September 2022

Exploring next generation sequencing in the diagnosis of inherited ataxia and other neurological diseases

Joshua Hersheson, MBBS, MRCP

This thesis is submitted to the University College of London for the degree of Doctor of Philosophy

Institute of Neurology
University College London

Declaration:

I, Joshua Hersheson 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.

Abstract

Neurodegenerative diseases are a complex and heterogeneous group of disorders characterised by progressive loss of cells from the central nervous system which can result in a broad range of clinical phenotypes including cognitive impairment, motor dysfunction, epilepsy and visual loss. Next generation sequencing technologies have greatly facilitated the search for novel genes underlying neurodegenerative disorders. In this thesis, I present the results of my efforts to understand and clarify the genetic bases for a range of neurological disorders with a focus on cerebellar ataxia. I have used a range of techniques including whole exome sequencing, together with traditional genetic approaches to investigate several families with hitherto genetically undiagnosed neurological syndromes.

The first project I present is the identification of a novel gene (TUBB4A) underlying a form of dominantly inherited dystonia (DYT4). With the understanding I gained in the application of exome sequencing and mapping strategies, I go on to present my findings of a further novel gene underlying a recessive cerebellar ataxia (CAPN1). I describe the identification of the genetic cause in several families with recessive ataxia and explore functional studies to confirm the pathogenicity of novel mutations in a rare form of neuronal ceroid lipofuscinosis (CLN10). I present a study in which I developed a next-generation sequencing panel for large scale sequencing of a cohort of autosomal dominant ataxia patients. Finally, I present data on the identification of PNPLA6 mutations as the genetic cause of a rare form of spastic ataxia associated with hypogonadrotropic hypogonadism – Oliver McFarlane syndrome.

Impact Statement

Inherited neurodegenerative disorders such as cerebellar ataxia are relatively rare however patients suffering with these conditions, typically experience significant morbidity and genetic diagnostic uncertainty which can impact for example their ability to explore options for reproductive planning such as pre-implantation genetic diagnosis. Effective disease modifying therapies for neurodegenerative disorders have until recent years been extremely limited however there has been great optimism generated for example with the advent of genetic therapies such as those for spinal muscular atrophy which offer real hope for patients with incurable, progressive disorders. Key to being able to develop effective therapies is an understanding of the pathophysiology of these disorders which invariably begins with the elucidation of their genetic basis. The rapid progress in genetic sequencing technology and a greater understanding of the complexity of the human genome has underpinned many of the advances in the study and treatment of neurodegenerative disease over the past 10 years.

In this research I have employed these techniques to investigate the genetic cause of neurological disease in a range of families affected by cerebellar ataxia and dystonia. In my first project I was able to solve the genetic puzzle of a multi-generational family affected by an unusual form of generalised dystonia (DYT4) to identify a mutation in a gene called TUBB4A. Whilst DYT4 dystonia is quite rare, it was subsequently shown that mutations in this gene are also the cause of a severe neurodegenerative leukoencephalopathy called H-ABC syndrome which causes severe disability in affected children. The insights gained from these studies will benefit patients through a greater understanding of the mechanisms through which the TUBB4A mutation causes disease but also in being able to provide a genetic diagnosis in those affected which can be helpful in informing the prognosis and expected disease course.

In other sections I explore the genetic causes of disease in families affected by recessive cerebellar ataxia including the identification of a novel gene (CAPN1) and an examination of

patients with a form of neuronal ceroid lipofuscinosis (Batten's disease) who had mutations in a gene called CTSD. CTSD may yet prove to be a treatable disorder with enzyme replacement therapy offering hope to other children and families affected by these disorders.

All of the chapters in this thesis have led to published articles in high-impact journals. I have also presented data at several conferences including the Movement Disorders Congress 2012 and the Ataxia Society UK conference which brings together researchers and patients to highlight advances and developments in ataxia research.

Acknowledgements

I am extremely grateful to my supervisors Professor Henry Houlden and Professor Nicholas Wood for their support during my rather extended studies and throughout my time at the Institute of Neurology. I am indebted to Professor Houlden for his mentorship, advice and training in neurogenetics which laid the foundation for my subsequent career as a neurologist.

I particularly would like to thank my friend, colleague and mentor, Niccolo Mencacci for his support, counsel and infectious enthusiasm. I would also like to thank the following individuals: Una Sheerin, Arianna Tucci, Lucia Schottlaender, Gavin Charlesworth, Mina Ryten, Alan Pittman, Hallgeir Jonvik, Deborah Hughes, Paola Giunti and Mike Parkinson.

Finally, my most heartfelt thanks go to my wife Caroline and daughter Ellie, for their love and unwavering support in my endeavours and to whom this thesis is dedicated.

Table of Contents

Contents

Abstract 3

Impact Statement 4

Acknowledgements 6

Table of Contents 7

Abbreviations 12

Publications 14

Figures 16

Tables 17

1 Introduction 18

1.1 Genetic basis of neurodegenerative disease 18

1.2 A brief history of genetic research 20

1.3 Next generation sequencing 23

1.4 Exome sequencing in Mendelian disease gene discovery. 26

1.4.1 Identifying the signal of disease-causing mutations from the noise of human genetic variability 27

1.4.2 Prioritisation of variants based on rarity 28

1.4.3 Filtering based on mode of inheritance and intersection analysis 30

1.4.4 Use of mapping strategies 32

1.4.5 Prioritisation based on deleteriousness 34

1.4.6 Limitations of next generation sequencing 35

1.5 Genetics of inherited cerebellar ataxia 38

1.5.1 Introduction 38

1.5.2 The cerebellum 39

1.5.3 Classification of hereditary ataxias 40

1.5.4 Autosomal dominant cerebellar ataxias (ADCAs) 41

1.5.5 Episodic ataxias 56

1.5.6 Autosomal recessive cerebellar ataxias (ARCAs) 57

1.6 Genetics of dystonia 62

1.6.1 Introduction 62

1.6.2 Monogenic causes of isolated dystonia 65

2 Materials and methods 69

2.1 Case ascertainment and clinical assessment 69

2.2 Ethics and consent 69

2.3 Skin biopsy and fibroblast culture 69
2.4 DNA extraction, quality assessment and quantification 70
2.4.1 DNA extraction from blood 70
2.4.2 DNA quantification using spectrophotometry 71
2.4.3 DNA quantification using fluorescence 71
2.5 Primer design 72
2.6 Polymerase chain reaction (PCR) 72
2.7 Agarose Gel Electrophoresis 73
2.8 PCR purification 73
2.9 Sanger Sequencing 74
2.9.1 Sequencing reaction 74
2.9.2 Sequencing reaction purification 74
2.9.3 Sanger sequencing and analysis 75
2.10 Whole genome SNP genotyping using DNA SNP arrays 75
2.11 Genome-Wide Parametric Linkage Analysis 75
2.12 Homozygosity mapping 76
2.13 Whole Exome Sequencing (WES) 77
2.14 Bioinformatic analysis of whole exome sequencing data 77
3 Application of exome sequencing and linkage analysis to identify a novel gene causing autosomal dominant dystonia 79
3.1 Outline of the chapter 79
3.2 Statement of contribution 79
3.3 Identification of TUBB4a as the gene underlying DYT4 dystonia using linkage analysis and exome sequencing. 81
3.3.1 Introduction 81
3.3.2 Subjects, materials and methods 82
3.3.3 Results 87
3.3.4 Discussion 93
3.4 TUBB4a mutation screening in a series of patients with H-ABC 100
3.4.1 Introduction 100
3.4.2 Methods 101
3.4.3 Results 101
3.4.4 Discussion 104
4 Identification of CAPN1 as cause of recessive spastic ataxia 108
4.1 Statement of contribution 108
4.2 Introduction 109

4.3 Subjects, materials and methods 110
4.3.1 Subjects 110
4.3.2 Homozygosity mapping 110
4.3.3 Exome sequencing 110
4.3.4 cDNA sequencing 111
4.3.5 In-silico homology modelling of mutant CAPN1 protein 111
4.3.6 Fibroblast biopsy and culture 111
4.3.7 Calpain-1 western blot in patient fibroblasts (Baudry lab) 111
4.3.8 Calpain-1 activity assay in patient fibroblasts (Baudry lab) 112
4.4 Results 113
4.4.1 Family 1 113
4.4.2 Calpain-1 expression and calpain activity in patient fibroblasts (Baudry lab) .. 118
4.4.3 Family 2 119
4.4.4 Family 3 121
4.4.5 Family 4 122
4.5 Discussion 123
4.5.1 Cerebellar phenotype in calpain-1 knockout mice 124
4.5.2 Role of calpain-1 in neurodegeneration 125
4.5.3 Clinical spectrum associated with CAPN1 mutations 126
4.5.4 Conclusion 127
5 Targeted amplicon sequencing in a UK autosomal dominant ataxia cohort 129
5.1 Statement of contribution 129
5.2 Introduction 130
5.3 Materials and Methods 130
Patient selection 131
5.3.1 Sample preparation 131
5.3.2 Amplicon design 131
5.3.3 Library preparation and sequencing 132
5.3.4 Bioinformatic analysis 133
5.4 Results 134
5.4.1 Sequencing and coverage statistics 134
5.4.2 Genetic results 136
5.4.3 Clinical and genetic summary of patient with SCA14 and novel *PRKCG* p.A458T mutation 137
5.5 Discussion 138

6 Identification of two families with cathepsin-D deficiency associated neuronal ceroid lipofuscinosis 141
6.1 Statement of contribution 141
6.2 Introduction 142
6.3 Methods 143
6.3.1 Sample collection 143
6.3.2 Homozygosity mapping 143
6.3.3 Exome sequencing 144
6.3.4 Sanger sequencing 144
6.3.5 Immunohistochemistry 144
6.3.6 Electron microscopy 144
6.3.7 Cathepsin D (CatD) enzyme analysis 145
6.4 Results 147
6.4.1 Family A 147
6.4.2 Family B 149
6.4.3 Histopathology and electron microscopy findings 152
6.4.4 CatD kinetic analysis 154
6.5 Discussion 155
6.5.1 Diagnostic challenges 155
6.5.2 Neuronal ceroid lipofuscinosis type 10/CLN10 156
6.5.3 Conclusion 161
7 Identification of SYNE-1 associated cerebellar ataxia cases – genetic and clinical characterisation 163
7.1 Statement of contribution 163
7.2 Introduction 164
7.3 Methods 164
7.3.1 Patients 164
7.3.2 Genetic analysis 165
7.4 Results 166
7.4.1 Family 1 166
7.4.2 Family 2 167
7.4.3 Family 3 168
7.4.4 Discussion 169
8 A broad spectrum of clinical phenotypes is associated with mutations in *PNPLA6* 172
8.1 Statement of contribution 172
8.2 Introduction 173

Deleted: Syne

8.3 Materials and Methods 175
8.3.1 Patient identification and clinical characterisation 175
8.3.2 DNA and RNA sample collection and preparation 176
8.3.3 Exome sequencing 176
8.3.4 Homozygosity mapping 177
8.3.5 Sanger sequencing 177
8.3.6 Neuropathological assessment of post-mortem brain of A:1 177
8.4 Results 178
8.4.1 OMS/LMS cases: Clinical phenotype of A:1 178
8.4.2 Analysis of Family A, 4 other OMS families and 1 LMS family 181
8.4.3 Pure cerebellar ataxia case: Clinical phenotypes of B:1 and B:2 184
8.4.4 Analysis of Family B 185
8.5 Discussion 187
8.5.1 Clinical phenotypes associated with PNPLA6 mutations 189
8.5.2 Conclusion 190
9 Conclusions 192
9.1 Future Directions 198
10 References 200

Abbreviations

AD	Alzheimer's Disease
ADCA	autosomal dominant cerebellar ataxia
ALS	amyotrophic lateral sclerosis
ARCA	autosomal recessive cerebellar ataxia
ATP	adenosine triphosphate
BBS	Bardet Biedl Syndrome
BNS	Boucher -Neuhauser Syndrome
Bp	base pair
C9orf72	chromosome 9 open reading frame 72
CAPN1	calpain-1 gene
CatD	cathepsin D
cDNA	complementary deoxyribonucleic acid
CLN10	neuronal ceroid lipofuscinosis type 10
cM	centi-Morgan
CNV	copy number variant
CT	computed tomography
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
DRPLA	Dentatorubral-pallidoluysian atrophy
DYT4	dystonia type 4
EDTA	ethylenediaminetetraacetic acid
EEG	Electroencephalography
EVS	Exome Variant Server
EVS	Exome Variant Server
FBS	fetal bovine serum
FTD	frontotemporal dementia
GHS	Gordon-Holmes Syndrome
GWAS	genome wide association study
H-ABC	Hypomyelination with atrophy of the basal ganglia and cerebellum
HD	Huntington's disease
HSP	hereditary spastic paraplegia
HTT	huntingtin gene
IBD	identical by descent
ILOCA	idiopathic late onset cerebellar ataxia
IR	isolated retinitis
LMS	Laurence Moon Syndrome
LOD	logarithm of odds
LOF	loss of function
MAF	mean allele frequency
MAF	mean allele frequency
Mb	megabase
MDa	mega Dalton
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid

mtDNA	mitochondrial DNA
NCBI	National Centre for Biotechnology Information
NCL	neuronal ceroid lipofuscinosis
NGS	next generation sequencing
NHS	National Health Service
NIH	National Institute for Health
OMS	Oliver Mcfarlane Syndrome
OPIDN	organophosphate-induced delayed neuropathy
PCR	polymerase chain reaction
PD	Parkinson disease
PRT	Parson's Russell Terrier
RAN	repeat associated Non-ATG translation
RefSeq	Reference Sequence Database
RNA	ribonucleic acid
SCA	Spinocerebellar Ataxia
SNP	single nucleotide polymorphisms
SNV	single nucleotide polymorphism
STR	short tandem repeat
UTR	untranslated region
UV	ultra-violet
WAIS	Wechsler Adult Intelligence Scale
WES	whole exome sequencing
WGS	whole genome sequencing
WND	Wilson's disease
Wt	wild type

Publications

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

2. Mutations in the autoregulatory domain of β-tubulin 4a cause hereditary dystonia **Hersheson J,** Mencacci NE, Davis M, MacDonald N, Trabzuni D, Ryten M, Pittman A, Paudel R, Kara E, Fawcett K, Plagnol V, Bhatia KP, Medlar AJ, Stanescu HC, Hardy J, Kleta R, Wood NW, Houlden H. *Ann Neurol*. 2013 Apr;73(4): 19

3. Autosomal-recessive cerebellar ataxia caused by a novel ADCK3 mutation that elongates the protein: clinical, genetic and biochemical characterisation. Liu YT, **Hersheson J**, Plagnol V, Fawcett K, Duberley KE, Preza E, Hargreaves IP, Chalasani A, Laurá M, Wood NW, Reilly MM, Houlden H. *J Neurol Neurosurg Psychiatry. 2014 May*

4. Insights from cerebellar transcriptomic analysis into the pathogenesis of ataxia. Bettencourt C, Ryten M, Forabosco P, Schorge S, **Hersheson J**, Hardy J, Houlden H; United Kingdom Brain Expression Consortium. JAMA Neurol. 2014 Jul 1;71(7):831-9.

5. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. Mencacci NE, Isaias IU, Reich MM, Ganos C, Plagnol V, Polke JM, Bras J, **Hersheson J**, Stamelou M, Pittman AM, Noyce AJ, Mok KY, Opladen T, Kunstmann E, Hodecker S, Münchau A, Volkmann J, Samnick S, Sidle K, Nanji T, Sweeney MG, Houlden H, Batla A, Zecchinelli AL, Pezzoli G, Marotta G, Lees A, Alegria P, Krack P, Cormier-Dequaire F, Lesage S, Brice A, Heutink P, Gasser T, Lubbe SJ, Morris HR, Taba P, Koks S, Majounie E, Raphael Gibbs J, Singleton A, Hardy J, Klebe S, Bhatia KP, Wood NW; International Parkinson's Disease Genomics Consortium and UCL-exomes consortium. Brain. 2014 Sep;137(Pt 9)

6. Extended phenotypic spectrum of KIF5A mutations: From spastic paraplegia to axonal neuropathy. Liu YT, Laurá M, **Hersheson J**, Horga A, Jaunmuktane Z, Brandner S, Pittman A, Hughes D, Polke JM, Sweeney MG, Proukakis C, Janssen JC, Auer-Grumbach M, Zuchner S, Shields KG, Reilly MM, Houlden H. Neurology. 2014 Aug 12;83(7):612-9

7. Cathepsin D deficiency causes juvenile-onset ataxia and distinctive muscle pathology**. Hersheson J**, Burke D, Clayton R, Anderson G, Jacques TS, Mills P, Wood NW, Gissen P, Clayton P, Fearnley J, Mole SE, Houlden H. Neurology. 2014 Nov 11;83(20):1873-5

8. Mutations in SNX14 cause a distinctive autosomal-recessive cerebellar ataxia and intellectual disability syndrome. Thomas AC, Williams H, Setó-Salvia N, Bacchelli C, Jenkins D, O'Sullivan M, Mengrelis K, Ishida M, Ocaka L, Chanudet E, James C, Lescai F, Anderson G, Morrogh D, Ryten M, Duncan AJ, Pai YJ, Saraiva JM, Ramos F, Farren B, Saunders D, Vernay B, Gissen P, Straatmaan-Iwanowska A, Baas F, Wood NW, **Hersheson J**, Houlden H, Hurst J, Scott R, Bitner-Glindzicz M, Moore GE, Sousa SB, Stanier P. Am J Hum Genet. 2014 Nov 6;95(5):611-21

9. Neuropathy target esterase impairments cause Oliver-McFarlane and Laurence-Moon syndromes. Hufnagel RB, Arno G, Hein ND, **Hersheson J**, Prasad M, Anderson Y, Krueger LA, Gregory LC, Stoetzel C, Jaworek TJ, Hull S, Li A, Plagnol V, Willen CM, Morgan TM, Prows CA, Hegde RS, Riazuddin S, Grabowski GA, Richardson RJ, Dieterich K, Huang T, Revesz T, Martinez-Barbera JP, Sisk RA, Jefferies C, Houlden H, Dattani MT,

Fink JK, Dollfus H, Moore AT, Ahmed ZM. J Med Genet. 2015 Feb;52(2):85-94. doi: 10.1136/jmedgenet-2014-102856. Epub 2014 Dec 5

10. H-ABC syndrome and DYT4: Variable expressivity or pleiotropy of TUBB4 mutations? Erro R, **Hersheson J,** Ganos C, Mencacci NE, Stamelou M, Batla A, Thust SC, Bras JM, Guerreiro RJ, Hardy J, Quinn NP, Houlden H, Bhatia KP. Mov Disord. 2015 May;30(6):828-33.

11. A novel TUBB4A mutation suggests that genotype-phenotype correlation of H-ABC syndrome needs to be revisited. Erro R, **Hersheson J**, Houlden H, Bhatia KP. Brain. 2015 Aug;138(Pt 8):e370

12. ADCY5 mutations are another cause of benign hereditary chorea. Mencacci NE, Erro R, Wiethoff S**, Hersheson J**, Ryten M, Balint B, Ganos C, Stamelou M, Quinn N, Houlden H, Wood NW, Bhatia KP. Neurology. 2015 Jul 7;85(1):80-8.

13. The 4H syndrome due to RNF216 mutation. Ganos C, **Hersheson J**, Adams M, Bhatia KP, Houlden H. Parkinsonism Relat Disord. 2015 Sep;21(9):1122-3

14. Syndromic associations and RNF216 mutations. Ganos C, **Hersheson J**, Adams M, Bhatia KP, Houlden H. Parkinsonism Relat Disord. 2015 Nov;21(11):1389-90

15. Pure Cerebellar Ataxia with Homozygous Mutations in the PNPLA6 Gene. Wiethoff S, Bettencourt C, Paudel R, Madon P, Liu YT, **Hersheson J**, Wadia N, Desai J, Houlden H. Cerebellum. 2017 Feb;16(1):262-26

16. Heterogeneity in clinical features and disease severity in ataxia-associated SYNE1 mutations. Wiethoff S**, Hersheson J,** Bettencourt C, Wood NW, Houlden H. J Neurol. 2016 Aug;263(8):1503-10

17. Defects in the CAPN1 Gene Result in Alterations in Cerebellar Development and Cerebellar Ataxia in Mice and Humans. Wang Y, **Hersheson J**, Lopez D, Hammer M, Liu Y, Lee KH, Pinto V, Seinfeld J, Wiethoff S, Sun J, Amouri R, Hentati F, Baudry N, Tran J, Singleton AB, Coutelier M, Brice A, Stevanin G, Durr A, Bi X, Houlden H, Baudry M. Cell Rep. 2016 Jun 28;16(1):79-91

18. Genotype-phenotype correlations, dystonia and disease progression in spinocerebellar ataxia type 14. Chelban V, Wiethoff S, Fabian-Jessing BK, Haridy NA, Khan A, Efthymiou S, Becker EBE, O'Connor E, **Hersheson J**, Newland K, Hojland AT, Gregersen PA, Lindquist SG, Petersen MB, Nielsen JE, Nielsen M, Wood NW, Giunti P, Houlden H.

19. Charcot-Marie-Tooth type 4B2 demyelinating neuropathy in miniature Schnauzer dogs caused by a novel splicing SBF2 (MTMR13) genetic variant: a new spontaneous clinical model PeerJ accepted Nicolas Granger Alejandro Luján Feliu-Pascual Charlotte Spicer Sally Ricketts Rebekkah Hitti Oliver Forman **Joshua Hersheson** Henry Houlden

Figures

FIGURE 1-1 DEVELOPMENTS IN DNA SEQUENCING OVER THE PAST 30 YEARS..... **ERROR! BOOKMARK NOT DEFINED.**
FIGURE 1-2 SEQUENCING WORKFLOW FOR BRIDGE-AMPLIFICATION/SEQUENCING BY SYNTHESIS 25
FIGURE 1-3 APPROACH TO REFINING LIST OF VARIANTS GENERATED FROM WHOLE EXOME-SEQUENCING PROJECTS 28
FIGURE 1-4 ESTIMATE OF PROBABILITY OF IDENTIFYING A CAUSAL GENE FOR A MONOGENIC DISORDER 30
FIGURE 1-5 POTENTIAL MECHANISMS FOR NON-CODING RNA TOXICITY 45
FIGURE 1-6 SUMMARY OF THE 2013 CLASSIFICATION OF DYSTONIA 63
FIGURE 1-7 DYSTONIA AS A NETWORK DISORDER 64
FIGURE 3-1- PEDIGREE OF THE DYT4 FAMILY 83
FIGURE 3-2 - MULTIPOINT PARAMETRIC LINKAGE ANALYSIS OF DYT4 PEDIGREE 87
FIGURE 3-3- SEQUENCE CHROMATOGRAM SHOWING P.R2G MUTATION 89
FIGURE 3-4 -EXPRESSION OF TUBB4A IN VARIOUS HUMAN TISSUES 91
FIGURE 3-5 GRAPH OF THE EXPRESSION OF TUBB4A GENE IN 10 BRAIN REGIONS 92
FIGURE 3-6 AXIAL AND SAGITTAL MR IMAGES SHOWING TYPICAL FEATURES OF HABC 103
FIGURE 3-7 SEQUENCE CHROMATOGRAM SHOWING THE C.745G>A P.ASP249ASN DE NOVO VARIANT 104
FIGURE 4-1 PHOTOGRAPH OF A PARSON'S RUSSELL TERRIER 109
FIGURE 4-3 (A) PEDIGREE OF FAMILY 1. 114
FIGURE 4-2- GRAPHICAL REPRESENTATION OF HOMOZYGOSITY MAPPING ANALYSIS 114
FIGURE 4-4 .BAM FILE ALIGNMENT FROM EXOME SEQUENCING SHOWING THE HOMOZYGOUS CAPN1 VARIANT 116
FIGURE 4-5 3D RENDER OF STRUCTURAL MODELLING OF WT AND MUTANT CAPN1 PROTEIN SEQUENCE 117
FIGURE 4-6 WESTERN BLOTS SHOWING VIRTUALLY ABSENT CALPAIN1 EXPRESSION IN PATIENT. 119
FIGURE 4-7 A) PEDIGREE OF FAMILY 2 120
FIGURE 4-8 A) PEDIGREE OF FAMILY 3 121
FIGURE 4-9 EXOME ALIGNMENT SHOWING CAPN1 MUTATIONS 122
FIGURE 5-1 WORKFLOW SHOWING MISEQ LIBRARY PREPARATION 133
FIGURE 5-2 GRAPHICAL REPRESENTATION OF COVERAGE FROM THE SEQUENCING RUN 135
FIGURE 5-3 A) PEDIGREE SHOWING FAMILY OF SCA14 PATIENT 138
FIGURE 6-1 PEDIGREE OF FAMILIES A & B. 147
FIGURE 6-2 BAM ALIGNMENT OF EXOME SEQUENCING DATA 149
FIGURE 6-3 PROTEIN SEQUENCE ALIGNMENT OF MEMBERS OF PEPTIDASE A1 FAMILY 151
FIGURE 6-4 MULTI-SPECIES PROTEIN SEQUENCE ALIGNMENT OF CATD 151
FIGURE 6-5 HISTOPATHOLOGY OF PATIENTS A:3 AND B:3 153
FIGURE 6-6 ELECTRON MICROGRAPH OF CULTURED FIBROBLASTS OF PATIENT B:3 154
FIGURE 6-7 A) KINETIC ANALYSIS OF CATD ACTIVITY IN PATIENT A:3 AND B:5 154
FIGURE 6-8 IMMUNOBLOT OF TRANSIENTLY OVEREXPRESSED CATD WILDTYPE AND CLN10 ASSOCIATED VARIANTS. 160
FIGURE 7-1 A) PEDIGREE OF FAMILY 1 167
FIGURE 7-2 A) SEQUENCE CHROMATOGRAMS SHOWING HOMOZYGOUS Q6633X MUTATION IN PATIENT II 169
FIGURE 8-1 NEWSPAPER ARTICLES SHOWING HEADLINES RELATED TO 'JAMAICA GINGER' OUTBREAKS 173
FIGURE 8-2 PEDIGREE OF PATIENT A:1 178
FIGURE 8-3 A) PHOTOGRAPH OF PATIENT A:1 179
FIGURE 8-4 SEQUENCE CHROMATOGRAMS SHOWING THE C.G3241A HETEROZYGOUS MUTATION 181
FIGURE 8-5 MUSCLE HISTOPATHOLOGY FROM PATIENT A:1 183
FIGURE 8-6 EXTENDED PEDIGREE OF FAMILY OF INDEX CASE B:1 184
FIGURE 8-8 SAGITTAL AND CORONAL MRI BRAIN SHOWING SEVERE CEREBELLAR ATROPHY 185
FIGURE 8-9 HOMOZYGOSITY MAPPING OF B:1 185

Tables

TABLE 1-1 SUMMARY OF AUTOSOMAL DOMINANT CEREBELLAR ATAXIAS 53
TABLE 1-2 SUMMARY OF EPISODIC ATAXIAS 54
TABLE 1-3 SUMMARY OF AUTOSOMAL RECESSIVE ATAXIAS 55
TABLE 3-1- CLINICAL CHARACTERISTICS OF SELECTED AFFECTED FAMILY MEMBERS. 83
TABLE 3-2 - RESULTS FROM EXOME SEQUENCING OF PATIENTS VI-2 AND VII-6 88
TABLE 3-3 - MULTISPECIES PROTEIN SEQUENCE ALIGNMENT OF BETA TUBULIN 90
TABLE 3-4 - THE CONSERVATION OF THE PROTEIN SEQUENCES IN THE DIFFERENT TUBULIN ISOTYPES 90
TABLE 3-5 EXPRESSION OF TUBB4A IN 10 BRAIN REGIONS FROM 134 NORMAL INDIVIDUALS 92
TABLE 5-1 SUMMARY OF SEQUENCING RESULTS 136
TABLE 6-1 SUMMARY OF CLINICAL FEATURES FOR THE AFFECTED INDIVIDUALS IN FAMILY A & B 147
TABLE 6-2 LIST OF EXONIC VARIANTS WITHIN REGION OF SHARED HOMOZYGOSITY IN PATIENT A:5 149
TABLE 6-3 SUMMARY OF THE CLINICAL, GENETIC AND HISTOLOGICAL FINDINGS IN THE REPORTED CASES OF CLN10 158
TABLE 8-1 FILTERING STRATEGY USED FOR EXOME SEQUENCING ANALYSIS FOR PATIENT A:1 182

1 Introduction

1.1 Genetic basis of neurodegenerative disease

Neurodegenerative diseases are a complex and heterogeneous group of disorders characterised by progressive loss of cells from the central nervous system which can result in a broad range of clinical phenotypes including cognitive impairment, motor dysfunction, epilepsy and visual loss.

Globally, neurodegenerative disease are a major cause of morbidity and mortality in addition to the concomitant social and economic impact on society. Alzheimer's and Parkinson's disease are the two most prevalent neurodegenerative disorders and account for approximately 25% of the burden of all neurological disease globally and is estimated will affect greater than 150 million people by 2050 (Feigin *et al.*, 2020). In the UK alone the societal cost of these disease is currently estimated at least £30bn annually (Fineberg *et al.*, 2013).

