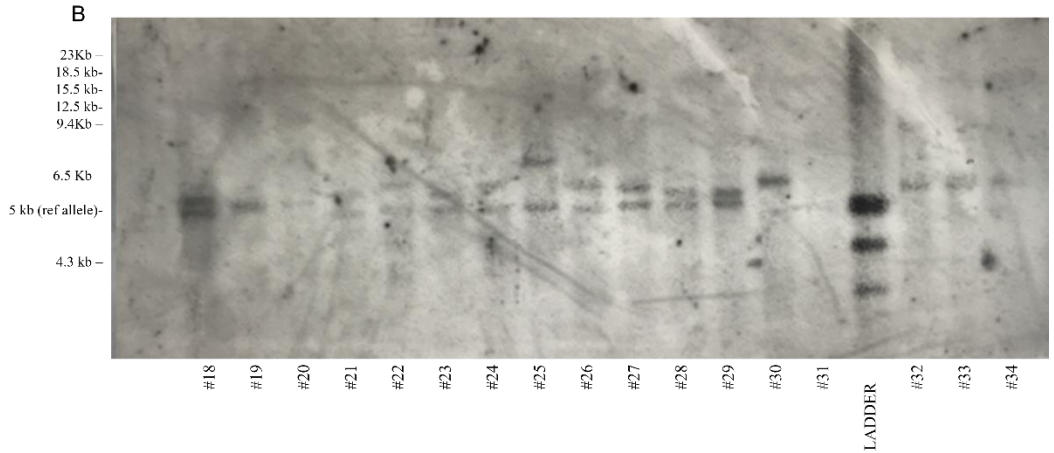
Conformation	Number of alleles with conformation	Frequency in normal population (%)	Average repeat size
Homozygous $AAGGG_{(n)}$	0	0	-
$AAGGG_{(n)}$	4	0.7%	740 ± 304
$AAAGG_{(n)}$	48	7.9%	173 ± 239
$AAAAG_{(11)}$	459	75.5%	-
$AAAAG_{(n)}$	79	13%	72 ± 43

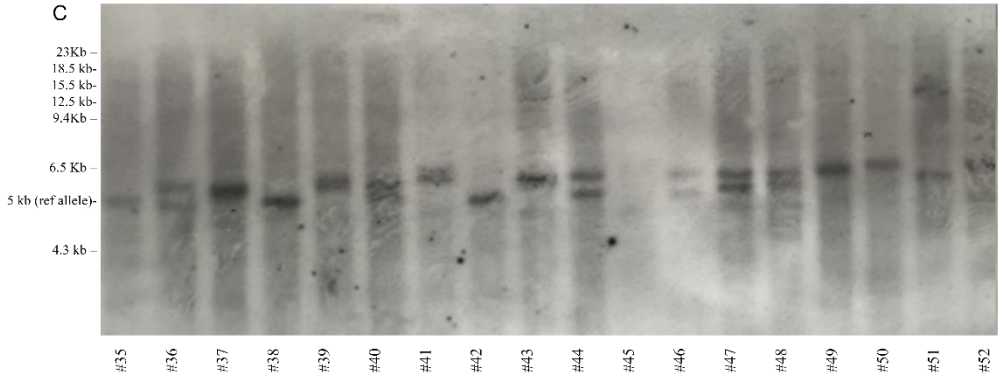
Table 3.1.2: Outline of allelic distribution and average repeat size in normal control population (n = 304). Mean and Standard deviation (S.D) used for average repeat size.

Biallelic expression	Number of cases
$AAAAG_{(n)}/AAGGG_{(n)}$	1
$AAAAG_{(n)}/AAAGG_{(n)}$	6
$AAAGG_{(n)}/AAGGG_{(n)}$	1
$AAAAG_{(n)}/AAAAG_{(n)}$	22
$AAAGG_{(n)}/AAAGG_{(n)}$	9

Table 3.1.3: Number of cases from healthy screen cohort with distinct biallelic conformations







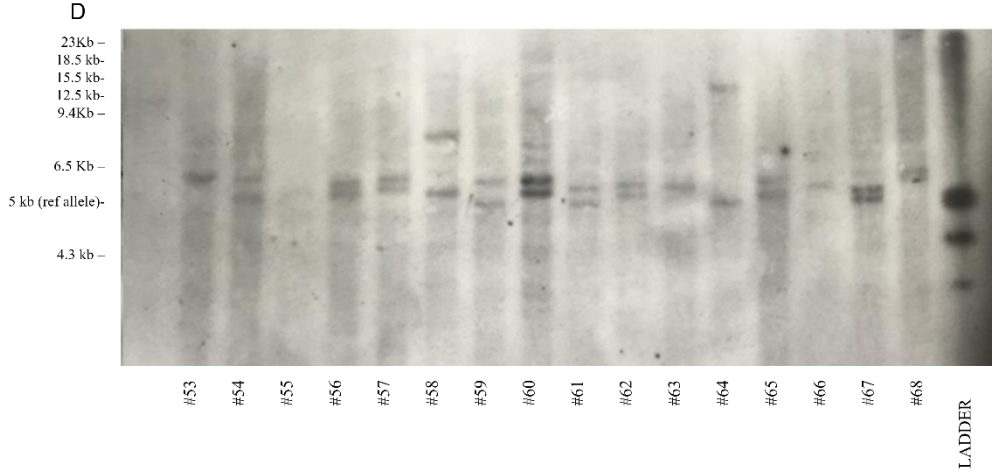
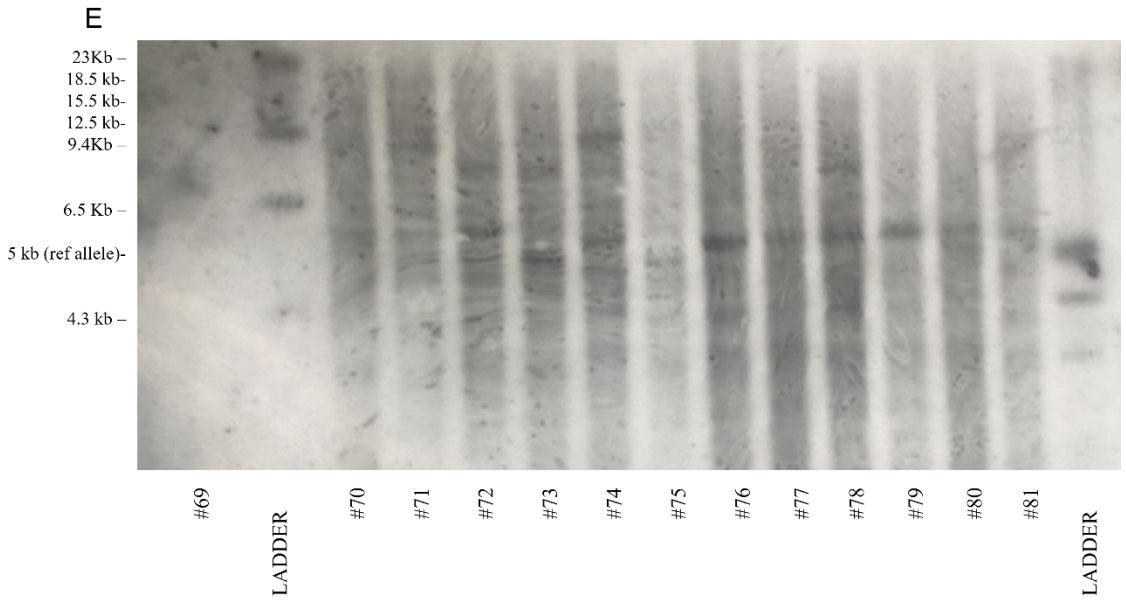


Figure 3.1.3 continued next page.

Figure 3.1.3 cont.



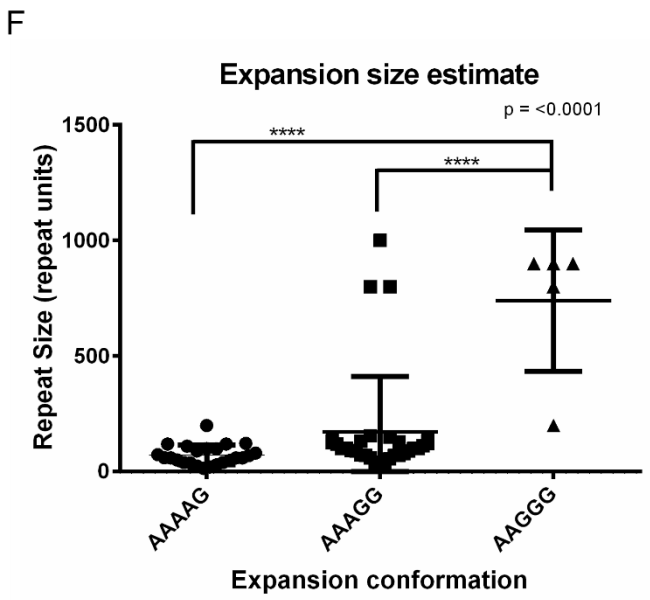


Figure 3.1.3: Southern blot sizing of expansions in selected healthy control samples. A-E, to estimate the size of expansion repeat in the healthy control population cohort Southern blots were performed for a selection of $AAAAG_{(n)}$, $AAAGG_{(n)}$ and $AAGGG_{(n)}$ samples. Expansion sizes were estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used. Ladders refer to DIG-labelled DNA Molecular Weight Marker II (Roche) that contain 8 fragments of base pair lengths 125, 564, 2,027, 2,322, 4,361, 6,557, 9,416, 23,130 bp and DIG-labelled DNA Molecular Weight Marker III (Roche) containing 13 fragments of base pair lengths 564, 831, 947, 1,375, 1,584, 1,904, 2,027, 3,530, 4,268, 4,973, 5,148, and 21,226 bp. F, to estimate expansion size, a selection of $AAAAG_{(n)}$, $AAAGG_{(n)}$, and all $AAGGG_{(n)}$ samples were Southern blotted, with estimated sizes correlated with primer clamp peak sizes for $AAAAG_{(n)}$ and $AAAGG_{(n)}$ for improved accuracy. Mean expansion size for $AAAAG_{(n)}$ was 72 ± 43, $AAAGG_{(n)}$ was 173 ± 239 and for $AAGGG_{(n)}$ the mean was 740 ± 304. S.D was used for error bars. One-way ANOVA with post-test Tukey performed, with $p < 0.05$ deemed significant.

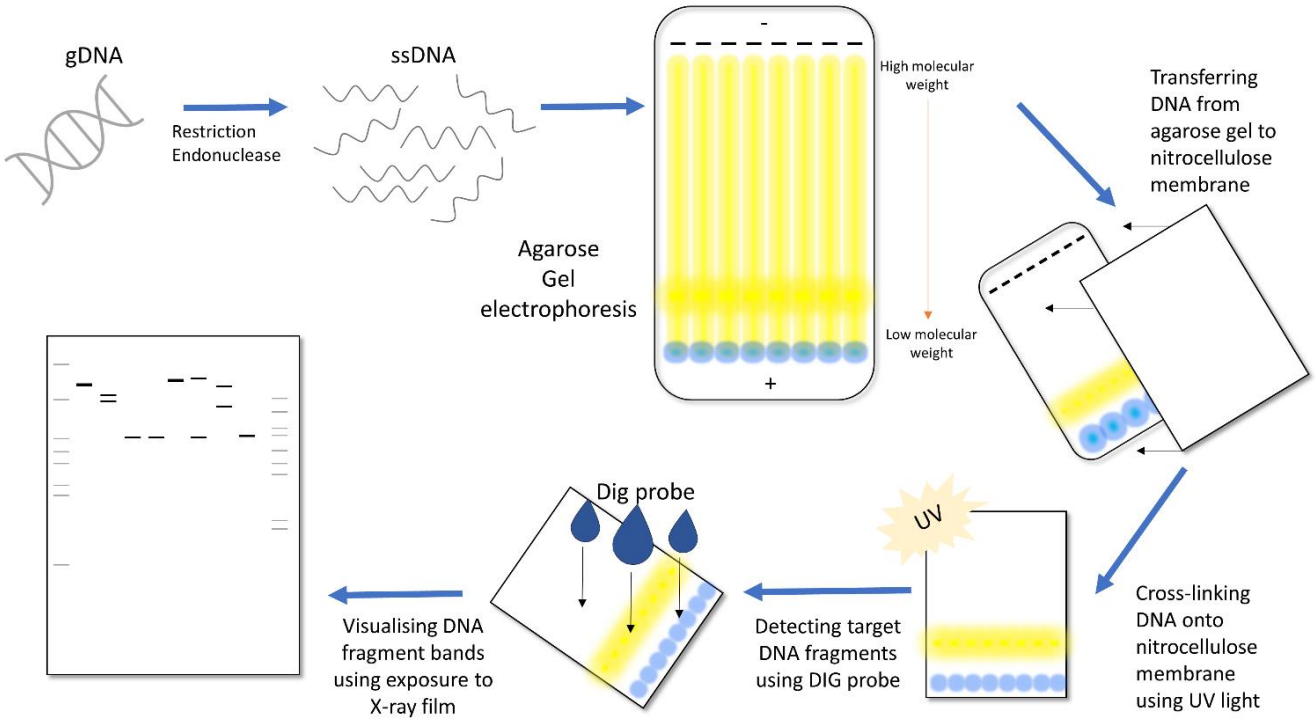


Figure 3.1.4: Schematic of Southern blotting protocol. Restriction endonucleases are used to turn genomic DNA, into single-stranded DNA, which then undergoes agarose gel electrophoresis overnight. The DNA within the agarose gel is then transferred onto a nitrocellulose membrane by using a concentration gradient that moves the DNA from a higher concentration gradient (agarose gel) to a lower concentration gradient (nitrocellulose membrane). The DNA is cross-linked to the membrane via UV light exposure. The membrane undergoes hybridisation with a DIG probe that targets the specific DNA region where the expansion lies. DIG is linked to the C-5 position of uridine nucleotides and can be incorporated into nucleic acid probes, which bind to the target region. It is detected using an anti-DIG antibody. Chemiluminescence and X-ray film exposure is the final step to visualise the bands that correspond to repeat expansions. Ladders are used to accurately size the expansions. *DIG, Digoxigenin*

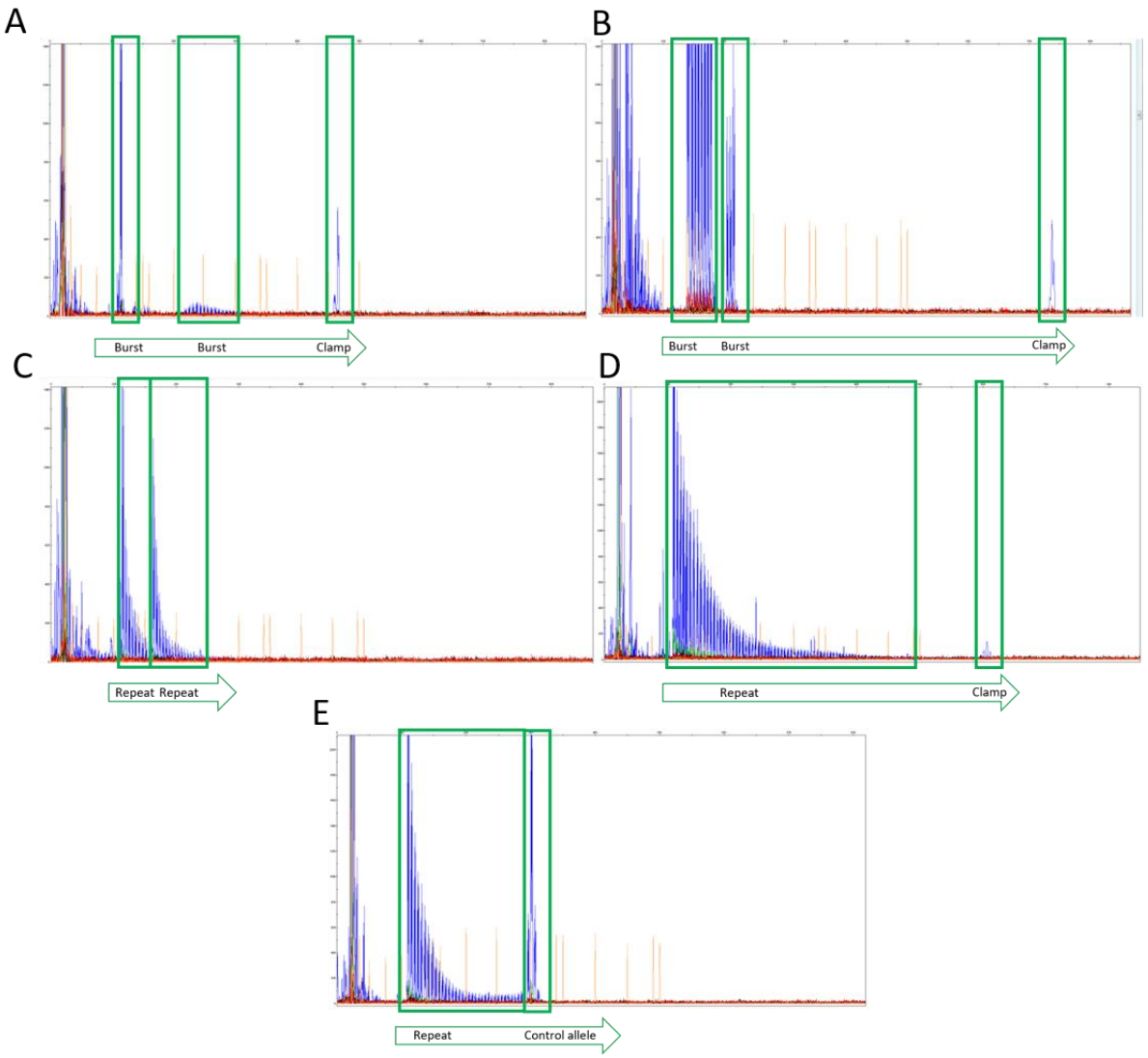


Figure 3.1.5: RPPCR trace examples from healthy control screen cohort. Peaks are sized along the X-axis in base pairs (bp), and the Y-axis indicates the intensity of the laser in relative fluorescent units (RFU). A, a sample with a positive AAAGG expansion, with bursts of amplification with bouts of 2 clear interruptions and primer clamp. B, a positive AAAGG RPPCR with two distinct amplification bursts followed by a long sequence interruption and then primer clamp. C, a sample with a clear homozygous $AAAGG_{(n)}$ expansion with both expansions picked up on RPPCR. D, a positive RPPCR result with the classic sawtooth pattern and the primer clamp further downstream. E, a sample with a positive RPPCR for $AAAAG_{(n)}$ expansion, with the normal $AAAAG_{(11)}$ allele being picked up, seen by a peak at the end of the expanded sawtooth pattern.

3.1.4.3. Southern blot optimisation

The Southern blotting method proved to be unreliable in giving clean, informative blots, requiring frequent reblotting of samples. Inadequate blots included large smearing of samples (figure 3.1.6 Ai) and significant background impairing visibility of expanded bands (figure 3.1.6 Aii). In order to reduce the need for repeating Southern blotting, optimisation was performed alongside Dr Wai Yan Yau, testing different concentrations of saline-sodium citrate (SSC) and sodium dodecyl sulphate (SDS) (figures 3.1.6 Bi and Bii). In addition, the incubation temperature of membrane hybridisation was also optimised. The combination of 0.1x SSC and 0.5% SDS proved to be the most effective at reducing background and producing clear blots with visible bands (figure 3.1.6 Bii & C).

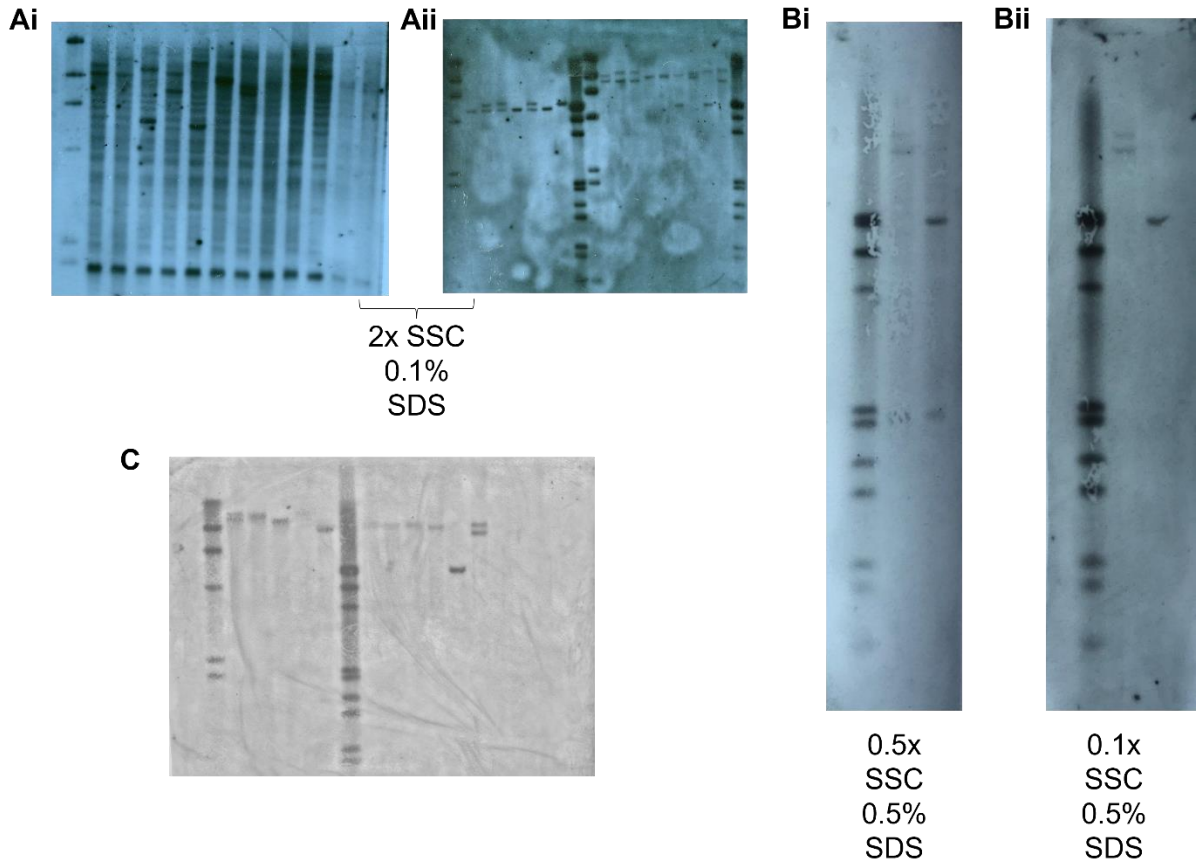


Figure 3.1.6: Optimisation of the Southern blotting protocol. Ai and Aii show poor-quality blots frequently yielded from the original protocol that included high background and smeared bands. Bi and Bii show two stages of optimisation, with the testing of SSC and SDS concentration combinations. C shows the use of 0.1x SSC and 0.5 % SDS. *SSC, saline-sodium citrate; SDS, sodium dodecyl sulphate*. This work was done in collaboration with Dr Wai Yan Yau.

3.1.4.4. SNP Haplotyping

To explore the haplotype most associated with a particular expansion, genotyping using two disease-associated haplotype markers rs11096992 (AA) and rs2066790 (AA) was performed on all healthy control screen samples, CANVAS patients with the expansion and late-onset ataxia patients also with the expansion. Dr Cortese also looked for the haplotype most associated in patients with the AAGGG repeat expansion within the 1000genome control population and found an allelic carrier frequency of 18%. Within the healthy screen cohort, both disease-associated haplotypes rs11096992 (AA) and rs2066790 (AA) were absent in recessive state in samples carrying two $AAAAG_{(11)}$ alleles, two $AAAAG_{(n)}$ alleles and a biallelic $AAAAG_{(11)}/AAAAG_{(n)}$ genotype (figure 3.1.7). However, it was observed in three subjects with two $AAAGG_{(n)}$ expansions and one subject with $AAGGG_{(n)}/AAAGG_{(n)}$ alleles. This suggests its possible involvement with both $AAGGG_{(n)}$ and $AAAGG_{(n)}$ conformations of the expansion, but not $AAAAG_{(11)}$ or $AAAAG_{(n)}$.

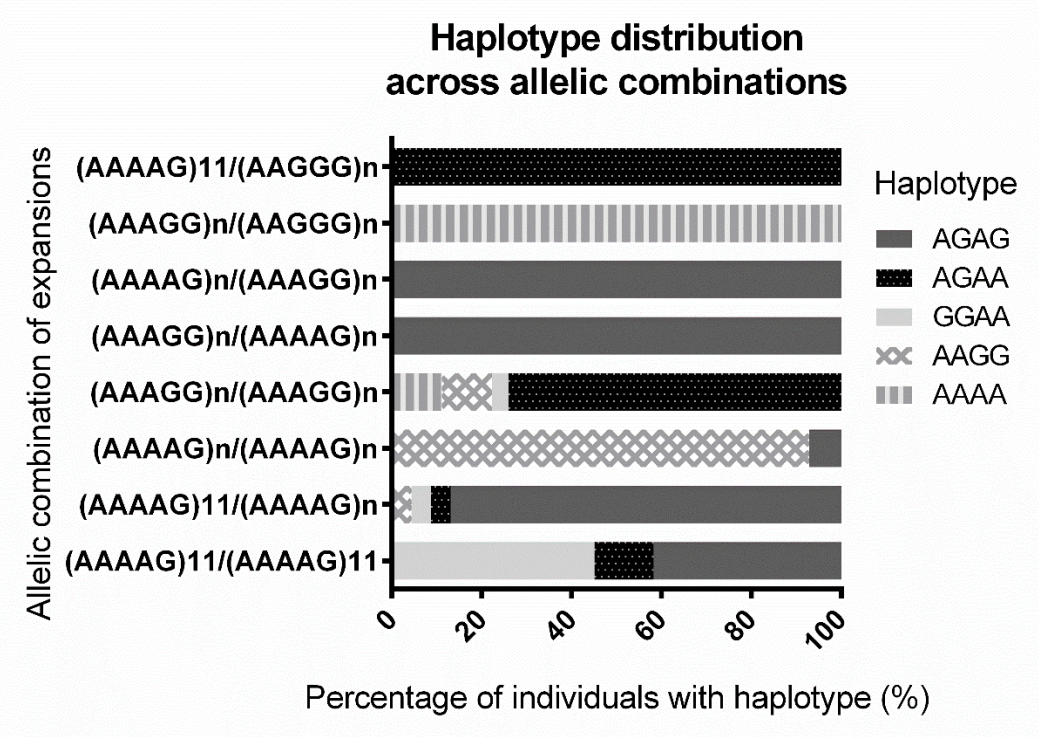


Figure 3.1.7: Haplotype distribution of biallelic combinations within the healthy control population. SNP haplotype screening was carried out on the healthy control population using two disease-associated haplotype markers rs11096992(AA) and rs2066790 (AA) to determine the percentage of individuals with certain haplotypes (x-axis), specific to certain allelic repeat motif combinations (y-axis). The recessive disease-causing AA and AA were absent in recessive state in samples carrying two $(AAAAG)_{11}$ alleles, two $(AAAAG)_n$ alleles and a biallelic $(AAAAG)_{11}/(AAAAG)_n$ genotype. However, it was observed in three subjects with two $(AAAGG)_n$ expansions and one subject with $(AAGGG)_n/(AAAGG)_n$ alleles. Figure key shows possible haplotype combinations of rs11096992(AA) and rs2066790 (AA).

3.1.4.5. Allelic and polymorphic conformations of the repeat expansion in late-onset ataxic population

To investigate the frequency of the AAAAG, AAAGG and AAGGG polymorphisms within a generalised late-onset ataxic population, targeted flanking PCR and RPPCR was employed on 665 samples (1330 chromosomes) from patients clinically diagnosed with ataxia, with a high proportion experiencing late-onset symptoms. From this cohort the frequency of the $AAGGG_{(n)}$ conformation was 3.1% (n = 41), 0.5% for $AAAGG_{(n)}$ (n = 7), 0.2% for $AAAAG_{(n)}$ (n = 2) and 91.9% for $AAAAG_{(11)}$ (n = 1222) (table 3.1.4). Within the cohort, 29 samples (58 chromosomes), frequency of 4.3%, had no product on flanking PCR and were RPPCR negative across the board, suggesting that they may have another expansion conformation not currently considered by RPPCR primers, similarly to the control screen cohort. 14 samples (2.1% of overall allelic distribution and 51.8% of AAGGG allelic distribution) were homozygous for the recessive $AAGGG9_{(n)}$. Southern blots were used to estimate expansion size for a selection of samples where gDNA was available (figure 3.1.8 A and B). These, alongside clamps for AAAGG and AAGGG conformations, showed that the expansion size for AAGGG was statistically significantly larger than AAAGG ($p = 0.008$), ranging from 300 to 2000 repeat units (mean 1183 ± 618), in contrast the size of AAAGG expansion ranged from 100 to 330 repeat units (mean 254 ± 111) and the AAAAG expansion was present with two 100 repeat units (figure 3.1.8).

Conformation	**Number of alleles with conformation**	**Frequency in ataxic population (%)**	**Average repeat size**
$(AAGGG)_n$	41	3.1%	1263 ± 762
$(AAAGG)_n$	7	0.5%	254 ± 111
$(AAAAG)_{11}$	1222	91.9%	-
$(AAAAG)_n$	2	0.2%	100
Undetermined	58	4.3%	-

Table 3.1.4: Outline of allelic distribution and average repeat size in ataxic population (n = 666). Mean and S.D used for average repeat size.

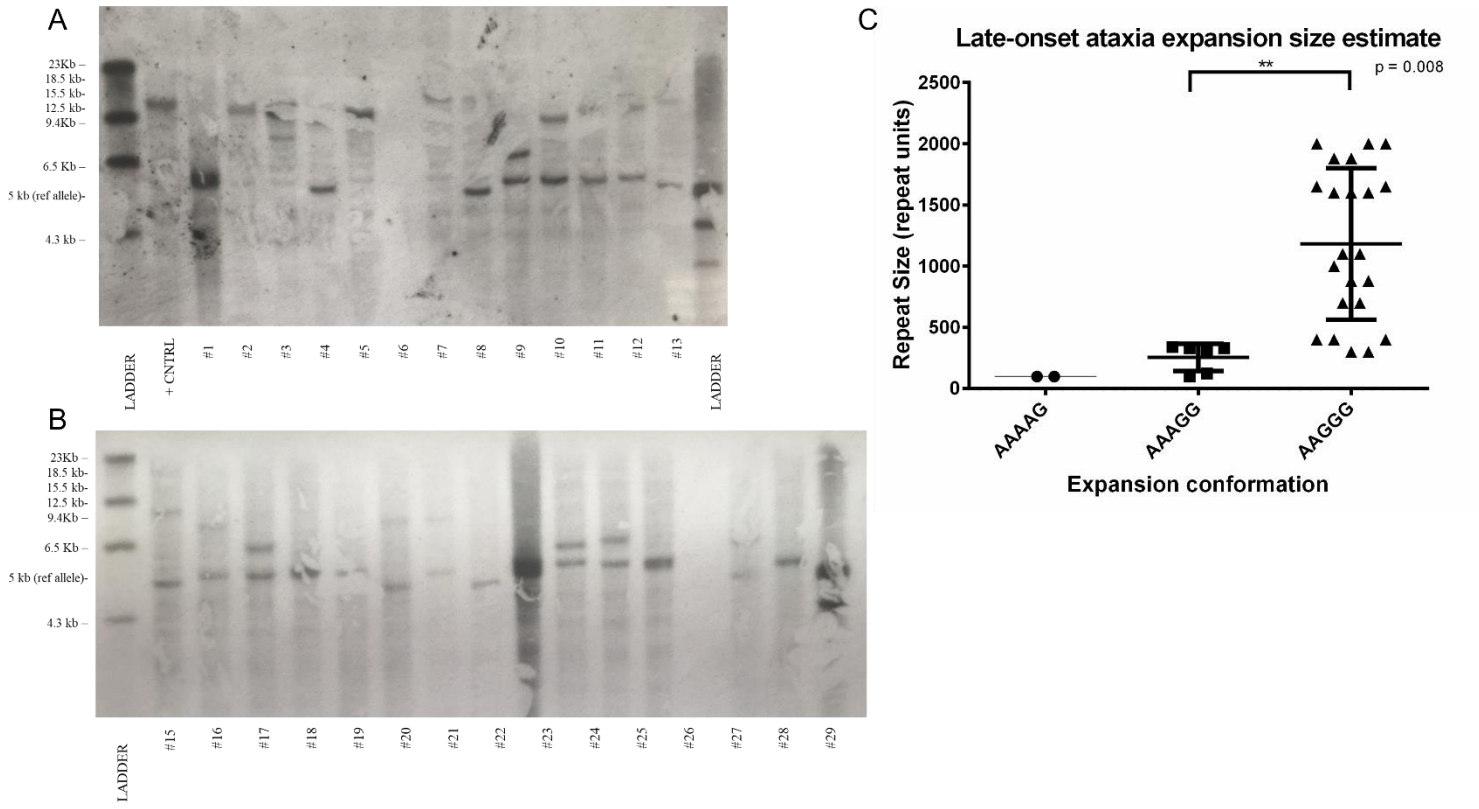


Figure 3.1.8: Southern blots of late-onset ataxia cohort. A & B, to estimate the size of the expansion repeat in late-onset ataxic populations Southern blots were performed for a selection of $(AAAAG)_N$, $(AAAGG)_N$ and $(AAGGG)_n$ samples. Expansion sizes were estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used. Ladders refer to DIG-labelled DNA Molecular Weight Marker II (Roche) that contain 8 fragments of base pair lengths 125, 564, 2,027, 2,322, 4,361, 6,557, 9,416, 23,130 bp and DIG-labelled DNA Molecular Weight Marker III (Roche) containing 13 fragments of base pair lengths 564, 831, 947, 1,375, 1,584, 1,904, 2,027, 3,530, 4,268, 4,973, 5,148, and 21,226 bp . C, expansion size for AAGGG ranged from 300 to 2000 repeat units (mean 1183 ± 618), the size of AAAGG expansion ranged from 100 to 330 repeat units (mean 254 ± 111) and the AAAAG expansion was present with two 100 repeat units. S.D. used for error bars. One-way ANOVA with post-test Tukey performed, with $p < 0.05$ deemed significant.

3.1.4.6. Māori and Cook Island Māori CANVAS patients have confirmed repeat expansions in RFC1, including a rare case with very early-onset CANVAS

Through the success of the optimised workflow for the detection of the *RFC1* repeat expansion, I collaborated with Dr Sarah Beecroft from the Laing research group at Harry Perkins Institute of Medical Research, Perth, Australia, to demonstrate and teach the workflow of *RFC1* expansion detection. In collaboration with Dr Wai Yan Yau, I carried out Southern blotting on four interesting Māori and Cook Island Māori patient samples from their research study. All fifteen of the cohort tested positive for the recessive pathogenic repeat motif $AAGGG_{(n)}$, as well as the benign $AAAGG_{(n)}$ on RPPCR, with a specific pattern of $AAAGG(_{10\text{-}25)}$ $AAGGG_{(n)}$ $AAAGG_{(n)}$ (data not shown, but available in the Beecroft et al., publication (1)). Figures 3.1.9 and table 3.1.5 depict the Southern blot for these samples, with expanded bands present for all patients: M2 III:4 showed two expanded bands of 12,231 kb and 13736k kb (2643 and 1944 repeating units respectively), CI1 11:1 showed a thick band at 9464 kb (1090 repeating unit), M5 I:1 a thick band at 11, 398 kb (1476 repeating units) and M5 I:1 a thick band at 8957 kb (988 repeating units). The mean age of onset (excluding M2 V:1) was 55 years, M2 V:1 had an earlier onset of 6 years of age, which was of notable interest due to the atypical earlier disease onset and progression.

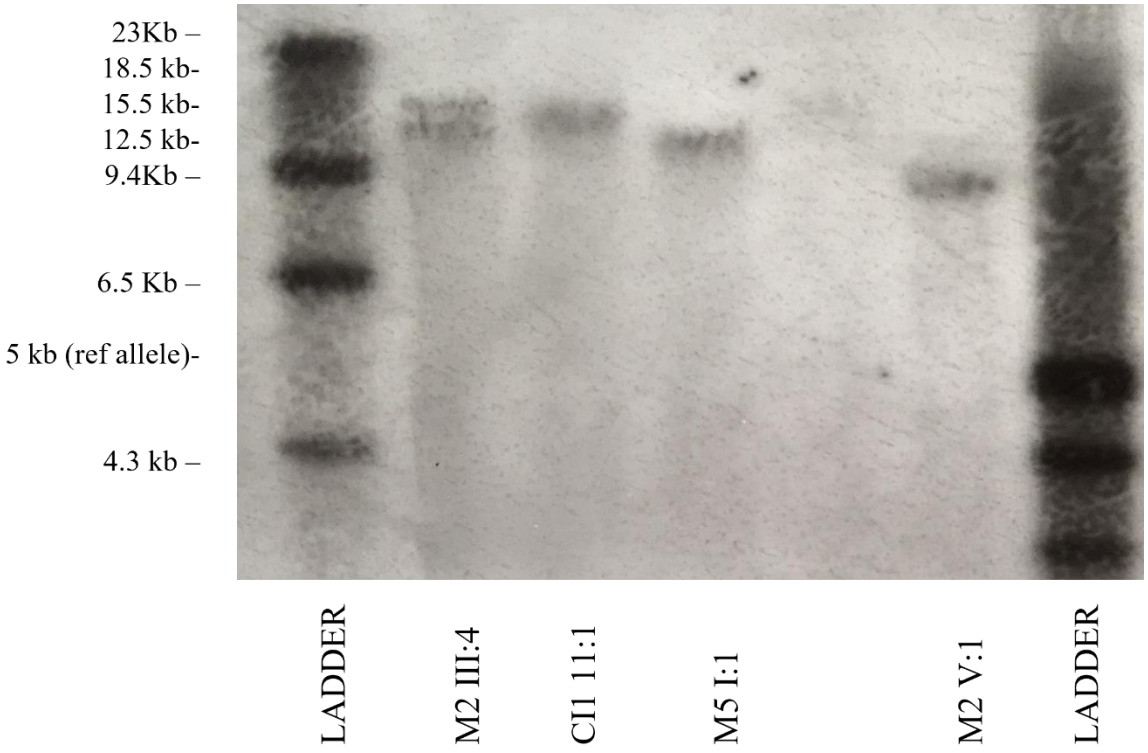


Figure 3.1.9: Southern blot of Māori /Cook Island Māori patients. Southern blot showing pathogenic expansions for Māori /Cook Island Māori patients. Ladders refer to DIG-labelled DNA Molecular Weight Marker II (Roche) that contain 8 fragments of base pair lengths 125, 564, 2,027, 2,322, 4,361, 6,557, 9,416, 23,130 bp and DIG-labelled DNA Molecular Weight Marker III (Roche) containing 13 fragments of base pair lengths 564, 831, 947, 1,375, 1,584, 1,904, 2,027, 3,530, 4,268, 4,973, 5,148, and 21,226 bp . Data published in Beecroft et al., (1)

Sample	Ethnicity	Family History	Allele size (kb)	Repeating unit size
M2 III:4	Māori	6 affected family members	12200 / 13700	1600 / 2000
CI1 11:1	Cook Island Māori	2 affected family members	9500	1000
M5 I:1	Māori	Singleton	11400	1500
MV V:1	Māori	Singleton	9000	1000

Table 3.1.5: Genetic information and expansion sizing for Māori /Cook Island Māori patients. Expansion sizes estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used. *kB, kilobases.* This work was done in collaboration with Dr Wai Yan Yau. Data published in Beecroft et al., (1).

3.1.5.Discussion

Through this work, a recessive pentanucleotide repeat expansion within intron 2 of *RFCI* was identified as the cause of CANVAS and late-onset ataxia (2). Within 11 families and 33 sporadic cases, 23 cases carried the recessive pathogenic AAGGG expansion repeat and further to this 46% of all cases and 65% of familial cases had the complete CANVAS syndrome, with all symptoms: cerebellar ataxia, vestibular areflexia and sensory neuropathy. The frequency of CANVAS patients with the biallelic AAGGG expansion is similar to that of Friedreich's ataxia expansion in recessive ataxia populations.

Within this study I determined that within the normal population (normal control cohort) the allelic carrier frequency was 0.7%. This is similar to the allelic carrier frequency of the *FXN* gene GAA expansion which is estimated at 1/120 in the Caucasian population (165). Since the publication of this study, the frequency within the normal population has been estimated to be between 0.7–6.8% (2, 166-169). The recessive GAA expansion, similarly to the AAGGG expansion which causes CANVAS, causes FRDA which is the most common recessive cerebellar ataxia.

A limitation of this study was the relatively small sample size (304 individuals) used as a representation of the general population. Allelic frequency estimation is often hampered by a lack of power, with larger cohorts always having a deviation from the true allelic frequency. This can be evidenced by the current carrier frequency estimation between 0.7% to 6.8%, which indicates variation between findings. Future work should endeavour to screen a much larger sample size as a truer representation of the general population.

Within this study I also identified that within a late-onset ataxia population, the frequency of the biallelic recessive $AAGGG_{(n)}$ frequency was 2.1%, with a carrier frequency of 0.9%. This suggests that the

recessive AAGGG expansion in *RFC1* may be a frequent cause of late-onset ataxia, and patients that harbour this expansion may be misdiagnosed. Detailed analysis of the clinical information of these patients positive for the biallelic expansion is crucial to understanding the true clinical spectrum of CANVAS, and therefore correct diagnosis. This work led to a subsequent follow-up study using the patients identified with biallelic *RFC1* expansions within the cohort of late-onset ataxia patients as well as further positive cases from centres in New Zealand, France, Italy, Brazil, Slovenia, Australia, and the UK. This was necessary to determine the full prevalence of the expansion in patients to fully characterise the phenotypic spectrum, currently the estimated prevalence of the biallelic recessive trait is approximately 1/20,000. The biallelic repeat expansion in *RFC1* was identified in 90% of patients with full CANVAS symptoms and 14% of patients with adult-onset ataxia (53). The clinical evaluation of these patients characterised the clinical spectrum seen in CANVAS patients and those with *RFC1*-associated ataxia, including an unexplained spasmodic dry cough in 60% of patients, MRI and clinical signs of cerebellar involvement in 66%, vestibular symptoms in 33% and signs of peripheral neuropathy in approximately 100% of patients on nerve conduction study (NCS) (53). This shows the wide-ranging spectrum of clinical features associated with *RFC1*-associated disease.

The initial CANVAS patients studied by Dr Cortese showed $AAGGG_{(n)}$ repeats ranging from 400 to 2000 repeats, with the majority having expansions of 1000 repeats (2). This mirrors the biallelic expansions seen within this study in the late-onset ataxia population, with the largest expansion encompassing 2000 repeats and the smallest 400. The healthy control cohort showed a mean $AAAGG_{(n)}$ expansion size of 700 repeats, suggesting larger repeat expansions, as well as being biallelically expressed, are necessary for pathogenesis. By contrast the $AAAAG_{(n)}$ and $AAAGG_{(n)}$ mean expansion sizes were smaller in size, with the mean $AAAGG_{(n)}$ expansion size of 173 in healthy controls and 254 in late-ataxic patients being larger than that of $AAAAG_{(n)}$; 74 and 100 mean repeats for healthy controls and late-ataxic patients, respectively. Further characterisation of the expansion locus is required, as despite targeting the additional $AAAAG_{(n)}$ and $AAAGG_{(n)}$ conformations identified in the control population, several samples with no product on flanking PCR were negative for these expansions, suggesting that there are more conformations yet to be identified. Indeed, sanger sequencing of these samples highlighted additional AAGAG, AAAAGG and AAAGGG, of which the latter two were interruptions of a sample with a $AAAGG_{(n)}$ repeat.

A key component of the workflow utilised within this study was the Southern blot which became a limitation due to the challenging and unreliable nature of the method. Southern blotting is regarded as the gold standard of detecting repeat expansions within genomic DNA, due to the ability to visualise and size expansions. The process of conducting a Southern blot spans four days and is extremely labour intensive, therefore getting a poor-quality blot which required repeating a blot became very time-limiting and inefficient. Examples of poor-quality blots include issues with faint bands, high background, and smeared bands, which made sizing of bands difficult and were not of publication

quality. In order to combat this, I undertook an optimisation of the method alongside Dr Wai Yan Yau. SDS is a detergent and SSC is a hybridization buffer and is used to control stringency for the washing steps that occur to wash off non-specific bound DIG probe during the Southern blotting protocol. We tried a number of different concentration combinations of SDS and SSC, including 0.5x SSC and 0.5% SDS and 0.1x SSC and 0.5% SDS, with the latter concentration combination yielding a clean background and clearly defined bands. The original concentrations were 2x and 0.5% respectively. The lower concentration of SSC (and lower stringency of washing) and higher concentration of detergent removed the non-specific DIG probe bound to non-target DNA fragments and the salmon sperm DNA used to block the membrane. In addition, we increased the hybridisation temperature from 46 ºC to 49 ºC to increase specificity of the DIG probe and aid hybridisation. These changes have now been implemented into the standard protocol for *RFC1* Southern blotting, with the method optimised even further to include an overnight transfer step (as opposed to 4 hours) and a 5 hour hybridisation incubation (as opposed to 3 hours) (170). One factor still remains that hinders Southern blotting reliability, and this is the quality and amount of starting gDNA. Southern blotting requires a very large amount of DNA, 5 µg, which is difficult to obtain from current the sample biobank. However, using less than this results in extremely faint bands that are difficult to size, and in circumstances where very low amounts are only available, no bands are seen at all. Currently, there is not any way of counteracting this effect apart from fresh blood collections, which remains a limitation of the method.

Due to the optimisation of the *RFC1* detection workflow, a collaboration with Dr Sarah Beecroft from the Laing research Group at the Harry Perkins Institute of Medical Research in Perth was set-up to teach the workflow and carry out the sizing of expansions in a Māori population. Previous to this, only European cohorts had been studied in the original discovery cohort and late-onset ataxia cohort within this study. All fifteen of the cohort tested positive for the *RFC1* pathogenic recessive $AAGGG_{(n)}$ motif on RPPCR and four of the thirteen individuals from the cohort with enough genomic DNA underwent Southern blotting, revealing large *RFC1* repeat expansions. Remarkably, a novel pathogenic motif configuration was identified, $(AAAGG)_{10\text{-}25}$ $AAGGG_{(n)}$ $AAAGG_{(n)}$ in all fifteen patients whereby a stretch of benign $AAAGG_{(n)}$, varying in unit number, preceded the pathogenic $AAGGG_{(n)}$ on at least one allele (80% harboured this novel configuration on both alleles) (1). One of the patients within the cohort (M2 V:1) was an interesting case, as the patient presented with initial symptoms at 6 years of age, which is significantly younger than the typical age of onset for CANVAS. It is speculated that this patient in question represents a variant, early onset form of the *RFC1*-associated CANVAS, presenting with two of the three distinctive signs of CANVAS: bilateral vestibular failure and ataxia (1). One patient from this cohort also presented with rapid eye movement sleep behaviour disorder, which has previously not been associated with CANVAS , or *RFC1*-associated disease (1). Since the publication of both of these studies, more conformations have been identified that include $AGAGG_{(n)}$ and

$AAGAG_{(n)}$, that currently have uncertainty regarding their possible pathogenicity (167). A likely pathogenic recessive $ACAGG_{(n)}$ was also identified in Asian Pacific and Japanese cohorts (171, 172).

A large proportion of the $AAAGG_{(n)}$ positive samples had interruptions on their RPPCR traces which suggest that further characterisation is clearly needed to identify if other expansion repeats, not already identified, exist within the normal and ataxic population. There did not appear to be a correlation between the size of repeat and the age of disease onset, however disease course progression is slow, and reports of initial symptoms may have been neglected by some patients. A follow-up study on a larger cohort of CANVAS and late-onset ataxic patients also found no correlation between size of repeat and age of onset (53).

The *RFC1* expansion is found at the 3'-end of a deep intronic AluSx3 element. Alu elements are highly conserved, primate specific and belong to a class of retroelements called short, interspersed elements (SINES). They continue to have evolutionary effects in modern humans through somatic insertion events, genetic diversifying, insertional mutagenesis which contributes to disease and non-allelic homologous recombination which causes copy number variation (173).They represent a 300-nucleotide long sequence that undergoes retro-positioning, which is the process whereby Alu elements are transcribed by RNA polymerase III and then reverse-transcribed by reverse transcriptase and inserted into a new genomic location (174). It is estimated that there is a new Alu insert per 20 human births (175) causing one in every 1000 new human genetic diseases (176). Alu elements are also active in somatic tissues, contributing to genetic instability throughout a lifetime of an individual and potentially contributing to age-related degeneration and disease (173).

Alu elements have a long A-rich region at the 3' end, which is critical for its amplification (177), whilst the entire element is flanked by direct repeats that are formed by sequence duplication at the site of insertion. The polyA tail at the end of the Alu element has been suggested to be an initiation site for microsatellite expansion (178).The amplification of Alu elements are dependent on an internal RNA polymerase III promotor that may initiate transcription and produce the RNA intermediates needed for amplification. Alu element transcription is terminated by a TTTT terminator sequence. These A-rich 3' ends are integral to the activity of Alu elements with the length and sequence heterogeneity determining the RNA polymerase III generated Alu RNA. Despite there being around one million Alu elements within the human genome, the majority are inactive due to rapid degradation on an evolutionary scale. This is caused by A-tail shortening and A-tail accumulation of heterogeneous base interruptions within active elements; eventually causing random mutation accumulation (173). It is a possibility that a G interruption of the polyA tail of ALUSx3 caused the mutated AAGGG repeat unit, with AAAGG and AAAAG being other non-pathogenic conformations arising from other G insertions (figure 3.1.10).

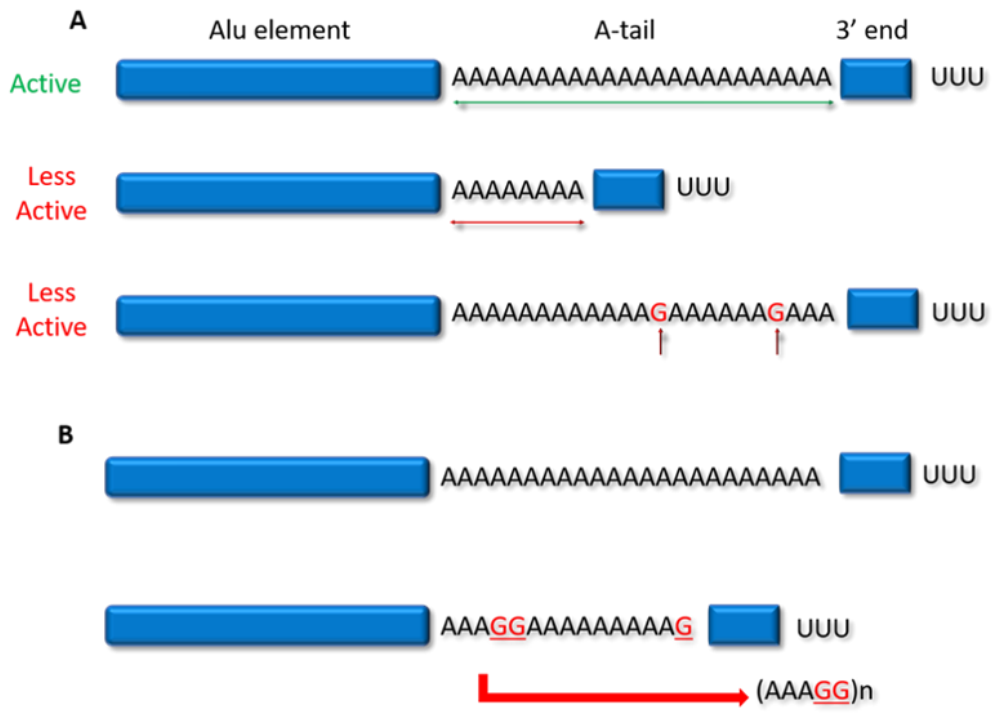


Figure 3.1.10: Structure of Alu element. A shows the active structure of Alu-elements, with intact polyA tail, with A-tail shortening and heterogenous base interruptions causing a reduction in activity. B, shows a possible event where G insertions during polyA tail shortening caused the pathogenic AAAGG conformation.

SNP haplotype screening revealed shared haplotype ancestry by most positive and familial cases and several healthy control carriers with two $AAAGG_{(n)}$ expansions. Therefore, I speculate that a potential ancestral founder event may have occurred whereby a nucleotide change in AAAAG produced the AAAGG or AAGGG conformation. The pathological expansion of the repeated unit would then have followed this event. However, Dr Cortese found that two patients shared only one allele of the common haplotype, highlighting that different genetic backgrounds can still harbour the mutated AAGGG repeated unit. Therefore, further studies are required to further define the exact frequency of the ancestral haplotype, which will be possible with increasing numbers of CANVAS and expansion diagnosed patients.

Remarkably Dr Cortese identified that there is no associated loss of function of the *RFC1* gene, which is different to the effects of the GAA expansion on frataxin that causes FRDA. FRDA shares similarities with CANVAS, in both symptoms and location of expansion, with both being progressive recessive disorders, that appear before the age of 25 years old and both expansions located in the 3' end of an Alu element. FRDA patients with a GAA expansions express lower levels of the frataxin protein (179), however no such effect was observed in CANVAS patients at the protein or transcriptional level, nor did RNA-seq highlight any effects on neighbouring and distant genomic regions. RFC1 is the 'large

subunit', containing both N- and C-terminal extensions, out of the 5 subunits of RFC. During DNA replication the DNA polymerase α-primase complex (Pol α) must be replaced by DNA polymerase δ (Pol δ) for DNA synthesis to occur, which is initiated by RFC and replication protein A (RPA), disrupting Pol α-RPA interactions and the removal of Pol α. RFC then loads PCNA onto the primed site, with RFC then displaced by PCNA, allowing Pol δ association (180). Additionally, *RFC1* has been linked to DNA repair enzyme interaction and associated roles in DNA damage recognition, DNA repair enzyme recruitment and positioning, including pathways such as nucleotide excision repair (NER)(181). Disrupted DNA repair pathways have been well documented in neurodegenerative diseases including Cockayne syndrome, AT, ataxia oculomotor apraxia 1 and 2 and xeroderma pigmentosum, which collectively clinically feature ataxia and neuropathy (182). This may suggest that the cerebellum and peripheral nerves are particularly susceptible to DNA damage. However, Dr Cortese did not observe any impaired response to DNA repair within patient fibroblasts, but pathological mechanisms involving AAGGG mediated effects on *RFC1* that disrupt other unidentified functions cannot be ruled out.

More than 40 neurological disorders are caused by microsatellite expansions including Huntington Disease (HD), DM and C9orf72 ALS and FTD, with differing mechanisms. Intronic mutations vary considerably in GC content (70), with GC-rich expansions associated with highly stable RNA secondary structure formation (183). They can also substantially increase and even multiply intron length, such as the $(GGCCTG)_n$ mutation in *NOP56* associated with SCA36, as well as form hairpins and G-quadruplexes. Expanded repeats in transcripts can also alter splicing machinery, as in the case of DM1 and DM2, which cause RNA foci in cells and RNA toxicity, implicating altered splicing that increases function of the CUG-binding protein and sequestration of muscleblind 1 (MBNL1) into the CUG repeat rich foci, reducing its function (184). Repeat-associated non-ATG (RAN) proteins can also accumulate in disease-affected tissues, causing various downstream effects including dipeptide proteins found in diseases such as C9orf72 ALS/FTD (185). Neither sense nor anti-sense unit RNA foci was observed by Dr Cortese in CANVAS brain tissue. However, remarkably intron retention of intron 2 in *RFC1* pre-mRNA was consistently found increased across all tissues of CANVAS patients. Intron retention was recently linked as a common event in GC-rich intronic expansions that cause disease, such as C9orf72 ALS/FTD and DM2 , but interestingly not associated with AT-rich expansion repeats such as FRDA and SCA10 where the expansion is located distally from the nearest splice site (186). Nuclear retention and nucleocytoplasmic transport of pre-mRNA could potentially be affected by intron retention and subsequent impaired pre-mRNA processing, which could be accessible to the translational machinery if inefficiently exported to the cytoplasm.

The neuropathology of CANVAS is characterised by widespread ganglionopathy (neuronopathy) that underpins the sensory deficits, dysautonomia and vestibular areflexia associated with the disorder (61, 64). This has been shown in oto- and neuropathological studies of CANVAS, with posterior column

atrophy and axonal neuropathy caused by the neuronal loss of sensory nerve ganglia of the dorsal root of the spinal cord (61, 64). Sensory neuropathy is a core feature of CANVAS and is pivotal to the process of neurodegeneration seen in the disease. It has been speculated that due to patients with the biallelic *RFC1* expansion not displaying insensitivity to pain, that that Aδ and C fibre sensory neurons are selectively spared (53). Evidence suggests that a sensory neuronopathy, with combined degeneration of peripheral and central projections of dorsal root ganglion neurons, plays a large part in the neurodegeneration seen (53).The majority of cerebellar pathology is seen through neuronal loss in the Purkinje cells, with atrophy of the anterior and dorsal vermis (lobules VI, VIIa, and VIIb) and hemispheric crus I, with the basal ganglia, diencephalon and cerebellar dentate nuclei remaining intact (61). The progression of the disease follows a pattern of pathological progression, from sensory neuron involvement early on, to the later contribution of cerebellar and vestibular dysfunction, which is mirrored in the progression and onset of symptoms. Gait ataxia and distal sensory disturbances appear earlier than the visual disturbances and oscillopsia, dysarthria and dysphagia (53). In addition, a higher proportion of patients have peripheral neuropathy (detected on NCS) in comparison to cerebellar dysfunction and vestibular failure, indicating that peripheral neuropathy is earlier onset as a clinical feature of CANVAS (53). Further validation is required to establish whether sensory neuropathy is the site of initial disease manifestation, despite it being a prominent feature of the disease (50).

One limitation of this study is that there was only one *RFC1*-positive CANVAS brain available for neuropathological examination, which is too small a sample size to make any meaningful conclusions (2). These findings need to be validated in order to further characterise the associated neuropathology with *RFC1*-associated CANVAS. Further to this, no neuropathology has been reported on any non-CANVAS cases that also harbour the pathogenic *RFC1* repeat expansion. This will be pivotal to understanding which phenotypes are directly caused by this expansion.

This work has identified an underlying genetic cause of CANVAS; a pathogenic $AAGGG_{(n)}$ expansion repeat, which is a significant finding and progression into the understanding of heritability in non-coding regions as well as a breakthrough in the genetic understanding of CANVAS. It has also contributed to the understanding of the clinical spectrum associated with *RFC1*-associated disorders, with the identification of an MSA mimic that harbours the biallelic repeat expansion, which are detailed in chapter 3. Recent findings confirm that *RFC1* is the principal gene underlying the genetic cause of CANVAS, however, the 68% of clinical CANVAS patients that harbour the expansion, and 90% of *RFC1*-positive patients with CANVAS with electrophysiological evidence, indicate that there are other genes and environmental causes of CANVAS not yet identified (50).

The optimisation of the Southern blotting protocol has resulted in a much improved, reliable technique. However future work is still needed to potentially find a way to confirm repeat expansions without needing to perform every component in the workflow. Recent advancements have successfully detected

the *RFC1* expansion using whole genome sequencing WGS and even WES, using a novel search algorithm that does not rely on the prior knowledge of existing *RFC1* repeats and is independent of prior PCR screening (50). If successfully validated, this will allow for future screening on existing patient datasets of unsolved cases, such as the 1kGP, furthering efficient and correct diagnosis of patients with *RFC1*-related disorders. This has led to further work, detailed in chapter 5, involving a novel bioinformatic tool that can be used on WGS datasets and has been tested on 100kGP data, on a large cohort of 62, 550 samples, successfully identifying patients with pathogenic recessive repeat expansions in *RFC1*.

Further work into the potential pathological mechanism of *RFC1* is ongoing by Dr Cortese and will be pivotal to understanding the neuropathology that leads to the symptoms seen. Due to the recessive inheritance of the pathogenic repeat expansion, a loss of function mechanism could be a possibility. However, all investigations do not show a reduced expression of RFC1 protein or canonical *RFC1* transcript (2). The interesting finding that intron 2 is retained in *RFC1*-preRNA across all tissues in affected patients provides a clear direction for future work to potentially elucidate the pathogenic mechanism (2). Chapter 4 details work carried out into the prevalence of the *RFC1* repeat expansion in other diseases, specifically MSA, and highlights additional diseases have been linked to the *RFC1* repeat expansion. However, no neuropathology has been reported on patients that have the repeat expansion disorder but do not have CANVAS, for example those patients that have a MSA-mimic, so this will be an important caveat of fully understanding which symptoms the neuropathological pathway involving *RFC1* directly contributes to. The ongoing screening, identification of *RFC1*-positive patients and subsequent broadening of the phenotypic spectrum associated with *RFC1* could make this work possible in the near future.

3.1.6. Conclusion

In conclusion this research has further developed understanding of the *RFC1* repeat expansion and its association with late-onset ataxia and CANVAS. This work has identified the allelic frequency of the pathogenic AAGGG repeat expansion as 0.7%, as well as further characterising the highly polymorphic region, detecting two additional expanded motifs prevalent within the general population, AAAAG and AAAGG. This work also carried out an *RFC1* repeat expansion screen on the largest cohort of late-onset ataxia patients (at the time of project and publishing), which found a pathogenic allelic frequency of 3.1%. The workflow involved in *RFC1* repeat expansion detection has been effectively optimised through this work, resulting in an efficient and accurate method of Southern blotting. The improved workflow has been utilised by collaborating laboratories and has led to the identification of a novel pathogenic conformation in the Māori and Cook Island Māori patient populations, $AAAGG_{10\text{-}25}$ $AAGGG_{(n)}$ $AAAGG_{(n)}$, that highlights the high level of heterogeneity within this region. It also suggests that there may be additional ethnically specific pathogenic motifs out there that have not yet been

identified. Genetic screening on larger ethnically diverse populations will be crucial for furthering our understanding of the region.

3.1.7. Chapter 4 associated publications

Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Cortese A, Simone R, **Sullivan R***, Vandrovcova J*, Tariq H, Yau WY, Humphrey J, Jaunmuktane Z, Sivakumar P, Polke J, Ilyas M, Tribollet E, Tomaselli PJ, Devigili G, Callegari I, Versino M, Salpietro V, Efthymiou S, Kaski D, Wood NW, Andrade NS, Buglo E, Rebelo A, Rossor AM, Bronstein A, Fratta P, Marques WJ, Züchner S, Reilly MM, Houlden H.Nat Genet. 2019 Apr;51(4):649-658. doi: 10.1038/s41588-019-0372-4. Epub 2019 Mar 29.

Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. Cortese A, Tozza S, Yau WY, Rossi S, Beecroft SJ, Jaunmuktane Z, Dyer Z, Ravenscroft G, Lamont PJ, Mossman S, Chancellor A, Maisonobe T, Pereon Y, Cauquil C, Colnaghi S, Mallucci G, Curro R, Tomaselli PJ, Thomas-Black G, **Sullivan R**, Efthymiou S, Rossor AM, Laurá M, Pipis M, Horga A, Polke J, Kaski D, Horvath R, Chinnery PF, Marques W, Tassorelli C, Devigili G, Leonardis L, Wood NW, Bronstein A, Giunti P, Züchner S, Stojkovic T, Laing N, Roxburgh RH, Houlden H, Reilly MM.Brain. 2020 Feb 1;143(2):480-490. doi: 10.1093/brain/awz418.

A Māori specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele. Beecroft SJ*, Cortese A*, **Sullivan R***, Yau WY*, Dyer Z, Wu TY, Mulroy E, Pelosi L, Rodrigues M, Taylor R, Mossman S, Leadbetter R, Cleland J, Anderson T, Ravenscroft G, Laing NG, Houlden H, Reilly MM, Roxburgh RH.Brain. 2020 Sep 1;143(9):2673-2680. doi: 10.1093/brain/awaa203.

A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families.
Scriba CK, Beecroft SJ, Clayton JS, Cortese A, **Sullivan R**, Yau WY, Dominik N, Rodrigues M, Walker E, Dyer Z, Wu TY, Davis MR, Chandler DC, Weisburd B, Houlden H, Reilly MM, Laing NG, Lamont PJ, Roxburgh RH, Ravenscroft G.Brain. 2020 Oct 1;143(10):2904-2910. doi: 10.1093/brain/awaa263.

3.2. Discovery of *RFC1* repeat expansions in an MSA mimic

3.2.1. Introduction

3.2.1.1. RFC1 repeat expansion in mimics of CANVAS

As identified within the previous chapter, a biallelic expansion repeat in *RFC1* is a genetic cause of late-onset ataxia and CANVAS. The clinical spectrum of CANVAS is broad, and onset of clinical features is varying among patients Subsequently CANVAS is often misdiagnosed within the clinic, often as multiple systems atrophy (MSA). CANVAS shares a clinical overlap with MSA, a disorder that presents with cerebellar impairment and autonomic dysfunction, but also parkinsonism, and cases are often differentially diagnosed as MSA.

3.2.1.2. Multiple Systems Atrophy

MSA is a rare progressive neurodegenerative disorder that is characterised by cerebellar ataxia, parkinsonism, autonomic failure, pyramidal features and autonomic failure (187). It is a fatal disease that leads to death in the years following symptom onset (188). The disease is divided into two subtypes based on their predominant clinical phenotype: parkinsonian (MSA-P) and cerebellar (MSA-C) (189). It is characterised by region-specific patterns of neuronal loss involving several brain regions, including the cerebellum, substantia nigra, inferior olive, putamen, striatum, globus pallidus and pons and the intermediolateral column of the spinal cord, with neurons in the basal ganglia nuclei and wide-spread oligodendrocyte numbers generally spared (190).

Pathophysiology of MSA

MSA is defined as a synucleinopathy, with gliosis and oligodendrogial cytoplasmic inclusions (GCIs) representing a major pathological disease hallmark (191, 192). Oligodendrocytes span a heterogenous population of glia and all derive from oligodendrocyte progenitor cells (OPCs) expressing chondroitin sulphate proteoglycan, ganglioside GD3, platelet derived growth factor-α (PDGFα) and membrane marker neuron glia 2 (NG2) (193, 194). Once fully matured, oligodendroglia express several myelin-associated markers, however differentiation into fully functioning cells only occurs once OPCs have migrated to their destination (194, 195). Oligodendroglia types I, II and III myelinate axons, type I and II specifically smaller diameter axons and type III larger diameter axons. Type IV myelinates individual large diameter axons by directly opposing its cell body to the axon (196, 197). The primary function of mature oligodendroglia is to produce, form and maintain myelin sheaths which facilitate action potentials, as well as the regeneration and maintenance of axons (198, 199).

Alpha-synuclein (α-synuclein) fibril aggregates comprise the glial cytoplasmic GCIs in MSA, which appear in triangular, sickle or flame shapes (200, 201). α-synuclein has predominant expression in the central nervous system and is localised principally at presynaptic terminals (202). GCIs are implicated in other synucleinopathies including PD and dementia with Lewy bodies, where misfolded,

hyperphosphorylated and fibrillar α-synuclein form pathological lesions leading to neurodegeneration (203-205). GCIs have also been found to contain tau and putative α-synuclein interacting 14-3-3 proteins and also microtubule-associated protein MAP5 in MSA brains (206, 207). Additionally GCIs in MSA brains have been found to have strong immunoreactivity to heat shock proteins (HSP) Hsc70 and Hsp70, which are stress-induced molecular chaperones, which accumulate in inclusions (208). Parkin and LRRK2, which are pathogenically implicated in PD via protein deposition in Lewy body inclusions, immunoreactivity has also been linked to MSA brain GCIs. Smaller observations of Parkin immunoreactivity in contrast to LRKK2 which widely colocalised to α-synuclein-immunopositive GCI and degenerating myelin sheaths and was observed in the majority of enlarged oligodendroglia in MSA (209). The oligodendroglial protein p25α, which is a constituent of myelin and has a high affinity for myelin basic protein (MBP), induces aggregation and colocalization of α-synuclein to GCIs (210). In MSA it has been found that p25α is relocalised from myelin to expanding oligodendroglial cell bodies alongside a reduction of full-length MBP, which occurs in parallel to the demyelination of small calibre axons in pontecerebellar and corticospinal pathways affected in MSA (211).

The pathophysiological mechanism of MSA is not well understood, however emerging clinical and post-mortem evidence points to a contributary combination of glial and neuronal dysfunction that leads to the development of the disease. Adult oligodendrocytes do not express α-synuclein normally, also true in the case of MSA, suggesting that it is the abnormal accumulation of α-synuclein that develops the disease (212). An early pathogenic step is the relocalisation of p25α to swelling oligodendroglia (211). This is followed by an increase in concentrations of α-synuclein in oligodendroglia, alongside an inability to carry out α-synuclein degradation (213, 214). Insoluble oligomers of α-synuclein form as a result of α-synuclein phosphorylation and aggregation, induced by its interactions with p25α. The subsequent formation of GCIs induces activation of microglia and the reactive release of cytokines and oxygen reactive species, inducing inflammation (215, 216). The fibril-shaped aggregates are taken up by neighbouring neurons, and compared to other synucleinopathies such as PD, have an increased propensity to induce neuronal cytoplasmic inclusions and cytoplasmic pathological changes (217). Neuronal death and reactive astrogliosis are induced via neuroinflammation, along with mitochondrial dysfunction caused by the α-synuclein inclusion and the loss of neurotrophic support usually supplied by glial cells (218, 219). It has been proposed that α-synuclein toxicity may propagate and spread to connected brain regions, causing the widespread degeneration seen in MSA (220).

Neuropathology of MSA

As aforementioned, a definitive diagnosis of MSA can only be made after post-mortem tissue examination, whereby the presence of GCIs is confirmed (189). These GCIs cause selective atrophy and neuronal loss in olivopontocerebellar and striatonigral systems, which correlate strongly to the two

main clinical phenotypes; striatonigral degeneration (SND) and MSA-P and olivopontocerebellar atrophy (OPCA) with MSA-C (218, 221).

SND is characterised macroscopically by iron and neuromelanin accumulation in the dorsolateral putamen and nucleus caudatus, and pallor of substantia nigra and locus coeruleus, alongside atrophy due to lipofuscin. OPCA presents with severe atrophy of the middle cerebellar peduncle, cerebellar folia and pontine basis, alongside cerebellar white matter discolouration and blurred inferior olivary nucleus (222). Microscopically; atrophy, GCIs, reactive astrogliosis and loss of neurons is observed, alongside less frequently seen neuronal cytoplasmic inclusions that are immunoreactive for α-synuclein (223). GCIs are most abundant at primary motor and premotor cortices, with the number of GCIs increasing with disease progression (221).

MSA-C presents with degeneration of the vermis, dentate nucleus, inferior olivary nucleus, pontine basis and Purkinje cells, whereas MSA-P is primarily affected in the dorsolateral putamen and caudate nucleus (223). There is an additional subtype of 'minimal MSA' which presents with widespread GCIs and localised neuronal loss in the locus coeruleus and substantia nigra. This type of MSA also shows early orthostatic hypotension (OH), major respiratory dysfunction and earlier disease onset (224).