The imperative to understand the aetiology and pathogenesis of these conditions is clear. By way of comparison, the cancer field has enjoyed an explosion of novel therapeutics and diagnostic tests as a result of the advances in cancer genomics and understanding of molecular pathogenesis over the last two decades. Identification and characterisation of the molecular aetiology underlying relatively common complex diseases such as sporadic Alzheimer's disease, Parkinson's disease and rarer Mendelian neurodegenerative diseases is essential for being able to establish a specific diagnosis and enabling the development of targeted therapeutic approaches. The past few years have seen the development of novel therapies, both licensed and under investigation, for hitherto untreatable and progressive, neurological disease. For example, effective gene therapies, which significantly mitigate and delay disease progression, now exist for spinal muscular atrophy, a progressive motor neuronopathy which is often fatal when presenting in childhood (Abreu and Waldrop, 2021). More recently adacanumab, a monoclonal antibody that can clear amyloid plaques from the brain, was the

first drug (albeit somewhat controversially) licensed for the treatment of Alzheimer's disease (Budd Haeberlein *et al.*, 2022).

Many of the mechanistic insights into neurodegeneration have only been possible with the identification of causative genes in families with Mendelian forms of neurodegenerative disease e.g. *APP* gene mutations in dominantly inherited Alzheimer's disease (Hardy and Selkoe, 2002)and *SNCA* gene mutations in PD families (Polymeropoulos *et al.*, 1997). Whilst monogenic causes for neurodegenerative disease are relatively rare, it is now well known that polymorphisms in genes which underly these disorders can contribute to the risk of causing sporadic disease in the general population e.g. *SNCA*, *LRRK2* and *GBA* polymorphisms in Parkinson's disease risk (Nalls *et al.*, 2019).

Identification of individuals with Mendelian forms of neurodegenerative disease may also facilitate the investigation into the efficacy of treatments for those diseases. One of the criticisms of some of the novel therapeutics for age-related diseases like Alzheimer's disease has been that they were only instituted in patients in a symptomatic phase of the illness (Frozza, Lourenco and de Felice, 2018) . Identification of pre-symptomatic individuals with pathogenic mutations in Alzheimer's disease associated genes may be offer the opportunity to modify the course of the disease by instituting treatment prior to the onset of disease before clinically significant neuronal loss has occurred.

Finally, and perhaps the most immediately, clinically relevant applications of our improved understanding of the genetic basis of neurodegenerative diseases are with regard to genetic counselling on inheritance risk, family planning e.g. pre-implantation genetic diagnosis, surveillance of at-risk individuals within the family and in certain cases better informing prognosis. In the UK, the NHS leads the world in offering whole genome sequencing as a part of routine clinical care via the pilot 100k genomes project and now instituted through the NHS Genomic Medicine Service (Smedley *et al.*, 2021) . Neurology was one the largest single

specialty contributing to the 100k genomes project which was overall able to establish a genetic diagnosis in 25% of families taking part.

The next section will examine the evolution and history of genetic techniques and technologies which led to the advent of high throughput next-generation sequencing, now the backbone of genetic research and clinical genetic diagnostics.

1.2 A brief history of genetic research

In 1865, Gregor Mendel – Augustinian friar, meteorologist, mathematician and otherwise polymath – published his seminal paper Versuche über Pflanzenhybriden" ("Experiments on Plant Hybridization") in which he demonstrates for the first time that heredity is transmitted in discrete units (Mendel, 1866) and provided mechanistic support for Darwin's proposals on natural selection espoused 6 years previously (Darwin, 1859). These discoveries were largely dismissed at the time, and it would not be until 1900 that Mendel's work would be independently rediscovered and shortly after in 1902 his hypotheses were proved to be true with the observation by William Sutton that the segregation of chromosomes during meiosis matched that of Mendel's Law of segregation of genes (Hegreness and Meselson, 2007).

Fast forward to 1950s and the discovery of the structure of DNA by Watson and Crick in 1953 and later in 1959, the first report of specific amino acid sequence alteration in human beta-globin protein as the cause of sickle cell disease (Ingram, 1956). It would not be until 1970s when the development occurred of gene cloning techniques in lambda bacteriophage vectors and most importantly rapid DNA sequencing methods by Sanger in 1975 (Casjens and Hendrix, 2015).

It was known by the late 1970s that there was considerable polymorphic variability throughout the human genome comprising either single nucleotide variants (SNVs) or short sequence repeats. This led to the development of a theoretical framework for linkage of a disease-

associated genomic locus with a polymorphic marker close enough that the likelihood of recombination between the marker and locus is minimal (Pulst, 1999). This theoretical technique was employed in 1983 in the first real world example with the identification of the previously unknown locus for Huntington's disease through linkage analysis of a set of polymorphic markers in two large kindreds in Venezuela, mapping the disease precisely to chromosome 4 (4p16.3) (Gusella, 1983). The *HTT* gene was finally sequenced using positional cloning i.e. cloning based on the position of polymorphic marker. The causative mutation – an expansion of a CAG trinucleotide repeat in *HTT* - was identified in 1993 (MacDonald *et al.*, 1993).

Linkage analysis, positional cloning and Sanger sequencing would be the mainstay of gene discovery for roughly the next 20 years until the completion of the first draft of the human genome sequence in 2001. Many of the genes responsible for the most common Mendelian diseases were cloned during this period (Claussnitzer *et al.*, 2020). The development of higher resolution maps of genetic variability to facilitate linkage and genome-wide association studies began with samples from the National Institute of Health/Centre D'etude Du Polymorphisme Humain (NIH/CEPH) consortium which was initiated in 1984 and the subsequent HapMap project which was initiated in 2002 and completed in 2010 yielded millions of polymorphic markers to generate a comprehensive linkage map of the human genome across a range of populations with ancestry from Africa, Asia and Europe (Claussnitzer *et al.*, 2020).

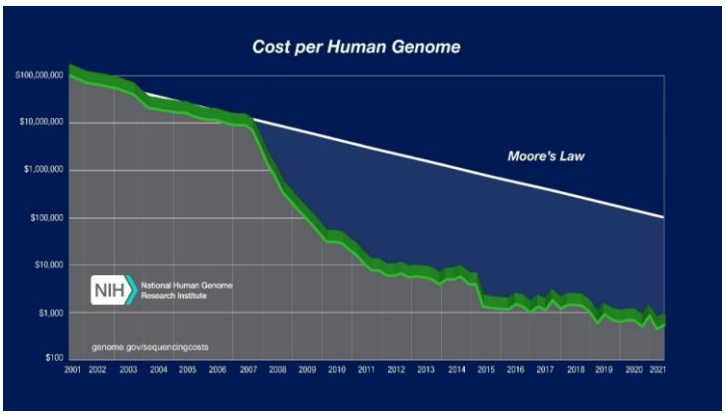


Figure 1-1 Graph showing fall in cost per genome between 2001-2021. The Y-axis is a logarithmic scale - from 2008 there was an exponential decrease in the cost of sequencing due to the transition between Sanger sequencing to NGS (Wetterstrand KA, 2021)

Deleted: *X*

Technologies such as high-resolution genotyping microarrays and high throughput DNA sequencing technologies converged in the early 2000s to hugely accelerate genotyping and gene discovery in Mendelian disease. By 2005, a range of technologies employing massively parallel DNA sequencing (next-generation sequencing/NGS) which were able sequence orders of magnitude more DNA per run compared to traditional Sanger sequencing techniques. More detail about the technical aspects of NGS is discussed in the next section. Further development in the technology has led to a precipitous fall in the cost of sequencing DNA and a concomitant increase in the speed. By way of illustration the first draft of the human genome took around 15 months to complete (Lander *et al.*, 2001), once significant genome sequencing was underway for the Human Genome Project, at an estimated cost of $300m. In 2021, the cost of sequencing at high accuracy, an entire human genome, cost between $600-$1000 and can by performed in less than two days (Wetterstrand KA, 2021).

The first successful application of a next generation sequencing technology in the discovery of a gene for a Mendelian disease of hitherto unknown cause came in 2009 with the identification using exome sequencing in *DHODH* as the cause of Miller syndrome, a rare

dominantly inherited disorder (Ng *et al.*, 2010). Exome sequencing, which examines the protein-coding portion of the genome and is discussed in more detail in the next section, represented the most tractable initial approach in the application of next generation sequencing technologies and has been responsible for a marked acceleration in the identification of genes causing Mendelian disease whilst revolutionising clinical genetic practice in establishing molecular diagnoses in patients with rare diseases.

Furthermore, where Mendelian diseases were previously characterised according to a canonical phenotype, next-generation sequencing led discoveries have identified significant phenotypic heterogeneity whereby mutations in a single gene can lead to both a spectrum of phenotypic effects or in many cases seemingly distinct clinical entities either through allelic or non-allelic (locus) heterogeneity (Bamshad, Nickerson and Chong, 2019). Such developments have led to a shift from phenotype-driven delineation of novel Mendelian diseases to genotype-driven diagnoses where the phenotypic spectrum is defined or expanded following identification of a causal gene.

1.3 Next generation sequencing

The development of next generation sequencing technologies represented a paradigm shift in how DNA and RNA are sequenced with respect to much costlier and labour-intensive Sanger sequencing. Depending on the application, NGS can be employed for DNA sequencing of the entire genome (whole genome sequencing/WGS), the whole exome sequencing/WES (sequencing only the protein coding portions of the genome), targeted gene sequencing (limited to specific set of genes or genomic regions), and also RNA specific applications (transciptomics/RNA-seq) or characterization of epigenetic changes such as methylation. In this thesis, I will only focus on DNA sequencing specific applications of NGS, targeted and whole exome sequencing.

Next-generation DNA sequencing encompasses a range of different technologies each of which differ in key technical aspects of sequence generation including read length, chemistries and sample preparation but share a basic conceptual workflow (Bamshad *et al.*, 2011).

In the first step, genomic DNA is fragmented into of small DNA fragments which range in size (depending on the platform) from ~100-250 base pairs (bp). These fragments may then be enriched for the desired target i.e. either the whole exome or specified genes/genomic regions. Whole genome sequencing does not require this enrichment step as all DNA is sequenced. Specific adaptors are added to each end of the fragments to create a library of short DNA fragments. Subsequent sequencing methods vary by sequencing platform but that can be hybridized typically to surface such as a microfluidic flow cell (e.g. Illumina) or to individual beads (e.g. Roche).

Deleted: which

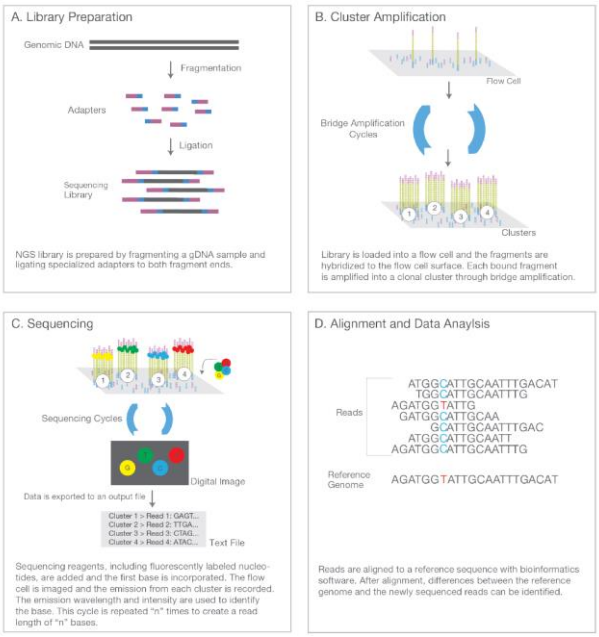


Figure 1-1 Sequencing workflow for bridge-amplification/sequencing by synthesis (reproduced from illumina.com)

Using the Illumina workflow as an example (see figure 1-1), the next step following hybridisation to the flow cell is the clonal amplification of single DNA fragments to form, over multiple cycles of amplification, a cluster of clonal DNA fragments comprising around 1000 clonal molecules. Around 50 x 10^6 clusters can be generated per flow cell, which are sequenced in parallel.

Deleted: 2

During sequencing, four different fluorescent dye terminators are incorporated with each nucleotide in the fragment. During each cycle of sequencing, the fluorophores are illuminated with laser excitation and an image of the flow-cell captured recording each successive nucleotide with every sequencing cycle. This technique is referred to as 'sequencing by synthesis'. This produced sequences of 75-100 bp from billions of clusters simultaneously.

After sequence image acquisition, the intensity data recorded are converted into nucleotide sequence reads and are then subjected to bioinformatic analysis.

1.4 Exome sequencing in Mendelian disease gene discovery.

Until the advent of next generation sequencing technologies, the majority of all Mendelian disease genes were identified using a combination of linkage or homozygosity mapping strategies, positional cloning and subsequent Sanger sequencing of candidate genes. Although expensive, laborious and time-consuming, these approaches led to the discovery of ~1200 Mendelian disease-associated genes. The majority (~90%) were mapped in pedigrees with a hypothesis of a dominant, co-dominant, recessive or X-linked pattern of inheritance.

Approximately 50% of Mendelian disease phenotypes are estimated to have been associated with a disease gene (Posey, 2019). Of these, it is estimated that 85% of disease-causing mutations are located within the coding regions of the genome or canonical splice sites. This figure is likely to be an overestimate due to bias imparted by the relative ease of identifying and interpreting the impact of coding nucleotide and thus protein sequence variation compared to other types of potentially pathogenic mutations e.g. exon, multiexon and gene deletions/duplications, complex chromosomal rearrangements, repeat variation, regulatory and other non-coding variations (Botstein and Risch, 2003).

Whilst the commonest mutation types are amenable to sequencing and analysis through a NGS approach, there are challenges (some experimental and relating to mode of sequencing and some relating to bioinformatic analysis) with identification of particular mutation types. These include chromosomal rearrangements, copy number variants and large repetitive DNA sequences e.g. repeat expansions which underly numerous important neurodegenerative disorders including Huntington's disease, several dominant cerebellar ataxias and more recently in a form of inherited frontotemporal dementia and motor neurone disease (hexanucleotide repeat in C9orf72),

Deleted: e.g.

Deleted: the trinucleotide and polynucleotide

Deleted: s

Deleted:

Despite these challenges, next generation sequencing of only the coding portion of the genome or 'exome' (which accounts for ~1% of the human genome) delivered a tractable means to identify disease-associated genes, at much lower cost and greater speed. Furthermore, this approach has been used to identify causes of disease in smaller families and singletons, where linkage analysis would not have been possible. Even when the entire genome is sequenced, a search for causal variants within the protein coding DNA sequences still represents the least challenging first line approach for searching for causal variants with most whole genome studies published, still focussing on the variants identified within the exome. (Gilissen *et al.*, 2012)

Deleted: was a

Deleted: much

Deleted: and even within individual probands

1.4.1 Identifying the signal of disease-causing mutations from the noise of human genetic variability

Starting with the hypothesis that disease-causing mutations *are* located within the exome and that it is technically possible to identify them using a NGS approach, it is a fundamental challenge to be able to identify the mutation/s responsible for a given disease phenotype, against the large number of benign polymorphisms found in an individual exome.

Depending on the ethnic background, human exomes may harbour between 20-24000 single nucleotide variants (SNVs) per individual (Bamshad *et al.*, 2011). Of these, approximately 50% would be expected to be non-synonymous i.e., would be expected to result in an amino acid sequence alteration and thus would more likely be associated with disease, although the majority will merely represent benign polymorphisms. Even taking into account the fact that variants most likely to be disease causing are novel or very rare missense mutations, nonsense mutations or canonical splice variants, the average human exome is estimated to have at least 200-300 variants that could potentially cause disease (Bamshad *et al.*, 2011)

The task of reducing a list of several thousand non-synonymous variants to identify the causal allele remains significant and I will discuss the different strategies that are employed to tackle

the problem. The choice of strategy needs to consider several factors including pedigree structure and mode of inheritance, prevalence of the disease or trait in the population, penetrance (i.e. the likelihood that a carrier of a pathogenic mutation will manifest disease) and also consider whether de novo variants may account for the observed phenotype.

Deleted: to

A schema of the overall approach is summarised in figure 1-3.

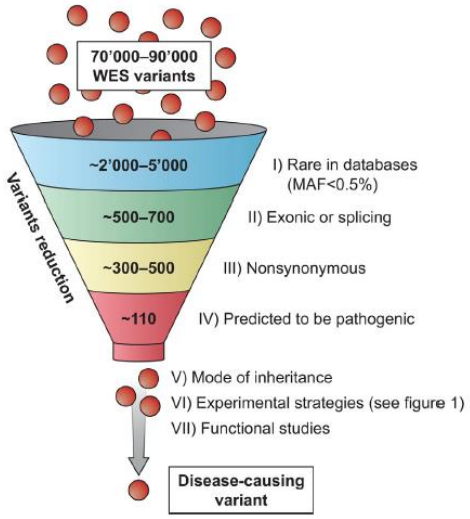


Figure 1-2 Approach to refining list of variants generated from whole exome-sequencing projects to identify the causal variant - reproduced from (Olgiati, Quadri and Bonifati, 2016)

1.4.2 Prioritisation of variants based on rarity

In view of the rarity of Mendelian disease, databases of population sequence variation allow for rare or novel variants to be distinguished from benign polymorphisms.

Over the past decade several databases have been published with the aim of documenting the range of genetic variation. These databases have, as sequencing technology has evolved become increasingly accurate and comprehensive with the integration of data from large scale sequencing projects.

At the time my research was undertaken, the databases that were available were dbSNP version 137 (https://www.ncbi.nlm.nih.gov/snp/), the Exome Variant Server (https://evs.gs.washington.edu/EVS/), 1000 genome project (https://www.internationalgenome.org/), Complete Genomic 69 database (https://www.completegenomics.com/public-data/69-genomes/) and an in-house database of exome sequencing data which is useful at filtering out systematic sequencing error and technical artifacts. Subsequently, and of relevance to current research and clinical diagnostics, the most comprehensive database in use as of 2022 was the gnomAD v2.1/3.1 Genomic Aggregation Databases (https://gnomad.broadinstitute.org/), comprising variants identified in >75000 whole genome and >125000 whole exome datasets across a much more diverse range of ethnicities and genetic ancestries than previous resources.

Deleted: and employed in my filtering strategies

An estimate of the population prevalence for specific Mendelian disorders is important, which informs the cut-off values for the mean allele frequency (MAF) of variants with regard to the plausibility of the variant being pathogenic for the trait in question. For rare dominant Mendelian diseases, a cut-off MAF of <0.001 is generally adopted as the prevalence of these disorders, is usually below 0.1% in the population (Koboldt, 2020) . In practice the first analysis of sequencing results would be to consider novel variants i.e. those not present in published databases of normal sequence variation. For recessive diseases, a MAF <0.01 is considered to have sufficient power to identify pathogenic alleles. Figure 3-4 demonstrates the probability of identifying a pathogenic allele in dominant and recessive disorders using such filtering models.

One of the issues with earlier releases of sequence variation databases was the limited ethnic ancestry of the populations submitted to the database. Variants absent from databases may be present in healthy individuals from a different population not captured in the dataset and therefore are much more likely to be considered benign polymorphisms. This is particularly relevant in recessive disorders which disproportionately affect populations with high levels of

parental consanguinity. With earlier releases of population sequence variant databases, ethnogenetic diversity in the sampled populations was limited. This presents issues both in gene discovery and perhaps more importantly in clinical genetic diagnosis/variant classification which relies, amongst other criteria, on determining likely pathogenicity with reference to the expected allele frequency in a particular population (Ellard *et al.*, 2020). In this situation a lack of population specific variation data may lead to variants erroneously being classified as pathogenic.

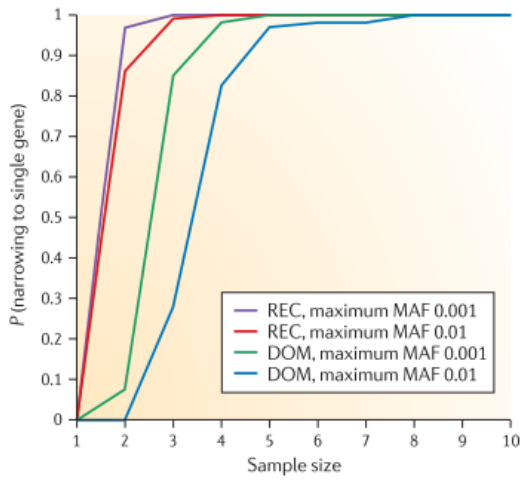


Figure 1-3 Estimate of the probability of identifying a causal gene for a monogenic disorder - reproduced from Bamshad et al., 2011

1.4.3 Filtering based on mode of inheritance and intersection analysis

In families with multiple affected individuals, the mode of inheritance is more easily defined. For dominantly inherited conditions, only heterozygous mutations are considered although in families with a high degree of consanguinity, it is possible for a pseudo-dominant pattern of inheritance to be observed for a recessive disorder.

Where the mode of inheritance is recessive, only homozygous and compound heterozygous mutations are considered although homozygous variants would be prioritised if there is known consanguinity. In analysis of results from a single proband, it is not possible to assess the phase of compound heterozygous variants. This requires analysis of parental samples or informative relatives to confirm whether the mutations are in 'cis' i.e. present on the same allele or in 'trans' i.e. present on differing alleles and therefore potentially relevant.

In sporadic cases, both dominant and recessive inheritance should be considered and where possible, de novo heterozygous mutations assessed with a trio analysis of both (unaffected) parents. Apparently sporadic cases can on occasion be explained by somatic mosaicism of the pathogenic mutation in an unaffected parent resulting in apparently de novo or a pseudo-recessive pattern of inheritance i.e., more than one affected child from (apparently) unaffected parents.

Where data are available from more than one affected individual, it is possible to significantly reduce the genomic search space and variant list by examining shared variants. The most informative results arise from assessing distantly related relatives. Based on calculations of kinship coefficients (THOMPSON, 1975), a measure of relatedness, 1st cousins, for example, share approximately 1/8 of their genome that is identical by descent and so restricting an analysis to only shared variants can be helpful in refining the candidate variant list. In addition, excluding variants in non-affected individuals can also reduce the list of private variants. This approach can also be used with multiple affected individuals from different families although in this application, the search is for a shared gene harbouring one or more potential pathogenic variants rather than a specific shared variant.

1.4.4 Use of mapping strategies

As was the case with the original paradigms of gene discovery, the use of linkage analysis in dominant and homozygosity mapping in recessive pedigrees has been a useful tool in gene discovery.

1.4.4.1 Linkage analysis:

Mendel's Law of independent assortment of traits held true as Mendel had selected traits that were not localised closely together on the same chromosome and therefore were observed to segregate independently from each other (Pulst, 1999). However, traits that are co-localised or linked may be inherited together and transmitted as a block of DNA known as a haplotype.

Genetic markers are said to be linked to a locus if they are inherited more often than would be expected than if they were independent of each other. The closer that the marker is to the locus of interest, the less likely it is that recombination will occur during meiosis. Where the likelihood of recombination is 1%, the markers are defined as being 1cM apart which is not a true distance but roughly equates to around 1Mb in the human genome (Pulst, 1999).

Historically, multiallelic short tandem repeats (STRs) comprising di, tri or tetranucleotide repeats were used as DNA markers as they were easily assessed using polymerase chain reaction (PCR) and gel electrophoresis to resolve the size of the marker (Fan and Chu, 2007). Currently, genotyping arrays which can determine the genotype of up to millions of genome-wide SNVs are one of the commonest means to generate the markers that can be used in linkage studies although techniques have been developed to use genotype data from exome or whole genome projects (Smith *et al.*, 2011).

Linkage studies generally rely on the calculation of logarithm of odds (LOD) score which is a statistical test comparing the likelihood that two loci are linked compared to whether the observed results could be obtained by chance. Conventionally, a LOD score > 3 is accepted as

evidence for linkage and equates to a probability that the loci are linked of 1000:1. A LOD score of -2 indicates a probability of 100:1 that two loci are not linked (Pericak-Vance, 2001). For Mendelian disease models, a parametric analysis is employed which requires specification of the mode of inheritance, estimated disease allele frequency and penetrance of the allele.

Linkage analyses rely on an accurate designation of an individual's affectation status and pedigree structure. Where this data is inaccurate, linkage analysis may fail. Other issues that can arise that can impact the results, relevant to Mendelian disease, are with incomplete penetrance, phenotypic and locus heterogeneity and phenocopies.

Limiting the variant list against those located within a linked genomic region is a powerful method used to identify variants and genes associated with a Mendelian trait and provides statistical evidence to support the pathogenicity of any putative variants or genes (Ott, Wang and Leal, 2015).

1.4.4.2 Homozygosity mapping

Developed as a technique to map recessive traits in children from consanguineous parents, homozygosity mapping allows identification of a disease locus since adjacent chromosomal regions are also likely to be homozygous i.e. homozygous by descent (Lander and Botstein, 1987). The assumption that the pathogenic, homozygous variant lies within a large stretch of homozygosity, allows variants and genes to be prioritised within these regions. Whilst this can be an effective strategy in a single affected proband, identifying shared regions of homozygosity in affected relatives, can further limit the genomic region of interest.

As has been true for linkage analysis, high resolution SNP arrays have enabled the use of homozygosity mapping as a standard tool in gene discovery for families with known parental consanguinity but also in individuals without known consanguinity but who are from isolated communities presumed to share a gene pool from common ancestors.

Tracts of homozygosity tend to be quite large however these data can still be extremely helpful at narrowing the region of interest when filtering variant lists in exome and whole genome sequencing projects and have been used successfully in the identification of many recessive Mendelian disease genes.

1.4.5 Prioritisation based on deleteriousness

Mutations that result in a truncated protein sequence e.g. frameshift, stop codon or canonical splice site variants are generally given greater weight with regards to the likelihood of pathogenicity. Missense variants are more challenging to interpret, and a range of tools have been developed to predict whether the observed variant is likely to be deleterious. In some cases, an understanding of the protein sequence and the location of important functional domains can be helpful in assessing the possible impact of a variant. This however requires there to be some understanding of the role and function of the gene/gene product in question in order to draw these inferences.

Evolutionary conservation of a given nucleotide site can be an important indication of potential pathogenicity as sites of deleterious mutations may show a relative absence of variation across a range of different species due to purifying selection. Various tools have been developed such as GERP (http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html) and PhyloP (Pollard *et al.*, 2010) which calculate the degree of conservation at each nucleotide site. Similarly, evolutionary conservation can be assessed using multi-species protein sequence alignment.

Other tools such as SIFT (Ng and Henikoff, 2001), MutationTaster (Schwarz *et al.*, 2010) and Polyphen2 (Adzhubei *et al.*, 2010) can make predictions based on the impact to protein sequences. These tools are reasonably sensitive (~70%) but have low specificity in predicting pathogenic variants and are generally more reliable in assessing *loss* rather than *gain* of function (Flanagan, Patch and Ellard, 2010).

1.4.6 Limitations of next generation sequencing

These approaches have been remarkably successful to date in the identification of Mendelian disease genes with estimates of potential success rates in identifying causal variant being between 60-80% (Gilissen *et al.*, 2012).

Despite this, several limitations, technical and analytical must be considered, both when designing studies and when this approach fails to identify a causal variant or gene.

1. Exome sequencing studies rely on the hypothesis that the causal variant is located within the coding region of the genome. Whilst, as discussed, this is estimated to be true in 85% of cases, the definition of what constitutes the exome is important. Uncertainty remains over which sequences in the genome code for protein sequences. Early versions of exome sequencing capture kits used the Consensus Coding Sequence Project (CCDS) definition of the exome with a target capture of ~31.1Mb (Pruitt *et al.*, 2009). This is a highly curated but relatively limited set of genes compared to the RefSeq database of genes which targets 67Mb of sequence (O'Leary *et al.*, 2016). Targeting CCDS sequences alone would miss several genes present within the more expansive RefSeq database (Steward *et al.*, 2017). This is not an issue with whole genome sequencing which does not require enrichment of genomic regions of interest.

2. Even though whole genome studies do not rely on enrichment of sequences of interest, the definition of the human genome is crucial to the process of aligning read sequences. In 2009, the GRCh37/hg19 build of the human genome was released by the Genome Reference Consortium and was the genome build on which the majority of the research in this thesis was based. This genome build was updated in 2013 to the GRCh38/hg38 build which was a more complete representation of the genome however significant gaps remained particularly with respect to coverage for genomic regions that were technically difficult to sequence e.g. sequence within

heterochromatic DNA. In March 2022, the T2T-CHM13 genome assembly was released which addressed the remained 8% of unsequenced genome (Nurk et al., 2022). This added an additional 1956 gene predictions, 99 of them which were protein coding. This now near complete human genome build is the current reference sequence used in human genomic research studies and hopefully with facilitate further understanding into disease pathogenesis now that previously 'hidden' areas of the genome have now been revealed.

3. Regions with high GC content may result in failed enrichment or a sequencing failure. If the variant is in a region which is poorly covered or is sequenced at low depth, then the variant may not be called accurately or at all.

4. Repeat expansions are an important class of mutations that are not easy to sequence or analyse using the relatively short reads obtained with the majority of 2nd generation NGS platforms. This is a relevant issue for both exome and whole genome studies. Expansions in both coding and non-coding portions of the genome are associated with a significant number of important neurodegenerative diseases. Where expansions are small, the expanded allele may be sequenced using short-read technology however many repeat expansions disorders are associated with >100 repeats (Bahlo *et al.*, 2018) which are beyond the limit of most short reads and therefore would not be able to be mapped or aligned to the genome in the absence of non-expanded flanking sequence. An example of this as an issue in practice relates to the eventual identification of a hexanucleotide repeat expansion as the cause of C9orf72-associated frontotemporal dementia (Renton *et al.*, 2011). In this case there was very strong evidence for the genetic locus from large genome-wide studies, yet whole genome sequencing had failed to identify the causative mutation. Analysis of mismatched sequencing reads from WGS data gave the initial clue as to the nature of the mutation

Deleted: was with the

which was subsequently confirmed using repeat-primed PCR which is a well-established method for identifying a polynucleotide repeat. 3rd generation NGS technology e.g. PacBio or Nanopore sequencing, which allow for read lengths in the 1000bp range rather than several hundred base pairs, may offer the means to address this issue in future (Depienne and Mandel, 2021).

5. Other mutation types such as copy number variants and structural rearrangements are also technically difficult to identify from next generation sequencing data, particularly exome data which has more limited genomic representation. Novel bioinformatic tools have been developed to address this however this remains a challenge. Whilst for individual patients and pedigrees, CNVs and structural rearrangements may be the causal mutation, they are overall a less common cause of Mendelian disease. Nonetheless it remains important to consider these mutation types in affected individuals for whom a causal variant is not identified.

6. Perhaps the most common reason for failing to identify a gene is due to studies undertaken in less informative families with few affected individuals or where the disease is particularly rare or phenotypically heterogeneous. It is not uncommon to be left with a list of candidate variants in genes that cannot be reduced further through intersection or prioritisation strategies. Thus, although the variant may have been sequenced and called in the data, it may not be possible to resolve the signal of pathogenicity from the noise of irrelevant variants.

Deleted: the

Deleted: traditional

Deleted: s

Deleted: of

Deleted: Although this was located

Deleted: s

Deleted: determine

1.5 Genetics of inherited cerebellar ataxia

1.5.1 Introduction

The inherited cerebellar ataxias are a group of neurodegenerative disorders in which the dominant feature is progressive cerebellar degeneration resulting in impairment of balance, gait, coordinated limb movements and speech. Ataxia may present as an isolated cerebellar syndrome or more often is associated with a broad spectrum of neurological manifestations including pyramidal, extrapyramidal, sensory and cognitive dysfunction. Given the significant clinical heterogeneity of these disorders and the complexity of the cerebellum and its associated connections it is unsurprising that there exists significant genetic heterogeneity.

The prevalence of genetic forms of ataxia has been difficult to accurately determine and several reviewers suggest that previous data have underestimated the extent of the problem (Klockgether, 2011). Worldwide prevalence of autosomal dominant ataxia is estimated to be 2.7/100000 and autosomal recessive ataxia 3.3/100000 (Ruano *et al.*, 2014).

Inheritance patterns of these disorders can be complex with autosomal dominant, recessive, X-linked and mitochondrial inheritance demonstrated by one or more ataxic syndromes. The broad range of mutation types found in the inherited ataxias contribute to the complex genetic aetiology and pathogenesis of these disorders.

Many of the more dominant ataxias are caused by a range of polynucleotide repeat expansions including tri, penta and hexanucleotide repeats however point mutations and deletions and duplications are also represented. Dominant ataxia loci are historically designated with the SCA (spinocerebellar ataxia) prefix. The precise number of loci is open to interpretation as the list of approved SCA designations has been 'polluted' with a variety of other related syndromes (SCA18 – predominantly sensory ataxia; SCA29 congenital non-progressive ataxia which shares the locus 3p26 with SCA15/16), allelic disorders (SCA15/16) and a SCA designation without a reported locus (SCA9). Taking these discrepancies aside there are a total of 49 separate SCA

loci with the associated genetic variant identified in 39 of these. Additionally, there are many other dominantly inherited neurological syndromes in which ataxia is a prominent feature including Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), Alexander disease and Gerstmann-Straussler-Scheinker disease (GSS).

The recessive ataxias are even more diverse with nearly 100 genes having been identified (Washington Neuromuscular Website - www.neuromuscular.wustl.edu/ataxia/recatax) and many of which are exceedingly rare, the commonest inherited ataxia worldwide is caused by recessively inherited mutations in the frataxin (*FXN*) gene resulting in Friedrich ataxia (FRDA).

1.5.2 The cerebellum

The cerebellum is a key structure in the hindbrain of all vertebrates located in the posterior cranial fossa. Whilst modulation of motor control is the cerebellum's most important function and most of the efferent connections are to the motor system, the cerebellum supports several important, albeit less well understood, non-motor and cognitive functions. The cerebellum occupies less than 10% of the total brain volume yet accounts for over 50% of the total number of neurons in the brain (Andersen *et al.*, 1992).

Anatomically, the cerebellum is divided into two hemispheres connected by a structure called the vermis. The cerebellum is subdivided antero-posteriorly three lobes – the anterior lobe, the posterior lobe and flocculonodular lobe (Roostaei *et al.*, 2014). The medial portion of the anterior and posterior lobes receive proprioceptive inputs from the spinocerebellar tract and visual pathways and serve primarily to modulate motor output in order to achieve coordinated and accurate limb movements and adaptive motor coordination. The flocculonodular lobe receives primary inputs from the vestibular system in addition to visual and other relevant sensory systems and is concerned primarily with balance and gait and postural maintenance. The lateral hemispheres are the largest functional subdivision of the cerebellum have extensive connections to the cerebral cortex via the pontine nuclei and thalamus. These

Deleted: and is

Deleted: as well as

structures are important in the planning and timing of movements but also many of the non-motor functions of the cerebellum (Roostaei *et al.*, 2014).

Histologically the cerebellar cortex is divided into 3 layers. The innermost layer is comprised of densely packed granule cells which account for around half of all cerebellar neurons. These project to the Purkinje cells in the Purkinje layer. Purkinje cells exhibit striking dendritic morphology and are arranged parallel to each other in a flattened plane. Purkinje cells receive more synaptic input than any other cell type in the brain (Hirano, 2018). Purkinje cells are GABA-ergic and exert an inhibitory effect on the deep cerebellar nuclei which is their only output.

1.5.3 Classification of hereditary ataxias

The pathological hallmark of cerebellar ataxia was first described in 1901 by Dejerine who introduced the term olivopontinecerebellar atrophy (OPCA) to describe the pathological findings in a case of progressive cerebellar ataxia. In 1954, Greenfield made the first attempt to classify the hereditary ataxias based on pathological criteria (Greenfield, 1954): 1) predominantly spinal forms (Friedreich's ataxia and hereditary spastic ataxia); 2) spino-cerebellar forms 3) predominantly cerebellar forms. This classification was updated by Anita Harding in 1981 based on a clinico-genetic framework to include the following categories (Harding, 1981a):

1. Congenital disorders of unknown aetiology
2. Ataxic disorders with known metabolic cause or due to defective DNA repair
3. Ataxic disorders of unknown aetiology – this was subdivided into early onset (<30 years) and includes most recessive cerebellar ataxias and late onset (>30 years) comprising the dominant cerebellar ataxias.

Dominant cerebellar ataxias were further divided into 6 categories:

1. ADCA type 1: autosomal dominant cerebellar ataxia with optic atrophy, ophthalmoplegia, dementia/extrapyramidal features and amyotrophy.
2. ADCA type 2: autosomal dominant cerebellar ataxia with pigmentary retinal degeneration and/or ophthalmoplegia and/or extrapyramidal features. This was subsequently attributed only to SCA7.
3. ADCA type 3: pure autosomal dominant cerebellar ataxia.
4. ADCA type 4: autosomal dominant cerebellar ataxia with myoclonus and deafness
5. Periodic autosomal cerebellar ataxia (now generally known as episodic ataxia)
6. Idiopathic late onset cerebellar ataxia (ILOCA)

Whilst this classification was certainly more useful clinically than previous pathological frameworks, due to the wealth of new genes identified with their associated phenotypic heterogeneity, a new framework for dominantly inherited ataxia was suggested by the International Parkinson's and Movement disorders task force (Marras *et al.*, 2016). This divides cases based on their being either being a pure or relatively pure ataxia or a complex ataxia i.e. having other non-cerebellar neurological features. For known genes, SCA prefix is appended to the gene associated with the disorder e.g. SCA1 is now designated SCA-ATXN1. For the purposes of this thesis, the older SCA designation will be used.

1.5.4 Autosomal dominant cerebellar ataxias (ADCAs)

With respect to the currently known genes associated with ADCA, the majority of dominantly inherited ataxia cases are caused by repeat expansions in either coding or non-coding parts of the relevant genes (Dueñas, Goold and Giunti, 2006). However, it should be noted that most

Deleted: s

newly discovered genes associated with ADCA, are caused by conventional rather than repeat expansion mutations although in general, of the most recently discovered genes, the proportion of cases that have been explained by these genes remains relatively small. It may transpire however that yet-to-be-discovered novel complex expansions, such as those which were identified in C9orf72-associated frontotemporal dementia or the more recently identified biallelic pentanucleotide repeat expansions in RFC1 underlying the cerebellar ataxia, sensory neuropathy, and vestibular nerve palsy syndrome (CANVAS) (and which may account for a relatively high proportion of late-onset sporadic ataxia cases), could potentially explain a significant proportion of genetically unexplained dominantly inherited ataxia.

Of the ataxia-associated repeat expansions, polyglutamine (CAG_n) expansions are the most common and underly the pathogenesis in SCAs 1, 2, 3, 6, 7, 17 and dentatorubral-pallidoluysian atrophy (DRPLA). Genotype/phenotype correlations of these disorders are well described (Schöls *et al.*, 2004) with the disease manifesting above a threshold of CAG repeats. The non-coding expansion SCAs comprise SCA8 (CTG_n), SCA10 ($ATTCT_n$), SCA12 (CAG_n), SCA31 ($TGGAA_n$) and SCA36 ($GGCCTG_n$).

Anticipation is a feature of repeat expansion disorders whereby an inverse correlation exists between the size of the repeat length and the age at onset and clinical severity with increasing repeat lengths seen as the disease passes through subsequent generations. Increasing repeat expansions are most frequently observed in paternal transmission due to in part to marked repeat expansion stability in sperm (Bates and Lehrach, 1994).

1.5.4.1 ADCA caused by trinucleotide expansions:

SCA1 is caused by a CAG repeat in *ATXN1* and was the first gene to be associated with dominantly inherited ataxia (Orr et al., 1993). The function of the gene is not known but may be involved in transcriptional regulation and RNA splicing (Kang and Hong, 2009). The

incidence varies depending on the population and is particularly common in South Africa. As with all the repeat expansion SCAs, the clinical course is variable depending on the repeat length. The normal repeat length range is 6-34 repeats. Where the repeat length is >45, a fully penetrant ataxic phenotype is observed. The age at onset varies between 4-74 years but most typically presents in the 4th decade. In addition to ataxia, ophthalmoparesis, lower limb spasticity and neuropathy are relatively common. Extrapyramidal features such as dystonia or chorea may be seen later in the disease course.

Deleted: with complete penetrance of the disease at with a repeat length >45.

SCA2 is caused by a CAG expansion in *ATXN2* (Magaña, Velázquez-Pérez and Cisneros, 2013). This is ubiquitously expressed although high expression levels are seen in cerebellar Purkinje cells. The protein has a role in RNA metabolism and translation. SCA2 is the second most common SCA accounting for an estimated 10% of dominantly inherited ataxia. There is clinical overlap with SCA1 & SCA3 however polyneuropathy is more common. Expanded alleles (>34 repeats) are also known to be a risk factor for amyotrophic lateral sclerosis (ALS). Patients may have significant extrapyramidal symptoms such as parkinsonism which can be responsive to L-Dopa.

SCA3 is caused by a repeat expansion in *ATXN3* (McLoughlin, Moore and Paulson, 2020). This is the most common ADCA worldwide although is particularly prevalent in Brazil where is accounts for 89% of cases (Ruano *et al.*, 2014). *ATXN3* functions as a deubiquitinase and transcriptional regulator. Repeats of >51 are associated with disease and very large repeats may be associated with dementia, an earlier onset and more rapid progression. Ophthalmoparesis, polyneuropathy and spasticity are common.

SCA6 is caused by a CAG repeat expansion in *CACNA1A* which encodes a voltage dependent calcium channel (Zhuchenko *et al.*, 1997). It generally manifests as a slowly progressive, pure cerebellar ataxia when repeat size >20. It is allelic with episodic ataxia type 2 and familial

hemiplegic migraine however these are caused by missense mutations in the *CACNA1A* gene (Mantuano *et al.*, 2003).

SCA7 is caused by a CAG expansion in *ATXN7* which encodes the protein ataxin-7 which is thought to have a role in transcriptional regulation (Garden and La Spada, 2008). SCA7 is unusual compared to the other trinucleotide repeat expansion associated SCAs as it is associated with a cone-rod dystrophy retinal degeneration phenotype. As a result, it was initially distinguished from the other known ADCAs by Harding as ADCA2. *ATXN7* exhibits significant instability in the repeat length. Infantile onset cases are typically associated with large repeats (>200 repeats). Retinal degeneration, usually resulting in visual loss, may be the first clinical feature manifesting where the repeat length is >59 repeat. A supranuclear ophthalmoplegia is common in ~70% of cases in addition to pyramidal features.

1.5.4.2 ADCA caused by non-coding repeat expansions

The mechanism for neuronal toxicity in non-coding repeat expansion cerebellar ataxias and other neurogenerative disorders including myotonic dystrophy type 1 and 2 and *C9orf72*-associated frontotemporal dementia/amyotrophic lateral sclerosis (FTD/ALS) is complex and in many instances not well understood or fully defined.

Three potential mechanisms have been suggested (Swinnen, Robberecht and Van Den Bosch, 2020) (See figure 1-5): 1) Expanded RNA repeats may impair normal function of RNA by interacting with RNA-binding proteins to impair their normal function, usually referred to as 'RNA toxicity'. 2) RNA may in certain circumstances be translated via a mechanism not independent of the ATG start signal to repetitive amino acid sequences that depend on the repeat motif. These proteins are referred to as RAN protein products and may be directly toxic and form protein aggregates. This mechanism is referred to as RAN-toxicity. 3) Finally, the large expansions may result in loss of function through a variety of potential mechanisms including impaired transcription or increased mRNA degradation.

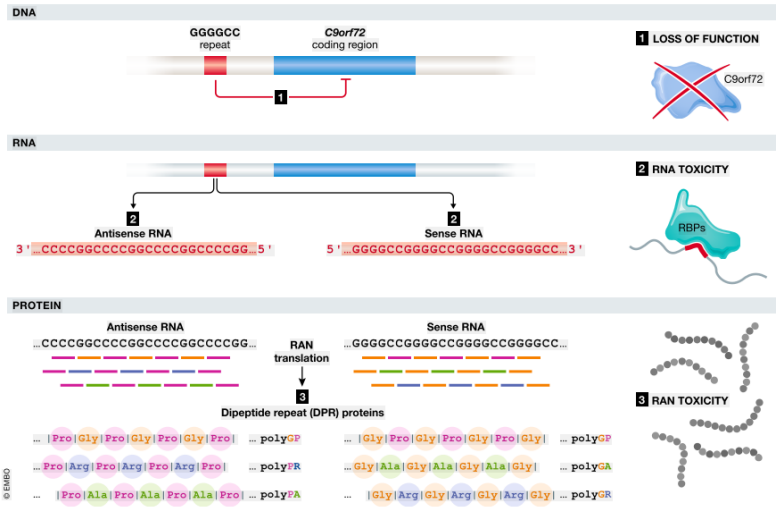


Figure 1-4 Potential mechanisms for non-coding RNA toxicity - reproduced from (Swinnen, Robberecht and Van Den Bosch, 2020)

The non-coding expansions SCAs comprise SCA8 (CTG_n), SCA10 ($ATTCT_n$), SCA12 (CAG_n), SCA31 ($TGGAA_n$), SCA36 ($GGCCTG_n$) and SCA37 ($ATTTC_n$). Anticipation has been demonstrated to a varying degree by most of the non-coding repeat expansion SCAs other than SCA36. Repeat expansions may be located in either promoter regions, the 5' or 3' untranslated regions (UTRs)

or introns. The repeat units are variable (tri, penta, hexanucleotide repeat motifs) and very large repeat expansions can be seen e.g. 800-4500 repeats seen in SCA10.

SCA 8 is caused by a CTG repeat in the 3' UTR of *ATXN8* which is thought to confer toxicity through a similar mechanism to that seen in myotonic dystrophy type 1 & 2, also caused by a CTG_n expansion. Whether this is the only or predominant mechanism is uncertain as ATG-dependent translation of the alternate strand (*ATXN8OS*) generates a toxic polyglutamine sequence seen in the coding polyglutamine repeat disorders. In addition to cerebellar and pyramidal features, cognitive impairment is seen in around 50% of individuals.

SCA10 is caused by an $ATTCT_n$ expansion in *ATXN10*. Initially identified in Mexican families, SCA10 is relatively uncommon outside of Latin American populations with most affected cases believed to share a common ancestral haplotype (Matsuura *et al.*, 2000). SCA10 is associated with seizures in addition to a range of other complex phenotypes including polyneuropathy, pyramidal signs and neuropsychiatric impairment. Very large repeat expansions are observed with the disease associated repeat range being between 800-4500 repeats. A loss of function mechanism is not believed to be the pathogenic mechanism as heterozygous knockout mice do not exhibit any abnormal phenotype (Wakamiya *et al.*, 2006). RNA foci have been observed in patient derived cells which co-located with the RNA binding protein hnRPK (Heterogeneous Nuclear Ribonucleoprotein K) which may be important in the pathogenicity of the pentanucleotide expansion in SCA10 (White *et al.*, 2012).

SCA12 is caused by a trinucleotide CAG_n repeat in *PPP2R2B* which encodes Protein phosphatase 2, regulatory subunit B, a brain specific subunit of protein phosphatase 2 whose precise function remains unknown. SCA12 was first identified in a single German family (Holmes *et al.*, 1999) however is relatively uncommon outside of India where it is estimated to account for 7-16% of ADCA (Bahl *et al.*, 2005). The mechanism for toxicity has not been fully evaluated however it is thought unlikely that SCA12 is a polyglutamine disorder due to the

absence of an open reading frame in repeat-containing transcripts and also the absence of ubiquitin-staining aggregates typically seen in polyglutamine disorders.

SCA31 is caused by a pentanucleotide $TGGAA_n$ repeat expansion in an intron shared between two genes - *BEAN1*, which encodes the protein Brain-Expressed, Associated with NEDD4 and *TK2* encoding thymidine kinase 2 (Sato *et al.*, 2009). SCA31 is relatively common in Japan (Sakai *et al.*, 2010), where it is third most common ADCA after SCA3 and SCA6, and extremely rare outside of surrounding populations. It typically presents as a late-onset pure cerebellar ataxia (ADCA3) with a mean age of onset between 56-61 years.

SCA36 is caused by a hexanucleotide $GGCCTG_n$ expansion in *NOP56*. The pathogenic expansions usually comprise >650 repeats however some affected cases have relatively low repeat lengths (14-30 repeats). SCA36 was originally identified in Japanese kindreds from the Chugoku region and was subsequently shown to be the cause of 'Costa de Morte' ataxia prevalent in Galicia, Spain (Kobayashi *et al.*, 2011; García-Murias *et al.*, 2012). The clinical features of the Japanese cases were similar with the majority exhibiting motor neuron degeneration manifesting with tongue fasciculation and atrophy. The age at onset is usually between 40-50 years with patients with larger expansions generally having a younger age at onset and more severe phenotype. In the Spanish SCA36 cohort, sensorineural deafness was a common feature (García-Murias *et al.*, 2012). The lower motor neurone features in both cohorts progressed relatively slowly and did not typically result in significant functional impairment with dysphagia, unlike typical amyotrophic lateral sclerosis cases. Genetic anticipation has not been conclusively demonstrated as the large repeat expansions can show significant somatic instability and consequent variability in the repeat length between generations (Lopez and He, 2022). Technical issues with accurately sizing very large repeats also have impacted the reliability of any conclusions about genetic anticipation in SCA36. In

SC36 post-mortem brain tissue, RNA foci were frequently seen supporting RNA gain of function as the pathogenic mechanism underlying SCA36 (Kobayashi *et al.*, 2011).

SCA37 is the 6th and most recent ADCA shown to be caused by an intronic repeat expansion. A pentanucleotide $ATTTC_n$ repeat in the reelin adaptor protein DAB1 was identified in Spanish and Portuguese families (Seixas *et al.*, 2017; Corral-Juan *et al.*, 2018). SCA37 is a slowly progressive pure ataxia with a mean age at onset of 48 and inverse correlation between repeat expansion size and age at onset. The pathogenic range for the expansion is between 31-75 repeats. Abnormal RNA foci in a SCA37 zebrafish embryo model were observed in addition to dysregulated DAB1 expression and abnormal upregulation of the reelin-DAB signalling pathway in the SCA37 cerebellum.

1.5.4.3 ADCA due to conventional mutations

23 dominant ataxia syndromes (SCA5, 11, 13, 14, 15, 19/22, 20, 21, 23, 25, 26, 27, 28, 34, 35, 38, 40, 41, 42, 43, 45, 46 and 47) are caused by conventional mutations. In a French ataxia series, conventional mutations accounted for 6% of all dominant ataxia, repeat expansions accounted for 45% with the remaining 48% genetically undiagnosed (Durr et al 2007). Genotype/phenotype correlations are much harder to determine in this group owing to the limited number of families affected by these mutations. Functional analysis of potassium channels in SCA6, SCA13 and EA2 has demonstrated a correlation between the degree of functional impairment and the severity of the phenotype. In contrast to the repeat-expansion SCAs, these disorders often have a 'purer' cerebellar phenotype (ADCA3) with a slower rate of progression.

SCA5 is caused by a mutation in the *SPTBN2* gene which encodes B3 spectrin (Ikeda *et al.*, 2006). Missense and in-frame deletions have been described resulting in a pure cerebellar syndrome with onset between 15-50 years. The first SCA5 kindred was reported in 1994 with

56 affected individuals over 10 generations who were descendants of the paternal grandparents of Abraham Lincoln. SCA5 has also been reported in French and German pedigrees (Zühlke *et al.*, 2007).

SCA11, initially reported in 2 British families, is caused by stop mutations, frameshift insertions or deletions in the *TTBK2* gene, resulting in a pure cerebellar syndrome with normal life expectancy (Houlden *et al.*, 2007). Pathogenic variants in *TTBK2* have also been reported in French and German families. (Bauer *et al.*, 2010)

SCA13 was initially reported in French and Filipino families and is caused by missense mutations in *KCNC3* which encodes a voltage-gated potassium channel (Figueroa *et al.*, 2010). There is a wide phenotypic spectrum that correlates with different missense mutations. The childhood-onset form, in which with motor and mental developmental delay is a common feature, has been associated with two variants: p.Arg423His and p.Phe448Leu described in a Filipino and European families respectively (Figueroa *et al.*, 2011). Females are more frequently affected and of the two missense mutations reported, the p.Phe448Leu variant results in the more severe phenotype. The p.Arg423His variant has also been reported in a Caucasian family in the United States.

SCA14 is caused by mutations in *PRKCG* (Yabe *et al.*, 2003), resulting in a variable ataxic phenotype which may include myoclonus, dystonia or peripheral neuropathy. The onset is usually in adulthood. The majority of mutations (missense) have been reported in exons 4, 5, 10 and 18. It has been reported in more than 20 families from Europe, Japan and Australia (Klebe *et al.*, 2005).

SCA15/16 is caused by heterozygous deletions of the 5' part of the *ITPR1* gene (van de Leemput *et al.*, 2007) although a missense mutation (c.1480G>A p.V494I) has been reported. The ITPR1 protein is highly expressed in cerebellar Purkinje cells and is an important modulator

of intracellular calcium signalling. SCA15/16 is characterized by a mild cerebellar ataxia with slow disease progression. In a French ataxia series, SCA15 was identified in 1.8% of patients (Cecilia Marelli et al 2011) SCA15/16 shares a locus with SCA29 raising the possibility that they are allelic disorders.

SCA19/22 are both caused by mutations in *KCND3* which encodes Kv4.3, an alpha subunit of a voltage-gated potassium channel important in membrane repolarization and expressed in the brain and the heart (Lee *et al.*, 2012). SCA19 and SCA22 were initially thought to be distinct disorders although there was overlap of the chromosomal locus on 1p21. SCA19 was a more complex ataxic syndrome associated with cognitive decline and polyneuropathy. SCA22 was described as pure cerebellar ataxia. Loss of function mutations are thought to underlie the spinocerebellar ataxic syndromes however gain of function mutations in KCND3 are associated with Brugada syndrome (Giudicessi *et al.*, 2011).

SCA20 has been described in a single Australian family of Anglo-Celtic descent and is the result of a 260kb duplicated region comprising >12 genes at 11q12 (Knight *et al.*, 2004). Bulbar symptoms including dysphonia and spasmodic cough in addition to dentate nucleus calcification are characteristic of this condition.

SCA23 is due to missense mutations of *PDYN* (Bakalkin *et al.*, 2010), which encodes prodynorphin protein, an opioid neuropeptide precursor. This causes a relatively pure cerebellar syndrome with a late onset (43-73 years) and slow progression. The disease has been reported in only a single large Dutch ataxia family and was not identified on screening a large German ataxia series (Schicks *et al.*, 2011).

SCA27 causes an early onset ataxia (Brusse *et al.*, 2006), associated cognitive deficits and head or limb tremor that can be exacerbated by stress or exercise. Missense and nonsense mutations in *FGF14* have been reported. Life expectancy is normal however most affected

patients are unable to walk by the 7th-8th decade. The disease has been reported in a large Dutch ataxia family and one additional German ataxia patient.

SCA28 is caused by a mutation in *AFG3L2* which encodes a mitochondrially located metalloprotease (Di Bella *et al.*, 2010). Missense mutations have been reported which are commonly located in the proteolytic domain of the protein with a mutation hotspot in exons 15-16. SCA28 has a typically early onset between 12-36 years and is characterized by a slowly progressive cerebellar ataxia with ophthalmoparesis and lower limb hyperreflexia. The disease is estimated to account for 1.5% of European ADCA cases (Cagnoli *et al.*, 2010).

SCA35 is caused by mutations in the cerebral transglutaminase *TGM6* and was the first dominant ataxia gene to be identified through exome sequencing (Wang *et al.*, 2010). Missense mutations were reported in two Chinese families in which a late onset cerebellar syndrome with associate upper motor neurone involvement was reported. There was moderate progression with patients commonly using a wheelchair 20 years after disease onset.

SCA38 is caused by mutations in *ELOVL5* which is an enzyme involved in the synthesis of long chain fatty acids. ELOVL5 is specifically expressed in cerebellar Purkinje cells 2 missense mutations (Leu72Val; Gly230Val) have been reported in French and Italian families. SCA38 is a pure cerebellar ataxia which presents in the 4th to 6th decade. Docosahexaenoic acid (DHA) supplementation has shown efficacy, in multiple patients from two affected families, in improving the clinical symptoms of cerebellar ataxia.

Other recently identified SCAs are summarised in the table below.