Autonomic dysfunction is a major clinical feature of both MSA-C and MSA-P and is caused by central autonomic network disruption (225). The sacral portion of the intermediolateral column of the spinal column, the Oluf's nucleus and the dorsal vagal nucleus are affected in MSA and cause in part urinary and erectile dysfunction (226). Urogenital symptoms are also caused by dysfunction of the substantia nigra, Purkinje cells, hypothalamus, pontine micturition center and the locus coeruleus (226). Degeneration of the brainstem autonomic nuclei, sympathetic ganglia, thoracic part of the intermediolateral column of the spinal cord and postganglionic fibres cause the OH seen in MSA (226). Cognitive impairment in MSA is underpinned by Lewy-body like inclusions, cortical and limbic GCIs and frontal striatal deafferentation (227, 228). The additional clinical feature of REM sleep behaviour disorder can develop due to pedunculopontine nucleus and locus coeruleus projection neurodegeneration, alongside central hypoventilation which is caused by an affection of the medullary respiratory centres (229).

Aetiology of MSA

MSA is considered a sporadic disease although there have been some reports in Europe and Asia of MSA pedigrees with both autosomal recessive and dominant inheritance (230-232). In the most recent genome-wide-association study (GWAS) no significant loci were identified, but SNPs in the genes *MAPT*, *FBXO47*, *ELOVL7* and *EDN1* were found, suggesting potential implication in MSA pathogenesis. However, the study was underpowered, and therefore definitive conclusions cannot be made (233).

Mutations in *COQ2*, which codes for coenzyme Q10, have been reported in two Japanese multiplex families (homozygous mutation M78V-V343A/M78V-V343A and heterozygous mutation R337X/V343A) and sporadic MSA cases (common variant V343A) (234). Variant V343A has been found exclusively within Asian populations and has not been replicated in European cohorts, suggesting it is region specific (233, 235). Coenzyme Q10 is a vital component of the electron transport chain, specifically electron transfer between the complexes I, II and III (219). It is reduced in MSA cerebellum which is associated with Coenzyme Q10 biosynthesis dysfunction and an increased vulnerability to oxidative stress (236).

SNCA, which codes for α-synuclein has been associated with MSA. A large European MSA series found significant associations with increased development of risk and two SNPs of the SNCA locus, which was subsequently replicated, but not found in the most recent GWAS (233, 237, 238). Additionally, *SNCA* triplication has been implicated in familial PD that presents with MSA traits and a G51D SNCA mutation in a British family with dominant juvenile Parkinsonism that presented with neuropathological traits of both MSA and PD (239, 240).

Variants in gaucher-disease-associated glucocerebrosidase (GBA) have been associated with MSA-C in both Japanese, European, and North American cohorts (241). *TMEM230* sequence changes have been reported in Southwest Chinese MSA patients (242). A discordant loss of copy numbers in *SHC2* have been reported in sporadic Japanese MSA patients and monozygotic twins but was not replicated in US populations (243, 244). Mutations in *LRRK2* have also been implicated in sporadic and familial MSA (245, 246). The G2019S-LRRK2 mutation has been reported to account for 1% and 4% of sporadic and familial PD, respectively, with these patients presenting with milder motor involvement, normal cognition, good levodopa responsiveness and less prominent dysautonomia (247, 248). Interestingly G2019S-*LRRK2* implicated MSA patients presented with unusual features, with early onset, early genitourinary and REM behaviour involvement, but with a prolonged disease course (246).

MSA has no currently known environmental risk factors. There have been two case study reports of an association between MSA and agricultural and occupational exposure to organic solvents, pesticides, metal dusts and organic solvents (249, 250).

Clinical overlap between MSA and CANVAS

MSA-mimicries are frequently being reported for other genetic conditions. Spinocerebellar ataxia 17 (SCA17) is a frequent example, with 7.3% of a clinical MSA cohort comprising SCA gene mutations (251). A recent brain bank series on patient's diagnosed using the second consensus criteria with either possible or probable MSA found that 38% of samples did not meet the pathological criteria; of these 37% had a true diagnosis of dementia with Lewy bodies (DLB), 29% had supranuclear palsy (PSP) and 15% had PD (228). The cogency of the second consensus for correct diagnosis was evaluated in a

clinicopathological study, with 41% and 18% correct diagnosis at first visit for possible and probable MSA respectively, which rose to 92% and 63% at last visit (189). It is therefore clear that the development of effective diagnostic tools in the early stages of MSA are crucial for correct diagnosis, disease prognosis estimation and patient evaluation of suitability for therapeutic clinical trials.

MSA and CANVAS share an overlap of symptoms at early stage, with autonomic dysfunction and cerebellar ataxia a main feature of both CANVAS and MSA-C (figure 3.2.1) (189). Crucially, the onset of CANVAS symptoms varies in timing, with a diagnosis only made after other causes such as FRDA or SCA3 are ruled out (189).

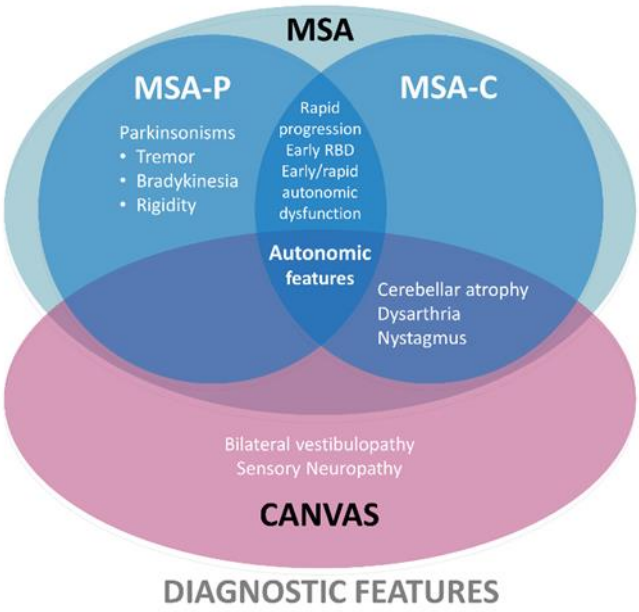


Figure 3.2.1: Phenotypic overlap between MSA and CANVAS. MSA, multiple systems atrophy; MSA-P, predominance of parkinsonian features; MSA-C, predominance of cerebellar features, CANVAS, cerebellar ataxia, neuropathy, vestibular areflexia syndrome. Figure created by RS and published in R Sullivan et al., JNNP, 2020 (252).

3.2.2. Aims of study

During the course of this project, I hypothesised that the repeat expansion in *RFC1* could also be implicated in other cases diagnosed with disorders that share an overlap in phenotype with CANVAS, particularly the presence of ataxia as a feature. CANVAS is difficult to diagnose in the clinic and quite often a differential diagnosis of MSA-C is made, especially during the early stages of disease. In order to test the prevalence of the recessive repeat expansion in *RFC1* within the MSA disease population, I screened two independent MSA cohorts. The first was a cohort of pathologically confirmed MSA patients and the second was a cohort of clinically diagnosed MSA patients. Due to the pathological hallmarks of MSA, pathologically confirmed cases of MSA were hypothesised to be negative for the

recessive *RFC1* repeat expansions. Both cohorts underwent the *RFC1* genetic testing workflow: flanking PCR, RPPCR and Southern blotting on suspected cases, and were subsequently clinically evaluated by Dr Wai Yan Yau and Prof. Henry Houlden.

3.2.3. Summary of methods

Study aim	Methods used	Contributors	Methods section
Identification of repeat expansions in pathologically confirmed patients	Flanking PCR RPPCR Southern blotting	Roisin Sullivan Roisin Sullivan Roisin Sullivan	2.2.2 2.2.3 2.2.4
Identification of repeat expansions in clinically diagnosed patients	Flanking PCR RPPCR Southern blotting	Roisin Sullivan Roisin Sullivan Roisin Sullivan	2.2.2 2.2.3 2.2.4
Clinical examination of patients	Neurological examination	Dr Wai Yan Yau, Professor Henry Houlden	2.1.6

3.2.4. Results

3.2.4.1. Recessive RFC1 repeat expansions are absent in pathologically confirmed MSA

gDNA from 336 patient cerebellar brain samples with a neuropathological diagnosis of MSA were screened for the pathogenic AAGGG expansion, alongside three other possible allelic confirmations; wildtype $AAAAG_{(11)}$, expanded $AAAAG_{(11)}$ and expanded $AAAGG_{(n)}$, using a combination of flanking PCR and RPPCR. These pathologically confirmed MSA cases were enrolled from various brain banks after request and approval that consisted of Queen Square Brain Bank, Cambridge Brain Bank, Newcastle Brain Tissue Resource, Oxford Brain Bank, Edinburgh Brain Bank, Southwest Dementia Brain Bank, London Degenerative Disease Brain Bank, Sheffield Brain Tissue Bank, Manchester Brain Bank, Maryland Brain and Tissue Bank, and the Harvard Brain Tissue Resource Center. The definitive neuropathological diagnosis of MSA is established based on MSA working criteria (189).

Of the 336 pathologically diagnosed MSA samples, none were positive for the biallelic pathogenic $AAGGG_{(n)}$ *RFC1* expansion, as defined by a lack of amplifiable product on PCR and a positive RPPCR result. The motif allelic frequencies, out of 672 chromosomes, were 0.7% (n = 5) for the pathogenic $AAGGG_{(n)}$ expansion, 3% (n = 22) for $AAAGG_{(n)}$, 7% (n = 47) for $AAAAG_{(n)}$ and 88% (n = 589) for $AAAAG_{(11)}$ (table 3.2.1). Of the five heterogenous carriers of the pathogenic expanded conformation $AAGGG_{(n)}$, four carried the wildtype $AAAAG_{(11)}$ conformation on their second allele and one carried the $AAAGG_{(n)}$ expanded conformation. In addition, there were homozygous combinations not involving the pathogenic $AAGGG_{(n)}$ expansion, with 10 patients biallelic for $AAAAG_{(n)}$ and one patient biallelic with $AAAGG_{(n)}$, as assumed by no amplifiable PCR product and a positive RPPCR. A total of 5 patients

had an unidentified conformation on one allele, and 2 cases had two alleles unidentified (1.3%, n = 9), suggesting that there are other possible expanded conformations present not targeted by the primer sets used.

The mean expansion size of the expanded $AAAAG_{(n)}$ was 85 (±152) repeat units, the $AAAGG_{(n)}$ conformation had a mean expansion size of 124 (±93) repeating units and the AAGGG(n) conformation had a mean expansion size of 1185 (±545), calculated by use of sequence clamps on RPPCR that target the sequence directly adjacent to the expansion and Southern blotting (figure 3.2.2). Taken together, these results show that there were no biallelic *RFC1* expansions within the pathologically diagnosed cohort of MSA samples.

Pathologically confirmed cohort (n = 336)				
Allelic Combination	Allelic frequency		Expansion sizing	
	No. of samples (n = 336)	Allelic (carrier) Frequency (n = 672)	Expansion repeat conformation	Expansion repeat size (repeating units)
$AAAAG_{(11)}/AAAAG_{(11)}$	272	81%	$AAAAG_{(n)}$	175 (±88)
$AAAAG_{(11)}/AAAAG_{(n)}$	30	8.9%		
$AAAAG_{(11)}/AAAGG_{(n)}$	7	2.1%		
$AAAAG_{(11)}/AAGGG_{(n)}$	3	0.9%		
$AAAGG_{(n)}/AAGGG_{(n)}$	2	0.6%		
			$AAAGG_{(n)}$	355 (±172)
$AAGGG_{(n)}/AAGGG_{(n)}$	***0***	***0%***		
$AAAAG_{(n)}/AAAAG_{(n)}$	5	1.5%		
$AAAGG_{(n)}/AAAGG_{(n)}$	3	0.9%		
$AAAAG_{(n)}/AAAGG_{(n)}$	7	2%		
			$AAGGG_{(n)}$	1185 (±545)
$AAAAG_{(11)}$ /undetermined	5	1.5%		
Biallelically undetermined	2	0.6%		

Table 3.2.1: Allelic frequencies of allelic conformations in pathologically confirmed MSA cohort (n = 336). Pathogenic **$AAGGG_{(n)}$** conformation underlined in bold.

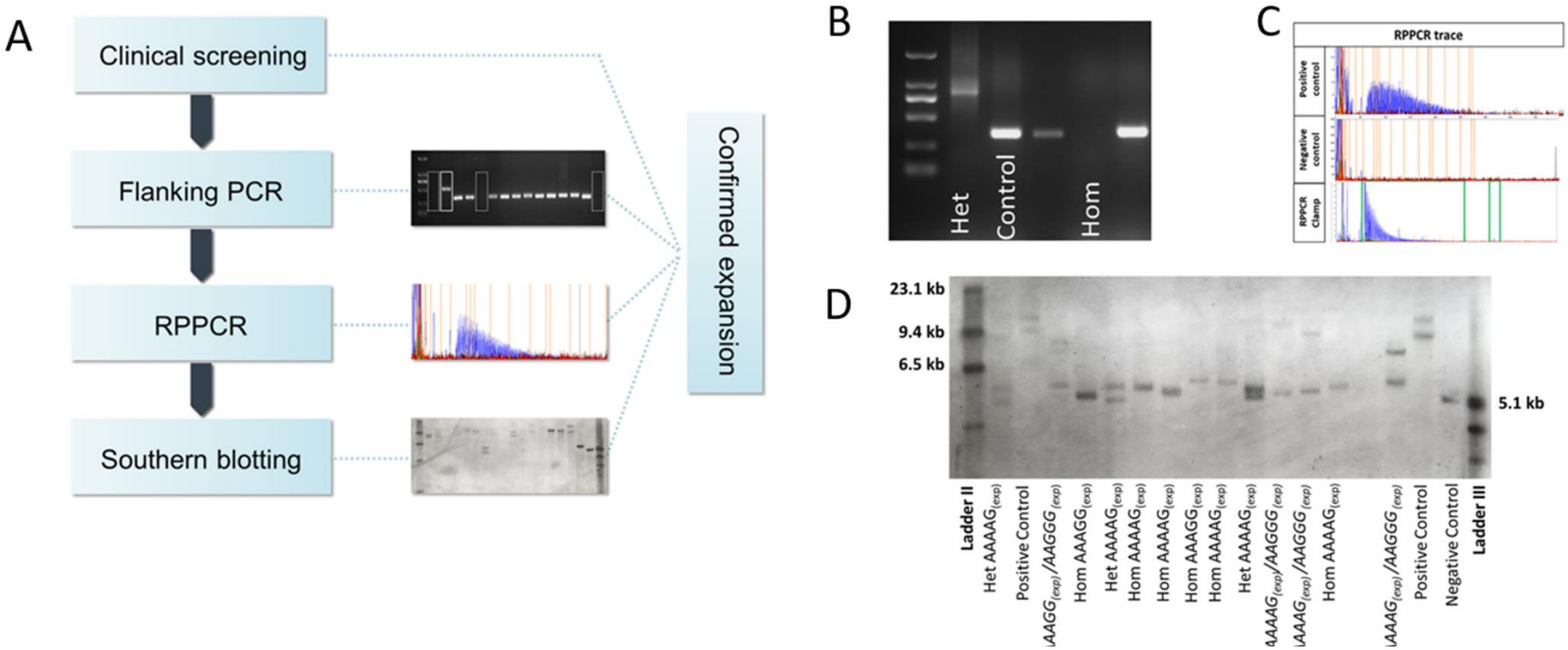


Figure 3.2.2: Screening of pathologically confirmed cohort. A**,** clinical screening occurs for identifying patients diagnosed with 'possible' or 'probable' MSA, whereas the pathological MSA cohort were identified by a prior neuropathological confirmation of MSA using Brain Bank searches. All samples underwent flanking PCR, with samples with no amplifiable product on PCR (suggesting a homozygous expansion) and those with a higher molecular weight (suggesting carrier status) taken forward for RPPCR (white dashed boxes). RPPCR was performed using primer sets targeting the three conformations, $AAGGG_{(n)}$, $AAAGG_{(n)}$ and $AAAAG_{(n)}$. A selection of samples of each RPPCR confirmed conformation (including all AAGGG cases) were then Southern blotted. B, PCR image with indicated carrier of *RFC1* expansion (Het), homozygous sample for AAAAG reference conformation (Control) and homozygous sample for an expanded conformation (Hom). **C**, RPPCR traces showing a positive control, negative control, and a positive RPPCR trace with a clamp at the end (green box). **D** A Southern blot of a selection of samples with labelled allelic conformations (*AAGGG(n)* expansions in italics; ladders in bold). Figure adapted from Sullivan et al., (252).

3.2.4.2. Recessive RFC1 repeat expansions identified in three clinically diagnosed MSA patients

In order to test for the prevalence of the recessive *RFC1* repeat expansion in clinically diagnosed MSA, a cohort of 207 clinically diagnosed 'possible' and 'probable' MSA cohort was genetically screened for the pathogenic AAGGG expansion, alongside three other possible allelic confirmations; wildtype $AAAAG_{(11)}$, expanded $AAAAG_{(11)}$ and expanded $AAAGG_{(n)}$, using a combination of flanking PCR and repeat-primed PCR. probable' MSA were enrolled from the National Hospital for Neurology and Neurosurgery. All subjects were assessed by neurologists with a special interest in MSA based on the second consensus criteria for MSA (189).

Three patients had biallelic *RFC1* expansions, as defined by no amplifiable product on PCR, and a positive repeat primed PCR for $AAGGG_{(n)}$. The allelic frequencies in clinical MSA cases were 1.7% (n=7) for pathogenic $AAGGG_{(n)}$, 2.4% (n=10) for $AAAGG_{(n)}$, 11.4% (n=47) for $AAAAG_{(n)}$ and 82.8% (n=343) for $AAAAG_{(11)}$ (table 3.2.2) Southern blotting revealed two large expansions, sized at 9434 kb (887 repeat units) and 12 477 kb (1495 repeat units) in patient 1 and 11 244 kb (1249 repeat units) in patient 2, corresponding to the two expanded alleles; both patients were initially diagnosed with 'possible' MSA (figure 3.2.3). There was insufficient DNA to perform a Southern blot for patient 3, initially diagnosed with 'probable' MSA.

Allelic Combination	**Clinical cohort (n=207)**	
	No. of patients (n = 207)	**Allelic Frequency (n = 414)**
$AAAAG_{(11)}/AAAAG_{(11)}$	144	69.57%
$AAAAG_{(11)}/AAAAG_{(n)}$	46	22.22%
$AAAAG_{(11)}/AAAGG_{(n)}$	4	1.93%
$AAAAG_{(11)}/AAGGG_{(n)}$	1	0.48%
$AAAGG_{(n)}/AAGGG_{(n)}$	1	0.48%
$AAGGG_{(n)}/AAGGG_{(n)}$	***3***	***1.45%***
$AAAAG_{(n}$ /$AAAAG_{(n)}$	0	0.00%
$AAAGG_{(n)}/AAAGG_{(n)}$	2	0.97%
$AAAAG_{(11)}$ /undetermined	4	1.93%
$AAGGG_{(n)}$/undetermined	1	0.48%
Biallelically undetermined	1	0.48%

Table 3.2.2: Frequency of allelic combinations in clinically diagnosed MSA cohort. Pathogenic **AAGGG** conformation allelic frequency highlighted in bold italics

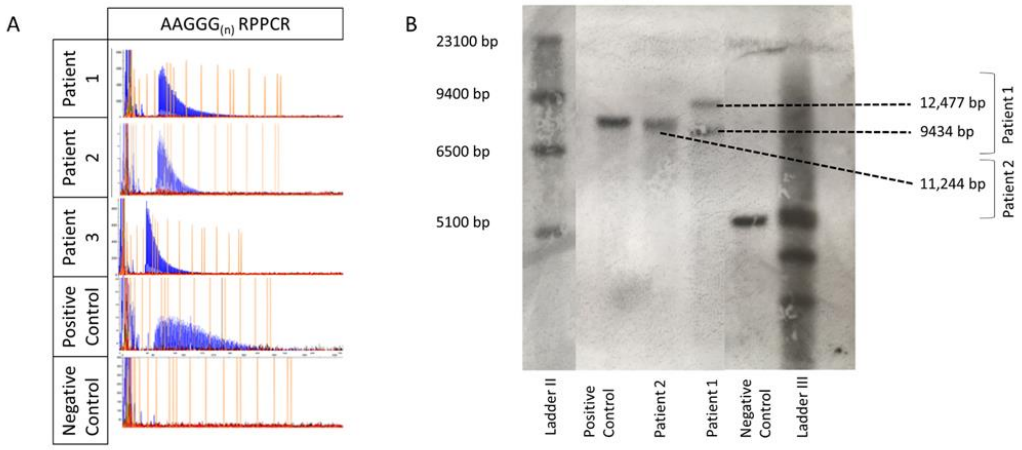


Figure 3.2.3: Genetic testing results for clinical MSA cohort (n = 207). A, RPPCR traces for patients 1, 2 and 3, identified as recessive for *RFC1* expansion, with hallmark 'sawtooth' pattern of decremental peaks corresponding to fluorescence (y-axis) and fragment size (x-axis, bp units), alongside a positive and negative control. B, Southern blot showing expansions for patient 1 and patient 2. Expansion sizes were estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used. For patient 1, an expansion is observed at 11, 200 bp, corresponding to two alleles) and for patient 2 expansions seen at 12, 500bp for one allele and 9400 bp for the other allele, alongside a positive control with band seen at 9500 bp (corresponding to both alleles) and a negative control with a band seen at 5000 bp. Figure adapted from Sullivan et al., (252).

3.2.4.3. Two patients identified with recessive RFC1 repeat expansion presented with clinical features representative of 'possible MSA', and one patient presented with clinical features of 'probable MSA'.

Two patients (patient 1 and 2) presented with clinical features typical of MSA-C including cerebellar ataxia and autonomic dysfunction. Patient 1 presented with atypical features of MSA, such as slow disease progression and sensory neuropathy. Patient 3 presented with symptoms suggestive of MSA-P, with parkinsonism and postural tremor, as well as cerebellar and autonomic dysfunction. Case studies 1, 2, and 3, figure 3.2.4 and table 2.2.3 depict the clinical features of each patient. Clinical examinations were carried out by Dr Wai Yan Yau.

3.2.4.4. Case Studies

Clinical research and examinations on patients 1, 2 and 3 were carried out by Dr Wai Yan Yau, a neurologist at the NHNN.

1. *Patient 1*

Patient 1 was clinically diagnosed with possible MSA: developed gait imbalance at 62/3 years of age with progression to wheelchair use at 67 years, parkinsonism, dysarthria, dysgraphia, multidirectional

diplopia, neurogenic bladder and presyncope. His past medical history was non-contributory. He did not report any preceding chronic cough or pins and needles. He did not smoke and there was no history of alcohol consumption. His brother also had ataxia, attributed to alcohol abuse and died at 62 years old from oesophageal cancer. On examination in our clinic the patient was 69 years old, and he had a poor shuffling and stiff broad-based gait with no arm swing. He was not able to perform tandem gait. His extraocular eye movements were full, with down-beating nystagmus at primary gaze, gaze evoked nystagmus, broken smooth pursuit and slow saccades. He had dysarthria, bradykinesia, cogwheel rigidity and appendicular dysmetria. His strength was preserved with absent Babinski's response. His reflexes were brisk and symmetrical with absent ankle jerks. Sensory examination demonstrated absent sensation to pinprick throughout, reduced vibration to right anterior superior iliac spine and left costal margin, reduced proprioception to right ankle and normal in left lower limb. Nerve conduction study showed absent sensory responses with normal motor unit action potentials and mild-length dependent denervation (figure 3.2.4D). MRI showed marked atrophy of cerebellum, involving middle cerebellar peduncle. There was also medulla atrophy and minor supratentorial atrophy without regional predominance. DaTscan showed asymmetrical dopamine transporter uptake in the striatum (figure 3.2.4A) Autonomic tests indicate mild cardiovascular autonomic dysfunction with orthostatic hypotension in tilt table test and liquid meal test, preserved pressor responses and blocked Valsalva manoeuvre. Neuro-otological review did not show vestibular areflexia. Routine investigations for ataxia and genetic tests (SCA 1, 2, 3, 6, 7, 12, 17, FRDA and whole exome sequencing) were negative. His parkinsonism did not respond to a trial of levodopa treatment.

2. Patient 2

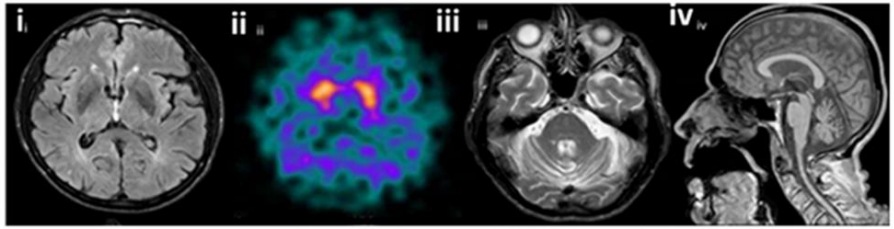
Patient 2 was clinically diagnosed with possible MSA: symptom onset at 62 years old, orthostatic decrease of blood pressures was <30 mm/15 mm Hg with symptoms of unexplained bladder urgency, levodopa non-responsive parkinsonism and a cerebellar syndrome. Started to experience difficulty in dancing and postural dizziness, and gait deterioration at age 62 years with rapid decline developing dysphagia, dysarthria, and mild neurogenic bladder over two years. Her balance was not worse in the dark and did not report any dysesthesia. Her past medical history included hysterectomy and hypertension; her only medication was amlodipine. She did not smoke and consumed two standard drinks of alcohol per week. Her family history was not contributory. On examination at age 66, she had a stooped posture and narrow-based gait with loss of bilateral arm swing. VORs were impaired on bedside testing. In additional to cerebellar ataxia, finger tapping was slow and showed decrements with interruptions. Her overall strength was preserved with absent plantar reflex. Her other reflexes were symmetrical and brisk. Sensory examination revealed loss of vibration to ankles and absent pinprick sensation in lower limbs and distal upper limbs. Routine investigations for ataxia were normal. Genetic tests (SCA 1, 2, 3, 6, 7, 12, 17, FRDA and whole exome sequencing) were negative. Her symptoms did

not respond to levodopa challenge. Patient declined to undergo DaTscan and further autonomic testing. figure 3.2.4B and 3.2.4D show the results of her MRI head and NSC.

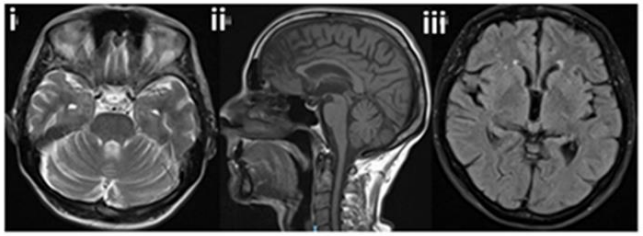
3. Patient 3

Patient 3 was clinically diagnosed with probable MSA: symptom onset at age 53 years old, orthostatic decrease of blood pressures was >30mm/15mmHg, levodopa non-responsive parkinsonism and a cerebellar syndrome. Her initial clinical manifestations were gait imbalance, dysarthria and increased extra pyramidal tone in her right hand. Her family history was not contributary. On examination at age 56, she had a broad-based gait, increased tone, and reduced tempo in her right hand, with a slight reduction in her left. She also had finger dysmetria bilaterally with impaired tandem gait. She showed no clear increased tone in her lower extremities. She had increased reflexes in her right arm, reduced reflexes in her lower extremities and upgoing plantar responses bilaterally. She had broken pursuit and hypometric saccadic eye movements, as well as postural tremor bilaterally. She had bradykinesia, head drop and flexion deformities. Autonomic testing showed parasympathetic impairment, with postural blood pressure drop, unexplained urinary urgency and incomplete bladder emptying. NCS results were normal (figure 3.2.4D). MRI head showed extensive atrophy of the superior and middle cerebellar peduncles, cerebellum, and pons (figure 3.2.4C). Genetic tests (SCA 1, 2, 3, 6, 7, 12, 17, FRDA and whole exome sequencing) were negative.

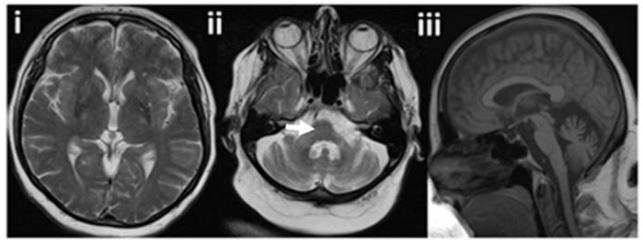
A



B



C



D

Patients	Age		Motor					Sensory		EMG	
			CMAP (mV)	CV (m/s)	DML (ms)	F wave latency (ms)		SAP (µV)	Latency (ms)		
1	63	Peroneus R	6.8	38	6.8	63	Suralis	Absent	n/a	Vastus Lateralis L	N
		Tibialis R	4.2	n/a	4.2	54	Median	Absent	n/a	TP L	Slight excess of polyphasic MUP
							Radial	Absent	n/a		
2	67	Median R	12	61	3.5	64	Suralis	Absent	n/a	Not performed	
		Peroneus R	8.2	43	3.7	46	Median	Absent	n/a		
		Tibialis R	12.8	41	3.8	43	Radial	4	5		
3		Median R	9.3	56	2.7	65	Median				
		Tibial R	12.2	47	3.7	46					

Figure 3.2.4: MRI images of patients 1, 2 and 3, diagnosed with MSA and positive for recessive RFC1 expansion. A**,** Neuroimaging of patient 1 at age 66 years i. MRI head T2 FLAIR axial sequence showed minimal changes in basal ganglia, ii. DaTscan showed asymmetrical reduction in putaminal dopamine transporter uptake (black arrow), iii. MRI head T2 axial sequence showed pontine and middle cerebellar peduncle atrophy without “hot cross bun” sign and iv. MRI head T1 sagittal sequence shows cerebellar and medulla atrophy. B**,** Neuroimaging of patient 2 at age 65 Years, i. MRI head T2 FLAIR axial sequence showed normal pontine appearance, ii. MRI head T1 sagittal sequence showed no evidence of cerebellar or brainstem atrophy, iii. MRI head T2 FLAIR axial sequence showed normal appearance of basal ganglia. C**,** Neuroimaging of patient 3 at age 56 years, i. MRI head T2 axial sequence showed symmetrical diminution of lentiform nuclei with subtle hyperintensity along dorsolateral aspect of the putamen ii. MRI T2 axial sequence showing "hot cross bun” sign (white arrow), iii. MRI T1 axial sequence showed pontine and cerebellar atrophy. D, Nerve conduction study (NCS) showed sensory neuronopathy with normal motor studies and mild distal denervation on EMG whilst for patient 1 and patient 2 showed sensory neuronopathy with normal motor studies. Patient 3 was normal. Adapted from Sullivan et al., (252).

Clinical Feature	Patient		
	Patient 1	**Patient 2**	**Patient 3**
Age at examination/sex (years)	69/m	66/f	56/f
AAO (years)	62	62	53
Sign at onset	Impaired gait	Impaired gait, dizziness	Impaired gait
Cerebellar dysfunction	+	+	+
Parkinsonism	+	+	+
Postural tremor	-	-	+
Abnormal head thrust test	-	+	na
Autonomic dysfunction	+	Urinary symptom only	+
Reflexes in LLs	Absent Ajs	Retained	Retained
Vibration at ankles	Impaired	Impaired	NA
Romberg's sign	+	+	NA
Plantar response	↓	↓	↑
Lower motor neuron sign	-	-	-
Cognitive impairment	-	-	-
Response to levodopa	Poor	Poor	Poor
Postural instability within 3 years of motor onset	+	+	+
Family history	Brother with ataxia (attributed to alcohol cerebellar degeneration)	-	-
NCS/ EMG	Absent UL/LL SAPs	Absent UL/LL SAPs	normal
MRI head	Pontine/ cerebellar atrophy	WNL	Pontine/ cerebellar atrophy and putaminal hyperintensity
DaTscan	Bilateral reduced pallidal reuptake	na	na
Autonomic testing	Sympathetic impairment	na	Parasympathetic impairment
Vestibular testing	Normal	na	NA
Repeat length	950/1500	1300/1300	NA

Table 3.2.3: Clinical features of patients 1, 2 and 3. Expansion sizes were estimated using a linear regression model that corresponded to the DNA molecular weight marker ladders used. *M, Male; F, female; NA, Not applicable, Cerebellar dysfunction: cerebellar gait, dysarthria, limb ataxia and eye signs; Parkinsonism: bradykinesia with rest tremor and or rigidity and/or postural instability; Autonomic dysfunction: postural blood pressure drop, unexplained urinary urgency or incomplete bladder emptying; Upper motor neuron sign: spasticity, hyperreflexia and positive babinski's sign;*

Autonomic testing: parasympathetic (HR during Valsalva manoeuvre, deep breath and standing), sympathetic (BP during change in posture and handgrip; SSR), combined parasympathetic and sympathetic; WNL within normal limits; AJs ankle jerks; AAO age at onset; LL lower limbs; SAPs sensory action potentials; UL upper limbs

3.2.5. Discussion

A recessive pentanucleotide repeat expansion within intron 2 of *RFCI* was identified by Cortese et al, and detailed in chapter 4, as the cause of CANVAS and late-onset ataxia (2).

Detailed analysis of the clinical information of patients positive for the biallelic expansion is crucial to understanding the true clinical spectrum of CANVAS, and therefore correct diagnosis. This work led to a subsequent follow-up study using the patients identified with biallelic *RFC1* expansions within the cohort of late-onset ataxia patients as well as further positive cases from centres in New Zealand, France, Italy, Brazil, Slovenia, Australia, and the UK. This was necessary to determine the full prevalence of the expansion in patients to fully characterise the phenotypic spectrum, currently the estimated prevalence of the biallelic recessive trait is approximately 1/20,000. The biallelic repeat expansion in RFC1 was identified in 90% of patients with full CANVAS symptoms and 14% of patients with adult-onset ataxia (53). The clinical evaluation of these patients characterised the clinical spectrum seen in CANVAS patients and those with *RFC1*-associated ataxia, including an unexplained spasmodic dry cough in 60% of patients, MRI and clinical signs of cerebellar involvement in 66%, vestibular symptoms in 33% and signs of peripheral neuropathy in approximately 100% of patients on nerve conduction study (NCS) (53). This shows the wide-ranging spectrum of clinical features associated with *RFC1*-associated disease.

As discussed in chapter 3.1, CANVAS is frequently misdiagnosed, often with a differential diagnosis of MSA. Within this study I have identified three patients that have biallelic *RFC1* repeat expansions, presenting with a combination of cerebellar ataxia and parkinsonism; two initially diagnosed with possible MSA and one with probable MSA. The absence of the biallelic repeat expansion in *RFC1* in pathologically confirmed MSA patients, suggests that the expansion is not involved in the pathogenesis of MSA, defined with characteristic cytoplasmic inclusions (253). However, the presence of the biallelic *RFC1* expansion in three patients diagnosed with either 'possible' or 'possible' MSA, suggests that the phenotype associated with *RFC1* repeat expansions could include mimics of MSA and other atypical parkinsonian disorders (252).

The difficulty in reaching an early diagnosis of MSA is well documented due to a broad collection of symptoms affecting patients (254). Interestingly, patient 1 identified biallelic for the *RFC1* expansion presented with an abnormal DaTscan, a supportive feature of the parkinsonism of MSA which has not been described in CANVAS. They also presented with cardiovascular autonomic dysfunction and cerebellar ataxia on head MRI, in line with the diagnosis of probable MSA. Furthermore, patient 1 did not develop vestibular areflexia, differing from a probable CANVAS diagnosis. However, they had additional clinical features that were not typical of MSA include sensory neuropathy and slower disease progression. The average survival age from diagnosis for MSA is reported as 6-10 years, although there have been reports of average 17-years survival of pathology-proven MSA-P cases (254, 255). Patient 2

also presented with a combination of cerebellar ataxia and distinct features of parkinsonism including mild bradykinesia, limb rigidity and a lack of arm swing. Patient 3 presented with autonomic dysfunction, cerebellar ataxia, and levodopa non-responsive parkinsonism which are classical signs of MSA-P, however it could be possible that this patient presents with an incomplete penetrance of the *RFC1* repeat expansion, which is explored further in chapter 3.3 All patients presented with an MSA mimic in the early stage, but slower clinical and radiological progressions indicate that they are not phenotypically representative of MSA and if these cases came to autopsy, they would not be expected to show MSA pathology.

This study has contributed to the broadening phenotypic spectrum implicated with *RFC1*, as well as confirming that pathogenically confirmed MSA is not associated with the expansion, but that clinical MSA is associated, which has been replicated in further studies (169, 256). The biallelic repeat expansion has also been implicated in unspecified late-onset ataxia, hereditary sensory neuropathy (HSN) with cough, idiopathic sensory neuropathy and Sjögren syndrome (257-259). It has also been associated with multisystemic phenotypes overlapping progressive supranuclear palsy and hyperkinetic movement disorders (16). It will be important in the future to continue to identify other diseases implicated with the *RFC1*-repeat expansion, in order to determine its true clinical scope. This will be an important caveat of mechanism discovery, determining which phenotypes result directly from the *RFC1*-expansion and will require neuropathological analysis of brain tissue of patients with varying presentations of *RFC1*-related disease.

A limitation of this study was the low numbers of patients within the sample cohort; a follow-up validation cohort would be crucial to replicate findings. However, MSA is a rare neurological disorder, so this would require possible centre collaboration in order to recruit a large enough cohort. Importantly, another study has replicated these findings, identifying 3 patients from 282 MSA cases with the biallelic repeat expansion, and one MSA patient with a $AAGGG_{(n)}/AAAGG_{(n)}$ genotype, that has uncertain pathogenicity (256). This again affirms the possibility that MSA might share the same genetic background as *RFC1*-associated CANVAS patients.

Despite the absence of amplifiable product on flanking PCR and positive RPPCR trace, the lack of Southern blot, due to insufficient DNA, for patient 3 was another limitation, as Southern blotting is the gold standard for visualising repeat expansions. In addition, patient 3 was not fit for a secondary clinical examination, unlike patients 1 and 2, where examinations were video recorded. Importantly, patient 3 presented with a more MSA-P phenotype, than patients 1 and 2, with more severe parkinsonism features. The *RFC1* repeat expansion has been reported in a case with dopa responsive parkinsonism, which could suggest that parkinsonism is associated with the repeat expansion (260). This highlights the need for larger cohort of patients with atypical parkinsonism to be screened for this expansion.

3.2.6.Conclusion

This study has confirmed the presence of the *RFC1* repeat expansion in a clinically diagnosed cohort of MSA patients, whilst confirming its absence in pathologically confirmed MSA. Three patients, two with 'possible' MSA and one with 'probable' MSA were confirmed to have the biallelic repeat expansion through flanking PCR, RPPCR and Southern blotting. They all share differing clinical features of MSA and CANVAS including cerebellar impairment, parkinsonism, and autonomic dysfunction, with atypical presentation for either disease. This highlights the prevalence of the *RFC1* repeat expansion in other disorders aside from late-onset ataxia and CANVAS. The phenotypic scope associated with the *RFC1* repeat expansion is broadening, evidenced by this research identifying the expansion in an MSA-mimic disorder. It will be important for future work to fully characterise this clinical spectrum to enable discovery of a pathological mechanism, which may directly cause a shared phenotype across these disorders. If research in the future enables *RFC1* to be targeted therapeutically, it will be crucial to ensure that all patients that harbour the mutation are identified in case they may benefit from this treatment. Neuropathological analysis of brain tissue from the host of diseases implicated with the repeat expansion will be crucial future work. The ongoing screening, identification of *RFC1*-positive patients and subsequent broadening of the phenotypic spectrum associated with *RFC1* could make this work possible in the near future.

3.2.7.Chapter 5 associated publications

RFC1 Intronic Repeat Expansions Absent in Pathologically Confirmed Multiple Systems Atrophy.
Sullivan R, Yau WY, Chelban V, Rossi S, O'Connor E, Wood NW, Cortese A, Houlden H.Mov Disord. 2020 Jul;35(7):1277-1279. doi: 10.1002/mds.28074. Epub 2020 Apr 24.

RFC1-related ataxia is a mimic of early multiple system atrophy.
Sullivan R, Yau WY, Chelban V, Rossi S, Dominik N, O'Connor E, Hardy J, Wood N, Cortese A, Houlden H.J Neurol Neurosurg Psychiatry. 2021 Feb 9;92(4):444-6. doi: 10.1136/jnnp-2020-325092. Online ahead of print.

Reply: Dopa-Responsive Parkinsonism in a Patient With Homozygous RFC1 Expansions.
Sullivan R, Yau WY, Chelban V, Rossi S, O'Connor E, Wood NW, Cortese A, Houlden H.Mov Disord. 2020 Oct;35(10):1890-1891. doi: 10.1002/mds.28279.

3.3. Development of the ExpansionHunter associated RFC1 Caller as a tool to identify *RFC1* repeat expansions

3.3.1. Introduction

3.3.1.1. Short Tandem Repeats

Short random repeats (STR) comprise at least 6.77% of the human genome (261). STRs, also referred to as microsatellites, are small sections of DNA that are repeated consecutively at a particular locus and are usually 2-6 nucleotides long in length. STR expansions play a pivotal role in many neurological diseases, including a number of rare inherited disorders. They are relatively unstable, due to an increased likelihood of their alteration via processes such as direct alteration of the DNA sequence, changes in the size of the tandem repeat array, changes in the organisation of the tandem repeat, and mutations in the sequence that codes for chromatin modifiers involved in epigenetic regulation (262). This instability introduces genetic variation within the population and increases disease susceptibility (263).

Large STRs are commonly more unstable and have a higher probability of expansion during DNA replication (264). They subsequently may become pathogenic, causing a wide range of neurological disease, including CANVAS, HD, ALS, FTD and multiple SCAs (table 3.3.1). Large STRs have also been implicated in polygenic, complex diseases including schizophrenia, bipolar disorder, major depressive disorder and heart disease, and may contribute to the missing heritability seen in these polygenic disorders (265).

STRs have a propensity to expand during DNA replication, explaining the presence of sporadic, as well as inherited disorders caused by STR expansions. The instability of STRs causes clinical anticipation, which describes the phenomenon of increasing severity of disease and/or earlier onset of symptoms through generations. An example of this occurs in HD, where expansion sizes increase through generations, becoming more unstable (266). Other examples include the paternal transmission pattern seen in SCA1, SCA2, SCA7 and DRPLA. Clinical anticipation is particularly severe in SCA7 with affected children reported to develop symptoms before an affected grandparent or parent (267).

Sequence interruptions have been found in various STRs, and have been found in HD, myotonic dystrophies, hereditary ataxias, and Fragile X (264). Although not fully understood, some of these interruptions have been shown to directly alter the phenotype of the resultant disease. Within the huntingtin gene (*HTT)* the loss of a ‘CAA’ interruption within expanded ‘CAG’ tracts has been shown to lead to an estimated 9.5 years earlier disease onset in HD (268, 269). The ‘CAA’ interruption has been implicated in polyglutamine repeat expansions including SCA17 and SCA2, occurring within ‘CAG’ coding tracts (270, 271). These still translate to glutamine, but form shorter, branching hairpin structures which increases stability of the repeat by reducing strand slippage (272, 273). Interestingly,

repeat interruptions of the sequences $AAAGG_{(n)}$ and $AAGGG_{(n)}$ were seen in the work identifying the pathogenic AAGGG repeat expansion in *RFC1* causal of CANVAS, detailed in chapter 4, however currently their potential influence on pathogenicity or phenotype is unknown (2). A novel pathogenic configuration, $AAAGG_{(10\text{-}25)}AAGGG_{(n)}$, was identified in Māori populations (detailed in chapter 4), involving a 10–25 repeating AAAGG unit interruption detected before the pathogenic $AAGGG_{(n)}$ configuration (1). Pentanucleotide STR regions, such as the biallelic AAGGG repeat expansion seen in *RFC1*, are particularly unstable and dynamic, displaying a wide range of heterogeneity within the normal population as well as patients. SCA10 has phenotypic differences influenced by the dynamic structure of the pathogenic ATTCT repeat in *ATXN10*; with pure ATTCT tracts associated with parkinsonism, whereas the interrupted ATCCT repeats are associated with epilepsy (274-276). The locus where the biallelic repeat expansion in *RFC1* is found is highly polymorphic and dynamic, with numerous repeat motifs identified including the novel pathogenic ACAGG identified in Asian-Pacific families, as well as some whereby their pathogenicity is yet to be concluded (1, 2, 171).

Gene	**Repeat Motif**	**Pathogenic repeat number**	**Mode of inheritance**	**Clinical phenotype**	**References**
RFC1	AAGGG ACAGG	400 - 2000	Autosomal Recessive	CANVAS	(1, 2, 166, 171)
C9orf72	GGGGCC	24 - 4000	Autosomal Recessive	Amyotrophic lateral sclerosis, frontotemporal dementia	(277-279)
FMR1	CGG	200 – 3000 55 - 200	X-Linked	Fragile X Syndrome Fragile X tremor/ataxia syndrome, premature ovarian failure 1	(280, 281)
FXN	GAA	66 - 1300	Autosomal Recessive	Friedreich ataxia	(20)
ATXN1	CAG	33 – 91	Autosomal Dominant	Spinocerebellar ataxia 1	(282)
ATXN2	CAG	33 – 200	Autosomal Dominant	Spinocerebellar ataxia 2	(283, 284)
ATXN3	CAG	53 - 87	Autosomal Dominant	Spinocerebellar ataxia 3	(285)
CACNA1A	CAG	19 - 33	Autosomal Dominant	Spinocerebellar ataxia 6	(284, 286)
ATXN7	CAG	34 – 460	Autosomal Dominant	Spinocerebellar ataxia 7	(287, 288)
ATXN8	CAG/TAG	74 – 1300	Autosomal Dominant	Spinocerebellar ataxia 8	(289)
ATXN10	ATTCT	280 – 4500	Autosomal Dominant	Spinocerebellar ataxia 10	(275, 290)
PPP2R2B	CAG	51 – 78	Autosomal Dominant	Spinocerebellar ataxia 12	(291, 292)

TBP	CAG	43 – 66	Autosomal Dominant	Spinocerebellar ataxia 17, Huntington disease-like 4	(293, 294)
BEAN1	TGGAA	500 – 760	Autosomal Dominant	Spinocerebellar ataxia 31	(295)
NOP56	GGCCTG	650 – 2500	Autosomal Dominant	Spinocerebellar ataxia 36	(74)
DAB1	ATTTC	31 - 75	Autosomal Dominant	Spinocerebellar ataxia 37	(296)
CSTB	CCCCGCCCCGCG	30 - 125	Autosomal Recessive	Progressive myoclonic epilepsy 1A	(297)
HTT	CAG	36 - 250	Autosomal Dominant	Huntington Disease	(19)
PRNP	PHGGGWGQ *24-base octapeptide	8 -14	Autosomal Dominant	Huntington Disease-like 1	(298)
JPH3	CTG	40 - 59	Autosomal Dominant	Huntington Disease-like 2	(299)
NOTCH2NLC	CGG	66 - 517	Autosomal Dominant	Neuronal intranuclear inclusion disease	(300, 301)
XLR	CAG	36 - 68	Autosomal Recessive	Spinal and bulbar muscular atrophy of Kennedy	(302-304)
LRP12	CGG	90 - 130	Autosomal Dominant	Oculopharyngodistal myopathy	(305)
GIPC1	CGG	70 - 164	Autosomal Dominant	Oculopharyngodistal myopathy	(306)
PABPN1	GCG	7 - 18	Autosomal Dominant	Oculopharyngeal muscular dystrophy	(307, 308)
NUTM2B-AS1	CGG	16-160	Autosomal Dominant	Oculopharyngeal myopathy with leukoencephalopathy 1	(305)
VWA1	GGCGCGGAGC	3	Autosomal Recessive	Hereditary axonal motor neuropathy	(309)
DMPK	CTG	50 - 10000	Autosomal Dominant	Myotonic dystrophy 1	(310, 311)
CNBP	CCTG	50 – 11000	Autosomal Dominant	Myotonic dystrophy 2	(311)
ARX	GCC	17 – 27	X-linked	X-linked lissencepaly 2, X-linked mental retardation 29, hydraencephaly with abnormal genitalia, developmental and epileptic encephalopathy 1	(312, 313)
ATN1	CAG	49 - 93	Autosomal Dominant	Dentatorubral-pallidoluysian atrophy	(314)

SAMD12	TTTCA	105 - 3680	Autosomal Dominant	Familial adult myoclonic epilepsy 1	(315, 316)
STARD7	ATTTC	150 - 460	Autosomal Dominant	Familial adult myoclonic epilepsy 2	(317)
MARCHF6	TTTCA	700 - 1035	Autosomal Dominant	Familial adult myoclonic epilepsy 3	(318)
TNRC6A	TTTCA	Unknown	Autosomal Dominant	Familial adult myoclonic epilepsy 6	(316)
RAPGEF2	TTTCA	Unknown	Autosomal Dominant	Familial adult myoclonic epilepsy 7	(316)
FMR2	CCG	>200	X-linked recessive	Mental retardation, X-linked, FRAXE type	(319)

Table 3.3.1: Summary table of neurological diseases causes by STR expansions

CANVAS, cerebellar ataxia, neuropathy and vestibular areflexia syndrome. Table adapted from Chintalaphani et al., (264).

3.3.1.2. Diagnostic identification of STR expansions

STR expansions can be detected with molecular biology techniques, as well as NGS. The established process for the molecular diagnosis of repeat expansion disorders involves flanking PCR, RPPCR and Southern blotting to accurately size large expansions. The work detailed in chapter 4 on the repeat expansion in *RFC1* as the genetic cause of CANVAS established an effective protocol for identifying and sizing repeat expansions within the laboratory (figure 3.3.1). Patients that fit the clinical phenotypes for *RFC1*-related disease are selected through clinical screening and gDNA from those patients is used to perform targeted flanking PCR. Samples with no amplifiable DNA seen from flanking PCR are selected for RPPCR. A positive RPPCR trace for the biallelic repeat expansion presents with a hallmark sawtooth pattern of decremental peaks, and samples showing this are selected for Southern blotting. Southern blotting allows for visualisation of the repeat expansion and accurate sizing (2, 252, 253). Southern blotting is regarded as the gold standard for detection of large repeat expansions; however, it is time-inefficient and requires a large amount of high quality of gDNA. RPPCR can be used for some repeat expansions, without Southern blotting, but this is less accurate at determining the length of the repeat due to the fragmentation of the repeated region into shorter reads. It is also challenging to establish an effective protocol for those repeats with a high GC content, and high level of polymorphism, such as the *RFC1* and *C9orf72* repeat expansions (79). Allelic dropout of PCR amplification and high levels of secondary structure formation in these polymorphic and interruption rich regions can lead to sequencing errors and inaccuracy in motif identification (320, 321).

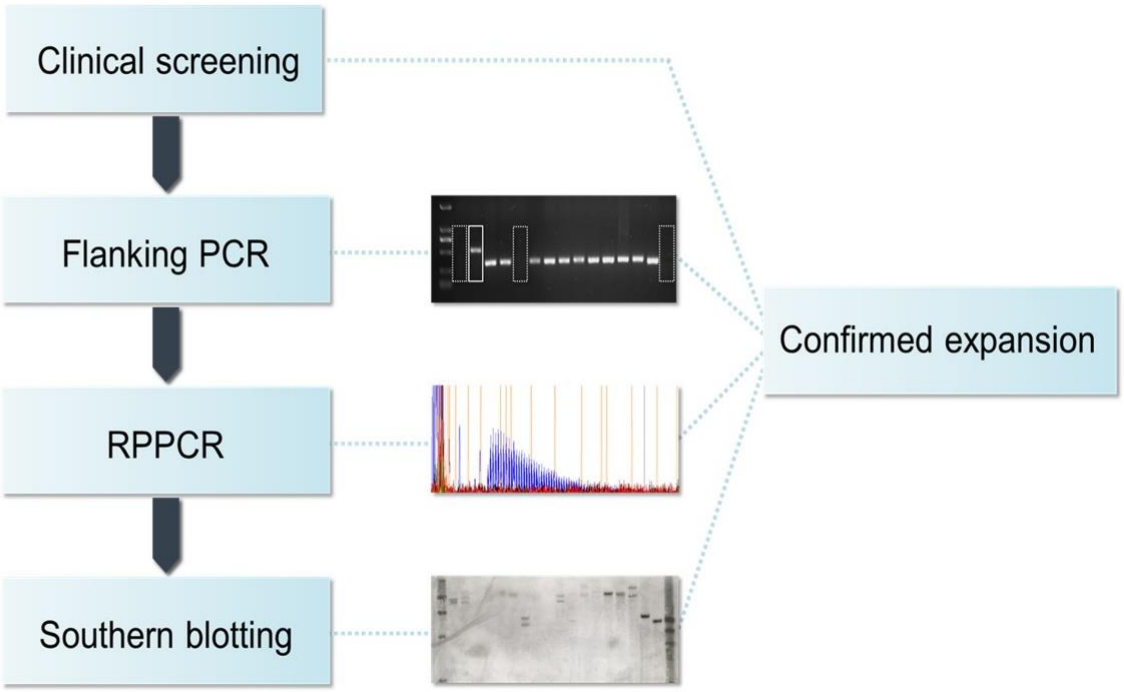


Figure 3.3.1: Workflow of molecular diagnosis of RFC1 repeat expansion. Patients with clinical phenotypes characteristic of *RFC1*-related disorders put forward for molecular diagnostic testing through clinical screening. Flanking PCR performed on samples, samples with no amplifiable product on PCR selected for RPPCR. Samples with a positive RPPCR trace showing hallmark sawtooth pattern of decremental peaks selected for Southern blotting. Southern blotting allows for estimation of expansion size. Figure made by R.S. for publication by Sullivan et al., (252). *RPPCR, repeat primed PCR*

3.3.1.3. Computational identification of RFC1 recessive repeat expansion

An alternative approach to identifying STR expansions employs NGS techniques that use short-read NGS platforms. Specific bioinformatic tools have been developed to analyse NGS data, including ExpansionHunter, HipSTR, GangSTR, RepeatSeq, and LobSTR (86, 322-324). These tools allow for investigations into allelic segregation, inheritance patterns in affected families and the identification of novel expansions. WGS allows for the profiling of the entire genome, along with non-STR mutations and STR contractions, which can also be implicated in disease (264). However, short-read NGS does have limitations to its ability to identify STR expansions, which has restricted its widespread adoption as a diagnostic tool for expansion repeat disorders. NGS library preparation can be hampered by GC rich regions of the genome, along with PCR amplification and sequencing, which can affect coverage. Stutter errors can also be introduced during the library preparation PCR amplification step (325). Misalignment and uncertain alignment to reference genome reads have also been shown to be caused by STR regions, which are highly repetitive (264). Larger STR regions are not covered by the short-read length of NGS (approximately 100 – 150bp), preventing the estimation of their full-length (264).

Additionally, epigenetic modifications cannot yet be detected using NGS, aside from the use of bisulphate sequencing, which must be carried out separately (264, 326).

Long read sequencing platforms that include Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT) have emerged as important tools to evaluate suspected repeat expansion disorders and surmount the limitations associated with short-read NGS platforms. A biological nanopore is used in ONT devices to measure the displacement of ionic current caused by a DNA strand passing through. This is subsequently translated into DNA sequence (264). It has the capability to span the entire length of large STR expansions and has successfully identified the expanded 'CCCCGG' repeats in *C9orf72* and 'GGCCTG' repeats in *NOP56*, spanning 80 - 99.5% of reads (327). Another advantage of ONT is that methylation status at a given locus can be determined simultaneously, providing another crucial level of diagnostic evidence for associated disorders (328). PacBio Single Molecule, Real-Time (SMRT) sequencing technology exploits the process of nucleotide incorporation into single DNA-template-polymerases, detecting fluorescent nucleotide signals in real time (329). PacBio reads are highly accurate and allow for the detection of possible variants in adjacent areas (264).

ExpansionHunter

A software package called ExpansionHunter was created as a method to genotype STRs using PCR-free, WGS data by Dolzhenko et al (86). ExpansionHunter has the ability to detect the size of repeats from both small repeats in length, to large expansions that may be longer than the read length. This gives the software the ability to identify both known and novel pathogenic repeat expansions within population and pedigree studies. An advantage of ExpansionHunter is its ability to estimate region length via read number, even in regions with a high GC content, which it does using in-repeat reads (IRRs) (figure 3.3.2 & 3.3.3). The software successfully detected and classified expanded samples in a results cohort of patients with *C9orf72* repeat expansions, previously identified by wet lab protocols (86). The classifications were expanded (208 samples), potentially expanded (4 samples) and wild type, of which 2786 of 2789 were correctly detected, showing the accuracy of the software (86). ExpansionHunter has also been successfully used to detect short repeat expansions, with a sensitivity of 97.3% and specificity of 99.6% (330). One single WGS run can be used to screened for all known pathogenic repeat expansions, which is a major advantage of Expansion Hunter, which, as the cost of WGS decreases and throughput increases, could become the principal method of detecting repeat expansion disorders (86).

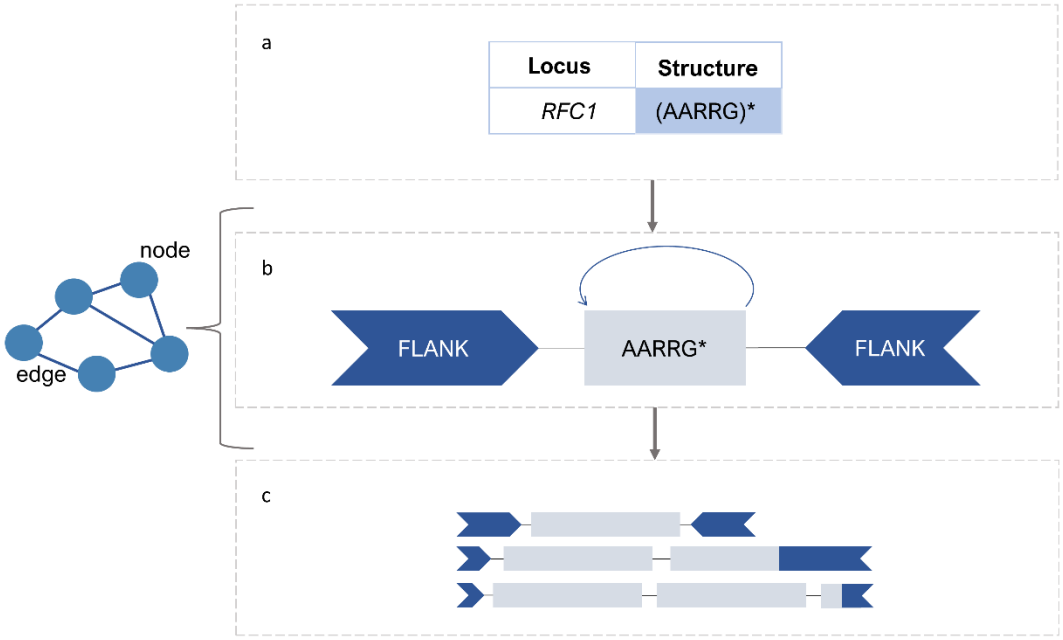


Figure 3.3.2: Schematic of ExpansionHunter prediction for RFC1 locus. A, Locus variant catalogue defines RFC1 locus definition (AARRG)*, which is a subset of the expression syntax. B, A sequence graph is constructed based on the locus definition within the variant catalogue. This consists of nodes that correspond to sequences, and edges define and direct the connection of these sequences. These edges and nodes are assembled to resemble the different allele predictions. C, reads are aligned to the sequence graph and alignments are assembled.

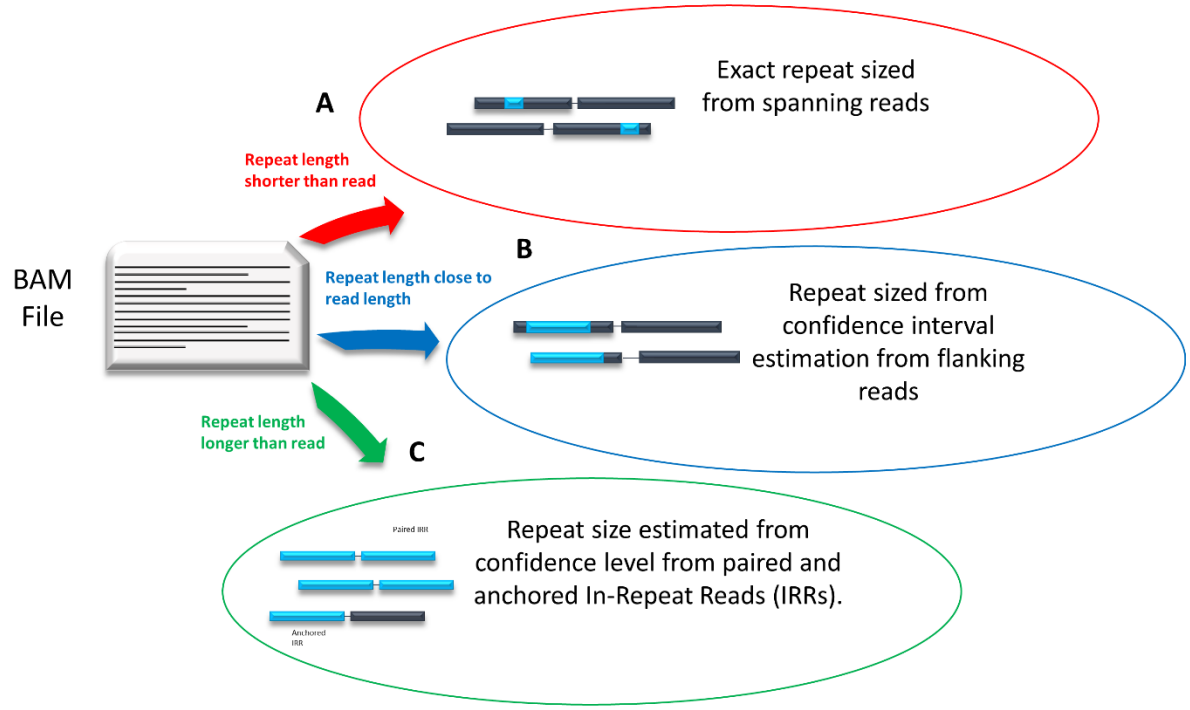


Figure 3.3.3: A schematic describing how ExpansionHunter produces repeat estimations from a binary alignment/map (BAM) file based on associated reads. A, Short repeats are sized exactly from spanning reads that are contained completely within the sequence. B, Size of repeats that are nearly the size of the read length are estimated using flanking reads that partially overlap the repeat and one of the repeat flanks. C, Repeats longer than the read length are estimated in size using paired IRRs that are anchored to the repeat region or when no repeats with the same repeat motif are found within the rest of the genome, paired IRRs are used to estimate long repeat size. *IRRs, In-Repeat-Reads*

3.3.1.4. Genomics England

Genomic England (GEL) has founded the 100kGP as a platform for using WGS data to advance the clinical application of genomics. The rare disease programme was created to enable WGS within the National Health Service (NHS) to identify novel genetic causes in rare inherited diseases. Within the 100kGP, thirteen centres across the England participate and provide tissue, which is prepared, DNA extracted and quantified, and then transferred to the national sequencing centre where it undergoes NGS (331).

The diagnostic rate for rare diseases that are likely to have a genetic cause is reported as 22%, with higher rates in specific diseases such as intellectual disability (331). 190 out of an estimated 8000 rare diseases were recruited into the 100kGP project, with approximately 39% of probands and their families recruited with neurological disease, figure 3.3.4 shows the range of these disorders (332). These diseases were selected on the basis that there was a large proportion of patients who did not have a known genetic molecular diagnosis for their disease. Trios were recruited, were possible, allowing for

variant filtration and inheritance pattern analysis. The majority of samples recruited for adult neurological disorders are singletons.

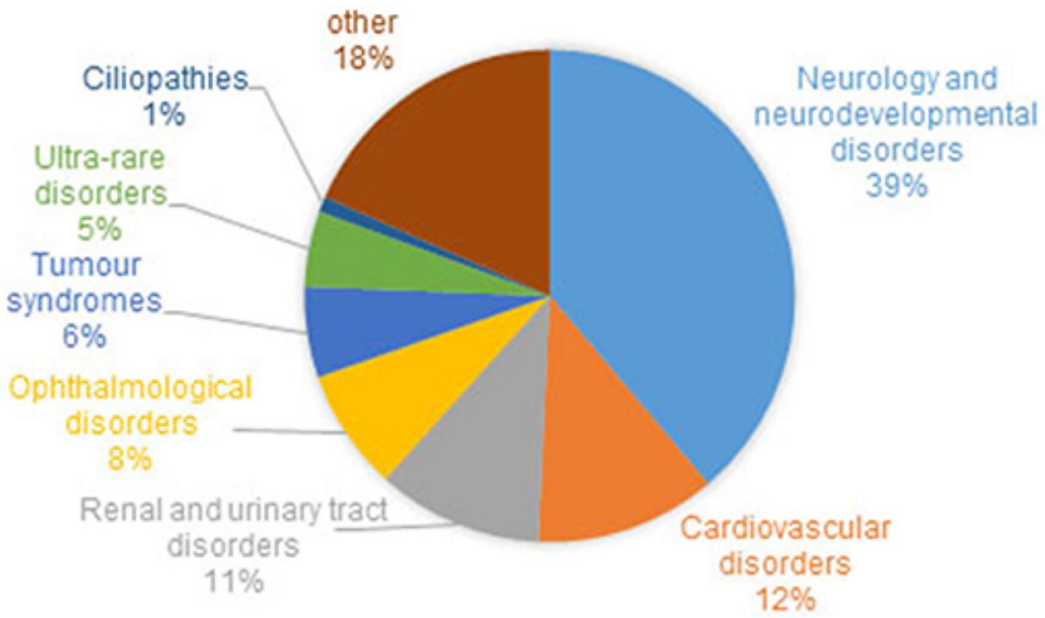


Figure 3.3.4: Proportion of families and probands with different rare diseases, recruited to the 100,000 Genome Project. 39% of probands and families were recruited with neurological and neurodevelopmental disorders. Figure taken from Wheway et al., 2019 (333).

3.3.2. Aims of study

The aim of this study was to aid in the development of a new algorithmic tool to detect pathogenic *RFC1* repeat expansions, that can be used in conjunction with ExpansionHunter. Currently expansion hunter can only effectively detect repeat expansions of short length and is unable to distinguish between repeat motifs. In collaboration with Dr Sai Chen and Dr Michael Eberle from Illumina, the wealth of data available in the 100kGP was utilised to screen a cohort of 62, 550 samples, including 9654 trio families (mother, father and proband) for the pathogenic *RFC1* repeat expansion, developing and refining the accuracy of the RFC1 caller.

The allelic frequency of the repeat expansion is estimated to be around 0.7%-6.8% in the normal population (2, 50, 166), which was confirmed in the rare diseases' cohort within the 100kGP project. In addition, *RFC1*-associated disease is thought to be a Mendelian disorder which was investigated by looking at segregation within trios. The sensitivity and specificity of the RFC1 caller was analysed using a cohort of patients put forward for *RFC1* diagnostic screening based on their clinical phenotype. The expansion status of these patients was confirmed using Southern blotting. The aim was to create a bioinformatic tool with the ability to detect individuals with the biallelic *RFC1* expansion, negating the

need for the screening combination of flanking PCR, RPPCR and Southern blotting. This will reduce the amount of time required for reporting molecular diagnosis, and in addition provide a platform in which to screen existing large WGS databases of patients, helping solve rare neurological cases with no currently known molecular diagnosis.

3.3.3. Summary of methods

Study aim	Methods used	Contributors	Methods section
RFC1 caller algorithm development	RFC1 caller algorithm script design	Dr Sai Chen, Dr Michael A. Eberle	2.4.4
Identification of *RFC1* repeat expansion in rare diseases GEL cohort	Running of ExpansionHunter in R statistical environment	Roisin Sullivan	2.4.4
	Script design for analysis in R Statistical environment	Roisin Sullivan	2.4.4, Appendix 8.1
Visual inspection of repeat expansion	Pile-up plot script	Roisin Sullivan	Appendix 8.1.2
Identification of *RFC1* repeat expansion in in validation cohort (unvalidated cases only on Southern blot)	Flanking PCR	Roisin Sullivan, Dr Wai Yan Yau, Natalia Dominik, Valentina Galassi Deforie	2.2.2
	RPPCR	Roisin Sullivan, Dr Wai Yan Yau, Natalia Dominik, Valentina Galassi Deforie	2.2.3
	Southern blotting	Roisin Sullivan	2.2.4
Haplotype investigations	Script design for analysis in R statistical environment	Roisin Sullivan	Appendix 8.1

3.3.4. Results

3.3.4.1. RFC1 caller workflow utilises ExpansionHunter output files

WGS data from the rare diseases' cohort from the 100kGP were run on ExpansionHunter (figure 3.3.5). The RFC1 caller was then run using the on BAM and VCF output files, generated from ExpansionHunter, assigning an allelic status to each sample, normal: no expanded alleles, normal: one expanded benign and one short reference, normal: no expansions observed, normal: two expanded alleles (possibly reference), carrier: one expanded pathogenic allele, one expanded benign allele, carrier: one expanded pathogenic allele, one short reference allele, potential carrier: two expanded alleles with <30% pathogenic kmers and patient: two expanded pathogenic alleles. Individuals identified as 'patient: two expanded pathogenic alleles' which had not previously been confirmed via diagnostic workflow and had enough gDNA, were Southern blotted to confirm the pathogenic repeat expansion.

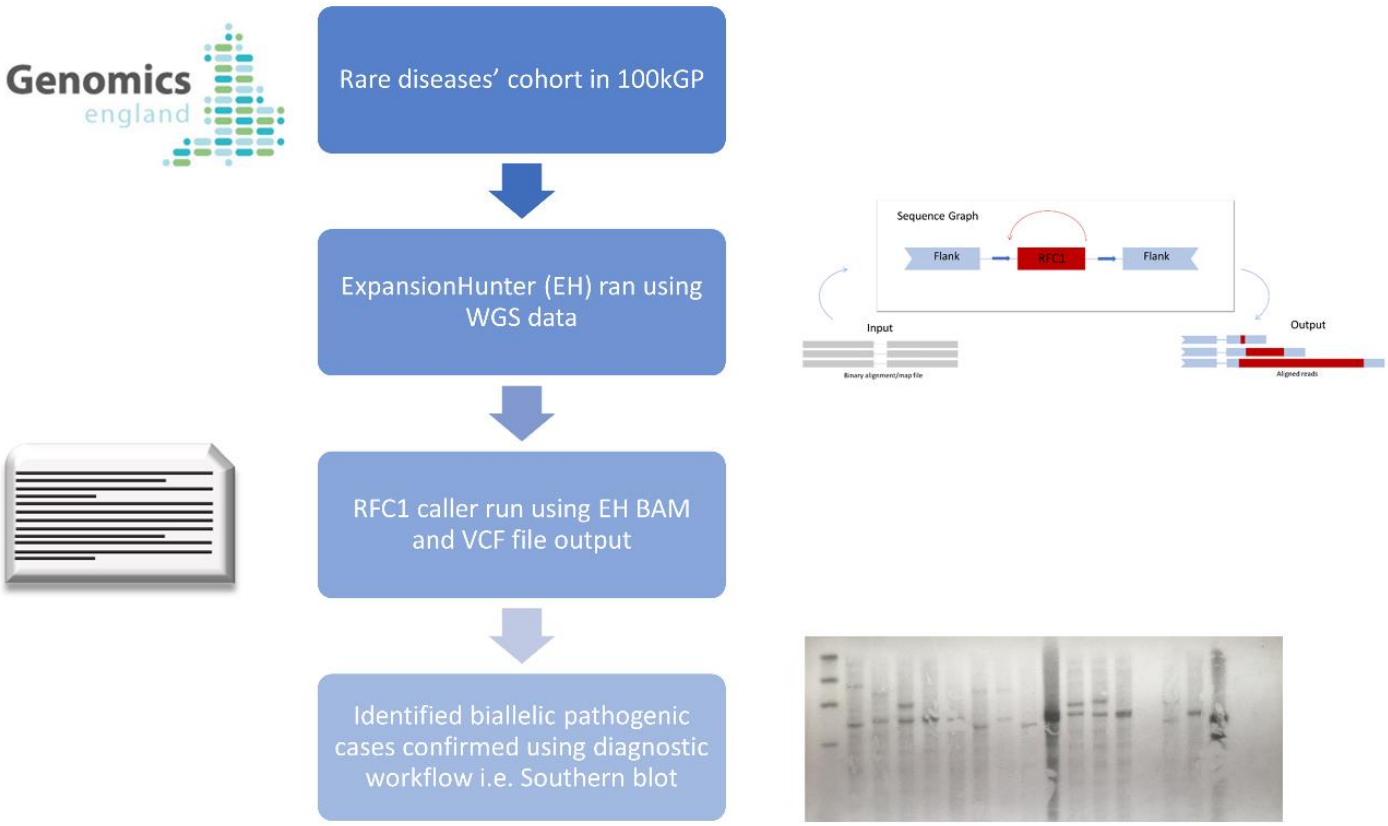


Figure 3.3.5: Workflow of RFC1 caller. WGS data from rare diseases cohort from 100kGP dataset run on ExpansionHunter (EH), which generates BAM and VCF output files. These files were run on the RFC1 caller which identified whether the sample contained one expansion, two expansions or no expansions using an aggressive definition of an expansion (125bp or 50bp longer than the read length). These samples were then classified as 'normal', 'carrier' or 'potential pathogenic'. An algorithm was then employed to detect the exact expansion motif and define the motif combinations on each allele.

3.3.4.2. Allelic distribution within the rare diseases GEL cohort, and results from the RFC1 caller.

In order to distinguish the allelic distribution of the *RFC1* repeat expansion within the 100kGP rare diseases population, the bam and VCF file outputs from ExpansionHunter for 62,550 samples within the rare diseases' cohort in GEL were run on the RFC1 Caller. The allelic distribution of the cohort (n = 125,100) was 55% for the reference AAAAG, 41% for an expanded benign expansion which includes both $AAAAG_{(n)}$ and $AAAGG_{(n)}$ among others, and 4% for the pathogenic expanded $AAGGG_{(n)}$ (figure 3.3.6A).

The results for the caller showed that the majority of samples did not have the pathogenic *RFC1* repeat on either allele; 28,125 (45%) had one expanded benign expansion and one short reference allele, 19, 041 (30%) had no expansions on either allele, and 10,716 (17%) had two expanded benign alleles (possibly $AAAAG_{(n)}$). 4124 samples were carriers of the pathogenic $AAGGG_{(n)}$ (6.6%), with 2493

(3.99%) of these having 1 expanded $AAGGG_{(n)}$ expansion on one allele, and one short reference allele, and 1631 (2.61%) with one expanded $AAGGG_{(n)}$ on one allele, and 1 expanded benign expansion on the other. 368 samples (0.6%) were identified as being potential carriers with greater than 30% kmers (substrings of length k), and 31 samples (0.05%) of samples were called as normal where an expanded allele was not observed. Importantly, 145 (0.2%) samples were identified as having a biallelic pathogenic $AAGGG_{(n)}$ expansion (figures 3.3.6B and C)

Within the 100kGP, samples are annotated with participant type, with 29, 494 (47%) probands and 33, 056 (53%) relatives. Normal with one expanded benign allele and one short refence allele made up 44.8% and 45.1% of probands and relatives respectively, normal, with no expanded allele 30.5% and 30.4%, normal with two expanded (possibly AAAAG) alleles 17.1% and 17.1%, carrier with one short reference allele 4% and 4%, carrier with one expanded benign allele 2.7 % and 2.6 %, potential carrier 0.5% and 0.6%, patient with two expanded pathogenic allele 0.35% and 0.13% and no observed expansion 0.05 % and 0.05% (figure 3.3.6D).

43 relatives were called as 'patient: 2 expanded pathogenic alleles', therefore having the biallelic *RFC1* repeat expansion, compared to 102 probands. The allelic frequency of the pathogenic $AAGGG_{(n)}$ motif was 3.7% and 3.9% for the relatives and probands respectively (figure 3.3.6D). In addition both cohorts had allelic frequencies of 55% for the AAAAG short reference motif, and 41.3% and 41.1% for an expanded benign motif for relatives and probands respectively.

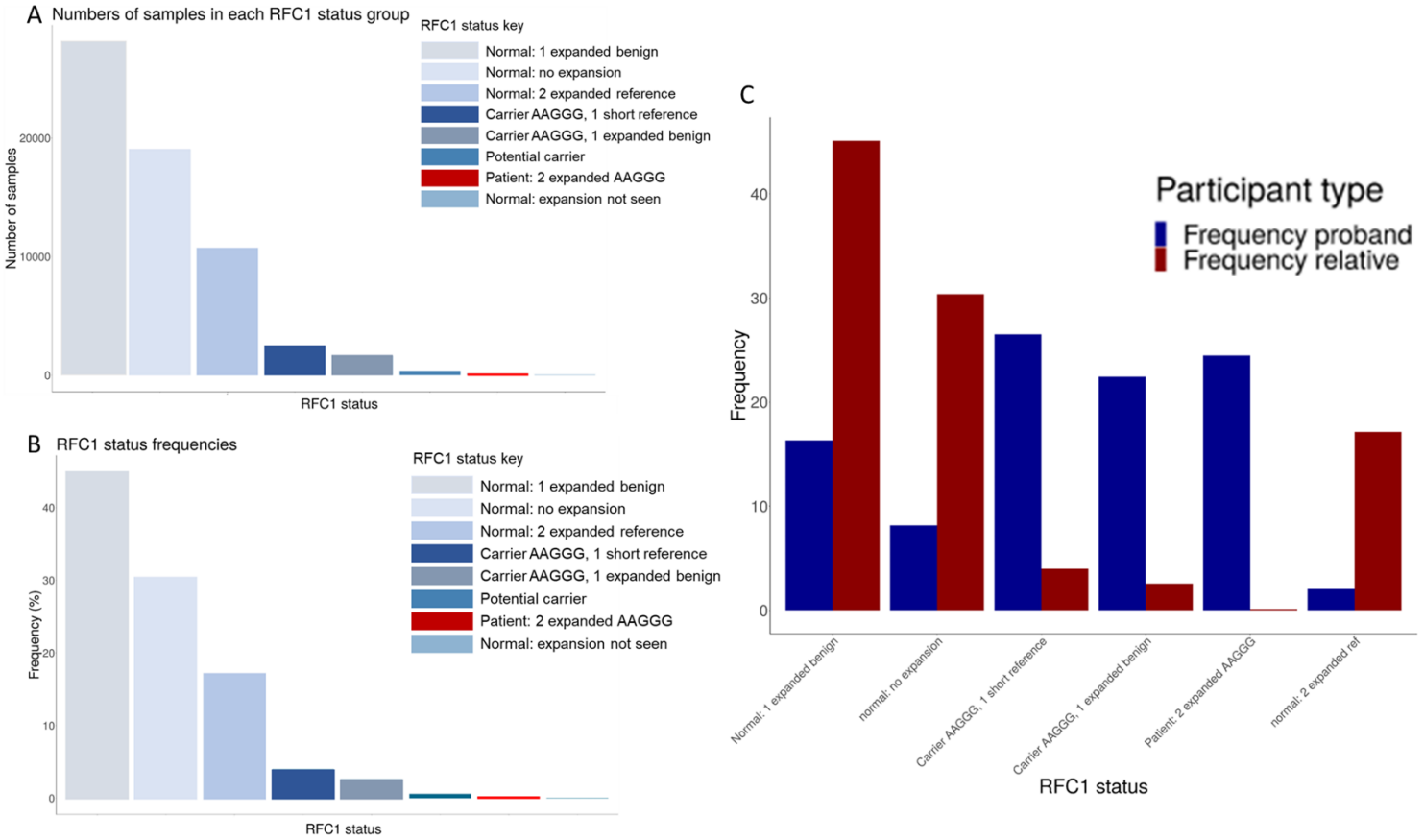


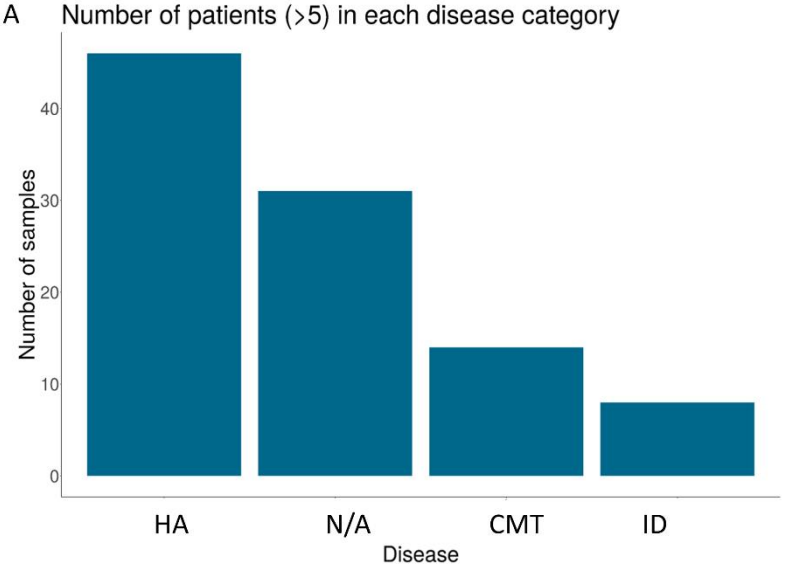
Figure 3.3.6: Allelic distribution within GEL rare diseases cohort, and results from RFC1 caller

A, Bar plot showing allelic distribution of the cohort (125,100 chromosomes) B, Bar plot showing frequency of individuals per *RFC1* status, as called by the RFC1 caller, C, Bar plot showing numbers of individuals in each *RFC1* status, as called by the RFC1 caller, D, Bar plot showing frequency of either probands (n = 29, 494) or relatives (n = 33, 056) per *RFC1* status.

3.3.4.3. Specific diseases involved with the biallelic RFC1 repeat expansion, as identified by the RFC1 caller.

In order to investigate which diseases are most highly associated with the biallelic *RFC1* repeat expansion, RFC1 caller results were sub-grouped into those samples that were identified as 'patient: two expanded pathogenic alleles'. The most prevalent diseases within the cohort were hereditary ataxia (31.7%) and CMT disease (9.7%). 21.4% of samples within this cohort were identified as relatives (picked up as 'N/A' in terms of Normalised Specific Disease). Intellectual disability (5.5%) was the next most prevalent within the cohort. Figure 3.3.7B presents the implicated disease subgroups with less than 5 individuals. Non-neurological disorders such as cystic kidney disease, Long QT syndrome, congenital myopathy and anomalies of the kidneys and urinary tract were also identified in more than one sample (but with <5 individuals) (figure 3.3.7B). As hereditary ataxia and CMT are phenotypically close to CANVAS, and have been associated with the *RFC1* repeat expansion, the RFC1 caller and allelic frequency results were looked at in more detail. Out of 1623 samples with hereditary ataxia and CMT, 53.7% carried the AAAAG short reference allele, 38.7% carried the expanded benign allele and 7.5% carried the pathogenic AAGGG allele (figure 3.3.8A). The majority of samples in the hereditary ataxia and CMT cohort were identified as 'normal' by the RFC1 caller, 60 samples were identified as 'patient: 2 expanded pathogenic alleles' and 113 carriers of the pathogenic expansion (figure 3.3.8B).

A



B

Patient disease subtypes with <5 individuals
Congenital anomaly of the kidneys and urinary tract (CAKUT)
Long QT syndrome
Congenital myopathy
Cystic kidney disease
Early onset dystonia
Epilepsy plus other features
Familial hypercholesterolaemia
Rod-cone dystrophy
Ultra-rare undescribed monogenic disorders
Cerebellar hypoplasia
Congenital heating impairment
Congenital muscular dystrophy
Early onset and familial Parkinson's disease
Hereditary spastic paraplegia
Hypotrophic cardiomyopathy
Inherited macular dystrophy
Insulin resistance (including lipodystrophy)
IUGR and IGF abnormalities
Ketotic hypoglycaemia
Meige disease
Neurofibromatosis type I
Non-CF bronchiectasis
Paediatric motor neuropathies
Primary lymphoedema
Renal tubular acidosis
Rhabdomyolysis and metabolic muscle disorders
Unexplained monogenic fetal disorders
Unexplained skeletal dysplasia
VACTERL-like phenotypes

Figure 3.3.7: Specific diseases involved with the biallelic RFC1 repeat expansion, as identified by the RFC1 caller. A, Bar plot showing number of individuals (>5) in each disease category, Table showing other disease groups with individuals >5 identified as patient: 2 expanded pathogenic alleles. **GEL policy does not permit graphical representation of any data presenting data of <5 individuals**. *HA, hereditary ataxia; N/A, relative; CMT, Charcot-Marie-Tooth disease, ID, intellectual disability; CAKUT, congenital abnormalities of the kidneys and urinary tract, IUGR, intrauterine growth restriction; IGF, insulin growth factor.*

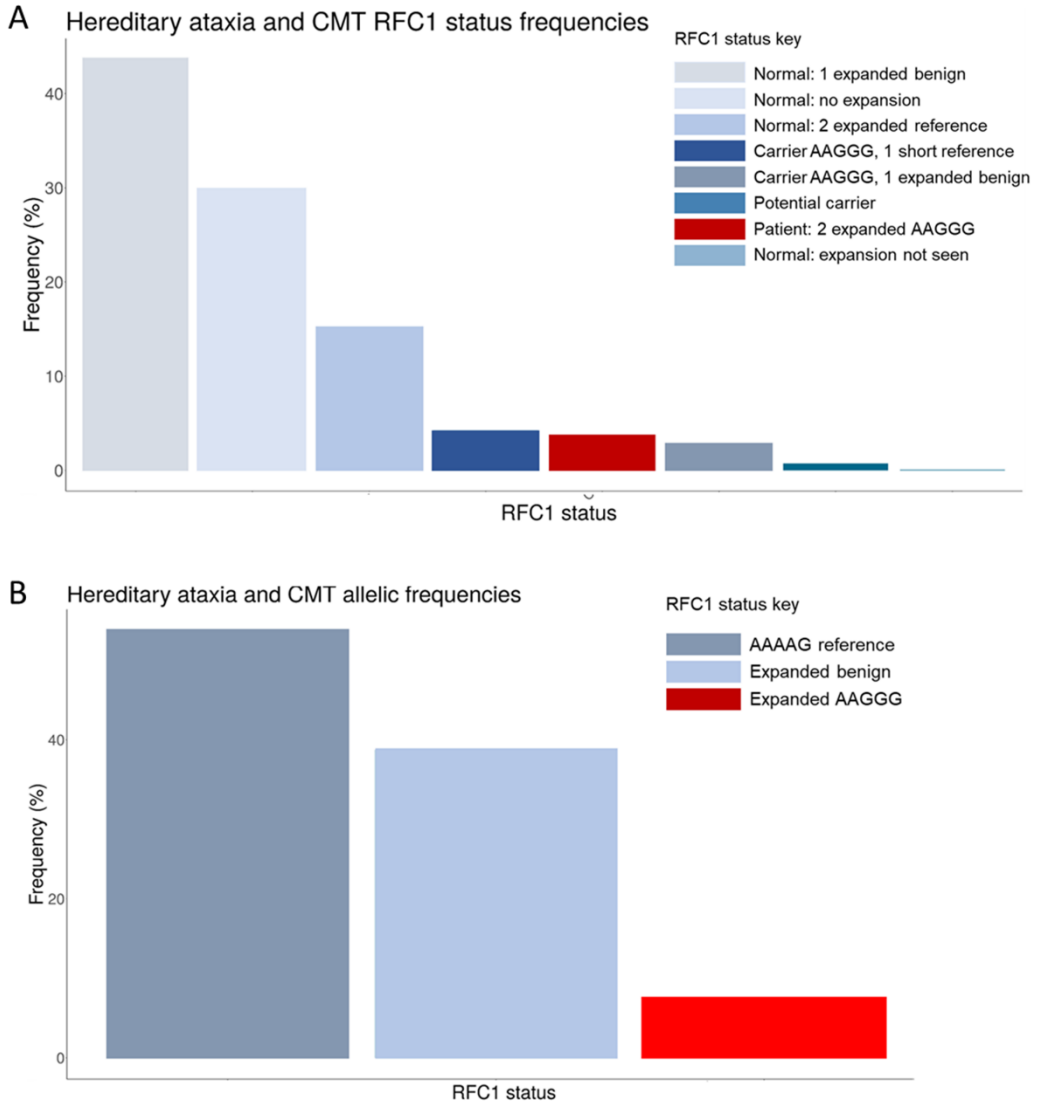


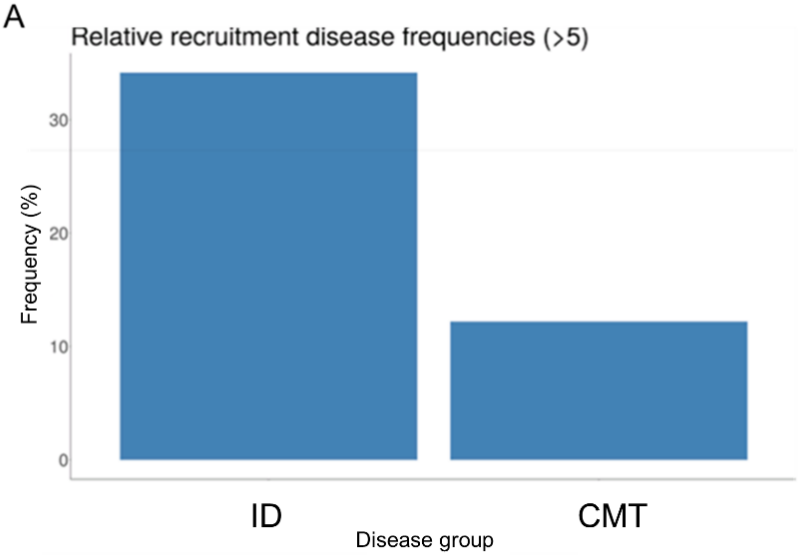
Figure 3.3.8: Frequency of RFC1 statuses in hereditary ataxia and CMT disease cohort. A, Bar plot showing frequency of RFC1 statuses in hereditary ataxia and CMT disease cohort. B, Bar plot showing allelic frequencies in hereditary ataxia and CMT disease cohorts. *HA, hereditary ataxia; N/A, relative; CMT, Charcot-Marie-Tooth disease.*

3.3.4.4. Associated disease groups of relatives subgroup identified as 'patient: 2 expanded pathogenic alleles'

The 'relatives' subgroup identified as 'patient: 2 expanded pathogenic alleles', were recruited into GEL via their associated related proband. As clinical information is not available for these individuals, the related disease group and recruitment domains of the initial related proband were analysed. 34. 2% of the initial probands were donated as intellectual disability, 12.2% as CMT, and the rest as various other diseases including ultra-rare undescribed monogenic disorders (<5), hereditary ataxia (<5) and

mitochondrial disorders (<5) amongst others (figure 3.3.9). In addition, the largest proportion of relatives identified as patient: 2 expanded pathogenic alleles have an associated proband recruited under the Paediatrics domain, (46.3%), followed by Neurology (26.8%), Cardiovascular (<5) and Hearing and sight (<5) (figure 3.3.10).

A

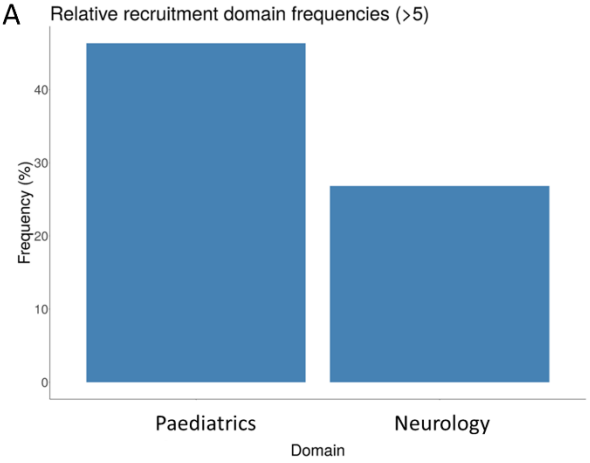


B

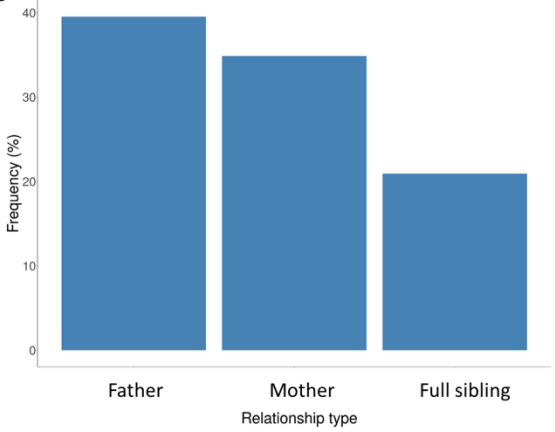
Disease subgroup (<5)
Ultra-rare undescribed mongenic disorders
Hereditary ataxia
Mitochondrial disorders
Undiagnosed monogenic disorder seen in a specialistic clinic
Arthrogryposis
Atypical Beckwith-Wiedemann syndrome
Congenital anomaly of the kidneys and urinary tract (CAKUT)
Congenital hearing impairment
Dilated cardiomyopathy
Epileptic encephalopathy
Familial hypercholesterolaemia
Idiopathic hypogonadotropic hypogonadism
Infantile nystagmus
Long QT syndrome
Non-CF bronchiectasis
Osterogenesis imperfecta
Primary immunodeficiency
Rhomboencephalosynapsis

Figure 3.3.9: A, Barplot of the number of individuals (>5) in each disease group, picked up as 'relative', within patient: 2 expanded pathogenic alleles cohort. B, Frequency of each disease subgroup in 'relatives' category. C, List of disease subtypes with <5 individuals from 'relatives' category of samples called as 'patient: 2 expanded pathogenic alleles. **GEL policy does not permit graphical representation of any data presenting data of >5 individuals**. *ID, intellectual disability; CMT, Charcot-Marie-Tooth disease; CF, cystic fibrosis; IUGR, intrauterine growth restrictions; IGF, insulin growth factor, VACTERL, Vertebral anomalies (V), anal atresia (A), cardiac malformation (C), tracheo-oesophageal fistula (TE) with or without oesophageal atresia, renal dysplasia (R) and limb abnormalities (L).*

A



B



C

Disease subgroup (<5)	Recruitment domains (<5)
Ultra-rare undescribed mongenic disorders	Cardiovascular
Hereditary ataxia	Hearing and sight
Mitochondrial disorders	Endocrine and metabolism
Undiagnosed monogenic disorder seen in a specialistic clinic	Immune disorders
Arthrogryposis	Musculoskeletal
Atypical Beckwith-Wiedemann syndrome	Renal
Congenital anomaly of the kidneys and urinary tract (CAKUT)	Respiratory
Congenital hearing impairment	
Dilated cardiomyopathy	
Epileptic encephalopathy	
Familial hypercholesterolaemia	
Idiopathic hypogonadotropic hypogonadism	
Infantile nystagmus	
Long QT syndrome	
Non-CF bronchiectasis	
Osterogenesis imperfecta	
Primary immunodeficiency	
Rhomboencephalosynapsis	

Figure 3.3.10: Summary of relatives data. A, Barplot of recruitment domain frequencies (>5 individuals) within cohort of relatives identified as 'patient: 2 expanded pathogenic alleles'. B, Frequencies of biological relationship of relatives to probands. C, List of recruitment domains with <5 individuals from relative category identified as patient: 2 expanded pathogenic alleles. **GEL policy does not permit graphical representation of any data presenting data of <5 individuals**. *CF, cystic fibrosis.*

3.3.4.5. *Motif distribution within the rare diseases GEL cohort*

In order to investigate the most prevalent motifs seen within biallelically expanded samples, a motif distribution plot was generated. There were 111 homozygous motifs identified, of which 26 comprise more than 10% of the expansion and 21 are novel. Novel motifs include ACGGG, ACAAG, CAAGG, GAGGG, AACAG, AAGGC, AGAGG, AAGAC, CAGGG, AAACG, AACGG, AAGGA and AAGGT. Figure 3.3.11 shows the prevalence of the 26 from a subset of individuals from the rare diseases' cohort,

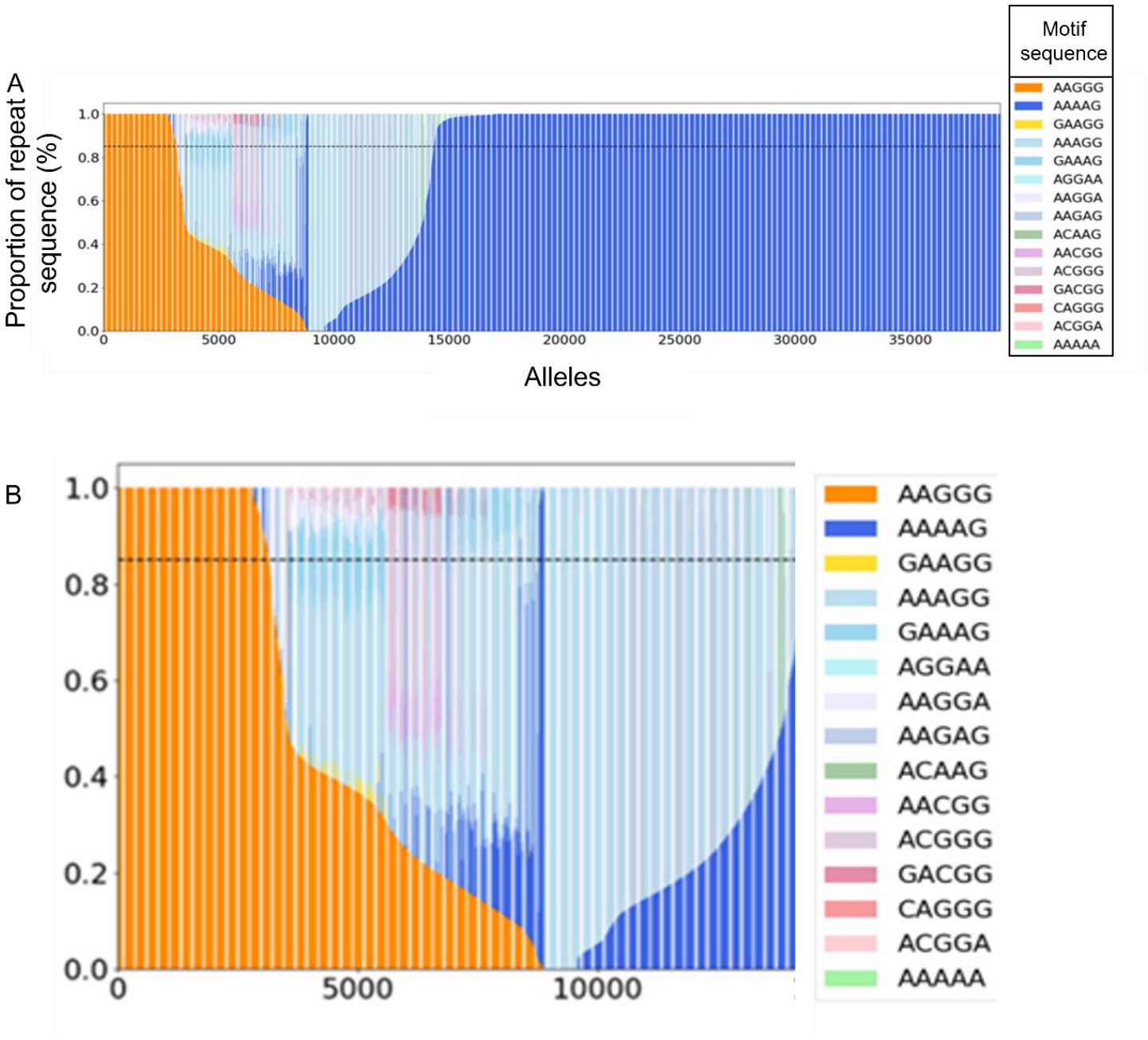


Figure 3.3.11: Distribution of repeat motifs within alleles, within the GEL rare diseases cohort. A, Proportion of top 15 repeat motif sequences that make up a certain percentage of each allele sequence

within a subset of alleles from the rare diseases' cohort. B, Magnified portion of allele numbers 0 – 15000 to allow closer inspection of allelic combinations.

3.3.4.6. Segregation of the expansion within trio families

In order to identify potential caller errors, I looked at the segregation of the expansion within trios: samples where a mother, father and proband were all recruited. Out of 9657 set of trios (28, 971 samples in total) 49 trios contained two carrier parents. In order to check Mendelian inheritance (presumed to be the case for the *RFC1* repeat expansion), I looked at the proband status in the 49 trio carrier parents. 24.5% of probands with two carrier parents were identified as 'patient: 2 expanded pathogenic alleles', 49% were carriers of the pathogenic allele and 26.5% were identified as normal (figure 3.3.12A) In order to further check for caller errors, segregation was investigated in trios where the proband was identified as a carrier of the pathogenic expansion, which occurred in 642 trios. 85% of the carrier probands had at least 1 carrier parent, 7.5% had at least one potential carrier parent, 2.8% had a parent who was a patient with two expanded pathogenic alleles and 4.7% had parents identified as 'other' which comprised normal alleles. Individuals from this 'other' subgroup are believed to be cases where the *RFC1* expansion has occurred *de novo* (figure 3.3.12B).

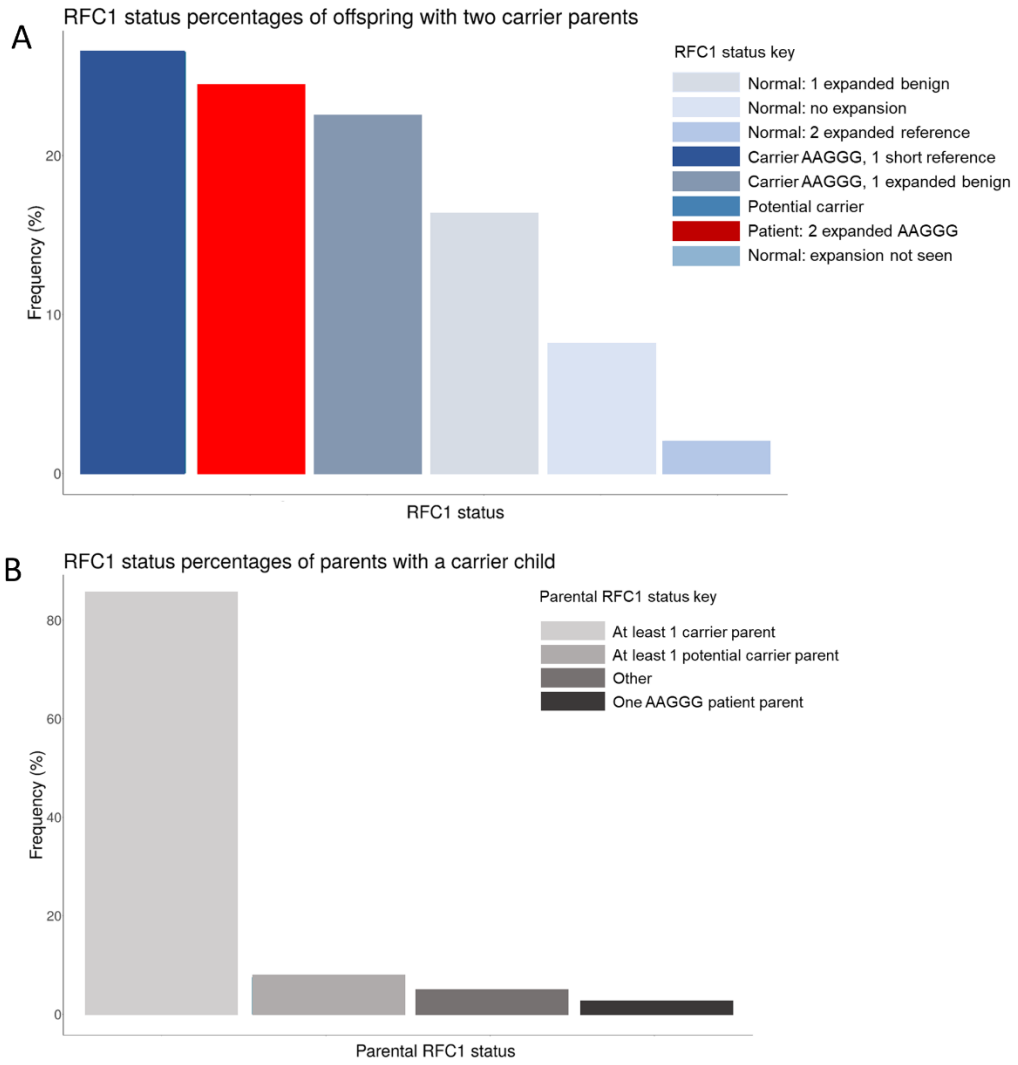


Figure 3.3.12: Segregation of the families within trio families. A, Bar plot showing percentage of *RFC1* statuses of probands that have two carrier parents, showing classical Mendelian recessive inheritance (n = 9657). B, Bar plot showing percentages of *RFC1* statuses of parents with a carrier proband child (n = 642).

3.3.4.7. Visual inspection of pile-up plots

In order to confirm the RFC1 caller results, pile-up plots were generated of the expanded region and visually inspected. This was to ensure that the RFC1 caller was correctly identifying atypical cases. Out of 145 cases identified as 'patient: 2 expanded pathogenic alleles', multiple samples did not have the expected phenotype associated with *RFC1*-associated disease. Pile-up plots were generated for these cases and visually inspected. Figure 3.3.13 shows two examples from NHNN, case clinically diagnosed with schwannomatosis (patient 1) and a plot for a case with myotonia congenita (patient 2). Both pile up plots showed the expanded $AAGGG_{(n)}$ spanning both alleles, consistent with the biallelic pathogenic *RFC1* repeat expansion.

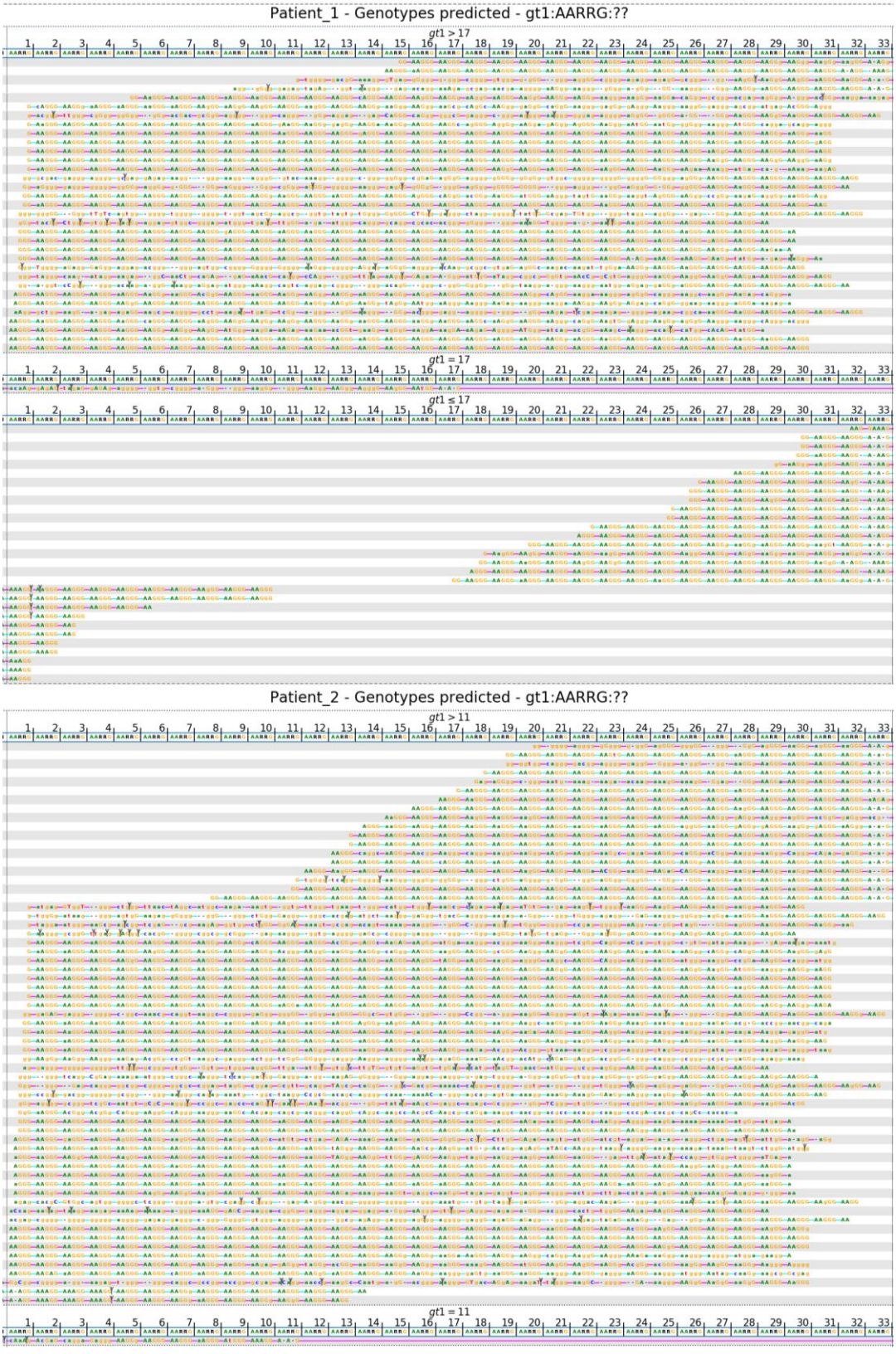


Figure 3.3.13: Visual inspection of cases with prior clinical diagnosis. Visual inspection of pile-up plot for patients with schwannamatosis (patient 1) and mytonia congenita (patient 2), both identified as 'Patient: two expanded pathogenic alleles' by *RFC1* Caller.

3.3.4.8. Southern blot confirmation of samples identified as patient: 2 expanded pathogenic alleles by the RFC1 caller.

Of the 145 samples identified as 'patient: 2 expanded pathogenic alleles', 59 were from the catchment area 'RP4', where access to DNA via the biobank at the NHNN is accessible. In order to confirm the presence of the biallelic repeat expansion, the samples were Southern blotted, as per the diagnostic screening workflow. 30 samples had successfully been confirmed via previous diagnostic screening of samples that fit the phenotypic criteria. We were unable to retrieve DNA for 20 samples. The remaining nine samples were Southern blotted. All nine patient samples showed visible expansions on Southern blot, with the mean expansion sized at 1341 (±727) repeating units and ranging from 237 to 1845 repeating units (figure 3.3.14, table 3.3.2).

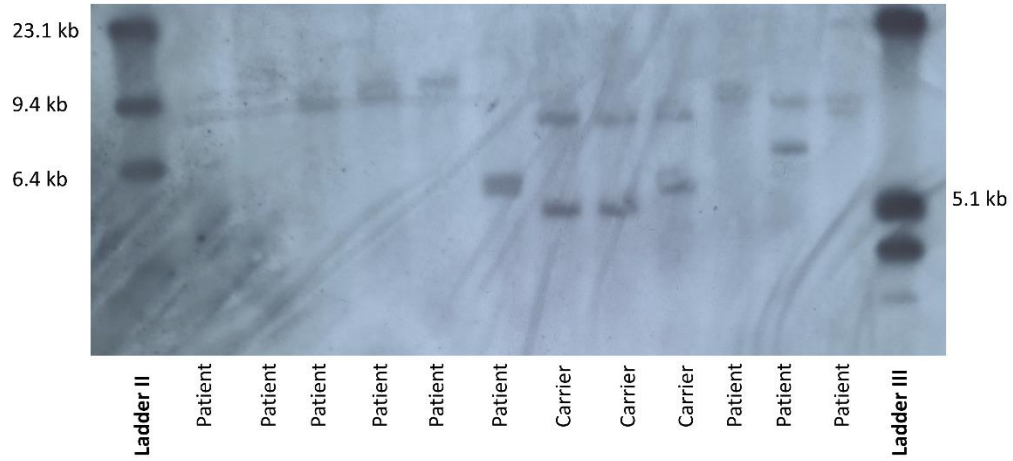


Figure 3.3.14: Southern blot of outstanding RP4 catchment area samples. Nine samples with sufficient available gDNA selected for Southern blotting that had not previously been blotted. Three suspected pathogenic carriers also included. Ladder II refer to DNA Molecular Weight Marker II (Roche) and ladder III to DNA Molecular Weight Marker III (Roche). *kb, kilobases*

Sample	Allele 1 (kb)	Allele 2 (kb)
1	900	1500
2	1000	1500
3	1100	1300
4	1300	1800
5	1300	1800
6	200	300
7	2700	3100
8	600	900
9	1200	1500

Table 3.3.2: Expansion sizes for Patients in Southern blot. Expansion sizes were estimated using a linear regression model that corresponded to DNA molecular weight marker ladders used. *kb, kilobases*

3.3.4.9. Sensitivity and specificity of RFC1 caller.

In order to test the sensitivity and specificity of the RFC1 caller, we used a cohort of 105 cases independently put forward by Dr Andrea Cortese for *RFC1* diagnostic screening, based on a suspected *RFC1*-associated clinical phenotype. We compared the RFC1 caller results to the confirmed Southern blotted results. I then calculated the sensitivity and specificity using 2 x 2 contingency table to find the True Positive rate (sensitivity) and True Negative rate (specificity) and a receiver operating characteristic (ROC) curve to assess its diagnostic ability. The 2 x 2 contingency table gave a sensitivity rate of 96.2% and sensitivity of 100%. The area under the curve (AUC) for the ROC curve was 0.962, with a *p* value of < 0.00000001 (figure 3.3.15). The singular RFC1 caller sample that did not correspond to the Southern blotting confirmed results was a carrier, that was picked up as 'normal: 1 expanded benign, 1 short ref'.

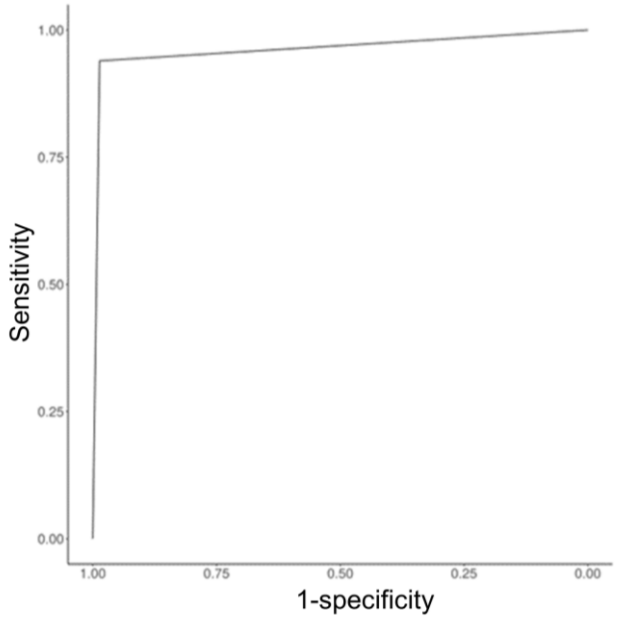


Figure 3.3.15: Sensitivity and specificity of RFC1 caller. ROC curve showing sensitivity of 96.2% and specificity of 100%, with an AUC of 0.962 (p = < 0.0000001).

3.3.4.10. Haplotype analysis of rare diseases cohort

Using eight SNPs present within the *RFC1* locus, rs2381378, rs6823497, rs6844176, rs17584703, r11096992, rs2066790, rs2062229 and rs2381873, haplotypes across the rare diseases' cohort were analysed in 8000 individuals. r11096992 and rs2066790 both are within a recombination hotspot and are the most informative SNPs. Individuals with a biallelic pathogenic repeat expansion were homozygous for rs11096992 (AA) and rs2066790 (AA), with normal non-expanded individuals having rs11096992 (GG) and rs2066790 (AA) (figure 3.3.16). Individuals that were carriers of the pathogenic repeat expansion with an expanded benign expansion on their other allele and individuals that have two expanded benign expansions were also homozygous for rs11096992 (AA), however this was not the case for rs2066790 (GG). This suggests that patients with the pathogenic $AAGGG_{(n)}$ have an (AA) (AA) combination for rs11096992 and rs2066790, and therefore a different haplotype to individuals with normal non-expanded alleles.

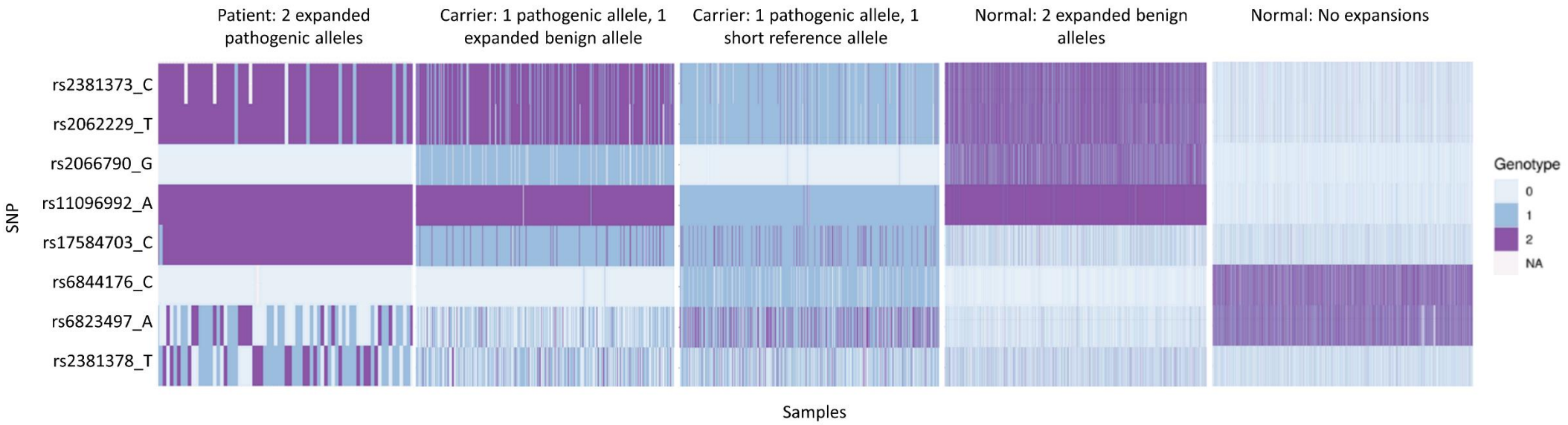


Figure 3.3.16: Heatmap of haplotypes. Heatmap showing haplotypes of 8000 individuals from rare diseases cohort, categorised by RFC1 caller into *RFC*1 status, and based on 8 SNPs; rs2381373 (C:T), rs2062229 (C:A), rs2066790 (A:G), rs11096992 (G:A), rs17584703 (T:C), rs6844176 (C:T), rs6823497 (C:A) and rs238138(T:C). Genotype key: 0 (pale blue) = absent on both alleles, 1 (medium blue) = present on one allele, 2 (purple) = present on both alleles, NA (light grey) no data available. *SNP, single nucleotide polymorphism; rs, reference SNP cluster ID; NA, not applicable.*

3.3.5. Discussion

A bioinformatic tool has been successfully developed that picks up the presence of a biallelic repeat expansion in *RFC1* with a sensitivity of 100% and specificity of 96.2%. Out of a total of 62, 550 samples within the rare diseases cohort in the 100kGP, 145 were picked up by the RFC1 caller as 'patient: 2 expanded pathogenic alleles`, identifying them as having the biallelic pathogenic $AAGGG_{(n)}$ repeat expansion.

The allelic frequency of 4% for the pathogenic $AAGGG_{(n)}$ expansion found within the rare diseases 100kGP population is within the range already published within the literature that ranges from 0.7 – 6.8% (50, 54, 166). Additionally, when looking at the hereditary ataxia and CMT patients as a separate cohort, the $AAGGG_{(n)}$ allelic frequency of 7.5% is higher than the 3.5% frequency found in a late-onset ataxia population that contained a proportion of hereditary ataxia and CMT patients, detailed in chapter 4. This could be due to the GEL cohort containing purely CMT and hereditary ataxia patients, compared to the late-onset ataxia cohort in chapter 2, which contained generalised late-onset ataxia samples.

Interestingly, 29.7% of the 145 samples identified as 'patient: 2 expanded pathogenic alleles' were relatives and not probands. Unfortunately, clinical information was not available for these samples, as they were recruited to check inheritance patterns in the specific normalised disease groups that their associated proband were recruited in. A potential explanation for the high frequency of the biallelic repeat expansion in relatives is that, depending on age, these individuals could still develop symptoms consistent with *RFC1*-associated disease. The most highly associated disorder with the biallelic *RFC1* repeat expansion CANVAS clinically presents with late-onset and slowly progressive imbalance with absent VVOR, combined with cerebellar dysfunction and bilateral vestibular dysfunction (54). In addition, late-onset ataxia, also associated with the *RFC1* repeat expansion, has a slower, later-onset progression (2). This could theoretically mean that a proportion of relatives identified with the expansion have not yet been diagnosed. The majority of relatives identified as being 'patient: 2 expanded pathogenic alleles' were recruited under paediatrics (46.2%), followed by neurology at 26.8%. Additionally, 74.4% of these relatives are either the father or mother of the proband in terms of biological relationship. It could be plausible that these individuals have not yet experienced the later-onset symptoms associated with the repeat expansion, especially if the proband at the time of recruitment was a child. It is also interesting that the biggest disease groups of the associated probands for the relatives identified as 'patient: 2 expanded pathogenic alleles' were intellectual disability and CMT disease, which are both neurological diseases.

This could represent an artifact of the 100kGP research environment, where focus is on the affected proband with a molecularly unsolved disease. The recruitment process is at the point of the proband and therefore limited information is available for relatives. Therefore, it is difficult to determine if these individuals have a clinical phenotype or not. Another limitation of the research environment is the

clinical annotations attached to each sample that are wide-ranging, often ambiguous and may not truly describe the disease that the proband has. This was exemplified by one family identified as 'patient: 2 expanded pathogenic alleles' who were categorised as having 'mitochondrial disease'. Upon inspection of clinical notes, these patients all had a neuropathy which is a feature associated with the *RFC1* repeat expansion

It could also be possible that the biallelic *RFC1* repeat expansion has a reduced penetrance, with the phenotypic trait not always expressed, despite the presence of the pathogenic $AAGGG_{(n)}$ repeat expansion on both alleles. To date, all literature suggests that the pathogenic repeat expansion is fully penetrant, so this would be an extremely novel finding (1, 50, 53). Low penetrance has been found in several other neurological repeat expansion disorders including SCA8, which has been reported to have incomplete penetrance, SCA12 that presents with reduced penetrance within the expansion range of 40 – 66 repeating units and *C9orf72* FTD/ALS, which has a low penetrance range of 25 repeating units or less (67). In addition, HD alleles with a CAG repeating length of 36-38 have a much-reduced penetrance and have been found to occur within the general population at a high frequency (334). Currently no reduced penetrance has been reported for the *RFC1* repeat expansion, and additionally no correlation has been observed between repeat length and age of onset or severity of disease (53).

33 of the probands identified as having the biallelic pathogenic repeat expansion were not recruited through 'Neurology' (22.8% of the 145 samples in total). This could therefore be suggestive of an expanding phenotypic spectrum associated with the disease. Indeed, it has been identified in clinically confirmed MSA patients, dopa-responsive parkinsonism, hereditary sensory neuropathy with cough, Sjögren syndrome, progressive supranuclear palsy and hyperkinetic movement disorders (50, 169, 252, 257, 258, 260). However, there have been no reports to date of any non-neurological disease implicated with the biallelic *RFC1* repeat expansion. Pile-up plots of these phenotypically atypical cases were generated for visual inspection, and all cases had convincing biallelic $AAGGG_{(n)}$ expansions. In addition, two cases from RP4 catchment area that presented with schwannomatosis and myotonia congenita were Southern blotted, confirming the presence of the biallelic $AAGGG_{(n)}$ expansion. These two diseases have not been associated with the *RFC1* repeat expansion until now. Additionally, intellectual disability and mitochondrial disorders, of which multiple samples present with from the 145 patients identified, have also not been associated with the expansion in *RFC1*.

Another possible theory of why a large proportion of individuals identified as 'patient: 2 expanded pathogenic alleles' do not potentially express the phenotype could be due to somatic mosaicism. Repeat expansion numbers have been shown to differ across different cells, contributing to differential phenotype expression and pathogenicity. Somatic expansion has been shown to be key to disease progression, with brain tissues experiencing higher levels, and therefore increased degeneration (335, 336). In the case of HD, somatic expansions are thought to occur from DNA damage, due to increased

levels of ROS, which can induce single and double stranded breaks. DNA mismatch repair genes have been shown, through GWAS, to be modifiers of age of onset of HD (337-340). SCA1 has also been shown to show genetic anticipation, along with HD, and DNA repair failure is also thought to be contributory, with high levels of repeat expansion instability in the cortex and neostriatum regions (341, 342). In the case of *RFC1* there could be differential expression of the *RFC1* repeat expansion in different tissues between individuals. Somatic expansion of the *C9ORF72* GGGGCC repeat expansion has been shown in different tissues of individuals of both ALS and FTD with germline repeat expansions (343, 344). Additionally in cases of sporadic early-onset Alzheimer's disease, somatic mosaicism was shown for expression of the presenilin 1 gene mutation, with degree of mosacism at the age of presentation at 8% in peripheral lymphocytes and 14% in cerebral cortical tissue (345). Blood is the only tissue collected for DNA extraction in GEL, and therefore is used to extract the DNA from all individuals picked up as 'patient: 2 expanded pathogenic alleles'. Somatic mosaicism, from epigenetic or other structural changes, could therefore cause relatives or phenotypically different patients from expressing the mutation in blood, but not in brain tissue. This would potentially alter the use of the RFC1 caller as a diagnostic predictive tool for the repeat expansion., The RFC1 caller detects the repeat expansion from WGS data, that is most commonly generated from blood. Therefore, a targeted approach would be necessary, whereby cases positive for the biallelic repeat expansion would require clinical evaluation and secondary diagnostic confirmation. Additional tissue is required to compare blood and brain tissue levels of *RFC1* biallelic repeat expansion, and due to the rare nature of this disease, brain tissue is a challenge to obtain. Investigating alterations at a epigenetic, transcriptomic, and proteomic level in *de novo* cases of the *RFC1* repeat expansion, would provide an opportunity to understand the functional consequences of genomic mosaicism.

To explore the possibility that the *RFC1* repeat expansion has a lower penetrance, based on the large numbers of relatives and non-neurological disease probands identified with the recessive pathogenic repeat expansion, large cohorts of healthy or non-neurological disease individuals are required. There is an upcoming release for the UK Biobank which will have WGS data available for healthy individuals, which would be an interesting cohort to screen using ExpansionHunter and the *RFC1* caller and could provide further insight into the penetrance and prevalence of the repeat expansion. Additionally, to further explore the possible involvement of genomic mosaicism, brain tissue is required to check for the *RFC1* biallelic repeat expansion to compare expansion size or expansion expression between blood and brain tissues. This has not yet been carried out but would be vital for investigating expression levels of the repeat expansion between tissue types, particularly in those patients who do not present with an associated phenotype.

Future work could benefit from performing nerve conduction studies on patients that do not have a typical phenotypic presentation, as sensory neuropathy, including idiopathic cases, is associated with the *RFC1* repeat expansion. All patients to date with *RFC1*-associated CANVAS have a sensory

neuropathy/neuronopathy, and 34% of sensory neuropathy patients were identified as having the biallelic repeat expansion in one study (259). Therefore, these non-neurological patients could have an undocumented sensory neuropathy, highlighting the need for further clinical examinations.

The RFC1 caller picked up 111 expanded motifs that appear on both alleles, that include the motifs already identified including $AAAAG_{(n)}$, $AAAGG_{(n)}$, $ACAGG_{(n)}$, $AAGAG_{(n)}$ and $AGAGG_{(n)}$, but also 106 novel motifs not yet published, of which 26 comprised more than 10% of the expansion (2, 170, 171). Interestingly, the $ACAGG_{(n)}$ motif was described as pathogenic in two affected Asian-Pacific families (171). 28 sample were found to be homozygous for the ACAGG motif, with 7 also being pathogenic carriers for $AAGGG_{(n)}$, two in particular had the motif sequence $AAGGG_{(n)}$ $ACAGG_{(n)}$ on one of their alleles. The high number of motifs identified emphasises the highly polymorphic locus associated with *RFC1*, and the motif distribution plot shows the heterogeneity within alleles, with multiple expanded motifs within the same allele.

Haplotype analysis using a heatmap representation of eight SNPs from the *RFC1* locus showed that the haplotype combination rs11096992 (AA) and rs2066790 (AA) is associated with the expanded pathogenic motif, which supports previous findings by Cortese et al (2). Interestingly, the data from 8000 individuals from the rare diseases cohort showed that expanded benign alleles are more strongly associated with rs2066790 (GG), which does not support the theory that expanded benign and expanded pathogenic alleles share the same haplotype which was suggested by Cortese et al (2). However, that cohort consisted of nine expanded benign ($AAAGG_{(n)}$) individuals, of which three had rs2066790 (AA), which is considerably smaller than the 8000 individuals included in this SNP analysis.

To check for errors with the RFC1 caller, segregation within the carrier trios was analysed, with 9657 complete sets of trios present within the rare diseases' cohort. Looking at trios where both parents were carriers of the pathogenic repeat expansion revealed that 24.5% of probands in these trios were patients with the biallelic pathogenic repeat and 49% of probands are carriers, which fits with the suggested autosomal Mendelian recessive inheritance for the *RFC1* repeat expansion (346). Mendelian recessive inheritance patterns of rare single gene mutations dictate that there is a 25% chance of two carriers producing an affected child, a 50 % chance of a carrier child and 25% chance of the child not inheriting the recessive gene at all. Looking at segregation in all trios where the proband had a carrier status also helped identify if there were any errors with the RFC1 caller. 85% of carrier probands had at least one carrier parent, 7.5% had at least one potential carrier parent, 2.8% had one patient parent (with the biallelic expanded pathogenic allele) and 4.7% of samples had two parents with a 'normal' *RFC1* status. This 4.7% of carrier probands represents the presumed proportion of samples where a repeat expansion in *RFC1* has occurred *de novo*. In HD, the *de novo* frequency is 10%, which is higher than the percentage of presumed *de novo* cases for *RFC1* (66)

The RFC1 caller has the potential to be used in a diagnostic setting, identifying patients with the biallelic pathogenic repeat expansion from WGS data. In order to it to be viable for diagnostic use it must have a high specificity and sensitivity. To assess this, a cohort of 105 independently recruited individuals that were put forward for *RFC1* diagnostic screening, that involves flanking PCR, RPPCR and Southern blotting, and were also part of GEL were analysed. The results from their diagnostic screening were compared to the RFC1 caller results, which showed that the caller has a sensitivity of 100% and specificity of 96.2%. A limitation of the study was that the cohort size was relatively modest, therefore future work would benefit from a larger cohort, including samples from other recruitment centres, to further confirm these findings.

ExpansionHunter has been reported as effective at detecting short repeat expansions, with a 97.3% sensitivity and 99.6% specificity, which the RFC1 Caller builds upon, detecting longer expansion lengths (330). Within diagnostic settings, predictive modelling tools are already being used, such as the PICADAR, used to detect primary ciliary dyskinesia, which has a sensitivity and specificity of 0.90 and 0.87 respectively (347). This suggests that the RFC1 caller has the sensitivity and specificity required to effectively and accurately be used in a diagnostic setting.

The one sample in question that was a false negative for the repeat expansion was a carrier of the pathogenic expansion, that was called as normal, with an expanded benign allele. In terms of identifying those individuals with the pathogenic $AAGGG_{(n)}$ expansion on both alleles, the RFC1 caller has a specificity of 100%. Therefore, the caller has the capability of picking up all patients with the biallelic *RFC1* repeat expansion from large WGS datasets. The carrier not successfully picked up has a complex motif configuration, with $AAGGG_{(n)}$ sequence motifs interspersed between benign $AAAGG_{(n)}$ motifs. The locus is highly polymorphic and heterogeneous and the presence of motif interruptions within the expansion have already been documented in chapter 4 and within the literature (2). Therefore, the capabilities of the RFC1 caller to correctly identify carriers of the pathogenic allele may be currently insufficient to reliably be used for genetic counselling in a clinical setting, requiring further refinement.

3.3.6. Conclusion

In conclusion, this work has developed a new bioinformatic tool that has the ability to detect biallelic pathogenic repeat expansions in *RFC1*, using VCF and BAM file outputs from ExpansionHunter. The tool has a sensitivity of 96.2% and specificity of 100%, demonstrating its accuracy at picking up biallelic pathogenic expansions. This *RFC1* caller utilises WGS data, which enables existing datasets to be screened for the expansion, potentially solving cases with no current molecular diagnosis. Additionally, it opens the possibility of a more scaled down workflow for diagnosing *RFC1* repeat expansions, which currently is a time-inefficient process that requires a large amount of gDNA. Interestingly these results suggest that there might be cases of reduced penetrance or involvement of somatic mosaicism, owing to the large number of relatives and non-neurological diseases implicated.

Further work needs to occur to investigate this with analysis of cases found in additional recruitment centres being a clear validatory next step. Validating individuals picked up as ‘patient: 2 expanded pathogenic alleles’ would help solve these cases of rare molecularly unsolved disease, which could be reported back to the respective recruitment centre. Additionally comparative studies of the presence, size, and functional expression of the *RFC1* biallelic repeat expansion between blood and brain tissues is required to explore levels of somatic mosaicism. It also might suggest that there is an expanded phenotypic spectrum associated with the repeat expansion that may include intellectual disability and epilepsy.

3.4. Acyl-coA binding domain containing 5 (ACBD5) and autosomal recessive ataxia

3.4.1. Introduction

Acyl-coA binding domain containing 5 (*ACBD5*) (MIM:616618) encodes a member of the acyl-coenzyme A binding protein family, ACBD5 which is a peroxisomal membrane protein. Peroxisomes are key metabolic organelles, first discovered in rodent kidney cells that have important roles in fundamental cell processes, including lipid and ROS metabolism, cell signalling and innate immunity. Their dysfunction or absence underlie multiple peroxisomal disorders caused by peroxisomal biogenesis and function. They are present in virtually all eukaryotic cells and been identified in microglia, astrocytes, oligodendrocytes within the central nervous system (CNS) (348, 349) and within Schwann cells within the peripheral nervous system (PNS) (350). Furthermore, they have also been identified in the myelinating fibres of white and grey matter, predominantly within myelin and rarely in axons (350). Their involvement in neurological development can be seen by the plethora of diseases that stem from peroxisomal dysfunction, including Zellweger Spectrum Disorder (ZSD).

These peroxisomal disorders can be further characterised into peroxisomal biogenesis disorders (PBDs), such as ZSD, and peroxisomal enzyme deficiencies (PEDs). PEDs define a group of disorders where wither a peroxisomal membrane protein or peroxisomal matrix enzyme is absent. ZSD, a PBD is characterised by four phenotypes; Zellweger syndrome (ZS), neonatal adrenoleukodystrophy (NALD), infantile Refsum disease (IRD) and Heimler syndrome, which was recently recognised as a ZSD, is a heterogenous group of autosomal recessive disorders. Peroxisomes are postulated to form via division of pre-existing peroxisomes which involves the biogenesis factors peroxins, encoded by *PEX* genes. ZSD is caused by mutations in *PEX* genes, thereby causing multiple metabolic pathway impairments, with the main features including developmental delay, adrenocortical dysfunction, hearing and vision impairment and liver dysfunction (351).

ZS presents with a range of clinical features including developmental defects, neonatal seizures, psychomotor delay, renal cysts, retinopathy, adrenal insufficiency, cataracts, impaired hearing, skeletal abnormalities, and dysmorphia (352). Neurologically, there are several pronounced abnormalities affecting the white matter that include hypomyelination, demyelination or dysmyelination, as well as neuronal migration abnormalities affecting the cerebral hemispheres and cerebellum (352). ZSD features mitochondrial dysfunction, with mitochondrial morphology changes, including distorted cristae and matrix appearances, and a reduction in mitochondrial respiration reported (353, 354). Mouse models of PEX protein deficiencies have shown mitochondrial complex 1 activity reduction, elevated levels of ROS, and increased neuronal cell death (355, 356). Peroxisomal and mitochondrial cross-talk has also been proposed to be affected in cases of ZSD, with peroxisomal dysfunction potentially causing altered redox control and mitochondrial metabolism (352). Most recently it has been shown that

peroxins are mislocalised to the mitochondria in cases of ZSD, causing an accumulation and subsequent morphological mitochondrial alterations. This phenotype has been shown to be rescued by overexpression of mammalian ortholog ATAD1, an AAA+ ATPase protein family member, in ZSD primary patient fibroblast lines (357).

The PEDs include X-linked adrenoleukodystrophy (ALD), a severe neurodegenerative disorder characterised by adrenal insufficiency, accumulation of very-long-chain-fatty-acids (VLCFA) and progressive demyelination within the central and peripheral nervous system. It is caused by mutations in ATP binding cassette subfamily D member 1 (*ABCD1)*, which codes for ALD, a peroxisomal transmembrane protein. ALD is thought to facilitate the import of VLCFA into the peroxisome, where they are subsequently degraded (358).

Peroxisomes have a key role in lipid metabolism, with ether lipid synthesis beginning inside peroxisomes with esterification of dihydroxyacetone phosphate (DHAP) with a long-chain fatty acid via the enzyme DHAP acyltransferase (DHAPAT). The fatty acid is subsequently exchanged by a fatty alcohol, forming alkyl-DHAP via alkyl-glycerone phosphate synthase (AGPS). Additionally, peroxisomes are the exclusive site of β-oxidation of very-long-chain fatty acids (VLCFA).

3.4.1.1. Peroxisomal biogenesis

The activity of 16 PEX proteins or peroxins encoded by *PEX* genes facilitate the process of peroxisomal biogenesis. Matrix protein import involves the production of matrix enzymes which contain a C-terminal peroxisomal targeting signal 1 (PTS1), which are synthesised on cytosolic ribosomes. Matrix enzyme cis acting targeting signals are recognised by cytosolic enzymes PEX5 and PEX7, which form a cytosolic complex that is then transported to the peroxisomal membrane. PEX13 and PEX14 bind to the complex which docks to the peroxisomal membrane. The interaction with PEX14 and PEX5 facilitates the formation of a large and dynamic membrane pore, which facilitates translocation of the cargo proteins and shuttling of the receptors through the matrix (359). Folded, cofactor bound, oligomeric proteins are transported through this transient peroxisomal pore. PEX2, PEX10, PEX12 form a complex which is also involved with the translocation of ligands through the peroxisomal matrix. PEX5 is polyubiquitinated by PEX10 and PEX12 on a lysine residue, which targets it for 26S proteome degradation (360). Additional rounds of import are enabled by recycling of the PTS1-receptor, PEX7 and PEX5, by the PEX1, PEX6 and PEX26 complex (361).

ATP is required for many interactions within peroxisomal biogenesis. PEX1 and PEX6 contain one or two ATPases associated with various cellular Activities (AAA) cassettes which govern ATP-hydrolysis and ATP-binding by forming hexameric rings that alter shape during the ATPase cycle that provides the force required for their cellular function (362, 363). PEX5 export is ATP driven and ATP-binding is required for PEX1 and PEX6 interaction. ATP hydrolysis enables the conformational alteration to

the PEX1 and PEX6 complex to remove monoubiquinated PEX5 out of the peroxisomal membrane (364).

New peroxisomal membrane formation is mediated by PEX3, PEX16 and PEX19. The endoplasmic reticulum (ER) provides membrane material and PEX3, PEX16 and PEX19 also enable membrane localisation. PEX11β and dynamin-like proteins (DLPs) are required for peroxisomal membrane fission. DLPs are shared between peroxisomes and mitochondria (365). Membrane elongation requires PEX11β, which binds to Fis1, resulting in DLP recruitment. Mechanisms such as autophagy and pexophagy contribute to peroxisomal turnover, which is estimated to be around 30% per day (366).

3.4.1.2. Peroxisomal pathways

Peroxisomal fatty acid beta-oxidation

A large variety of fatty acids can be beta-oxidised in peroxisomes, including saturated, hydroxylated, dicarboxylic, mono- and poly-unsaturated and 2- and 3- methyl branched-chain fatty acids (367) However, peroxisomes are also exclusively able to β-oxidise some other fatty acid substrates including; VLCFAs, particularly C22:0, C24:0 and C26:0, di-and trihydroxycholestanoic acid, pristanic acid (2 ,6, 10, 14,-tetramethylpentadecanoic acid), long-chain dicarboxylic acids, 2-hydroxy fatty acids, and a variety of prostanoids including prostaglandin F2 (PGF2-α), 8-iso-FGF2-alpha, thromboxane-B2, monohydroxy-eicosatrienoic acid (12-HETE, 15-HETE), cysteinyl leoukotriene LTE4 and leukotriene-B4 (368).

Peroxisomes perform one or more cycles of β-oxidation, with products being released into the cytosol to undergo further metabolism. However, the number of cycles of β-oxidation needed for some fatty acids such as VLCFAs C22:0, C24:0 and C26:0 is not yet fully established; C8-or C6-medium-chain acyl-coA are known to be generated. Two different acyl-coA oxidases, bifunctional proteins and two thiolases are all catalases for peroxisomal β-oxidation. Acyl-coA oxidase-1 (ACOX1) is thought to primarily catalase saturated acyl-coAs such as C22:0-coA, C24:0-CoA and C2:0-CoA, whereas acyl-CoA oxidase 2 (ACOX2) is involved with pristanoyl-CoA and di- and trihydroxycholestanoyL-CoA. Deficiencies in ACOX1 have been reported with associated VLCFA accumulation (369), however no deficiencies in ACOX2 have yet been described (367). Multifunctional proteins 1- and 2 (MFP1 and MFP2), multifunctional enzymes-1 and -2 (MFE-1 and MFE-2), and the L- and D- bifunctional proteins (LBP and DBP) also catalyse peroxisomal β-oxidation, with DPB specifically being involved in VLCFA β-oxidation as seen in DBP-deficient patients (370, 371). Sterol carrier protein X (SCPx) and 3-ketoacyl-coA thiolase are also involved in the β-oxidation of various fatty acids (367). Auxiliary enzymes are required for mono- and poly-unsaturated fatty acids, and these include different di-enoyl-coA reductases and isomerases (372, 373).

Fatty-acid α-oxidation

Fatty acid α-oxidation is localised to peroxisomes and is the main mechanism of oxidation for some fatty-acids, such as phytanic acid and other fatty acids with a methyl group at the 3-position. Refsum patients, which are α-oxidation pathway deficient have an accumulation in phytanic acid (368). The pathway occurs over 5 steps; activation to an acyl-coA, hydroxylation of the acyl-coA to a 2-hydroyacyl-CoA, cleavage of the 2-hydroxyacyl-CoA, oxidation of the resultant (n-1) aldehyde, oxidation of the aldehyde to the relevant acid and formation of an acyl-CoA ester (368).