Name	OMIM #	Locus	Gene	Protein	Mutation	% of ADCA	Geographical distribution	Distinguishing features
Repeat Expansions: coding								
SCA1	164400	6p22.3	ATXN1	Ataxin 1	CAG repeat	6-27%	Common: South Africa, Japan, India, Italy, Australia	Hyperreflexia, sensory neuropathy, mild cognitive impairment
SCA2	183090	12q24.12	ATXN2	Ataxin 2	CAG repeat	13-18%	Common: US Spain, India, Mexico, Italy	Polyneuropathy, parkinsonism, dysphagia
SCA3	109150	14q32.12	ATXN3	Ataxin 3	CAG repeat	20-50%	Most common worldwide	Spasticity, polyneuropathy, dystonia, parkinsonism
SCA6	183086	19p13.2	CACNA1A	Calcium channel, voltage dependent, P/Q type, α_{1A} subunit	CAG repeat	13-15%	Common: US, Germany, Australia, Taiwan	Late onset, pure ataxia
SCA7	164500	3p14.1	ATXN7	Ataxin 7	CAG repeat	3-5%	Finland, Mexico, South Africa	Retinal degeneration
SCA17	607136	6q27	TBP	TATA box-binding protein	CAG repeat	rare	UK, Belgium, France, Germany, Japan	Dementia
DRPLA	125370	12p13.31	ATN1	Atrophin 1	CAG repeat	0.8:100000 (Japan) Rare worldwide	Japan, Portugal, United States	Dementia, epilepsy
HD	143100	4p16.3	HTT	Huntingtin	CAG repeat	3-7:100000	Worldwide	Chorea, dementia
Repeat Expansions: non-coding								
SCA8	608768	13q21.33	*ATXN8OS*	Ataxin 8 opposite strand	CTG repeat	3%	Common: Finland	Pure ataxia
SCA10	603516	22q13.31	*ATXN10*	Ataxin 10	ATTCT repeat	Unknown	Mexico, Brazil	Seizures
SCA12	604326	5q32	*PPP2R2B*	Protein phosphatase 2, regulatory subunit B, β	CAG repeat	Rare worldwide 7% in India	Common: India	Tremor, polyneuropathy
SCA31	117210	16q21	*BEAN*	Brain-Expressed Associated with NEDD4	TGGAA repeat	8-40% in Japan Rare worldwide	Japan esp. Nagano prefecture	Spasmodic torticollis
SCA36	614153	20p13	*NOP56*	Nuclear protein 56	GGCCTG repeat	6.3 % Galicia, Spain 9 Japanese families	Spain, Japan	Motor neurone involvement
SCA37	615945	1p32	*DAB1*	Disabled, drosophila, homolog of, 1	ATTTC repeat	Rare	Spain, Portugal	Pure ataxia
Conventional mutations								
SCA5	600224	11q13	*SPTBN2*	Beta 3 Spectrin	Deletions, missense mutations	Rare	United States, Germany, France	Pure ataxia, facial myokymia, gaze palsy
SCA11	611695	15q15.2	*TTBK2*	Tau tubulin kinase 2	Nonsense, frameshift deletions/insertions	Rare	UK, France, Germany	Pure ataxia
SCA13	605259	19q13.3-q13.4	*KCNC3*	Potassium channel, voltage-gated, shaw-related subfamily, member 3	Missense	1% (France)	France, Philippines	Early onset, mental

Name	OMIM #	Locus	Gene	Protein	Mutation	% of ADCA	Geographical distribution	Distinguishing features
SCA14	176980	19q13.4	*PRKCG*	Protein kinase C gamma	Missense, deletion	2% (France)	UK, France, Netherlands, US, Japan, Australia	Myoclonus, dystonia
SCA15	606658	3p26-p25	*ITPR1*	Inositol 1, 4, 5-triphosphate receptor type 1	Multi-exon or whole gene deletion, missense	1.8% (France) 0.3% (Japan)	UK, France	Pure ataxia
SCA19/22	607346	1p21-q21	*KCND3*	Kv4.3, an alpha subunit of a voltage-gated potassium channel	Missense (typically loss of function)	Rare	Diverse ethnic origins	Relatively pure ataxia or associated with cognitive impairment/dementia
SCA20	608687	11q12	-	-	260kb duplication	Rare	Australia	Dentate calcification, bulbar symptoms
SCA21	607454	1p36	*TMEM240*	Transmembrane protein 240	Missense	2% French SCA	European, South America, China	Ataxia with hyper and hypokinetic movement disorder, cognitive impairment
SCA23	610245	20p13	*PDYN*	Prodynorphin	Missense	Rare	Netherlands	Pyramidal signs
SCA25	608703	2p16	*PNPT1*	Polyribonucleotide Nucleotidyltransferase, Mitochondrial 1	Missense	Rare	France, Australia	Ataxia with sensory neuropathy
SCA26	609306	19p13.3	*EEF2*	Eukaryotic translation elongation factor 2	Missense	Rare	USA	Late onset pure cerebellar ataxia
SCA27	609307	13q33.1	*FGF14*	Fibroblast growth factor 14	Missense	Rare	Netherlands	Onset with tremor, psychiatric episodes
SCA28	610246	18p11.21	*AFG3L2*	ATPase FAMILY GENE 3-LIKE 2	Missense, deletion	1.5% ADCA (European)	Italy, France, UK	Ophthalmoplegia, spasticity
SCA35	613908	20p13	*TGM6*	Transglutaminase 6	Missense	Rare	China	Pure ataxia
SCA38	615957	6p22.2-q14.1	*ELOVL5*	Elongation of very long chain fatty acids-like 5	Missense	Rare	France, Italy	Late onset ataxia with polyneuropathy; low serum docosahexanoic acid
SCA40	616053	14q32.11	*CCDC88C*	Coiled-coil domain containing 88C	Missense	Rare	Hong Kong	Late onset spastic ataxia
SCA42	616795	17q21	*CACNA1G*	Calcium Voltage-Gated Channel Subunit Alpha1 G	Missense	Rare	France, Japan	Pure ataxia – age at onset highly variable
SCA44	617691	6q24	*GRM1*	metabotropic glutamate receptor 1	Missense	Rare		Pure ataxia; spasticity in 1 patient
SCA45	617769	5q33.1	*FAT2*	FAT atypical cadherin 2	Missense	Rare		Late onset ataxia – slowly progressive
SCA47	617931	1p35.2	*PUM1*	Pumilio, Drosophila, homolog of, 1	Missense, deletion	Rare		Missense mutations associated with late-onset ataxia, deletions associated with severe phenotype and developmental delay
SCA48	618093	16p13	*STUB1*	STIP1 homologous and U box-containing protein 1	Missense	Common in Italy (23.5% of familial cases in a series of 235 patients)	Europe esp. Italy	Late onset ataxia with extrapyramidal signs – chorea, parkinsonism; frontal lobe dementia

Table 1-1 Summary of autosomal dominant cerebellar ataxias

Name	OMIM #	Locus	Gene	Protein	Mutation	Geographical distribution	Characteristic features
EA1	160120	12p13.32	*KCNA1*	Potassium Channel, Voltage-Gated, Shaker-Related Subfamily, Member 1	Missense	worldwide	Attacks last seconds – minutes; myokymia
EA2	108500	19p13.2	*CACNA1A*	Calcium Channel, Voltage Dependent, P/Q Type, A_{1a} Subunit	Missense, nonsense, large deletions	worldwide	Attacks last hours; allelic with SCA6, Familial Hemiplegic Migraine
EA5	613855	2q23.3	*CACNB4*	Calcium Channel, Voltage-Dependent, Beta-4 Subunit	Missense	French-Canadian family	Attacks last hours - days; late onset, seizures
EA6	612656	5p13.2	*SLC1A3*	Solute Carrier Family 1 (Glial High Affinity Glutamate Transporter), Member 3	Missense	United States, Netherlands	Alternating hemiplegia, seizures

Table 1-2 Summary of episodic ataxias

Disease	OMIM #	Gene	Protein	Mutation	Incidence/Carrier frequency	Geographic distribution	Characteristic features
Friedreich's ataxia	606829	*FXN*	frataxin	GAA repeat expansions; point mutations in compound heterozygotes	Incidence: 1:30-50000 Carrier frequency: 0.9-1.6%	Worldwide except natives to: Far East, sub-Saharan Africa, Australia, America	Spasticity, neuropathy, cardiac involvement
Ataxia-telangiectasia	607585	*ATM*	Ataxia telangiectasia mutated	Deletions: splice-site related; nonsense; missense	Incidence: 1:400000-450000 live births Carrier frequency: 0.35-1%	Reported in many worldwide populations	Oculomotor apraxia; extrapyramidal features; increased cancer risk/ radiosensitivity
Ataxia-telangiectasia like disorder (ATLD)	604391	*MRE11A*	Meiotic recombination 11, S. cerevisiae, homolog of	missense	25 reported cases worldwide	Saudi Arabia (15 cases), Japan (4 cases), UK (4 cases), Italy (2 cases)	Similar to ATM but milder phenotype
Ataxia-Oculomotor Apraxia type 1 (AOA1)	208920	*APTX*	Aprataxin	Insertion, deletion, missense	Rare worldwide– More common in Portuguese and Japanese populations	Portugal, Japan, France, Tunisia	Oculomotor apraxia, peripheral neuropathy
Cerebellar ataxia with muscle coenzyme Q10 deficiency	607426	*APTX*	Aprataxin	missense	Rare	Single Italian family	Low coenzyme Q10 levels; late onset hypergonadotropic hypogonadism
Ataxia-Oculomotor Apraxia type 2	606002	*SETX*	Senataxin	Nonsense, missense	Carrier frequency: 2.1-3.5 % Incidence: 1:400000 (Alsace)	Commoner in French-Canadian populations	Oculomotor apraxia (variable); extrapyramidal features; peripheral neuropathy
Spastic Ataxia of Charlevoix-Saguenay (ARSACS)	270550	*SACS*	Sacsin	Stop-gain deletions and point mutations most common	Carrier frequency (Quebec): 4.5% Incidence: 1/1930	Most common in Quebec. Tunisian, Turkish, Italian, Japanese families reported	Myelinated retinal fibers; prominent lower limb spasticity
Cerebellar ataxia, seizures and ubiquinone deficiency	612016	*ADCK3*	aarF domain containing kinase 3	Missense, splice site, frame shift, deletion	Rare	French, Dutch, British families reported	Mental retardation, seizures, low co-enzyme q10 levels
Spinocerebellar ataxia with axonal neuropathy (SCAN1)	607250	*TDP1*	Tyrosyl DNA phosphodiesterase 1	Missense	Rare	Saudi Arabian family	Axonal neuropathy
Autosomal recessive spinocerebellar ataxia type 8	610743	*SYNE1*	Synaptic Nuclear Envelope protein 1	Splice site, intronic	Rare worldwide 3rd most common ARCA in Quebec	Canada	Hyperreflexia
Autosomal recessive spinocerebellar ataxia type 10	613728	*ANO10*	Anoctamin 10	Missense, splice site, deletion	Rare	French, Dutch, Serbian families	Tortuous conjunctival vessels

Table 1-3 Summary of autosomal recessive ataxias

1.5.5 Episodic ataxias

The episodic ataxias are a group of heterogeneous channel disorders characterized by attacks of ataxia, which may be associated with a range of other neurological manifestations including myokymia, migraine, seizures or chorea. 8 episodic-ataxia syndromes have been described: EA1-7 and episodic ataxia with paroxysmal choreoathetosis & spasticity (CSE). EA1 & EA2 are the most common and best characterised of these. The genes for EA1, 2, 5 and 6 have been identified with linkage loci mapped in EA3, 7 and CSE. Episodic ataxia is rare with a combined incidence of <1:100000.

EA1 is primarily due to missense mutations in *KCNA1* (Browne, 1994) although truncating mutations have been reported. The disease is characterised by brief periods of ataxia (lasting seconds to minutes) and interictal myokymia. The degree of channel impairment correlates with the severity of the phenotype. Mutations associated with severe phenotypes that may be poorly treatment responsive or associated with seizures or neuromyotonia show the most significant impairment of potassium channel function.

EA2 is due to a range of mutations in *CACNA1A* (Ophoff *et al.*, 1996) which include missense, nonsense, aberrant splicing and nucleotide insertions and deletions. EA2 typified by longer periods of ataxia lasting several hours with baseline nystagmus and progressive ataxia. There is a wide spectrum of phenotypes associated with mutations in *CACNA1A*. EA2 is allelic with SCA6 and familial hemiplegic migraine (FHM). Most of the mutations that cause EA2 disrupt the open reading frame whereas FHM is caused primarily by missense mutations.

EA5 has been described in a single French-Canadian family that were heterozygous for a missense mutation in the *CACNB4* gene resulting in a phenotype similar to EA2 (Escayg *et al.*, 2000). This mutation appears to have pleiotropic effects given that the same variant was identified in a German family with gene with generalised epilepsy but no ataxia.

Deleted:

Deleted: The precise functional effects of this mutation are not clear as the …

Deleted: mutation

EA6 was initially reported in a patient from the US presenting with characteristic episodes of hemiplegia, seizures and ataxia. A de novo mutation was identified in the *SLC1A3* gene, which results in complete loss of function of the protein EAAT1- a glutamate transporter localised to astrocytes. Other cases have been reported in the Netherlands with the p.C186S variant that resulted in a milder phenotype without the manifestations of seizures or alternating hemiplegia (de Vries *et al.*, 2009).

1.5.6 Autosomal recessive cerebellar ataxias (ARCAs)

The recessive ataxias are a particularly diverse group of disorders that are generally early onset with significant variation in clinical phenotype, which can include neuropathy, ophthalmological issues, seizures, and a range of other neurological and non-neurological manifestations. Some common pathological pathways have been described in the recessive ataxias including DNA repair dysfunction, mitochondrial dysfunction, defects in lipoprotein metabolism and protein chaperone dysfunction. There is significant overlap of clinical phenotypes with a range of metabolic ataxias, which are invariably complex early-onset multisystem disorders that can result in severe disability. A non-exhaustive summary of recessive ataxia genes is listed in table 2. There is significant geographic variability in the prevalence rates and genetic aetiology for recessive ataxias worldwide. The most comprehensive assessment of global prevalence from a meta-analysis of studies since 2000 estimates the prevalence of ARCAs to be 2.3-4.5/100000 (Ruano *et al.*, 2014). The same study estimated proportion of patients with an ARCA who do not have a genetic diagnosis after systematic testing to be 45-67%. This section will summarise the most common ARCA subtypes.

Deleted: is variably associated with

Deleted: disturbance

1.5.6.1 Friedreich's ataxia (FRDA)

FRDA is the most common recessive ataxia worldwide (Palau and Espinós, 2006) and is mainly due to homozygous GAA expansions in intron 1 of *FXN. FXN* encodes frataxin which is a

mitochondrial protein involved in trafficking of iron within mitochondria but whose precise role is has not been fully elucidated. Biallelic loss of *FXN* is incompatible with life in a murine model and it is only homozygous expanded *FXN* alleles or an expanded allele together with a second point mutation or deletion that have been shown to be associated with the FRDA phenotype. The cause for this is unclear but it seems likely that the expanded *FXN* allele interferes with *FXN* gene transcription.

Deleted: s

FRDA was first described by Nikolaus Friedreich in a series of monographs in 1863 although it would be 20 years before the disease was confirmed by other prominent neurologists who agreed that Friedreich's ataxia was a distinct entity. The prevalence of FRDA is estimated between 0.8-5/100000 with a carrier frequency between 1:60 -1:100 (Vankan, 2013). It is much less common in patients from sub-Saharan Africa and the Far East. FRDA is a multisystem neurodegenerative disorder characterised by progressive cerebellar ataxia, spasticity, distal areflexia and dorsal column sensory loss. Non-neurological features such as scoliosis and cardiac conduction defects and cardiomyopathy are common. Typically FRDA presents as an early onset syndrome around or before puberty with a mean age at onset of 10.5 years (Harding, 1981b).

1.5.6.2 Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)

ARSACS is caused by mutations in the *SACS* gene (Engert *et al.*, 2000). The disease was first described in families of French ancestry in Quebec where the disease is particularly prevalent due to a common founder mutation in the *SACS* gene (estimated frequency 1/22 in this region) (Thiffault *et al.*, 2013). Clinically ARSACS is characterised by a progressive cerebellar ataxia with prominent pyramidal signs, spasticity and a demyelinating sensorimotor neuropathy. Analysis of retinal fibre thickness on OCT shows a characteristic thickening of the fibre layer which can distinguish ARSACS from other cerebellar syndromes (Parkinson *et al.*, 2018).

Deleted: ngs and

SACS encodes sacsin, an extremely large 4,579 amino acid/520kDa protein which is highly expressed in the cerebellum, thalamus, midbrain, brainstem and pyramidal neurones. The function of sacsin is unknown however the protein contains multiple difference functional domains that may implicate roles as a molecular chaperone, in regulation of tau phosphorylation, protein quality control and microtubule function (Bagaria, Bagyinszky and An, 2022).

1.5.6.3 SYNE1 ataxia (ARCA1)

SYNE1 ataxia (ARCA1) is caused by mutations in *SYNE1* which encodes the Spectrin repeat containing nuclear envelope protein 1, a very large 1Mda protein believed to have role in linking organelles to the actin cytoskeleton. ARCA1 was first identified in several French-Canadian families from the Beauce and St Laurent regions of Quebec (Gros-Louis *et al.*, 2007). It is likely to be one of the more common recessive ataxias and may account for up to 5% of recessive ataxia after FRDA is excluded (Synofzik *et al.*, 2016).

SYNE1 mutations are associated with a spectrum of neurological phenotypes from pure ataxia to ataxia plus motor neurone and multisystem neuromuscular phenotypes (Synofzik *et al.*, 2016).

1.5.6.4 RFC1/CANVAS

Recently biallelic intronic pentanucleotide (AAGGG) repeat expansions were identified as a common cause of CANVAS syndrome (cerebellar ataxia, neuropathy and vestibular areflexia syndrome) (Cortese, 2020). The study identified biallelic expansions in *RFC1* in 90% of cases who met the criteria for full CANVAS syndrome and ~15% of late-onset sporadic ataxia cases which is an astonishingly high proportion and confirmed in other multi-centre studies (Traschütz et al., 2021) . The biallelic carrier rate is estimated to between 1:400 to 1:10000 which would make RFC1 expansions one of the most prevalent causes of ataxia. In additional

to the typical features of CANVAS, an unexplained dry cough was observed in 60% of patients which may be a useful diagnostic clue in cases of late-inset ataxia.

RFC1 expansions are polymorphic, and several repeat motifs have been subsequently shown to be associated with a disease phenotype ($AAGGG_n$, $ACAGG_n$, $AAAGG_{(10-25)}$, $AAGGG_n$, $AAAGG_{(4-6)}$). Although the pathogenic mechanism has not been fully elucidated, RFC1 encodes the large subunit of replication factor C, a DNA polymerase accessory protein involved in DNA replication and repair and is associated with several other pathways involved in DNA damage response. Furthermore, unlike other complex repeat expansion disorders (e.g. C9orf72, SCA36), accumulation of RNA foci or RAN toxicity has not been observed in RFC1 studies.

Deleted: polymorphic

1.5.6.5 Recessive ataxia due to genes associated with DNA repair

Several recessive syndromes, some of which relatively common, are now known to be associated with genes involved in DNA repair. Both double-stranded DNA repair genes (*ATM*, *MRE11*) and single stranded DNA repair defects (*APTX*, *SETX*,) are associated with early-onset cerebellar ataxia and usually a range of other phenotypes including neuropathy, spasticity and cancer predisposition.

Ataxia-telangiectasia was the first such syndrome to be reported and has been observed in most populations worldwide with an estimated incidence of 1:40000-300000 (Amirifar *et al.*, 2020). The disease is caused by mutations in the *ATM* gene which encodes the protein 'ataxia-telangiectasia mutated' which has an important role in double-stranded DNA repair. The clinical syndrome encompasses a range of manifestations including early-onset progressive ataxia (typically between 1-4 years of age), cutaneous telangiectasia, primary immunodeficiency and cancer predisposition. Life expectancy is significantly reduced with most deaths attributable to pulmonary infection.

A milder phenotype can be seen in patients with mutations in *MRE11* which causes ataxia telangiectasia-like disorder type 1 (ATLD1). MRE11 also encodes a protein with a role in double stranded DNA repair. Typically, the cerebellar syndrome is later onset and patients also show increased sensitivity to ionising radiation.

Ataxia with oculomotor apraxia type 1 (AOA1) is caused by mutations in aprataxin (APTX). Aprataxin is a protein with an important role in DNA repair caused by oxidative damage. AOA1 is relatively common in some populations in particular Portugal and Japan where it is the most common cause of ARCA. The onset is typically in early childhood with a progressive ataxic syndrome in association with prominent extrapyramidal features including dystonia. Oculomotor apraxia in addition to cerebellar eye signs is always present and patients also exhibit severe peripheral motor and sensory neuropathy. Coenzyme Q levels may be significantly reduced in the blood.

1.6 Genetics of dystonia

1.6.1 Introduction

Dystonias comprise a heterogeneous group of movement disorders and can be considered as an impairment or failure of motor programmes controlling semiautomatic movements or postures. A review of the nomenclature and phenomenology defined dystonia to be the "sustained or intermittent muscle contractions resulting in repetitive movements or postures causing abnormal, often repetitive movements or both. Dystonic movements are typically patterned, twisting and may be tremulous" (Albanese *et al.*, 2013) . It is the third most common movement disorder worldwide after essential tremor and Parkinson's disease (Steeves *et al.*, 2012).

There have been several attempts to categorise and classify dystonia however the most recently proposed classification schema is the least ambiguous (Albanese *et al.*, 2013). This schema uses two axes to define the condition:

Axis 1 – *Clinical characteristics of dystonia*.

These describe the phenomenology of a given patient and use the following categories: age at onset, body distribution, temporal pattern and co-existence of other movement disorders or neurological manifestations. A controversial aspect of this classification has been to eliminate the terms 'primary' and 'secondary' in describing the aetiology. The new approach seeks to classify broadly in terms of the presence or absence of associated features resulting in the term 'isolated' to imply the presence of dystonia as the only motor feature (but including tremor) or 'combined' where there are other neurological manifestations e.g. myoclonus or chorea.

Axis 2 - *This addresses the varied and often obscure aetiology of dystonia.*

1) Inherited (where a genetic cause has been proven) comprises autosomal dominant, autosomal recessive, X-Linked, and mitochondrial. Previously the DYT classification was used to define clinico-genetic syndromes and is still useful to designate the genetic subtypes.

2) Acquired - dystonia due to a known, non-genetic cause including perinatal brain injury, infection, drug-induced, toxic, vascular, neoplastic, non-perinatal brain injury and psychogenic.

3) Idiopathic (unknown cause) – these may be sporadic or familial and account for the majority of late-onset focal or segmental dystonia however it may be that cases previously classified as idiopathic may be subsequently identified as being inherited as new genes are discovered.

Table 1 2013 classification of dystonia[1]

Axis I: clinical characteristics	Axis II: etiology
Age at onset	**Neuropathology**
Infancy (0–2 years)	Degenerative
Childhood (3–12 years)	Structural lesion
Adolescence (13–20 years)	No degeneration/structural lesion
Early adulthood (21–40 years)	**Etiology**
Late adulthood (>40 years)	Inheritance pattern
Body distribution	Autosomal dominant
Focal	Autosomal recessive
Segmental	X-linked recessive
Multifocal	Mitochondrial
Generalized	Acquired cause
Hemidystonia	Perinatal brain injury
Temporal pattern	Infection
Disease course	Drug
Static	Toxic
Progressive	Vascular
Variability	Neoplastic
Persistent	Brain injury
Action-specific	Psychogenic
Diurnal	Idiopathic
Paroxysmal	Sporadic
Associated features	Familial
Isolated	

Figure 1-5 Summary of the 2013 Classification of dystonia - reproduced from (Berman & Jinnah 2015)

The pathophysiology of dystonia is not well understood and although has traditionally been viewed as a basal ganglia disorder. However, increasing evidence points toward dystonia as a network disorder involving sensory input pathways via the spinocerebellar tracts and cerebellar projections and basal ganglia pathways including nigrostriatal circuits (Schirinzi *et al.*, 2018). Anatomical studies have demonstrated that lesions of the basal ganglia, thalamus, sensorimotor cortex, cerebellum and brainstem can induce dystonia (Jinnah, Neychev and Hess, 2017). The role of the cerebellum in dystonia is increasingly recognised and there is evidence from studies of genetic ataxias showing that many patients exhibit both cerebellar ataxia and dystonia (van Gaalen, Giunti and Van de Warrenburg, 2011). Dystonia is frequently seen in SCA2 (14%), SCA3 (24%) and SCA17 (53%) and has been reported in SCAs 1, 6, 12, 14, 15 and 20. Furthermore, gene co-expression network analysis has identified 99 shared genes that may play a role in the pathogenesis of both disorders, and which implicated two main pathways – synaptic transmission and nervous system development (Nibbeling *et al.*, 2017).

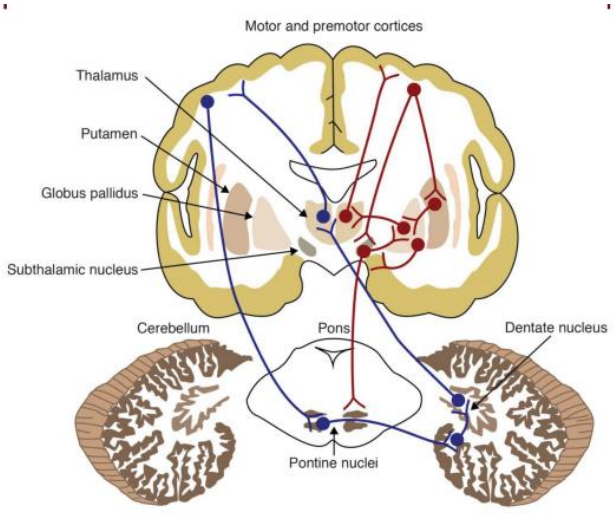


Figure 1-6 Dystonia as a network disorder - representation of basal ganglia and cerebellar circuits - reproduced from (Berman & Jinnah 2015)

1.6.2 Monogenic causes of isolated dystonia

There is strong evidence supporting a genetic basis for a significant proportion of both early and late-onset dystonia cases. Mendelian inheritance is most common in early-onset cases however variable penetrance is common and many apparently sporadic cases have been demonstrated to have a monogenic cause. A study examining the clinical features of first-degree relatives with cervical dystonia have shown clear evidence of dystonia in between 18-50% of individuals examined (Rubio-Agusti *et al.*, 2013).

The currently known genes associated with isolated dystonia are summarised below:

1.6.2.1 TOR1A (DYT1/Oppenheim's disease)

This was first described in 1911 by Oppenheim as a childhood-onset syndrome affected by jerking, twisting movements and abnormal gait which usually progressed to a fixed postural deformity (Klein and Fahn, 2013). Linkage to chromosome 9q32-34 was reported in 1990 and in 1997, a heterozygous in-frame GAG deletion in exon 5 of the *TOR1A* gene was identified (Ozelius *et al.*, 1997). The mutation accounts for about 50% of cases of generalised dystonia worldwide however it is particularly prevalent in the Ashkenazi Jewish population where it accounts for 80% of early-onset dystonia. (Valante *et al.*, 1998). The penetrance of the GAG deletion is around 30-40% and in carriers of the mutation there is a wide range of severity (Opal *et al.*, 2002). Typically, carriers would present in childhood with dystonic posturing of the lower limbs progressing to generalised dystonia however late-onset and focal presentations have been described (Opal *et al.*, 2002).

TOR1A encodes the protein TorsinA, a member of the AAA+ superfamily of ATPases. TorsinA is only expressed in neuronal cells throughout the central nervous system but at the highest levels in the dopaminergic neurones of the substantia nigra pars compacta, Purkinje cells, cerebellar dentate nucleus, thalamus, and frontal cortex.

1.6.2.2 TUBB4A (DYT4)

This gene was identified as part of this project and a detailed summary of DYT4 and the role of *TUBB4A* mutations is presented in chapter 3. In brief, a large Australian kindred, originating from Norfolk, UK, was identified with a dominantly inherited, severe generalised dystonia with prominent dysphonia and a characteristic 'hobby horse gait' (Parker, 1985).

Through a combination of linkage analysis and whole exome sequencing, we were able to identify a heterozygous missense p.Arg2Gly mutation in *TUBB4A* that segregated with the disease phenotype within area of linkage (Hersheson *et al.*, 2012). This was independently and concurrently verified by a separate group using whole genome and linkage data (Lohmann *et al.*, 2013).

TUBB4A encodes β-tubulin-4a which is a constituent of axonal microtubules and expressed exclusively within the CNS. The mutation is located within a highly conserved MREI (methionine-arginine-glutamic acid-isoleucine) domain which prior in vitro studies had shown to be essential for the autoregulation of β -tubulin-4a mRNA levels (Yen, Machlin and Cleveland, 1988).

Subsequent screening of a large cohort of ~200 dystonia cases did not identify any further *TUBB4a* mutations however a further study of 336 multi-ethnic dystonia patients identified a potentially pathogenic in-frame deletion c.1015_1017delAGC in a single case (Vulinovic *et al.*, 2017). It is likely therefore that *TUBB4A* mutations are an extremely rare cause of isolated dystonia.

1.6.2.3 THAP1 (DYT6)

DYT6 is an autosomal dominant dystonia first identified in 3 Mennonite families in the USA. A range of different mutations in *THAP1* have subsequently been identified in affected patients from diverse ethnic backgrounds (Bressman *et al.*, 2009; Fuchs *et al.*, 2009). The penetrance of

the *THAP1* mutations in Mennonite populations is around 60% but has not been measured in other populations.

Clinically, DYT6 patients develop symptoms in adolescence or adulthood and typically present with mainly upper body involvement including the upper limbs, oromandibular, cranio-cervical and laryngeal areas with later generalisation.

THAP1 is a member of a family of cellular factors sharing a highly conserved THAP domain with DNA binding properties. It has been speculated that *THAP1* mutations cause dystonia by altered regulation of key genes. In vitro studies have shown that THAP1 binds to the promotor of *TOR1A* and that *THAP1* mutations abolishes this binding effect suggesting a mechanistic link between DYT6 and DYT1 dystonia (Gavarini *et al.*, 2010).

1.6.2.4 GNAL (DYT25)

In 2012, heterozygous mutations in *GNAL* (guanine nucleotide-binding protein subunit alpha L) were identified as a novel cause of adult onset cranio-cervical dystonia (Fuchs *et al.*, 2012; Vemula *et al.*, 2013) . Subsequently around 30 different loss-of function *GNAL* mutations have been reported however the prevalence is low at between 0.5-1% of affected cases (Balint and Bhatia, 2015). Although it is usually inherited as a dominant trait, recessive cases with homozygous mutations have been documented, presenting as a severe childhood-onset dystonia. Typically, most mutation carriers present with cervical dystonia with progression to other body segments in around 50% of cases.

GNAL encodes the stimulatory α subunit of the heterotrimeric G protein Golf (Gαolf) which is highly expressed in the striatum where it couples D1 dopamine (D1R) and A2a adenosine (A2AR) receptors to downstream effector molecules.