Etherphospholipid biosynthesis

The first stage of etherphospholipid biosynthesis pathway occurs solely in the peroxisome, therefore several key components of etherphospholipid synthesis are peroxisomal. Alkylihydroxyacetone phosphate synthase (ADHAPS) and its two required substrates, long-chain alcohol and acyldihydroxyacetone phosphate (acyl-DHAP), are synthesised in the peroxisome by the enzymes NADPH oxidoredutase and dihydroxyacetone phosphate acyltransferase (DHAPAT), bound on the inner face of the peroxisomal membrane in a complex (374). The subsequent steps of etherphospholipid biosynthesis take place in the ER membrane, with ER catalysis enzymes(375).

Due to the functions of peroxisomes, interaction and exchange of metabolites and signalling with the ER, mitochondria, lysosomes, and lipid droplets are necessary, requiring a dynamic spatial organisation (376-378). Peroxisomes interact with the ER, during the biosynthesis of etherphospholipids and polyunsaturated fatty acids, as well as during peroxisomal biogenesis through the delivery of phospholipids from the ER (379). Interaction between the peroxisome and ER are thought to be physical, with ultrastructural studies showing peroxisomes localised opposite to ER tubules (380) and individual peroxisomes and ER tubules sharing short electron-dense cross bridges (381). The structure and formation of these cross-bridges is unknown in eukaryotes, however in yeast, the ER-peroxisomal contact has been found to involve Pex3 and Inp1p (382).

Hydrogen Peroxide Degradation

Peroxisomes contain enzymes that produce hydrogen peroxide (H_2O_2), nitric oxide ($NO\cdot$) and superoxide ($O_2\cdot^-$), which all readily react to form ROS and reactive nitrogen species (RGN) which can mediate gene expression, cell proliferation, programmed cell death, cell senescence and angiogenesis (383). However, an imbalance in ROS or RGN homeostasis is a known causative of neurodegeneration (384). Many of the metabolic pathways that occur in peroxisomes generate specific ROS or RGN by-products and so any impairment in one of several of these pathways could cause an imbalance in ROS, causing oxidative stress. Neurogenerative diseases that feature impaired neuronal migration, demyelination and axonal degeneration have been shown to likely involve oxidative stress (385). Interestingly a mouse model of X-linked ALD, whereby antioxidant treatment was applied, was reported to have experienced a halt and possible reversal in axonal damage (386).

Peroxisomal ROS production

Peroxisomal respiration produces H_2O_2, and up to 35% of total mammalian H_2O_2 is generated within peroxisomes (387). Peroxisomes are also involved with other types of ROS including superoxide and nitric oxide radicals (385). In order to control the physiological state of oxidative stress, peroxisomes contain abundant levels of the antioxidant enzyme catalase which reduces H_2O_2 and produces H_2O; as well as other antioxidant enzymes that target superoxide and nitric oxide radicals (388). Peroxisomal-mediated ROS production has been linked to mitochondrial integrity, including increased levels of mitochondrial fragmentation, as well as perturbation of the mitochondrial redox balance (389, 390).

Pexophagy

Selective autophagic degradation of peroxisomes, pexophagy, aids the maintenance of peroxisomal homeostasis and accounts for 80% of peroxisomal degradation (391). There are three known pathways of pexophagy; ubiquitin mediated, NBR1-induced and Pex3-induced pexophagy. The former is induced by ubiquitin-fused construction of PMPs, such as PMP34, and Pex3-ubiqutin.

3.4.1.3. ACBD5 and peroxisomes

ACBD5 is a member of the acyl-CoA binding domain-containing protein (ACBD) family, which are characterised by an acyl-coA binding domain at their N terminus (figure 3.4.1). ACBD5 has been found to be exclusively localised to peroxisomes in quantitative mass spectrometry studies (392, 393). This advocates that ACBD5 binds to acyl-CoAs, exerting its function in a peroxisomal-dependent metabolic pathway, suggesting that ACBD5 dysfunction could cause peroxisomal disease.

Figure 3.4.1: Schematic representation of ACBD5 protein domains.

ACB, acyl co-enzyme A binding domain; CC, coiled-coil region; FFAT, FFAT-like motif; TMD, transmembrane domain/anchor signalling domain.

A homozygous truncating mutation in *ACBD5* has been reported in a family with severe neurological impairment and retinal dystrophy (OMIM: 616618) that show features associated with peroxisomal disorders such as spastic paraparesis and white matter disease (394). Another novel homozygous deleterious c.626-689_937_234delins936+1075_c936+1230inv mutation in *ACBD5* was recently described in a patient with progressive leukodystrophy, ataxia and retinal dystrophy (145). Interestingly the patient was found to have impaired VLCFA metabolism, with an accumulation in VLCFAs in both

blood and fibroblast profiles. It has been postulated that ACBD5 binds to C26-CoA in the cytosol through its acyl-CoA binding domain and passes it onto ABCD1 which is a VLCFA transporter for β-oxidation within the peroxisome. Patients with X-ALD are ABCD1-deficient and have been found to have similar reduced β-oxidation activity of C26:0 to ACBD5-deficient cells, lending support to this theory (145). VLCFA accumulation is not an exclusive biochemical impairment to either X-ALD or ACBD5-deficiency and features in other peroxisomal disorders including ZSD, ACOX1 deficiency and DBP deficiency.

3.4.2. Aims of study

Through WES, Dr David Lynch from the NHNN identified a novel homozygous variant in *ACBD5* in a family with progressive recessive cerebellar ataxia. Using a range of biochemical and molecular techniques including cDNA sequencing, Western blotting, confocal microscopy, immunofluorescent staining, mitochondrial function testing, and RNA sequencing I have aimed to explore the functional effects that may explain the pathological mechanism. Whilst performing our preliminary studies and through the GeneMatcher portal, I have established a global collaboration with Professor Nancy Braverman from the McGill Institute, Canada, Dr Irman Forghani from Miami Miller School of Medicine, USA, Dr Adeline Verdeverer from the Children's Hospital of Philadelphia, USA, and Dr Luisa Averdunk from Heinrich-Heine-University, Germany, collecting several patients with ACBD5-deficiency. Our additional collective aim was to further characterise the phenotype, collaborating with Dr Sacha Ferdinandusse, Professor Michael Shrader's research group from Exeter University and Professor Andrey Abramov's research group from UCL.

3.4.3.Summary of methods

Study aim	Methods used	Contributors	Methods section
Genetic analysis	Whole-exome-sequencing	Dr David Lynch	2.3.11
	DNA extraction	Roisin Sullivan	2.2.5
	Flanking PCR	Roisin Sullivan	2.2.8
	Sanger sequencing	Roisin Sullivan	2.2.7
	Reverse transcription	Roisin Sullivan	
Investigation into loss-of-function mutational effect	Cell culturing	Roisin Sullivan	2.3.1
	Immunofluorescent microscopy	Roisin Sullivan	2.3.9
	Whole fibroblast cell protein lysate generation	Roisin Sullivan	2.3.4
	Bradford assay/BCA assay	Roisin Sullivan	2.3.5, 2.3.6
	Polyacrylamide gel electrophoresis and Western blotting	Roisin Sullivan	2.3.7
Metabolic screening	Cell culturing	Roisin Sullivan	2.3.1
	Metabolic screening assay	Dr Sacha Ferdinandusse	2.3.3
Mitochondrial investigations	Cell culturing	Roisin Sullivan	2.3.1
	Measurement of mitochondrial membrane by TMRM	Roisin Sullivan	2.3.10
	Measurement of NADH redox	Roisin Sullivan	2.3.11
Oxidative stress and lipid peroxidation investigations	Cell culturing	Roisin Sullivan	2.3.1
	Measurement of ROS production via Glutathione concentration	Roisin Sullivan	2.3.12
	Measurement of lipid peroxidation	Roisin Sullivan	2.3.13
Peroxisomal quantification	Cell culturing	Mrs Tina A Schrader, Miss Rechal Kumar	2.3.14

	Peroxisomal number and biogenesis analysis	Professor Michael Schrader, Dr Christian Hacker	2.3.14
Quantification of Peroxisomal-ER contacts	Cell culturing Transmission electron microscopy and quantification	Mrs Tina A Schrader, Miss Rechal Kumar Professor Michael Schrader, Dr Christian Hacker	2.3.15 2.3.15
Patient clinical examinations	Patients 1 and 2 Patient 3 Patient 4 Patient 5 Patient 6	Dr Emer O'Connor, Professor Henry Houlden Professor Adeline Vanderver Dr Imran Forghani Professor Nancy Braverman, Dr Sarah Alsubhi Dr Luisa Averdunk	2.1.6
RNA sequencing	RNA sequencing	UCL Genomics Facility	2.2.10
RNA sequencing analysis	Kallisto processing Sleuth programme used in R statistical environment Pathway analysis using STRING Gene set enrichment analysis Over-representation analysis	Dr Jana Vandrovcova Roisin Sullivan Roisin Sullivan Roisin Sullivan Roisin Sullivan	2.4.5 2.4.5 2.4.1 2.4.2 2.4.3
RNA mutational effect prediction	Mutation Taster	Roisin Sullivan	2.4.1

3.4.4. Results

3.4.4.1. Genetic Analysis

WES carried out by Dr David Lynch identified a homozygous mutation in exon 7 of gene *ACBD5* which codes for acyl-coA binding domain containing 5 (ACBD5) (MIM:616618). Sanger sequencing (figure 3.4.2) confirmed a AATG deletion in exon 7 of ACBD5 (ACBD5:ENST00000396271.3:exon7:c.726_729del:p.242_243del) in both patients, whereas the control and unaffected sister had a normal genomic sequence. Both parents had a normal sequence until the mutational position, whereby two alleles then became apparent, due to their carrier status. To check if nonsense mediated decay (NMD) was occurring, RNA was extracted from patient, carrier and control line fibroblasts and cDNA was synthesised and subsequently sanger sequenced to view the mRNA transcript (figure 3.4.2 and figure 3.4.3). Results showed evidence of NMD, with the carrier having the same mRNA transcript as the control and unaffected sister, despite the Sanger sequencing on cDNA showing both the mutated and normal allelic sequences. Both patients had no detectable results after cDNA sequencing, suggesting a loss-of-function of ACBD5, due to NMD.

Patients 3, 4, 5 and 6 were all identified using WES by our collaborators across the globe. Patient 3 has a homozygous deletion of exons 7-8 caused by a large intragenic deletion (c.626-689_937-234delins936+1075_c.936+1230inv; p.W489X), patient 4 has a homozygous missense mutation changing Tryptophan to a stop codon (c.1467 G>A; p.W489X), patient 5 has a homozygous missense mutation changing a lysine to a stop codon (c.1105A>T; p.K369X) and patient 6 has a compound heterozygous mutation with an insertion (c.1210dup) on one allele and a deletion on the other (c.1019_1020del) (table 3.4.1).

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
Method of variant detection	WES	WES	WES	WES	WES	WES
Gene	ACBD5	ACBD5	ACBD5	ACBD5	ACBD5	ACBD5
Transcript	NM_145698.5	NM_145698.5	NM_145698.4	NM_145698.5	NM_145698.5	NM_145698.3
Variant 1						
cDNA	c.726-729del:p.242-243del.	c.726-729del:p.242-243del.	c.626-689_937-234delins936+1075_c.936+1230inv	c.1467 G>A	c.1105A>T	c.1210dup
Coding Effect	p.I242Mfs*6.	p.I242Mfs*6.	p.D208Vfs*30	p.W489X	p.K369X	p.R404Kfs*19
Predictions	Frameshift	Frameshift	Frameshift	Nonsense mutation	Nonsense mutation	Frameshift
Exon	Exon 7	Exon 7	Exon 6	Exon 12/13	Exon 9	Exon 10
Variant 2						
cDNA	*NA*	*NA*	*NA*	*NA*	*NA*	c.1019_1020del
Coding Effect	*NA*	*NA*	*NA*	*NA*	*NA*	p.Gly340Valfs*3
Predictions	*NA*	*NA*	*NA*	*NA*	*NA*	Frameshift
Inheritance						
Inheritance mode	Homozygous Recessive	Homozygous Recessive	Homozygous Recessive	Homozygous Recessive	Homozygous Recessive	Compound heterozygous
Status Parents	Unaffected	Unaffected	Unaffected	Unaffected	Unaffected	Unaffected
Consanguinity	Y	Y	Y	Y	Unknown	N

Table 3.4.1: Genetic information of ACBD5-deficient patients

WES, whole exome sequencing, ACBD5, acyl-coenzyme A binding protein domain 5; Y, yes; N, no; NA, not applicable; del, deletion; dup, duplication; inv, inversion; c, coding; p, protein; mb; megabases; I, isoleucine; M, methionine; G, glycine; A, alanine; D, aspartic-acid; K, lysine; W, tryptophan

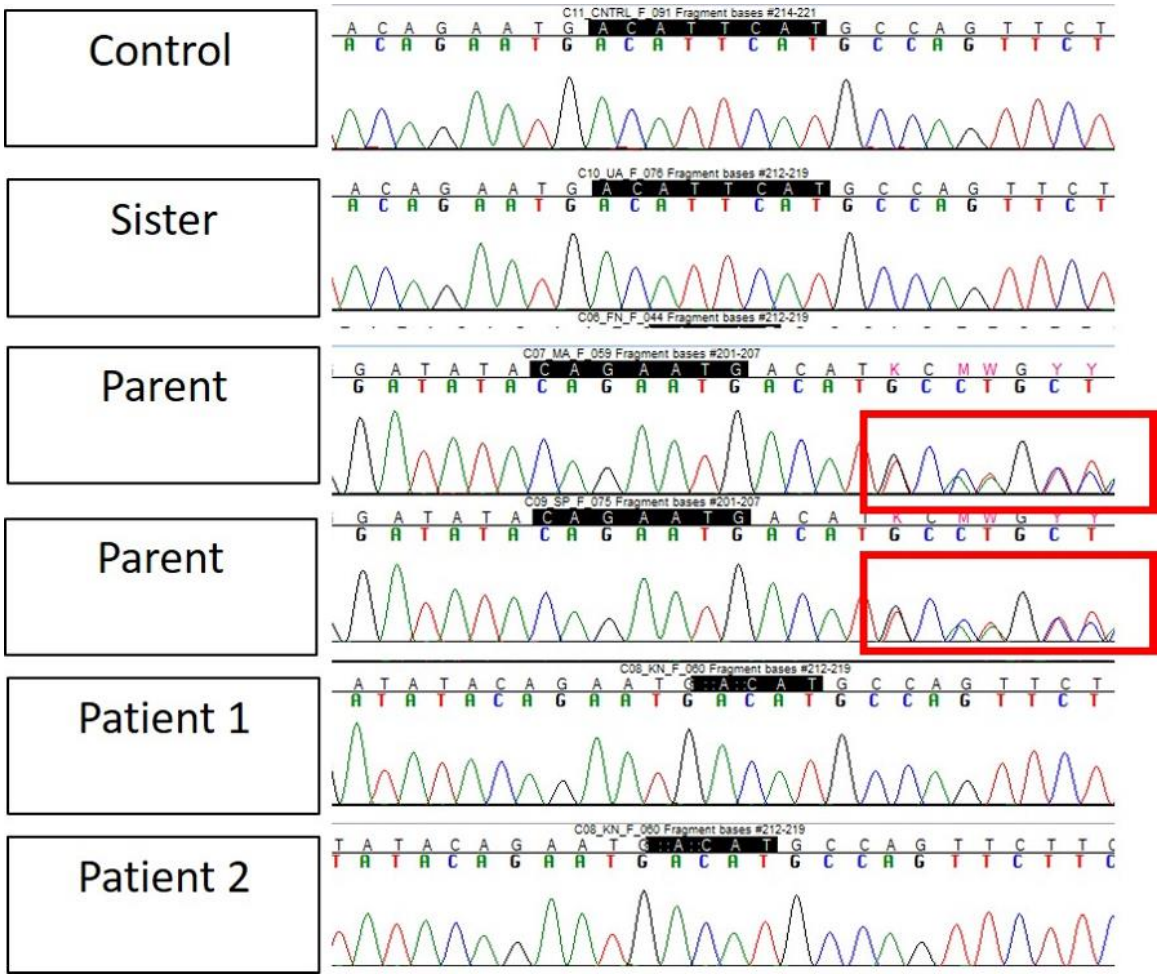


Figure 3.4.2: Sanger sequencing to identify ACBD5 in family 1. The control and unaffected sister fibroblast lines have the normal reference sequence for *ACBD5*. Both parent fibroblast lines have a normal *ACBD5* transcript, until the AATG sequence whereby the normal and abnormal allele can be seen (red box), characteristic of a carrier. Both patients have an AATG deletion in the *ACBD5* sequence. *ACBD5*, *acyl-coenzyme A binding protein domain 5*

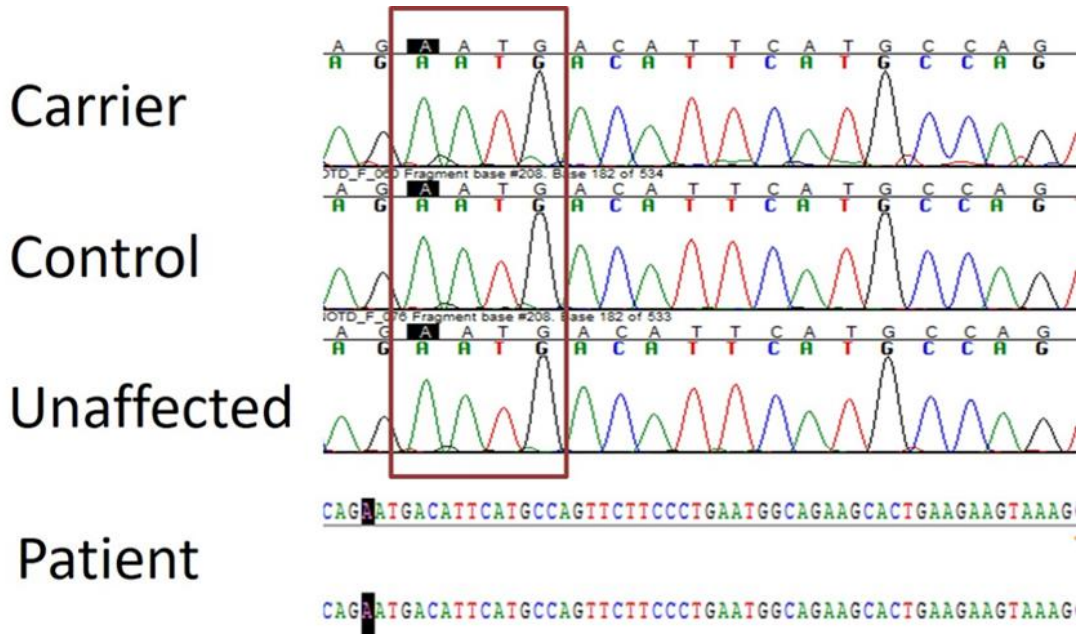


Figure 3.4.3: NMD of ACBD5 transcript shown via cDNA sanger sequencing for family 1. The carrier fibroblast line has the same *ACBD5* mRNA transcript as the control and unaffected sister fibroblast lines, therefore showing evidence of NMD, whereby only the normal *ACBD5* transcript is undergoing translation. Both patients had no detectable *ACBD5* transcript due to both alleles undergoing NMD. *ACBD5, acyl-coenzyme A binding protein domain 5; NMD, nonsense-mediated-decay*

3.4.4.2. Clinical presentation of patients

Two initial patients (family 1) from the NHNN were found to have homozygous mutations in *ACBD5* from WES. Four additional patients were recruited to the study via Genematcher and three additional collaborations. Case studies 1 – 5 depict the clinical presentation of the patients. There is limited clinical information for patient 5.

3.4.4.3. Case studies of patients

1. *Family 1: Patients 1 and 2*

Two brothers, patient 1 and patient 2, with consanguineous parents experienced onset of symptoms around their late teens/ early twenties. They are both very similar in clinical presentation. Patient 1 and patient 2 have a jerky broad-based ataxic gait, nystagmus on upgaze, moderate limb ataxia and dysarthria. Patient 1 also has a history of epilepsy. Both patients have reported mild cognitive impairment but no evidence of neuropathy. Brain MRI showed cerebellar atrophy (table 3.4.2, figure 3.4.5).

2. *Family 2: Patient 3*

The patient, a girl, presented with abnormal eye movement and was diagnosed with retinal rod cone dystrophy at aged 7 months. She had delayed gains of motor skills and by two years old had

progressively abnormal gait. By aged four she had limited vocabulary and dysarthria. She developed progressive microcephaly, along with facial dimorphisms that included rotatory nystagmus, tubular nose, prominent ears, bilateral ptosis and hypotelorism. By aged nine, she could not walk without two-handed assistance, had a reduced walking endurance and great difficulty descending stairs. Brain MRI showed hypomyelination, with diffuse T2 signal abnormality within the deep white matter. The long tracts of the brainstem also showed signal abnormality (table 3.4.2, figure 3.4.5) (145).

3. *Family 3: Patient 4*

The female patient, now 36 years old, did not have medical issues up to her first birthday, however progressive deterioration of visual perception and nystagmus began to present at the age of three years old. She presented with progressive lower extremity weakness and developmental delay. She underwent a 'transfer factor vaccine' to combat her eye issues, which was thought to have been responsible for her neurological symptoms. She experienced rapid cognitive decline between the ages of eleven and thirteen, with rapid motor and speech regression. She was, and remains, wheelchair bound by the age of thirteen. She presented with upper extremity weakness and impaired fine motor movement by the age of fifteen. She has confirmed severe retinal rod-cone dystrophy. At age twenty-nine, she received an intrathecal autologous stem cell injection, which led to partial improvement in gross motor ability, speech, and vision; however, this was transient. She currently presents with dysarthria, spastic paraparesis and impaired cognitive ability. Brain MRI shows infratentorial and supratentorial atrophy with diffuse white matter signal malformations. (table 3.4.2, figure 3.4.5) (146)

4. *Family 5: Patient 6*

The male patient, now aged twenty-two and a half years old, had a normal development until the age of five, and presented with unbalanced, ataxic gait by the age of seven years. He has been wheelchair bound since the age of fifteen and can only walk with the support of an Arjohuntleigh (ARJO) walker with a physical therapist. The patient presents with mild cognitive impairment. He has dysarthria, ataxia, dysphagia, and nystagmus. He does not have retinal dystrophy, has a normal ocular fundus, but does have myopia magna and astigmatism. He has spasticity and parkinsonism. The patient also has the inability to control urination, with intermittent pain of his legs, which affect him more severely at night, and recurrent headaches. Brain MRI shows cerebellar atrophy (table 3.4.2, figure 3.4.5D).

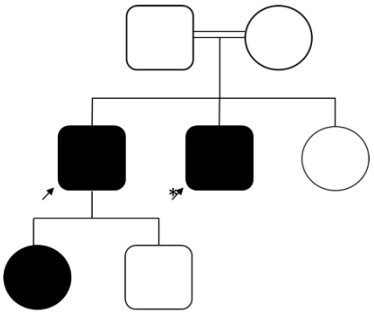
A detailed clinical history for family 4: patient 5 was not available.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
Sex	Male	Male	Female	Female	Female	Male
Age at evaluation			7 months	Unknown	Unknown	21
Age at Presentation	20	20	6	36	Unknown	5
					Unknown	
Neurological symptoms - developmental						
Intellectual disability	Mild	Mild	Y	Y	Unknown	Mild
Epilepsy/seizures	N	Y	N	N	Unknown	N
Autism	N	N	N	N	N	N
Delayed Developmental milestones						
Gross-Motor	Normal	Normal	Delayed	Delayed	Unknown	Normal
Fine-Motor	Normal	Normal	Delayed	Delayed	Unknown	Normal
Speech	Normal	Normal	Delayed	Delayed	Unknown	Normal
Neurological Symptoms						
Peripheral Neuropathy	N	N	N	N		N
Other behaviour/psychiatric abnormality	N	N	N	N	N	N
DaTScan	Normal	Normal	Not performed		Unknown	Not performed
EEG					Unknown	Normal
Movement						
Parkinsonism	Y	N		N		Y
Ataxia (sensory/peripheral)	Y	Y	Y	Y	Y	Y
Tone	Increased	Increased	Increased	Increased	Unknown	Increased
Dysarthria	Y	Y	Y	Y	Unknown	Y

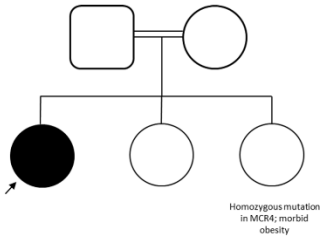
Dysphagia	Y	Y	N	Y	Unknown	Y
Nystagmus	Y	Y	Y	Y	Unknown	Y
Retinal rod-cone dystrophy	N	N	Y	Y	Y	N
Spasticity	Y	N	Y	Y	Unknown	Y (lower body/legs)
Spastic parapesis	N	N	Y	Y	Unknown	N
Bradykinesia	Y	N	N	N	Unknown	Y
Diploplia	N	N	Not available	N	Unknown	N; Sixth nerve palsy (abducens); stabismus covergens
Tremor	N	N	N	N	Unknown	Y (intentional tremor)
Auditory neuropathy	N	Y	N	N	Unknown	N
Head titubation	Y	N	N	N	Unknown	N
Truncal titubation	N	N	Y	N	Unknown	N
Other						
Facial dysmorphologies (specific)	N	N	Cleft palate, tubular nose, hypotelorism, ptosis	N	N	N
Cardiac anomalies	N	N	N	N	N	N
Kidney anomalies	N	N	N	N	N	N
Other organ anomalies	N	N	N	N	N	N

Table 3.4.2: Clinical Features of Patients 1 – 6. Clinical examinations for patients 1 and 2 conducted by Dr Emer O'Connor and Prof. Henry Houlden. Clinical data for other patients provided by their consulting clinician.

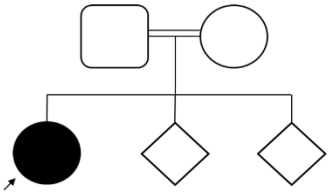
Y, yes; N, no, DaT Scan; dopamine transporter scan; EEG, electroencephalogram



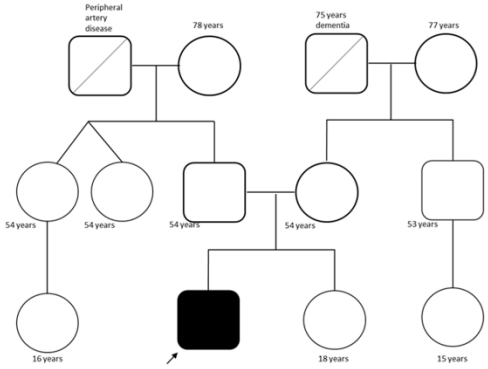
Family 1: Patient 1 (arrow) and 2 (arrow with asterix)



Family 2: Patient 3 (arrow)



Family 3: Patient 4 (arrow)



Family 5: Patient 6 (arrow)

Figure 3.4.5: Pedigrees from clinical information available for five patients, spanning four families. Squares indicate males, circles indicate females, black coloured shapes indicate affected family members, arrow and Asterisks represent patients, double lines indicate consanguineous marriage, diagonal lines indicate twins.

3.4.4.4. ACBD5 mutation causes a loss-of-function of ACBD5 protein in preliminary experiments on patients 1 and 2

To confirm the results of the cDNA sequencing and functional effects of the mutation, western blotting and immunofluorescence imaging were carried out on patient 1. Western blotting showed that there is a complete loss of ACBD5 protein for the patient 1 fibroblast line, compared to a control (figure 3.4.6). To visualise the expression of ACBD5, confocal microscopy was carried out on patient 1 using immunofluorescent tagged anti-ACBD5 antibodies (figure 3.4.6) which show green puncta staining for ACBD5 expression in the control, but a complete absence in the patient.

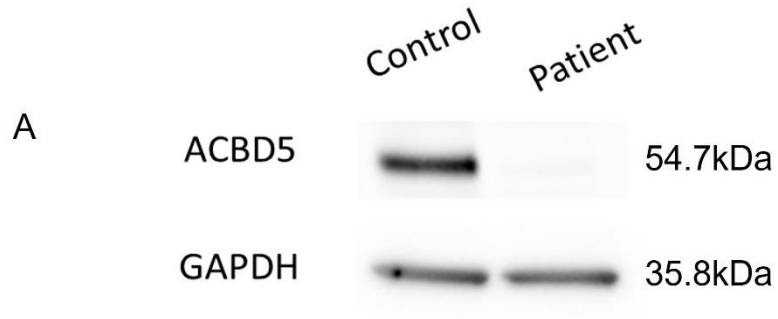


Figure 3.4.6: Loss of function of ACBD5 protein. A, western blot of patient 1 and a control; there is a clear complete loss of expression of ACBD5 protein (54.7kDa) in the patient, compared to the control. B,

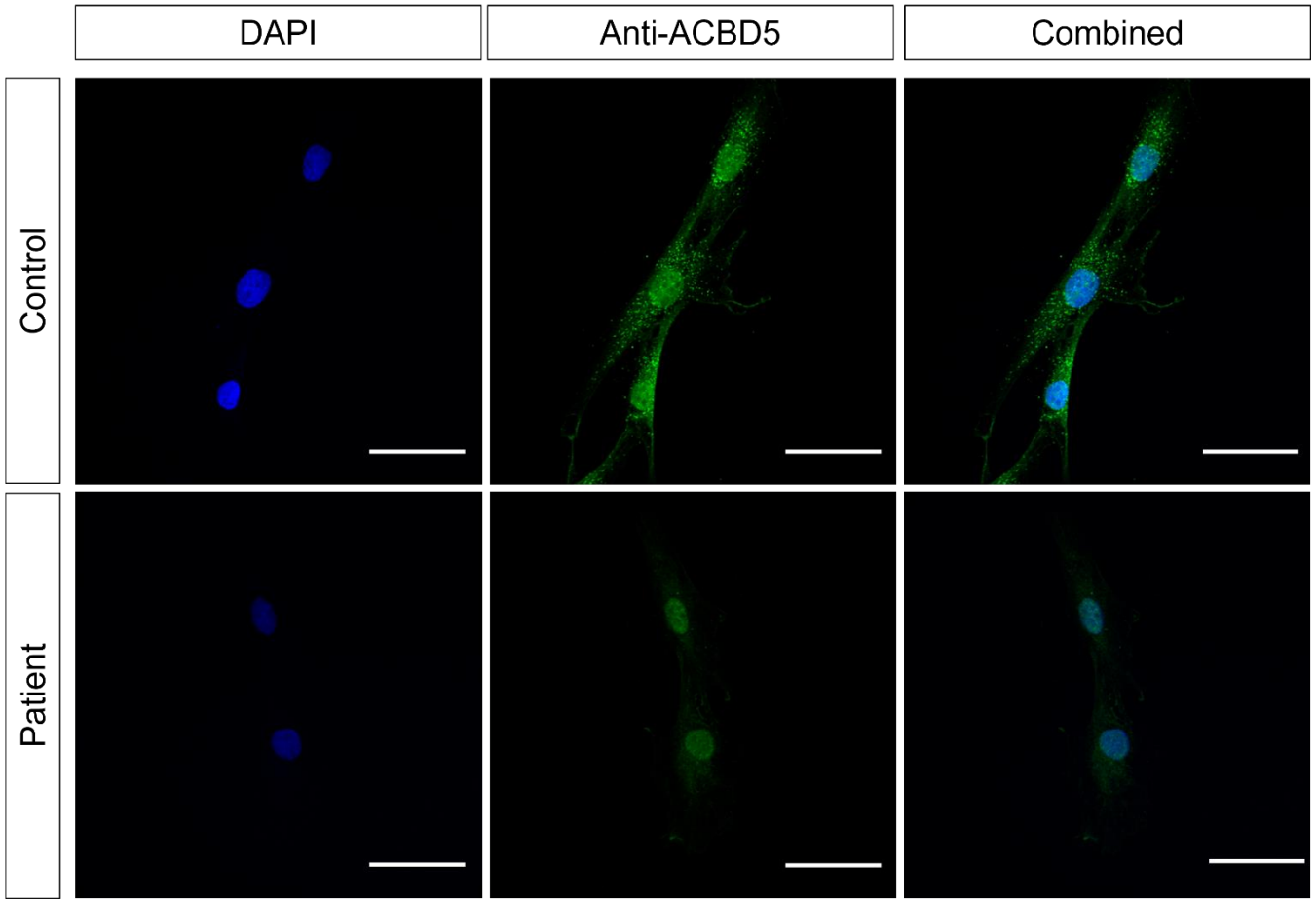


Figure 3.4.7: Immunofluorescence showing loss of function of ACBD5 protein in patient. Immunofluorescent imaging of patient and control fibroblasts showing DAPI staining of the nucleus, anti-ACBD5 staining, and both channels combined. An antibody targeting ACBD5 protein was followed by a secondary donkey anti-rabbit antibody (Alexa Fluoro 488). ACBD5 protein is tagged in green and can be seen as green puncta in the control fibroblasts. There is a complete absence of ACBD5 protein in the patient fibroblast line, indicated by an absence of green puncta. Scale bar indicated in bottom right corner indicates 50µm.

3.4.4.5. Loss-of-function present in additional patient fibroblast lines and ACBD5 present in all additional experimental control lines.

Once additional patients were recruited to the study, and additional fibroblast cell lines collected, western blotting was repeated on all fibroblast cell lines and two healthy controls, using β-actin as a control target gene (figure 3.4.8A). All patients showed a complete loss of protein, with no quantifiable protein bands present, compared to healthy controls. Two additional fibroblast lines that had ZSD and ALD, and therefore peroxisomal dysfunction, were used within experiments as controls. To confirm the presence of ACBD5 in these fibroblast lines, Western blotting was performed (figure 3.4.8B).

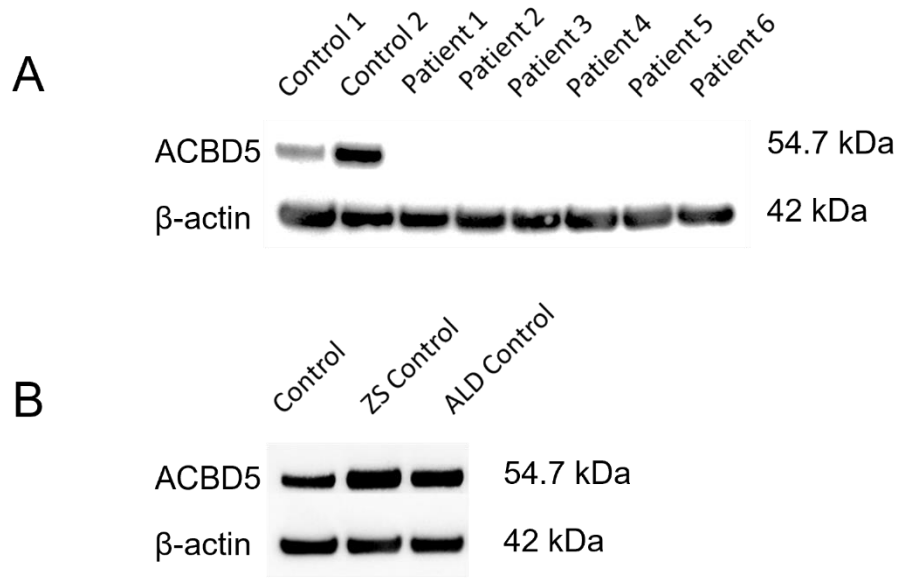


Figure 3.4.8: Western blot of all patient fibroblast and control cell lines. A, Bands corresponding to ACBD5 at 54.7 kDa present in two healthy controls, compared to complete absence of protein bands for all 6 patient fibroblast lines. B, presence of ACBD5 in ZS and ALD control lines, alongside a healthy control. All cell lines, healthy controls, and patients showed protein bands for control gene β-actin at 42 kDa. B, control. *kDa, kilodaltons; ZS, Zellweger syndrome; ALD, adrenoleukodystrophy*

3.4.4.6. ACBD5-deficient patients have an accumulation of very-long-chain fatty acids

To assess whether the loss of ACBD5 causes an effect on VLCFA β-oxidation, patient, carrier, and control fibroblasts underwent metabolic screening, which was carried out by Dr Sacha Ferdinandusse. The results show increased C26:0 phospholipid fractions, increased chain elongation from D3C22 and impaired β-oxidation of VLCFAs in the form of increase C26:0 and increased C26:0/C22:0 phospholipid ratio in both patients. Both carriers had results within the normal range across the board (table 3.4.3). Metabolic screening has been carried out for patient 5 by Prof. Nancy Braverman's research group, with confirmation of VLCFA β-oxidation impairment (unpublished findings by Braverman et al., data not yet shared).

		Carrier 1	Carrier 2	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
C26:0 lyso PC (pmol/mg protein	**C26:0 lyso PC** (Ref range: 2 - 14)	4	6	**27 +**	**19 +**	**32 +**	19 +	NA	**28 +**
D3-C22 Loading Test (µmol/g protein)	**D3-C26 (chain elongation)** (ref range: 0.15 – 0.51)	0.27	0.19	**1.16 +**	**0.94 +**	**1.53 +**	**0.72 +**	NA	**1.32 +**
	Ratio D3-C16/D3-C22 (β-oxidation) (ref range: 0.88 – 1.93)	1.49	1.64	1.46	1.46	NA	1.19	NA	1.31
Very-Long chain fatty acids (µmol/g protein)	**C22:0** (ref range: 2.46- 6.59)	4.16	4.01	3.97	3.81	3.01	5.63	NA	3.2
	C24:0 (ref range: 6.37 – 13.87)	10.51	10.48	10.46	11.29	12.87	12.50	NA	9.73
	C26:0 (ref range: 0.16 – 0.41)	0.31	0.37	**0.81 +**	**0.81 +**	0.26	**1.05 +**	NA	**1.11 +**
	Ratio C24:0/C22:0 Ref range: 1.68 – 2.92)	2.53	2.62	2.64	**2.96 +**	2.54	2.22	NA	**3.04 +**
	Ratio C26:0/C22:0 (ref range: 0.03 – 0.10)	0.08	0.09	**0.20 +**	**0.21 +**	**0.15 +**	**0.19 +**	NA	**0.35 +**

Table 3.4.3: Metabolic screening results of two carrier and two patient fibroblast lines Reference ranges shown. + signifies an abnormal increase compared to reference range. All data generated by Dr Sacha Ferdinandusse.

3.4.4.7. Preliminary results show that ACBD5-deficient fibroblasts have a higher contribution of mitochondrial complex II in the electron-transport chain than control fibroblasts and that there are differences in mitochondrial membrane potential in fibroblasts derived from patients with retinal dystrophy.

To investigate any effects of ACBD5-deficiency on mitochondrial membrane potential a preliminary experiment using fibroblasts from patients 1, 2 and 3 were incubated in tetramethylrhodamine (TMRM) and then viewed using confocal microscopy. Upon pooling the measurements for controls, carriers and patients, patients (n=50) had a lower mitochondrial membrane potential ($\Delta\Psi_m$) figure 3.4.10A) than control (n=45) and carrier (n=45) fibroblasts (n=45), p = 0.02). Carrier fibroblast had a slightly lower $\Delta\Psi_m$ than control fibroblasts, but this did not show statistical significance. At a patient level only patients 3 showed a significantly lower mitochondrial membrane potential ($\Delta\Psi_m$) than controls ($p < 0.0001$) and carriers ($p < 0.001$). Additionally patient 3 had a lower $\Delta\Psi_m$ than both patients 1 and 2 ($p < 0.001$) (figure 3.4.10C).

TMRM intensity changes over time were then measured on control fibroblasts (n = 6), carrier fibroblasts (n=7) and fibroblasts from patients 1, 2 and 3 (n=7), using inhibitors of the different mitochondrial complexes; Oligomycin, Rotenone and FCCP (figure 3.4.10B and figure 3.4.9). Oligomycin, an inhibitor of complex V, did not significantly affect the $\Delta\Psi_m$ of either the control, carrier, or patient fibroblasts. Rotenone, which is an inhibitor of complex I, caused decreased depolarisation of the mitochondrial membrane in all fibroblasts as expected. However, it caused less of a decrease in patients compared to control fibroblasts ($p < 0.0001$ (Figure 3.4.10B). FCCP caused a similar complete depolarisation of the mitochondrial membrane in all fibroblast lines (figure 3.4.10D). At a patient level, there was a similar trend where patients 1, 2 and 3 had less of a decrease in membrane depolarisation, however this lost significance. When separated based on ACBD5-deficient associated phenotype, the patients without retinal dystrophy (patients 1 and 2) showed a significant reduction in depolarisation ($p < 0.0001$) than patients with retinal dystrophy (patient 3), when compared to control fibroblasts (figure 3.4.10E). The slight decrease in depolarisation, compared to controls, and similar depolarization upon addition of FCCP suggests that complex I may not be contributing to the maintenance of the $\Delta\Psi_m$ at as high a level in patient and carrier fibroblasts, compared to controls.

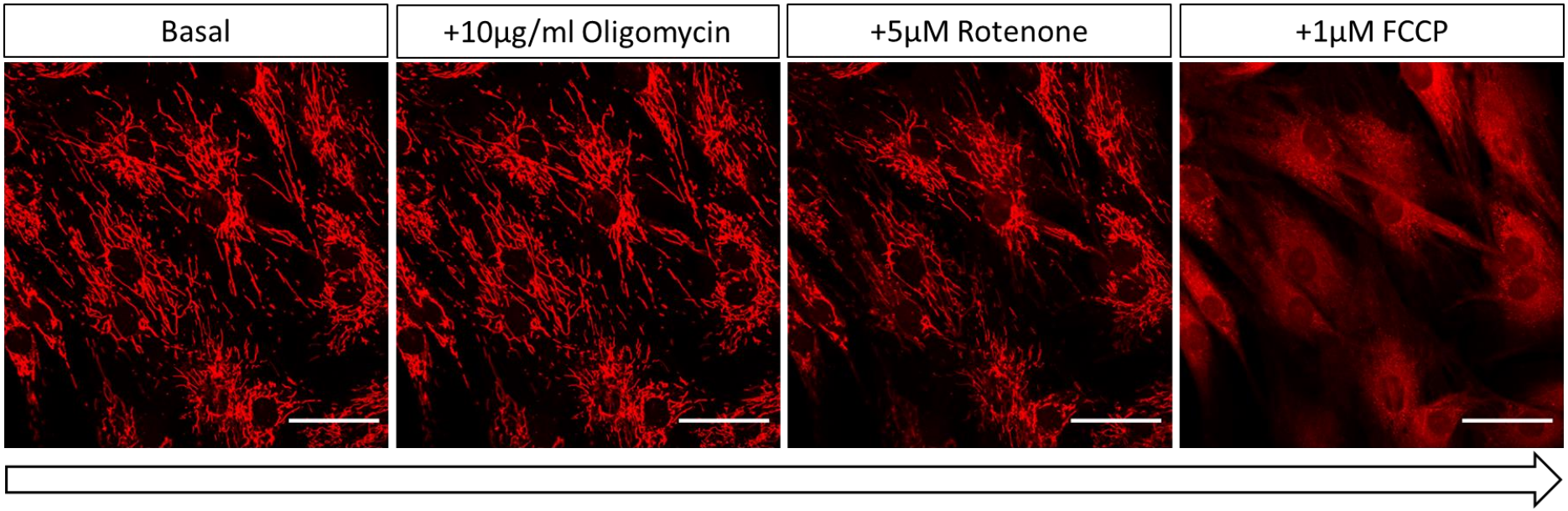


Figure 3.4.9: Mitochondrial complex function using TMRM intensity changes over time with toxin addition that target specific complexes. Basal TMRM intensity measured, 10µg/ml of Oligomycin is then added that targets complex V and TMRM intensity is measured, 5µM of Rotenone which targets complex I is added and TMRM intensity is measured, 1µM of FCCP is then added which causes complete depolarisation of the mitochondrial membrane. *TMRM, tetramethylrhodamine; FCCP, Carbonyl cynanide-p- trifluoromethoxyphenylhydrazone.*

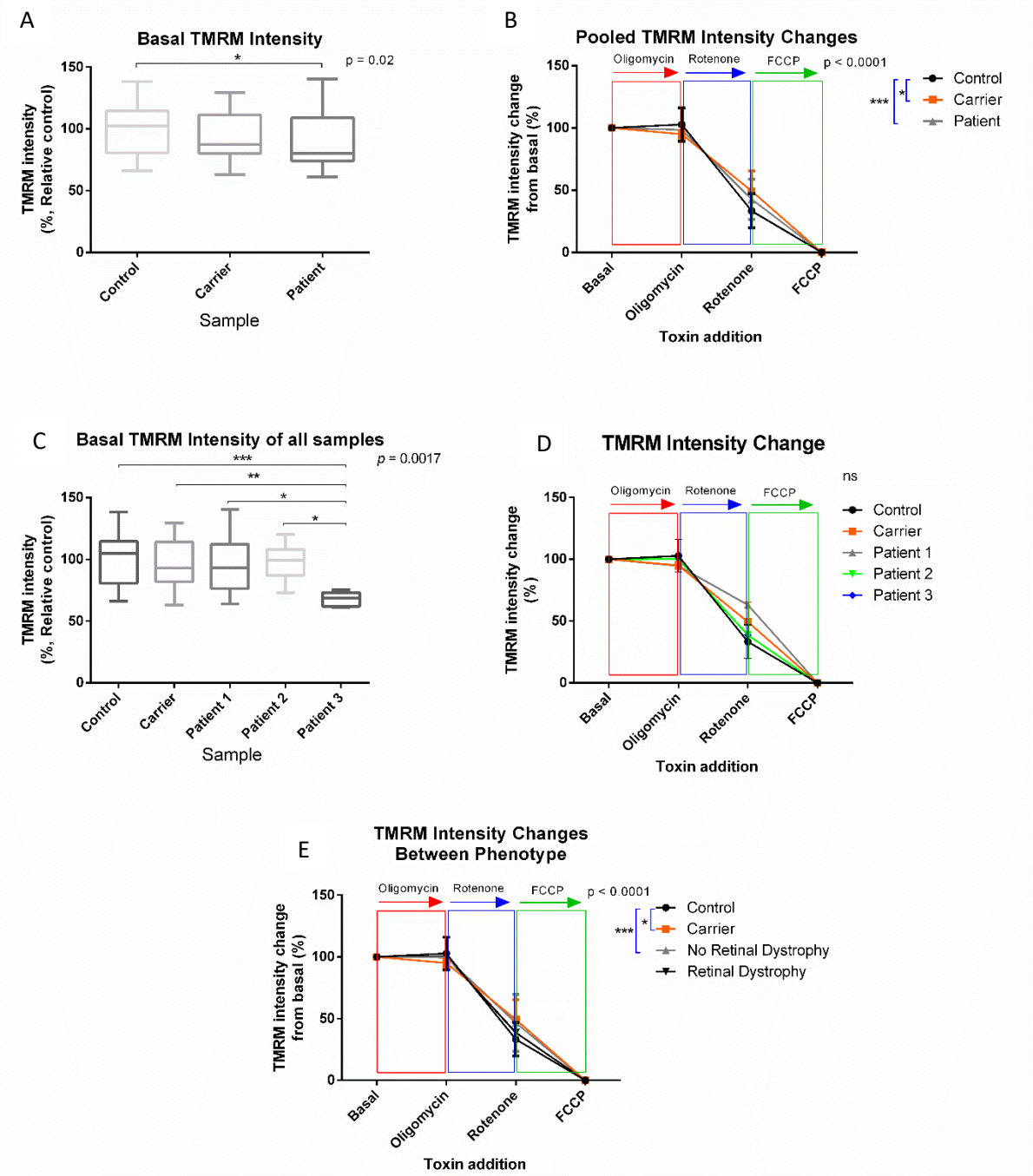


Figure 3.4.10: Assessing changes in mitochondrial membrane potential. A, Pooled basal TMRM intensity, as a percentage of the relative control, Patient (n=3), control (n=2) and carrier (n=2) (p = 0.02). B, pooled changes in TMRM intensity of controls (n=2), carriers (n=2) and patients 1, 2 and 3 as a percentage of the basal intensity level. C, Pooled basal TMRM intensity of pooled healthy controls, pooled carriers, and separate patients. D, Live imaging TMRM intensity changes after toxin addition of pooled controls, pooled carriers, and separate patient fibroblasts. E, live imaging TMRM intensity changes after toxin addition of pooled patients with retinal dystrophy and patients without retinal dystrophy, alongside pooled controls, and pooled carriers (p < 0.0001). One Way ANOVA performed, with Tukey HSD post-hoc test.

3.4.4.8. Validation investigations on second cohort of ACBD5-deficient fibroblasts show that patients have a higher contribution of mitochondrial complex II in the electron-transport chain than control fibroblasts.

To validate previous results showing that ACBD5-deficient fibroblasts have a higher contribution of complex II in the electron transport chain than control fibroblasts, a second separate cohort of 3 ACBD5-deficient fibroblasts, 2 controls and 1 positive control from a patient with ZSD were tested. Fibroblasts were incubated in TMRM and then viewed using confocal microscopy. At a patient level, patients 4 and 5 had a lower membrane potential ($\Delta\Psi_m$) than controls, however patient 6 (n=45) had a higher $\Delta\Psi_m$. All measurements were insignificant ($p = 0.6$) (figure 3.4.11A). When pooled with the results from the preliminary experiment using fibroblasts from patients 1, 2 and 3, membrane potentials were varied between all patients compared to controls ($p = 0.05$) (figure 3.4.11B).

TMRM intensity changes over time were then measured on two control fibroblasts lines (n=6), a positive control line from a patient with ZSD and fibroblasts from patients 4, 5 and 6 (n=6), using inhibitors of the different mitochondrial complexes; Oligomycin, Rotenone and FCCP (figure 3.4.11C). Oligomycin, an inhibitor of complex V, did not significantly affect the $\Delta\Psi_m$ of either the control or patient fibroblasts. Rotenone, an inhibitor of complex I, caused decreased depolarisation of the mitochondrial membrane in all fibroblasts as expected. As seen in the preliminary study (figure 3.4.11D) patients showed less of a decrease in all patients compared to control fibroblasts, particularly in patient 6 ($p < 0.001$) and in the positive control ($p < 0.0001$). FCCP caused a similar complete depolarisation of the mitochondrial membrane in all fibroblast lines (figure 3.4.11C).

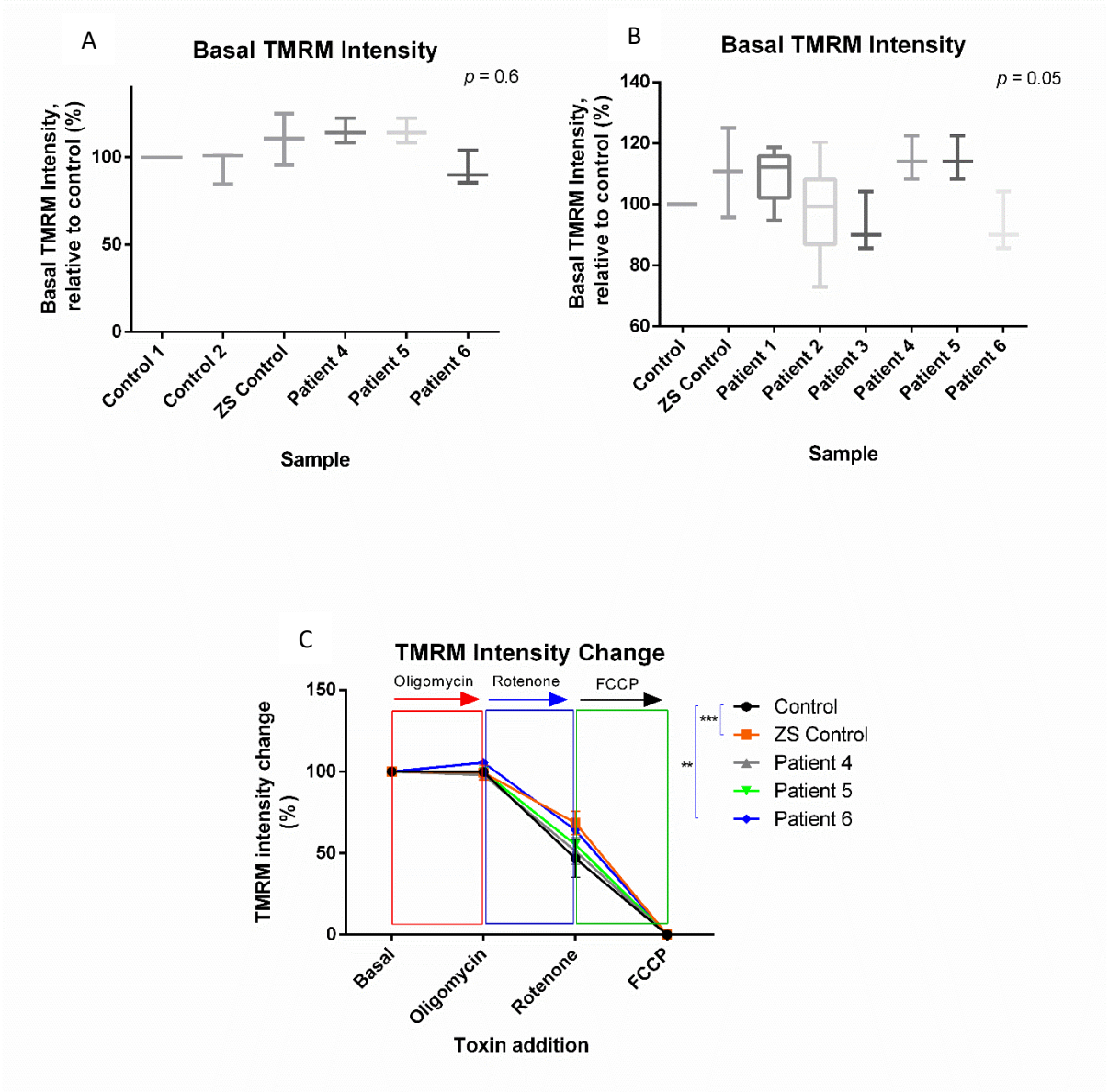


Figure 3.4.11: Results from second round of TMRM basal and live imaging. A, Basal TMRM intensity between two healthy controls, one ZS positive control and three ACBD5-deficient fibroblast cell lines (p = 0.6). B, Basal TMRM intensity of all ACBD5-deficient cell lines, pooled healthy controls and one ZS positive control (p = 0.05). C, Live TMRM imaging following toxin addition for pooled healthy controls, one ZS positive control and three ACBD5-deficient patient fibroblast cell lines. Coloured boxes correspond to toxin addition. *ZS, Zellweger syndrome.*

3.4.4.9. ACBD5-deficient fibroblasts presenting with retinal dystrophy have a lower mitochondrial membrane potential and a higher contribution of mitochondrial complex II in the electron-transport chain than ACBD5-deficient fibroblasts presenting without retinal dystrophy.

Patients were grouped into their presenting phenotypic differences: retinal dystrophy versus no retinal dystrophy and their resulting measurements for basal TMRM level and live TMRM measurements to test the mitochondrial complexes analysed. Fibroblasts from patients 3, 4 and 5 presenting with retinal dystrophy had a significantly higher mitochondrial membrane potential ($\Delta\Psi_m$) than those patient fibroblasts presenting without retinal dystrophy ($p < 0.0001$) and also the positive control line with ZSD ($p < 0.01$), suggesting differences in $\Delta\Psi_m$ (figure 3.4.12A). Live TMRM results showed that both patients with and without retinal dystrophy had less of a decrease in depolarisation than control fibroblast lines, but those without retinal dystrophy had less of a decrease than those with retinal dystrophy ($p < 0.001$). The positive control line presenting with ZSD had the smallest response to rotenone, compared to control fibroblast lines ($p < 0.00001$), patients without retinal dystrophy ($p < 0.001$) and patients with retinal dystrophy ($p < 0.01$) figure 3.4.12B). Taken together, these results suggest that patients without retinal dystrophy have a more significantly reduced contribution of complex I than patients with retinal dystrophy.

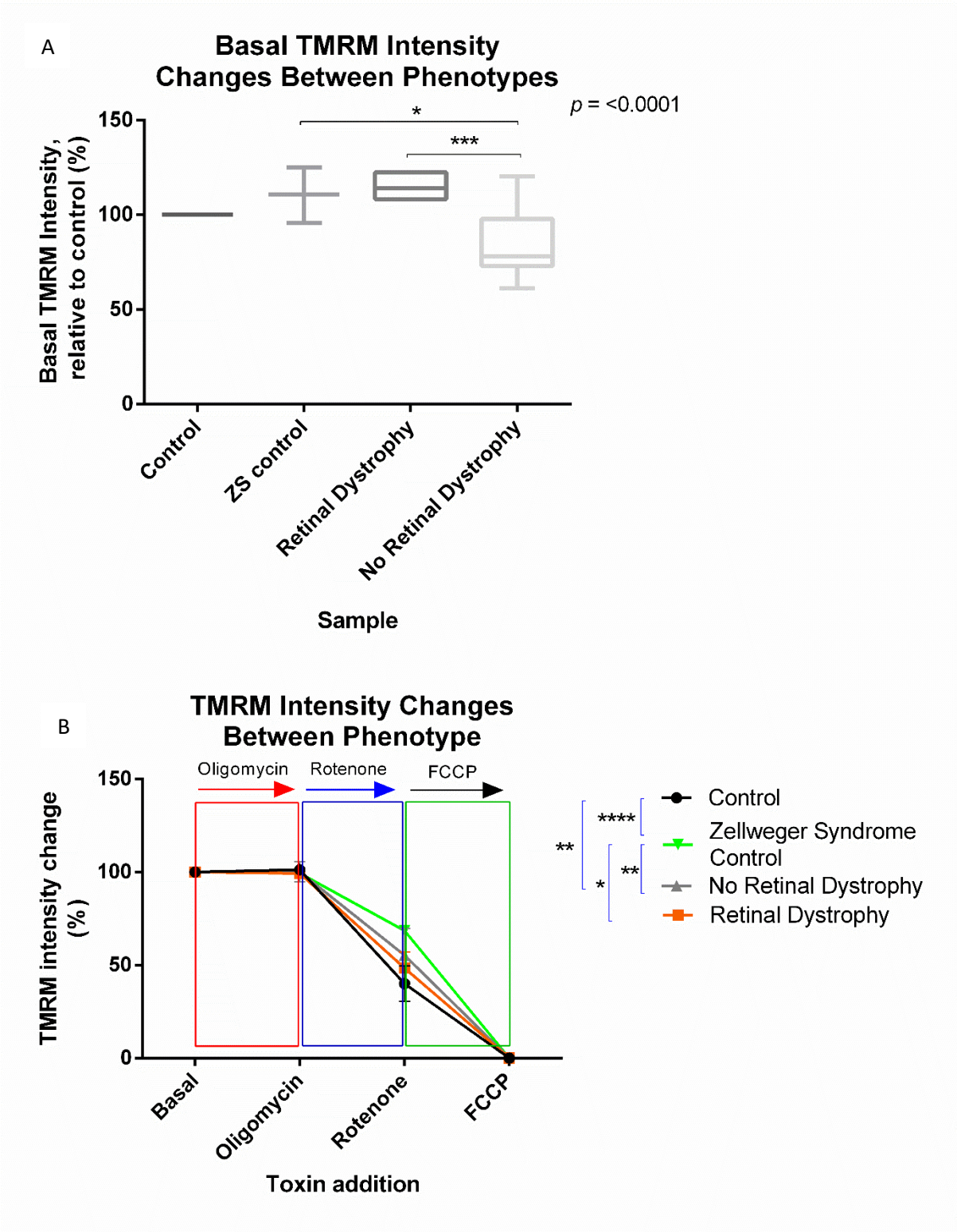


Figure 3.4.12: TMRM intensity changes between different ACBD5 disease phenotypes. A, Pooled basal TMRM intensity changes between patients with retinal dystrophy, and patients without, along with healthy controls and one ZS positive control ($p < 0.0001$). B, Pooled results from live TMRM imaging of patients with retinal dystrophy, patients without, along with healthy controls and one ZS positive control. Coloured boxes correspond to different toxin addition. *ZS, Zellweger syndrome*

3.4.4.10. Preliminary experiments on patient fibroblasts show intense patches of NADH fluorescence, suggesting a higher basal level of NADH in patients

To examine respiratory function of respiratory complex I, the redox state of NADH was measured. Figure 3.4.13A shows images of the basal level of NADH, with intense patches of fluorescence present in the three patients (highlighted in red) that did not respond to FCCP or NaCN but absent in the control and carrier. This suggests a higher basal level of NADH in patients, pertaining to a possible impairment of complex I respiration. Basal autofluorescence of NADH was measured prior to the addition of FCCP to depolarize the mitochondria which stimulated respiration and caused complete oxidation of NADH to NAD^+, diminishing fluorescence, which was then considered as 0%. NaCN was then added, inhibiting respiration, and allowing for the complete generation of NADH, considered as 100%. Basal NADH redox state (%) was extrapolated from the minimum and maximum autofluorescence and controls seemed to have a significantly reduced redox index (i.e. more reduced redox state) compared to carriers and patient 3 (figure 3.4.13B). There seemed to be a more reduced redox state for all patients, compared to controls. NADH Pool was calculated as the difference between maximum and minimum autofluorescence (%, relative to control) and is reduced in patients and carriers (figure 3.4.13C).

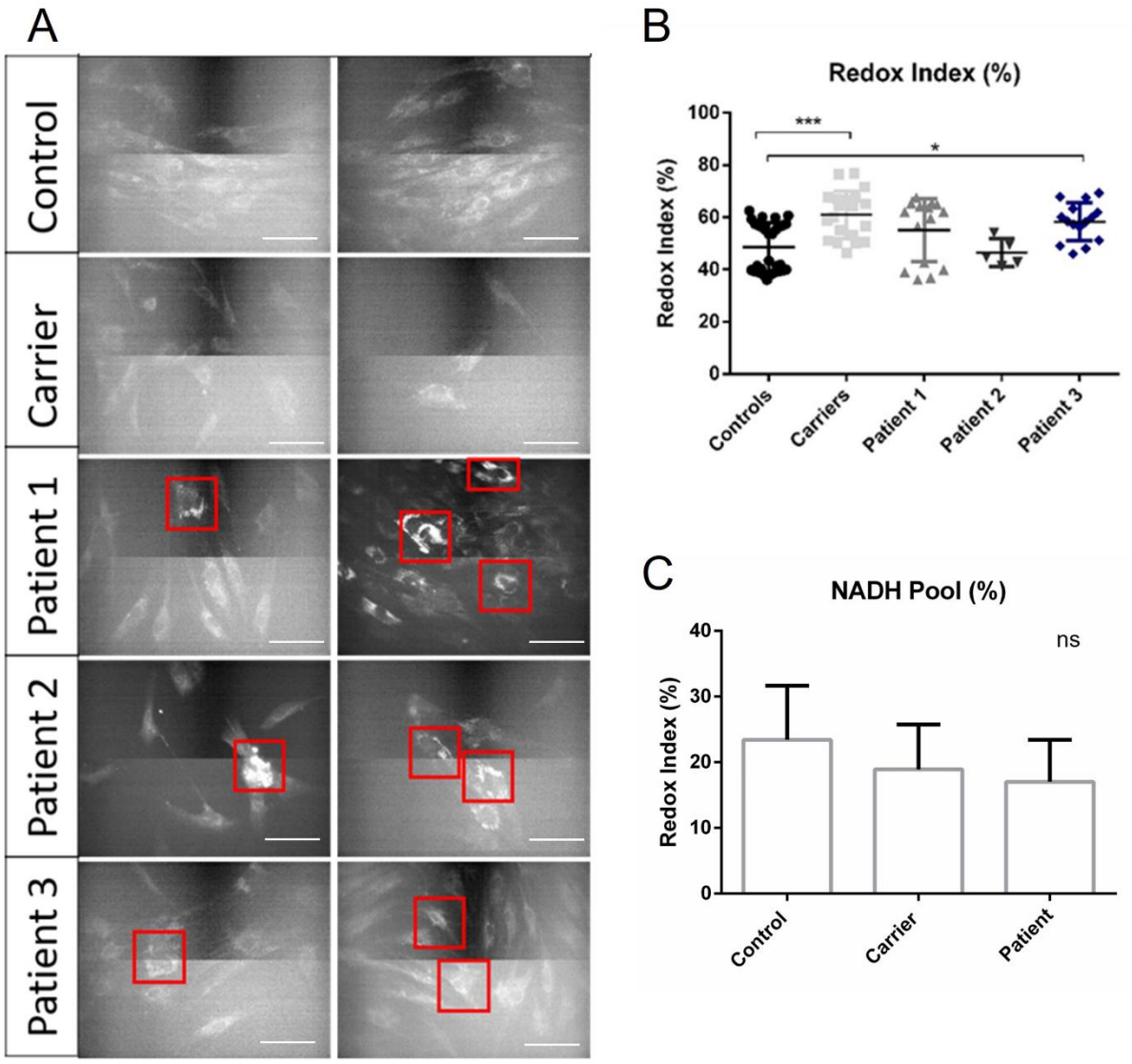


Figure 3.4.13: Confocal microscopy using NADH-autofluorescence. A, shows images taken at oil-immersed 40x magnification for control, carrier and three patients. Intense patches of NADH fluorescence, indicating high levels of basal NADH, can be seen in the three control fibroblast cells (indicated in red squares) in comparison to the control and carrier fibroblast cells. Two-tone colour in images is an artifact of the microscope. B, shows the redox index (%) for controls, carriers, and patients 1, 2 and 3, showing a deceased redox index for controls, compared to carriers and patients. One-way ANOVA performed, with Tukey multiple comparison test, * = $p<0.05$, *** = $p< 0.005$. **C** shows the NADH Pool normalised to control levels, controls have a higher NADH pool compared to carriers and patients. One-way ANOVA performed, with Tukey HSD post-hoc test, no significance found. Scale bar represents 50µM.

3.4.4.11. Validation cohort of patient fibroblasts show no significant differences in basal NADH fluorescence or redox index but suggest trends of a higher redox index and reduced NADH pools in patients compared to controls.

To further look at mitochondrial respiration in patient fibroblasts, preliminary NADH autofluorescence were repeated on a second cohort of ACBD5-deficient patient fibroblasts (patients 4, 5 and 6), compared to three controls and one positive control with ZSD. As in the preliminary experiments (figure 3.4.14A) controls had a reduced redox index (i.e., more reduced redox state) compared to patients, however this was not significant (figure 3.4.14B). The positive ZSD control had a significantly increased redox index, compared to controls ($p < 0.001$). NADH autofluorescence showed a significantly reduced NADH pool in patients, compared to control fibroblasts ($p = 0.04$), which mirrors preliminary findings (figure 3.4.14D). At a patient level, measurements for NADH pool were varied between patients, when preliminary and validation cohorts were combined (figure 3.4.14E). No significant differences in NADH pool and redox index were seen between patients with retinal dystrophy and those without (figure 3.4.14F).

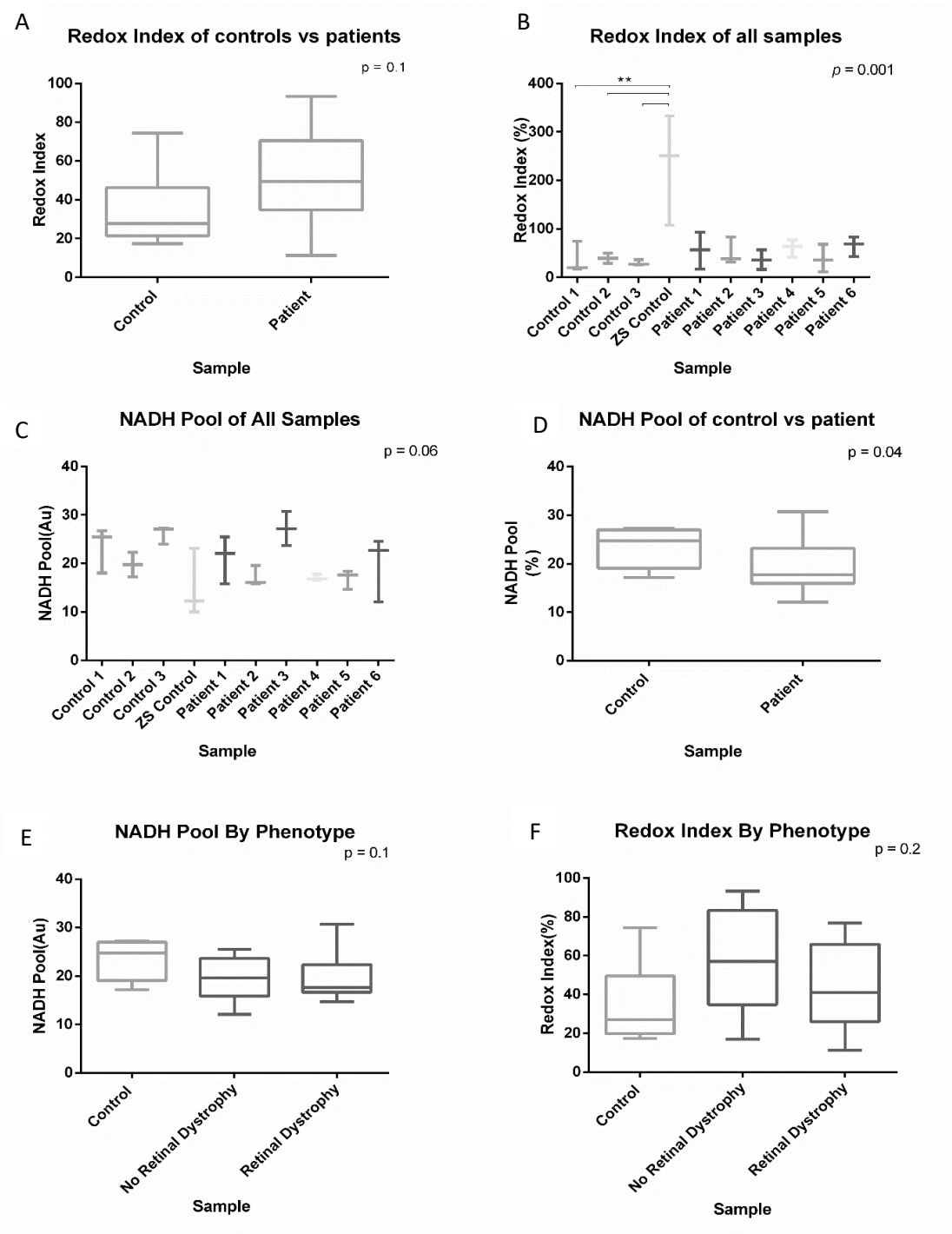


Figure 3.4.14: Results from NADH autofluorescence. A, Pooled redox index, showing basal redox state of controls and patients (p = 0.1). B, Redox index of healthy controls (n =3), ZS positive control and patients (n = 6) (p = 0.001). C, NADH pool of all cell line fibroblasts. D, Pooled NADH pools of controls and patients (p = 0.04). E, Pooled NADH pool of healthy controls and patients with and without retinal dystrophy (p = 0.1), F, Pooled redox indexes for healthy controls, and patients with and without retinal dystrophy (p = 0.02). One-way ANOVA performed, with Tukey HSD, post-hoc test, T-test performed for D. *ZS, Zellweger syndrome*.

3.4.4.12. There are no significant differences in the rate of lipid peroxidation between patients and controls

The rate of lipid peroxidation was analysed in three control, one control line of a patient with adrenoleukodystrophy (ALD) and three patient fibroblasts using BODIPY (655/675, 2 µM, Molecular Probes) which is a lipid peroxidation sensor (figure 3.4.15). BODIPY was excited by 563 and 630 nm lasers and measured from 580 to 610 and above 650 nm and data measured using ZEN2009 software. There were no significant differences in the rate of lipid peroxidation between patient and control fibroblasts (p = 0.3) and no differences in fold-change when patient results were normalised to control (p = 0.07) (figures 3.4.15A and 3.5.15B). Representative microscopy images of rate of lipid peroxidation investigations can be seen in figure 3.4.16.

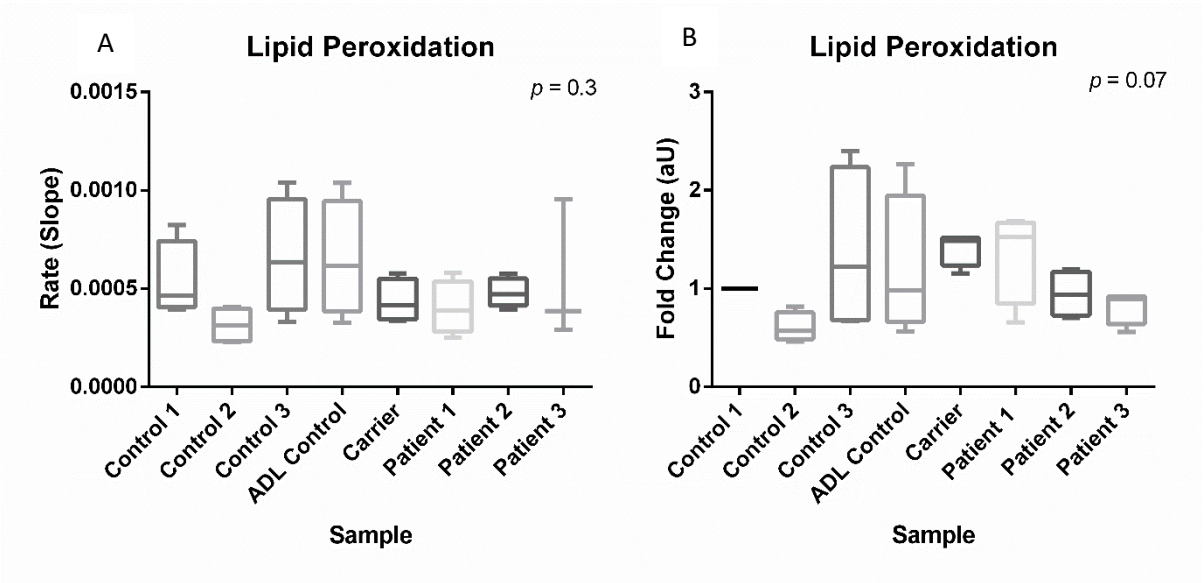


Figure 3.4.15: Changes in mitochondrial lipid peroxidation across cell lines. A, rate of lipid peroxidation of all ACBD5-deficient patient lines, three healthy controls and one positive ALD control (p = 0.3). B, fold changes of lipid peroxidation rates between all ACBD5-deficient patient lines, three healthy controls and one positive ALD control (p = 0.07). Sample values normalised to control 1. *ALD, adrenoleukodystrophy*

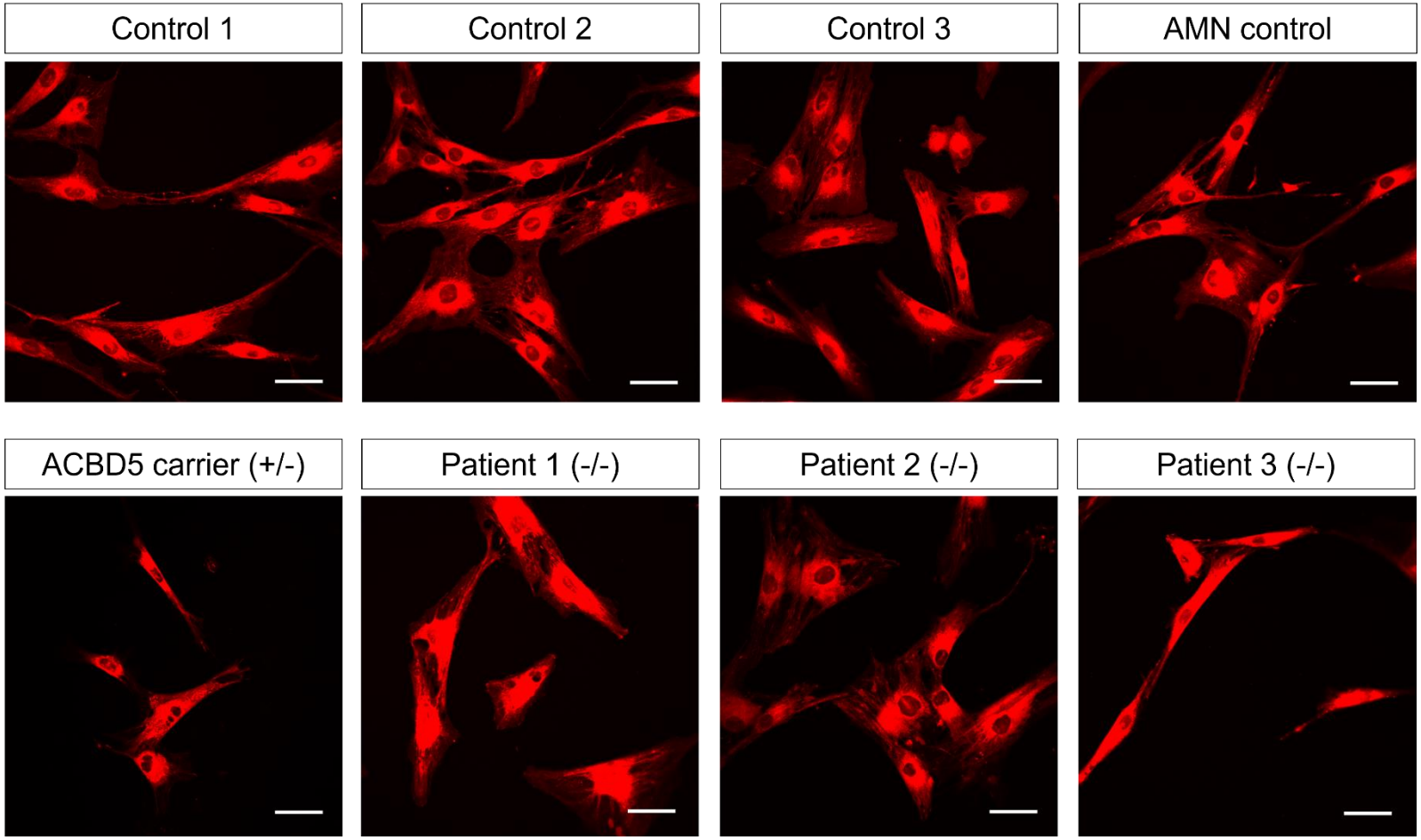


Figure 3.4.16: Representation of cells under the microscope during lipid peroxidation analysis. Rate of lipid peroxidation measured in three controls, an ABCD1-deficient patient with Adrenomyeloneuropathy, a carrier of ACBD5-deficiency, and patients 1, 2 and 3. BODIPY (655/675, 2 µM, Molecular Probes) was excited by 563 and 630 nm lasers and measured from 580 to 610 and above 650 nm. Intensity of pixels was measured over time. + = ACBD5 present on allele, - = ACBD5 allelic deficiency. Scale bars represent 50µM.

3.4.4.13. ACBD5-deficient fibroblast lines have higher glutathione concentrations ([GSH]) compared to control fibroblast lines, with higher elevations in patients without retinal dystrophy.

In order to measure levels of mitochondrial oxidative stress, glutathione levels were measured in three control fibroblast lines, one positive patient control line with adrenoleukodystrophy and six ACBD5-deficient patient fibroblast lines (figure 3.4.17). To measure glutathione concentration ([GSH]), cells were incubated with 50 μm monochlorobimane (MCB) in HBSS at room temperature for 40 min, or until a steady state had been reached before images were acquired for quantitation. Monochlorobimane (MCB) is a non-fluorescent bimane which freely permeates across the cell membrane and forms a fluorescent adduct when combined with reduced glutathione (GSH) in a reaction catalyzed by glutathione s-transferase. Across all cell lines, ACBD5-deficient patients had raised GSH levels, compared to controls, however only patient 1 and patient 6 had significant increases ($p < 0.01$). The ALD control line had reduced GSH levels, compared to controls (not significant) (figure 3.4.18A). When taken together, ACBD5-deficient patients had significantly elevated levels of GSH compared to control fibroblast lines ($p = 0.038$) (figure 3.4.18B), and additionally those patients without retinal dystrophy had significant increases to controls ($p < 0.001$) (figure 3.4.18C). The ZSD control line also had significantly increased levels of GSH compared to control lines ($p < 0.01$), whereas the ALD control line had significantly reduced levels of GSH, compared to ZSD control line ($p < 0.001$). The patients without retinal dystrophy had significantly elevated levels of GSH compared to the ALD control line ($p < 0.01$) (figure 3.4.18C). Taken together, these data suggest that ACBD5-deficient patient fibroblasts have an elevation in GSH levels, similarly to the ZSD control line compared to control fibroblast lines. Within the patient cohort, those without retinal dystrophy have elevated GSH levels when compared to controls, but patients with retinal dystrophy showed no differences when compared to controls.

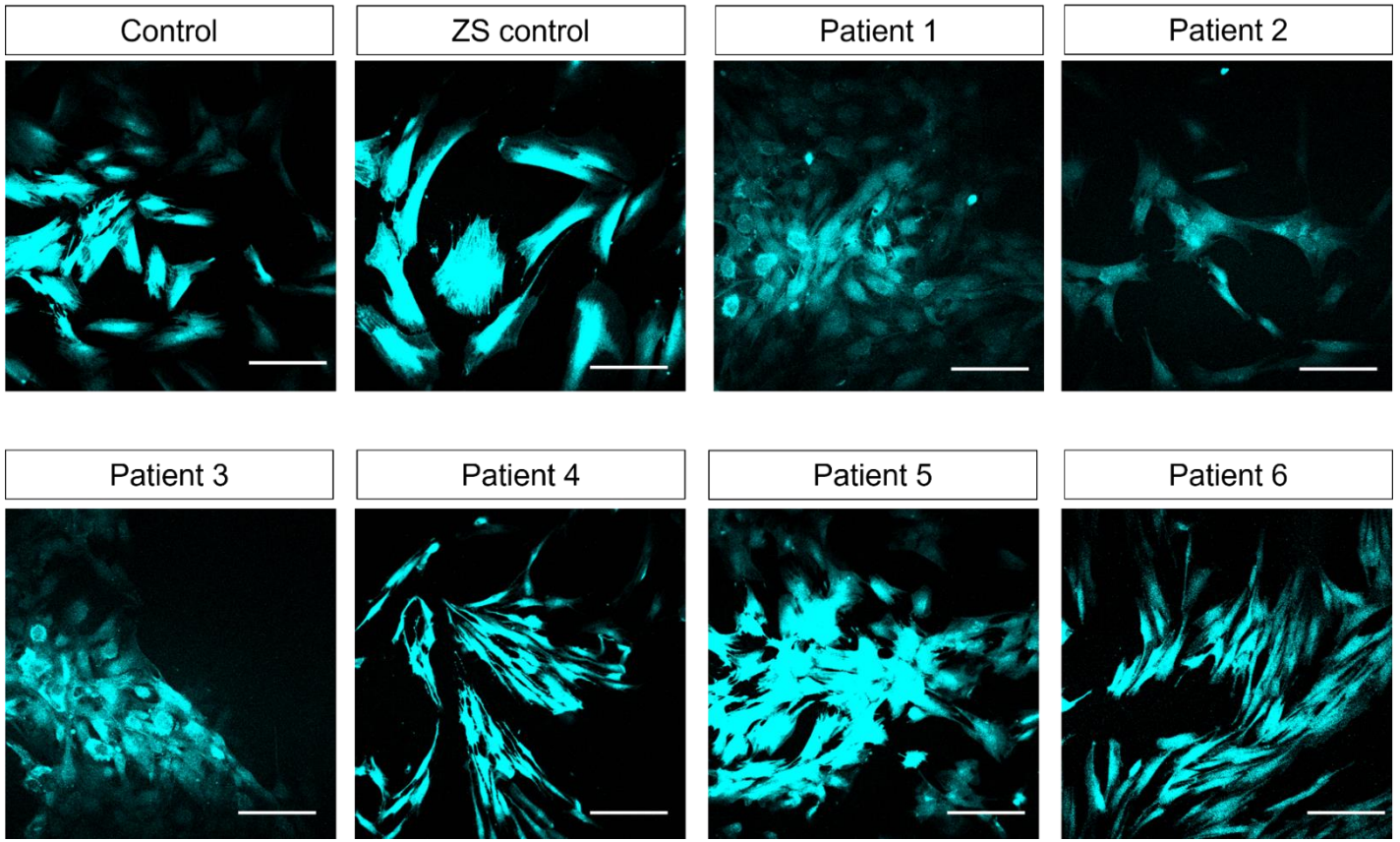


Figure 3.4.17: Representative microscopic images of glutathione concentration quantification. Fibroblast cells for controls, a ZS positive control and all patients were incubated with MCB until a steady state had been reached. MCB is a non-fluorescent bimane which freely permeates across the cell membrane and forms a fluorescent adduct when combined with reduced GSH in a reaction catalysed by glutathione s-transferase. Pixel intensity was quantified for each microscopic image. 5 images taken per cell line; analysis repeated three times using separate cell passages. Scale bars represent 50 µM. *MCB, Monochlorobimane; GSH, glutathione*

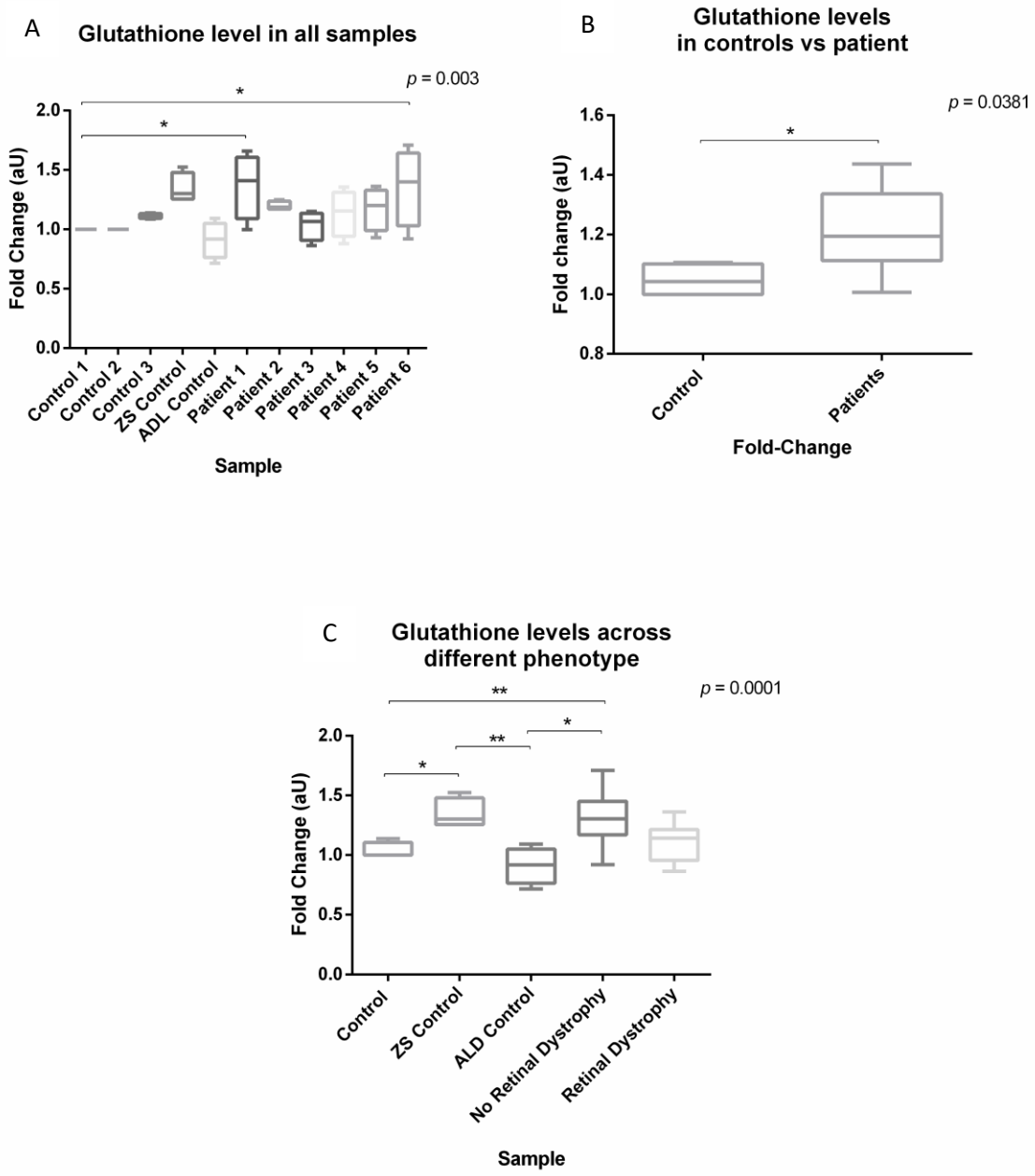


Figure 3.4.18: Glutathione concentrations across cell lines. GSH concentrations were measured in all six ACBD5-deficient patient lines, as well as two healthy controls, one positive ZS control and one positive control with ALD. A, glutathione levels across all cell lines (p = 0.003). B, Pooled glutathione levels in controls and patients (p = 0.0381). C, pooled glutathione levels in patients with retinal dystrophy phenotype and without, alongside pooled healthy controls, and ZS and ALD positive controls (p = 0.0001). One-way ANOVA performed, with Tukey HSD post-hoc test. T-test performed for B. *ZS, Zellweger syndrome; ALD, adrenoleukodystrophy.*

3.4.4.14. Patients without retinal dystrophy do not show a reduction in peroxisomal number, consistent with patients with the retinal dystrophy phenotype.

Previous studies have shown that peroxisomal number is unchanged between HeLa *ACBD5* knockout cell lines and an ACBD5-deficient patient cell line (patient 3) with the phenotype of retinal dystrophy, and control lines. To investigate if ACBD5-deficient patient lines without the phenotype of retinal dystrophy also had similar peroxisomal numbers to controls, fibroblast cells from patients 1 and 6, along with two age-matched control lines, were sent to Prof. Michael Schrader's team for further analysis. Cells were processed for immunofluorescence using antibodies against the peroxisomal membrane marker PEX14 by Mrs Tina A. Schrader and Miss Rechal Kumar from the Schrader Group at Exeter University, and peroxisomal number was calculated using ImageJ, utilising the Analyze Particles function. No significant differences were found in the number of peroxisomes in both patients 1 and 6 (306 and 176, respectively) and control lines (254 and 184, respectively), suggesting that ACBD5-deficiency and the phenotype a retinal dystrophy, do not show a reduction in peroxisomal biogenesis (figure 3.4.19).

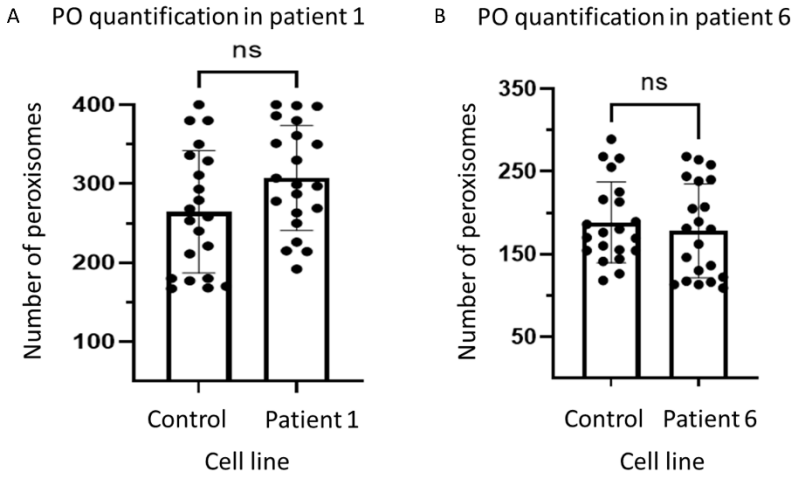


Figure 3.4.19: Peroxisomal quantification in fibroblast cell lines from ACBD5-deficient patients without retinal dystrophy and control lines. A, PO quantification in fibroblast cells from patient 1 and an age-matched control. B, PO quantification in fibroblast cells from patient 6 and an age-matched control. Error bars represent SEM. *PO, peroxisome; ns, not significant; SEM, standard error of the mean.* All work was carried out by Mrs Tina A. Schrader and Miss Rechal Kumar, from the Schrader Group at Exeter University, and data kindly provided by Prof. Michael Schrader.

3.4.4.15. ACBD5-deficient patients without retinal dystrophy have impaired peroxisomal-endoplasmic reticulum contacts.

Previous studies have shown that contacts between peroxisomes and the ER are reduced in *ACBD5* knockout cell lines and peroxisomal motility away from the ER is increased (395). To investigate if this effect was also present in ACBD5-deficient patients without retinal dystrophy, fibroblast cells from patients 1 and 6 underwent peroxisomal-ER attachment assays and electron microscopy. All data were generated by Dr. Christian Hacker from the Exeter Bioimaging Centre, and cell cultures were grown and maintained by Mrs Tina A Schrader from the Schrader group. Organelle contacts between PO's and ER were quantified, with a systematic uniform random (SUR) approach to ensure unbiasedness of the organelle selection. Peroxisomes were identified as single membrane-bound organelles of round to oval shape, displaying a fine granular homogenous matrix and matching the size range previously published. Grid lines intersecting with PO membranes were categorized as being an organelle contact site (distance between PO and ER membrane ≤ 15 nm) or non-associated (yielding 232 ± 34 [STDEV] intersections in total per experiment).

Both patients had significantly reduced peroxisomal-ER membrane contact; 12% (±1.2%) and 7.3% (±1.3%) for patients 1 and 6 respectively ($p < 0.0005$), compared to healthy control fibroblast lines (21.2% (±0.74%) and 22.9% (±1.5%), respectively) (figure 3.4.20A). Additionally, both patients had significantly less peroxisomal-ER attachment compared to healthy control lines, with 53% (±6.4) attachment in patient 1 and a more severe reduction in patient 6 at 34.6% (±5.7%, $p < 0.0005$), compared to controls (75.6% (±6.4%) and 80.5% (± 3.5%) respectively) (figure 3.4.20B and 3.4.20C).

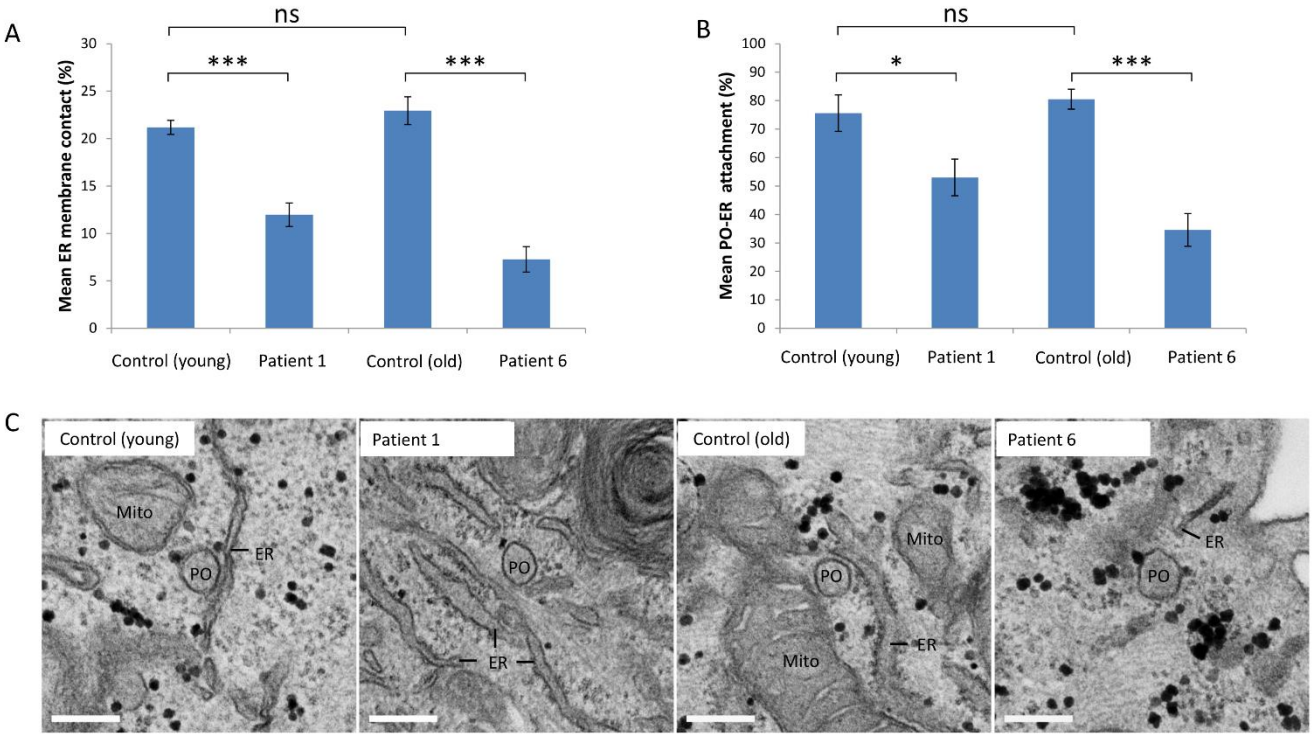


Figure 3.4.20: PO-ER association is impaired in ACBD5 patient cells. (A) Quantitation of PO-ER contacts using spatial stereology. The mean fraction of total PO membrane surface in direct contact with ER (<15 nm distance). (B) The mean proportion of PO's associated with ER. (C) Representative EM micrographs of PO's and their interaction with ER in dACBD5 patient and age-matched control cell lines. Quantitative results were analysed statistically by one-way analysis of variance with Tukey's multiple comparison test; ns, not significant; * ≤ 0.05; ***, P ≤ 0.001. Error bars represent the SEM from 4 experiments per cell line (from 2 individual culture dishes and 2 embedded EM blocks per dish). *PO, peroxisome; Mito, mitochondrion; ER, endoplasmic reticulum*; *SEM, standard error of the mean.* Scale bars: 200 nm. All data generated by Dr Christian Hacker from the Exeter Bioimaging Centre and kindly provided by Prof. Michael Schrader from Exeter University.

3.4.4.16. All ACBD5-deficient patient mRNA transcripts are predicted to undergo nonsense-mediated decay (NMD).

Due to the differences in phenotype seen between patients and the differences in severity of mitochondrial function, it was hypothesised that some patient mRNA transcripts may be undergoing NMD-escape, a phenomenon whereby parts of transcripts intended for NMD are translated into truncated proteins that then have the ability to produce deleterious downstream effects. A Western blot was previously carried out to confirm protein loss-of-function in all cases, with no truncated proteins visible. To further support this all mutations were run through the online portal Mutation Taster, and all were predicted to undergo NMD (table 3.4.4).

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
Zygosity	Homozygous	Homozygous	Homozygous	Homozygous	Homozygous	Compound heterozygous
Variant 1						
Alteration Type	Deletion	Deletion	Deletion/Inversion	Single Base Exchange	Single Base Exchange	Insertion
AA Changes	l243Mfs*6	l243Mfs*6	p.D208Vfs*30	W489*	K369*	R404Kfs*19
Frameshift	Yes	Yes	YES	No	No	Yes
NMD Predicted	Yes	Yes	N/A	Yes	Yes	Yes
Position of stop codon in WT/patient cDNA	1705/871	1705/871	N/A	1705/1467	1705/1234	1705/1393
Splice site abrogation predicted	Yes	Yes	N/A	No	No	Yes
Variant 2						
Alteration Type						Deletion
AA Changes						F34Sfs*2
Frameshift						Yes
NMD Predicted						Yes
Position of stop codon in WT/patient cDNA						1705/1153
Splice site abrogation predicted						No

Table: 3.4.4 Results from Mutation Taster online portal

All patients have NMD predicted, with the position of the stop codon brought forward in all cases, compared to reference cDNA sequence. Three patients are predicted to have splice site alterations, these also have a frameshift predicted as a result of the mutation. Mutation for patient 3 was too complex to be run through the Mutation Taster portal

*NMD, nonsense-mediated decay; AA, amino acid; fs, frameshift; I, isoleucine; M, methionine; W, tryptophan; K, leucine; R, arginine, F, phenylalanine; S, serine; *, premature termination codon*

3.4.4.17. Mutations in four patients could cause NMD escape.

To investigate the potential presence of NMD-escape and that could result in truncated protein transcripts, the protein sequence was predicted via online portal Expasy Translate. Predicted amino acid alterations and predicted premature termination codons were visualised in relation to the targeted antibody. Mutational alteration of amino acids occurs for patients 1, 2, 3, and 6 occur before the antibody target sequence for prior Western blotting (figure 3.4.21). In addition, predicted premature termination codons also occur before the antibody target sequence for patients 1, 2, 3, and 6. Mutational amino acid alterations and predicted premature stop codons for patients 4 and 5 occur within the antibody binding region. This raises the possibility that patients 1, 2, 3 and 6 could experience NMD escape and form small, truncated transcripts, which may not be identifiable on Western blot (figure 3.4.21).

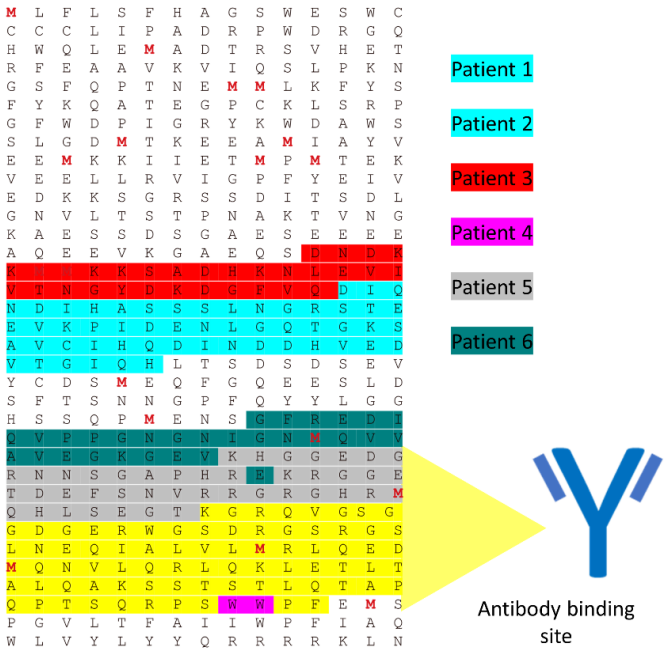


Figure 3.4.21: Amino acid transcript for ACBD5 protein, with denoted mutational amino acid changes and position of new termination codon for each patient. Antibody binding site highlight in yellow, spanning sequence encompassed by the yellow arrow. Patients highlighted in differing colours, patient 1 and patient 2 (blue), patient 3 (red), patient 4 (pink), patient 5 (grey) and patient 6 (teal). Start of highlighted colour indicates amino acid change as a result for the mutation, end of highlighted colour indicates new sequence position of termination codon, resulting from mutation. Mutations in patients 1, 2, 3, and 4 cause premature stop codons that occur before the sequence where the antibody immunogenic sequence binds, which raises the possibility that they may experience NMD escape on potentially truncated transcripts. Red 'M' signifies the presence of a methionine start codon. *NMD, nonsense-mediated-decay*

3.4.4.18. RNA sequencing of six ACBD5-deficient patients

To assess *ACBD5* transcript decay and potential differences in gene expression RNA-sequencing (RNA-seq) was performed on all six ACBD5-deficient patients and 33 controls (comprising patients recruited for a cluster headache study) using RNA extracted from fibroblast cell lines. The raw data was analysed using Kallisto by Dr Jana Vandrovcova and Sleuth software by myself. Two sets of analyses were performed: firstly ACBD5-deficient patients (n=6) against controls (n=33), and secondly ACBD5-deficient patients with a retinal dystrophy phenotype (n=3) against ACBD5-deficient patients without a retinal dystrophy phenotype (n=3). In order to determine whether the majority of variants within the transcriptomics reflected the disease status, dimensionality reduction was performed (principal component analysis, PCA) which did not show any clustering by disease type (figure 3.4.22B). This suggests that at a transcriptomic level both groups were very similar, which may indicate that the pathway differences are limited to a small number of genes that does not affect a large proportion of the transcriptome. Differential gene expression analyses revealed that six gene transcripts were identified as either significantly or close to significantly ($p < 0.1$, after multiple testing correction) downregulated in ACBD5-deficient patients, when compared to control lines. These genes were *ACBD5* (three transcripts with $p = 0.076$, $p = 0.014$ and $p = 0.003$), Ankyrin repeat domain 11 (*ANKRD11*, $p = 0.004$), Dynamin 2 (*DNM2*, $p = 0.029$) and Drosha Ribonuclease III (*DROSHA*, $p = 0.004$) (figure 3.4.23). The remainder of the genes fell below significance after false discovery rate correction (figure 3.4.22A Appendix figure 8.3.1). This confirms that there is a loss of *ACBD5* mRNA transcript in all six ACBD5-deficient patients and suggests that there could be candidate genes also implicated in the pathophysiology of ACBD5-deficiency.

When looking at ACBD5-deficient patients with retinal dystrophy against those without retinal dystrophy no significant differentially expressed gene transcripts were identified after false discovery rate correction (figure 3.4.24A, appendix figure 8.3.1). Additionally, dimensionality reduction did not show any clustering by disease phenotype (figure 3.4.24B). The top 5 genes ($p < 1.0$, when corrected for multiple testing) were Zinc Finger and BTB Domain Containing 22 (*ZBTB22*, $p = 0.15$), Human Major Histocompatibility Complex (MHC) Class I Chain-Related gene A (*MICA*, two transcripts both with $p = 0.261$), Tripartite Motif Containing 26 (*TRIM26*, $p = 0.891$) and Major Histocompatibility Complex, Class I, A (*HLA-A*, $p = 0.244$) (figure 3.4.25). This was an extremely small cohort and therefore the lack of significance when correcting for multiple testing is to be expected, however it does identify some candidate genes that are differentially expressed between these two phenotypic groups, which are significant before multiple testing correction.

There were zero significantly differentially expressed transcripts of ACBD5 within the analyses between patients with retinal dystrophy, and those without. Only two transcripts were detected (other potential transcripts were therefore too low to detect) and these were at nominal levels (figure 3.4.26) in terms of transcripts per million (TPM) in both those patients with retinal dystrophy and those without

(figure 3.4.26). This suggests that NMD-escape is unlikely to be occurring and all ACBD5 transcripts across the six ACBD5-deficient cohort are being degraded via NMD. Comparing ACBD5 expression between cases and controls (n=33) showed an overall reduction in expression at the gene level (sum of all isoforms), with a fold-change of 0.17, representing more than a 5-fold reduction in expression (p= 6.32945E-07), confirming what was previously observed using Western blotting, at a protein level.

Gene Set Enrichment was performed using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) to look for enriched pathways within ACBD5-deficient patients and control, and also retinal-dystrophy phenotype and non-retinal dystrophy datasets. The top 200 ranked genes were included in both analyses. No significantly enriched (FDR > 0.05) pathways were present within either cohort (figures 3.4.27A and A). The top upregulated pathways enriched within the ACBD5-deficient patients and control dataset included coenzyme metabolic process (FDR = 0.54) that involves the genes dicarbonyl and L-xylulose reductase (*DCXR*), ecdysoneless cell cycle regulator (*ECD*), solute carrier family 2 member 1 (*SLC2A1*), solute carrier family 2 member 3 (*SLC2A3*), solute carrier family 25 member 1 (*SLC25A1*) and transketolase (*TKT*). Additionally, protein localization to endoplasmic reticulum (FDR = 0.39), that involves the genes golgi brefeldin A resistant guanine nucleotide exchange factor 1 (*GBF1*), ribosomal protein L10 (*RPL10*), ribosomal protein L13 (*RPL13*), ribosomal protein L14 (*RPL14*), ribosomal protein L27a (*RPL27A*), ribosomal protein L4 (*RPL4*), ribosomal protein S8 (*RPS8*) and signal sequence receptor subunit 3 (*SSR3*). The top downregulated pathways enriched within the data included regulation of transporter activity (FDR = 0.55) that includes the genes *DNM2,* and peroxisome proliferator activated receptor gamma (*PPARG*), and granulocyte activation (FDR = 0.39) that includes proteasome 26S subunit, non-ATPase 2 (*PSMD2*) and ubiquitin protein ligase E3 component n-recognin 4 (*UBR4*). Appendix table 8.3.1 shows the top 200 ranked genes, and appendix table 8.3.3 shows the full Gene Set Enrichment analysis results.

Within the dataset comparing patients with retinal dystrophy and without, the top upregulated pathways enriched within the data included regulation of protein catabolic process (FDR = 0.18) that includes the genes ubiquitin protein ligase E3A (*UBE3A*), proteasome 26S subunit, ATPase 3 (*PSMC3*), BTG3 associated nuclear protein (*BANP*), AKT serine/threonine kinase 1 (*AKT1*), casein kinase 2 alpha 1 (*CSNK2A1*), neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase (*NEDD4L*), and nardilysin convertase (NRDC), and positive regulation of transport (FDR = 0.37) that includes the genes AKT serine/threonine kinase 1 (*AKT1*), WNK lysine deficient protein kinase 1 (*WNK1*), *NEDDL4* and sirtuin 6 (*SIRT6*). The top downregulated enriched pathways included cellular amino acid metabolic processes (FDR = 0.04) that include the genes cystathionine-beta-synthase like (*CBSL*), serine hydroxymethyltransferase 2 (*SHMT2*), valyl-tRNA synthetase 2, mitochondrial (*VARS2*) and dihydrolipoamide S-succinyltransferase (*DLST*), and sulphur compound metabolic process (FDR = 0.05) that includes the genes cystathionine-beta-synthase like (*CBSL*), dihydrolipoamide S-succinyltransferase (*DLST*), glucose-6-phosphate dehydrogenase (*G6PD*),

diacylglycerol O-acyltransferase 2 (*DGAT2*), gamma-glutamyltransferase 5 (*GGT5*), hydroxysteroid 17-beta dehydrogenase 4 (*HSD17B4*) and ATP binding cassette subfamily D member 1 (*ABCD1*) (figure 3.4.28A). Appendix table 8.3.2 shows the top 200 ranked genes included in the analysis, and appendix table 8.3.4 shows the full Gene Set Enrichment data results.

Over-Representation Analysis was also performed using Web Gestalt, using the top 200 ranked genes across both datasets. In the ACBD5-deficient patients and control dataset 10 significantly enriched pathways were identified that included SRP-dependent co-translational protein targeting to membrane (FDR = 0.0009), that includes the genes encoding ribosomal protein L10, L11, L13, L14, L27A, L37A and L4, (*RPL10*, *RPL11*, *RPL13*, *RPL14*, *RPL27*, *RPL37A* and *RPL4*), ribosomal protein S8 (*RPS8*) and signal sequence receptor subunit 3 (*SSR3*), co-translational protein targeting to membrane (FDR = 0.0009), that includes the genes ribosomal proteins L10, L11, L13, L14, L27A, L37A, L4, *RPS8*, *SSR3*, and protein targeting to the ER (FDR=0.001) also including the genes ribosomal proteins L10, L11, L13, L14, L27A, L37A, *RPS8* and *SSR3* (figure 3.4.28B, see appendix table 8.1.5 for full results). Within the dataset comparing patients with retinal dystrophy and those without, three significantly enriched pathways (FDR $\leq$ 0.05) were identified that included developmental cell growth (FDR = 0.03) that includes the genes brain derived neurotrophic factor (*BDNF*), discoidin domain receptor tyrosine kinase 1 (*DDR1*), DExD-box helicase 39B (*DDX39B*), ephrin A5 (*EFNA5*), fibronectin 1 (*FN1*), glucose-6-phosphate dehydrogenase (*G6PD*) and *NEDD4L*, cell growth (FDR = 0.002) that includes the genes *AKT1*, *BDNF*, caprin family member 2 (*CAPRIN2*), cellular communication network factor 4 (*CCN4*), *CSNK2A1*, catenin beta 1 (*CTNNB1*), discoidin domain receptor tyrosine kinase 1 (*DDR1*), *DDX39B*, DEAD-box helicase 3 X-linked (*DDX3X*) and *EFNA5*, and regulation of cell growth (FDR = 0.02) that includes the genes *AKT1*, *BDNF*, *CAPRIN2*, *CCN4*, *CSNK2A1*, *DDR1*, *DDX39B*, *DDX3X*, *EFNA5* and *FN1*. Additionally, gene sets involved in oxidation of fatty acids, β-oxidation for fatty acids and lipid modification were enriched in ACBD5-deficient patients with retinal dystrophy, however these were not significant (figure 3.4.28B, see appendix table 8.3.6 for full results).

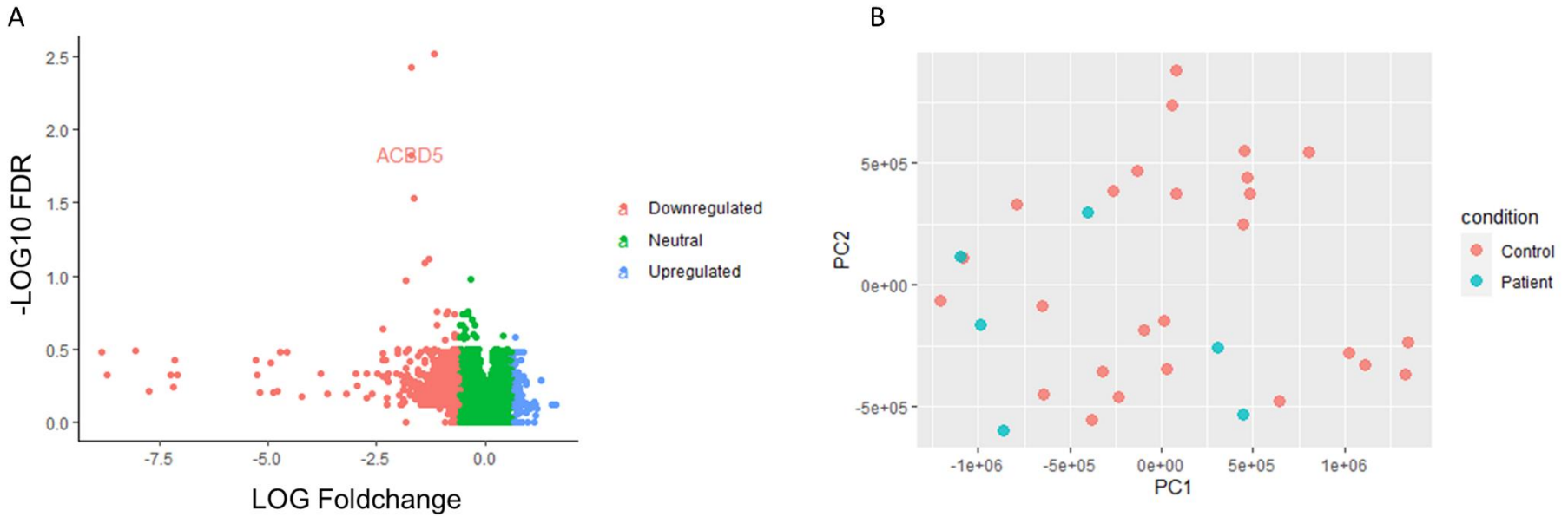


Figure 3.4.22: Differentially expressed gene analysis following RNA sequencing for controls and patients. A, Volcano plot showing differentially expressed genes in patients compared to controls. Y-axis shows LOG foldchange, and X-axis shows -LOG10 of corrected p value (FDR). Colours donate direction of gene expression change: upregulation, downregulation or neutral. B, PCA plot showing dimensionality reduction applied to RNA sequencing data, coloured by patients and controls. X-axis shows principal component one (PC1) and y-axis shows principal component two (PC2). *FDR, false discovery rate; PCA, principal component analysis*

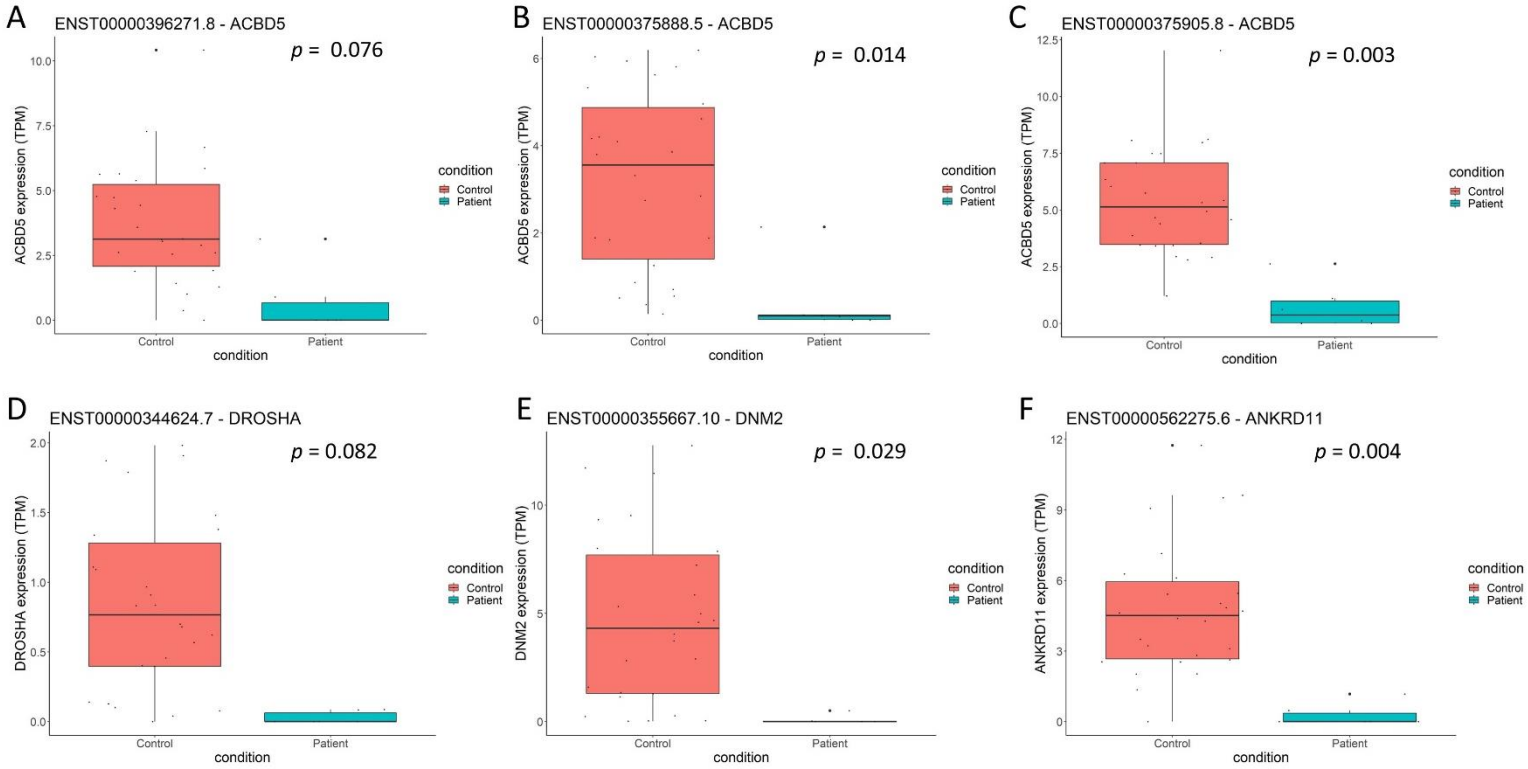


Figure 3.4.23: Differentially expressed gene transcripts in ACBD5-deficient patients, compared to controls. Gene transcripts ($p < 0.1$) downregulated in ACBD5-deficient patients, in comparison to controls. **A**, **B** & **C**, ACBD5 transcripts, **D**, DROSHA transcript, **E**, DNM2 transcript, **F**, ANKRD11. p values correspond to q value, after multiple testing correction using Benjamini–Hochberg correction. *ACBD5, Acyl-CoA Binding Domain Containing 5; DROSHA, Drosha Ribonuclease III; DNM2, Dynamin 2; ANKRD11, Ankyrin Repeat Domain 11; TPM, transcripts per million.*

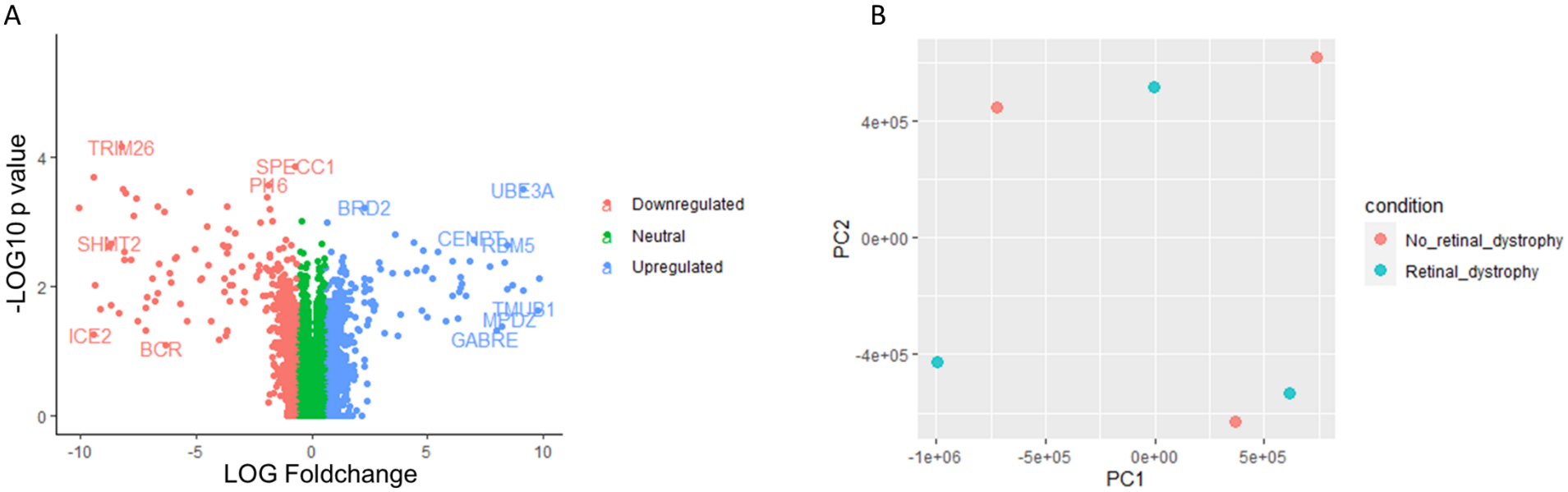


Figure 3.4.24: Differentially expressed gene analysis following RNA sequencing for patients with retinal dystrophy and those without. A, Volcano plot showing differentially expressed genes in patients with retinal dystrophy compared to patients without retinal dystrophy. Y-axis shows LOG foldchange, and X-axis shows -LOG10 of p value FDR. Colours donate direction of gene expression change: upregulation, downregulation or neutral. B, PCA plot showing dimensionality reduction applied to RNA sequencing data, coloured by patients without retinal dystrophy (No_retinal_dystrophy) and patients with retinal dystrophy (Retinal_dystrophy). X-axis shows principal component one (PC1) and y-axis shows principal component two (PC2). *PCA, principal component analysis*

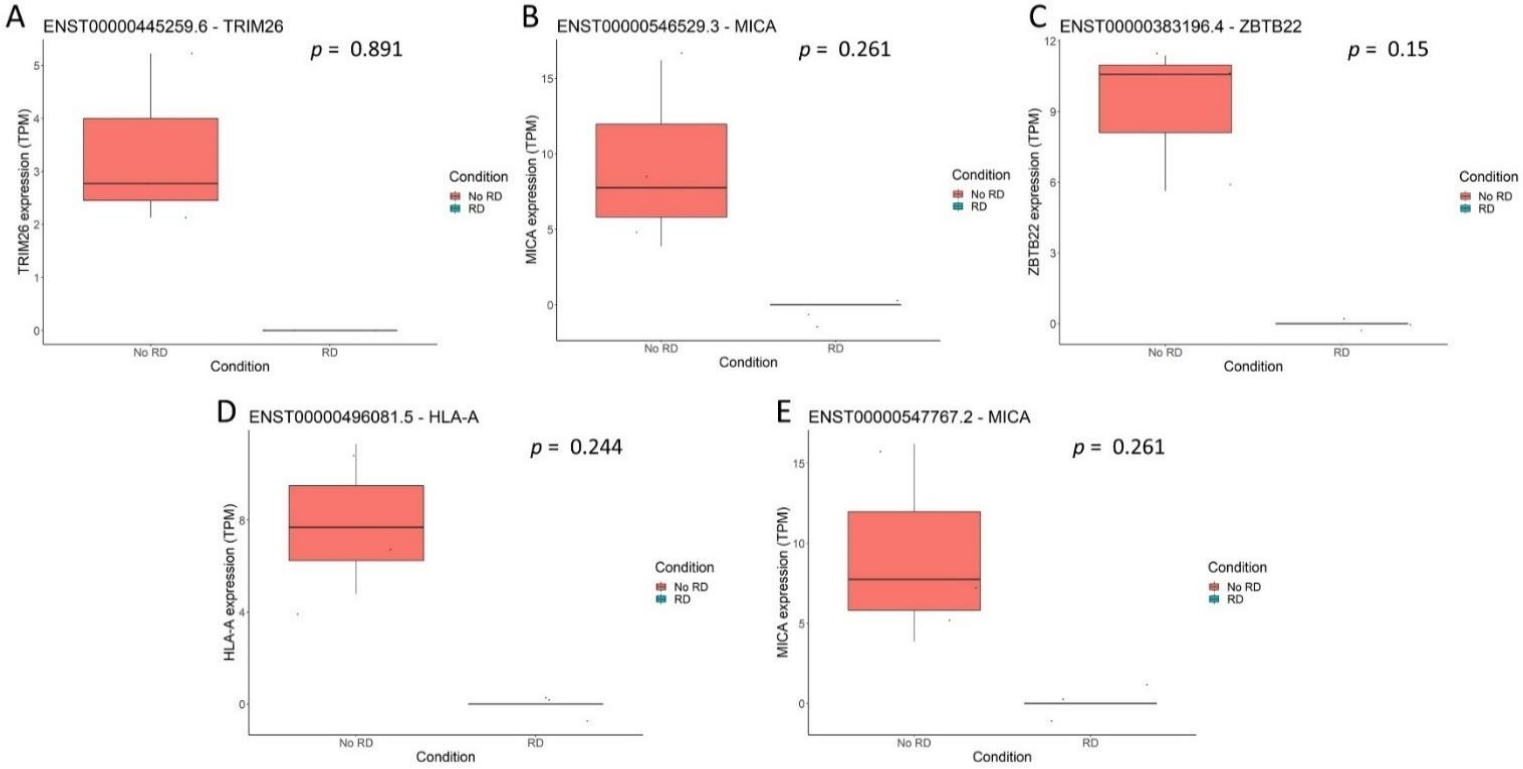


Figure 3.4.25: Top 5 differentially expressed genes between ACBD5-deficient patients with retinal dystrophy and those without. Top 5 ($p < 1.0$) gene transcripts ($p < 1.0$) differentially expressed in ACBD5-deficient patients with retinal dystrophy, and those without. **A**, *TRIM26* ($p = 0.891$), **B**, *MICA* ($p = 0.261$), **C**, Z*BTB22* ($p = 0.15$), **D**, *HLA-A* ($p = 0.244$), **E**, *MICA* ($p = 0.261$). p values correspond to q value, after multiple testing correction using Benjamini–Hochberg correction. *TRIM26, Tripartite Motif Containing 26; MICA, Human Major Histocompatibility Complex (MHC) Class I Chain-Related gene A; ZBTB22, Zinc Finger And BTB Domain Containing 22, TPM, transcripts per million; RD, retinal dystrophy.*

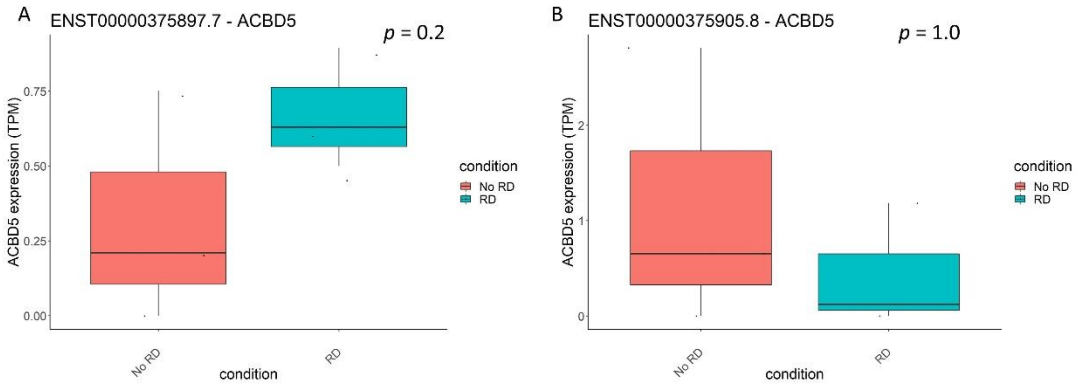


Figure 3.4.26: ACBD5 transcript expression in patients with retinal dystrophy and those without. *ACBD5* transcript ENST00000375897.7 (**A**, *p*=0.2) and ENST00000375905.8 (**B**, *p*=1.0) expression in ACBD5-deficient patients with retinal dystrophy (RD), and those without (No RD). p values correspond to q value, after multiple testing correction using Benjamini–Hochberg correction. *ACBD5, Acyl-CoA Binding Domain Containing 5, TPM, transcripts per million; RD, retinal dystrophy.*

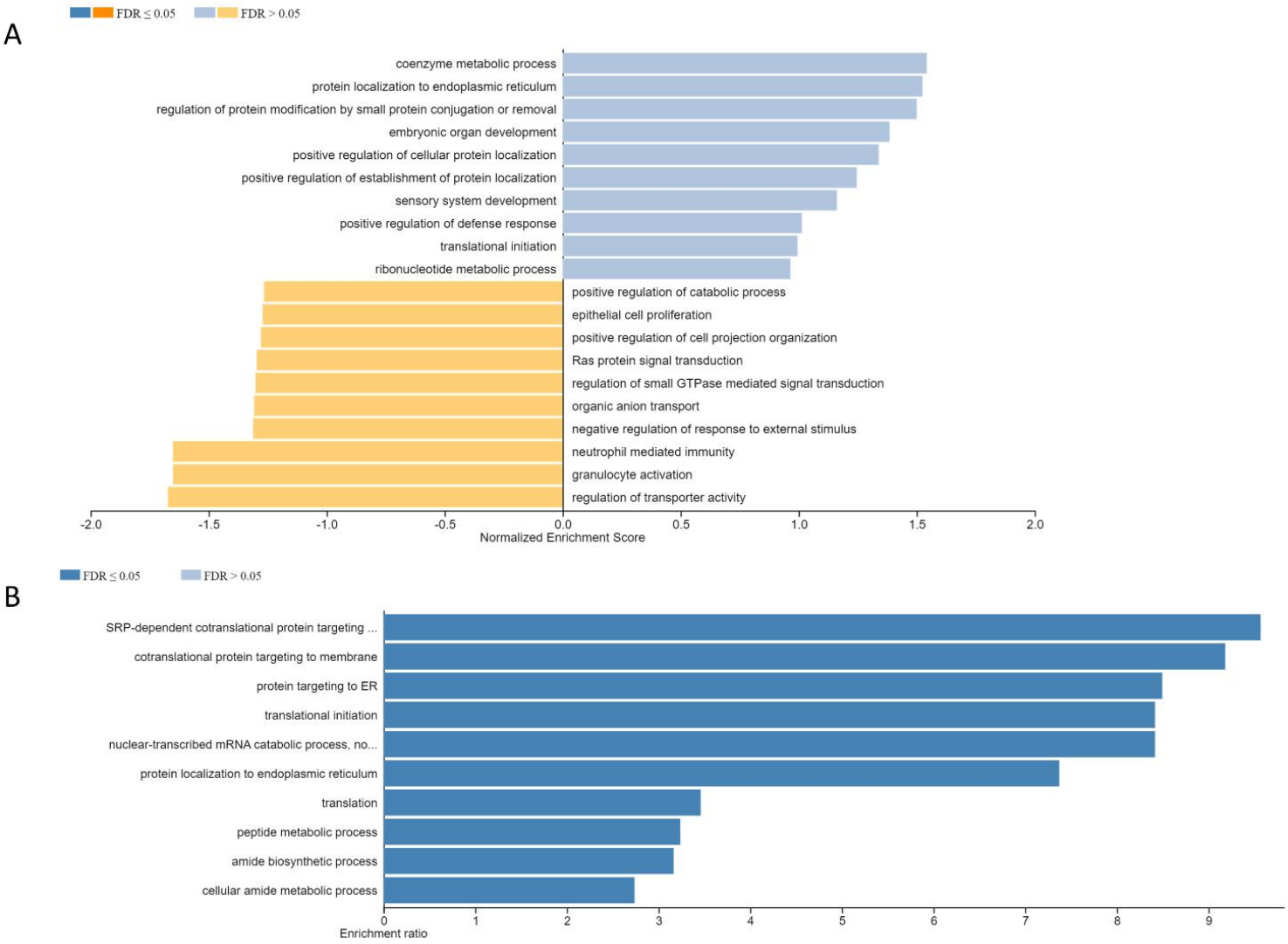


Figure 3.4.27: Gene Set Enrichment Analysis and Over Representation Analyses (ORA) for ACBD5-deficient pathways compared to controls. A, Gene set enrichment analysis results showing top upregulated (positive normalised enrichment score) and downregulated (negative normalised enrichment score) pathways. B, ORA analysis results showing top 10 significantly enriched pathways, based on FDR. FDR ≤ 0.05 (dark blue and dark orange), FDR > 0.05 (light blue and light orange). Analysis performed using WebGestalt. *FDR, false discovery rate; ORA, Over Representation Analysis.*

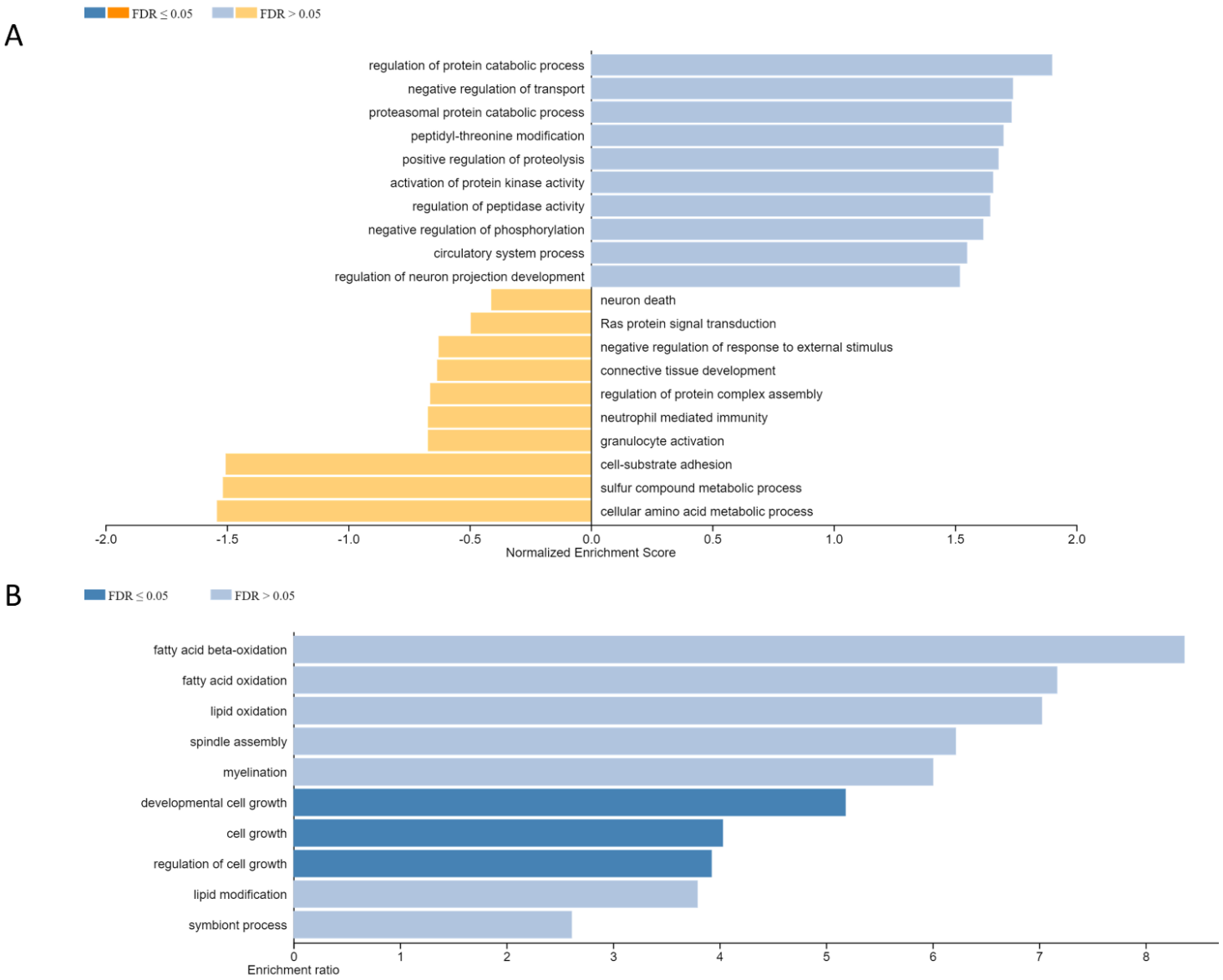


Figure 3.4.28: Gene Set Enrichment Analysis and Over Representation Analyses (ORA) for ACBD5-deficient patients with retinal dystrophy and those without. Gene set enrichment analysis results showing top upregulated (positive normalised enrichment score) and downregulated (negative normalised enrichment score) pathways. B, ORA analysis results showing top 10 enriched pathways, three of which are significant based on FDR. FDR ≤ 0.05 (dark blue and dark orange), FDR > 0.05 (light blue and light orange). Analysis performed using WebGestalt. *FDR, false discovery rate; ORA, Over Representation Analysis.*

3.4.5. Discussion

Two patients presenting with pure ataxic symptoms were identified, through WES, to have a recessive mutation in *ACBD5* which codes for the protein ACBD5, a peroxisomal membrane protein. Through multiple international collaborations, I have collected a further four confirmed ACBD5-deficient patients from four families. These patients present with a phenotype that includes retinal dystrophy which has also been reported in two previous cases: both with different variants. Both have differing clinical presentations, although ataxia is a recurring feature in both cases (145, 394). Interestingly there is no evidence of retinal dystrophy in the two original patients from the NHNN, which is a feature in all other cases currently. In addition, there seems to be differences in mitochondrial assay results between the patients that present with retinal dystrophy and those patients who do not, with a more severe effect on mitochondrial function seen in the latter group.

Using sanger sequencing I segregated the mutation within the family, confirming the biallelic loss of AATG in both patients 1 and 2. Characteristic of a carrier, both parents show a normal DNA transcript up to the point of mutation, whereby two alleles appear, showing the normal and affected allele. I have shown evidence that the recessive AATG deletion in exon 7, causes nonsense-mediated decay, which is a pathway that recognises and degrades mRNAs through translation termination codons that appear through abnormal contexts. These premature termination codons (PTCs) appear commonly as a result of frameshift and nonsense mutations, producing transcripts that undergo NMD and ultimately generating a loss of function phenotype (396). This was re-confirmed using cDNA sequencing, which showed that the carrier expresses only the normal mRNA transcript, identical to that of the controls, as the abnormal transcript has undergone degradation by means of NMD. Interestingly, all four other patients collected for this study have different mutations in *ACBD5*, affecting various domains of the gene, but all causing loss-of-function.

To confirm the loss of function phenotype across all patients, I performed Western blotting initially using only fibroblast-derived protein from patient 1, and then on all patient fibroblast lines, which showed a clear total absence of protein for the patients, compared to the control which showed a clearly defined band at 55kDa. In addition, immunofluorescent staining using anti-ACBD5 antibodies showed a complete absence of ACBD5 protein (seen as green, fluorescent puncta) in patient 1, compared to control. This suggests a loss of function phenotype, due to NMD. This was further confirmed using RNA sequencing data, where each patient had nominal TPM for *ACBD5*. Loss of function phenotypes are frequently implicated in both ataxias and other PEDs, with loss of function mutations in *FXN* being one of the main causes of FRDA, the most common recessive ataxia. Other examples include the loss of function of GLS, which codes for glutaminase, in recessive spastic ataxia (397), small vasohibin binding protein (SVBP) in recessive intellectual disability with ataxia (398), calcineurin like EF-hand protein 1 (CHP1) in complex recessive ataxia (399) and spectrin repeat containing nuclear envelope

protein 1 (SYNE1) in pure progressive ARCA (400). In addition the loss of ABCD1 causes ALD, which like ACBD5-deficiency, is the loss of a peroxisomal membrane protein (401).

ACBD5 is a peroxisomal membrane protein that currently does not have a fully defined function, however during the course of this research multiple studies were published that define a possible mechanism for ACBD5-deficiency. Peroxisomes have integral roles in essential metabolic pathways including the α- and β-oxidation of fatty acids, ROS generation and ether phospholipid biosynthesis (367). Peroxisomal impairment due to defects in peroxisomal proteins can cause peroxisomal disorders, of which many are documented and fall into two main classes: the PBDs and the single PEDs (402). Examples of these are the ZSDs and ALD, the latter of which is caused by a loss of ABCD1, another peroxisomal membrane protein that facilitates very-long chain fatty acid uptake into peroxisomes.

VLFCA elevations is a feature of PEDs caused by impaired β-oxidation, including ALD which utilises raised VLCFA levels to clinically diagnose it, with typically raised C26:0, C24:0 and C26:0/C22:0 ratio levels (403). In addition, patient 3 and patient 4, were reported as featuring an accumulation of VLCFA levels as a defining feature, suggesting impairment of β-oxidation of VLCFAs (145, 146). To explore whether the other four patients also have an accumulation of VLCFA levels, and therefore a possible impairment of VLCFA β-oxidation, I sent fibroblasts of patients 1, 2, 4 and 6 , and also parent carriers 1 and 2, to another one of my collaborators, Dr Sacha Ferdinadusse in Amsterdam. Dr Ferdinandusse has successfully set up a sensitive metabolic assay in order to detect VLCFA in fibroblasts. All patients had similarly increased VLCFA levels, with increases in C26:0 lysophosphatidylcholine (lyso PC), D3-C26 levels, hexacosanoic acid (C26:0) levels and C26:0/C22:0 and slightly raised C24:0/C22:0 ratio levels. These results all point to impaired β-oxidation of VLCFA, as the C26:0 level was increased, as well as C26:0/C22:0 ratio fraction. A D3-C26:0 loading study was also performed, whereby deuterium-labelled C22:0 (D3-C22:0) was incubated with the fibroblasts for three days, followed by tandem mass spectrometry fatty acid analysis. D3-C26 levels were increased in patients, indicating increased chain elongation from the substrate D3-C22:0, likely due to increased substrate availability from impaired breakdown of D3-C22:0. Peroxisomes are exclusively able to β-oxidise C22:0, C24:0 and C26:0, so the increased levels in all patients, compared to controls and the carriers which were in normal control range, indeed suggest impaired VLCFA metabolism, particularly β-oxidation. The end products of peroxisomal β-oxidation are transferred to the mitochondria as either carnitine ester or as a free ester, where they undergo final oxidation to CO_2 or H_2O (401).