1.6.2.5 ANO3 (DYT24)

Missense mutations in *ANO3* were identified in a British kindred with autosomal dominant cranio-cervical dystonia (Charlesworth *et al.,* 2012). Additional *ANO3* variants were identified in a screening study of 188 patients with dystonia. Further screening identified 3 families with dystonia and myoclonic jerks (Stamelou *et al.,* 2014).

The clinical characteristics are of a cervical dystonia and may include head or limb tremor. Laryngeal dystonia has been reported and around 10% of cases with progress to generalised dystonia. Hyposmia has been reported in some cases.

ANO3 encodes anoctamin 3 which is a transmembrane protein belonging to a family of calcium-activated chloride channels which is highly expressed in the striatum which may have a role in the modulation of neuronal excitability (Charlesworth *et al.,* 2012) .

2 Materials and methods

2.1 Case ascertainment and clinical assessment

Most of the pedigrees and individuals that I studied over the course of my research were identified from the Neurogenetics clinic at the National Hospital for Neurology and Neurosurgery by Prof Henry Houlden and Prof Nicholas Wood. In several cases, clinical characterisation was undertaken by collaborators or clinicians who had sent DNA samples to NHNN for analysis and or DNA storage. All DNA samples were obtained from the library of samples held at the department of neurogenetics with research consent.

Where possible and practicable, I undertook clinical evaluation including history and neurological examination of the patients either in the neurogenetics clinic or on occasion, in the patient's home.

Where patients were deceased or it was not possible to arrange contemporaneous clinical evaluation, available medical notes were obtained with consent and reviewed.

2.2 Ethics and consent

The ethics for all next generation sequencing and fibroblast studies was drafted by me and submitted to and approved by the research ethics committee at UCL. This was an update of previously approved ethics for the EUROSCA project (European Integrated Project on Spinocerebellar Ataxias).

Other ethical approvals were utilised for non-ataxia research namely the approval for NIHR Bioresource Study into rare diseases.

2.3 Skin biopsy and fibroblast culture

In selected patients, skin biopsies were obtained for subsequent fibroblast culture. Fibroblast samples were used where applicable for RNA extraction/cDNA synthesis and enzymatic assays.

Where fibroblasts have been used in this project, I performed the skin biopsy with the subsequent culture and storage undertaken by technicians in the cell-culture laboratory in the Molecular Neuroscience department at ION.

Written consent was obtained for skin biopsy and sample storage. Usually, the upper arm was selected for the biopsy site due to accessibility and ease of wound care. The area was cleaned using a 2% chlorhexidine sponge and 1% lignocaine infiltrated subcutaneously. A 3mm punch biopsy set was used with the sample immediately transferred to a container of biopsy medium (Gibco Roswell Park Memorial Institute medium (RPMI) 1640 (Life Technologies, 11875101) with 10% foetal bovine serum (FBS, CorningTM, 11-648-647) and 0.2% penicillin-streptomycin solution (Life Technologies, 15140-122). The wound was closed with a steri-strip and a Tegaderm dressing applied. Standard wound-care advice was given to the participant.

The biopsies were transferred into in vitro dishes where fibroblasts start growing from the skin samples after several weeks. These were passaged and fibroblasts were stored for subsequent use.

2.4 DNA extraction, quality assessment and quantification

For patients attending clinic or whom I was able to visit, 10mls fresh blood would be collected by standard venepuncture into an ethylenediaminetetraacetic acid (EDTA) tube for later DNA extraction. If this were not convenient, EDTA tubes with the relevant study information and consent, would be sent by post to be returned for subsequent DNA extraction

2.4.1 DNA extraction from blood

Samples were extracted from whole blood by the diagnostic genetics clinic at UCL Institute for Neurology. DNA was extracted using a Flexigene kit (Qiagen).

300µl of whole blood is mixed with 750 µl of cell-lysis buffer (FG1). The samples are centrifuged for 20 seconds at 10000g. The supernatant is then discarded, and the resulting

Deleted: obtained
Deleted: and
Deleted: was performed
Deleted: fetal

pellet mixed with a protease buffer (FG2). The sample is homogenised using a vortex and then centrifuged for 5 seconds before heating in a block at 65°C for 5 minutes. Precipitation of DNA is achieved by adding 150ul of isopropanol and inverting the tube several times. The samples are centrifuged for 3 minutes at 10000g. The supernatant is discarded, and the pellet allowed to dry. The DNA is re-dissolved by adding 200ul of buffer (FG3) and mixing for 5 seconds followed by heating in a block to 65°C.

The sample is then assessed for DNA concentration and quality using the method described below.

2.4.2 DNA quantification using spectrophotometry

The concentration and quality of DNA in all samples used in the project was assessed using a Nanodrop ND-1000 (Thermo Scientific) spectrophotometer. This technique relies on the fact that nucleic acids absorb ultraviolet light at a wavelength of 260nm. When a sample is exposed to this wavelength, the less UV light transmitted through the sample indicates a higher optical density and therefore a higher concentration of nucleic acid.

1µl of sample was used in the analysis. The concentration was reported in ng/µl. The purity of the sample was estimated by the absorbance ratio at 260/280nm and 260/230nm. Values of 1.8 and 2.0 respectively were indicative of good DNA quality with low levels of impurities.

These samples were then diluted with autoclaved distilled H20 (dH20) to the desired concentration.

2.4.3 DNA quantification using fluorescence

For next generation sequencing applications, a higher degree of accuracy is needed for DNA quantification. For this purpose, DNA was further quantified using a technique which relies on a fluorochrome (Picogreen) which can bind selectively to double stranded DNA. Picogreen has excitation/emission spectra at 485/530nm and does not bind to contaminants such as protein

or RNA. Whilst this is a more precise estimation of DNA quantity, the requirement for reagents makes this more expensive to use than the spectrophotometric method.

Samples are prepared by mixing 199ul of Qubit buffer with 1ul of Qubit reagent and 1ul of DNA solution to be tested. After mixing, the concentration is read on the Qubit sample reader (Invitrogen).

2.5 Primer design

Primers were designed using Primer3 (Untergasser *et al.*, 2012) which is a web-based tool for designing flanking primers based on an inputted nucleotide sequence. For any given variant or exon to be sequenced, the sequence flanking 200bp upstream and downstream was selected from the CCDS sequence using Ensembl (https://www.ensembl.org/index.html). Where exons were being sequenced, the longest transcript was selected from the Ensembl data for that gene.

Primers were designed with an optimum primer size set to 20bp and optimum melting temperature between 55-65°C. Primer GC content was set at ~50%. Primers were excluded if they were complementary to sequences containing SNPs as this can affect primer binding if a polymorphism is present in the primer binding site.

The primer sequences were checked using primer-BLAST (Ye *et al.*, 2012) which is a tool for checking for regions of nucleotide sequence similarity to ensure that the designed primers would only bind the target region.

2.6 Polymerase chain reaction (PCR)

PCR was performed in 96 well plates using 10µl Roche Fast start Master mix, 5µl dH20, 2µl of forward and reverse primers (5pmol/µl), 1µl genomic DNA at a concentration of 25-50ng/µl. The reaction mix was loaded onto an Eppendorf Mastercycler thermal cycler. The PCR conditions were optimised for each set of primers using control genomic DNA.

2.7 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to confirm the presence of a PCR product. When loaded into a gel, negatively charged DNA will travel across and electrical gradient when a voltage is applied across the gel.

10X tris-borate-EDTA (TBE) solution was prepared using 121.1g of Trizma base (Sigma), 61.8g boric acid (Sigma), 7.4g of EDTA (Sigma) and dissolved in 1L of distilled water. This was diluted to a working solution of 1X concentration with distilled water. To make the gel, 1X TBE buffer was combined with 1.5% Ultrapure Agarose (Invitrogen) and 20ul of Gel Red (Cambridge Biosciences) and allowed to set in a tray. Once set, 3µl of PCR product with 3µl X6 Orange loading dye (Thermo Scientific) were mixed and loaded into a single well. A DNA ladder (Midrange 100-2000bp (Qiagen) was run alongside the samples to enable approximate sizing of the PCR products. The gel was run at 100mV for 30 minutes and the DNA fragments visualised under an ultraviolet transilluminator.

2.8 PCR purification

PCR products were purified using either an enzymatic or filtration method. This is in order to remove any unbound primers and remaining dideoxynucleotides (dNTPs) in the PCR reaction mixture.

For enzymatic clean-up, an enzyme mix comprising Exonuclease 1 (Thermo Scientific) and Fast-Alkaline phosphatase (Thermo Scientific) was made by combining 50µl Exonuclease I, 200µl Fast-Alkaline phosphatase and 750µl of purified water. 2µl of enzyme mix was added to 5µl of PCR product. This was heated in a thermal cycler at 37°C for 30minutes then 80°C for 15minutes.

For filtration clean-up, 80µl of distilled water was added to each well of PCR product in the 96 well PCR sample plate and transferred to a 96 well PCR filter plate (Millipore, Co. Cork,

Ireland). This plate was then transferred to a vacuum for 5 minutes until the filter membrane was dry. The PCR product was then re-suspended in 50µl of purified water by gently agitating on a plate shaker for 30 minutes.

2.9 Sanger Sequencing

2.9.1 Sequencing reaction

Sanger sequencing was performed on a 3730 DNA analyser (Applied Biosystems, Ca, USA) using BigDye 3.1 (Applied Biosystems) sequencing chemistry. The sequencing reaction mix was prepared with 0.6µl BigDye Terminator v3.1 (Applied Biosystems), 2µL 5X sequencing buffer (Applied Biosystems), 1 µL forward or reverse primers (5pmol/ µL), 3 µL of purified PCR product and 2.15 µL of dH20. The following program is run on a thermal cycler: denaturation at 94°C for 1 minute, followed by x25 cycles of: denaturation at 94°C for 30 seconds, annealing at 50°C for 15 seconds, and elongation at 60°C for 4 minutes.

2.9.2 Sequencing reaction purification

The sequencing reaction was purified using a gel filtration method. Sephadex is a cross-linked dextran gel which can be used for the separation of unincorporated terminator dyes from the sequencing product. A hydrated Sephadex solution was prepared by mixing 40mls dH20 and 2.9g of Sephadex G-50 powder (Sigma Aldrich). This was mixed thoroughly for at least 30 minutes. 350µl of Sephadex solution was transferred to each well of a Corning FiltrEXTM 96 well plate. This was then placed on top of an empty collection plate and centrifuged for 3 minutes at 700g. The prepared filter was placed onto a new sample plate. The contents of the sequencing reaction plate were then transferred to the prepared filter plate and centrifuged at 910g for 5 minutes.

2.9.3 Sanger sequencing and analysis

The purified sequencing product was sequenced on a 3730 DNA analyser (Applied Biosystems) and the sequence chromatograms were viewed using Sequencher software (Gene Codes Corporation, MI, USA)

2.10 Whole genome SNP genotyping using DNA SNP arrays

Whole-genome SNP genotyping was performed using IlluminaHumanCytoSNP-12 v2.1 arrays (Illumina). These data were used for the purposes of linkage analysis and homozygosity mapping. The samples were processed according to the manufacturer's protocols at UCL Genomics.

Using the Human CytoSNP arrays, approximately 220,000 genome-wide polymorphic markers were genotyped. The data generated were imported into Genome Studio 2010 (Illumina) which was used to perform clustering, normalisation of calls and genotype calling. Files were then exported in the necessary formats for linkage analysis and homozygosity mapping.

2.11 Genome-Wide Parametric Linkage Analysis

Initial quality control of the genotype files was performed using the command line tool PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). SNPs with low call rates (<90%) and monomorphic SNPs (MAF <0.5%) were removed from the genotype list using the following command parameters:

plink -maf 0.05 -geno 0.1.

Around 5000 randomly selected SNPs were selected which were evenly distributed throughout the genome using the command parameter *-thin 0.03*

The resulting files were imported into MERLIN (http://www.sph.umich.edu/csg/abecasis/merlin/tour/linkage.html) which was used to

perform the linkage analysis. The following files were generated to be used in the actual analysis: 1) .ped file: this contains details about the affectation status of the individuals used in the analysis (1=unaffected, 2=affected, 3=unknown) and the sex (1=male 2=female). The pedigree structure is encoded into the .ped file by assigning unique codes for each individual with the associated codes that designate their parents. 2) .model file - this specifies the mode of inheritance (e.g. dominant or recessive), the estimated disease allele frequency and penetrance. 3 .map file – this contains the list of SNP markers and their genetic position in cM.

PEDSTATs (http://www.sph.umich.edu/csg/abecasis/PedStats/download/) was then used to validate the input formats, ensure pedigree consistency and to check genotyping errors. The following command was used:

pedstats –d file.dat –p file.ped.

MERLIN was used to perform a parametric genome-wide linkage analysis with command: *merlin –d file.dat –p file.ped –m file.map –model file.model –markernames –-pdf –-tabulate.* This generated a table with the maximum LOD scores for each marker and a graphical representation in a .pdf file. The analysis was repeated at least twice using different randomly selected markers to ensure that the results were consistent.

2.12 Homozygosity mapping

Homozygosity mapping was performed using the online HomozygosityMapper tool (Seelow *et al.*, 2009). Genotype data were first exported from GenomeStudio into the required format. This file contained genotype calls of all the probe markers in the genotyping array for each individual in the analysis. Where possible, multiple affected individuals were genotyped with the software reporting regions of shared homozygosity.

2.13 Whole Exome Sequencing (WES)

WES was performed at the UCL Institute of Neurology. I performed DNA sample preparation, quantification and quality control and library preparation and sequencing was performed by Deborah Hughes. The library preparation and enrichment were performed according to the manufacturer's instructions using the TruSeq Exome Enrichment kit which targets 62Mb of nucleotide sequence from the RefSeq database. This covers 45Mb of exonic content (≥98% of RefSeq, CCDS, and Ensembl coding content).

Samples were then sequenced using an Illumina HiSeq 2000. Sequencing data were demultiplexed using the CASAVA tool (Illumina) into a FASTQ file which is a text-file which stores unaligned sequence reads used for all downstream bioinformatic analysis. Each FASTQ file is approximately 600GB in size.

2.14 Bioinformatic analysis of whole exome sequencing data

1) The first stage is a quality control step to remove low quality sequencing reads based on a metric known as a Phred score which is a logarithmic score indicating the probability of an incorrect base call. A base with 1 in 10 probability (90% accurate) of being incorrect has a Phred score of 10 and if there is a 1 in 100000 probability (99.999% accurate) the Phred score is 50. In NGS, a Phred score of >30 is considered acceptable quality.

2) The reads are then aligned to the hg19 human reference genome using Novoalign generating a sequence aligned file (.SAM) which is converted into a binary format (.BAM) for ease of compression.

3) PCR duplicates occur when PCR amplification forms part of the library preparation protocol as is the case for all exome and some whole genome library preparation methods. These duplicates may interfere with accurate variant calling and are removed using Picard software (http://picard.sourceforge.net/).

4) Once a clean set of aligned reads has been generated, SAMTools is used to recalibrate base quality scores, perform local realignments around possible insertion/deletions (indels) (Li *et al.*, 2009). BCFtools performs the variant calling with the data stored in a variant call file (.vcf) for subsequent annotation.

5) Variant annotation is performed using ANNOVAR which annotates the variant calls with details of the gene, mutation type e.g. synonymous, nonsynonymous, stop gain, frameshift, splice site etc, amino acid change (Wang, Li and Hakonarson, 2010). Pathogenicity prediction scores are annotated using SIFT, GERP, MutationTaster, Polyphen, CADD. Allele frequencies are annotated according to the presence of the variant in a range of databases including 1000genomes, EVS, dbSNP and EXAC.

The data are presented in a .csv file which can be opened in Microsoft Excel for further filtering, variant review and prioritisation. Where relevant, the variant calls in the associated .bam file can be viewed in the Integrated Genomics Viewer software (Broad Institute).

3 Application of exome sequencing and linkage analysis to identify a novel gene causing autosomal dominant dystonia

3.1 Outline of the chapter

This chapter describes my genetic analysis of a large autosomal dominant kindred affected with a severe generalized dystonia and subsequent investigation into patients with an allelic syndrome (H-ABC). The specific aims of this study were:

1. To identify the causal gene underlying disease phenotype in the DYT4 dystonia pedigree

2. To investigate the role of TUBB4a mutations in a series of H-ABC syndrome patients

3.2 Statement of contribution

Section 3.2 The kindred were originally investigated in 1992 by Anita Harding who supervised the extraction of DNA from affected members of the DYT4 kindred and stored in the Institute of Neurology sample library. I performed a clinical re-evaluation based on available records and videos of affected patients. I performed the linkage analysis based on genotype data obtained by Dr Katherine Fawcett. This was confirmed and repeated by Dr Robert Kleta. I performed the analysis of exome sequencing data and subsequent expression study of TUBB4a cDNA in healthy controls. Analysis of brain expression data was based on data obtained by Prof Mina Ryten and Dr Trabzuni.

Section 3.3 Case ascertainment and clinical assessment was performed Dr Erro and Prof Bhatia within the movement disorder clinic at NHNN. I performed and analysed the genetic studies in these affected individuals.

Publications arising from this section:

- Mutations in the autoregulatory domain of β-tubulin 4a cause hereditary dystonia **Hersheson J,** Mencacci NE, Davis M, MacDonald N, Trabzuni D, Ryten M, Pittman A, Paudel R, Kara E, Fawcett K, Plagnol V, Bhatia KP, Medlar AJ, Stanescu HC, Hardy J, Kleta R, Wood NW, Houlden H. *Ann Neurol*. 2013 Apr;73(4): 19

- H-ABC syndrome and DYT4: Variable expressivity or pleiotropy of TUBB4 mutations? Erro R, **Hersheson J,** Ganos C, Mencacci NE, Stamelou M, Batla A, Thust SC, Bras JM, Guerreiro RJ, Hardy J, Quinn NP, Houlden H, Bhatia KP. Mov Disord. 2015 May;30(6):828-33.

- A novel TUBB4A mutation suggests that genotype-phenotype correlation of H-ABC syndrome needs to be revisited. Erro R, **Hersheson J**, Houlden H, Bhatia KP. Brain. 2015 Aug;138(Pt 8):e370

3.3 Identification of TUBB4a as the gene underlying DYT4 dystonia using linkage analysis and exome sequencing.

3.3.1 Introduction

In 1985, forensic psychiatrist Neville Parker described the clinical characteristics of a large Australian kindred in which at least 20 members of the family were affected with a severe generalised dystonia, laryngeal dysphonia, dysphagia resulting in significant weight loss and an unusual 'hobby horse' gait (Parker, 1985). Several of the affected individuals had been given a diagnosis of hysterical mutism or dysphonia. Two members of the family were also affected by Wilson's disease (WND).

The family originated from the United Kingdom and genealogical investigation revealed that two affected sisters had emigrated from the Heacham area of Norfolk to Townsville, Australia in 1886. Although several family members appeared to remain in England, no similar cases have ever been reported in the UK.

Deleted: showed

Deleted: they

A more recent genetic and clinical re-examination of nine surviving affected family members provided some helpful additional information about the disease phenotype (Wilcox *et al.*, 2011). The patients examined had a symptom onset between 17-42 years. Nearly all the affected cases presented with a progressive adductor dysphonia, which was intermittent in some individuals. Some patients found it possible to talk more normally during periods of high emotional and could laugh normally even when otherwise severely affected. Progression to generalized and gait dystonia was a common feature with 5 out of 9 cases exhibiting the unusual gait pattern described by Parker. A minority of cases were reported to be alcohol-responsive however, response to propranolol was variable and one patient was reported to be responsive to tetrabenazine.

Several genetic analyses had been undertaken to identify the genetic locus in this family and to exclude known dystonia loci (Jarman *et al.*, 1999; Wilcox *et al.*, 2011). These studies had

previously excluded linkage in this family to DYT1, DYT6, DYT7, DYT11, DYT13, DYT15, and the ATP7B (WND). A complicating feature of the family was that two of the affected individuals were known to have had WND leading initially to the suspicion that the disease was linked to the WND locus.

Using genome wide linkage analysis and whole exome-sequencing, I was able to show that a segregating heterozygous mutation in the TUBB4a gene was the cause of the disease in this family.

3.3.2 Subjects, materials and methods

3.3.2.1 Index family

The clinical characteristics of the affected family members were ascertained from research notes and video recordings obtained by Dr Heather Waddy in the 1980s. These records, together with more recent published clinical assessment of the family formed the basis of the clinical re-evaluation and summary shown below. Blood samples were collected, and DNA extracted from 26 unaffected 12 affected and individuals from the DYT4 kindred using standard techniques.

Ethical approval was obtained for the cases used in the study (see section 2.2).

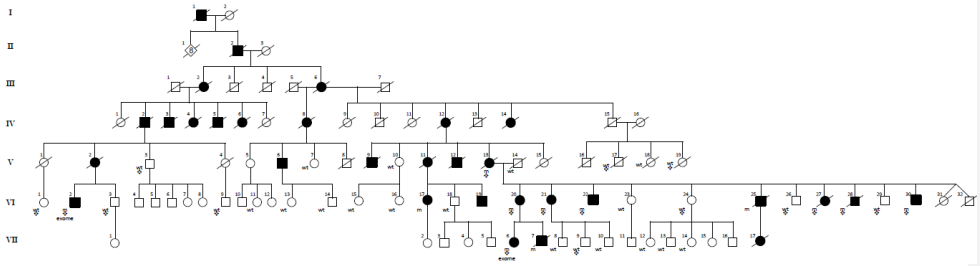


*Figure 3-1- Pedigree of the DYT4 family: * symbol indicates DNA available and mutation screening undertaken; ** indicates exome sequencing performed; ⍰ indicates individuals included in linkage analysis; V-26 and V-27 were known to have Wilson's disease and were also heterozygous for the R2G variant*

Patient ID	Age at onset	Age at examination	Clinical characteristics
V-16	30	44	Dysphonia: progressed over 2 years until unable to speak; cervical dystonia (35yrs)
V-2	21	42	Dysphonia (rapid progression resulting in psychiatric referral); swallowing (25yrs), cervical dystonia (34yrs), gait (37yrs)
V-24	23	31	Dysphonia, progression over 6 years to involve cervical muscles, tongue followed by limb dystonia
V-14	37	60	Onset with stooped posture; progressive dysphonia with swallowing difficulties over 5 years; cervical and oral dystonia; wheelchair bound (53yrs)
V-26	13	29	Severe dystonic gait, hepatitis and haemolytic anaemia; KF rings; ataxia. Diagnosed as having Wilson's disease but with additional dystonic features typical of DYT4.
V-27	15	29	Dysarthria; KF rings; upper limb dystonia. Diagnosed as having Wilson's disease but with additional dystonic features typical of DYT4.
V-20	28	37	Cervical dystonia; progressive dysphonia (30 yrs.); no swallowing difficulties; left hemidystonia (32yrs
V-18	13	35	Dysphonia; cervical dystonia (14yrs)

Table 3-1- Clinical characteristics of selected affected family members. The patient ID refers to the position of the individual on the family tree as per Figure 1A

3.3.2.2 Linkage analysis

This was carried out on 20 family members (marked on pedigree with ❖ in Figure 1), comprising 11 unaffected and 9 affected cases. These were genotyped using Illumina CytoSNP12 arrays with 301,232 genome-wide markers and the raw data was processed in GenomeStudio (Illumina, San Diego, US). Genotypes were examined with the use of a multipoint parametric linkage analysis and haplotype reconstruction performed with SwiftLink with the results confirmed with SimWalk. There were 24,000 informative SNPs, equally spaced 0.1cM apart used in the analysis. Data were formatted with Mega2 through ALOHMORA. Mendelian inconsistencies were checked with PedCheck. An autosomal dominant model was specified with an estimated allele frequency of 0.00001 and 90% penetrance.

3.3.2.3 Whole exome sequencing

3ug of DNA was used for exome sequencing using TruSeq SBS chemistry in 2 distantly related, affected relatives. The methods used were as described in the Methods section. Annotated variant files were generated which included a comparison to several reference databases including dbSNP (www.ncbi.nlm.nih.gov/projects/SNP), 1000genomes (www.1000genomes.org), NHLBI Exome Variant Server (evs.gs.washington.edu) and Complete Genomics cg69 database (www.completegenomics.com/public-data/69-Genomes). In-silico predictions of pathogenicity were performed using SIFT (Kumar, Henikoff and Ng, 2009), PolyPhen2 and MutationTaster (Schwarz *et al.*, 2010). Protein sequence alignment was performed using ClustalOmega (www.ebi.ac.uk/Tools/msa/clustalo).

3.3.2.4 TUBB4a mutation screening

To validate the results of exome sequencing, and to sequence other dystonia families and control individuals, primers were designed using Primer3 to amplify the entire *TUBB4a* gene. The following reaction conditions were used for PCR amplification: 12µl FastStart Mastermix (Roche), 1µl (10pM) forward and reverse primers, 1µl DMSO, 4µl PCR grade water, 1µl DNA

(30ng/µl). PCR products were visualized on a 2% agarose gel and cleaned with Millipore filter plates. Clean PCR product was sequenced using BigDye Terminator 3.1 chemistry (Life Technologies, Carlsbad, CA, USA) and cleaned with Millipore filter plates prior to capillary electrophoresis on an ABI 3130XL Genetic Analyzer (ABI Biosystems, Foster City, CA, USA). Sequencing data was visualized using Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA).

3.3.2.5 TUBB4a expression studies

Expression of TUBB4a in various human tissues was determined using gene-specific primers against cDNA generated from tissue-specific RNA in the FirstChoice Human Total RNA Survey Panel (Life Technologies, Carlsbad, USA). The cDNA was synthesised from the mRNA panel with SuperScript II reverse transcriptase according to the manufacturer's protocol, 1000ng of mRNA was used as template with random oligonucleotide primers. cDNA was made from the following regions, trachea, thyroid, prostate, skeletal muscle, spleen, small intestine, thymus, lung, placenta, kidney, adipose tissue, brain, oesophagus, colon, heart, liver, ovary, cervix, bladder and testes. One microliter of the resulting cDNA product was then used as a template for the RT-PCR reaction at 30 cycles with primers (forward CTCGCCTGCACTTCTTCAT and reverse CAGCCTCTTTGTTCCCAGTC) specific to *TUBB4a* cDNA and a comparative reaction with a housekeeping gene. This was visualised on a 2% agarose gel.

Regional distribution of *TUBB4a* mRNA expression in the normal human brain was determined using microarray analysis of human post-mortem brain tissue from the UK Human Brain Expression Consortium. Brain tissues originating from 134 control Caucasian individuals were collected by the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank (Edinburgh, UK). The following brain regions were included in the analysis: cerebellum (CRBL), frontal cortex (FCTX), hippocampus (HIPP), medulla (MEDU), occipital cortex (OCTX), putamen (PUTM), substantia nigra (SNIG), temporal cortex (TCTX), thalamus (THAL) and white matter

(WHMT). Total RNA was isolated from these tissues using mRNeasy 96-well kit (Qiagen, UK) before processing with the Ambion WT Expression Kit and Affymetrix GeneChip Whole Transcript Sense Target Labelling Assay, and hybridization to the Affymetrix Exon 1.0 ST Array. The probe-set defining TUBB4a mRNA was determined using the Affymetrix Netaffx annotation file (HuEx-1_0-st-v2 Probe set Annotations, Release 31). The combined signal of the TUBB4a probe sets was used to determine mRNA expression.

3.3.3 Results

3.3.3.1 Linkage analysis:

Linkage to the known DYT genetic loci was excluded by multipoint parametric linkage analysis which identified, across the whole autosomal genome, a single significant linked region on chromosome 19p13.12-13 between SNP markers rs12977803 and rs2303099. The maximum LOD score between these markers was 6.33. There were no other regions of linkage in the genome (Fig 5-2). To find a single significant LOD peak such as this is uncommon in linkage studies however the analysis included a large number of distantly affected cases in a fully penetrant disease.

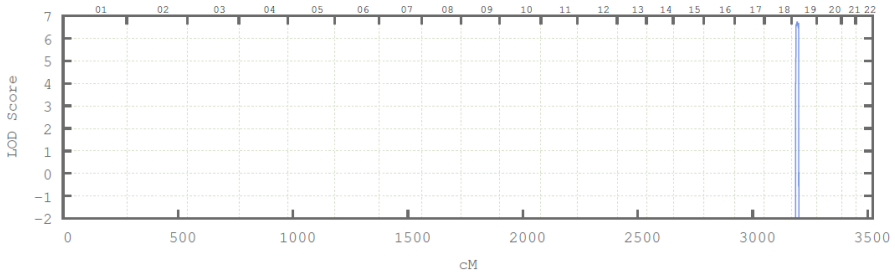


Figure 3-2 - Multipoint parametric linkage analysis of DYT4 pedigree - Y axis shows single LOD peak. Lower X axis shows genetic distance in cM, upper X axis - autosomal chromosomes

Deleted:

3.3.3.2 Exome sequencing:

Exome sequencing was performed on 2 affected cousins VI-2 and VII-6. A summary of the read data can be found in Table 3-2.

Following alignment and quality assessment of the data, 23398 and 23701 variants were identified in exome data for VI-2 and VII-6, respectively. The filtering strategy undertaken initially excluded homozygous and synonymous variants. Variants were then filtered against several control data sets, including the 1000 Genomes, Exome Variant Server, UCL Exome, and cg69 databases, but not dbSNP because of concerns about pathological SNPs having been uploaded to it. In total, 153 and 156 novel variants were identified in the 2 respective patients, 3 of which were located within the linkage region. Both individuals shared 2 novel variants within the linkage region: TUBB4a (c.4C>G, p.R2G) and FCER2 (c.947C>T, p.S316F).