One hypothesis is that ACBD5 has a role in binding through its acyl-coA domain to fatty acids such as C26-CoA in the cytosol and then passing it onto ABCD1 for transportation into the peroxisome for degradation through β-oxidation (figure 3.4.29) (146). Interestingly, X-linked ALD patients with ABCD1-deficiency have similarly increased C26:0 levels compared to ACBD5-deficient fibroblasts (145) making a potential overlap in mechanism possible. The accumulation of VLCFAs due to impaired

β-oxidation could cause toxicity to the cell and several different key biological processes that peroxisomes carry out. In order to explore any potential mechanistic overlap between ABCD1- and ACBD5-deficiency, an ALD fibroblast line, with loss-of-function in ABCD1, was used as a control line in assays to test ROS production, lipid peroxidation and glutathione levels ([GSH]) which tests for mitochondrial oxidative stress. The results for both experiments suggest that there is not a potential overlap in mechanisms, as the ALD fibroblast line had an increase in oxidative stress from GSH concentration assay. However, a limitation of this study is that only one ALD patient was used as a positive control to test this, therefore observed differences in phenotypic assays cannot be formally tested and future replications should include additional samples.

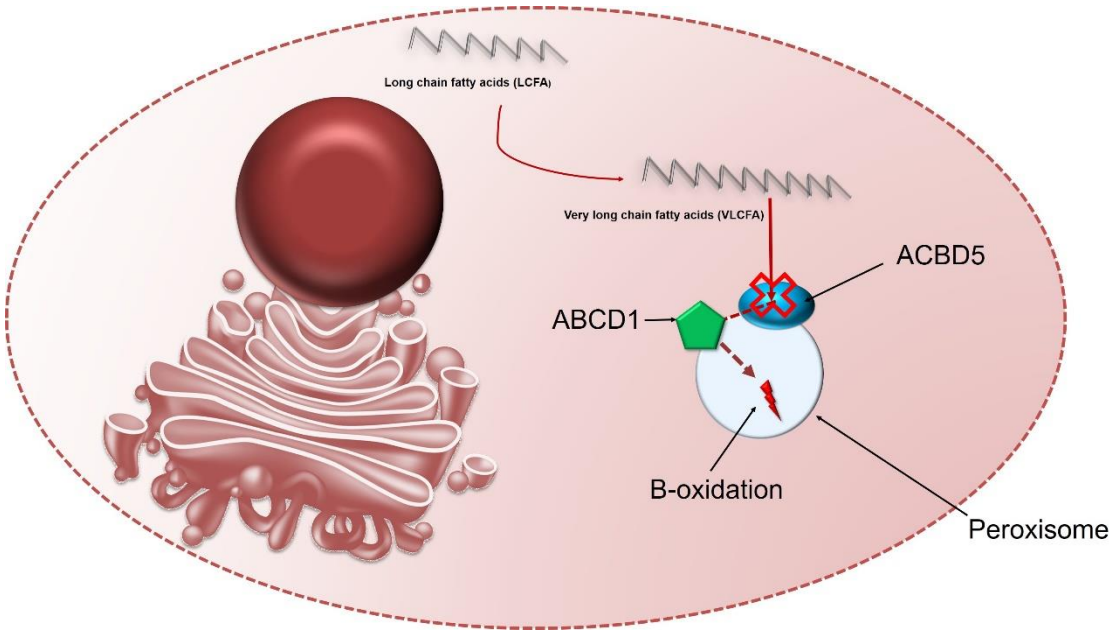


Figure 3.4.29: Schematic of potential peroxisomal β-oxidation impairment. One hypothesis is that ACBD5 binds to fatty acids in the cytosol via its acyl-coA domain, passing them onto ABCD1, a VLFCA-transporter, gaining them access into the peroxisome for subsequent β-oxidation. Therefore, without ACBD5, VLFCAs do not enter the peroxisome and accumulate in the cytosol, creating cell toxicity. *ABCD1, acyl-CoA binding protein 1; ACBD5, acyl-CoA binding protein 5*; *LCFA, long-chain-fatty-acids; VLCFA, very-LCFA*

Oxidative stress occurs from increases in the concentration of ROS and RNS, and occurs naturally due to IV irradiation, inflammation, exposure to environmental toxins and oxidants and heat shock, amongst other sources (404). It has been proven to be a pathogenic contribution to a variety of neurological diseases including PD and Alzheimer's disease, as well as other pathologies including atherosclerosis, cystic fibrosis, cancer and inflammation (405). However, ROS also function as mediators in multiple cellular processes and signalling pathways such as apoptosis (406). Peroxisomes play a key role in the

scavenging and production of ROS , consuming high levels of O_2 and producing H_2O_2, ·OH, $O_2\cdot^-$ and ·NO (407-409). The principle sources of endogenous ROS production are in the peroxisomes, mitochondria, plasma membrane and ER (410). Due to the interplay between mitochondria and peroxisomes, and the established impairment in fatty acid metabolism in ACBD5-deficient peroxisomes, we hypothesised that there may be an increased level of mitochondrial ROS, and subsequent increase in oxidative stress, which could cause degenerative downstream effects. Patients 1, 2, 3 and 4 all had an increased fold change of mitochondrial ROS production rate, compared to controls, as did the positive ALD control. However, the results were not significant and greatly varied at both a control and patient level. The patients with retinal dystrophy appeared to have a slightly higher ROS production rate, however this was not significant. These data suggest that there could be a trend of a slight increase in mitochondrial ROS production in ACBD5-deficient fibroblasts and ALD control, but we cannot conclude that this is linked to ACBD5 loss-of-function due to the lack of significance and inter-sample variability. Additionally, we cannot conclude that ABCD1- and ACBD5- deficient fibroblasts have similar levels of oxidative stress, with regard to a possible overlap in mechanism. Excessive ROS production can disrupt multiple cellular processes, causing cellular damage including damaging the lipid membrane and increasing the permeability and fluidity of the lipid membrane (411). It can also lead to protein damage in the form of peptide chain fragmentation, enzyme deactivation and amino acid modifications at specific sites, and DNA damage through deoxyribose oxidation, base modification, strand breakage and removal of nucleotides (411, 412). However, significant ROS overproduction was not found in patient fibroblasts, compared to controls, suggesting that these downstream mechanisms may not be contributing to pathology in the case of ACBD5-deficiency.

Lipid peroxidation is a process in which oxidants target lipids that contain carbon-carbon double bonds and insert oxygen, which produces hyperoxides and lipid peroxyl radicals. It has been shown that ROS overproduction leads to the oxidation of a variety of different molecules that include lipids, proteins, and DNA (413). Therefore we hypothesised that oxidative stress and ROS production might be elevated in ACBD5-deficient fibroblasts, and if so, would also increase rates of lipid peroxidation, as free radicals involved in oxidative stress are pivotal to the chemistry of lipid peroxidation (412). Similarly, to the ROS production assay, the patient fibroblasts showed a great deal of variability between controls and between patient lines. Findings show a trend of increased lipid peroxidation rate in the ALD positive control, and no increase of ACBD5-deficient fibroblasts. Increases in lipid peroxidation rates have been reported in X-linked ALD symptomatic patients, suggesting that an increase in oxidative stress could contribute to the pathophysiology of X-linked ALD, which fits into the elevated peroxidation rate in the ALD positive control (414).

Interestingly the variability seen in lipid peroxidation rates between fibroblast lines, especially control lines, could be due to diet and lifestyle. Smoking, inadequate, or inappropriate diets and exercise have all been linked to oxidative stress levels. Overtraining or performing exhausting long exercise can

induce a significant response to oxidative stress, whereas moderate exercise with low-intensity training has been shown to improve endogenous antioxidant status (411). Cigarette smoke also increases the production of local oxidants, due to it comprising a series of free radicals, oxidants and organic components which accumulate in the lungs (415). Diet also had a significant effect on the ability to combat oxidative stress and prevent associated damage. Exogenous antioxidants have to be ingested through the diet, as they cannot be synthesised in eukaryotic cells. Important antioxidants include polyphenols and vitamin C which are water soluble antioxidants and can be absorbed by the body through vegetables and fruit (416). The lifestyle status is unknown for the three controls used within this study, so it is not known if they smoke, maintain a healthy diet and exercise regularly. However, controls 1 and 3 have a trend of a raised rate of lipid peroxidation, compared to control 2, suggesting higher levels of mitochondrial oxidative stress. Patients 1 and 2 have very different lifestyles, with patient 1 making a conscious effort to maintain an exercise regime and eat a healthy diet in order to help manage their symptoms. Contrastingly, patient 1's affected sibling, patient 2, does not maintain a particularly healthy diet or exercise regime, and is classified as obese. Interestingly this is reflected in the lipid peroxidation rates, where patient 1 has a trend of a lower peroxidation rate, compared to patient 2.

During this study, several important articles were published that shone some more light on the function of ACBD5 as a component of peroxisomes. Peroxisomes have been shown to interact with the ER, through tail-anchored ER membrane protein VAMP Associated Protein B (VAPB), which binds directly to ACBD5 (395, 417). ACBD5 has been found to be an important component of peroxisomal motility and formation of peroxisomal-ER tethers, which are crucial for lipid transfer. The transfer of lipids between peroxisomes and the ER is crucial to peroxisomal function, with peroxisomes depending on the ER for lipid composition, and membrane proteins (379). Likewise, there is crucial trafficking and transfer of lipids and lipid precursors from the peroxisome to the ER. Plasmalogen synthesis begins in the peroxisome, and is completed in the ER, and additionally cholesterol precursors are also thought to be synthesised in peroxisomes (376, 418, 419). Plasmalogens are a class of membrane glycerophospholipids that comprise 20% of total plasmalogen mass in humans. Plasmalogen associated defects cause the malformation syndrome Rhizomelic Chondrodysplasia Punctata (RCDP), which leads to defects in bone, brain, lung, kidney and lens development (418).

Reports have shown that *ACBD5*-mutants, that localise to mitochondria, increase association of mitochondria to VAPB-labelled ER (395). VAPB is a ER tail-anchored protein that is involved with membrane tethering and inter-organellar lipid exchange (420). A major sperm protein (MSP) domain on VAPB facilitates its interactions by recognising FFAT-like motifs (two phenylalanines (FF) in an acidic tract). Importantly ACBD5 has an FFAT-like motif, which is thought to bind to the VAPB domain (421). Mutated *ACBD5* lines that did not contain an FFAT-like-motif prevented VAPB-ACBD5 interaction, supporting this theory (395). Discrete peroxisomal structures have been shown to form

between peroxisomes and the ER, when ACBD5 and VAPB are co-expressed, suggesting a close peroxisomal-ER contact. When *ACBD5*-mutated lines that do not have a FFAT-domain were co-expressed with VAPB, this effect was lost, suggesting that the FFAT-like domain is crucial for this peroxisomal-ER tether (395). Recent work has highlighted that these peroxisomal-ER tethers are regulated by phosphorylation, specifically of a serine-269 residue within the ACBD5 FFAT motif (422).

Furthermore, studies have shown that the loss of ACBD5 causes an increase in peroxisomal motility, a reduction of peroxisomal-ER contacts and shorter peroxisomal membrane tubules, suggesting it has a role in peroxisomal membrane elongation (395). It is suggested that the reduction of membrane lipid transfer from the ER could contribute to the impact on peroxisomal elongation (395). The same study also found that ACBD5 is not crucial for peroxisomal biogenesis, with control fibroblasts showing ACBD5 localisation with PEX14, which was completely absent in ACBD5-deficient lines. Remarkably however, PEX14-positive structures were seen in ACBD5-deficient lines, as well as other peroxisomal matrix proteins and import machineries, suggesting that they are unaffected in the absence of ACBD5. This is also seen in the work carried out by Prof. Schrader's group in Exeter University, with ACBD5-deficient patients without the retinal dystrophy phenotype showing no differences in peroxisomal numbers and no reduction of PEX14-structures. Additionally, the Schrader group also showed that the ACBD5-deficient patients without retinal dystrophy had reduced peroxisomal-ER contact and attachment, similarly to ACBD5 knockout HeLa cell lines, previously published (395). This suggests that the patient group without retinal dystrophy do not have any additional mechanistic pathogenesis that may perturb peroxisomal biogenesis and share the reduction in peroxisomal-ER tether formation, attachment and increase in peroxisomal motility shown in ACBD5 knockout HeLa cell lines.

Peroxisomal membrane dynamics and interactions are important for peroxisomal biogenesis, with lipid exchange from the ER important for the growth, elongation, and eventual fission of the peroxisomal membrane (417). This exchange could be mediated by the peroxisomal-ER tethers, with the loss of ACBD5 directly affecting peroxisomal membrane elongation (figure 3.4.30) (395). The loss of lipid exchange between the ER and peroxisomes, due to the increase in peroxisomal motility, loss of peroxisomal-ER tether attachment and perturbed peroxisomal membrane exchange caused by ACBD5-deficiency likely causes great dysfunction to peroxisomal lipid metabolism and biogenesis and could contribute to the impaired β-oxidation of VLCFAs seen in all ACBD5-deficient patients. Additionally, the increased association of ACBD5-mutants with the mitochondria, could potentially result in mitochondrial dysfunction.

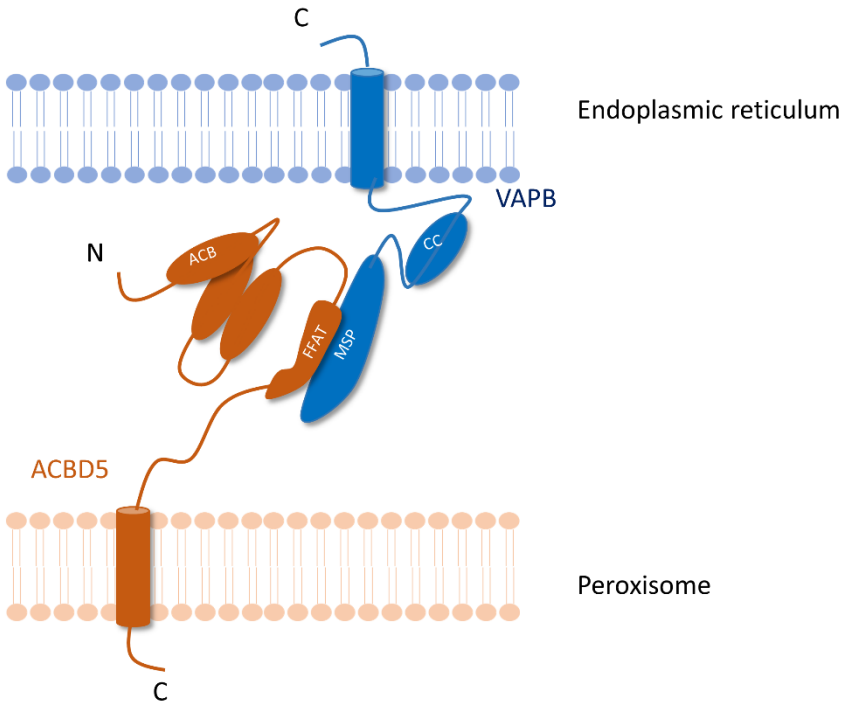


Figure 3.4.30: Schematic of ACBD5-VAPB mediated peroxisomal tether with the ER

VAPB, Vesicle-associated membrane protein-associated protein B/C; ACBD5, Acyl-coA binding domain containing 5, N, N-terminus; C, C-terminus; ACB, acyl-coenzyme binding domain; FFAT; FFAT-like motif, MSP, major sperm protein; CC, coiled-coil region

ACBD5, and VAPB, a tail-anchored ER membrane protein bind via their FFAT and MSP binding sites, respectively. This binding facilitates the formation of discrete peroxisomal-ER tethers that suggest a close peroxisomal-ER contact that could mediate the exchange of lipids and metabolites between the two organelles.

The role of peroxisomes permits exchange of metabolites and lipids that require signalling and interaction of peroxisomes with other organelles including the mitochondria. The products of β-oxidation are transferred to the mitochondria for a final round of oxidation, and metabolic screening has shown that the β-oxidation of VLCFAs is impaired in all ACBD5-deficient patients. To explore any effects of ACBD5-deficiency on mitochondrial function, mitochondrial tests such as TMRM and NADH/NAD+ autofluorescence were carried out.

Mitochondrial dysfunction is a feature of multiple diseases that feature neurodegeneration including Alzheimer's disease, HD and PD and therefore could also be associated with ACBD5-deficiency due to the cerebellar atrophy seen on MRI for the majority of patients (423). In order to control for the mitochondrial phenotype seen and ensure it is not a secondary effect of disrupted metabolism in the patients, or degeneration, a ZSD positive control was also added to the cohort. ZSD patients are reported

to have secondary mitochondrial changes, due to their peroxisomal defect and disturbed functional interplay between peroxisomes and mitochondria, therefore impaired mitochondrial function is expected (424). Preliminary studies were first carrier out on patients 1, 2 and 3 using NADH autofluorescence, which examines respiratory function. Results revealed intense patches of NADH in patients 1, 2 and 3, which were absent in controls and carrier. These patches also stayed relatively consistent throughout addition of FCCP and NaCN. Patients and carriers had an increased redox index, therefore a more reduced redox state (an increase in NADH, compared to NAD^+), than controls. This suggested that complex I-driven respiration may be impaired in control and carrier fibroblasts compared to controls.

A repeat of NADH autofluorescence was carried out on patients 4, 5 and 6 and, as in the preliminary experiments, controls, had an increased redox index (i.e. more reduced redox state) compared to patients, however findings were not significant. A significantly reduced NADH Pool was also observed in patients compared to controls. Importantly the positive ZS control had a significantly increased redox index, compared to controls, showing a more severe impairment of mitochondrial respiration.

As peroxisomes do not possess a respiratory chain, they rely on redox shuttles composed of cytosolic and peroxisomal malate dehydrogenases to regenerate NAD^+ from NADH (425). Impaired peroxisomal function could lead to an accumulation of NADH in mitochondria if this redox shuttle is also disrupted, which could explain the patches of NADH seen on the NADH auto-fluorescence assay. Another theory that would fit into this reasoning is that peroxisomal β-oxidation produces NADH as a by-product, therefore impaired β-oxidation could produce less NADH and negate the requirement, or at least reduce it, for the redox shuttles (401).

Mitochondrial membrane potential $\Delta\Psi_m$ was also measured, which is a good indicator of the health of cells and could also be affected by peroxisomal dysfunction due to the peroxisomal-mitochondrial interactions and exchange of metabolites (376, 378). Preliminary experiments using patients 1, 2 and 3 were conducted prior to the extensive collaboration. Patient fibroblasts had a statistically significant slight reduction in basal $\Delta\Psi_{m.}$ as measured by higher TMRM signal, than control and carrier fibroblasts. In healthy mitochondria, the $\Delta\Psi_m$ is maintained by pumping protons from the mitochondrial matrix to the intermembrane space through action of the electron transport chain (ETC). Complex V was first tested, which pumps protons back to the matrix and can pathologically work in reverse, pumping protons back into the intermembrane space therefore reducing the energy derived from the protein-motive force that is used to generate ATP. Oligomycin, an inhibitor of complex V, did not cause much of a change in TMRM signal in any of the cell lines, suggesting that complex V is not working in reverse. Rotenone, an inhibitor of complex I, was then added, which caused depolarisation in all cell lines, however the reduction was less pronounced in all ACBD5-deficient fibroblast lines and the carriers, suggesting potential impairment of carrier I. FCCP induced a similar effect of complete

depolarisation in all fibroblast lines, which may suggest that complex II is overcompensating for the possible impairment of complex I in the patient and carrier fibroblasts. An impairment of complex I would lead to an accumulation of NADH, and a more reduced redox state, which is exactly the trend we see from the NAD^+/NADH autofluorescence assay in all ACBD5-deficient patients.

In order to validate these findings, this assay was repeated on three more ACBD5-deficient fibroblast lines that were kindly sent by our global collaborators, patients 4, 5 and 6. Similarly a ZSD positive control was also added to the cohort in order to investigate if any changes seen were likely a result of the ACBD5-deficiency, or a secondary effect of impaired metabolism. The previous finding of a lower basal $\Delta\Psi_m$ control fibroblasts was replicated in two of the three patients, patient 6 presented with a higher basal $\Delta\Psi_{m,}$ although none of these findings were significant, and when all patients were compared against each other, basal $\Delta\Psi_m$ was extremely varied.

Live imaging was also repeated on the three additional patients to validate TMRM intense changes over time. The same result was seen as the preliminary study, with patients 4, 5 and 6 all showing less of a decrease in membrane depolarisation after addition of rotenone, an inhibitor of complex I. Patient 6 and the ZSD positive control showed a significant difference to controls, suggesting a more severe complex I impairment. Importantly, the ZSD positive control also had a lower basal $\Delta\Psi_{m,}$ and a significant decrease in response to rotenone in the live imaging, compared to controls. This suggests that the mitochondrial phenotype seen in the patients could be a secondary downstream effect of peroxisomal dysfunction, and not directly caused by the loss of ACBD5. A limitation of this aspect of the study is the variability of fibroblasts and in addition the singular ZSD positive control that was included. In order to draw conclusive comparisons between the ZSD positive control and the ACBD5-deficient patients, a larger cohort of ZSD controls should be added.

Glutathione, also known as γ-l-glutamyl-l-cysteinyl-glycine, is a tripeptide, and a vital antioxidant. It has a crucial role in the removal of reactive species including ROS and the ratio of GSH:GSSH (GSSH is the oxidised form) is a good indicator of the oxidative stress levels within a cell. Monochloramine (MCB) was used to measure cellular glutathione [GSH] levels, as MCB enters the cell and forms a fluorescent GSH-MCB adduct, catalysed by glutathione S-transferase, that is then measured fluorometrically (426). The ratio of reduced glutathione (GSH) to oxidised glutathione (GSSH) is a measure of cellular oxidative stress, where the oxidation of free and protein bound GSH to GSSH by glutathione reductase is induced by oxidative stress (427). Interestingly all ACBD5-deficient patients and the ZSD positive controls had significantly elevated levels of GSH compared to controls, suggesting lower levels of oxidative stress, or reactive oxygen species. Contrastingly the ALD positive control had significantly reduced levels of GSH (a higher level of GSSH to GSH) compared to controls and also the ZSD positive control, suggesting higher levels of oxidative stress. It has already been shown that X-linked ALD lymphoblasts and erythrocytes have reduced glutathione levels, supporting these

findings (428). One theory of why the ZSD positive control and ACBD5-deficient fibroblasts showed elevated levels of GSH is the potential impairment in complex I that was found through the TMRM mitochondrial assay. A large producer of intracellular ROS is the mitochondrial respiratory chain, specifically electron transfer. Two key sources of this production are complex I (NADH:ubiquinone oxidoreductase) and complex III (ubiquinol:cytochrome *c* oxidoreductase). Complex I in particular generates H_2O_2 and $O_2\cdot^-$ within the intermembrane (429). Therefore, an impairment of complex I (and compensatory effect of complex II indicated from addition of FCCP) would cause a reduction in H_2O_2 and $O_2\cdot^-$ production, and therefore less mitochondrial oxidative stress. The insignificant ROS production rate and lipid peroxidation findings also support this theory. Contrastingly, the ALD positive control showed increased rates of ROS production, lipid peroxidation and reduced GSH, all supportive of increased levels of oxidative stress and mitochondrial free radicals. In addition, the patients without retinal dystrophy had higher GSH levels than those with retinal dystrophy, furthering the distinct differences between these two phenotypic groups.

The clinical phenotypic spectrum of these ACBD5-deficient patients is extremely broad, with one of the main phenotypic differences being the presence of retinal dystrophy in half of the cases. Retinal dystrophy has been previously described in 5 patients from 3 families that also had leukodystrophy; three siblings of a Saudi family with a truncating mutation, patient 3 from this study with a homozygous frameshift mutation, and patient 4 from this study, a 36 year old Brazilian woman with a nonsense variant, presenting with leukodystrophy and retinal dystrophy (145, 146, 394). The remaining three patients within this study (patients 1, 2 and 6) do not present with retinal dystrophy. Retinal dystrophy has a range of clinical presentations that are reported to affect 1 in 4000 (430). Cases can cause varied vision impairment, from poor night vision or poor peripheral vision to complete blindness, and progressive severity is reported to increase with age (431). Retinal dystrophy is characterised into different groups by the manifestation and degree of retinal atrophy and by the type of photoreceptor affected.

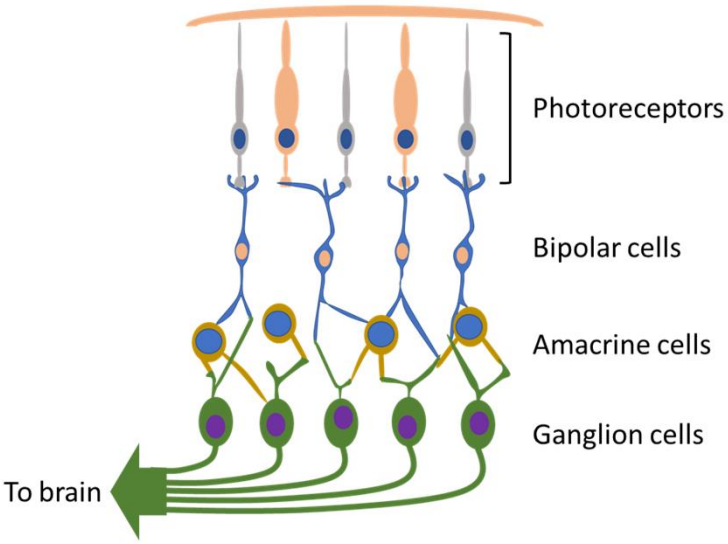


Figure 3.4.31: Schematic of retinal layer structure. Visual information is submitted via an excitatory vertical chain, photoreceptors, to bipolar cells, to ganglion cells. The amacrine cells provide luminance gain control. These layers are affected during retinal dystrophy, impairing vision.

An excitatory vertical chain from photoreceptors to ganglion cells enables the submission of visual information (figure 3.4.31). Rod and cone photoreceptors are the principal cellular components that enable the conversion of light energy to an excitatory neural action potential within the retina; this process facilitates an image to be understood by the brain. Retinal dystrophy can encompass diseases that affect predominantly cones, rods or a generalised degeneration that involves both (431). Interestingly half of the patients within this cohort presented with some form of retinal dystrophy (patients 3, 4 and 5), whereas the other half presented with a more severe ataxic phenotype (patients 1, 2 and 6). These clinical findings suggest that there is a clinical spectrum and variable expressivity associated with ACBD5-deficiency. There have been previous reports of clinical variability in PEDs that involve β-oxidation of VLCFAs. Furthermore, neurological and vision impairments, as well as white matter disease, have been reported a shared clinical presentation within these PEDs that involve β-oxidation of VLCFAs (401). White matter disease and retinal dystrophy are a feature in three of the six patients within this study. Therefore, the additional cases (patients 1, 2 and 6) represent a novel phenotype associated with ACBD5-deficiency and PEDs that involve β-oxidation of VLCFAs, as well as a broadening of the associated clinical spectrum associated with ACBD5-deficiency. One limitation was that detailed clinical information for patient 5 was unable to be shared by the consulting collaborating clinician due to delays associated with the Covid-19 pandemic, however the presence of retinal dystrophy and ataxia were confirmed. This clinical information will be included in the future publication of this work.

Remarkably, there seemed to be a shared feature of mitochondrial dysfunction by all patients within this study, seen from the mitochondrial membrane potential assays, but also a clear distinction between those patients with retinal dystrophy and those without for several of the assays tested within this study. This suggests some genotype-phenotype variations between these two phenotypic groups. The patients within this study come from 5 families, and all have different mutations, that affect different regions of ACBD5. Despite the range of mutations, all have a clear loss-of-function of the ACBD5 protein on Western Blot, suggesting NMD, which has also been predicted by Mutation Taster, and has been further confirmed in patients 1 and 2 from cDNA sequencing.

NMD is a safeguard against aberrant mRNAs and the subsequent products of their translation. Within normal transcripts there exist exon junction complexes (EJCs) that are marks left behind by spicing machinery, and RNA binding proteins, which are expelled from the translational process by the ribosome as it moves across the mRNA transcript to the intended stop codon. When there is an aberrant RNA transcript containing a premature termination codon (as is the case for all of the ACBD5-deficient patient transcripts), the RNA stops translation at the premature termination codon, thus leaving the EJCs and RNA binding proteins across the remaining untranslated transcript. These EJCs and RNA binding proteins recruit NMD machinery and trigger the destruction of the aberrant mRNA transcript (432). NMD escape is the phenomenon of NMD inefficiency and has been shown to occur via multiple routes. Alternative or inefficient splicing could exclude the exon containing the premature termination codon from the transcript, or inefficiently splice a downstream intron of the premature termination codon, which will alter the EJCs downstream of the premature stop codon, preventing initiation of NMD (432). Prevention of EJC assembly and subsequent inhibition of NMD has been shown in a transcript containing a premature termination codon through use of an anti-sense oligonucleotide targeting the downstream EJC site, showing that specific mRNA transcripts do undergo NMD escape when lacking an EJC site (433).

Importantly, reports have suggested that although proteins resultant from nonsense alleles that undergo NMD escape in loss-of-function mutations may have varying degrees of functionality, RNA processing and translation plasticity could convert these variants into silent, hypomorphic or neomorphic alleles (434). This could cause pathological variation in disease and phenotypic presentation. Facioscapulohumeral muscular dystrophy (FSHD) is an example of a disease where NMD plays a key role in pathology, although NMD escape occurs across the board in this case. FSHD is caused by the activation of transcription factor *DUX4* in skeletal muscle cells by epigenetic de-repression of a microsatellite repeat (435). The *DUX4* transcript is usually an endogenous target of NMD, so increased expression of DUX4 causes muscle deterioration in FSHD (436). It has also been shown that NMD occurs to different degrees between individuals, for example in choroideremia, an X-linked recessive chorioretinal dystrophy, which has up to 40 % variability in transcript expression (437). This suggests

that genetic variation in NMD factors could influence the penetrance, genotype-phenotype correlation, and severity of diseases. Importantly western blotting showed no protein expression across all six ACBD5-deficient patients, however the mutations involved span different regions of *ACBD5*, and could theoretically, if NMD escape occurs and splicing changes occur, produce smaller abnormal transcripts that could have access to mRNA processing and translational machinery.

RNA sequencing was performed to investigate differentially expressed genes in ACBD5-deficient patients and the possibility of NMD escape or differential expression of compensatory pathway activation in those patients with retinal dystrophy in comparison to those without. Four gene transcripts were identified as significantly downregulated in ACBD5-deficient patients in comparison to controls, when corrected for multiple testing: *ACBD5, DNM2* and *ANKRD11* (*p* = 0.014, 0.003, 0.029 and 0.003 respectively). Two other gene transcripts were nearing significance after multiple testing correction, another *ACBD5* transcript (p = 0.076 after correction, p = 0.000006 before correction) and *DROSHA* (p = 0.082 after correction, p = 0.000008 before correction). These data suggest that all ACBD5 loss of function is occurring in all six ACBD5-deficient patients, with nominal transcripts per million detected that support the occurrence of NMD, as also suggested by Western blot, cDNA Sanger sequencing and mutation taster analysis. In addition, when comparing ACBD5-deficient patients with retinal dystrophy, to those without, there were zero significantly differentially expressed ACBD5 transcripts, with the two transcripts detectable presenting with nominal transcripts per million. This lends further support for ACBD5 NMD and suggests that NMD-escape is unlikely to occur and does not contribute to the phenotypic variability.

Two candidate genes were identified by RNA sequencing that could potentially be involved in mechanistic pathogenicity: *ANKRD11* and *DNM2*. ANKRD11 is a large nuclear protein that is involved with transcription regulation. It is thought to bind chromatin modifying enzymes, such as histone deacetylases, which are crucial in neural precursor development and function (438). It has been shown to regulate histone acetylation and precursor proliferation, therefore determining neuronal positioning and neurogenesis, and is highly expressed in the brain (439, 440). *ANKRD11* haploinsufficiency has been implicated in autism spectrum disorder (ASD), intellectual disability and KBG syndrome (named after the initials of the first families diagnosed). KBG, is characterised by ASD, intellectual disability and seizures (441). Interestingly, intellectual disability has been reported in all six ACBD5-deficient patients. Protein interaction analysis using STRING did not show any direct or indirect interaction between ACBD5 and ANKRD11 (153). qPCR confirmation of *ANKRD11* downregulation is required in all ACBD5-patient patients, compared to controls, as the first investigation into its potential pathogenic role.

DNM2 is a ubiquitously expressed member of the GTPase family and is involved in membrane fission, particularly organelle fusion or division, and endocytosis. It has been implicated in CMT, with

autosomal dominant mutations in DNM2 leading to intermediate and axonal forms of the disease, predominantly affecting skeletal muscle and peripheral nerves (442). Membrane fission during endocytic vesicle release is known to be catalysed by dynamin (443, 444). Membrane vesicles facilitate the transfer of cellular cargo between organelles. Dynamins assemble into helical polymers and oligomerize into helices that wrap around the neck of plasma membrane buds, and using a GTP hydrolysis lever, sever membrane tubules, releasing the contents of the vesicle. This is an important process in both membrane fission and endocytosis (445). Mutations in *DNM2* affect multiple pathways that are involved in CMT pathology including endocytosis, T-tubule biogenesis, autophagy, and actin dynamics (442). Interestingly, protein interaction analysis using STRING identified that DNM2 indirectly interacts with both VAPA and VAPB, which are both implicated in ACBD5 deficiency pathogenesis (153). ACBD5 does not interact with DNM2 (153). Both VAPA and VAPB have been shown to be crucial for peroxisomal-ER tethering, which is lost when ACBD5 is deficient (395). StAR-related lipid transfer protein 3 (STARD3) and STARD3-N-terminal-like protein (STARD3NL) both are involved in tethering at ER contact sites, via contact with VAPA and VAPB. STARD3NL has been shown to interact with amphiphysin (AMPH) which is involved with regulated exocytosis and the membrane associated cytoskeleton. AMPH has been shown to interact with DNM2. Therefore, it is possible that DNM2 is downregulated in ACBD5-deficiency and this effect has a role to play in peroxisomal-ER tethering and the vesicle trafficking that is facilitated by this tether. Interestingly, enriched gene pathways found in ACBD5-deficient patients compared to controls, via RNA sequencing Gene Set Enrichment and Over-Representation Analysis, included upregulation of protein localisation to the ER and downregulation of regulation of transporter activity. This could indicate upregulated protein localisation as a compensatory mechanism, against the reduced trafficking and transporter activity caused by the loss of this ACBD5-PO tether. Further work is necessary to investigate *DNM2* as a candidate gene, including qPCR confirmation of *DNM2* downregulation in ACBD5-deficient patients compared to controls.

No significant genes were identified as being differentially expressed between ACBD5-deficient patients with retinal dystrophy, and those without. However, each cohort contained only three samples which is an extremely small sample size and may explain the lack of significance. The top five gene transcripts ($p < 1.0$, after multiple testing correction) contained three transcripts involved in immunity: two transcripts of *MICA*, and *HLA-A*. These were downregulated in patients with retinal dystrophy, in comparison to those without. MICA is a C-Type Lectin-Like NK Receptor (NKG2D) ligand and is associated with natural killer (NK) cell-mediated immunity. MICA and NKG2D interaction activate NK and T-cell surveillance, leading to the elimination of pathogens (446). HLA-A is a component of the HLA complex, located on chromosome 6, which forms the HLA haplotype that varies between individuals. HLA antigens are involved in self-recognition and immune response to antigenic stimulus. HLA-A belongs to the HLA class I genes and are expressed on the majority of somatic cells. They

present peptides that are recognised by CD8+ T cells (447). *HLA*-related polymorphisms have been linked to age-related macular degeneration (AMD) (448). Killer cell immunoglobulin-like receptors (KIRs), regulatory molecules that are expressed by NK and T cells, are thought to be mechanistically involved (449). Inhibitory KIRs that interact with HLA class I molecules modulate NK function and are dominantly phenotypically expressed in comparison to KIRs that activate NK cells (450). Healthy cells avoid spontaneous NK cell activation and killing via HLA class I molecules that act as ligands for inhibitory KIR receptors expressed on NK cells. Reduced HLA class I expression, for example reduced HLA-A, could result in spontaneous NK cell destruction, leading to AMD, which has been shown for HLA-C (448). Therefore, there could be immune involvement involved in the retinal dystrophy seen in three of the six ACBD5-deficient patients. Gene Set Enrichment highlighted neutrophil mediated immunity and granulocyte activation as enriched downregulated pathways in ACBD5-deficient patients with retinal dystrophy, however neither pathway was significantly enriched. The small cohort size is limiting and therefore RNA sequencing needs to be repeated on a larger cohort when more patients with ACBD5-deficiency are identified.

Gene Set Enrichment and Over-Representation Analysis using WebGestalt was used to identify enriched pathways within the ACBD5-deficient patient group compared to control results dataset and patients with retinal dystrophy comparison to those without results dataset. No significantly upregulated or downregulated enriched pathways were detected using gene set enrichment in either dataset, based on FDR. Over-Representation Analysis identified 10 significantly enriched pathways in ACBD5-deficient pathways compared to controls that include pathways involved in protein targeting and localisation to the ER, as well as protein targeting and signal recognition particle (SRP) co-translational protein targeting to membranes. These pathways involve a number of ribosomal proteins and overlapping genes. The current known mechanism of reduced membrane tethering between the peroxisome and ER could also have effects on other ER processes including ER protein localisation and targeting. However, following analysis using STRING the genes involved in these pathways do not directly or indirectly interact with ACBD5, VAPA or VAPB, and therefore these enriched pathways could be a secondary effect (appendix figure 8.3.1 and 8.3.2). Three significantly enriched pathways were also identified in the patients with retinal dystrophy compared to those without which focused-on cell growth and its regulation, with an overlap of genes involved within these biological processes. STRING analysis showed that there is no direct or indirect interaction between these genes and ACBD5, however they do show significantly differentially enriched gene sets in patients with retinal dystrophy, compared to those without (Appendix figure 8.3.3). Interestingly, fatty acid β-oxidation, fatty acid oxidation and lipid modification gene sets were all enriched, although not significantly, in patients with retinal dystrophy, suggesting differences in these pathways between the two groups of patients (appendix figure 8.3.4). Additionally, *PEX7* comprised these gene sets and is important in peroxisomal biogenesis. Interestingly however, work by the Schrader group did not find any differences in

peroxisomal number between patients without retinal dystrophy, similarly to those patients with retinal dystrophy, suggesting that peroxisomal biogenesis is not impaired. Impaired β-oxidation of very-long chain fatty acids is a universal feature of ACBD5-deficiency. Additionally plasmalogen synthesis begins in the peroxisome and is completed in the ER, therefore relying on efficient trafficking via the peroxisomal-ER tether, which is universally reduced in ACBD5-deficient patients (418). There could be subtle differences in these mechanistic pathways between patients with retinal dystrophy and those without. Future work into detailing the metabolome and lipid profiles of these patients could reveal more subtle differences that may explain the phenotypic differences seen.

However, a limitation of both datasets is the low number of patient samples (n=6), especially in comparing patients with retinal dystrophy (n=3) and those without (n=3). Therefore, RNA sequencing would need to be repeated in the future with larger numbers of patient samples; this is not currently possibly due to the low number of known cases globally. There are some potential candidate genes and pathways with p values < 0.05 before multiple testing correction, that would be interesting to investigate further if validated with a larger RNA sequencing cohort.

3.4.6. Conclusion

In conclusion, this work has furthered the understanding of ACBD5-deficiency and identified a novel associated phenotype. Two patients from the NHNN with cerebellar ataxia were identified as having a homozygous mutation in *ACBD5*, a peroxisomal transmembrane protein, causing a loss of function. Using Genematcher four more cases of ACBD5-deficiency were recruited, including two published cases, confirming loss of function and abnormalities in VLCFA β-oxidation in all cases. Interestingly three of the six patients presented with retinal dystrophy, as well as cerebellar ataxia. Mitochondrial function was explored in the six patients, due to the interaction between mitochondria and peroxisomes, and their linked metabolic functions. Mitochondrial assays suggested a very subtle impairment in complex I of the OXPHOS machinery in patients compared to controls, with patients without retinal dystrophy appearing to have a more severe phenotype. However, with addition of a positive control in the form of a patient with ZSD, and therefore mitochondria dysfunction, this impairment is most likely a secondary effect of perturbed metabolic processes. Throughout this study another research group identified a potential mechanism involved with ACBD5-deficiency that implicates two additional proteins VAPA and VAPB. Membrane tethers were identified between peroxisomes and ER, thought to facilitate the trafficking of lipids and other molecules. These tethers are lost in the case of ACBD5-deficiency, increasing peroxisomal motility away from the ER, and potentially reducing the transfer of lipids and other molecules. Peroxisomal motility assays conducted by Dr Michael Schrader's group on the three additional patients with the novel non-retinal dystrophy phenotype all were confirmed to show reduced peroxisomal-ER attachments, with a reduction in tethering, similarly to those patients with retinal dystrophy. RNA sequencing revealed some new potential candidate genes that are significantly

downregulated in ACBD5-deficient patients compared to controls. These include *DNM2*, *ANKRD11* and *DROSHA*, as well as downregulation of *ACBD5* in all patients, further confirming loss of function. Pathways concerning protein localisation to the ER and downregulation of transporter activity were enriched in ACBD5-deficient patients in comparison to controls, potentially adding further support to the downregulation of vesicular trafficking and cargo transfer between peroxisomes and the ER in cases of ACBD5-deficiency. Potential candidate genes were also identified in ACBD5-deficient patients with retinal dystrophy, compared to those without that include downregulation of *HLA-A* and *MICA*, involved in immunity. Additionally, fatty acid β-oxidation and lipid modification were enriched in patients with retinal dystrophy, albeit not significantly, suggesting that there could be mechanistic differences in these pathways between the two different phenotypic groups. However, RNA sequencing requires validation with a larger cohort and validation of targets using qPCR. This work builds upon existing evidence related to the potential mechanism of peroxisomal-ER contact reduction and establishes some intriguing preliminary targets for investigation in future studies.

4. General Discussion and Future Directions

4.1. Covid-19 Impact Statement

The covid-19 pandemic introduced a variety of challenges throughout the last year of experimental time within this research. The laboratory was inaccessible for 6 months, with intermittent and restricted use thereafter. This proved challenging as this work was primarily lab-based and relied on culturing cell-lines for functional tests and equipment including thermocyclers, fragment analysis sequencers, Western blotting apparatus, and confocal and electron-transmission microscopes. Large delays in work with fibroblast lines also occurred due to UCL closure, with additional time required to re-culture when access to the laboratory were reinstated. This delay was also applicable to the use of DNA samples from the NHNN Biobank, of which this work was largely reliant, and of which samples became inaccessible for a large period of time.

Additionally, this work relied on close collaborations with the McGill University, Montreal, Exeter University, UK and Illumina, USA. All these institutes were shut due to Coronavirus, with additional clinical collaborators refocused on aiding the clinical response to Covid-19 within their respective countries. This meant that detailed clinical information for one patient included in the project into ACBD5 mutations was unable to be sent from the consulting clinicians on this case. Additionally, metabolic screening results for this patient have not yet been shared by the respective collaborator due to delays.

One positive aspect from this challenging period was the increased use of bioinformatic techniques. This resulted in an opportunity to collaborate with Illumina, helping develop and refine a bioinformatic tool used to identify *RFC1* repeat expansions. The time where access to the laboratory was restricted was redirected into learning how to code in the RStudio statistical environment and additionally learn Bash and Python coding languages. Permission to access the Genomics England 100,000 genomes project data, which also greatly aided this project.

4.2. Identification of RFC1 repeat expansions as a cause of late-onset ataxia and CANVAS

The research presented in chapter 3, details the discovery of the *RFC1* repeat expansion found to be causative of late-onset ataxia and CANVAS. This work has been instrumental to the genetic understanding of CANVAS, resulting in the first report of a genetic cause of the disease. The utilisation of NGS techniques on 29 affected individuals from 11 families, revealed the biallelic repeat expansion mapped to the position 4q14, with WGS revealing a STR $AAAAG_{(11)}$ in intron 2, that was mutated to $AAGGG_{(n)}$ and expanded in affected individuals (2). The region is extremely heterogeneous, with multiple motif conformations present, including expanded reference $AAAAG_{(n)}$ and $AAAGG_{(n)}$. The genomic techniques flanking PCR, RPPCR and Southern blotting confirmed the pathogenic repeat

expansions in all patients from the 11 families and an additional 16 sporadic cases of CANVAS, with expansions size from 400 to 2000 repeating units. Understanding relative genotype frequency and relative allele frequency are extremely important in gauging genetic variation and determining the effects of multigenic mechanisms. This work determined that the allelic frequency of the pathogenic conformation was 0.7%. The allelic frequencies of $AAAAG_{(n)}$, $AAAGG_{(n)}$ and $AAAAG_{(11)}$ were 13%, 7.9% and 75.5% respectively. Mean expansion size for expanded AAGGG, AAAGG and AAAAG were 740, 72 and 173 repeating units respectively. Motif interruption was identified in several cases, evidencing the high level of polymorphism within the region, and suggesting that alternative pathogenic motifs may occur which was confirmed by collaboration with the Laing research group from the Harry Perkins Institute of Medical Research in Perth Australia. The screening of a Māori and Cook Island Māori CANVAS cohort identified a novel pathogenic repeat motif confirmation of $AAAGG_{(10\text{-}25)}$ $AAGGG_{(n)}$ $AAAGG_{(n)}$, with pathogenic repeat expansions confirmed on Southern blot. The allelic frequency within a late-onset ataxia cohort was also identified as 3.1% for the pathogenic AAGGG(n) motif, and 91.9%, 0.5% and 0.2% for the short reference $AAAAG_{(11)}$, $AAAGG_{(n)}$ and expanded $AAAAG_{(n)}$ motifs respectively.

Importantly this work optimised the Southern blot protocol, a key component of *RFC1* diagnostic workflow, which is a time-consuming but necessary step in sizing repeat expansions. This has enabled a more efficient and reliable protocol to be utilised in the identification of *RFC1* repeat expansions, vastly reducing the requirement of repeats and additional use of large quantities of gDNA due to low quality or failed blots.

4.2.1. Future work

The mechanism behind the pathogenicity of the biallelic *RFC1* repeat expansion is currently unknown and will account for a large proportion of future work, currently undertaken by Dr Cortese. No loss of function of the *RFC1* gene has been found and additionally no reduction in RFC1 protein expression has been found. RNA sequencing did not highlight any neighbouring or distant genomic regions that may affect RFC1 at a protein or transcriptional level, and no impaired response to DNA repair suggestive of potential genetic modifiers, was observed in *RFC1*-positive patient fibroblasts. Additionally, no sense or anti-sense RNA foci were observed in CANVAS brain tissue. An important functional finding, however, was the retention of intron 2 in *RFC1* pre-mRNA found increased across all tissues of CANVAS patients, which could point to a potential pathogenic mechanism. Abnormal pre-mRNA transcripts could cause aberrant affects to downstream translational processes, and future work will focus on further investigating this finding.

Future neuropathological analysis of *RFC1*-positive CANVAS brain tissue will be instrumental in elucidating a potential mechanism. Only one *RFC1*-positive CANVAS brain to date has been included in this research, which is too small a sample size to draw meaningful conclusions. Further validation

will be needed to characterise the associated pathology of *RFC1*-associated CANVAS, defining the downstream pathophysiology caused by the expansion. Since the publication of these results, clinical examination of larger cohorts of *RFC1*-positive CANVAS patients has been carried out, and sensory neuropathy seems to be a largely universal feature (53). Therefore, future further validation using brain tissue is required to determine if sensory neuropathy is the site of initial disease manifestation.

Since the publication of these results, the understanding of the *RFC1* repeat expansion has expanded with multiple studies and screenings on larger and more diverse cohorts, including Asian Pacific and Japanese populations (171, 172). However, a frequent limitation of allelic frequency estimations is a lack of power to define true allelic frequency. The cohort size for normal population used within this research was 304 individuals. Larger studies will always have a smaller deviation from the true frequency. Therefore, ongoing screenings of more diverse ethnicities and larger sample sizes will be required to gauge the true allelic frequency, which from numerous screenings to date is reported within the range of 0.7% to 6.8% (50, 53, 169, 171, 172, 256). In addition a meta-analysis of all reported allelic frequencies would be valuable in defining the true allelic frequency across various populations. Chapter 5 details future work I have carried out on a much larger cohort of 62, 550 individuals enrolled in the 100k GP, identifying the allelic frequency within the largest sample cohort to date.

Future work into finding bioinformatic solutions for *RFC1* repeat expansion identification are ongoing. Chapter 5 details the development of a novel algorithmic tool that utilises ExpansionHunter output data to detect the pathogenic *RFC1* repeat expansion. This, alongside reported work identifying the expansion using WES data, opens the possibility for future screenings on existing WGS and WES datasets, independent of PCR screening (38).

4.3. *RFC1* repeat expansion identification in other disorders

Chapter 4 details research undertaken to identify the prevalence of the *RFC1* repeat expansion in a disease cohort that heavily clinically overlaps with CANVAS, MSA. Two separate cohorts were screened, 336 cases that were neuropathologically confirmed using the MSA working criteria and 207 clinically diagnosed cases with either ‘possible’ or ‘probable’ MSA. No cases positive for the biallelic *RFC1* repeat expansion were found in the pathologically confirmed cohort. The allelic frequencies found were 0.7% for the $AAGGG_{(n)}$ motif, 3% for $AAGGG_{(n)}$, 7% for expanded $AAAAG_{(n)}$ and 88% for the short reference $AAAAG_{(11)}$. The mean expansions sizes for the different expanded motif conformations were 85 repeating units for $AAAAG_{(n)}$, 124 repeating units for $AAAGG_{(n)}$ and 1185 repeating units for pathogenic $AAGGG_{(n)}$.

Within the clinically diagnosed cohort, three patients were identified with the recessive *RFC1* repeat expansion. The allelic frequencies for the different motifs were 1.7% for pathogenic $AAGGG_{(n)}$, 2.4% for expanded $AAGGG_{(n)}$, 11.4% for expanded $AAAAG_{(n)}$ and 82.8% for the short reference $AAAAG_{(11)}$. Two of the *RFC1*-positive patients were diagnosed with ‘possible’ MSA, and one with ‘probable’ MSA.

Southern blotting was carried out for two of the patients that had sufficient gDNA, revealing expansions at 887 and 1495 repeating units for one patient, and 1249 repeating units for the other.

These three patients are thought to have an MSA-mimic disorder, as they all have classical MSA features of autonomic dysfunction and parkinsonism, but also share CANVAS features such as sensory neuropathy in two cases and cerebellar impairment. Importantly the core triad of CANVAS symptoms was missing in all cases, such as vestibular areflexia, therefore ruling out CANVAS as a diagnosis. CANVAS is frequently misdiagnosed and MSA is often given as a differential diagnosis. This work has importantly highlighted that the *RFC1* repeat expansion is implicated in diseases other than CANVAS, that may however share a phenotypic overlap. Since the publication of this data, further studies have confirmed the *RFC1* repeat expansion in MSA, and additionally a range of other diseases including dopa responsive parkinsonism, supranuclear palsy and hyperkinetic movement disorders (202, 256, 260). The clinical spectrum associated with the repeat expansion is broadening, expanded by this research and reports of other diseases linked to the expansion. Importantly this work has highlighted that patients that present with atypical features of both MSA and CANVAS, should be screened for the *RFC1* repeat expansion within the clinic.

4.3.1. Future work

The identification of the *RFC1* repeat expansion has broadened the phenotypic scope of the repeat expansion, which still has not been fully characterised. Further studies are required to determine its full spectrum, especially on disorders that share clinical features of CANVAS, such as sensory neuropathy and have an atypical presentation of their diagnosis. Work detailed in chapter 5 outlines the first study of non-neurological disease associated with the *RFC1* repeat expansion, using a rare diseases cohort from the 100kGP dataset. Within this cohort are also diseases not yet associated with the expansion, including intellectual disability and epilepsy.

Additionally, neuropathological analysis of non-CANVAS brain tissue that harbours the repeat expansion will be crucial to furthering our understanding of how the expansion leads to downstream phenotypic presentation. With the continued investigations into other possible diseases associated with the repeat expansion, and the increasing numbers of implicated disorders, the capacity to make advances in our understanding of *RFC1*-related neuropathology will increase. Determining the specific phenotypes that are directly caused by the repeat expansion will aid in the discovery of a pathogenic mechanism. This will be important if *RFC1* or downstream pathways are targeted therapeutically, to ensure all patients that harbour this repeat expansion can receive therapeutic benefit.

4.4. Development of bioinformatic tool to detect *RFC1* expansions

The work detailed in chapter 5 has developed a bioinformatic tool that utilises ExpansionHunter output data (that works on WGS data) to effectively identify individuals with the *RFC1* recessive pathogenic repeat expansion. The project capitalised on the wealth of data available from the 100kGP and specifically used a rare diseases cohort of 62, 550 samples. The RFC1 caller results indicated that the pathogenic $AAGGG_{(n)}$ motif had an allelic frequency of 4%, 41% for a benign expanded motif (such as $AAAGG_{(n)}$ and $AAAAG_{(n)}$) and 55% for the short reference $AAAAG_{(11)}$. This is the largest cohort that has been screened for the expanded motifs involved in *RFC1* expansions to date, and therefore may reflect the truest allelic frequency values. Remarkably, of the 145 individuals called as 'patients: two expanded pathogenic alleles', 43 were relatives, compare to 102 which were probands. This could suggest that the *RFC1* recessive repeat expansion has a lower penetrance than once thought, which has not been reported in any other studies to date. 22.8% of probands called as 'patient: two expanded pathogenic alleles' were also not recruited through 'Neurology' and therefore have non-neurological disease. This suggests that the *RFC1* repeat expansion could be involved in a broader range of phenotypes, or, again, that there is a reduced penetrance. The RFC1 caller also identified 106 novel motif configurations, 26 of which comprise greater than 10% of the repeating sequencing, emphasising the high level of polymorphism within the *RFC1* locus.

Of the 145 individuals identified with the pathogenic repeat expansion, 59 probands were recruited from the NHNN catchment area and 30 samples had previously been confirmed to have the repeat expansion via flanking PCR, RPPCR and Southern blotting. In order to confirm the remaining samples, Southern blotting was carried out for 12 individuals, where gDNA was available, and all showed visible bands corresponding to the pathogenic repeat expansion. Interestingly it confirmed two cases that had been detected as having the recessive, pathogenic expansion, despite having a previous diagnosis of schwannomatosis and myotonia congenita, further suggesting reduced penetrance or suggesting more complex digenic cases. Using a cohort of 105 cases independently put forward for *RFC1* diagnostic screening, based on their clinical phenotype, I compared their diagnostic workflow results and the results from the RFC1 caller. The RFC1 caller showed a sensitivity of 96.2% and specificity of 100%, which although carried out on a modest cohort size, shows the efficacy of the caller for predicting the correct *RFC1* allele status. Segregation using the trios available within the rare diseases' cohort showed a classic pattern of Mendelian inheritance, and a *de novo* frequency of 4.7%.

4.4.1. Future work

The results from the RFC1 caller have highlighted a potential lower penetrance for the pathogenic repeat expansion, with 30% of individuals identified as carrying the recessive pathogenic expansion as being relatives, and 22.8% of remaining probands not having neurological disease. This requires further

screening on larger molecularly confirmed neurological and non-neurological disease cohorts to confirm. It also requires the in-depth analysis of clinical features of these patients that do not fit the criteria for neurological disease categorisation. The data suggests that epilepsy and intellectual disability, among other diseases not yet associated with the repeat expansion, may be linked the *RFC1* expansion, however this could also be explained by reduced penetrance. Therefore if available WGS data exists for these diseases cohorts it would be beneficial to carry out an RFC1 caller screen. If WGS data is unavailable, then gDNA from these disease cohorts would be interesting candidates for the diagnostic screening workflow to explore this further. For these patients who do not fit the expected phenotype, it would be interesting to perform nerve conduction studies, in order to see if they have some form of sensory neuropathy, which is reported as a main associated feature of *RFC1*-associated CANVAS. This would also aid in the understanding of the pathophysiology underpinning the mechanism of the *RFC1* repeat expansion.

A limitation of the GEL environment is that no information is available for relatives recruited into the scheme, therefore the ability to check relatives that have been identified as having the recessive pathogenic expansion is diminished. However, screening much larger cohorts of individuals, either healthy or not with neurological disease, is required to ascertain the true allelic frequency of $AAGGG_{(n)}$, and further whether the recessive pathogenic expansion does have a lower penetrance. It will be crucial to screen larger, independent cohorts, that comprise a wider range of diseases and have clinical data available, to investigate the suspected reduced penetrance. Longitudinal studies on individuals identified as having the biallelic *RFC1* repeat expansion, but with atypical or no associated symptoms, would provide insight into whether they go onto develop associated clinical features later on in life.

The only recruitment area for 100kGP with immediate access for genomic DNA samples is the 'RP4' catchment area, which the NHNN falls under. Therefore future work will include collaborations with other centres for which patients with the pathogenic recessive repeat expansion have been identified, to report back molecular diagnostic findings and also to further confirm the accuracy of the RFC1 caller, via secondary Southern blotting confirmation. This would expand the cohort used to test specificity and sensitivity, likely increasing the current sensitivity value.

4.5. Recessive loss of function mutations in ACBD5 as a cause of autosomal recessive cerebellar ataxia

The work detailed in chapter 6 has further characterised a subset of ARCA caused by mutations in peroxisomal membrane protein *ACBD5*. Within this study, six patients with mutations occurring in *ACBD5* and ARCA were recruited; initially two brothers were identified at the NHNN, and then four additional patients were recruited via collaboration and the online portal Genematcher. This provides the largest cohort of ACBD5-deficient patients to date, all of which were found to have a complete loss-

of-function of *ACBD5*. Interestingly there was a distinct separation between the phenotypic presentation of the six patients, three presented with pure ataxic symptoms and three presented with ataxia and retinal dystrophy. All six patients underwent metabolic screening and were identified with having impaired β-oxidation of VLCFAs.

Due to the metabolic function of peroxisomes and peroxisomal interaction with the mitochondria, mitochondrial assays were employed to test mitochondrial function. Results suggested a subtle impairment of complex I of the mitochondrial oxidative phosphorylation machinery, and a decreased NADH pool across all patients, compared to healthy controls. Using a positive control with ZSD, another peroxisomal disorder, it was determined that the mitochondrial dysfunction seen was likely a secondary effect of initial peroxisomal dysfunction. Interestingly, the impairment in complex I and reduced redox state of the mitochondria seemed to be more severe in those patients who did not present with retinal dystrophy.

Whilst this study was ongoing, advancements in the understanding of ACBD5 and its role with regard to peroxisomal function were made, with peroxisomal motility reported to be mediated by the presence of ACBD5. When ACBD5 is lost, peroxisomal motility increases, and additionally ACBD5 mediates the formation of discrete contact bridges between peroxisomes and the ER (395, 417). In order to explore if there were any differences in the function of peroxisomes, the samples without retinal dystrophy were sent to a collaborator Prof Michael Schrader, who performed peroxisomal biogenesis and peroxisomal motility assays. There were no significant differences in peroxisomal number between patients and controls, but significantly reduced peroxisomal-ER contacts and attachment compared to controls, fitting the results seen for patients with retinal dystrophy. This suggests that there are no differences in peroxisomal function between those patients with retinal dystrophy and those without, ruling this out as a mechanistic contributor to the phenotypic differences seen.

Sanger sequencing of cDNA revealed that NMD was present in two patients, with western blotting revealing no protein for all six patients and Mutation Taster predicting NMD in the remaining four patients. By analysing the mutational position and the new position of premature termination codon caused by the resulting frameshift in all six patients, four of the patients presented as candidates for NMD escape, which could explain the phenotypic differences seen as three of the four had no retinal dystrophy. Samples were sent for RNA sequencing in order to assess if any truncated transcripts may escape Western blotting, and to assess if there were any differences in compensatory pathway activation. No significant differences in *ACBD5* transcripts were observed between patients with retinal dystrophy and those without, however significantly reduced *ACBD5* transcripts were observed across all ACBD5-deficient patients. These data rule out NMD escape as a potential mechanistic contributor and confirms the loss-of-function phenotype in all patients.

RNA sequencing of all ACBD5-deficient patients and controls identified some potential candidate genes that included *DNM2*, *ANKRD11* and *DROSHA*, as well as potentially important pathways implicated in the mechanism that include protein localisation to the ER and transporter regulation. Additional candidate genes were identified in RNA sequencing comparison between patients with retinal dystrophy and those without, including immunity genes *HLA-A* and *MICA*, however the cohort size was extremely small (n=3 for both groups). These results identify potential new targets for investigation and highlight pathways that could be interesting to explore further in terms of mechanistic involvement.

4.5.1. Future work

The mitochondrial impairment of complex I shown in this study requires further validation to confirm whether it is a secondary phenotype caused by perturbed metabolic activity, or part of the primary mechanism. Only one positive control with ZSD and one with ALD were included, which is not enough to draw reliable conclusions. Therefore, investigations into electron transport chain function requires validation with additional ZSD and ALD positive samples.

RNA sequencing data has identified some potential new targets for investigation in ACBD5-deficient patients. Whilst these genes are significant, their involvement in the mechanism is still unclear. These genes require validation using qPCR to confirm their downregulation in patients compared to controls before any further investigation. Ideally RNA sequencing requires repeating to confirm results and with a much larger cohort of ACBD5-deficient patients. However, there are very low known numbers of these patients, with this study using six of the eight currently known cases worldwide, so it may be a long timeframe until a larger study could be conducted. One potential solution to this would be to use RNA extracted from ACBD5 knockdown cell lines, in a more controlled experiment. Comparing these results to the RNA sequencing patient data could identify whether the observed gene changes are solely due to the loss of ACBD5. The pathways highlighted from the RNA sequencing work included in this study have the potential to further our understanding of the mechanism, as pathways involving the ER were enriched in ACBD5-deficient patients. This fits with the currently known mechanism involving the loss of peroxisomal-ER tethering in cases of ACBD5-deficiency.

Over-Representation Analysis highlighted fatty acid β-oxidation and lipid modification as enriched in patients with retinal dystrophy. These enrichments were not significant; however cohort numbers were very small, and therefore this could suggest mechanistic differences between the two phenotypic groups. It is very likely that disruption to lipid transfer from the peroxisome to the ER and vice versa is implicated in pathogenicity in all patients, and therefore subtle mechanistic differences could contribute to the differences in phenotype. Professor Schrader's group in Exeter is planning on carrying out more in-depth lipidomic work on ACBD5-deficient patients and the patients identified in this study with the novel clinical feature of cerebellar ataxia with no retinal dystrophy have been sent to be

included. This work has the potential to further elucidate the mechanism and identify any differences in lipid profile between patients with retinal dystrophy and those without.

5. Publication list

Genome wide association study identifies risk loci for cluster headache.
Emer O'Connor, Carmen Fourier, Caroline Ran, Prasanth Sivakumar, Franziska Liesecke, Laura Southgate, Aster V.E. Harder, Lisanne S. Vijfhuizen, Yip Hau Ying, Nicola Giffin, Nicholas Silver, Fayyaz Ahmed, Isabel C Hostettler, Brendan Davies, M Zameel Cader, Benjamin S Simpson, **Roisin Sullivan** et al., Annals of Neurology 2021 Aug;90(2):193-202.
Author contribution: Middle author; performed genetic testing, data analysis and helped edit manuscript.

Cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS): Genetic and Clinical Insights.
Sullivan R, Kaiyrzhanov R and Houlden H. Current Opinion in Neurology August 2021 34(4):556-564
Author Contribution: Joint first author; researched and wrote manuscript.

Reply: Genetic heterogeneity of neuronal intranuclear inclusion disease. What about the infantile variant?
Yau WY, Chen Z, **Sullivan R**, Vandrovcova J, Houlden H. Ann Clin Transl Neurol. 2021 Apr;8(4):1002-1004. doi: 10.1002/acn3.51330. Epub 2021 Mar 29.
Author contribution: Third author; helped write and edit reply to Letter to the Editor

RFC1-related ataxia is a mimic of early multiple system atrophy.
Sullivan R, Yau WY, Chelban V, Rossi S, Dominik N, O'Connor E, Hardy J, Wood N, Cortese A, Houlden H.J Neurol Neurosurg Psychiatry. 2021 Feb 9;92(4):444-6. doi: 10.1136/jnnp-2020-325092. Online ahead of print.
Author contribution: Joint first author; developed primary hypothesis, analysis of all genomic data, data visualisation and writing of manuscript.

NOTCH2NLC Intermediate-Length Repeat Expansion and Parkinson's Disease in Patients of European Descent.
Yau WY, **Sullivan R**, Rocca C, Cali E, Vandrovcova J, Wood NW, Houlden H.Ann Neurol. 2021 Mar;89(3):633-635. doi: 10.1002/ana.26003. Epub 2021 Jan 11.
Author contribution: Joint first author; developed primary hypothesis, analysis of all genomic data, data visualisation and writing of manuscript.

A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families.
Scriba CK, Beecroft SJ, Clayton JS, Cortese A, **Sullivan R**, Yau WY, Dominik N, Rodrigues M, Walker E, Dyer Z, Wu TY, Davis MR, Chandler DC, Weisburd B, Houlden H, Reilly MM, Laing NG, Lamont PJ, Roxburgh RH, Ravenscroft G.Brain. 2020 Oct 1;143(10):2904-2910. doi: 10.1093/brain/awaa263.
Author contribution: Middle author; performed Southern blots included in manuscript, analysed genomic data.

Dopa-Responsive Parkinsonism in a Patient With Homozygous RFC1 Expansions.
Sullivan R, Yau WY, Chelban V, Rossi S, O'Connor E, Wood NW, Cortese A, Houlden H.Mov Disord. 2020 Oct;35(10):1890-1891. doi: 10.1002/mds.28279.
Author contribution: First author; wrote and edited reply to Letter To The Editor

Low Prevalence of NOTCH2NLC GGC Repeat Expansion in White Patients with Movement Disorders.

Yau WY, Vandrovcova J, **Sullivan R**, Chen Z, Zecchinelli A, Cilia R, Duga S, Murray M, Carmona S; Genomics England Research Consortium, Chelban V, Ishiura H, Tsuji S, Jaunmuktane Z, Turner C, Wood NW, Houlden H.Mov Disord. 2021 Jan;36(1):251-255. doi: 10.1002/mds.28302. Epub 2020 Oct 7.
Author contribution: Third author; developed primary hypothesis, contributed to analysis of genomic data, data visualisation and writing of manuscript.

A Māori specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele.
Beecroft SJ, Cortese A, **Sullivan R**, Yau WY, Dyer Z, Wu TY, Mulroy E, Pelosi L, Rodrigues M, Taylor R, Mossman S, Leadbetter R, Cleland J, Anderson T, Ravenscroft G, Laing NG, Houlden H, Reilly MM, Roxburgh RH.Brain. 2020 Sep 1;143(9):2673-2680. doi: 10.1093/brain/awaa203.
Author contribution: Joint second author; contributed to primary hypothesis, performed Southern blotting and analysis of genomic data, helped write and edit manuscript.

Neuronal intranuclear inclusion disease is genetically heterogeneous.
Chen Z, Yan Yau W, Jaunmuktane Z, Tucci A, Sivakumar P, Gagliano Taliun SA, Turner C, Efthymiou S, Ibáñez K, **Sullivan R**, Bibi F, Athanasiou-Fragkouli A, Bourinaris T, Zhang D, Revesz T, Lashley T, DeTure M, Dickson DW, Josephs KA, Gelpi E, Kovacs GG, Halliday G, Rowe DB, Blair I, Tienari PJ, Suomalainen A, Fox NC, Wood NW, Lees AJ, Haltia MJ; Genomics England Research Consortium, Hardy J, Ryten M, Vandrovcova J, Houlden H.Ann Clin Transl Neurol. 2020 Sep;7(9):1716-1725. doi: 10.1002/acn3.51151. Epub 2020 Aug 10.
Author contribution: Middle author; performed and analysed genetic tests and helped edit manuscript

De Novo and Bi-allelic Pathogenic Variants in NARS1 Cause Neurodevelopmental Delay Due to Toxic Gain-of-Function and Partial Loss-of-Function Effects.
Manole A, Efthymiou S, O'Connor E, Mendes MI, Jennings M, Maroofian R, Davagnanam I, Mankad K, Lopez MR, Salpietro V, Harripaul R, Badalato L, Walia J, Francklyn CS, Athanasiou-Fragkouli A, **Sullivan R**, Desai S, Baranano K, Zafar F, Rana N, Ilyas M, Horga A, Kara M, Mattioli F, Goldenberg A, Griffin H, Piton A, Henderson LB, Kara B, Aslanger AD, Raaphorst J, Pfundt R, Portier R, Shinawi M, Kirby A, Christensen KM, Wang L, Rosti RO, Paracha SA, Sarwar MT, Jenkins D; SYNAPS Study Group, Ahmed J, Santoni FA, Ranza E, Iwaszkiewicz J, Cytrynbaum C, Weksberg R, Wentzensen IM, Guillen Sacoto MJ, Si Y, Telegrafi A, Andrews MV, Baldridge D, Gabriel H, Mohr J, Oehl-Jaschkowitz B, Debard S, Senger B, Fischer F, van Ravenwaaij C, Fock AJM, Stevens SJC, Bähler J, Nasar A, Mantovani JF, Manzur A, Sarkozy A, Smith DEC, Salomons GS, Ahmed ZM, Riazuddin S, Riazuddin S, Usmani MA, Seibt A, Ansar M, Antonarakis SE, Vincent JB, Ayub M, Grimmel M, Jelsig AM, Hjortshøj TD, Karstensen HG, Hummel M, Haack TB, Jamshidi Y, Distelmaier F, Horvath R, Gleeson JG, Becker H, Mandel JL, Koolen DA, Houlden H.Am J Hum Genet. 2020 Aug 6;107(2):311-324. doi: 10.1016/j.ajhg.2020.06.016. Epub 2020 Jul 31.
Author contribution: Middle author; performed and analysed Western blots and helped edit manuscript

GGC Repeat Expansion in NOTCH2NLC Is Rare in European Leukoencephalopathy.
Yau WY, **Sullivan R**, Chen Z, Lynch DS, Vandrovcova J, Wood NW, Houlden H.Ann Neurol. 2020 Sep;88(3):641-642. doi: 10.1002/ana.25818. Epub 2020 Jul 8.
Author contribution: Joint first author; developed primary hypothesis, carried out and analysed all genomic data, wrote and edited manuscript.

RFC1 Intronic Repeat Expansions Absent in Pathologically Confirmed Multiple Systems Atrophy.
Sullivan R, Yau WY, Chelban V, Rossi S, O'Connor E, Wood NW, Cortese A, Houlden H.Mov Disord. 2020 Jul;35(7):1277-1279. doi: 10.1002/mds.28074. Epub 2020 Apr 24.
Author contribution: First author; developed primary hypothesis, performed genetic tests, analysed all genomic data and wrote manuscript.

Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion.
Cortese A, Tozza S, Yau WY, Rossi S, Beecroft SJ, Jaunmuktane Z, Dyer Z, Ravenscroft G, Lamont PJ, Mossman S, Chancellor A, Maisonobe T, Pereon Y, Cauquil C, Colnaghi S, Mallucci G, Curro R,

Tomaselli PJ, Thomas-Black G, **Sullivan R**, Efthymiou S, Rossor AM, Laurá M, Pipis M, Horga A, Polke J, Kaski D, Horvath R, Chinnery PF, Marques W, Tassorelli C, Devigili G, Leonardis L, Wood NW, Bronstein A, Giunti P, Züchner S, Stojkovic T, Laing N, Roxburgh RH, Houlden H, Reilly MM.Brain. 2020 Feb 1;143(2):480-490. doi: 10.1093/brain/awz418.
Author contribution: Middle author; performed Southern blotting, analysed data and helped edit manuscript.

PDXK mutations cause polyneuropathy responsive to pyridoxal 5'-phosphate supplementation.
Chelban V, Wilson MP, Warman Chardon J, Vandrovcova J, Zanetti MN, Zamba-Papanicolaou E, Efthymiou S, Pope S, Conte MR, Abis G, Liu YT, Tribollet E, Haridy NA, Botía JA, Ryten M, Nicolaou P, Minaidou A, Christodoulou K, Kernohan KD, Eaton A, Osmond M, Ito Y, Bourque P, Jepson JEC, Bello O, Bremner F, Cordivari C, Reilly MM, Foiani M, Heslegrave A, Zetterberg H, Heales SJR, Wood NW, Rothman JE, Boycott KM, Mills PB, Clayton PT, Houlden H; Care4Rare Canada Consortium and the SYNaPS Study Group.
Author contribution: SYNaPs Study Group author

Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia.
Cortese A, Simone R, **Sullivan R**, Vandrovcova J, Tariq H, Yau WY, Humphrey J, Jaunmuktane Z, Sivakumar P, Polke J, Ilyas M, Tribollet E, Tomaselli PJ, Devigili G, Callegari I, Versino M, Salpietro V, Efthymiou S, Kaski D, Wood NW, Andrade NS, Buglo E, Rebelo A, Rossor AM, Bronstein A, Fratta P, Marques WJ, Züchner S, Reilly MM, Houlden H.Nat Genet. 2019 Apr;51(4):649-658. doi: 10.1038/s41588-019-0372-4. Epub 2019 Mar 29.
Author Contribution: Third author, carried out genetic screen on general population and wrote parts of manuscript.

Spinocerebellar ataxia: an update.
Sullivan R, Yau WY, O'Connor E, Houlden H.J Neurol. 2019 Feb;266(2):533-544. doi: 10.1007/s00415-018-9076-4. Epub 2018 Oct 3.
Author contribution: First author, researched and wrote manuscript

DNA repair in trinucleotide repeat ataxias.
Yau WY, O'Connor E, **Sullivan R**, Akijian L, Wood NW.FEBS J. 2018 Oct;285(19):3669-3682. doi: 10.1111/febs.14644. Epub 2018 Sep 10.
Author contribution: Created figures and helped write manuscript

6. Other achievements

Awarded associate fellowship of the Higher Education Academy (AFHEA), from MSc teaching carried out at the Institute of Neurology.

Awarded first place in poster competition at 2019 12th UK Neuromuscular Translational Research Conference.

Awarded runner up place in poster competition at 2018 11th UK Neuromuscular Translational Research Conference.

Awarded first place in poster competition at the 2019 UCL Institute of Neurology research symposium.

Awarded first place in poster competition at the 2018 MRC-DTP conference.

Gave oral presentation at the 2020 International Parkinson Disease Genomics Consortium (IPDGC) virtual meeting.

7. References

1. Beecroft SJ, Cortese A, Sullivan R, Yau WY, Dyer Z, Wu TY, et al. A Maori specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele. Brain. 2020;143(9):2673-80.
2. Cortese A, Simone R, Sullivan R, Vandrovcova J, Tariq H, Yau WY, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet. 2019;51(4):649-58.
3. Sullivan R, Yau WY, O'Connor E, Houlden H. Spinocerebellar ataxia: an update. J Neurol. 2019;266(2):533-44.
4. Carter CO. Monogenic disorders. J Med Genet. 1977;14(5):316-20.
5. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. Am J Hum Genet. 1988;42(5):677-93.
6. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet. 2007;80(4):588-604.
7. Cooper DN, Chen JM, Ball EV, Howells K, Mort M, Phillips AD, et al. Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. Hum Mutat. 2010;31(6):631-55.
8. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013;14(10):681-91.
9. Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic Analysis in the Age of Human Genome Sequencing. Cell. 2019;177(1):70-84.
10. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet. 2011;48(9):580-9.
11. Kuhlenbaumer G, Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. Hum Mutat. 2011;32(2):144-51.
12. Hernandez DG, Reed X, Singleton AB. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. J Neurochem. 2016;139 Suppl 1:59-74.
13. Fernandez-Marmiesse A, Gouveia S, Couce ML. NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment. Curr Med Chem. 2018;25(3):404-32.
14. Fahn S, Jankovic J, Hallett M. Principles and practice of movement disorders. 2nd ed. Edinburgh ; New York: Elsevier/Saunders; 2011. vii, 548 p. p.
15. Fahn S. Classification of movement disorders. Mov Disord. 2011;26(6):947-57.
16. Schirinzi T, Canevelli M, Suppa A, Bologna M, Marsili L. The continuum between neurodegeneration, brain plasticity, and movement: a critical appraisal. Rev Neurosci. 2020;31(7):723-42.
17. Murthy M, Cheng YY, Holton JL, Bettencourt C. Neurodegenerative movement disorders: an epigenetics perspective and promise for the future. Neuropathol Appl Neurobiol. 2021.
18. De Jager PL, Yang HS, Bennett DA. Deconstructing and targeting the genomic architecture of human neurodegeneration. Nat Neurosci. 2018;21(10):1310-7.
19. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell. 1993;72(6):971-83.
20. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science. 1996;271(5254):1423-7.
21. Ashizawa T, Xia G. Ataxia. Continuum (Minneap Minn). 2016;22(4 Movement Disorders):1208-26.
22. Ruano L, Melo C, Silva MC, Coutinho P. The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. Neuroepidemiology. 2014;42(3):174-83.
23. Paulson HL, Shakkottai VG, Clark HB, Orr HT. Polyglutamine spinocerebellar ataxias - from genes to potential treatments. Nat Rev Neurosci. 2017;18(10):613-26.

24. de Castilhos RM, Furtado GV, Gheno TC, Schaeffer P, Russo A, Barsottini O, et al. Spinocerebellar ataxias in Brazil--frequencies and modulating effects of related genes. Cerebellum. 2014;13(1):17-28.
25. Zaltzman R, Sharony R, Klein C, Gordon CR. Spinocerebellar ataxia type 3 in Israel: phenotype and genotype of a Jew Yemenite subpopulation. J Neurol. 2016;263(11):2207-14.
26. Paradisi I, Ikonomu V, Arias S. Spinocerebellar ataxias in Venezuela: genetic epidemiology and their most likely ethnic descent. J Hum Genet. 2016;61(3):215-22.
27. Harding AE. The clinical features and classification of the late onset autosomal dominant cerebellar ataxias. A study of 11 families, including descendants of the 'the Drew family of Walworth'. Brain. 1982;105(Pt 1):1-28.
28. Bird TD. Hereditary Ataxia Overview. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. GeneReviews((R)). Seattle (WA)1993.
29. Nibbeling EAR, Duarri A, Verschuuren-Bemelmans CC, Fokkens MR, Karjalainen JM, Smeets C, et al. Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia. Brain. 2017;140(11):2860-78.
30. Winkelmann J, Lin L, Schormair B, Kornum BR, Faraco J, Plazzi G, et al. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum Mol Genet. 2012;21(10):2205-10.
31. Pfeffer G, Blakely EL, Alston CL, Hassani A, Boggild M, Horvath R, et al. Adult-onset spinocerebellar ataxia syndromes due to MTATP6 mutations. J Neurol Neurosurg Psychiatry. 2012;83(9):883-6.
32. Gennarino VA, Palmer EE, McDonell LM, Wang L, Adamski CJ, Koire A, et al. A Mild PUM1 Mutation Is Associated with Adult-Onset Ataxia, whereas Haploinsufficiency Causes Developmental Delay and Seizures. Cell. 2018;172(5):924-36 e11.
33. Marras C, Lang A, van de Warrenburg BP, Sue CM, Tabrizi SJ, Bertram L, et al. Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force. Mov Disord. 2016;31(4):436-57.
34. Rossi M, Perez-Lloret S, Doldan L, Cerquetti D, Balej J, Millar Vernetti P, et al. Autosomal dominant cerebellar ataxias: a systematic review of clinical features. Eur J Neurol. 2014;21(4):607-15.
35. Galatolo D, Tessa A, Filla A, Santorelli FM. Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis. Neurogenetics. 2018;19(1):1-8.
36. Koht J, Tallaksen CM. Cerebellar ataxia in the eastern and southern parts of Norway. Acta Neurol Scand Suppl. 2007;187:76-9.
37. Anheim M, Tranchant C, Koenig M. The autosomal recessive cerebellar ataxias. N Engl J Med. 2012;366(7):636-46.
38. Burk K. Friedreich Ataxia: current status and future prospects. Cerebellum Ataxias. 2017;4:4.
39. Pandolfo M. Friedreich ataxia. Arch Neurol. 2008;65(10):1296-303.
40. Epplen C, Epplen JT, Frank G, Miterski B, Santos EJ, Schols L. Differential stability of the (GAA)n tract in the Friedreich ataxia (STM7) gene. Hum Genet. 1997;99(6):834-6.
41. Chen OS, Hemenway S, Kaplan J. Inhibition of Fe-S cluster biosynthesis decreases mitochondrial iron export: evidence that Yfh1p affects Fe-S cluster synthesis. Proc Natl Acad Sci U S A. 2002;99(19):12321-6.
42. Yoon T, Cowan JA. Frataxin-mediated iron delivery to ferrochelatase in the final step of heme biosynthesis. J Biol Chem. 2004;279(25):25943-6.
43. Bradley JL, Blake JC, Chamberlain S, Thomas PK, Cooper JM, Schapira AH. Clinical, biochemical and molecular genetic correlations in Friedreich's ataxia. Hum Mol Genet. 2000;9(2):275-82.
44. Vermeer S, van de Warrenburg BP, Willemsen MA, Cluitmans M, Scheffer H, Kremer BP, et al. Autosomal recessive cerebellar ataxias: the current state of affairs. J Med Genet. 2011;48(10):651-9.
45. Ouahchi K, Arita M, Kayden H, Hentati F, Ben Hamida M, Sokol R, et al. Ataxia with isolated vitamin E deficiency is caused by mutations in the alpha-tocopherol transfer protein. Nat Genet. 1995;9(2):141-5.

46. Engert JC, Berube P, Mercier J, Dore C, Lepage P, Ge B, et al. ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. Nat Genet. 2000;24(2):120-5.
47. Woods CG, Bundey SE, Taylor AM. Unusual features in the inheritance of ataxia telangiectasia. Hum Genet. 1990;84(6):555-62.
48. Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science. 1995;268(5218):1749-53.
49. Beaudin M, Matilla-Duenas A, Soong BW, Pedroso JL, Barsottini OG, Mitoma H, et al. The Classification of Autosomal Recessive Cerebellar Ataxias: a Consensus Statement from the Society for Research on the Cerebellum and Ataxias Task Force. Cerebellum. 2019;18(6):1098-125.
50. Traschutz A, Cortese A, Reich S, Dominik N, Faber J, Jacobi H, et al. Natural History, Phenotypic Spectrum, and Discriminative Features of Multisystemic RFC1 Disease. Neurology. 2021;96(9):e1369-e82.
51. Umeh CC, Polydefkis M, Chaudhry V, Zee DS. Sweat Gland Denervation in Cerebellar Ataxia with Neuropathy and Vestibular Areflexia Syndrome (CANVAS). Mov Disord Clin Pract. 2017;4(1):46-8.
52. Szmulewicz DJ, McLean CA, MacDougall HG, Roberts L, Storey E, Halmagyi GM. CANVAS an update: clinical presentation, investigation and management. J Vestib Res. 2014;24(5-6):465-74.
53. Cortese A, Tozza S, Yau WY, Rossi S, Beecroft SJ, Jaunmuktane Z, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. Brain. 2020;143(2):480-90.
54. Sullivan R, Kaiyrzhanov R, Houlden H. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome: genetic and clinical insights. Curr Opin Neurol. 2021;34(4):556-64.
55. Szmulewicz DJ, Waterston JA, MacDougall HG, Mossman S, Chancellor AM, McLean CA, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS): a review of the clinical features and video-oculographic diagnosis. Ann N Y Acad Sci. 2011;1233:139-47.
56. Infante J, Garcia A, Serrano-Cardenas KM, Gonzalez-Aguado R, Gazulla J, de Lucas EM, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS) with chronic cough and preserved muscle stretch reflexes: evidence for selective sparing of afferent Ia fibres. J Neurol. 2018;265(6):1454-62.
57. Burke D, Halmagyi GM. Normal tendon reflexes despite absent sensory nerve action potentials in CANVAS: a neurophysiological study. J Neurol Sci. 2018;387:75-9.
58. Szmulewicz DJ, Waterston JA, Halmagyi GM, Mossman S, Chancellor AM, McLean CA, et al. Sensory neuropathy as part of the cerebellar ataxia neuropathy vestibular areflexia syndrome. Neurology. 2011;76(22):1903-10.
59. Szmulewicz DJ, Seiderer L, Halmagyi GM, Storey E, Roberts L. Neurophysiological evidence for generalized sensory neuronopathy in cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome. Muscle Nerve. 2015;51(4):600-3.
60. Migliaccio AA, Halmagyi GM, McGarvie LA, Cremer PD. Cerebellar ataxia with bilateral vestibulopathy: description of a syndrome and its characteristic clinical sign. Brain. 2004;127(Pt 2):280-93.
61. Szmulewicz DJ, McLean CA, Rodriguez ML, Chancellor AM, Mossman S, Lamont D, et al. Dorsal root ganglionopathy is responsible for the sensory impairment in CANVAS. Neurology. 2014;82(16):1410-5.
62. Wu TY, Taylor JM, Kilfoyle DH, Smith AD, McGuinness BJ, Simpson MP, et al. Autonomic dysfunction is a major feature of cerebellar ataxia, neuropathy, vestibular areflexia 'CANVAS' syndrome. Brain. 2014;137(Pt 10):2649-56.
63. Paley GL, Kung NH, Bucelli RC, Margolis TP, Van Stavern GP. Cerebellar Ataxia With Neuropathy and Vestibular Areflexia Syndrome Presenting With Neurotrophic Keratopathy. J Neuroophthalmol. 2018;38(3):342-3.
64. Szmulewicz DJ, Merchant SN, Halmagyi GM. Cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome: a histopathologic case report. Otol Neurotol. 2011;32(8):e63-5.
65. Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, et al. De novo expansion of a (CAG)n repeat in sporadic Huntington's disease. Nat Genet. 1993;5(2):168-73.

66. Houge G, Bruland O, Bjornevoll I, Hayden MR, Semaka A. De novo Huntington disease caused by 26-44 CAG repeat expansion on a low-risk haplotype. Neurology. 2013;81(12):1099-100.
67. Paulson H. Repeat expansion diseases. Handb Clin Neurol. 2018;147:105-23.
68. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell. 1992;68(4):799-808.
69. Brussino A, Graziano C, Giobbe D, Ferrone M, Dragone E, Arduino C, et al. Spinocerebellar ataxia type 12 identified in two Italian families may mimic sporadic ataxia. Mov Disord. 2010;25(9):1269-73.
70. Zhang N, Ashizawa T. RNA toxicity and foci formation in microsatellite expansion diseases. Curr Opin Genet Dev. 2017;44:17-29.
71. Wojciechowska M, Krzyzosiak WJ. Cellular toxicity of expanded RNA repeats: focus on RNA foci. Hum Mol Genet. 2011;20(19):3811-21.
72. Lin X, Ashizawa T. Recent progress in spinocerebellar ataxia type-10 (SCA10). Cerebellum. 2005;4(1):37-42.
73. White M, Xia G, Gao R, Wakamiya M, Sarkar PS, McFarland K, et al. Transgenic mice with SCA10 pentanucleotide repeats show motor phenotype and susceptibility to seizure: a toxic RNA gain-of-function model. J Neurosci Res. 2012;90(3):706-14.
74. Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y, et al. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet. 2011;89(1):121-30.
75. Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, Ebner TJ, et al. RNA gain-of-function in spinocerebellar ataxia type 8. PLoS Genet. 2009;5(8):e1000600.
76. Cho DH, Thienes CP, Mahoney SE, Analau E, Filippova GN, Tapscott SJ. Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. Mol Cell. 2005;20(3):483-9.
77. Rudnicki DD, Holmes SE, Lin MW, Thornton CA, Ross CA, Margolis RL. Huntington's disease--like 2 is associated with CUG repeat-containing RNA foci. Ann Neurol. 2007;61(3):272-82.
78. Corral-Juan M, Serrano-Munuera C, Rabano A, Cota-Gonzalez D, Segarra-Roca A, Ispierto L, et al. Clinical, genetic and neuropathological characterization of spinocerebellar ataxia type 37. Brain. 2018;141(7):1981-97.
79. Akimoto C, Volk AE, van Blitterswijk M, Van den Broeck M, Leblond CS, Lumbroso S, et al. A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. J Med Genet. 2014;51(6):419-24.
80. Ashley EA. The precision medicine initiative: a new national effort. JAMA. 2015;313(21):2119-20.
81. Cagnoli C, Brussino A, Mancini C, Ferrone M, Orsi L, Salmin P, et al. Spinocerebellar Ataxia Tethering PCR: A Rapid Genetic Test for the Diagnosis of Spinocerebellar Ataxia Types 1, 2, 3, 6, and 7 by PCR and Capillary Electrophoresis. J Mol Diagn. 2018;20(3):289-97.
82. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol. 2013;14(7):405.
83. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4(4):265-70.
84. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics. 2012;13:375.
85. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif. 2015;3:1-8.
86. Dolzhenko E, van Vugt J, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. 2017;27(11):1895-903.
87. Johnson J, Mercado-Ayon E, Mercado-Ayon Y, Dong YN, Halawani S, Ngaba L, et al. Mitochondrial dysfunction in the development and progression of neurodegenerative diseases. Arch Biochem Biophys. 2021;702:108698.

88. Jornayvaz FR, Shulman GI. Regulation of mitochondrial biogenesis. Essays Biochem. 2010;47:69-84.
89. Virbasius JV, Scarpulla RC. Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: a potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis. Proc Natl Acad Sci U S A. 1994;91(4):1309-13.
90. Wu Z, Puigserver P, Andersson U, Zhang C, Adelmant G, Mootha V, et al. Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. Cell. 1999;98(1):115-24.
91. Baar K, Wende AR, Jones TE, Marison M, Nolte LA, Chen M, et al. Adaptations of skeletal muscle to exercise: rapid increase in the transcriptional coactivator PGC-1. FASEB J. 2002;16(14):1879-86.
92. Meirhaeghe A, Crowley V, Lenaghan C, Lelliott C, Green K, Stewart A, et al. Characterization of the human, mouse and rat PGC1 beta (peroxisome-proliferator-activated receptor-gamma co-activator 1 beta) gene in vitro and in vivo. Biochem J. 2003;373(Pt 1):155-65.
93. Hardie DG. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. Nat Rev Mol Cell Biol. 2007;8(10):774-85.
94. Reznick RM, Zong H, Li J, Morino K, Moore IK, Yu HJ, et al. Aging-associated reductions in AMP-activated protein kinase activity and mitochondrial biogenesis. Cell Metab. 2007;5(2):151-6.
95. Nisoli E, Clementi E, Paolucci C, Cozzi V, Tonello C, Sciorati C, et al. Mitochondrial biogenesis in mammals: the role of endogenous nitric oxide. Science. 2003;299(5608):896-9.
96. Lagouge M, Argmann C, Gerhart-Hines Z, Meziane H, Lerin C, Daussin F, et al. Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1alpha. Cell. 2006;127(6):1109-22.
97. Wu Z, Huang X, Feng Y, Handschin C, Feng Y, Gullicksen PS, et al. Transducer of regulated CREB-binding proteins (TORCs) induce PGC-1alpha transcription and mitochondrial biogenesis in muscle cells. Proc Natl Acad Sci U S A. 2006;103(39):14379-84.
98. Seth A, Steel JH, Nichol D, Pocock V, Kumaran MK, Fritah A, et al. The transcriptional corepressor RIP140 regulates oxidative metabolism in skeletal muscle. Cell Metab. 2007;6(3):236-45.
99. White R, Morganstein D, Christian M, Seth A, Herzog B, Parker MG. Role of RIP140 in metabolic tissues: connections to disease. FEBS Lett. 2008;582(1):39-45.
100. Akimoto T, Pohnert SC, Li P, Zhang M, Gumbs C, Rosenberg PB, et al. Exercise stimulates Pgc-1alpha transcription in skeletal muscle through activation of the p38 MAPK pathway. J Biol Chem. 2005;280(20):19587-93.
101. Marmolino D, Manto M, Acquaviva F, Vergara P, Ravella A, Monticelli A, et al. PGC-1alpha down-regulation affects the antioxidant response in Friedreich's ataxia. PLoS One. 2010;5(4):e10025.
102. Ciron C, Zheng L, Bobela W, Knott GW, Leone TC, Kelly DP, et al. PGC-1alpha activity in nigral dopamine neurons determines vulnerability to alpha-synuclein. Acta Neuropathol Commun. 2015;3:16.
103. St-Pierre J, Drori S, Uldry M, Silvaggi JM, Rhee J, Jager S, et al. Suppression of reactive oxygen species and neurodegeneration by the PGC-1 transcriptional coactivators. Cell. 2006;127(2):397-408.
104. Yeung AWK, Tzvetkov NT, Georgieva MG, Ognyanov IV, Kordos K, Jozwik A, et al. Reactive Oxygen Species and Their Impact in Neurodegenerative Diseases: Literature Landscape Analysis. Antioxid Redox Signal. 2021;34(5):402-20.
105. Loane DJ, Faden AI. Neuroprotection for traumatic brain injury: translational challenges and emerging therapeutic strategies. Trends Pharmacol Sci. 2010;31(12):596-604.
106. Madesh M, Hajnoczky G. VDAC-dependent permeabilization of the outer mitochondrial membrane by superoxide induces rapid and massive cytochrome c release. J Cell Biol. 2001;155(6):1003-15.
107. Oliveira PJ, Wallace KB. Depletion of adenine nucleotide translocator protein in heart mitochondria from doxorubicin-treated rats--relevance for mitochondrial dysfunction. Toxicology. 2006;220(2-3):160-8.
108. Halestrap AP, Woodfield KY, Connern CP. Oxidative stress, thiol reagents, and membrane potential modulate the mitochondrial permeability transition by affecting nucleotide binding to the adenine nucleotide translocase. J Biol Chem. 1997;272(6):3346-54.

109. Shi Y, Nikulenkov F, Zawacka-Pankau J, Li H, Gabdoulline R, Xu J, et al. ROS-dependent activation of JNK converts p53 into an efficient inhibitor of oncogenes leading to robust apoptosis. Cell Death Differ. 2014;21(4):612-23.
110. Ott M, Robertson JD, Gogvadze V, Zhivotovsky B, Orrenius S. Cytochrome c release from mitochondria proceeds by a two-step process. Proc Natl Acad Sci U S A. 2002;99(3):1259-63.
111. Jasoliya MJ, McMackin MZ, Henderson CK, Perlman SL, Cortopassi GA. Frataxin deficiency impairs mitochondrial biogenesis in cells, mice and humans. Hum Mol Genet. 2017;26(14):2627-33.
112. Lin H, Magrane J, Rattelle A, Stepanova A, Galkin A, Clark EM, et al. Early cerebellar deficits in mitochondrial biogenesis and respiratory chain complexes in the KIKO mouse model of Friedreich ataxia. Dis Model Mech. 2017;10(11):1343-52.
113. Purroy R, Medina-Carbonero M, Ros J, Tamarit J. Frataxin-deficient cardiomyocytes present an altered thiol-redox state which targets actin and pyruvate dehydrogenase. Redox Biol. 2020;32:101520.
114. Pastore A, Tozzi G, Gaeta LM, Bertini E, Serafini V, Di Cesare S, et al. Actin glutathionylation increases in fibroblasts of patients with Friedreich's ataxia: a potential role in the pathogenesis of the disease. J Biol Chem. 2003;278(43):42588-95.
115. Bulteau AL, Planamente S, Jornea L, Dur A, Lesuisse E, Camadro JM, et al. Changes in mitochondrial glutathione levels and protein thiol oxidation in yfh1 yeast cells and the lymphoblasts of patients with Friedreich's ataxia. Biochim Biophys Acta. 2012;1822(2):212-25.
116. Abeti R, Uzun E, Renganathan I, Honda T, Pook MA, Giunti P. Targeting lipid peroxidation and mitochondrial imbalance in Friedreich's ataxia. Pharmacol Res. 2015;99:344-50.
117. Marmolino D, Acquaviva F, Pinelli M, Monticelli A, Castaldo I, Filla A, et al. PPAR-gamma agonist Azelaoyl PAF increases frataxin protein and mRNA expression: new implications for the Friedreich's ataxia therapy. Cerebellum. 2009;8(2):98-103.
118. Meyer JN, Leuthner TC, Luz AL. Mitochondrial fusion, fission, and mitochondrial toxicity. Toxicology. 2017;391:42-53.
119. Tilokani L, Nagashima S, Paupe V, Prudent J. Mitochondrial dynamics: overview of molecular mechanisms. Essays Biochem. 2018;62(3):341-60.
120. Braschi E, Zunino R, McBride HM. MAPL is a new mitochondrial SUMO E3 ligase that regulates mitochondrial fission. EMBO Rep. 2009;10(7):748-54.
121. Pagliuso A, Cossart P, Stavru F. The ever-growing complexity of the mitochondrial fission machinery. Cell Mol Life Sci. 2018;75(3):355-74.
122. Zaja I, Bai X, Liu Y, Kikuchi C, Dosenovic S, Yan Y, et al. Cdk1, PKCdelta and calcineurin-mediated Drp1 pathway contributes to mitochondrial fission-induced cardiomyocyte death. Biochem Biophys Res Commun. 2014;453(4):710-21.
123. Li J, Wang Y, Wang Y, Wen X, Ma XN, Chen W, et al. Pharmacological activation of AMPK prevents Drp1-mediated mitochondrial fission and alleviates endoplasmic reticulum stress-associated endothelial dysfunction. J Mol Cell Cardiol. 2015;86:62-74.
124. Lee H, Yoon Y. Mitochondrial Membrane Dynamics-Functional Positioning of OPA1. Antioxidants (Basel). 2018;7(12).
125. Lee H, Smith SB, Yoon Y. The short variant of the mitochondrial dynamin OPA1 maintains mitochondrial energetics and cristae structure. J Biol Chem. 2017;292(17):7115-30.
126. Chan DC. Mitochondria: dynamic organelles in disease, aging, and development. Cell. 2006;125(7):1241-52.
127. Hu C, Huang Y, Li L. Drp1-Dependent Mitochondrial Fission Plays Critical Roles in Physiological and Pathological Progresses in Mammals. Int J Mol Sci. 2017;18(1).
128. Itoh K, Nakamura K, Iijima M, Sesaki H. Mitochondrial dynamics in neurodegeneration. Trends Cell Biol. 2013;23(2):64-71.
129. Corrado M, Scorrano L, Campello S. Mitochondrial dynamics in cancer and neurodegenerative and neuroinflammatory diseases. Int J Cell Biol. 2012;2012:729290.
130. Bernard-Marissal N, van Hameren G, Juneja M, Pellegrino C, Louhivuori L, Bartesaghi L, et al. Altered interplay between endoplasmic reticulum and mitochondria in Charcot-Marie-Tooth type 2A neuropathy. Proc Natl Acad Sci U S A. 2019;116(6):2328-37.

131. Santos D, Esteves AR, Silva DF, Januario C, Cardoso SM. The Impact of Mitochondrial Fusion and Fission Modulation in Sporadic Parkinson's Disease. Mol Neurobiol. 2015;52(1):573-86.
132. Lefevre S, Sliwa D, Rustin P, Camadro JM, Santos R. Oxidative stress induces mitochondrial fragmentation in frataxin-deficient cells. Biochem Biophys Res Commun. 2012;418(2):336-41.
133. Liu Y, Xu Z, Jin T, Xu K, Liu M, Xu H. Ferroptosis in Low-Grade Glioma: A New Marker for Diagnosis and Prognosis. Med Sci Monit. 2020;26:e921947.
134. Rhodes SL, Buchanan DD, Ahmed I, Taylor KD, Loriot MA, Sinsheimer JS, et al. Pooled analysis of iron-related genes in Parkinson's disease: association with transferrin. Neurobiol Dis. 2014;62:172-8.
135. Walker JE. The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. Q Rev Biophys. 1992;25(3):253-324.
136. Mitchell P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature. 1961;191:144-8.
137. Cocheme HM, Quin C, McQuaker SJ, Cabreiro F, Logan A, Prime TA, et al. Measurement of H2O2 within living Drosophila during aging using a ratiometric mass spectrometry probe targeted to the mitochondrial matrix. Cell Metab. 2011;13(3):340-50.
138. Palomero J, Pye D, Kabayo T, Spiller DG, Jackson MJ. In situ detection and measurement of intracellular reactive oxygen species in single isolated mature skeletal muscle fibers by real time fluorescence microscopy. Antioxid Redox Signal. 2008;10(8):1463-74.
139. Droge W. Free radicals in the physiological control of cell function. Physiol Rev. 2002;82(1):47-95.
140. Zorov DB, Bannikova SY, Belousov VV, Vyssokikh MY, Zorova LD, Isaev NK, et al. Reactive oxygen and nitrogen species: friends or foes? Biochemistry (Mosc). 2005;70(2):215-21.
141. Dai DF, Rabinovitch PS, Ungvari Z. Mitochondria and cardiovascular aging. Circ Res. 2012;110(8):1109-24.
142. Cornelius N, Wardman JH, Hargreaves IP, Neergheen V, Bie AS, Tumer Z, et al. Evidence of oxidative stress and mitochondrial dysfunction in spinocerebellar ataxia type 2 (SCA2) patient fibroblasts: Effect of coenzyme Q10 supplementation on these parameters. Mitochondrion. 2017;34:103-14.
143. Zorov DB, Juhaszova M, Sollott SJ. Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. Physiol Rev. 2014;94(3):909-50.
144. Zoccarato F, Toscano P, Alexandre A. Dopamine-derived dopaminochrome promotes H(2)O(2) release at mitochondrial complex I: stimulation by rotenone, control by Ca(2+), and relevance to Parkinson disease. J Biol Chem. 2005;280(16):15587-94.
145. Ferdinandusse S, Falkenberg KD, Koster J, Mooyer PA, Jones R, van Roermund CWT, et al. ACBD5 deficiency causes a defect in peroxisomal very long-chain fatty acid metabolism. J Med Genet. 2017;54(5):330-7.
146. Bartlett M, Nasiri N, Pressman R, Bademci G, Forghani I. First reported adult patient with retinal dystrophy and leukodystrophy caused by a novel ACBD5 variant: A case report and review of literature. Am J Med Genet A. 2021;185(4):1236-41.
147. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928-30.
148. Passmore JB, Carmichael RE, Schrader TA, Godinho LF, Ferdinandusse S, Lismont C, et al. Mitochondrial fission factor (MFF) is a critical regulator of peroxisome maturation. Biochim Biophys Acta Mol Cell Res. 2020;1867(7):118709.
149. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-82.
150. Bishop A, Kamoshita M, Passmore JB, Hacker C, Schrader TA, Waterham HR, et al. Fluorescent tools to analyse peroxisome-ER interactions in mammalian cells. Contact (Thousand Oaks). 2019;2.
151. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-2.
152. Ison J, Kalas M, Jonassen I, Bolser D, Uludag M, McWilliam H, et al. EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics. 2013;29(10):1325-32.

153. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49(D1):D605-D12.
154. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199-W205.
155. Dolzhenko E, Deshpande V, Schlesinger F, Krusche P, Petrovski R, Chen S, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics. 2019;35(22):4754-6.
156. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525-7.
157. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14(7):687-90.
158. Evidente VG, Gwinn-Hardy KA, Caviness JN, Gilman S. Hereditary ataxias. Mayo Clin Proc. 2000;75(5):475-90.
159. Schols L, Szymanski S, Peters S, Przuntek H, Epplen JT, Hardt C, et al. Genetic background of apparently idiopathic sporadic cerebellar ataxia. Hum Genet. 2000;107(2):132-7.
160. Abele M, Burk K, Schols L, Schwartz S, Besenthal I, Dichgans J, et al. The aetiology of sporadic adult-onset ataxia. Brain. 2002;125(Pt 5):961-8.
161. Muzaimi MB, Thomas J, Palmer-Smith S, Rosser L, Harper PS, Wiles CM, et al. Population based study of late onset cerebellar ataxia in south east Wales. J Neurol Neurosurg Psychiatry. 2004;75(8):1129-34.
162. Zingler VC, Cnyrim C, Jahn K, Weintz E, Fernbacher J, Frenzel C, et al. Causative factors and epidemiology of bilateral vestibulopathy in 255 patients. Ann Neurol. 2007;61(6):524-32.
163. Tang H, Hilton B, Musich PR, Fang DZ, Zou Y. Replication factor C1, the large subunit of replication factor C, is proteolytically truncated in Hutchinson-Gilford progeria syndrome. Aging Cell. 2012;11(2):363-5.
164. Hommelsheim CM, Frantzeskakis L, Huang M, Ulker B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci Rep. 2014;4:5052.
165. Chattopadhyay B, Gupta S, Gangopadhyay PK, Das SK, Roy T, Mukherjee SC, et al. Molecular analysis of GAA repeats and four linked bi-allelic markers in and around the frataxin gene in patients and normal populations from India. Ann Hum Genet. 2004;68(Pt 3):189-95.
166. Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NLM, Pope K, Smith KR, et al. Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. Am J Hum Genet. 2019;105(1):151-65.
167. Akcimen F, Ross JP, Bourassa CV, Liao C, Rochefort D, Gama MTD, et al. Investigation of the RFC1 Repeat Expansion in a Canadian and a Brazilian Ataxia Cohort: Identification of Novel Conformations. Front Genet. 2019;10:1219.
168. Aboud Syriani D, Wong D, Andani S, De Gusmao CM, Mao Y, Sanyoura M, et al. Prevalence of RFC1-mediated spinocerebellar ataxia in a North American ataxia cohort. Neurol Genet. 2020;6(3):e440.
169. Fan Y, Zhang S, Yang J, Mao CY, Yang ZH, Hu ZW, et al. No biallelic intronic AAGGG repeat expansion in RFC1 was found in patients with late-onset ataxia and MSA. Parkinsonism Relat Disord. 2020;73:1-2.
170. Dominik N, Galassi Deforie V, Cortese A, Houlden H. CANVAS: a late onset ataxia due to biallelic intronic AAGGG expansions. J Neurol. 2021;268(3):1119-26.
171. Scriba CK, Beecroft SJ, Clayton JS, Cortese A, Sullivan R, Yau WY, et al. A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families. Brain. 2020;143(10):2904-10.
172. Tsuchiya M, Nan H, Koh K, Ichinose Y, Gao L, Shimozono K, et al. RFC1 repeat expansion in Japanese patients with late-onset cerebellar ataxia. J Hum Genet. 2020;65(12):1143-7.
173. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011;12(12):236.
174. Rogers J. Retroposons defined. Nature. 1983;301(5900):460.
175. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, et al. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 2009;19(9):1516-26.

176. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab. 1999;67(3):183-93.
177. Dewannieux M, Heidmann T. Role of poly(A) tail length in Alu retrotransposition. Genomics. 2005;86(3):378-81.
178. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA. Alu repeats: a source for the genesis of primate microsatellites. Genomics. 1995;29(1):136-44.
179. Lazaropoulos M, Dong Y, Clark E, Greeley NR, Seyer LA, Brigatti KW, et al. Frataxin levels in peripheral tissue in Friedreich ataxia. Ann Clin Transl Neurol. 2015;2(8):831-42.
180. Tomida J, Masuda Y, Hiroaki H, Ishikawa T, Song I, Tsurimoto T, et al. DNA damage-induced ubiquitylation of RFC2 subunit of replication factor C complex. J Biol Chem. 2008;283(14):9071-9.
181. Overmeer RM, Gourdin AM, Giglia-Mari A, Kool H, Houtsmuller AB, Siegal G, et al. Replication factor C recruits DNA polymerase delta to sites of nucleotide excision repair but is not required for PCNA recruitment. Mol Cell Biol. 2010;30(20):4828-39.
182. McKinnon PJ. Maintaining genome stability in the nervous system. Nat Neurosci. 2013;16(11):1523-9.
183. Ciesiolka A, Jazurek M, Drazkowska K, Krzyzosiak WJ. Structural Characteristics of Simple RNA Repeats Associated with Disease and their Deleterious Protein Interactions. Front Cell Neurosci. 2017;11:97.
184. Philips AV, Timchenko LT, Cooper TA. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science. 1998;280(5364):737-41.
185. Cleary JD, Pattamatta A, Ranum LPW. Repeat-associated non-ATG (RAN) translation. J Biol Chem. 2018;293(42):16127-41.
186. Sznajder LJ, Thomas JD, Carrell EM, Reid T, McFarland KN, Cleary JD, et al. Intron retention induced by microsatellite expansions as a disease biomarker. Proc Natl Acad Sci U S A. 2018;115(16):4234-9.
187. Heras-Garvin A, Stefanova N. MSA: From basic mechanisms to experimental therapeutics. Parkinsonism Relat Disord. 2020.
188. Fanciulli A, Wenning GK. Multiple-system atrophy. N Engl J Med. 2015;372(3):249-63.
189. Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ, Trojanowski JQ, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology. 2008;71(9):670-6.
190. Salvesen L, Ullerup BH, Sunay FB, Brudek T, Lokkegaard A, Agander TK, et al. Changes in total cell numbers of the basal ganglia in patients with multiple system atrophy - A stereological study. Neurobiol Dis. 2015;74:104-13.
191. Quinn N. Multiple system atrophy--the nature of the beast. J Neurol Neurosurg Psychiatry. 1989;Suppl:78-89.
192. Papp MI, Kahn JE, Lantos PL. Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy-Drager syndrome). J Neurol Sci. 1989;94(1-3):79-100.
193. Gonzalez-Perez O, Alvarez-Buylla A. Oligodendrogenesis in the subventricular zone and the role of epidermal growth factor. Brain Res Rev. 2011;67(1-2):147-56.
194. McTigue DM, Tripathi RB. The life, death, and replacement of oligodendrocytes in the adult CNS. J Neurochem. 2008;107(1):1-19.
195. Bradl M, Lassmann H. Oligodendrocytes: biology and pathology. Acta Neuropathol. 2010;119(1):37-53.
196. Bunge RP. Glial cells and the central myelin sheath. Physiol Rev. 1968;48(1):197-251.
197. Burke RE, Dauer WT, Vonsattel JP. A critical evaluation of the Braak staging scheme for Parkinson's disease. Ann Neurol. 2008;64(5):485-91.
198. Baumann N, Pham-Dinh D. Biology of oligodendrocyte and myelin in the mammalian central nervous system. Physiol Rev. 2001;81(2):871-927.
199. Nave KA, Trapp BD. Axon-glial signaling and the glial support of axon function. Annu Rev Neurosci. 2008;31:535-61.
200. Spillantini MG, Crowther RA, Jakes R, Cairns NJ, Lantos PL, Goedert M. Filamentous alpha-synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. Neurosci Lett. 1998;251(3):205-8.

201. Wakabayashi K, Yoshimoto M, Tsuji S, Takahashi H. Alpha-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. Neurosci Lett. 1998;249(2-3):180-2.
202. Iwai A, Masliah E, Yoshimoto M, Ge N, Flanagan L, de Silva HA, et al. The precursor protein of non-A beta component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system. Neuron. 1995;14(2):467-75.
203. Duda JE, Lee VM, Trojanowski JQ. Neuropathology of synuclein aggregates. J Neurosci Res. 2000;61(2):121-7.
204. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. Nature. 1997;388(6645):839-40.
205. Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. Proc Natl Acad Sci U S A. 1998;95(11):6469-73.
206. Giasson BI, Mabon ME, Duda JE, Montine TJ, Robertson D, Hurtig HI, et al. Tau and 14-3-3 in glial cytoplasmic inclusions of multiple system atrophy. Acta Neuropathol. 2003;106(3):243-50.
207. Arai N, Nishimura M, Oda M, Morimatsu Y, Ohe R, Nagatomo H. Immunohistochemical expression of microtubule-associated protein 5 (MAP5) in glial cells in multiple system atrophy. J Neurol Sci. 1992;109(1):102-6.
208. Kawamoto Y, Akiguchi I, Shirakashi Y, Honjo Y, Tomimoto H, Takahashi R, et al. Accumulation of Hsc70 and Hsp70 in glial cytoplasmic inclusions in patients with multiple system atrophy. Brain Res. 2007;1136(1):219-27.
209. Huang Y, Song YJ, Murphy K, Holton JL, Lashley T, Revesz T, et al. LRRK2 and parkin immunoreactivity in multiple system atrophy inclusions. Acta Neuropathol. 2008;116(6):639-46.
210. Lindersson E, Lundvig D, Petersen C, Madsen P, Nyengaard JR, Hojrup P, et al. p25alpha Stimulates alpha-synuclein aggregation and is co-localized with aggregated alpha-synuclein in alpha-synucleinopathies. J Biol Chem. 2005;280(7):5703-15.
211. Song YJ, Lundvig DM, Huang Y, Gai WP, Blumbergs PC, Hojrup P, et al. p25alpha relocalizes in oligodendroglia from myelin to cytoplasmic inclusions in multiple system atrophy. Am J Pathol. 2007;171(4):1291-303.
212. Miller DW, Johnson JM, Solano SM, Hollingsworth ZR, Standaert DG, Young AB. Absence of alpha-synuclein mRNA expression in normal and multiple system atrophy oligodendroglia. J Neural Transm (Vienna). 2005;112(12):1613-24.
213. Reyes JF, Rey NL, Bousset L, Melki R, Brundin P, Angot E. Alpha-synuclein transfers from neurons to oligodendrocytes. Glia. 2014;62(3):387-98.
214. Bukhatwa S, Zeng BY, Rose S, Jenner P. A comparison of changes in proteasomal subunit expression in the substantia nigra in Parkinson's disease, multiple system atrophy and progressive supranuclear palsy. Brain Res. 2010;1326:174-83.
215. Fellner L, Irschick R, Schanda K, Reindl M, Klimaschewski L, Poewe W, et al. Toll-like receptor 4 is required for alpha-synuclein dependent activation of microglia and astroglia. Glia. 2013;61(3):349-60.
216. Stefanova N, Kaufmann WA, Humpel C, Poewe W, Wenning GK. Systemic proteasome inhibition triggers neurodegeneration in a transgenic mouse model expressing human alpha-synuclein under oligodendrocyte promoter: implications for multiple system atrophy. Acta Neuropathol. 2012;124(1):51-65.
217. Peng C, Gathagan RJ, Covell DJ, Medellin C, Stieber A, Robinson JL, et al. Cellular milieu imparts distinct pathological alpha-synuclein strains in alpha-synucleinopathies. Nature. 2018;557(7706):558-63.
218. Ubhi K, Lee PH, Adame A, Inglis C, Mante M, Rockenstein E, et al. Mitochondrial inhibitor 3-nitroproprionic acid enhances oxidative modification of alpha-synuclein in a transgenic mouse model of multiple system atrophy. J Neurosci Res. 2009;87(12):2728-39.
219. Monzio Compagnoni G, Di Fonzo A. Understanding the pathogenesis of multiple system atrophy: state of the art and future perspectives. Acta Neuropathol Commun. 2019;7(1):113.
220. Watts JC, Giles K, Oehler A, Middleton L, Dexter DT, Gentleman SM, et al. Transmission of multiple system atrophy prions to transgenic mice. Proc Natl Acad Sci U S A. 2013;110(48):19555-60.

221. Ozawa T, Paviour D, Quinn NP, Josephs KA, Sangha H, Kilford L, et al. The spectrum of pathological involvement of the striatonigral and olivopontocerebellar systems in multiple system atrophy: clinicopathological correlations. Brain. 2004;127(Pt 12):2657-71.
222. Jellinger K. Neuropathology. In: Wenning G., Fanciuli A. (ed). In: Wenning G, Fanciuli, editor. Multiple Systems Atrophy. Vienna: Springer; 2014.
223. Jellinger KA. Neuropathology of multiple system atrophy: new thoughts about pathogenesis. Mov Disord. 2014;29(14):1720-41.
224. Batla A, De Pablo-Fernandez E, Erro R, Reich M, Calandra-Buonaura G, Barbosa P, et al. Young-onset multiple system atrophy: Clinical and pathological features. Mov Disord. 2018;33(7):1099-107.
225. Braak H, de Vos RA, Bohl J, Del Tredici K. Gastric alpha-synuclein immunoreactive inclusions in Meissner's and Auerbach's plexuses in cases staged for Parkinson's disease-related brain pathology. Neurosci Lett. 2006;396(1):67-72.
226. Jecmenica-Lukic M, Poewe W, Tolosa E, Wenning GK. Premotor signs and symptoms of multiple system atrophy. Lancet Neurol. 2012;11(4):361-8.
227. Cykowski MD, Coon EA, Powell SZ, Jenkins SM, Benarroch EE, Low PA, et al. Expanding the spectrum of neuronal pathology in multiple system atrophy. Brain. 2015;138(Pt 8):2293-309.
228. Koga S, Aoki N, Uitti RJ, van Gerpen JA, Cheshire WP, Josephs KA, et al. When DLB, PD, and PSP masquerade as MSA: an autopsy study of 134 patients. Neurology. 2015;85(5):404-12.
229. Gama RL, Tavora DG, Bomfim RC, Silva CE, de Bruin VM, de Bruin PF. Sleep disturbances and brain MRI morphometry in Parkinson's disease, multiple system atrophy and progressive supranuclear palsy - a comparative study. Parkinsonism Relat Disord. 2010;16(4):275-9.
230. Wullner U, Abele M, Schmitz-Huebsch T, Wilhelm K, Benecke R, Deuschl G, et al. Probable multiple system atrophy in a German family. J Neurol Neurosurg Psychiatry. 2004;75(6):924-5.
231. Itoh K, Kasai T, Tsuji Y, Saito K, Mizuta I, Harada Y, et al. Definite familial multiple system atrophy with unknown genetics. Neuropathology. 2014;34(3):309-13.
232. Hara K, Momose Y, Tokiguchi S, Shimohata M, Terajima K, Onodera O, et al. Multiplex families with multiple system atrophy. Arch Neurol. 2007;64(4):545-51.
233. Sailer A, Scholz SW, Nalls MA, Schulte C, Federoff M, Price TR, et al. A genome-wide association study in multiple system atrophy. Neurology. 2016;87(15):1591-8.
234. Multiple-System Atrophy Research C. Mutations in COQ2 in familial and sporadic multiple-system atrophy. N Engl J Med. 2013;369(3):233-44.
235. Zhao Q, Yang X, Tian S, An R, Zheng J, Xu Y. Association of the COQ2 V393A variant with risk of multiple system atrophy in East Asians: a case-control study and meta-analysis of the literature. Neurol Sci. 2016;37(3):423-30.
236. Barca E, Kleiner G, Tang G, Ziosi M, Tadesse S, Masliah E, et al. Decreased Coenzyme Q10 Levels in Multiple System Atrophy Cerebellum. J Neuropathol Exp Neurol. 2016;75(7):663-72.
237. Scholz SW, Houlden H, Schulte C, Sharma M, Li A, Berg D, et al. SNCA variants are associated with increased risk for multiple system atrophy. Ann Neurol. 2009;65(5):610-4.
238. Al-Chalabi A, Durr A, Wood NW, Parkinson MH, Camuzat A, Hulot JS, et al. Genetic variants of the alpha-synuclein gene SNCA are associated with multiple system atrophy. PLoS One. 2009;4(9):e7114.
239. Gwinn K, Devine MJ, Jin LW, Johnson J, Bird T, Muenter M, et al. Clinical features, with video documentation, of the original familial lewy body parkinsonism caused by alpha-synuclein triplication (Iowa kindred). Mov Disord. 2011;26(11):2134-6.
240. Kiely AP, Asi YT, Kara E, Limousin P, Ling H, Lewis P, et al. alpha-Synucleinopathy associated with G51D SNCA mutation: a link between Parkinson's disease and multiple system atrophy? Acta Neuropathol. 2013;125(5):753-69.
241. Mitsui J, Matsukawa T, Sasaki H, Yabe I, Matsushima M, Durr A, et al. Variants associated with Gaucher disease in multiple system atrophy. Ann Clin Transl Neurol. 2015;2(4):417-26.
242. Yang X, An R, Xi J, Zhen J, Chen Y, Huang H, et al. Sequence TMEM230 gene in patients with multiple system atrophy in a southwest Chinese population: A pilot study. J Neurol Sci. 2017;375:264-5.
243. Ferguson MC, Garland EM, Hedges L, Womack-Nunley B, Hamid R, Phillips JA, 3rd, et al. SHC2 gene copy number in multiple system atrophy (MSA). Clin Auton Res. 2014;24(1):25-30.

244. Sasaki H, Emi M, Iijima H, Ito N, Sato H, Yabe I, et al. Copy number loss of (src homology 2 domain containing)-transforming protein 2 (SHC2) gene: discordant loss in monozygotic twins and frequent loss in patients with multiple system atrophy. Mol Brain. 2011;4:24.
245. Lee K, Nguyen KD, Sun C, Liu M, Zafar F, Saetern J, et al. LRRK2 p.Ile1371Val Mutation in a Case with Neuropathologically Confirmed Multi-System Atrophy. J Parkinsons Dis. 2018;8(1):93-100.
246. Riboldi GM, Palma JA, Cortes E, Iida MA, Sikder T, Henderson B, et al. Early-onset pathologically proven multiple system atrophy with LRRK2 G2019S mutation. Mov Disord. 2019;34(7):1080-2.
247. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol. 2008;7(7):583-90.
248. Marras C, Alcalay RN, Caspell-Garcia C, Coffey C, Chan P, Duda JE, et al. Motor and nonmotor heterogeneity of LRRK2-related and idiopathic Parkinson's disease. Mov Disord. 2016;31(8):1192-202.
249. Nee LE, Gomez MR, Dambrosia J, Bale S, Eldridge R, Polinsky RJ. Environmental-occupational risk factors and familial associations in multiple system atrophy: a preliminary investigation. Clin Auton Res. 1991;1(1):9-13.
250. Vanacore N, Bonifati V, Fabbrini G, Colosimo C, De Michele G, Marconi R, et al. Case-control study of multiple system atrophy. Mov Disord. 2005;20(2):158-63.
251. Kim HJ, Jeon BS, Shin J, Lee WW, Park H, Jung YJ, et al. Should genetic testing for SCAs be included in the diagnostic workup for MSA? Neurology. 2014;83(19):1733-8.
252. Sullivan R, Yau WY, Chelban V, Rossi S, Dominik N, O'Connor E, et al. RFC1-related ataxia is a mimic of early multiple system atrophy. J Neurol Neurosurg Psychiatry. 2021.
253. Sullivan R, Yau WY, Chelban V, Rossi S, O'Connor E, Wood NW, et al. RFC1 Intronic Repeat Expansions Absent in Pathologically Confirmed Multiple Systems Atrophy. Mov Disord. 2020.
254. Wiblin L, Durcan R, Galna B, Lee M, Burn D. Clinical Milestones Preceding the Diagnosis of Multiple System Atrophy and Progressive Supranuclear Palsy: A Retrospective Cohort Study. J Mov Disord. 2019.
255. Petrovic IN, Ling H, Asi Y, Ahmed Z, Kukkle PL, Hazrati LN, et al. Multiple system atrophy-parkinsonism with slow progression and prolonged survival: a diagnostic catch. Mov Disord. 2012;27(9):1186-90.
256. Wan L, Chen Z, Wan N, Liu M, Xue J, Chen H, et al. Biallelic Intronic AAGGG Expansion of RFC1 is Related to Multiple System Atrophy. Ann Neurol. 2020;88(6):1132-43.
257. Gisatulin M, Dobricic V, Zuhlke C, Hellenbroich Y, Tadic V, Munchau A, et al. Clinical spectrum of the pentanucleotide repeat expansion in the RFC1 gene in ataxia syndromes. Neurology. 2020;95(21):e2912-e23.
258. Kumar KR, Cortese A, Tomlinson SE, Efthymiou S, Ellis M, Zhu D, et al. RFC1 expansions can mimic hereditary sensory neuropathy with cough and Sjogren syndrome. Brain. 2020;143(10):e82.
259. Curro R, Salvalaggio A, Tozza S, Gemelli C, Dominik N, Galassi Deforie V, et al. RFC1 expansions are a common cause of idiopathic sensory neuropathy. Brain. 2021.
260. da Silva Schmitt G, Martinez ARM, da Graca FF, de Lima FD, Bonadia LC, Amorim BJ, et al. Dopa-Responsive Parkinsonism in a Patient With Homozygous RFC1 Expansions. Mov Disord. 2020;35(10):1889-90.
261. Shortt JA, Ruggiero RP, Cox C, Wacholder AC, Pollock DD. Finding and extending ancient simple sequence repeat-derived regions in the human genome. Mob DNA. 2020;11:11.
262. Francastel C, Magdinier F. DNA methylation in satellite repeats disorders. Essays Biochem. 2019;63(6):757-71.
263. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. Trends Genet. 2010;26(2):59-65.
264. Chintalaphani SR, Pineda SS, Deveson IW, Kumar KR. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. Acta Neuropathol Commun. 2021;9(1):98.

265. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet. 2018;19(5):286-98.
266. Ridley RM, Frith CD, Farrer LA, Conneally PM. Patterns of inheritance of the symptoms of Huntington's disease suggestive of an effect of genomic imprinting. J Med Genet. 1991;28(4):224-31.
267. La Spada AR. Trinucleotide repeat instability: genetic features and molecular mechanisms. Brain Pathol. 1997;7(3):943-63.
268. Wright GEB, Black HF, Collins JA, Gall-Duncan T, Caron NS, Pearson CE, et al. Interrupting sequence variants and age of onset in Huntington's disease: clinical implications and emerging therapies. Lancet Neurol. 2020;19(11):930-9.
269. Findlay Black H, Wright GEB, Collins JA, Caron N, Kay C, Xia Q, et al. Frequency of the loss of CAA interruption in the HTT CAG tract and implications for Huntington disease in the reduced penetrance range. Genet Med. 2020;22(12):2108-13.
270. Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari SK. CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. Hum Mol Genet. 2001;10(21):2437-46.
271. Gao R, Matsuura T, Coolbaugh M, Zuhlke C, Nakamura K, Rasmussen A, et al. Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. Eur J Hum Genet. 2008;16(2):215-22.
272. Sobczak K, Krzyzosiak WJ. CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. J Biol Chem. 2005;280(5):3898-910.
273. Xu P, Pan F, Roland C, Sagui C, Weninger K. Dynamics of strand slippage in DNA hairpins formed by CAG repeats: roles of sequence parity and trinucleotide interrupts. Nucleic Acids Res. 2020;48(5):2232-45.
274. Schule B, McFarland KN, Lee K, Tsai YC, Nguyen KD, Sun C, et al. Parkinson's disease associated with pure ATXN10 repeat expansion. NPJ Parkinsons Dis. 2017;3:27.
275. Landrian I, McFarland KN, Liu J, Mulligan CJ, Rasmussen A, Ashizawa T. Inheritance patterns of ATCCT repeat interruptions in spinocerebellar ataxia type 10 (SCA10) expansions. PLoS One. 2017;12(4):e0175958.
276. McFarland KN, Liu J, Landrian I, Zeng D, Raskin S, Moscovich M, et al. Repeat interruptions in spinocerebellar ataxia type 10 expansions are strongly associated with epileptic seizures. Neurogenetics. 2014;15(1):59-64.
277. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011;72(2):245-56.
278. Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Engelborghs S, De Bleecker J, et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry. 2016;21(8):1112-24.
279. Iacoangeli A, Al Khleifat A, Jones AR, Sproviero W, Shatunov A, Opie-Martin S, et al. C9orf72 intermediate expansions of 24-30 repeats are associated with ALS. Acta Neuropathol Commun. 2019;7(1):115.
280. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991;65(5):905-14.
281. Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology. 2001;57(1):127-30.
282. Orr HT, Chung MY, Banfi S, Kwiatkowski TJ, Jr., Servadio A, Beaudet AL, et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet. 1993;4(3):221-6.
283. Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, et al. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet. 1996;14(3):277-84.
284. Sequeiros J, Seneca S, Martindale J. Consensus and controversies in best practices for molecular genetic testing of spinocerebellar ataxias. Eur J Hum Genet. 2010;18(11):1188-95.
285. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet. 1994;8(3):221-8.

286. Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet. 1997;15(1):62-9.
287. David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat Genet. 1997;17(1):65-70.
288. Cagnoli C, Stevanin G, Michielotto C, Gerbino Promis G, Brussino A, Pappi P, et al. Large pathogenic expansions in the SCA2 and SCA7 genes can be detected by fluorescent repeat-primed polymerase chain reaction assay. J Mol Diagn. 2006;8(1):128-32.
289. Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, et al. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet. 1999;21(4):379-84.
290. Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet. 2000;26(2):191-4.
291. Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C, et al. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nat Genet. 1999;23(4):391-2.
292. Lone WG, Khan IA, Poornima S, Shaik NA, Meena AK, Prabhakar Rao K, et al. Exploration of CAG triplet repeat in nontranslated region of SCA12 gene. J Genet. 2016;95(2):427-32.
293. Maltecca F, Filla A, Castaldo I, Coppola G, Fragassi NA, Carella M, et al. Intergenerational instability and marked anticipation in SCA-17. Neurology. 2003;61(10):1441-3.
294. Nakamura K, Jeong SY, Uchihara T, Anno M, Nagashima K, Nagashima T, et al. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. Hum Mol Genet. 2001;10(14):1441-8.
295. Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, et al. Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. Am J Hum Genet. 2009;85(5):544-57.
296. Seixas AI, Loureiro JR, Costa C, Ordonez-Ugalde A, Marcelino H, Oliveira CL, et al. A Pentanucleotide ATTTC Repeat Insertion in the Non-coding Region of DAB1, Mapping to SCA37, Causes Spinocerebellar Ataxia. Am J Hum Genet. 2017;101(1):87-103.
297. Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, et al. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature. 1997;386(6627):847-51.
298. Moore RC, Xiang F, Monaghan J, Han D, Zhang Z, Edstrom L, et al. Huntington disease phenocopy is a familial prion disease. Am J Hum Genet. 2001;69(6):1385-8.
299. Holmes SE, O'Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, et al. A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. Nat Genet. 2001;29(4):377-8.
300. Ogasawara M, Iida A, Kumutpongpanich T, Ozaki A, Oya Y, Konishi H, et al. CGG expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. Acta Neuropathol Commun. 2020;8(1):204.
301. Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, et al. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. Nat Genet. 2019;51(8):1215-21.
302. Kuhlenbaumer G, Kress W, Ringelstein EB, Stogbauer F. Thirty-seven CAG repeats in the androgen receptor gene in two healthy individuals. J Neurol. 2001;248(1):23-6.
303. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991;352(6330):77-9.
304. Fratta P, Collins T, Pemble S, Nethisinghe S, Devoy A, Giunti P, et al. Sequencing analysis of the spinal bulbar muscular atrophy CAG expansion reveals absence of repeat interruptions. Neurobiol Aging. 2014;35(2):443 e1-3.
305. Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, et al. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat Genet. 2019;51(8):1222-32.
306. Xi J, Wang X, Yue D, Dou T, Wu Q, Lu J, et al. 5' UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. Brain. 2021;144(2):601-14.

307. Brais B, Bouchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, et al. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat Genet. 1998;18(2):164-7.
308. Richard P, Trollet C, Stojkovic T, de Becdelievre A, Perie S, Pouget J, et al. Correlation between PABPN1 genotype and disease severity in oculopharyngeal muscular dystrophy. Neurology. 2017;88(4):359-65.
309. Pagnamenta AT, Kaiyrzhanov R, Zou Y, Da'as SI, Maroofian R, Donkervoort S, et al. An ancestral 10-bp repeat expansion in VWA1 causes recessive hereditary motor neuropathy. Brain. 2021;144(2):584-600.
310. He F, Todd PK. Epigenetics in nucleotide repeat expansion disorders. Semin Neurol. 2011;31(5):470-83.
311. Yum K, Wang ET, Kalsotra A. Myotonic dystrophy: disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes. Curr Opin Genet Dev. 2017;44:30-7.
312. Kato M, Saitoh S, Kamei A, Shiraishi H, Ueda Y, Akasaka M, et al. A longer polyalanine expansion mutation in the ARX gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome). Am J Hum Genet. 2007;81(2):361-6.
313. Stromme P, Mangelsdorf ME, Shaw MA, Lower KM, Lewis SM, Bruyere H, et al. Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. Nat Genet. 2002;30(4):441-5.
314. Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, et al. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet. 1994;6(1):9-13.
315. Cen Z, Jiang Z, Chen Y, Zheng X, Xie F, Yang X, et al. Intronic pentanucleotide TTTCA repeat insertion in the SAMD12 gene causes familial cortical myoclonic tremor with epilepsy type 1. Brain. 2018;141(8):2280-8.
316. Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, et al. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. Nat Genet. 2018;50(4):581-90.
317. Corbett MA, Kroes T, Veneziano L, Bennett MF, Florian R, Schneider AL, et al. Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. Nat Commun. 2019;10(1):4920.
318. Florian RT, Kraft F, Leitao E, Kaya S, Klebe S, Magnin E, et al. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. Nat Commun. 2019;10(1):4919.
319. Gu Y, Shen Y, Gibbs RA, Nelson DL. Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. Nat Genet. 1996;13(1):109-13.
320. Hoijer I, Tsai YC, Clark TA, Kotturi P, Dahl N, Stattin EL, et al. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. Hum Mutat. 2018;39(9):1262-72.
321. Kebschull JM, Zador AM. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Res. 2015;43(21):e143.
322. Halman A, Oshlack A. Accuracy of short tandem repeats genotyping tools in whole exome sequencing data. F1000Res. 2020;9:200.
323. Kumar KR, Cowley MJ, Davis RL. Next-Generation Sequencing and Emerging Technologies. Semin Thromb Hemost. 2019;45(7):661-73.
324. Mousavi N, Shleizer-Burko S, Yanicky R, Gymrek M. Profiling the genome-wide landscape of tandem repeat expansions. Nucleic Acids Res. 2019;47(15):e90.
325. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet. 2016;135(3):359-62.
326. Russ J, Liu EY, Wu K, Neal D, Suh E, Irwin DJ, et al. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. 2015;129(1):39-52.
327. Ebbert MTW, Farrugia SL, Sens JP, Jansen-West K, Gendron TF, Prudencio M, et al. Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease. Mol Neurodegener. 2018;13(1):46.

328. Giesselmann P, Brandl B, Raimondeau E, Bowen R, Rohrandt C, Tandon R, et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. Nat Biotechnol. 2019;37(12):1478-81.
329. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13(5):278-89.
330. Ibanez K, Polke J, Hagelstrom T, Dolzhenko E, Pasko D, Thomas E, et al. Whole genome sequencing for diagnosis of neurological repeat expansion disorders. bioRxiv. 2020.
331. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ. 2018;361:k1687.
332. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet. 2017;100(5):695-705.
333. Wheway G, Mitchison HM, Genomics England Research C. Opportunities and Challenges for Molecular Understanding of Ciliopathies-The 100,000 Genomes Project. Front Genet. 2019;10:127.
334. Kay C, Collins JA, Miedzybrodzka Z, Madore SJ, Gordon ES, Gerry N, et al. Huntington disease reduced penetrance alleles occur at high frequency in the general population. Neurology. 2016;87(3):282-8.
335. Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, Veitch NJ, et al. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. Hum Mol Genet. 2007;16(10):1133-42.
336. Labadorf A, Hoss AG, Lagomarsino V, Latourelle JC, Hadzi TC, Bregu J, et al. RNA Sequence Analysis of Human Huntington Disease Brain Reveals an Extensive Increase in Inflammatory and Developmental Gene Expression. PLoS One. 2015;10(12):e0143563.
337. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol. 2016;79(6):983-90.
338. Lee JM, Chao MJ, Harold D, Abu Elneel K, Gillis T, Holmans P, et al. A modifier of Huntington's disease onset at the MLH1 locus. Hum Mol Genet. 2017;26(19):3859-67.
339. Moss DJH, Pardinas AF, Langbehn D, Lo K, Leavitt BR, Roos R, et al. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol. 2017;16(9):701-11.
340. Tome S, Manley K, Simard JP, Clark GW, Slean MM, Swami M, et al. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet. 2013;9(2):e1003280.
341. Kraus-Perrotta C, Lagalwar S. Expansion, mosaicism and interruption: mechanisms of the CAG repeat mutation in spinocerebellar ataxia type 1. Cerebellum Ataxias. 2016;3:20.
342. Mouro Pinto R, Arning L, Giordano JV, Razghandi P, Andrew MA, Gillis T, et al. Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. Hum Mol Genet. 2020;29(15):2551-67.
343. Buchman VL, Cooper-Knock J, Connor-Robson N, Higginbottom A, Kirby J, Razinskaya OD, et al. Simultaneous and independent detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. Mol Neurodegener. 2013;8:12.
344. Dols-Icardo O, Garcia-Redondo A, Rojas-Garcia R, Sanchez-Valle R, Noguera A, Gomez-Tortosa E, et al. Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. Hum Mol Genet. 2014;23(3):749-54.
345. Beck JA, Poulter M, Campbell TA, Uphill JB, Adamson G, Geddes JF, et al. Somatic and germline mosaicism in sporadic early-onset Alzheimer's disease. Hum Mol Genet. 2004;13(12):1219-24.
346. Blencowe H, Moorthie S, Petrou M, Hamamy H, Povey S, Bittles A, et al. Rare single gene disorders: estimating baseline prevalence and outcomes worldwide. J Community Genet. 2018;9(4):397-406.
347. Behan L, Dimitrov BD, Kuehni CE, Hogg C, Carroll M, Evans HJ, et al. PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia. Eur Respir J. 2016;47(4):1103-12.

348. Kassmann CM, Lappe-Siefke C, Baes M, Brugger B, Mildner A, Werner HB, et al. Axonal loss and neuroinflammation caused by peroxisome-deficient oligodendrocytes. Nat Genet. 2007;39(8):969-76.
349. Kassmann CM, Quintes S, Rietdorf J, Mobius W, Sereda MW, Nientiedt T, et al. A role for myelin-associated peroxisomes in maintaining paranodal loops and axonal integrity. FEBS Lett. 2011;585(14):2205-11.
350. Holtzman E, Teichberg S, Abrahams SJ, Citkowitz E, Crain SM, Kawai N, et al. Notes on synaptic vesicles and related structures, endoplasmic reticulum, lysosomes and peroxisomes in nervous tissue and the adrenal medulla. J Histochem Cytochem. 1973;21(4):349-85.
351. Bowen P, Lee CS, Zellweger H, Lindenberg R. A Familial Syndrome of Multiple Congenital Defects. Bull Johns Hopkins Hosp. 1964;114:402-14.
352. Crane DI. Revisiting the neuropathogenesis of Zellweger syndrome. Neurochem Int. 2014;69:1-8.
353. Shinde AB, Baboota RK, Denis S, Loizides-Mangold U, Peeters A, Espeel M, et al. Mitochondrial disruption in peroxisome deficient cells is hepatocyte selective but is not mediated by common hepatic peroxisomal metabolites. Mitochondrion. 2018;39:51-9.
354. Argyriou C, Polosa A, Cecyre B, Hsieh M, Di Pietro E, Cui W, et al. A longitudinal study of retinopathy in the PEX1-Gly844Asp mouse model for mild Zellweger Spectrum Disorder. Exp Eye Res. 2019;186:107713.
355. Baumgart E, Vanhorebeek I, Grabenbauer M, Borgers M, Declercq PE, Fahimi HD, et al. Mitochondrial alterations caused by defective peroxisomal biogenesis in a mouse model for Zellweger syndrome (PEX5 knockout mouse). Am J Pathol. 2001;159(4):1477-94.
356. Muller CC, Nguyen TH, Ahlemeyer B, Meshram M, Santrampurwala N, Cao S, et al. PEX13 deficiency in mouse brain as a model of Zellweger syndrome: abnormal cerebellum formation, reactive gliosis and oxidative stress. Dis Model Mech. 2011;4(1):104-19.
357. Nuebel E, Morgan JT, Fogarty S, Winter JM, Lettlova S, Berg JA, et al. The biochemical basis of mitochondrial dysfunction in Zellweger Spectrum Disorder. EMBO Rep. 2021;22(10):e51991.
358. Aubourg P. [X-linked adrenoleukodystrophy]. Ann Endocrinol (Paris). 2007;68(6):403-11.
359. Meinecke M, Cizmowski C, Schliebs W, Kruger V, Beck S, Wagner R, et al. The peroxisomal importomer constitutes a large and highly dynamic pore. Nat Cell Biol. 2010;12(3):273-7.
360. Braverman NE, D'Agostino MD, Maclean GE. Peroxisome biogenesis disorders: Biological, clinical and pathophysiological perspectives. Dev Disabil Res Rev. 2013;17(3):187-96.
361. Platta HW, Debelyy MO, El Magraoui F, Erdmann R. The AAA peroxins Pex1p and Pex6p function as dislocases for the ubiquitinated peroxisomal import receptor Pex5p. Biochem Soc Trans. 2008;36(Pt 1):99-104.
362. Hanson PI, Whiteheart SW. AAA+ proteins: have engine, will work. Nat Rev Mol Cell Biol. 2005;6(7):519-29.
363. Vale RD. AAA proteins. Lords of the ring. J Cell Biol. 2000;150(1):F13-9.
364. Miyata N, Fujiki Y. Shuttling mechanism of peroxisome targeting signal type 1 receptor Pex5: ATP-independent import and ATP-dependent export. Mol Cell Biol. 2005;25(24):10822-32.
365. Opalinski L, Veenhuis M, van der Klei IJ. Peroxisomes: membrane events accompanying peroxisome proliferation. Int J Biochem Cell Biol. 2011;43(6):847-51.
366. Huybrechts SJ, Van Veldhoven PP, Brees C, Mannaerts GP, Los GV, Fransen M. Peroxisome dynamics in cultured mammalian cells. Traffic. 2009;10(11):1722-33.
367. Wanders RJ, Waterham HR. Biochemistry of mammalian peroxisomes revisited. Annu Rev Biochem. 2006;75:295-332.
368. Wanders RJ, Poll-The BT. "Role of peroxisomes in human lipid metabolism and its importance for neurological development". Neurosci Lett. 2017;637:11-7.
369. Poll-The BT, Roels F, Ogier H, Scotto J, Vamecq J, Schutgens RB, et al. A new peroxisomal disorder with enlarged peroxisomes and a specific deficiency of acyl-CoA oxidase (pseudo-neonatal adrenoleukodystrophy). Am J Hum Genet. 1988;42(3):422-34.