Deleted: being

Patient	V-2	VI-6
Unique reads	136291642	137233190
Aligned Reads (%)	86.2	85.4
Mean depth	102	106
Total variants	23398	23701
Heterozygous variants	14333	14207
Excluding synonymous variants	7294	7119
Novel variants	153	156
Variants in linkage region	3	3
Shared variants	2	2

Table 3-2 - Results from exome sequencing of patients VI-2 and VII-6 with the variants that were identified. The patient ID refers to the position of the individual on the family tree as per Figure 1

3.3.3.3 Mutation analysis

The TUBB4a variant is located in the highly conserved autoregulatory MREI domain in exon 1 of TUBB4a and results in an arginine to glycine (p.R2G) amino acid substitution. The R2G

mutation co-segregated perfectly with the disease phenotype, with all affected individuals having the R2G variant.

The genotypes of individuals screened in the segregation analysis are indicated in Figure 1 with either wt (wild type) or m (mutant allele: c.4C>G, p.R2G heterozygote). There were no known unaffected carriers, although individuals younger than 18 years were not analysed.

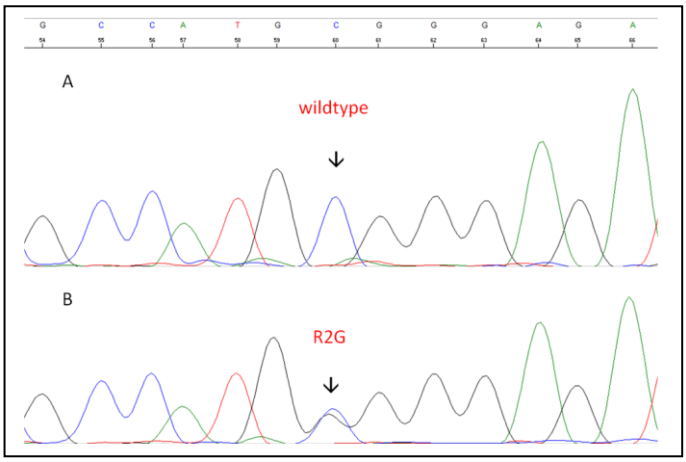


Figure 3-3- Sequence chromatogram showing (A) unaffected family member with the wildtype sequence (B) affected family member with a heterozygous c.4C>G: p.R2G mutation

This variant was not found in any of the reference SNP, exome, or in-house exome databases queried, and was not present in 1,045 ethnically matched UK control individuals. Variants in exon 1 were not identified in 124 mixed dystonia cases.

The variant was predicted to be deleterious by in silico analysis (SIFT and PolyPhen2) and was highly conserved in multi-species and multi-tubulin alignment.

The FCER2 gene which codes for an immunoglobulin E receptor (CD23) is expressed solely on lymphocytes and is involved in allergy and resistance to parasites. The variant was predicted to be deleterious by only one out of three prediction programs (SIFT) and there was no significant evolutionary conservation at this locus.

Tubulin	Tissue specificity	Protein sequence alignment (human)
TUBB4a	Brain-specific	MREIVHLQAGQCGNQIGAKFWEVISDEHGIDPTGTYH
TUBB	Ubiquitous	MREIVHIQAGQCGNQIGAKFWEVISDEHGIDPTGTYH
TUBB1	Haematopoietic cells	MREIVHIQIGQCGNQIGAKFWEMIGEEHGIDLAGSDR
TUBB2a	Brain-specific	MREIVHIQAGQCGNQIGAKFWEVISDEHGIDPTGSYH
TUBB2b	Brain-specific	MREIVHIQAGQCGNQIGAKFWEVISDEHGIDPTGSYH
TUBB3	Neuron-specific	MREIVHIQAGQCGNQIGAKFWEVISDEHGIDPSGNYV
TUBB4b	Ubiquitous	MREIVHLQAGQCGNQIGAKFWEVISDEHGIDPTGTYH
TUBB6	Ubiquitous	MREIVHIQAGQCGNQIGTKFWEVISDEHGIDPAGGYV
TUBB8	Ubiquitous	MREIVLTQIGQCGNQIGAKFWEVISDEHAIDSAGTYH

Table 3-3 - Multispecies protein sequence alignment of beta tubulin showing the highly conserved MREI sequence (highlighted in red)

Species	Protein sequence alignment (beta tubulin)
Homo (human)	MREIVHLQAGQCGNQIGAKFWEVISDEHGIDPTGTYHGD
Macaca (macaque)	MREIVHLQAGQCGNQIGAKFWEVISDEHGIDPTGTYHGD
Bos (cow)	MREIVHIQAGQCGNQIGAKFWEVISDEHGIDPTGTYHGD
Mus (mouse)	MREIVHIQAGQCGNQIGAKFWEVISDEHGIDPTGTYHGD
Xenopus (frog)	MREIVHLQAGQCGNQIGAKFWEVISDEHGIDPTGAYHGD
Arabidopsis (cress)	MREILHIQGGQCGNQIGSKFWEVICDEHGIDSTGRYSGD
Dictyostellida (slime mould)	MREIVQIQAGQCGNQIGSKFWEVISEEHGIQSDGFHAGG

Table 3-4 - The conservation of the protein sequences in the different tubulin isotypes. The MREI sequence is indicated in red

3.3.3.4 TUBB4a expression analysis:

Expression of TUBB4a was determined using a panel of cDNA from various body tissues. TUBB4a expression is largely confined to the brain with very low expression in all other body tissues assayed although weak expression was demonstrated in the testes (Figure 5-4).

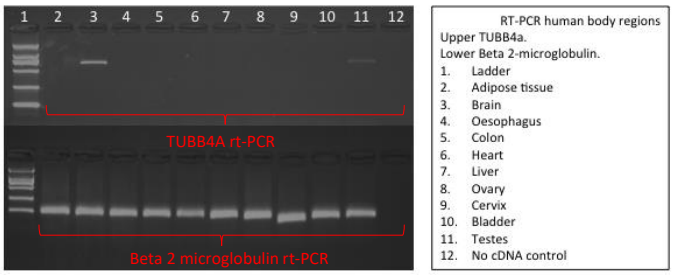


Figure 3-4 -Expression of TUBB4a in various human tissues determined using gene-specific primers against cDNA generated from tissue-specific RNA as compared to the house keeping gene beta2-microglobulin. Expression was high in the brain and very low in other tissues except for moderate expression in the testes

Expression of TUBB4a gene in 10 brain regions from 134 normal individuals was assessed using Affymetrix (Santa Clara, CA) Exon 1.0 ST Arrays, which identified high expression in the brain (see table 3-5, figure 3-5). DNA extracted from brain tissue for these individuals were negative for defects in the TUBB4a gene using Sanger sequencing. The highest expression was in the cerebellum, followed by putamen and white matter. There was a 2-fold difference between the cerebellum and the thalamus, with the lowest brain expression.

Region sampled	Brain regions (n)	Mean *TUBB4a* expression (log2 scale)
CRBL	130	10.84
FCTX	127	10.2
HIPP	122	10.2
MEDU	119	10.47
OCTX	129	10.19
PUTM	129	10.18
SNIG	101	9.94
TCTX	119	10.09

Table 3-5 Expression of TUBB4a in 10 brain regions from 134 normal individuals assessed using the Affymetrix Exon 1.0 ST Array. The values for TUBB4a expression are corrected for brain bank, batch effect and gender effects. See Trabzuni et al, J Neurochem. 2011 Oct;119(2):275-82 for detailed methods

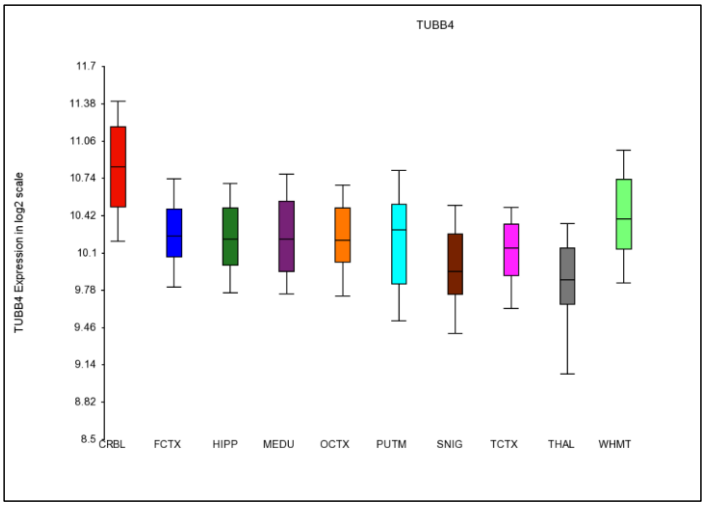


Figure 3-5 Graph of the expression of TUBB4a gene in 10 brain regions from 134 normal individuals was assessed using the Affymetrix Exon 1.0 ST Array. The level of TUBB4a is given as a log scale with range bars. This showed very high expression in the cerebellum

3.3.4 Discussion

I have demonstrated, through a combination of linkage analysis and exome sequencing, that the causative mutation in the DYT4 kindred is a missense R2G variant in exon 1 of TUBB4a. The mutation segregates perfectly with the disease phenotype and is the most plausible candidate from the limited list of novel exome variants located within the linkage region.

3.3.4.1 TUBB4a/Beta 4 Tubulin

Tubulin is a globular protein and the main constituent of microtubules, a major cytoskeletal component. Tubulins are formed from heterodimers of α and β subunits and are expressed in all eukaryotic cells (Mandelkow and Mandelkow, 1994). Multiple isotypes of both subunits are present in vertebrates and there is a high degree of homology between isotypes, which differ only in the last 10-15 amino acids at the C-terminal domain (Sullivan and Cleveland, 1986). These sequence differences have allowed the classification of tubulins into different classes, and which have been shown to be differentially expressed according to the tissue type. Class II, III and IV β-tubulin are all highly expressed in brain with class IV isotype shown to be neuron-specific. Microtubules are co-polymers of the various tubulin isotypes expressed within a particular cell (Lopata and Cleveland, 1987). There is some evidence to support the hypothesis that differential regulation of β-tubulin isotypes is important during the development of the central nervous system in the mammalian brain (Moskowitz and Oblinger, 1995).

3.3.4.2 The TUBB4a p.R2G variant may impair autoregulation of gene transcript

The location of this mutation within the gene is significant as it is located within the highly conserved autoregulatory MREI (Methionine – Arginine – Glutamic acid – Isoleucine) domain of the TUBB4a sequence. The R2G variant results in a glycine at position 2 instead of the wild-type arginine resulting in the tetrapeptide sequence MGEI.

Indeed, all β-tubulin proteins share the MREI tetrapeptide sequence at the start of the N-terminal domain, which has been demonstrated to be necessary for the autoregulated instability of the β-tubulin mRNA transcripts (Yen, Machlin and Cleveland, 1988)

Autoregulated instability is a regulatory mechanism of eukaryotic cells, which, establishes the appropriate level of expression of several genes, and it is based on the principle that mRNA stability is regulated by its own protein product (Cleveland, 1988). The discovery of this mechanism arose from the observation that microtubule depolymerizing agents, causing microtubule disassembly and a consequent rise in the concentration of unassembled tubulin units, led to a significant reduction in β-tubulin mRNA. This was later confirmed with a similar observation in the reduction of mRNA following injection of purified tubulin subunits into cultured mammalian cells (Cleveland, Pittenger and Feramisco, 1983).

The precise nature of the autoregulatory mechanism was determined experimentally by transfection of progressively smaller chimeric β tubulin genes into cultured fibroblasts (Yen *et al.*, 1988). The experiment showed that the first 13 translated nucleotides (encoding the MREI sequence) are sufficient to confer autoregulated instability onto the mRNA transcripts of β-tubulins.

Subsequent studies employing site-directed mutagenesis at codons 2 (Arg) and 3 (Glu), demonstrated that only the wild-type sequence had the ability to confer autoregulation (Yen et al. 1988). The authors hypothesized that the nascent amino acid terminus of the β-tubulin peptide acts as a co-effector in activating a ribonuclease that leads to the degradation of the mRNA transcript.

Expression analysis in fibroblasts collected from affected family members in the DYT4 kindred demonstrated lower TUBB4a mRNA levels in mutation carriers (Lohmann *et al.*, 2013). A recent investigation into the effect of the R2G variant on in-vitro synthesized mRNA transfected into

Deleted: by

neuroblastoma (SH-SY5Y) cells, showed no difference in mRNA stability compared to cells transfected with WT mRNA. This led to the speculation that reduced levels of mRNA seen in cells obtained from DYT4 patients was due to ribosome-related alterations in TUBB4a transcription rather than mRNA autoregulated instability. The precise mechanisms have yet to be conclusively elucidated.

3.3.4.3 Role of cytoskeleton in dystonia pathogenesis

The pathogenesis of primary dystonia is poorly understood in despite the discovery of the 8 or so genes that have been associated with the disorder. Few common pathways have emerged between the dystonia-related genes. The identification of a mutation in TUBB4a neuron-specific cytoskeletal protein is of particular interest in light of increasing evidence of the role of torsin-A in the nucleocytoskeletal network (Atai *et al.*, 2012).

Nucleocytoskeletal connections are important for a variety of neurodevelopmental events including neurogenesis, neural tube closure and neural migration. Torsin-A has been shown to interact with a Nesprin 3a, a protein involved in cytoskeletal network-bridging between the cytoplasm and the nucleus (Nery *et al.*, 2008). Nesprin 3a has been suggested to be a common interacting protein with both torsin-a and dystonin-a2, a plakin protein which interacts with microtubules via a C-terminal microtubule binding domain (Atai *et al.*, 2012).

Following on from our identification of the TUBB4a gene as the cause of DYT4 dystonia, several functional studies have investigated the pathogenic mechanism of the p.R2G variant. Analysis of tubulin network formation in neuronal and non-neuronal cell models expressing the p.R2G variant showed impaired formation of radial tubulin networks and impaired neuronal process growth (Watanabe *et al.*, 2018).

Deleted: the investigating

A detailed study of tubulin dynamics and cellular morphology in a neuronal model transfected with the p.R2G variant has shown the mutation does not affect the steady state of the

expressed TUBB4a protein nor did it affect the rate of tubulin polymerisation. There were clear morphological abnormalities in cerebellar granule neurones but not in oligodendrocytes. The p.R2G mutation may therefore cause defects in tubulin function through mechanisms that are unrelated to the dynamic state of the protein (Curiel *et al.*, 2017). This finding has been replicated in iPS-derived neurons transfected with the R2G variant which demonstrated unaltered microtubule dynamics but reduced microtubule stability and altered mitochondrial motility (Vulinovic *et al.*, 2018).

Deleted:

3.3.4.4 Heterozygous mutations in other beta tubulins are associated with neurological disease

Several human diseases are caused by heterozygous mutations in several genes encoding α and β tubulin isotypes. Missense mutations in *TUBA1A* (class 1a α-tubulin), *TUBB2B* (class 2b β-tubulin) and *TUBB3* (class 3 β-tubulin) have all been reported and result in a range of severe neurological manifestations (Tischfield *et al.*, 2011). Mutations were all reported as occurring de novo.

The phenotypes associated with non-TUBB4a tubulinopathies are variable but typically manifest severe structural brain abnormalities including lissencephaly, pachygyria, polymicrogyria and malformations of cortical development. Certain TUBB3 missense mutations have been associated with congenital fibrosis of the extra-orbital muscles type 3 (CFEOM3) with an otherwise relatively normal brain structural phenotype (Tischfield *et al.*, 2010).

The molecular aetiology of these diseases is the result of a range of effects on cytoskeletal pathways including disrupted binding to chaperone molecules, disrupted GTP binding, impaired motor protein interactions and impaired microtubule biogenesis (Tischfield *et al.*, 2011).

Correct microtubule dynamic and functioning has a fundamental role during brain development for neuron migration, differentiation and axon guidance and the fact that many congenital neurological disorders are caused by mutations in different tubulin genes further supports this view (Wade, 2009).

3.3.4.5 TUBB4a mutations are a very rare cause of isolated dystonia

In this study I screened 124 dystonia cases for mutations in exon 1 and did not identify any further mutations. Two large scale screening studies were undertaken by other groups after our identification of the gene. A large multi-ethnic screening study of 492 isolated dystonia cases did not identify any mutations in TUBB4A (Vulinovic *et al.*, 2017) . In the same study analysis of copy number variants in 336 dystonia patients did not identify any gene dosage variants but did find a potentially pathogenic in-frame deletion in an Italian patient (c.907_909delGAG). In a separate study, no TUBB4a mutations were identified in a population of 709 isolated dystonia patients of German/Austrian ancestry (Zech *et al.*, 2015).

3.3.4.6 Clinical spectrum of DYT-TUBB4A dystonia

A further 12 cases of dystonia associated with TUBB4a (DYT-TUBB4A) mutations have been reported (Bally *et al.*, 2021; Delorme *et al.*, 2021). A recent review examined the clinical phenotype in the 22 cases previously described in the original DYT4 kindred (Parker, 1985; Hersheson *et al.*, 2012; Lohmann *et al.*, 2013) and the 12 subsequent cases (Bally *et al.*, 2022).

Clinically, prominent laryngeal involvement (spasmodic dysphonia) is the most common feature, observed in ~75 % of cases. Whilst spasmodic dysphonia was common in DYT-TUBB4a, screening for TUBB4A mutations in two series of isolated spasmodic dysphonia cases did not identify any mutations (De Gusmão *et al.*, 2016; Putzel *et al.*, 2016). Cranio-cervical dystonia was present in ~60% of cases with the characteristic 'hobby-horse' gait, first described by Parker in the original kindred, present in ~25% of cases.

The age at onset was variable and ranged from early childhood to late adulthood. Similarly, the clinical course was also variable, although the age-at onset- did not predict subsequent severity or progression. Penetrance was very high in the original DYT4 kindred however in the subsequent families investigated, pathological variants were seen in unaffected family members.

Most of the patients had a poor response to medical treatment. 7 patients responded to alcohol, 4 responded to propranolol and 1 responded to tetrabenazine. Botulinum toxin was effective in 3 cases. In 4 cases, globus pallidus internus-deep brain stimulation was effective at treating either spasmodic dysphonia or cranio-cervical dystonia.

The conclusion of this review was that routine screening of TUBB4A in isolated spasmodic dysphonia is not recommended however should be considered in individuals with laryngeal dystonia and dystonia in another location.

3.3.4.7 Mutations in TUBB4a cause a spectrum of neurodegenerative phenotypes

To date, more than 30 heterozygous mutations in TUBB4a have been identified associated with an increasingly broad spectrum of clinical phenotypes including generalised dystonia (DYT4), hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC) (Simons *et al.*, 2013) , progressive spastic paraparesis (Blumkin *et al.*, 2014), isolated hypomyelination (Pizzino *et al.*, 2014) and infantile encephalopathy (Duncan *et al.*, 2017) .

This genotypic and phenotypic heterogeneity will be discussed in more detail in the next sub-chapter

3.3.4.8 Conclusion

The data I presented was the first report of TUBB4a being associated with a human disease. Whilst DYT4 dystonia is rare, the discovery of TUBB4a as a causal gene in both dystonia and

the neurodegenerative H-ABC syndrome has facilitated further research and understanding into the role of microtubule function and physiology in the pathogenesis of these disorders.

3.4 TUBB4a mutation screening in a series of patients with H-ABC

3.4.1 Introduction

Shortly after the identification of a mutation in TUBB4a as the cause of DYT4 dystonia, Simons et al published an investigation into a series of 11 patients with a rare hereditary leukoencephalopathy called "Hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC)". All patients were found to carry the same de novo nonsynonymous variant in TUBB4a (p.Asp249Asn) revealing for the first time the genetic aetiology of this rare leukoencephalopathy.

H-ABC is a neurodegenerative disorder radiologically characterized by leukodystrophy/hypomyelination, cerebellar atrophy, severe atrophy or absence of the putamen. Clinically this is a very severe, childhood-onset disorder associated with developmental delay, progressive generalized spasticity and a range of extrapyramidal movement disorders including dystonia and choreoathetosis. A number of other hypomyelinating leukodystrophies have been genetically defined (Charzewska *et al.*, 2016). Nearly all are associated with a very severe clinical phenotype comprising severe motor and cognitive delay, spasticity, pyramidal and cerebellar signs as is true for H-ABC however the specific radiologically defining features and the presence of prominent dystonia differentiates H-ABC from the other hypomyelinating leukodystrophies.

Together with my colleagues in the movement disorders service (Dr Erro and Prof Bhatia) we assessed the clinical and genetic characteristics of 3 patients with dystonia who met the radiological criteria for H-ABC syndrome.

3.4.2 Methods

3.4.2.1 Case ascertainment and clinical assessment

Patients were identified from a clinical database of dystonia patients who had been reviewed in the movement disorders service at NHNN. All Written consent was obtained by Dr Erro from all participating subjects. Clinical assessment was undertaken by Dr Erro and other clinicians in the movement disorders clinic.

Ethical approval was obtained for the patients in this study (see section 2.2).

3.4.2.2 TUBB4a mutation analysis

DNA was extracted from peripheral leucocytes using standard techniques. Primers were designed, using Primer3, to amplify the exons and adjacent splice sites for the TUBB4a gene. PCR products were sequenced using BigDye Terminator 3 chemistry and analysed using capillary electrophoresis on an ABI 3130XL sequencer.

3.4.3 Results

3.4.3.1 Clinical and radiological results:

Three patients were identified who met the radiological criteria for H-ABC.

3.4.3.1.1 Patient 1

Patient 1 was male with a normal birth history and no family history of any neurological conditions. Motor development was delayed with regard to right upper limb function, however in the first 12 months of life he was thought to have developed normally. There was regression of motor skills from 18 months, and he never achieved any expressive verbal communication although comprehension was preserved. He developed severe generalized dystonia from 3 years, significantly affecting swallow which required percutaneous gastronomy. He remained severely physically disabled and later developed painful dystonic craniocervical and left upper limb spasms.

Deleted: months

Examination in clinic at age 25 years showed normal voluntary eye movements but frequent oculogyric episodes. There was markedly increased axial tone, dystonic posturing of all four limbs and bilateral fixed flexion deformity of the knees due to contracture. There were pyramidal signs with sustained clonus at the right ankle.

His neurological disability progressed, and he suffered recurrent lower respiratory tract infections. He was initiated on non-invasive ventilation age 28 years and died from respiratory complications age 31 years.

MRI brain showed typical appearances of H-ABC with supratentorial hypomyelination, atrophy of the basal ganglia and cerebellum.

3.4.3.1.2 Patient 2

Patient 2 was male with a normal birth and early developmental history however ambulation was delayed, and he had frequent falls as a young child. Verbal and communication development was normal. After age 7, his gait, limb function and speech deteriorated such that by 16yrs he required a wheelchair to mobilise.

He was examined at age 40 and exhibited severe generalized dystonia. Verbal comprehension was preserved however he had a severe spasmodic dysphonia and was virtually anarthic with orofacial spasms on attempting to speak. There was bilateral leg weakness and fixed flexion at the knees due to contracture. Reflexes in the lower limbs were very brisk with extensor plantar reflexes and sustained ankle clonus.

MRI brain showed typical appearances of H-ABC.

3.4.3.1.3 Patient 3

Patient 3 was female with normal birth and early developmental history and no family history of neurological disorders. Whilst there was no motor delay, she started having falls from age 3

and difficulty writing from age 4. Her speech deteriorated from age 3 years with a 'whispering' quality to her speech and difficulties with phonation. Her dystonic symptoms progressed to a generalized dystonia from age 10 predominantly affecting gait, right arm function and speech. When examined age 10 she had upper limb cerebellar signs although no pyramidal signs. She was reassessed at age 28 and demonstrated severe generalized dystonia, pyramidal signs with brisk reflexes and extensor plantar reflexes and severe dys/aphonia with orolingual dystonia.

MRI showed typical appearances of H-ABC although the degree of basal ganglia and cerebellar atrophy was less marked than in the first two cases.

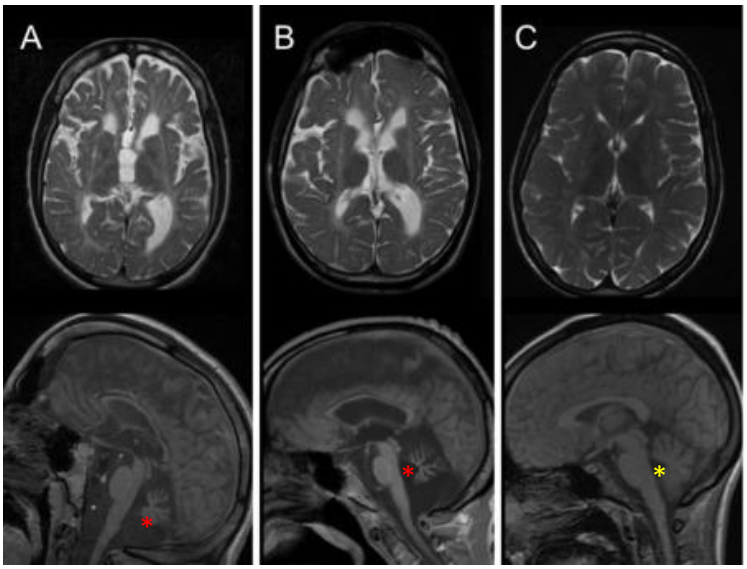


Figure 3-6 Axial and sagittal MR images showing typical features of HABC in cases 1 (A) and 2 (B) with hypomyelination, atrophy of the basal ganglia and cerebellum. The severe cerebellar atrophy is marked with a red asterisk. A milder radiological phenotype with less marked hypomyelination and basal ganglia atrophy preserved cerebellar volume (marked with a yellow asterisk) is seen in case 3 (C). Adapted from published figure (Erro *et al.*, 2015)

3.4.3.2 Genetic results

In all three patients, heterozygous mutations in TUBB4a were identified. In patients 1 & 2, a heterozygous c.745G>A; p.Asp249Asn variant was identified which matches the canonical

variant first reported to be associated with the disorder. Both variants were de novo mutations and were not identified in either parent of the two affected individuals.

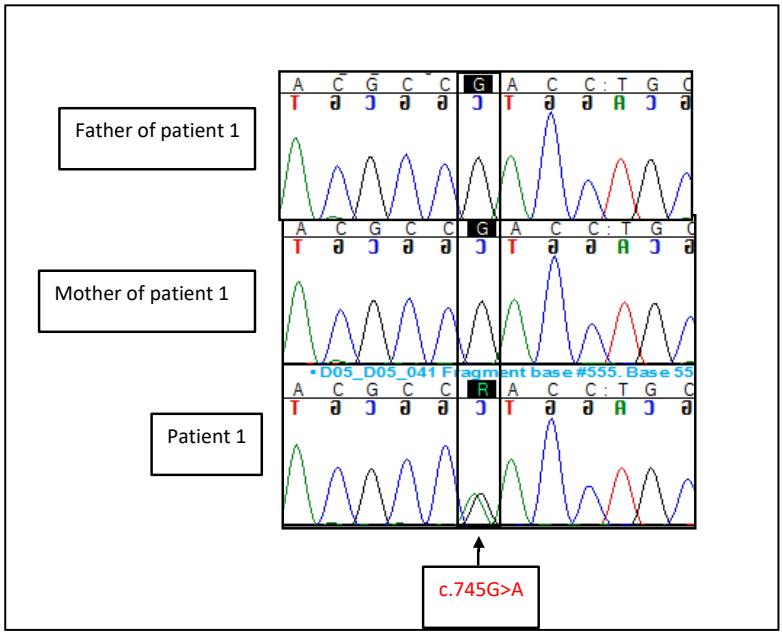


Figure 3-7 Sequence chromatogram showing the c.745G>A p.Asp249Asn de novo variant in patient 1 but absent in both parents.

In patient 3 a heterozygous de novo c.941C>T; p.Ala314Val variant was identified. This mutation was predicted to be pathogenic by multiple software prediction tools (SIFT, Mutation Taster) and was absent from the NHLBI ESP Exome Variant Server and the 1000genome project databases.

3.4.4 Discussion

Here are described 3 individuals with typical radiological and clinical features of H-ABC syndrome. There are some striking similarities between the H-ABC cases and the DYT4 patients, most notably the generalized dystonia and severe dysphonia/anarthria. H-ABC cases however are associated with a much more severe neurodegenerative phenotype with progressive atrophy in white matter tracts, basal ganglia and the cerebellum consistent with

the clinical symptoms and signs. In the individuals with DYT4 with documented brain imaging, no significant changes were noted which is more consistent with the findings in primary dystonia (Wilcox *et al.*, 2011).

The most common mutation associated with H-ABC is the canonical p.Asp249Asn variant which accounts for around 40% of all reported cases (Hamilton *et al.*, 2014). It is typically associated with a milder phenotype than cases associated with other TUBB4a mutations. For p.Asp249Asn cases, the median age at onset was 1.5 years (3 months- 3years). Ambulation either independently or with support was achieved by 96% of individuals with p.Asp249Asn compared to 12% being able to stand or walk with support for other mutations. The clinical severity in general reflected more severe loss or lack of myelin from an early stage.

Histopathological studies of H-ABC cases demonstrate near total loss of neuronal cells and subtotal degeneration of the putamen (Van Der Knaap *et al.*, 2007). In contrast the thalamus and globus pallidus were largely unaffected. Cerebellar cortical neuronal loss, particularly in the granule layer, was noted. Loss of myelin indicated by imaging was confirmed histopathologically and is thought to reflect both hypomyelination and progressive myelin loss.

The pattern of cell loss may in part be explained by the differential expression measured in our study investigating the DYT4 kindred (see figure 3-5). Expression of TUBB4a was highest in the cerebellum followed by the white matter and within the basal ganglia structures, expression was highest in the putamen.

3.4.4.1 Is there an explanation for divergent phenotypes associated with TUBB4a mutations?

It is intriguing that whilst there are some clinical similarities between DYT4 dystonia and H-ABC, there is no evidence of neurodegeneration on the available imaging studies for selected DYT4 patients (Wilcox *et al.*, 2011), in contrast to H-ABC in which neurodegeneration in the cerebellum and basal ganglia is pathognomonic. As previously mentioned, TUBB4a mutations

are known to be associated with a range of phenotypes in addition to DYT4 and H-ABC to include isolated hypomyelination and early infantile encephalopathy.

Histological, cell and animal models support the hypothesis that the divergence in phenotypes may be explained by the extent to which the underlying pathology affects either neuronal or glial cells. Of the conditions above, DYT4 is perhaps the most distinctive in exhibiting an adult-onset phenotype without any radiological evidence of either neurodegeneration or hypomyelination. Cell models using both cerebellar granule neurons and oli-neu cells (derived from primary mouse oligodendrocytes) transfected with R2G showed abnormal morphology of only neuronal cells and no effect on the expression of myelin genes (Curiel *et al.*, 2017). Conversely when transfected with the canonical D249N variant, abnormal neuronal *and* oligodendrocyte morphology with significant reduction in myelin gene expression. For variants associated with isolation hypomyelination and infantile encephalopathy, a purely glial pattern was seen with abnormal oligodendrocyte morphology and myelin expression but normal neuronal morphology.

Protein structural analysis and molecular modelling of TUBB4A mutants confirms the importance of the MREI tubulin domain due to the observation that the D249N residue is structurally adjacent to the n-terminal MREI domain and may interact via a salt bridge (Simons *et al.*, 2013). Of note two other mutations within the MREI domain have been reported (R2Q and R2W) which are both associated with an H-ABC phenotype suggesting that even very subtle changes in microtubule function can result in quite different effects on vulnerable neuronal and oligodendroglial cell populations (Krajka *et al.*, 2022).

There is strong evidence that microtubule dynamics and function are affected by mutations in TUBB4a. In a cell model, the R2G variant showed no effect of the expressed mutant protein on binding of motor protein (KIF-5) to microtubules (Vulinovic *et al.*, 2018). The D249N variant showed the reduced binding of KIF-5 compared to controls. Subsequent analysis of

Deleted: t

mitochondrial transport showed increased motility for both variants. It was hypothesized that as mitochondrial transport is an ATP-consuming proves that this may alter the energy balance of the cells and increase cellular stress resulting in the observed cellular phenotype. In a cell model of transfected HELA cells, microtubule dynamics assays showed no changes to polymerization of beta 4 tubulin in mutants transfected with R2G and abnormal (increased) polymerization in D249N mutants.

4 Identification of CAPN1 as cause of recessive spastic ataxia

4.1 Statement of contribution

The index family (Family 1) was under the care of the neurogenetics service at NHNN. I undertook a clinical revaluation of the index case, exome sequencing and homozygosity mapping based on genotype data obtained by Dr Katherine Fawcett. I collected fibroblasts from the index patient in order to undertake cDNA sequencing. I performed the protein sequence modelling of the mutant protein sequence. My initial findings of a probable causative homozygous mutation in the CAPN1 gene led to an international collaboration with Prof Michel Baudry and colleagues at Western University of Health Sciences who had been assessing learning and memory in a CAPN1 knockout mouse model. His team performed CAPN1 expression studies and activity assays in the patient fibroblasts and control samples.

Additional cases (Families 2-4) were identified from the Department of Molecular Neurobiology, National Institute of Neurology, Tunisia and sequenced at the National Institute of Health, Bethesda, USA (Monia Hammer, Faycal Hentati and Rim Amouri) and also by colleagues at Centre de Reference de Neurogenetique, Hopital de la Pitie-Salpetriere, Paris. (Alexis Brice, Giovanni Stevanin)

At the time of the initial study and subsequent collaboration, CAPN1 was not associated with human Mendelian disease however prior to publication of our work in June 2016, a study was published in May 2016 of two families with complex HSP associated with CAPN1 mutations. We were as they say: 'pipped to the post'.

Publications arising from this section:

- Defects in the CAPN1 Gene Result in Alterations in Cerebellar Development and Cerebellar Ataxia in Mice and Humans. Wang Y, **Hersheson J**, Lopez D, Hammer M, Liu Y, Lee KH, Pinto V, Seinfeld J, Wiethoff S, Sun J, Amouri R, Hentati F, Baudry N, Tran J, Singleton AB, Coutelier M, Brice A, Stevanin G, Durr A, Bi X, Houlden H, Baudry M. Cell Rep. 2016 Jun 28;16(1):79-91

4.2 Introduction

For many years spinocerebellar ataxias and hereditary spastic paraplegias have been separated and defined by clinico-genetic disease classifications. Most commonly these disorders were designated 'pure' if the phenotype comprised only cerebellar ataxia (or spastic paraplegia respectively or 'complex' where additional phenotypes such as spasticity/myelopathy (in the case of cerebellar ataxia), cerebellar ataxia/atrophy (in the case of HSP), neuropathy, epilepsy, parkinsonism, white matter disease or a range of other clinical phenotypes. Taken together there are approximately 145 genes associated with dominant or recessively inherited ataxia and HSP both 'pure' and 'complex' (*Washington Neuromuscular Website*). It is perhaps more helpful to consider that both cerebellar dysfunction and degeneration and corticospinal tract dysfunction/spasticity are both features that reflect a shared vulnerability of the cerebellum and corticospinal tracts and that these conditions exist on a phenotypic continuum.

Figure 4-1 Photograph of a Parson's Russell Terrier (source: Wikipedia

A mutation in CAPN1 was first reported as a cause of neurological disease in dogs in 2013 (Forman, De Risio and Mellersh, 2013). Spinocerebellar ataxia had been known to affect certain breeds of dog including Parson's Russel Terriers (PRTs) since the 1950s. Foreman et al used a whole genome association study of affected PRTs combined with next generation

sequencing of the disease-associated region to identify a homozygous missense mutation in CAPN1.

Here I describe the clinical and genetic analysis of two individuals from a consanguineous family with a spastic-ataxia phenotype, together with functional studies in patient fibroblasts and additional cases of CAPN1 associated spastic ataxia identified by collaborators to demonstrate that CAPN1 mutations are a novel cause of cerebellar ataxia in humans.

4.3 Subjects, materials and methods

4.3.1 Subjects

The index patient was recruited from the neurogenetics clinic at NHNN. DNA from her affected 1st cousin was available for the analysis. Her sister was not clinically assessed but relevant information was provided by the index patient. Additional cases (Families 2-4) were identified by Dr Monia Hammer from the National Institute for Health, Maryland and National Institute of Neurology, Tunisia and by Alexis Brice at Centre de Référence de Neurogénétique, Salpetriere.

4.3.2 Homozygosity mapping

For family 1, whole genome genotyping was performed Illumina CytoSNP arrays, and the genotype data was used for homozygosity mapping to identify shared regions of homozygosity. Homozygosity mapping was performed using HomozygosityMapper. The corresponding regions of shared homozygosity were used in subsequent variant filtering steps for exome sequencing analysis.

Deleted: arrays

4.3.3 Exome sequencing

Whole exome sequencing on patients R1 and R2 was performed as described in section 4.14. Respective nucleotide and protein positions for CAPN1 are based on the following accession

numbers from the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) NM_001198868 and NP_00185797.

4.3.4 cDNA sequencing

To test the effect of the c.337+1 G>A splice-site variant, messenger RNA (mRNA) was extracted from the peripheral blood of patient R1 using the PAXgene Blood RNA System (Qiagen/PreAnalytix). Following RNA extraction, complementary DNA (cDNA) was synthesised using the Applied Biosystems High-Capacity cDNA reverse transcription kit. The transcription products were separated by agarose gel electrophoresis and sequenced using intronic primers which flanked the exon 3 splice sites. Variant positions within the cDNA are numbered using the A of the translation initiation codon as position 1.

4.3.5 In-silico homology modelling of mutant CAPN1 protein

Sequencing of the cDNA products showed that the splice mutation results in retention of part of the intron leading to an in-frame insertion of 9 amino acids. The 3-D structure of the mutant amino acid sequence was modelled using Swiss-Model, which is an automated protein structure homology-modelling server, using the 1KFX protein crystal structure file from the Protein Data Bank in Europe (PDBe). The results were visualised using PyMol (Schrodinger).

4.3.6 Fibroblast biopsy and culture

With informed consent, fibroblasts were isolated from a skin punch biopsy from patient R1. The fibroblasts were cultured in Dulbecco's modified Eagle's medium GlutaM X supplemented with 10% heat-inactivated foetal bovine serum and 1% penicillin-streptomycin.

4.3.7 Calpain-1 western blot in patient fibroblasts (Baudry lab)

Human fibroblast lysates were homogenized in RIPA buffer (89901, Thermo) with protease inhibitor cocktail (78446, Thermo) and 1mM EGTA. Protein concentrations were measured using the BCA protein assay kit (Thermo), Sample buffer was added, and samples were

separated in 8% SDS-PAGE and transferred onto PVDF membrane (IPFL00010, Millipore). After blocking with Odyssey Blocking Buffer (LiCOR) for 1hr, membranes were incubated with primary antibodies overnight at 4°C and then with secondary antibodies for 2h at room temperature. Protein bands were visualised with the Odyssey imaging system (LI-COR). The primary antibodies used were calpain-1) 1:1000, 2556, CST), calpain-2 (1:1000, LSB12657, LSBio), PHLPP1 (1:1000, 07-1341, Millipore), phosphoAkt Ser473 (1:3000, 4060, CST) and Akt (1:2000, 2920). The secondary antibodies were IRDye secondary antibodies (1:10000, LI-COR)

Ratios of the level of PHLPP1 to actin, and pAkt to Akt were calculated. The experiments were replicated four times from difference cell batches. P values were calculated using one-way ANOVA followed by a Bonferroni test

4.3.8 Calpain-1 activity assay in patient fibroblasts (Baudry lab)

The activity of calpain-1 was measured based on the rate of hydrolysis of the fluorogenic substrate Suc-Leu-Tyr-AMC by calpains in human fibroblast lysates. 250ug of fibroblast lysates (200µl) were incubated with Suc-Leu-Tyr-AMC (100uM) in the reaction buffer (pH 7.4, containing 25mM Tris, 145mM NaCl and 1mM EGTA in the presence or absence of 200nM C2I. The reaction was initiated by adding 1mM (20 µM free Ca^{2+}) or 3mM (2mm free Ca^{2+} + $CaCl_2$ and continued at 30°C for 20 mins, while the fluorescence of 7-amino-4-methylcoumarin (Excitation 380 nm/Emission 450nm) was measured ever 30secs in a POLARstar Omega fluorescence polarisation microplate reader (BMG Labtech). The rate of hydrolysis (increase in fluorescence/second was determined by the linear portion of the curve.)

4.4 Results

6 patients from 4 families with homozygous or compound heterozygous mutations in CAPN1 were identified.

4.4.1 Family 1

The index patient was 43 years old at the time of her most recent assessment. She was born in Bangladesh. Her parents were 1st cousins, and she was one of 7 siblings and was the only affected individual from her immediate family. She has three unaffected children. Her maternal 1st cousin who is also the product of a 1st cousin marriage was also thought to be affected with balance problems and speech difficulties. The index patient had a normal birth and early developmental history. She had an onset of gait difficulties in her 20s with frequent falls. She later developed upper limb ataxia with difficulty with fine dexterity. Dysarthria has been present since her late 20s and now has dysphagia and weight loss. She had two presentations with multiple cerebral infarcts affecting the right and left insular areas and right occipital cortex. She was investigated at the time with a full vascular work-up including trans-oesophageal echocardiogram, thrombophilia screen and cerebral angiogram but no cause was identified.

Her symptoms have been slowly progressive, and she now can only mobilise in a wheelchair. Examination showed slow saccades and a severe bulbar dysarthria. There was upper limb and gait ataxia. Reflexes were brisk in the upper and lower limbs and there was moderate lower limb spasticity.

MRI showed mild cerebellar atrophy. Nerve conduction studies and electromyography showed no evidence of neuropathy or myopathy. Prior to exome sequencing, she had undergone genetic testing in clinic and was negative for SCA 1, 2, 3, 6, 7, 12, 17, FRDA, AOA1, AOA2, ATM and common nuclear mitochondrial mutations (tested in DNA from peripheral leucocytes).

Deleted: cousins

Deleted: been

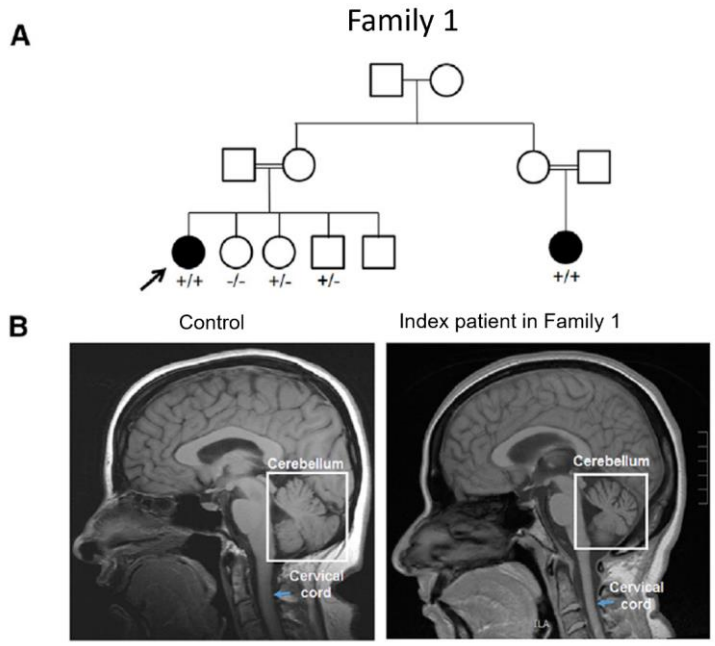


Figure 4-3 (A) Pedigree of family 1. Affected individuals are marked in black. (B) Sagittal MR images showing atrophy of the cerebellum in the index patient (reproduced from Wang et al 2016)

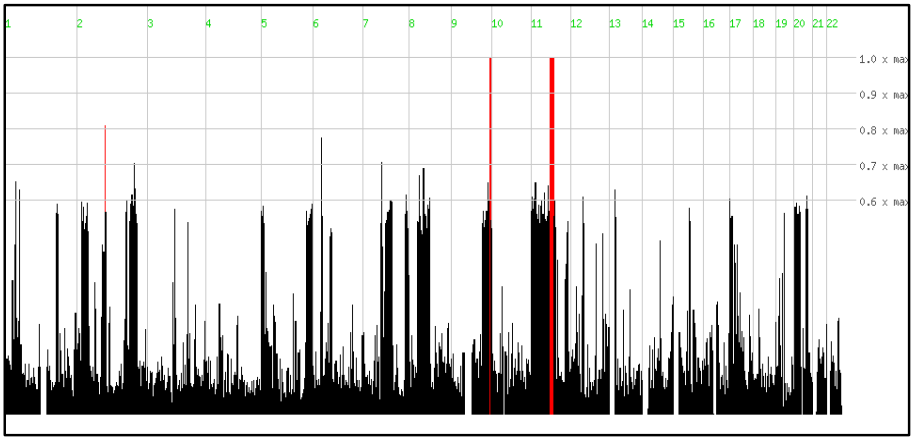


Figure 4-2- Graphical representation of homozygosity mapping analysis using genotype data for both affected individuals. The red lines indicate stretches of shared homozygosity

4.4.1.1 Homozygosity mapping

Figure 4-3 shows regions of shared homozygosity were identified from whole genome SNP arrays using Homozygosity mapper. 3 regions of shared homozygosity were identified on chromosome 2 (96904127-97554706 = 650kb), chromosome 9 (133037273 – 136458404 = 3.4Mb) and chromosome 11 (63826317 – 76852039 = 13.0 MB).

4.4.1.2 Exome sequencing

Exome sequencing was carried out to a depth of 50x coverage on average. Variants were filtered according to exclude synonymous variants, heterozygous variants, variants outside of the defined regions of shared homozygosity and variants present in publicly available databases (dbSNP version 129, 1000 Genomes project, NHLBI Exome Variant Server and Complete Genomics 69).

3 homozygous variants remained after filtering against the above criteria all located on chromosome 11: CAPN1 c.337+1G>A, CORO1B c.C488G:p.P163R and GPR152 c.G659A:p.C220Y.

CORO1B codes for coronin 1B which is a ubiquitously expressed protein that binds to actin and interacts with microtubules. The gene was not associated with any known mammalian disease phenotype. GPR152 codes for a probable G-protein coupled receptor 152 and mainly expressed in lymphoid tissue. The p.C220Y amino acid residue change was at a poorly conserved site in the protein sequence. The CAPN1 variant was the most interesting as this would be expected to interfere with normal splicing of the gene product.

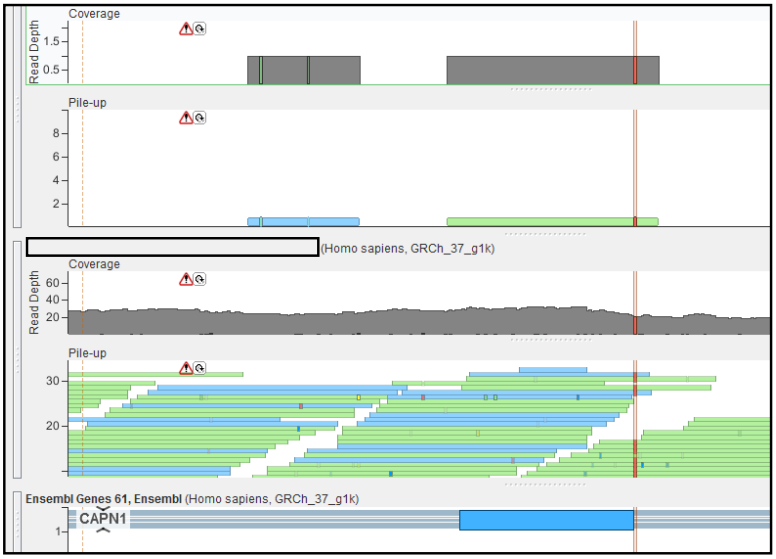


Figure 4-4 .BAM file alignment from exome sequencing showing the homozygous CAPN1 c.337+1G>A splice variant identified in the index patient from Family 1

4.4.1.3 Analysis of effect of CAPN1 c.337+1G>A on splicing

To investigate the effect of the c.337+1G>A on splicing, primers were designed to amplify cDNA spanning the CAPN1 exon 3 splice site. cDNA sequencing yielded two transcripts – one which led to utilisation of a cryptic 5' splice site resulting in an in-frame insertion of 9 amino acids. The second transcription product resulted in skipping of exon 3 and a subsequent frame shift mutation leading to a significantly truncated protein product.

Of particular interest and relevant to the decision to pursue further investigation of this gene as a (then) novel gene causing cerebellar ataxia, was the data from Forman et al identifying a missense C115Y mutation as the cause of a recessive ataxia syndrome in Parson's Russell Terriers. The C115 residue is a highly conserved site and forms part of a triad of residues comprising the active site of the calpain 1 protein. Residue 115 is highlighted in red above adjacent to the 9 amino acid insertion identified from sequencing of one of the CAPN1 c.337+1G>A splice transcripts.

4.4.1.4 Structural modelling of insertion in CAPN1 protease domain

The crystal structure of CAPN1 has been experimentally determined which allows for much more accurate structural modelling of sequence variants. Figure 4-5 shows the wild-type protein structure with the active site cleft indicated with a yellow *. The second image shows 9 amino acid insertion loop indicated by the yellow arrow. This may suggest that if the mutant cDNA sequence were translated, the amino acid may sterically interfere with the active site and impair normal enzymatic function.

MSEEIITPVYCTGVSAQVQKQRARELGLGRHENAIKYLGQDYEQLRVRCLQSGTLFR
DEAFPPVPQSLGYKDLGPNSSKTYGIKWKRPTELLSNPQFIVDGATRTDICQGAL[D
RPPGGRVP]GD**C**WLLAAIASLTLNDTLLHRVVPHGQSFQNGYAGIFHFQLWQFGE
WVDVVVDDLLPIKDGKLVFVHSAEGNEFWSALLEKAYAKVNGSYEALSGGSTSEGF
EDFTGGVTEWYELRKAPSDLYQIILKALERGSLLGCSIDISSVLDMEAITFKKLVKGHA
YSVTGAKQVNYRGQVVSLIRMRNPWGEVEWTGAWSDSSSEWNNVDPYERDQL
RVKMEDGEFWMSFRDFMREFTRLEICNLTPDALKSRTIRKWNTTLYEGTWRRGST
AGGCRNYPATFWVNPQFKIRLDETDDPDDYGDRESGCSFVLALMQKHRRRERRF

Figure 4 Protein sequence of human calpain-1 showing the in-frame insertion (red text) that results in the loop structure around the active site seen below in figure 6-6. The active site residue is shown in bold text.

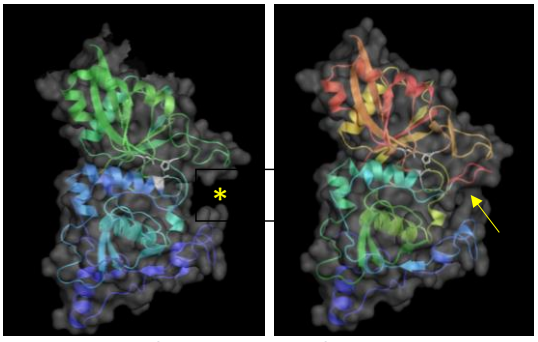


Figure 4-5 3D render of structural modelling of WT and mutant CAPN1 protein sequence. The arrow indicates the additional loop within the active site cleft (marked with yellow asterisk) that may interfere with normal enzyme activity.

4.4.2 Calpain-1 expression and calpain activity in patient fibroblasts (Baudry lab)

Levels of calpain-1, calpain-2, PHLPP1 and phospho-Akt Ser 473 (pAKT) were assessed in fibroblast lysates from the index patient in family 1 and two control samples. PHLPP1 is a specific substrate of calpain-1, which can phosphorylate and thus activate Akt is an important part of a signal transduction pathway associated with neuronal survival. Activation of Akt is associated with increased neuronal survival.

Calpain-1 protein expression was absent on western blot in patient fibroblasts compared to controls. This suggests that either the transcript which resulted in exon 3 skipping and a subsequent stop codon was the dominant transcript or else the extended Calpain-1 protein with the 9 amino acid insertion was unstable and degraded rapidly after synthesis. Calpain-2 protein levels were the same in the patient as the two control fibroblast samples.

The level of PHLPP1 was slightly increased compared to controls but this was not statistically significant. Levels of pAkt were significantly decreased in the patient fibroblasts.

Calpain-1 and 2 activities were measured by determining the hydrolysis rate of a fluorescent substrate Suc-Leu-Tyr-AMC in the presence of two difference concentrations of Ca2+ - 20uM (to activate Calpain-1) and 2mM (to activate Calpain-2). Calpain-1 activity was present in all control fibroblasts but was absent in patient fibroblasts. Calpain-2 activity was present in all control and patient samples.

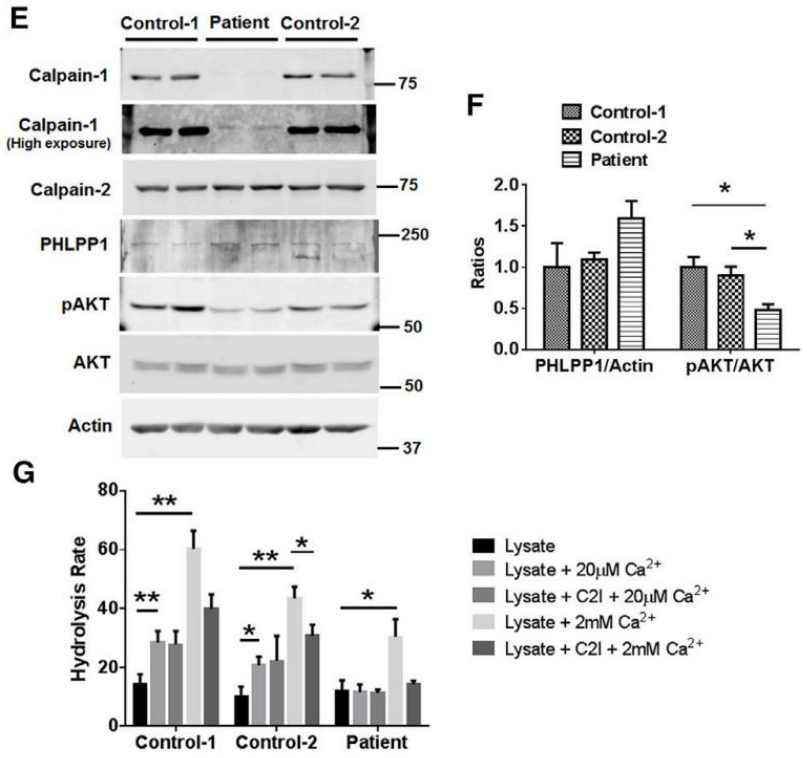


Figure 4-6 E) upper two Western blots show virtually absent calpain1 expression in patient controls F) Bar-chart showing signficantly reduced pAKT/AKT expression in patient fibroblasts from index case in family 1 – note signficantly reduced pAKT/AKT compared to controls G) bar chart showing hydrolysis rate of calpain substrate – no increase in hydrolysis rate is observed with activation of calpain 1 with 20uM Ca 2+ indicating absent calpain-1 activity. Calpain-2 activity is normal in patient samples – adapted from Wang et al 2016).

4.4.3 Family 2

The pedigree for this family in shown in figure 6-7 A. The two affected sisters IV:5 and IV: 6 are Tunisian and are the product of consanguineous parents. They both presented with gait impairment at age 23 and 20 yrs. respectively. Both had gait ataxia on examination, cerebellar dysarthria, spasticity and brisk tendon reflexes in the upper and lower limbs. Both had finger-nose and heel shin incoordination. Sensory examination was normal. MRI showed mild cerebellar atrophy.

Homozygosity mapping identified 13 homozygous regions including one on chromosome 11 spanning rs11602911 (chr11: 63596063 bp) to rs2290959 (chr11: 67410314 bp). Within that region, a homozygous variant was identified in CAPN1 - c.1534C>T p.R512C. This variant is absent in the 1000 genomes project, dbSNP and Exome Sequencing Database and is predicted to be damaging by in silico prediction tools (Polyphen2, SIFT). The variant is at a highly conserved amino acid site.

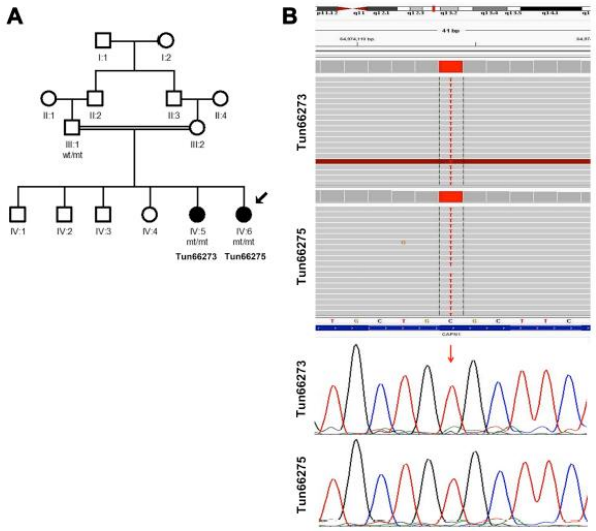


Figure 4-7 A) Pedigree of Family 2 B) .BAM alignment showing homozygous - c.1534C>T p.R512C variant and sequence chromatogram confirming exome result (reproduced from Wang et al 2016)

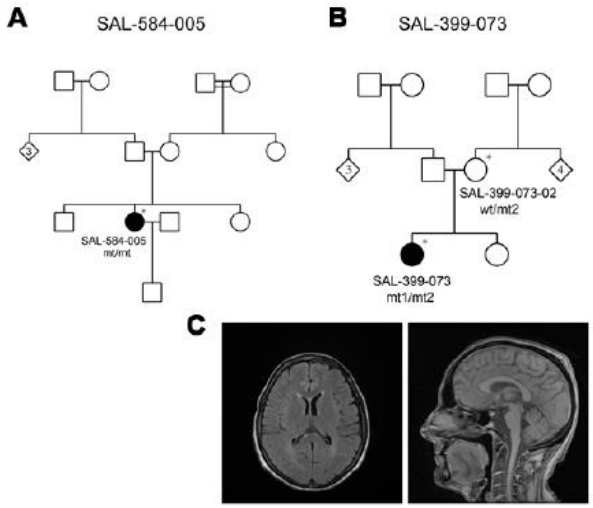


Figure 4-8 A) Pedigree of family 3 B) Pedigree of Family 4 C) MRI of index case in family 4 showing moderate cerebellar atrophy (reproduced from Wang et al 2016)

4.4.4 Family 3

The proband was an Italian 67-year-old woman. The pedigree can be seen in figure 4-8 A. Her parents were from the same village and consanguinity could not be excluded. She developed walking difficulties at age 25, required a walking stick at age 44 and used a wheelchair by age 52. Neurological examination at age 52 showed spastic paraparesis with upper limb ataxia, cerebellar dysarthria and mild dysphagia. She had generally brisk limb reflexes with bilateral Hoffman's and extensor plantar responses. MRI at age 48 and nerve conduction studies were normal.