370. Ferdinandusse S, Denis S, Mooyer PA, Dekker C, Duran M, Soorani-Lunsing RJ, et al. Clinical and biochemical spectrum of D-bifunctional protein deficiency. Ann Neurol. 2006;59(1):92-104.
371. Huyghe S, Mannaerts GP, Baes M, Van Veldhoven PP. Peroxisomal multifunctional protein-2: the enzyme, the patients and the knockout mouse model. Biochim Biophys Acta. 2006;1761(9):973-94.
372. Van Veldhoven PP. Biochemistry and genetics of inherited disorders of peroxisomal fatty acid metabolism. J Lipid Res. 2010;51(10):2863-95.
373. Poulos A, Sharp P, Fellenberg AJ, Danks DM. Cerebro-hepato-renal (Zellweger) syndrome, adrenoleukodystrophy, and Refsum's disease: plasma changes and skin fibroblast phytanic acid oxidase. Hum Genet. 1985;70(2):172-7.
374. Biermann J, Just WW, Wanders RJ, Van Den Bosch H. Alkyl-dihydroxyacetone phosphate synthase and dihydroxyacetone phosphate acyltransferase form a protein complex in peroxisomes. Eur J Biochem. 1999;261(2):492-9.
375. da Silva TF, Sousa VF, Malheiro AR, Brites P. The importance of ether-phospholipids: a view from the perspective of mouse models. Biochim Biophys Acta. 2012;1822(9):1501-8.
376. Chu BB, Liao YC, Qi W, Xie C, Du X, Wang J, et al. Cholesterol transport through lysosome-peroxisome membrane contacts. Cell. 2015;161(2):291-306.
377. Gao Q, Goodman JM. The lipid droplet-a well-connected organelle. Front Cell Dev Biol. 2015;3:49.
378. Schrader M, Godinho LF, Costello JL, Islinger M. The different facets of organelle interplay-an overview of organelle interactions. Front Cell Dev Biol. 2015;3:56.
379. Hettema EH, Erdmann R, van der Klei I, Veenhuis M. Evolving models for peroxisome biogenesis. Curr Opin Cell Biol. 2014;29:25-30.
380. Geuze HJ, Murk JL, Stroobants AK, Griffith JM, Kleijmeer MJ, Koster AJ, et al. Involvement of the endoplasmic reticulum in peroxisome formation. Mol Biol Cell. 2003;14(7):2900-7.
381. Zaar K, Hartig F, Fahimi HD, Gorgas K. Peroxisomal aggregates forming large stacks in the lipid segment of the canine kidney. Acta Histochem Suppl. 1984;29:165-8.
382. Knoblach B, Sun X, Coquelle N, Fagarasanu A, Poirier RL, Rachubinski RA. An ER-peroxisome tether exerts peroxisome population control in yeast. EMBO J. 2013;32(18):2439-53.
383. Dowling DK, Simmons LW. Reactive oxygen species as universal constraints in life-history evolution. Proc Biol Sci. 2009;276(1663):1737-45.
384. Scherz-Shouval R, Elazar Z. Regulation of autophagy by ROS: physiology and pathology. Trends Biochem Sci. 2011;36(1):30-8.
385. Fransen M, Nordgren M, Wang B, Apanasets O. Role of peroxisomes in ROS/RNS-metabolism: implications for human disease. Biochim Biophys Acta. 2012;1822(9):1363-73.
386. Lopez-Erauskin J, Fourcade S, Galino J, Ruiz M, Schluter A, Naudi A, et al. Antioxidants halt axonal degeneration in a mouse model of X-adrenoleukodystrophy. Ann Neurol. 2011;70(1):84-92.
387. Legakis JE, Koepke JI, Jedeszko C, Barlaskar F, Terlecky LJ, Edwards HJ, et al. Peroxisome senescence in human fibroblasts. Mol Biol Cell. 2002;13(12):4243-55.
388. Antonenkov VD, Grunau S, Ohlmeier S, Hiltunen JK. Peroxisomes are oxidative organelles. Antioxid Redox Signal. 2010;13(4):525-37.
389. Ivashchenko O, Van Veldhoven PP, Brees C, Ho YS, Terlecky SR, Fransen M. Intraperoxisomal redox balance in mammalian cells: oxidative stress and interorganellar cross-talk. Mol Biol Cell. 2011;22(9):1440-51.
390. Hwang I, Lee J, Huh JY, Park J, Lee HB, Ho YS, et al. Catalase deficiency accelerates diabetic renal injury through peroxisomal dysfunction. Diabetes. 2012;61(3):728-38.
391. Honsho M, Yamashita S, Fujiki Y. Peroxisome homeostasis: Mechanisms of division and selective degradation of peroxisomes in mammals. Biochim Biophys Acta. 2016;1863(5):984-91.
392. Islinger M, Luers GH, Li KW, Loos M, Volkl A. Rat liver peroxisomes after fibrate treatment. A survey using quantitative mass spectrometry. J Biol Chem. 2007;282(32):23055-69.
393. Wiese S, Gronemeyer T, Ofman R, Kunze M, Grou CP, Almeida JA, et al. Proteomics characterization of mouse kidney peroxisomes by tandem mass spectrometry and protein correlation profiling. Mol Cell Proteomics. 2007;6(12):2045-57.

394. Abu-Safieh L, Alrashed M, Anazi S, Alkuraya H, Khan AO, Al-Owain M, et al. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. Genome Res. 2013;23(2):236-47.
395. Costello JL, Castro IG, Hacker C, Schrader TA, Metz J, Zeuschner D, et al. ACBD5 and VAPB mediate membrane associations between peroxisomes and the ER. J Cell Biol. 2017;216(2):331-42.
396. Lykke-Andersen S, Jensen TH. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol. 2015;16(11):665-77.
397. Lynch DS, Chelban V, Vandrovcova J, Pittman A, Wood NW, Houlden H. GLS loss of function causes autosomal recessive spastic ataxia and optic atrophy. Ann Clin Transl Neurol. 2018;5(2):216-21.
398. Iqbal Z, Tawamie H, Ba W, Reis A, Halak BA, Sticht H, et al. Loss of function of SVBP leads to autosomal recessive intellectual disability, microcephaly, ataxia, and hypotonia. Genet Med. 2019.
399. Mendoza-Ferreira N, Coutelier M, Janzen E, Hosseinibarkooie S, Lohr H, Schneider S, et al. Biallelic CHP1 mutation causes human autosomal recessive ataxia by impairing NHE1 function. Neurol Genet. 2018;4(1):e209.
400. Synofzik M, Smets K, Mallaret M, Di Bella D, Gallenmuller C, Baets J, et al. SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: a large multi-centre study. Brain. 2016;139(Pt 5):1378-93.
401. Wanders RJ. Metabolic functions of peroxisomes in health and disease. Biochimie. 2014;98:36-44.
402. Waterham HR, Ferdinandusse S, Wanders RJ. Human disorders of peroxisome metabolism and biogenesis. Biochim Biophys Acta. 2016;1863(5):922-33.
403. Steinberg S, Jones R, Tiffany C, Moser A. Investigational methods for peroxisomal disorders. Curr Protoc Hum Genet. 2008;Chapter 17:Unit 17 6.
404. Ermak G, Davies KJ. Calcium and oxidative stress: from cell signaling to cell death. Mol Immunol. 2002;38(10):713-21.
405. Schrader M, Fahimi HD. Peroxisomes and oxidative stress. Biochim Biophys Acta. 2006;1763(12):1755-66.
406. Liguori I, Russo G, Curcio F, Bulli G, Aran L, Della-Morte D, et al. Oxidative stress, aging, and diseases. Clin Interv Aging. 2018;13:757-72.
407. De Duve C, Baudhuin P. Peroxisomes (microbodies and related particles). Physiol Rev. 1966;46(2):323-57.
408. Zwacka RM, Reuter A, Pfaff E, Moll J, Gorgas K, Karasawa M, et al. The glomerulosclerosis gene Mpv17 encodes a peroxisomal protein producing reactive oxygen species. EMBO J. 1994;13(21):5129-34.
409. Stolz DB, Zamora R, Vodovotz Y, Loughran PA, Billiar TR, Kim YM, et al. Peroxisomal localization of inducible nitric oxide synthase in hepatocytes. Hepatology. 2002;36(1):81-93.
410. Moldovan L, Moldovan NI. Oxygen free radicals and redox biology of organelles. Histochem Cell Biol. 2004;122(4):395-412.
411. Sharifi-Rad M, Anil Kumar NV, Zucca P, Varoni EM, Dini L, Panzarini E, et al. Lifestyle, Oxidative Stress, and Antioxidants: Back and Forth in the Pathophysiology of Chronic Diseases. Front Physiol. 2020;11:694.
412. Ayala A, Munoz MF, Arguelles S. Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. Oxid Med Cell Longev. 2014;2014:360438.
413. Gandhi S, Abramov AY. Mechanism of oxidative stress in neurodegeneration. Oxid Med Cell Longev. 2012;2012:428010.
414. Vargas CR, Wajner M, Sirtori LR, Goulart L, Chiochetta M, Coelho D, et al. Evidence that oxidative stress is increased in patients with X-linked adrenoleukodystrophy. Biochim Biophys Acta. 2004;1688(1):26-32.
415. Valavanidis A, Vlachogianni T, Fiotakis K. Tobacco smoke: involvement of reactive oxygen species and stable free radicals in mechanisms of oxidative damage, carcinogenesis and synergistic effects with other respirable particles. Int J Environ Res Public Health. 2009;6(2):445-62.

416. Lazzarino G, Listorti I, Bilotta G, Capozzolo T, Amorini AM, Longo S, et al. Water- and Fat-Soluble Antioxidants in Human Seminal Plasma and Serum of Fertile Males. Antioxidants (Basel). 2019;8(4).
417. Hua R, Cheng D, Coyaud E, Freeman S, Di Pietro E, Wang Y, et al. VAPs and ACBD5 tether peroxisomes to the ER for peroxisome maintenance and lipid homeostasis. J Cell Biol. 2017;216(2):367-77.
418. Braverman NE, Moser AB. Functions of plasmalogen lipids in health and disease. Biochim Biophys Acta. 2012;1822(9):1442-52.
419. Kovacs WJ, Tape KN, Shackelford JE, Duan X, Kasumov T, Kelleher JK, et al. Localization of the pre-squalene segment of the isoprenoid biosynthetic pathway in mammalian peroxisomes. Histochem Cell Biol. 2007;127(3):273-90.
420. Lev S, Ben Halevy D, Peretti D, Dahan N. The VAP protein family: from cellular functions to motor neuron disease. Trends Cell Biol. 2008;18(6):282-90.
421. Murphy SE, Levine TP. VAP, a Versatile Access Point for the Endoplasmic Reticulum: Review and analysis of FFAT-like motifs in the VAPome. Biochim Biophys Acta. 2016;1861(8 Pt B):952-61.
422. Kors S, Hacker C, Bolton C, Maier R, Reimann L, Kitchener EJA, et al. Regulating peroxisome-ER contacts via the ACBD5-VAPB tether by FFAT motif phosphorylation and GSK3beta. J Cell Biol. 2022;221(3).
423. Johri A, Beal MF. Mitochondrial dysfunction in neurodegenerative diseases. J Pharmacol Exp Ther. 2012;342(3):619-30.
424. Salpietro V, Phadke R, Saggar A, Hargreaves IP, Yates R, Fokoloros C, et al. Zellweger syndrome and secondary mitochondrial myopathy. Eur J Pediatr. 2015;174(4):557-63.
425. van Roermund CW, Elgersma Y, Singh N, Wanders RJ, Tabak HF. The membrane of peroxisomes in Saccharomyces cerevisiae is impermeable to NAD(H) and acetyl-CoA under in vivo conditions. EMBO J. 1995;14(14):3480-6.
426. Kamencic H, Lyon A, Paterson PG, Juurlink BH. Monochlorobimane fluorometric method to measure tissue glutathione. Anal Biochem. 2000;286(1):35-7.
427. Zitka O, Skalickova S, Gumulec J, Masarik M, Adam V, Hubalek J, et al. Redox status expressed as GSH:GSSG ratio as a marker for oxidative stress in paediatric tumour patients. Oncol Lett. 2012;4(6):1247-53.
428. Petrillo S, Piemonte F, Pastore A, Tozzi G, Aiello C, Pujol A, et al. Glutathione imbalance in patients with X-linked adrenoleukodystrophy. Mol Genet Metab. 2013;109(4):366-70.
429. Peoples JN, Saraf A, Ghazal N, Pham TT, Kwong JQ. Mitochondrial dysfunction and oxidative stress in heart disease. Exp Mol Med. 2019;51(12):1-13.
430. Haim M. Epidemiology of retinitis pigmentosa in Denmark. Acta Ophthalmol Scand Suppl. 2002(233):1-34.
431. Nash BM, Wright DC, Grigg JR, Bennetts B, Jamieson RV. Retinal dystrophies, genomic applications in diagnosis and prospects for therapy. Transl Pediatr. 2015;4(2):139-63.
432. Dyle MC, Kolakada D, Cortazar MA, Jagannathan S. How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. Wiley Interdiscip Rev RNA. 2020;11(1):e1560.
433. Nomakuchi TT, Rigo F, Aznarez I, Krainer AR. Antisense oligonucleotide-directed inhibition of nonsense-mediated mRNA decay. Nat Biotechnol. 2016;34(2):164-6.
434. Jagannathan S, Bradley RK. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. Genome Res. 2016;26(12):1639-50.
435. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camano P, Dauwerse JG, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 2010;329(5999):1650-3.
436. Rickard AM, Petek LM, Miller DG. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Hum Mol Genet. 2015;24(20):5901-14.
437. Sarkar H, Mitsios A, Smart M, Skinner J, Welch AA, Kalatzis V, et al. Nonsense-mediated mRNA decay efficiency varies in choroideremia providing a target to boost small molecule therapeutics. Hum Mol Genet. 2019;28(11):1865-71.

438. Lilja T, Heldring N, Hermanson O. Like a rolling histone: epigenetic regulation of neural stem cells and brain development by factors controlling histone acetylation and methylation. Biochim Biophys Acta. 2013;1830(2):2354-60.
439. Gallagher D, Voronova A, Zander MA, Cancino GI, Bramall A, Krause MP, et al. Ankrd11 is a chromatin regulator involved in autism that is essential for neural development. Dev Cell. 2015;32(1):31-42.
440. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics. 2014;13(2):397-406.
441. Ka M, Kim WY. ANKRD11 associated with intellectual disability and autism regulates dendrite differentiation via the BDNF/TrkB signaling pathway. Neurobiol Dis. 2018;111:138-52.
442. Zhao M, Maani N, Dowling JJ. Dynamin 2 (DNM2) as Cause of, and Modifier for, Human Neuromuscular Disease. Neurotherapeutics. 2018;15(4):966-75.
443. Mettlen M, Chen PH, Srinivasan S, Danuser G, Schmid SL. Regulation of Clathrin-Mediated Endocytosis. Annu Rev Biochem. 2018;87:871-96.
444. Durieux AC, Prudhon B, Guicheney P, Bitoun M. Dynamin 2 and human diseases. J Mol Med (Berl). 2010;88(4):339-50.
445. Antonny B, Burd C, De Camilli P, Chen E, Daumke O, Faelber K, et al. Membrane fission by dynamin: what we know and what we need to know. EMBO J. 2016;35(21):2270-84.
446. Sekiba K, Yamagami M, Otsuka M, Suzuki T, Kishikawa T, Ishibashi R, et al. Transcriptional activation of the MICA gene with an engineered CRISPR-Cas9 system. Biochem Biophys Res Commun. 2017;486(2):521-5.
447. Tumer G, Simpson B, Roberts TK. Genetics, Human Major Histocompatibility Complex (MHC). StatPearls. Treasure Island (FL)2021.
448. Goverdhan SV, Khakoo SI, Gaston H, Chen X, Lotery AJ. Age-related macular degeneration is associated with the HLA-Cw*0701 Genotype and the natural killer cell receptor AA haplotype. Invest Ophthalmol Vis Sci. 2008;49(11):5077-82.
449. Khakoo SI, Carrington M. KIR and disease: a model system or system of models? Immunol Rev. 2006;214:186-201.
450. Carrington M, Martin MP. The impact of variation at the KIR gene cluster on human disease. Curr Top Microbiol Immunol. 2006;298:225-57.

8. Appendix

8.1. RFC1 Caller script

8.1.1. Functions created to bind RFC1 Caller outputs

```
extractSummary <- function(f) -{
  j <- fromJSON(f) #loads new file
  dat <- NULL #creates new clean dataframe called dat
  out <- as.data.frame(cbind(j$rfc1_status, j$GT)) #extracting the rfc1 status, saves in a new
dataframe called 'out'
  dat <- plyr::rbind.fill(dat, out) #adding the new sample to the empty database
  colnames(dat) <- c("rfc1_status") #names the column 'rfc1 status'
  dat$SAMPLE <- gsub(".*/","",f) #removes everything before last '/' and stores the remaining text in
a new sample column
  return(dat)
}

processEH <- function(i) {
  files <- list.files(path = i, pattern = ".json", full.names = TRUE, recursive = TRUE)
  d <- lapply(files, extractSummary)
  d <- do.call(rbind, d)


# loop through the files

for (i in 1:length(names)){
  setwd("Filepath")
  dat <- processEH(paste0("Z:",names[i]))
  setwd("Filepath")
  saveRDS(dat, paste0("Z:", names[i],".rds"))
}
f <- list.files(pattern = ".rds", path = "", full.names = T, recursive = F)
d <- lapply(f, readRDS)
d <- do.call(rbind, d)
```

8.1.2. Generating *RFC1* pile-up plots

Pile-up plots generated from Graph Alignment:

python3 ../caller/caller.py --sample HG00982 --bam data/HG00982.eh_realigned.bam --json data/HG00982.eh.json --out-dir HG00982_output

Data was created, including file pathways.

8.2. Sleuth script used in RNA sequencing analysis

```
suppressMessages({
  library("sleuth")
})
sample_id <- dir(file.path("Z:/Roisin"))
sample_id = sample_id[-31]
sample_id

kal_dirs <- file.path("Z:/Roisin", sample_id)
kal_dirs
kal_dirs <- as.data.frame(kal_dirs)
kal_dirs <- as.vector(kal_dirs[order(kal_dirs$kal_dirs),])

s2c <- read.table("",header=T,sep="\t",stringsAsFactors=FALSE)
s2c <- as.data.frame(s2c [order(s2c$sample),])
s2c <- dplyr::mutate(s2c, path = kal_dirs)
print(s2c)

s2c$condition <- as.factor(s2c$condition)
so <- sleuth_prep(s2c, extra_bootstrap_summary = TRUE)
sleuth_save(so, "so_all")

so <- sleuth_fit(so, ~condition, 'full') #fitting the error model

so <- sleuth_fit(so, ~1, 'reduced') #fitting reduced model

so <- sleuth_lrt(so, 'reduced', 'full')
sleuth_save(so, "so_model2")

models(so)

sleuth_table <- sleuth_results(so, 'reduced:full', 'lrt', show_all = T)
sleuth_significant <- dplyr::filter(sleuth_table, qval <= 0.1)


extract_model_1 <- extract_model(so, "full")
```

8.2.1.1. Extracting TPM

```
sleuth_matrix <- sleuth_to_matrix(so, 'obs_norm', 'tpm')
head(sleuth_matrix,2)


# subset matrix for genes of interest
genes <- sleuth_significant$target_id


# subset only IDs which are significant:
sleuth_mat <- sleuth_matrix[c(genes), ]
write.csv(sleuth_mat,file=" ")


# transpose matrix for boxplot
```

```
sleuth_mat_t <- as.data.frame(t(sleuth_mat))


# merge the ACBD5 status onto the matrix
library(readr)
R <- read_delim("",
    delim = "\t", escape_double = FALSE,
    trim_ws = TRUE)

# make the sample column
sleuth_mat_t$sample <- rownames(sleuth_mat_t)


# merge
sleuth_mat_t <- merge(sleuth_mat_t,RNA_info_final,by.x="sample",by.y="sample",all.x=TRUE)
```

8.3. RNA sequencing data

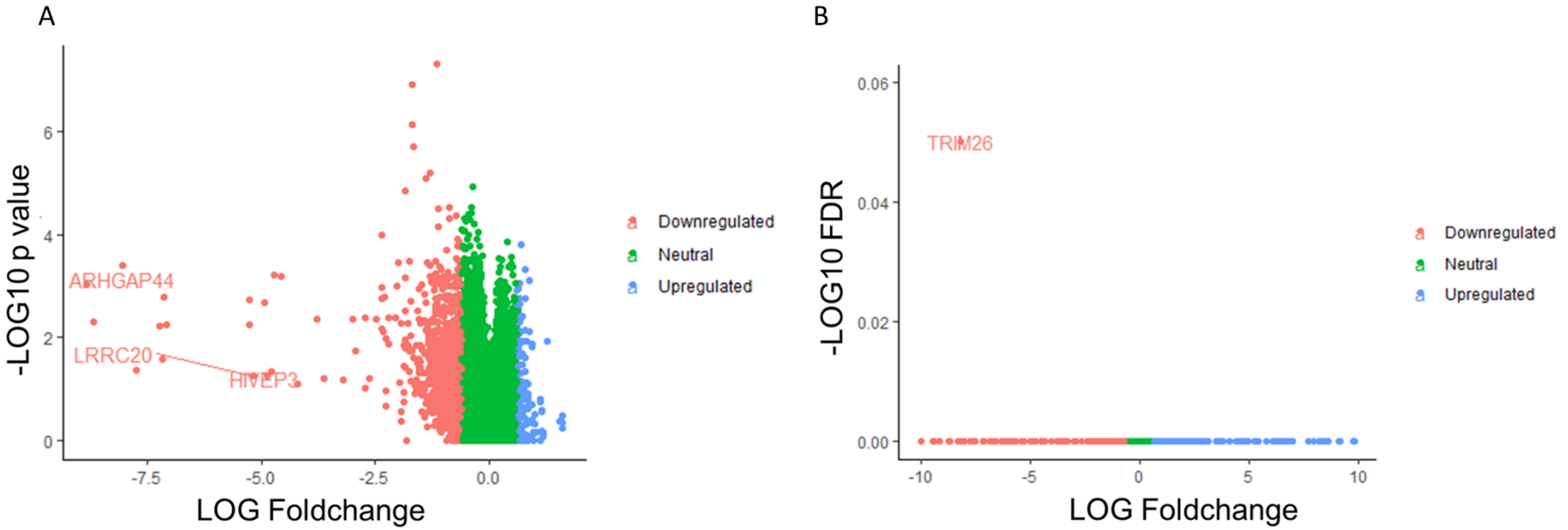


Figure 8.3.1: Transcriptomic summary. A, Volcano plot showing differentially expressed genes in patients compared to controls. Y-axis shows LOG foldchange, and X-axis shows -LOG10 of p value. B, Volcano plot showing differentially expressed genes in patients without retinal dystrophy, compared to those without. Y-axis shows Log foldchange, and X-axis shows -LOG10 of the corrected p value (FDR). Colours donate direction of gene expression change: upregulation, downregulation or neutral. *FDR, false discovery rate*

Gene Symbol	Rank
ACBD5	-6.26469
ANKRD11	-8.54424
DNM2	-9.78662
DROSHA	-7.47002
PDCD4	-1.52932
SRSF5	-2.68604
ITPR3	-3.09449
MRPL35	-2.10644
PHB2_ENST00000440277	-3.51735
GDAP1	-1.50498
SAAL1	-1.46451
REV3L	-3.33774
NDUFB3	-1.92672
TMEM132A	-2.57985
TBX15	-1.4201
EIF4G3	-1.99679
RPS8	-1.23554
MATN2	-4.94469
ZMYM2	-2.75704
LRP4	-1.89981
REPS2	-2.50813
EHBP1	-1.1018
ADA2	-3.94227
RBIS	-2.00432
IGDCC4	-9.68047
LAMA4	-1.91314
DUXAP9	-1.42506
SGCE	-3.08256
GPX1	-2.32151
PRPH2	-2.77564
HIVEP1	1.361117
MGAT5	2.38103
SEMA3F	-0.90149
NA	-2.72339
RARG	-1.0221
GLT8D2	-2.26809
CAD	-1.67734
UTP14C	-2.60511
SEM1	-2.23906
FCGRT	-2.17438
DDX60	-2.36017
UBE3A	-1.2731
GABBR1	-0.79072
OXCT2P1	-1.34764
MICAL2	1.700671

NACA	-0.75688
DDX3X	-2.87026
EIF3F	-1.80198
ST7L	-1.39349
SEC24A	0.671899
POPDC3	-0.89519
SSR3	0.973049
TFDP1	-1.5784
LRBA	-2.69406
PPP1R11	-6.41275
KIZ	-1.48125
ABCA6	-4.36389
PCM1	-1.41778
EIF4EBP3	-1.07278
HDAC6	-5.03777
RPL13	-0.77523
PPP5C	-1.96021
ASB13	-6.3502
ECD	-1.44204
HS3ST3B1	-1.56197
CDO1	-1.67084
UBR4	-27.6916
SLC2A4RG	-0.61481
B3GNT5	-2.0213
MARCHF2	-1.10316
OR52N5	1.366991
LAMB1	-3.95535
SLC35E1	0.741604
RPL37A	-0.58797
KRBOX1	-2.12431
EBAG9	-1.17577
ACOX3	-3.80264
EFNB3	-1.53204
SLC2A1	2.480608
ABCC10	-2.6002
EIF4A2	-1.35703
FAU	-1.64433
EBF1	-1.93046
PPARG	-16.4457
TIMP4	-2.06017
FAM156B	-1.11014
LPAR4	-6.48088
NOL8	-3.26296
S100A10	-0.83315
AQR	-1.54425
PTBP2	-1.84309
PHTF2	0.684599
ALKBH4	-0.95198
FYN	-0.80507

HECTD1	-0.62569
PARK7	-0.61748
PHLDA2	-1.54524
CDH4	-1.86158
SERTAD4	1.461264
PDE7A	-0.75962
CHD1L	-11.9336
TATDN3	-1.93055
FKBP9	-2.67635
RPL27A	-0.70899
SAP130	-9.89946
PRPF39	-0.97577
PGF	-2.42012
TSC1	-4.3786
MLLT10	-2.4337
HLA-A	-3.36356
EPHB2	-0.87671
PSMD2	-15.8301
CAPS	-6.09236
TKT	-0.84432
ABCB4-202	-1.74044
HMGB2	-0.80351
ESPNL	-0.79566
SMDT1	-1.11314
SERINC3	-3.51979
EFHB	-2.64894
DST	-2.11164
BROX	-1.29484
ATF4	-1.00721
ROBO3	-1.31379
PXK	-0.66886
ANPEP	-1.81642
MCM3AP	0.710785
EIF3E	-1.60014
TLR4	-1.563
CCNL1	-2.97904
MDK	-1.47623
BAALC	-0.84319
RPL11	-0.64289
ZC3HAV1L	0.904239
ZFAND5	-1.23912
AOX1	-4.1691
NFIA	-1.1604
TOP2B	-0.8262
CEP170	-2.97993
GBF1	0.511695
ST3GAL5	-1.15869
CEP41	-12.2047
TMEM154	-0.90265

SLC4A2	0.510498
RHOJ	2.061566
RUNX1	-2.40577
RPL14	-0.6584
MTARC1	-1.87955
ARHGAP44	-2.87372
ZNF154	0.798901
DHX57	-3.86113
UBXN1	-0.77213
NUDT10	-2.28636
IGFBP6	-1.04041
ARMCX5	-3.56553
COPA	-3.68476
SPTBN1	-1.78986
TMEM87B	-3.19159
SPTBN5	-1.54767
ERCC6	1.051114
BCL11A	-1.02173
MCM6	-2.30428
SLTM	-2.88705
SNRPA1	-1.1922
MED24	-6.26185
ZKSCAN1	-7.32355
RPL10	-1.04073
WDFY3	0.618245
PIBF1	-3.20014
EHMT2	-1.60912
EIF4A1	-2.38993
DAAM2	1.241475
SMG9	1.960501
KLHL12	-1.49674
EPHA4	-3.88428
DCLK1	-2.43463
GTF2IP1	-0.89357
MED17	-1.09544
CYBA	-0.83191
CSF2RB	1.623226
TCAF1	0.927161
SLC2A3	0.509321
PPIG	-1.46238
PACSIN2	2.027178
SLC25A1	-1.39301
RPL4	-1.7336
DAB2IP	0.57253
DMXL1	0.602172
TNC	1.019501
CNBP	-0.94259
MRPL32	-0.83021
NDN	-0.52546

MRRF	-1.04124
MPRIP	0.846762
ATP5MJ	-1.40871
CDCA2	-0.6982
CCNL2	-0.6932
DCXR	-0.92405
CIRBP	-0.44012
ZFAND2A	-1.70988

Table 8.3.1: Top 200 ranked genes included in Gene Set Enrichment using for ACBD5-deficient patients and controls. Genes uploaded to WebGestalt for analysis.

Gene Symbol	Rank
ZBTB22	-27.9264
HLA-A	-25.1045
MICA	-23.8201
TRIM26	-34.1266
SPECC1	-2.7769
CBSL	-34.8789
VARS2	-15.3101
SH3YL1	16.11695
PI16	-6.60987
UBE3A	32.08864
CIRBP	-28.5984
CTNNB1	-18.3732
TMEM164	-27.7624
HAPLN1	-16.1478
TUBB	-13.1617
BANP	13.92965
STING1	-25.5322
U2AF1	-16.4473
ST7L	13.34384
DDR1	-21.6069
MARCHF7	14.93431
POU2F1	-11.7809
BRD2	7.410499
RGS12	-32.3946
SH3BGR	-5.91169
DPP8	12.68407
LIMS1	-23.8968
ILF3	12.89867
ZC3H14	12.73456
CCN4	-1.25001
RPL3P2	-5.09518
PUF60	-6.71566
ACTB-213	1.907562
PGAP2	12.16635
KRBOX1	-13.2319
RPP21	13.6573

CAPRIN2	-13.2776
RHOG	-12.407
ENSA	-12.7698
RIOK3	-12.5519
PPWD1	-12.4537
EMP1	-9.40103
NLRP1	10.4224
MBP	10.1063
HLA-C	-11.4782
NEK6	-12.6377
CENPT	19.08261
AFF2	-3.04101
NT5E	-12.7782
MOK	-11.3085
ARHGAP6	11.8362
DDX49	12.74455
SHMT2	-23.2098
ZC3H6	1.107972
TSPAN13	-2.39989
RBM5	22.30686
SAMD11	-10.1522
TUBGCP3	9.345096
PEX7	11.00753
TIA1	-11.5128
MXI1	-9.47718
AKT1	12.11588
CLN8	-9.96903
RP13-1032I1.10	-22.9388
ZFP2	-9.66706
NCAM1	-9.60648
TANGO2	-10.44
FAM155A	-3.17304
Gm13872	-10.2798
GALNT13	-3.6162
SLC9A6	12.30943
TMEM167A	11.75055
LENG8	-20.6102
COL6A3	13.82606
ATF6B	-13.3609
FN1	-1.26365
OLFML1	2.066963
IFT80	-9.22269
CTXN1	-0.91545
TBC1D3D	-9.74762
CEP57L1	-4.50118
DDX3X	-6.5443
RBM39	-4.49391
FOXP2	-14.4563

ABCA5	3.311803
ZNF559-ZNF177	9.023245
DLST	-14.4821
ADGRG6	-1.9986
GOLGA8A	1.327156
KPNA1	-11.1031
TGFB1I1	-10.0819
TLCD2	-18.9003
PIGN	-19.591
BMPR1A	-2.62835
G6PD	-10.7723
CALHM2	-10.2802
GULP1	16.40496
CDC14B	0.618109
SLC25A3	-9.59028
PDGFRA	3.14567
PSMC3	14.57563
COA8	10.99048
CSNK2A1	11.35157
CPT1C	19.80716
ZEB1	6.911521
FOSB	-13.3305
SMN2	-3.45567
DDX39B	-3.46889
HRCT1	-1.62144
SYCP2	-15.5347
IFNAR1	-9.72699
RAB23	-0.48046
NEDD4L	8.867427
USP53	-5.34195
DCUN1D2	-10.392
B3GNTL1	-9.01484
NRDC	8.668385
SMARCA4	-8.28533
ABCD1	-7.02062
SLC1A7	3.180883
STAG1	-8.12694
OSBPL6	-1.11188
SPHK2	17.77921
CCDC144A	-1.13672
E2F6	-8.45617
RUNX3	1.403426
PABIR3	-4.56309
PTPRN	-7.33722
RASA4	0.700763
MAPKAP1	8.480005
PARD3	11.30835
DGCR8	8.51585

CALM2	-0.55808
SRGAP2B	2.911576
DGAT2	-8.47888
CSNK2A2	-3.69029
TFPI	6.695975
GGT5	-8.40452
SPG7	1.010273
PPT1	-8.74168
ALDH7A1	11.04047
FCHSD2	0.733502
INTS6L	10.23369
DIAPH1	-5.16142
EPS15L1	8.884727
GAB3	-7.53273
RASA1	-7.71625
ZNF638	2.823552
SLC16A4	3.205876
RPSAP53	3.331147
STXBP1	7.760459
SIRT6	8.633707
TBCA	8.339617
NMD3	9.013877
TMEM132A	-1.11791
NBAS	8.785204
ELOA	-2.55575
RGS16	-1.0369
CC2D1B	9.423048
WNK1	9.507005
TLR4	-2.75065
TRANK1	7.073135
UBAP1	-8.08086
CBFA2T2	8.647666
TPCN1	0.711537
PLEKHB2	-0.58146
MIEF1	-5.17671
NREP	-0.48897
TAF1C	7.786628
PMEPA1	-0.85477
PTPA-222	9.411526
SEMA6D	13.73274
LZTS2	9.41881
HLA-DPB1	-5.08381
PCSK5	-1.45827
NXF1	7.842391
ECI1	-0.60588
SERPINB6	4.878063
PLEKHG5	-10.1416
DTX3L	0.496492
UBC	0.790754

CD59	-8.80002
NFE2L1	-14.6165
RP11-61L23.2	4.758098
CCDC88A	20.86831
NPR3	-0.96343
HSD17B4	-7.9636
FAM117A	-8.25578
ZSWIM8	5.269046
TPTEP1	2.020311
NFYC	7.792351
HOXB6	-2.55752
SCAP	-10.1102
ANK2	2.368678
POM121L1	-2.06278
BDNF	7.110552
CELSR1	-2.08513
EFNA5	1.897925
MTMR1	2.695449
CANX	-0.51146

Table 8.3.2:Top 200 ranked genes included in Gene Set Enrichment Analysis in ACBD5-deficient patients with retinal dystrophy and those without. Genes uploaded to WebGestalt for analysis.

Gene Set	ES	FDR	P Value	Gene Symbol	Gene names
Translational initiation	0.99	0.69	0.47	ATF4	activating transcription factor 4
				DDX3X	DEAD-box helicase 3 X-linked
				EIF3E	eukaryotic translation initiation factor 3 subunit E
				EIF3F	eukaryotic translation initiation factor 3 subunit F
				EIF4G3	eukaryotic translation initiation factor 4 gamma 3
				EIF4A1	eukaryotic translation initiation factor 4A1
				EIF4A2	eukaryotic translation initiation factor 4A2
				EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3
				RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
Protein localization to endoplasmic reticulum	1.5	0.39	0.04	GBF1	golgi brefeldin A resistant guanine nucleotide exchange factor 1
				RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
				RPL13	ribosomal protein L13
				RPL14	ribosomal protein L14
				RPL27A	ribosomal protein L27a
				RPL37A	ribosomal protein L37a

				RPL4	ribosomal protein L4
				RPS8	ribosomal protein S8
				SSR3	signal sequence receptor subunit 3
Positive regulation of cellular protein localization	1.3	0.46	0.12	EPHB2	EPH receptor B2
				FYN	FYN proto-oncogene, Src family tyrosine kinase
				LRP4	LDL receptor related protein 4
				PARK7	Parkinsonism associated deglycase
				PCM1	pericentriolar material 1
				SPTBN1	spectrin beta, non-erythrocytic 1
				TFDP1	transcription factor Dp-1
				TCAF1	TRPM8 channel associated factor 1
Ras protein signal transduction	-1.3	0.55	0.19	DNM2	dynamin 2
				LPAR4	lysophosphatidic acid receptor 4
Positive regulation of defence response	1.01	0.7	0.41	CYBA	cytochrome b-245 alpha chain
				DAB2IP	DAB2 interacting protein
				DDX60	DExD/H-box helicase 60
				FYN	FYN proto-oncogene, Src family tyrosine kinase
				HMGB2	high mobility group box 2

				PARK7	Parkinsonism associated deglycase
				PDCD4	programmed cell death 4
				TLR4	toll like receptor 4
Neutrophil mediated immunity	-1.65	0.39	0.01	PSMD2	proteasome 26S subunit, non-ATPase 2
				UBR4	ubiquitin protein ligase E3 component n-recognin 4
Granulocyte activation	-1.65	0.36	0.13	PSMD2	proteasome 26S subunit, non-ATPase 2
				UBR4	ubiquitin protein ligase E3 component n-recognin 4
Epithelial cell proliferation	-1.27	0.56	0.2	LAMB1	laminin subunit beta 1
				PPARG	peroxisome proliferator activated receptor gamma
Regulation of small GTPase mediated signal transduction	-1.3	0.56	0.18	DNM2	dynamin 2
				LPAR4	lysophosphatidic acid receptor 4
Positive regulation of establishment of protein localization	1.25	0.58	0.19	FYN	FYN proto-oncogene, Src family tyrosine kinase
				PCM1	pericentriolar material 1
				SEC24A	SEC24 homolog A, COPII coat complex component
				SPTBN1	spectrin beta, non-erythrocytic 1
				TLR4	toll like receptor 4

				TFDP1	transcription factor Dp-1
				TCAF1	TRPM8 channel associated factor 1
Ribonucleotide metabolic process	0.96	0.71	0.47	CAD	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase
				ECD	ecdysoneless cell cycle regulator
				NDUFB3	NADH:ubiquinone oxidoreductase subunit B3
				NUDT10	nudix hydrolase 10
				PARK7	Parkinsonism associated deglycase
				PDE7A	phosphodiesterase 7A
				SLC25A1	solute carrier family 25 member 1
Coenzyme metabolic process	1.54	0.54	0.04	DCXR	dicarbonyl and L-xylulose reductase
				ECD	ecdysoneless cell cycle regulator
				SLC2A1	solute carrier family 2 member 1
				SLC2A3	solute carrier family 2 member 3
				SLC25A1	solute carrier family 25 member 1
				TKT	transketolase
Organic anion transport	-1.31	0.56	0.18	PPARG	peroxisome proliferator activated receptor gamma

Regulation of protein modification by small protein conjugation or removal	1.5	0.33	0.06	FYN	FYN proto-oncogene, Src family tyrosine kinase
				PARK7	Parkinsonism associated deglycase
				RPL11	ribosomal protein L11
				UBE3A	ubiquitin protein ligase E3A
				UBXN1	UBX domain protein 1
Embryonic organ development	1.38	0.45	0.11	EPHB2	EPH receptor B2
				MICAL2	microtubule associated monooxygenase, calponin and LIM domain containing 2
				RARG	retinoic acid receptor gamma
				RPL10	ribosomal protein L10
				TBX15	T-box 15
Regulation of transporter activity	-1.67	0.55	0.003	DNM2	dynamin 2
				PPARG	peroxisome proliferator activated receptor gamma
Positive regulation of cell projection organization	-1.28	0.55	0.21	DNM2	dynamin 2
				EPHA4	EPH receptor A4
Positive regulation of catabolic process	-1.27	0.55	0.23	HDAC6	histone deacetylase 6
				TSC1	TSC complex subunit 1
Negative regulation of response to external stimulus	-1.31	0.56	0.16	EPHA4	EPH receptor A4

				PPARG	peroxisome proliferator activated receptor gamma
Sensory system development	1.17	0.69	0.23	RHOJ	ras homolog family member J
				SMG9	S+A1:F88MG9, nonsense mediated mRNA decay factor

Table 8.3.3: Gene Set Enrichment results showing enriched pathways in ACBD5-deficient patients, compared to controls. ER, normalised enrichment score; FDR, false discovery rate. Data analysed using WebGestalt

Gene Set	**ES**	**FDR**	**P Value**	**Gene Symbol**	**Gene names**
Regulation of protein catabolic process	1.9	0.25	0.008	UBE3A	ubiquitin protein ligase E3A
				PSMC3	proteasome 26S subunit, ATPase 3
				BANP	BTG3 associated nuclear protein
				AKT1	AKT serine/threonine kinase 1
				CSNK2A1	casein kinase 2 alpha 1
				NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase
				NRDC	nardilysin convertase
Regulation of peptidase activity	1.64	0.31	0.02	COL6A3	collagen type VI alpha 3 chain
				AKT1	AKT serine/threonine kinase 1
				CSNK2A1	casein kinase 2 alpha 1

				NLRP1	NLR family pyrin domain containing 1
				MBP	myelin basic protein
				NRDC	nardilysin convertase
				TFPI	tissue factor pathway inhibitor
				SERPINB6	serpin family B member 6
Proteasomal protein catabolic process	1.73	0.32	0.018	UBE3A	ubiquitin protein ligase E3A
				PSMC3	proteasome 26S subunit, ATPase 3
				AKT1	AKT serine/threonine kinase 1
				NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase
Positive regulation of proteolysis	1.68	0.32	0.03	PSMC3	proteasome 26S subunit, ATPase 3
				AKT1	AKT serine/threonine kinase 1
				NLRP1	NLR family pyrin domain containing 1
				MBP	myelin basic protein
				NRDC	nardilysin convertase
Activation of protein kinase activity	1.66	0.32	0.024	CCDC88A	coiled-coil domain containing 88A
				AKT1	AKT serine/threonine kinase 1
				WNK1	WNK lysine deficient protein kinase 1

				MAPKAP1	mitogen-activated protein kinase associated protein 1
Peptidyl-threonine modification	1.7	0.33	0.28	AKT1	AKT serine/threonine kinase 1
				CSNK2A1	casein kinase 2 alpha 1
				PARD3	par-3 family cell polarity regulator
				WNK1	WNK lysine deficient protein kinase 1
Negative regulation of phosphorylation	1.62	0.33	0.57	AKT1	AKT serine/threonine kinase 1
				PARD3	par-3 family cell polarity regulator
				WNK1	WNK lysine deficient protein kinase 1
				SIRT6	sirtuin 6
Circulatory system process	1.55	0.41	0.05	AKT1	AKT serine/threonine kinase 1
				WNK1	WNK lysine deficient protein kinase 1
				NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase
Negative regulation of transport	1.74	0.41	0.18	AKT1	AKT serine/threonine kinase 1
				WNK1	WNK lysine deficient protein kinase 1
				NEDD4L	neural precursor cell expressed, developmentally down-

					regulated 4-like, E3 ubiquitin protein ligase
				SIRT6	sirtuin 6
Regulation of neuron projection development	1.52	0.45	0.58	CCDC88A	coiled-coil domain containing 88A
				SEMA6D	semaphorin 6D
				AKT1	AKT serine/threonine kinase 1
				NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase
				CBFA2T2	CBFA2/RUNX1 translocation partner 2
				BDNF	brain derived neurotrophic factor
				ZSWIM8	zinc finger SWIM-type containing 8
Negative regulation of response to external stimulus	-0.63	0.99	0.92	NT5E	5'-nucleotidase ecto
				RIOK3	RIO kinase 3
Ras protein signal transduction	-0.5	0.99	0.99	RHOG	ras homolog family member G
				PLEKHG5	pleckstrin homology and RhoGEF domain containing G5
				RASA1	RAS p21 protein activator 1
Regulation of protein complex assembly	-0.66	0.99	0.87	CTNNB1	catenin beta 1
				RIOK3	RIO kinase 3

Neutrophil mediated immunity	-0.67	0.99	0.88	TUBB	tubulin beta class I
				RHOG	ras homolog family member G
				CD59	CD59 molecule (CD59 blood group)
				DDX3X	DEAD-box helicase 3 X-linked
				DIAPH1	diaphanous related formin 1
Granulocyte activation	-0.67	0.99	0.88	TUBB	tubulin beta class I
				RHOG	ras homolog family member G
				CD59	CD59 molecule (CD59 blood group)
				DDX3X	DEAD-box helicase 3 X-linked
				DIAPH1	diaphanous related formin 1
Connective tissue development	-0.63	0.99	0.9	CTNNB1	catenin beta 1
				IFT80	intraflagellar transport 80
				DGAT2	diacylglycerol O-acyltransferase 2
Neuron death	-0.41	0.99	0.99	CTNNB1	catenin beta 1
				G6PD	glucose-6-phosphate dehydrogenase
				PPT1	palmitoyl-protein thioesterase 1
				RASA1	RAS p21 protein activator 1
				USP53	ubiquitin specific peptidase 53

Cellular amino acid metabolic process	-1.54	1	0.03	CBSL	cystathionine-beta-synthase like
				SHMT2	serine hydroxymethyltransferase 2
				VARS2	valyl-tRNA synthetase 2, mitochondrial
				DLST	dihydrolipoamide S-succinyltransferase
Cell-substrate adhesion	-1.5	1	0.04	LIMS1	LIM zinc finger domain containing 1
				DDR1	discoidin domain receptor tyrosine kinase 1
				CTNNB1	catenin beta 1
Sulfur compound metabolic process	-1.5	1	0.05	CBSL	cystathionine-beta-synthase like
				DLST	dihydrolipoamide S-succinyltransferase
				G6PD	glucose-6-phosphate dehydrogenase
				PPT1	palmitoyl-protein thioesterase 1
				DGAT2	diacylglycerol O-acyltransferase 2
				GGT5	gamma-glutamyltransferase 5
				HSD17B4	hydroxysteroid 17-beta dehydrogenase 4
				ABCD1	ATP binding cassette subfamily D member 1

Table 8.3.4: Gene Set Enrichment Analysis results showing enriched pathways in ACBD5-deficient patients with retinal dystrophy, compared to those without retinal dystrophy. *ES, normalised enrichment score; FDR, false discovery rate.* Data analysed using WebGestalt

Gene Set	ER	FDR	P Value	Gene Symbol	Gene Name
SRP-dependent co-translational protein targeting to membrane	9.6	0.0009	4.17E-07	RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
				RPL13	ribosomal protein L13
				RPL14	ribosomal protein L14
				RPL27A	ribosomal protein L27a
				RPL37A	ribosomal protein L37a
				RPL4	ribosomal protein L4
				RPS8	ribosomal protein S8
				SSR3	signal sequence receptor subunit 3
Cotranslational protein targeting to membrane	9.2	0.0009	5.93E-07	RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
				RPL13	ribosomal protein L13
				RPL14	ribosomal protein L14
				RPL27A	ribosomal protein L27a
				RPL37A	ribosomal protein L37a
				RPL4	ribosomal protein L4
				RPS8	ribosomal protein S8
				SSR3	signal sequence receptor subunit 3
Protein targeting to ER	8.5	0.001	0.000001	RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
				RPL13	ribosomal protein L13
				RPL14	ribosomal protein L14
				RPL27A	ribosomal protein L27a
				RPL37A	ribosomal protein L37a

				RPL4	ribosomal protein L4
				RPS8	ribosomal protein S8
				SSR3	signal sequence receptor subunit 3
Translational initiation	8.4	7.10E-07	7.81E-11	ATF4	activating transcription factor 4
				DDX3X	DEAD-box helicase 3 X-linked
				EIF3E	eukaryotic translation initiation factor 3 subunit E
				EIF3F	eukaryotic translation initiation factor 3 subunit F
				EIF4A1	eukaryotic translation initiation factor 4A1
				EIF4A2	eukaryotic translation initiation factor 4A2
				EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3
				EIF4G3	eukaryotic translation initiation factor 4 gamma 3
				RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	8.4	0.0009	3.13E-07	EIF3E	eukaryotic translation initiation factor 3 subunit E
				RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
				RPL13	ribosomal protein L13
				RPL14	ribosomal protein L14

				RPL27A	ribosomal protein L27a
				RPL37A	ribosomal protein L37a
				RPL4	ribosomal protein L4
				RPS8	ribosomal protein S8
				SMG9	SMG9, nonsense mediated mRNA decay factor
Protein localization to endoplasmic reticulum	7.4	0.001	1.0723E-06	GBF1	golgi brefeldin A resistant guanine nucleotide exchange factor 1
				RPL10	ribosomal protein L10
				RPL11	ribosomal protein L11
				RPL13	ribosomal protein L13
				RPL14	ribosomal protein L14
				RPL27A	ribosomal protein L27a
				RPL37A	ribosomal protein L37a
				RPL4	ribosomal protein L4
				RPS8	ribosomal protein S8
				SSR3	signal sequence receptor subunit 3
Translation	3.5	0.001	7.23E-07	ATF4	activating transcription factor 4
				CIRBP	cold inducible RNA binding protein
				DDX3X	DEAD-box helicase 3 X-linked
				EIF3E	eukaryotic translation initiation factor 3 subunit E
				EIF3F	eukaryotic translation initiation factor 3 subunit F

				EIF4A1	eukaryotic translation initiation factor 4A1
				EIF4A2	eukaryotic translation initiation factor 4A2
				EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3
				EIF4G3	eukaryotic translation initiation factor 4 gamma 3
				MRPL32	mitochondrial ribosomal protein L32
Peptide metabolic process	3.2	0.0009	2.01E-07	ANPEP	alanyl aminopeptidase, membrane
				ATF4	activating transcription factor 4
				CIRBP	cold inducible RNA binding protein
				DDX3X	DEAD-box helicase 3 X-linked
				EIF3E	eukaryotic translation initiation factor 3 subunit E
				EIF3F	eukaryotic translation initiation factor 3 subunit F
				EIF4A1	eukaryotic translation initiation factor 4A1
				EIF4A2	eukaryotic translation initiation factor 4A2
				EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3

				EIF4G3	eukaryotic translation initiation factor 4 gamma 3
Amide biosynthetic process	3.2	0.0009	5.45E-07	ATF4	activating transcription factor 4
				CAD	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase
				CIRBP	cold inducible RNA binding protein
				DDX3X	DEAD-box helicase 3 X-linked
				EIF3E	eukaryotic translation initiation factor 3 subunit E
				EIF3F	eukaryotic translation initiation factor 3 subunit F
				EIF4A1	eukaryotic translation initiation factor 4A1
				EIF4A2	eukaryotic translation initiation factor 4A2
				EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3
				EIF4G3	eukaryotic translation initiation factor 4 gamma 3

Table 8.3.5: Over-Representation Analysis results showing enriched pathways in ACBD5-deficient patients compared to controls. ER, normalised enrichment score; FDR, false discovery rate. Data analysed using WebGestalt.

Gene Set	ER	FDR	P Value	Gene Symbol	Gene Name
Fatty acid beta-oxidation	8.4	0.1	0.00008	ABCD1	ATP binding cassette subfamily D member 1
				AKT1	AKT serine/threonine kinase 1
				CPT1C	carnitine palmitoyltransferase 1C
				ECI1	enoyl-CoA delta isomerase 1
				HSD17B4	hydroxysteroid 17-beta dehydrogenase 4
				PEX7	peroxisomal biogenesis factor 7
Fatty acid oxidation	7.2	0.1	0.00006	ABCD1	ATP binding cassette subfamily D member 1
				AKT1	AKT serine/threonine kinase 1
				CPT1C	carnitine palmitoyltransferase 1C
				DGAT2	diacylglycerol O-acyltransferase 2
				ECI1	enoyl-CoA delta isomerase 1
				HSD17B4	hydroxysteroid 17-beta dehydrogenase 4
				PEX7	peroxisomal biogenesis factor 7
Lipid oxidation	7	0.1	0.00006	ABCD1	ATP binding cassette subfamily D member 1
				AKT1	AKT serine/threonine kinase 1
				CPT1C	carnitine palmitoyltransferase 1C

				DGAT2	diacylglycerol O-acyltransferase 2
				ECI1	enoyl-CoA delta isomerase 1
				HSD17B4	hydroxysteroid 17-beta dehydrogenase 4
				PEX7	peroxisomal biogenesis factor 7
Spindle assembly	6.2	0.1	0.0001	CDC14B	cell division cycle 14B
				GOLGA8A	golgin A8 family member A
				LZTS2	leucine zipper tumor suppressor 2
				NEK6	NIMA related kinase 6
				STAG1	stromal antigen 1
				TUBB	tubulin beta class I
				TUBGCP3	tubulin gamma complex associated protein 3
Myelination	6	0.1	0.0001	ABCD1	ATP binding cassette subfamily D member 1
				ADGRG6	adhesion G protein-coupled receptor G6
				AKT1	AKT serine/threonine kinase 1
				ANK2	ankyrin 2
				CTNNB1	catenin beta 1
				MBP	myelin basic protein
				PARD3	par-3 family cell polarity regulator
Developmental cell growth	5.2	0.03	0.00001	BDNF	brain derived neurotrophic factor
				CTNNB1	catenin beta 1

				DDR1	discoidin domain receptor tyrosine kinase 1
				DDX39B	DExD-box helicase 39B
				EFNA5	ephrin A5
				FN1	fibronectin 1
				G6PD	glucose-6-phosphate dehydrogenase
				NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase
				SEMA6D	semaphorin 6D
				SIRT6	sirtuin 6
Cell growth	4	0.002	2.58E-07	AKT1	AKT serine/threonine kinase 1
				BDNF	brain derived neurotrophic factor
				CAPRIN2	caprin family member 2
				CCN4	cellular communication network factor 4
				CSNK2A1	casein kinase 2 alpha 1
				CTNNB1	catenin beta 1
				DDR1	discoidin domain receptor tyrosine kinase 1
				DDX39B	DExD-box helicase 39B
				DDX3X	DEAD-box helicase 3 X-linked
				EFNA5	ephrin A5

Regulation of cell growth	3.9	0.02	0.000003	AKT1	AKT serine/threonine kinase 1
				BDNF	brain derived neurotrophic factor
				CAPRIN2	caprin family member 2
				CCN4	cellular communication network factor 4
				CSNK2A1	casein kinase 2 alpha 1
				DDR1	discoidin domain receptor tyrosine kinase 1
				DDX39B	DExD-box helicase 39B
				DDX3X	DEAD-box helicase 3 X-linked
				EFNA5	ephrin A5
				FN1	fibronectin 1
Lipid modification	3.8	0.1	0.0002	ABCD1	ATP binding cassette subfamily D member 1
				AKT1	AKT serine/threonine kinase 1
				CPT1C	carnitine palmitoyltransferase 1C
				DGAT2	diacylglycerol O-acyltransferase 2
				ECI1	enoyl-CoA delta isomerase 1
				HSD17B4	hydroxysteroid 17-beta dehydrogenase 4
				MTMR1	myotubularin related protein 1
				PDGFRA	platelet derived growth factor receptor alpha

				PEX7	peroxisomal biogenesis factor 7
				RHOG	ras homolog family member G
Symbiont process	2.6	0.1	0.00008	BRD2	bromodomain containing 2
				CTNNB1	catenin beta 1
				DDX39B	DExD-box helicase 39B
				DDX3X	DEAD-box helicase 3 X-linked
				FN1	fibronectin 1
				ILF3	interleukin enhancer binding factor 3
				KPNA1	karyopherin subunit alpha 1
				MICA	MHC class I polypeptide-related sequence A
				NCAM1	neural cell adhesion molecule 1
				NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase

Table 8.3.6: Over-Representation Analysis results showing enriched pathways in ACBD5-deficient patients with retinal dystrophy, compared to those without retinal dystrophy. *ER, normalised enrichment score; FDR, false discovery rate*. Data analysed using WebGestalt.

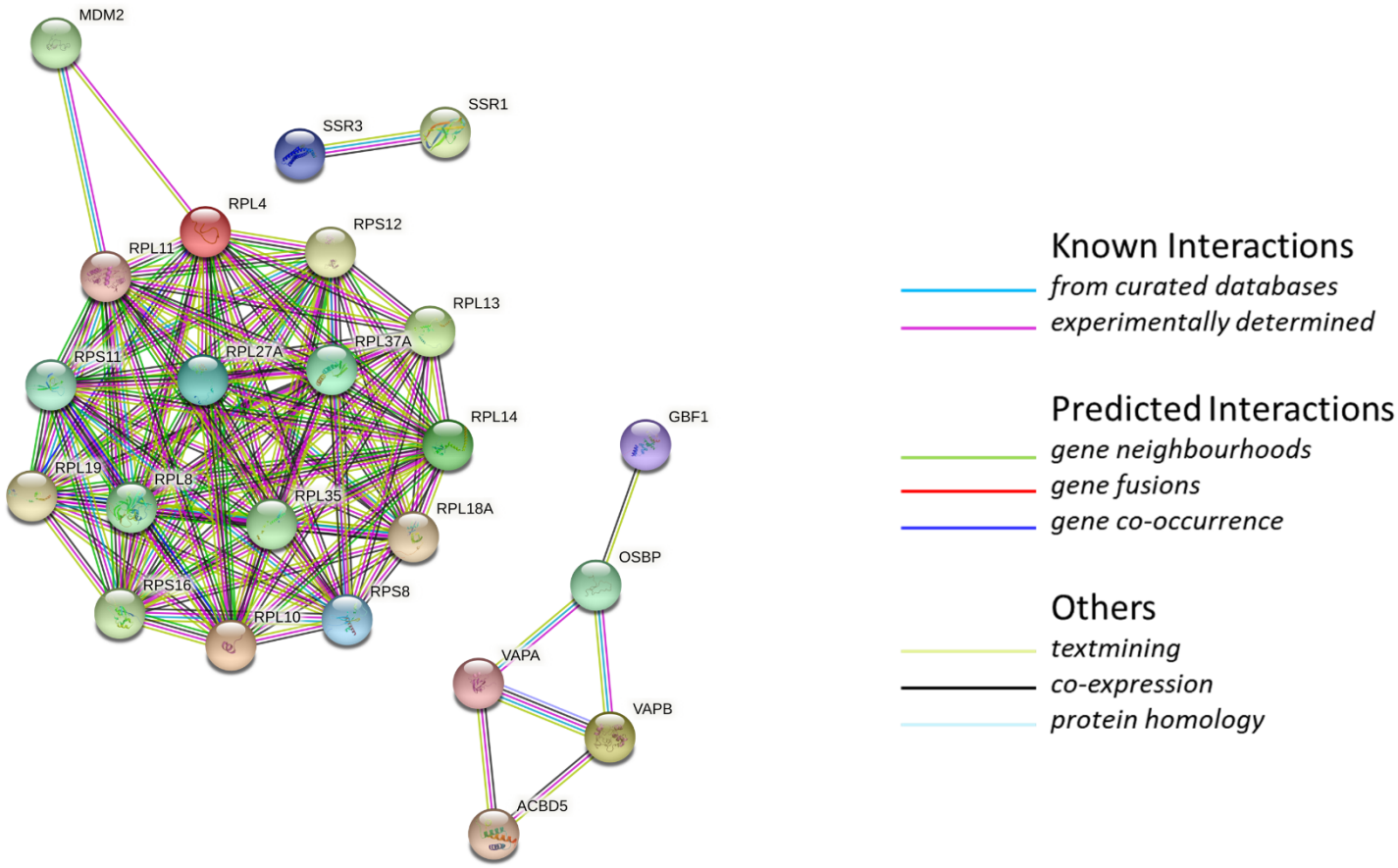


Figure 8.3.2: Protein-protein interaction from gene set involved in protein localisation to the ER, showing no direct interaction with ACBD5, VAPA or VAPB. Image generated using STRING.

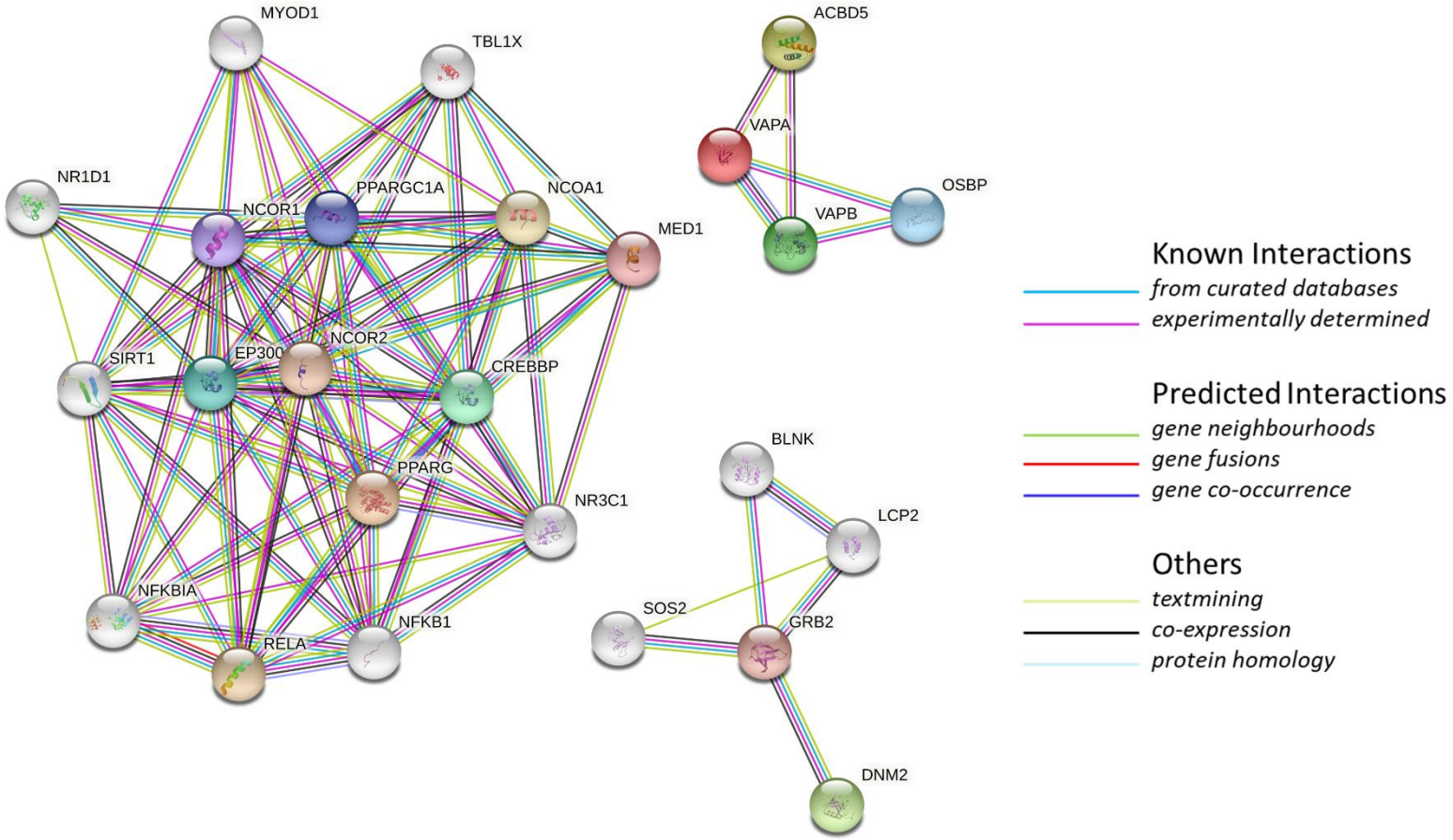


Figure 8.3.2: Protein-protein interactions for gene set involved in transporter activation, showing no interaction with ACBD5, VAPA or VAPB. Image generated using STRING.

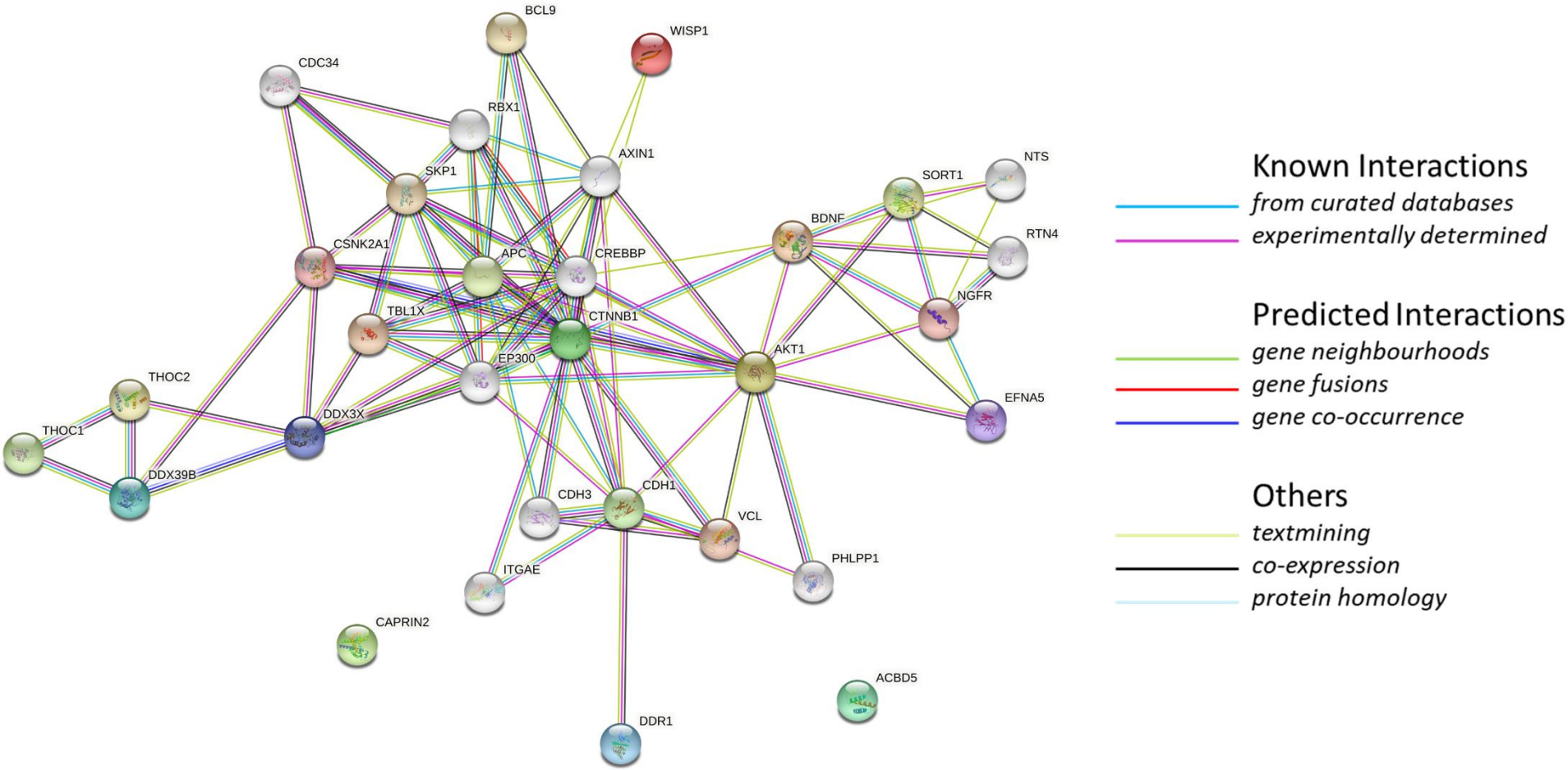


Figure 8.3.4: Protein-protein interaction network for gene set involved in cell growth, showing no interaction with ACBD5. Image generated using STRING.

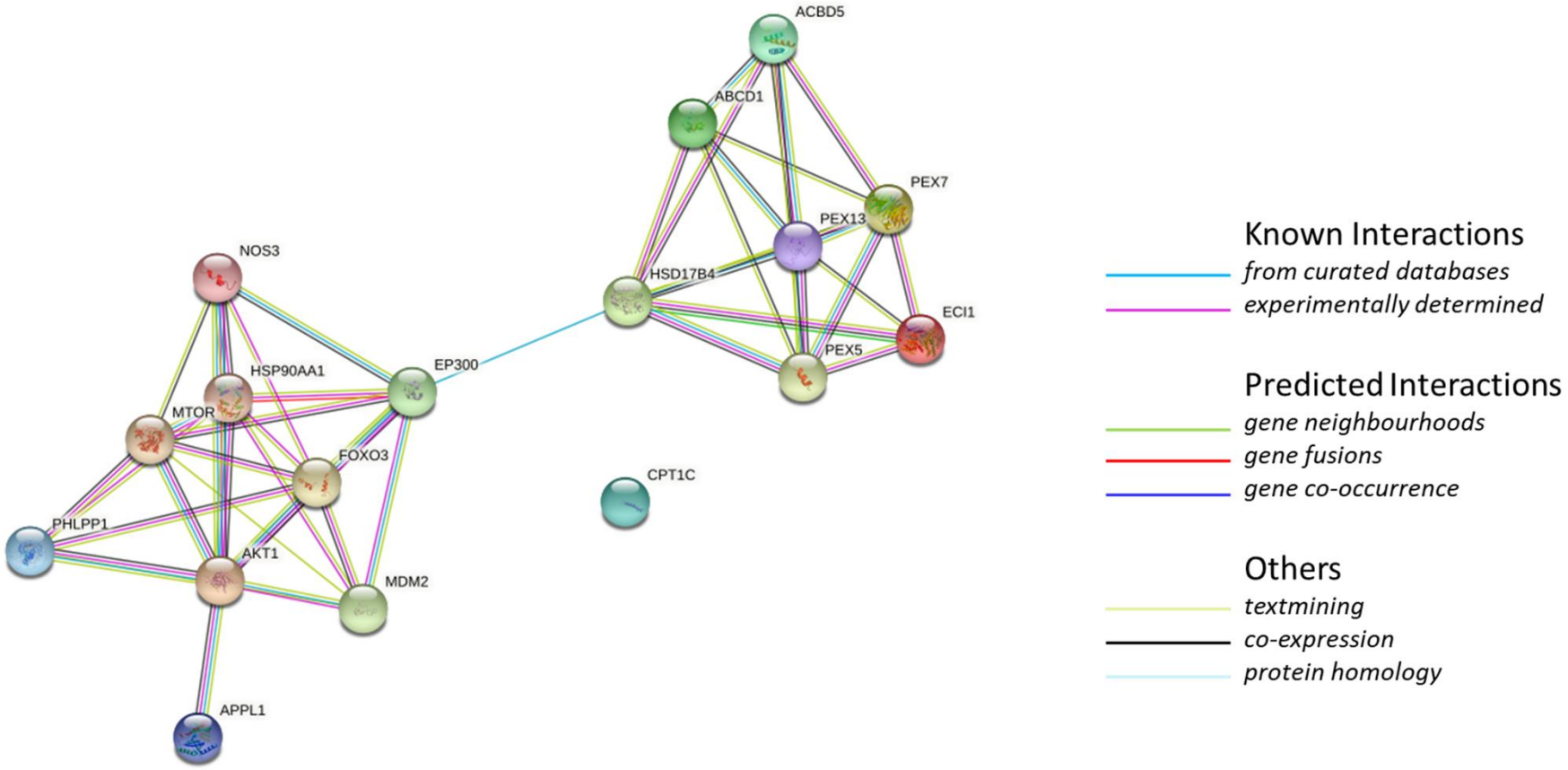


Figure 8.3.5: Protein-protein interaction network for gene set involved with fatty acid β-oxidation, fatty acid oxidation and lipid modification, showing direct and indirect interaction with ACBD5. Image generated using STRING.