Exome sequencing identified a frameshift mutation in CAPN1 - c.183dupC p.F61fs, which was confirmed on Sanger sequencing.

Deleted: 6

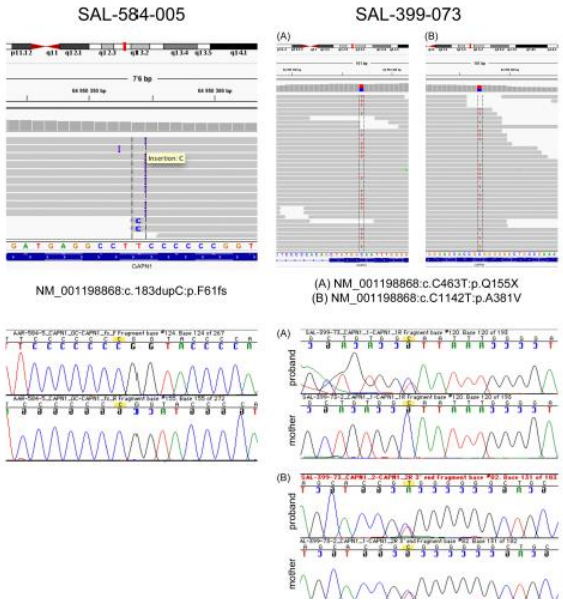


Figure 4-9 Exome alignment showing p.F61fs mutation in family 3 and p.Q155X; p.A381V compound heterozygous mutations in family 4 (reproduced from Wang et al 2016)

4.4.5 Family 4

The proband was of French origin. She was 46 years old at the time of her last assessment. Her maternal grandparents were 1st cousins. There was no family history of neurological disease. She developed progressive gait imbalance and speech difficulties from age 20. She required a wheelchair to mobilise from age 30. On examination she had limb and gait ataxia and cerebellar dysarthria. Oculomotor examination showed saccadic pursuit movements and slow horizontal and vertical saccades. There were generally brisk reflexes with spastic paraparesis associated with bilateral Hoffman's sign, sustained ankle clonus and extensor plantar responses. There was distal upper limb wasting with mild weakness. MRI showed moderate cerebellar atrophy and mild white matter changes on FLAIR weighted images. Nerve

conduction studies showed a mild axonal sensory peripheral neuropathy. Muscle biopsy was normal.

Exome sequencing identified two heterozygous mutations - p.A381V and p.Q155X. Sequencing of the mutations in the patient's mother showed that the two mutations were compound heterozygous i.e. in trans. The A381V missense mutation is predicted to be deleterious by SIFT and Polyphen and is at a conserved amino acid site. The Q155X mutation is a truncating mutation and is presumed pathogenic.

4.5 Discussion

I report the identification of a novel gene associated with spastic ataxia – *CAPN1*. 4 families carrying likely pathogenic mutations in *CAPN1* are reported here and all share a similar phenotype of adult-onset gait ataxia and spasticity. The journey for this discovery led from the neurogenetics clinic at NHNN, via the canine genetics laboratory at the now sadly closed Animal Health Trust in Newmarket (where I met the team responsible for the identification of *CAPN1* as the cause of ataxia in Parson's Russell Terriers), to an international collaboration with Michel Baudry's team at Western University of Health Sciences, California and finally to our other international collaborators who identified further families with CAPN1-associated spastic ataxia.

Just prior to publication of our work, *CAPN1* was reported in several families with very similar phenotypes to those reported here (Gan-Or *et al.*, 2016) although these families were described as having a complex HSP phenotype rather than a spastic ataxia however as previously discussed in the introduction, this distinction is moot and it is probably better to consider that these phenotypes exist on a spectrum between ataxia and HSP. As such I will refer to the neurological phenotype as CAPN1-neurodegeneration and will discuss the phenotypic spectrum of CAPN1 neurodegeneration further in this section.

Calpains are ubiquitously expressed calcium-dependent non-lysosomal cysteine proteases. The two main isoforms are mu-calpain and m-calpain, so named because of the micromolar or millimolar concentration of Ca2+ required for activation and are referred to as calpain 1 and 2 respectively (Goll *et al.*, 2003). Calpain-1 is known to have an important role in synaptic long-term potentiation and plasticity. In mammalian species calpain activity is highest in the cerebellum and brainstem compared to other brain areas (Wang *et al.*, 2013, 2020).

The data presented here, from functional studies of CAPN1 activity in fibroblasts from the index family, show that the splice mutation results in an almost complete abrogation of Calpain-1 activity and subsequent effects on downstream pathways (PHLPP1/Akt) that may provide an explanation for the neurodegenerative phenotype observed.

The data presented by Forman et al. on the PRTs were strongly supportive of the role of CAPN1 in causing cerebellar degeneration in dogs (Forman, De Risio and Mellersh, 2013). Although functional studies of the homozygous C115Y active site mutation or histopathological studies were not carried out in the affected animals, the presence of a mutation at the highly conserved active site is sufficiently compelling to infer that the variant would result in the ataxia phenotype seen in the dogs.

4.5.1 Cerebellar phenotype in calpain-1 knockout mice

Prior to our collaboration, Prof Baudry's team was focussed on examining the role of calpains in mediating long-term potentiation and synaptic plasticity. Cerebellar degeneration in calpain-1 knockout mice had not been explicitly assessed however our work and collaboration prompted a reappraisal of the cerebellar clinical and neuropathological phenotype in these mice.

Calpain-1 knockout mice exhibited a clinical ataxic phenotype with shorter latency to falling on rotarod assessment and shorter stride length compared to wild-type controls. These mice also

exhibited enhanced apoptosis of cerebellar granule density and altered Purkinje cell synaptic transmission (Wang *et al.*, 2016). The knockout mice also showed similar alterations in PHLPP1 and Akt levels to those observed in the fibroblast lysates from the index patient indicating that the mouse model was a good molecular model for CAPN1-neurodegeneration in humans. Of particular interest was the rescue of the observed cerebellar granule loss and motor deficits through activation of Akt using a PTEN inhibitor (bisperoxovanadium) which resulted in increased Akt levels and provides support for a role of this pathway in the pathogenesis of CAPN1-associated neurodegeneration.

4.5.2 Role of calpain-1 in neurodegeneration

In recent years, the role of the calpain system in both neuroprotection and neurodegeneration has been further evaluated. Prior to our report, calpain-1 *over*activation had been associated with cell loss in conditions such as Alzheimer's disease and traumatic brain injury. More recent studies indicate that calpain-1 activation may be a response to and neuroprotective against cell death (Wang *et al.*, 2018). Indeed previous work in Alzheimer's disease showing an association of calpain-1 overactivity in Alzheimer's disease brains at Braak stages III and IV, this was not seen in earlier Braak stages and may suggest that calpain-1 activation occurs secondary to cellular damage (Kurbatskaya *et al.*, 2016). This is important as calpain-1 inhibitors have been suggested as a possible treatment for Alzheimer's disease, an approach which may prove to be misguided (Fa *et al.*, 2015). Calpain-1 deficient animal models in humans, dogs, mice, zebrafish, fruit flies and Caenorhabditis elegans have now been shown to be associated with a neurodegenerative phenotype (Gan-Or *et al.*, 2016).

As demonstrated in the Calpain-1 knockout mouse models, activation of the PHLPP1/Akt pathway appears to be neuroprotective and reduction in Akt levels is one of the mechanisms through which neuronal loss is mediated. PHLPP1 negatively regulates the Akt and Erk

pathways. Disruption to these pathways has been associated with several neurodegenerative conditions including Alzheimer's, Parkinson's and Huntington's disease (Rai *et al.*, 2019).

The calpain system is also implicated in other pathways associated with cerebellar ataxia. Studies investigating the role of type 1 metabotropic glutamate receptors (mGluR) in synaptic plasticity demonstrated that calpain-1 deletion (in calpain-1 knock-out mice) impairs mGluR-dependent long-term potentiation (Zhu *et al.*, 2017). Mutations in GRM1 which codes for the mGluR1 receptor are associated with early-onset or congenital ataxia in mouse (crv4) and dog (Coton de Toulear) models and in humans (Guergueltcheva *et al.*, 2012)

4.5.3 Clinical spectrum associated with CAPN1 mutations

Over 80 patients with CAPN1-associated neurodegeneration have now been reported. All patients have either compound heterozygous or homozygous mutations in CAPN1. Patients typically present in teens to early adulthood. Most patients would be considered either to be 'complex HSP' or 'spastic ataxia'. The clinical spectrum was reported in a summary of 83 reported cases (Méreaux *et al.*, 2021). Lower limb-spasticity is the most commonly reported clinical feature and is seen in 100% of patients. Cerebellar ataxia (either gait or limb ataxia, dysarthria, dysphagia or cerebellar eye signs) was present in 85% of cases. A pure HSP phenotype was estimated to be seen in approximately 36% (range 26-47%) however this is likely to be an overestimate. Spasticity may well be an early presenting feature and so the timing of the assessment is important in determining whether there is any evidence of cerebellar involvement clinically or radiologically. Skeletal abnormalities were reported in 16/37 (43%) of cases. Peripheral sensory impairment was seen in 12/53 (23%) of cases. Neurophysiology in a limited number of cases showed sensory or sensorimotor axonal neuropathy.

The overall frequency of CAPN1 mutations in a European HSP series was reported at 1.4% however given the issues of defining the at-risk population based on clinical phenotype, the

utility of determining mutation frequency is limited (Méreaux *et al.*, 2021). It would be interesting however to know what proportion of patients with both ataxia and spasticity carried a CAPN1 mutation.

Reported mutations appear to be spread across the CAPN1 gene and no obvious mutation hotspots have been noted (Méreaux *et al.*, 2021). Various types of mutations have been reported including missense, nonsense, frameshift and splice site variants. Truncating variants account for 29/43 (67%) CAPN1 variants now reported. There does not appear to be any difference in frequency of the mutation type (truncating vs non-truncating) between the pure spastic phenotype and those with ataxia. Contrary to what might be expected, truncating mutations were associated with a statistically significant ($p<0.001$) earlier age of onset (mean 26.9 +-7.2 yrs.) vs missense mutations (17.5 +-8.1yrs).

More recently, compound heterozygous mutations in CAPN1 have been reported to cause spinal muscular atrophy type 4 (SMA4) (Perez-Siles *et al.*, 2022). Whole exome sequencing identified 2 mutations (p.G492R/p.F610C) in two siblings with SMA4 which is an adult onset and typically less severe form of SMA. Fibroblast studies showed a reduction in calpain 1 protein levels and protease activity. The mechanisms through which these mutations confer pathogenicity is yet to be determined and to date, no reported heterozygous carriers of spastic ataxia cases are known to have a degenerative muscle phenotype.

4.5.4 Conclusion

Together with international collaborators, I have reported a case series of patients with mutations in CAPN1 associated with a spastic ataxia phenotype. Functional studies in a patient with a probable loss of function mutation results in significantly reduced calpain-1 activity and effects on downstream pathways important in neuroprotection. Various animal models support the hypothesis that loss of calpain-1 activity is associated with a neurodegenerative phenotype and cerebellar degeneration. Many more cases of human CAPN1-associated

neurodegeneration have now been reported with exhibiting a range of phenotypes from pure HSP to spastic ataxia.

Calpain-1 is now clearly established as having a role in human neurogenerative disease and shared pathways appear to be important in the pathophysiology of several other neurodegenerative disorders.

5 Targeted amplicon sequencing in a UK autosomal dominant ataxia cohort

5.1 Statement of contribution

I performed all sample preparation, MiSeq amplicon design, library preparation and analysis.

The patients identified with PRKCG/SCA14 mutations were published as part of a larger case series of SCA14 patients (Chelban *et al.*, 2018). The clinical reappraisal of one of the SCA14 cases was undertaken by Dr Chelban (section 5.4.3).

Publications arising from this section:

- Genotype-phenotype correlations, dystonia and disease progression in spinocerebellar ataxia type 14. Chelban V, Wiethoff S, Fabian-Jessing BK, Haridy NA, Khan A, Efthymiou S, Becker EBE, O'Connor E, **Hersheson J**, Newland K, Hojland AT, Gregersen PA, Lindquist SG, Petersen MB, Nielsen JE, Nielsen M, Wood NW, Giunti P, Houlden H.

5.2 Introduction

Next generation DNA sequencing platforms have been widely available since 2005 and have led to a dramatic reduction in the cost and speed of DNA sequencing. This technology has been used to enable the development of whole genome and exome sequencing as both a research and diagnostic tool however significant challenges remain with the application of both these techniques. The same technology can be used to enable rapid sequencing of smaller amplicons that is well suited to support current research and diagnostic sequencing requirements. This allows multiple genes to be sequenced rapidly and accurately and at significantly reduced cost compared to Sanger sequencing.

At the time of this study, panel sequencing of large numbers of genes for ataxia and other neurodegenerative disorders was not in widespread use as the technology was in its infancy. I adopted a NGS approach to develop a screening panel for multiple genes that are known to cause autosomal dominant cerebellar ataxia (ADCA), the majority of which are rarely tested for primarily due to cost limitations. Genetic testing in a clinical setting generally consists of fragment analysis of the SCA 1, 2, 3, 4, 6 ,7 ,12 & 17 expansion loci. Despite testing for these genes, 40-50% of patients with inherited cerebellar ataxia remain genetically undiagnosed (Schöls *et al.*, 2004). This may be partly accounted for by a lack of comprehensive genetic testing although it is likely that some cases will be the result of mutations in genes yet to be identified. This project sought to identify the spectrum of mutations in the known dominant ataxia genes in a cohort of patients who have previously been screened for mutations in the repeat expansion SCAs.

5.3 Materials and Methods

I designed our panel of genes for use with our in-house Illumina MiSeq platform using. Amplicons defining the coding regions of the following genes were designed: SCA 11, 13, 14, 20, 23, 27, 35. Genes in which the predominant mutation was a nucleotide repeat expansion

were excluded from the panel due to limitations of the sequencing technology in identifying such mutations.

Patient selection

90 patients were selected for targeted sequencing with a diagnosis of cerebellar ataxia and a positive family history consistent with a dominant inheritance pattern. All patients had previously been screened and were negative for expansions in SCA 1, 2, 3, 6, 7, 12 and 17. 4 patients with conventional mutations in the known ataxia genes were selected as positive controls.

5.3.1 Sample preparation

Genomic DNA was extracted from venous blood using standard procedures by the Neurogenetics diagnostic laboratory, NHNN. DNA concentration was quantified using a fluorometric method (Qbit) and samples diluted to 50ug/µl (See methods 4.3.3)

5.3.2 Amplicon design

Amplicons defining the coding and splice site regions for the following genes were designed using Illumina DesignStudio software: *AFG3L2* (SCA28), *DNMT1* (Cerebellar ataxia with deafness and narcolepsy), *FGF14* (SCA27), *GFAP* (Alexander's disease), *IFRD1* (SCA18), *ITPR1* (SCA 15/16), *KCNC3* (SCA13), *KCND3* (SCA19), *PDYN* (SCA23), *POLG* (mitochondrial DNA depletion syndrome), *PRKCG* (SCA14), *PRNP* (GSS), *SPTBN2* (SCA5), *TGM6* (SCA35), *TTBK2* (SCA11), *EEF2* (SCA26).

Genomic coordinates were determined using the UCSC Genome Bioinformatics Table browser function and uploaded to the DesignStudio for amplicon design.

5.3.3 Library preparation and sequencing

Hybridisation of oligonucleotide pool: 5ul of the custom oligonucleotide pool reagent were mixed with 5 µL of DNA in a 96 well plate. 40µL of hybridisation solution was then added and the plate incubated at 95°C and left to cool over 90 minutes to 40°C

Removal of unbound oligos: The plate contents were transferred to a filter plate and placed on top of a deep well plate forming the filter-plate assembly unit (FPU). This was spun at 2400g for 2 minutes and then washed twice with stringent wash solution and then universal buffer solution. Hybridised DNA and oligonucleotide remained adherent to the filter plate membrane.

Extension-ligation of bound oligos: 45 µL of extension-ligation mix was added to the FPU which was then incubated at 37°C for 45 minutes.

PCR amplification: After centrifugation of the FPU, hybridised DNA was eluted in 25ul of 50mM NaOH and transferred to a clean 96 well plate. 22 µL of PCR master mix and indexing primers were added to each well which then underwent PCR amplification on a thermal cycler under the following conditions: 95°C for 3 minutes; 25 cycles of (95°C for 30s, 62°C for 30s, 72°C for 60s); 72°C for 5 minutes; hold at 4°C.

PCR clean-up: PCR products were purified using AMPure XP clean-up beads. Samples were mixed with 45ul of bead suspension liquid, agitated on a plate shaker and then placed on a magnetic stand to allow aspiration of the supernatant which was discarded. Samples were then twice washed with 800µL of ethanol and eluted into 30µL of elution buffer.

Library-normalisation: The resulting single-stranded DNA library was normalised using normalisation beads. 45µL of beads were incubated with the clean PCR product and agitated on a plate-shaker and the supernatant aspirated and discarded. The beads were washed twice with 45µL of library normalisation buffer, with the DNA then eluted from the beads into 30µL

of 0.1M NaOH. This eluate was then added to 30µL of library normalisation storage buffer for storage prior to pooling and sample loading:

Library pooling and sample loading: To ensure adequate sequencing depth, the 48 samples at a time were pooled and sequenced in a single MiSeq run. 5µL of each sample was loaded into an Eppendorf tube and mixed thoroughly. 10ul of the pooled library was added to 590ul of hybridisation buffer and incubated at 95°C for 2 minutes before being placed on ice. This was then transferred to the sample well of the MiSeq reagent cartridge and loaded onto the MiSeq for sequencing.

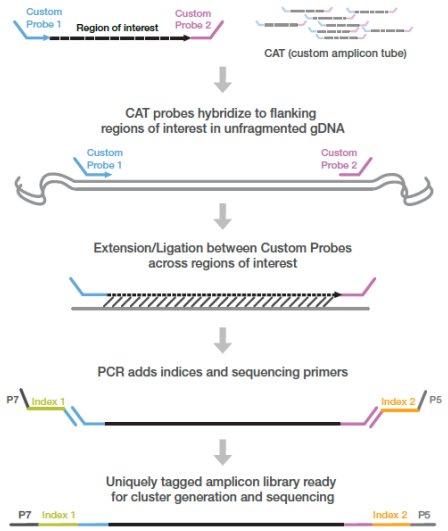


Figure 5-1 Workflow showing Miseq library preparation (reproduced from illumina.com)

5.3.4 Bioinformatic analysis

Sequence alignment and variant calling was performed with GATK through the automated Illumina pipeline. The resulting variant files in VCF format were annotated using ANNOVAR. The following filters were applied in order to select out technical artefacts and to identify

potentially pathogenic variants: heterozygous variants only, depth >20 reads, sequencing quality metric > 50, alternate read ratio: 0.30-0.65

Coverage analysis: Average coverage metrics were calculated for each exon using the Bedtools coverageBed function. A custom script was written to enable this data to be visualised in a single exon using the Bedtools coverageBed function. A custom script was written to enable this data to be visualised in a single table in Excel.

5.4 Results

5.4.1 Sequencing and coverage statistics

The two sequencing runs generated 5.3 gigabases and 5.1 gigabases of sequencing data respectively. Cluster density 565K/mm^2 and 550K/mm^2, which was slightly below the specification for the TruSeq Custom Amplicon kit although the data generated was of generally good quality.

Of the 95 samples, 3 samples failed completely, and one sample had markedly low coverage throughout. The remaining 92 samples gave good quality data. A visualisation of the coverage data from the runs is shown in figure x below. The samples are aligned along the x-axis and the exons aligned along the y axis. Cells are shaded green if 100% of the coverage is > depth 20, they are shaded yellow 0-99.5% > depth 20 and red if they are 100% < depth 20 (i.e. a failed exon). Overall, 89% of exons were covered 100% > depth 20, 7% were 0-99.5% covered > depth 20 and 5% were 0% covered > depth 20 (see figure 5-2).

Deleted: completely

Deleted: Overall

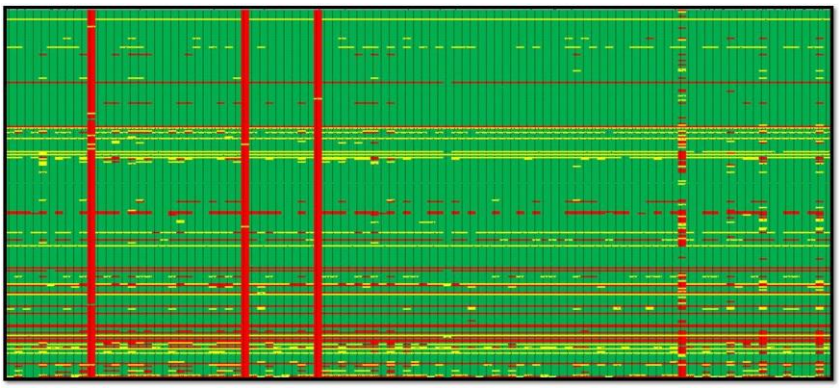


Figure 5-2 Graphical representation of coverage from the sequencing run. Amplicons for all sequenced genes are plotted sequentially along the x axis, patient samples are along the y-axis. Green shading indicates 100% amplicon coverage at depth > 20, yellow indicates partial coverage (0-99.5% at depth > 20 and red indicates zero coverage

5.4.2 Genetic results

Patient number	Clinical features	Gene	Mutation	Mutation novel or reported	Positive control?	Evidence for pathogenicity	Comments
1	cerebellar ataxia; leucodystrophy	GFAP	c.T704C:p.L235P	reported	yes	previously reported	
2	pure cerebellar ataxia	PRKCG	c.A76G:p.R26G	reported	yes	previously reported	
3	pure cerebellar ataxia	PRKCG	c.G392A:p.C131Y	reported	yes	previously reported	
4	pure cerebellar ataxia	TTBK2	c.1330_1331insA:p.R444fs	reported	yes	previously reported	
5	cerebellar ataxia, dementia	PRNP	c.C305T:p.P102L	reported	no	previously reported	Known mutation causing GSS
6	pure cerebellar ataxia	POLG	c.T391C:p.Y131H	novel	no	pathogenicity uncertain; POLYPHEN2 probably damaging, SIFT 0.42, MutationTaster neutral; not conserved	
7	cerebellar ataxia, white matter lesions on MRI	GFAP	c.C625G:p.R209G	novel	no	likely pathogenic; POLYPHEN2 damaging, SIFT 0.79, MutationTaster damaging; conserved	
8	pure cerebellar ataxia	ITPR1	c.G2817A:p.M939I	novel	no	pathogenicity uncertain; no in silico pathogenicity scores; moderately conserved	
9	pure cerebellar ataxia	TGM6	c.C1115T:p.A372V	novel	no	pathogenicity uncertain; SIFT 1, POLYPHEN2 damaging, MutationTaster damaging	
10	cerebellar ataxia, narcolepsy, dementia, optic atrophy, neuopathy	DNMT1	c.G1768T:p.V590F	novel	no	likely pathogenic; SIFT 1, POLYPHEN2 damaging, Mutationtaster damaging; conserved	located in exon 21 where all cerebellar ataxia-DN mutations are located
11	pure cerebellar ataxia	PRKCG	c.A302G:p.H101R	reported	no	previously reported	
12	cerebellar ataxia, neuropathy	PRKCG	c.G1372A:p.A458T	novel	no	pathogenicity uncertain; SIFT 0.86, POLYPHEN2 probably damaging, MutationTaster damaging	
13	pure cerebellar ataxia	PRKCG	c.G70A:p.A24T	novel	no	likely pathogenic, segregates with family; SIFT 1, POLYPHEN2 probably damaging, MutationTaster damaging; conserved	
14	cerebellar ataxia, childhood onset, learning disability	KCNC3	c.T1342C:p.F448L	reported	no	previously reported	single French pedigree reported
15	pure cerebellar ataxia	KCND3	c.AG1045CC p.S349P	novel	no	likely pathogenic; SIFT 1, POLYPHEN2 damaging, MutationTaster damaging; segregates with family	dinucleotide missense mutation
16	pure cerebellar ataxia	KCND3	c.A665G:p.Y222C	novel	no	pathogenicity uncertain; SIFT 0.97, POLYPHEN2 damaging, MutationTaster damaging, conserved	
17	cerebellar ataxia, opthalmoplegia, spasticity	AFG3L2	c.C2062T:p.P688S	novel	no	likely pathogenic; SIFT 1, POLYPHEN2 damaging, MutationTaster damaging	
18	pure cerebellar ataxia, opthalmoplegia	AFG3L2	c.T2009C:p.V670A	novel	no	likely pathogenic; SIFT 1, POLYPHEN2 damaging, MutationTaster damaging	

Table 5-1 Summary of sequencing results

A summary of the results can be seen in figure 5-1 above. 4/4 variants in were identified from the positive controls included in the run.

Novel or possibly pathogenic heterozygous mutations were identified in 14/87 (15%) of cases.

3 patients had mutations in *PRKCG*/SCA14 - 1 mutation had previously been reported in another patient with SCA14 (patient 11 – p.H101R) and two mutations were novel and predicted to be pathogenic based on in-silico predictions (patient 12 – p.A458T, patient 13 –

p.A24T). A clinical and genetic summary for the patient 12 can be seen in section 7.4.3. These data were published as part of a larger case series (Chelban *et al.*, 2018)

1 patient had a mutation in *KCNC3*/SCA13 (Patient 14). The p.448L variant had previously been reported in a French kindred (Figueroa *et al.*, 2011) although there was no known French ancestry in the affected individual.

2 patients had novel mutations in *KCND3* which were predicted to be pathogenic. The mutation in patient 15 (p.S349P) segregated with the disease phenotype in the family.

2 patients had novel mutations in *AFG3L2*/SCA28 which were predicted to be pathogenic and had phenotype compatible with SCA28 (ataxia and ophthalmoplegia).

Other pathogenic mutations likely to be pathogenic based on in-silico predictions and a compatible phenotype were found in *PRNP* (patient with late onset ataxia and dementia/GSS prion phenotype), DNMT1 (patient with ataxia, neuropathy and narcolepsy) and *GFAP* (progressive ataxia and white matter lesions on MRI compatible with Alexander's disease)

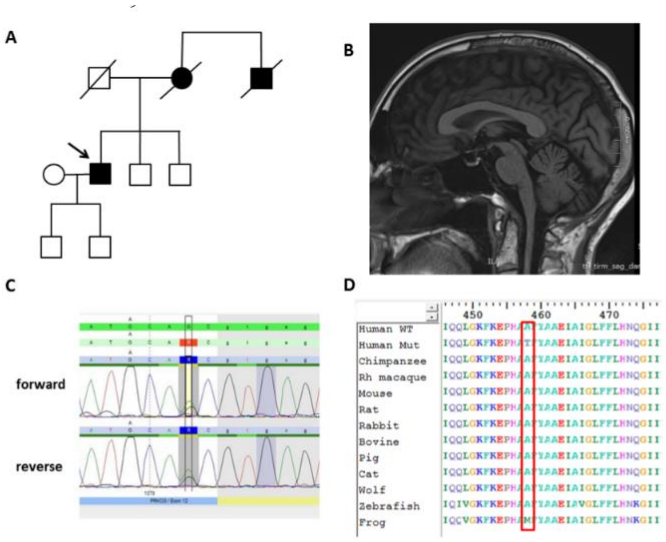
5/87 of cases had mutations that were of uncertain pathogenicity.

5.4.3 Clinical and genetic summary of patient with SCA14 and novel *PRKCG* p.A458T mutation

Patient 12 developed a slowly progressive ataxia from the mid-30s. At the age of 55 years, he developed a progressive dystonic head tremor associated with severe, constant dystonic posturing of the head and right hand. Examination showed a dystonic head tremor, laterocollis to the left and hypertrophy of right sternocleidomastoid together with cerebellar ataxia. Nerve conduction studies showed a severe length dependent axonal sensory motor neuropathy. Brain MRI shows cerebellar hemisphere and vermian atrophy. Neuropsychometry showed mild anterior and subcortical cognitive impairment.

Deleted: years

He was shown to have a novel mutation in *PRKCG* – p.Ala458Thr which was likely pathogenic based on pathogenicity prediction and evolutional conservation.



E

Chr	Position (hg19)	Transcript	Variant	dbSNP, 1000g, ExAC, EVS, cg69	GERP score[a]	CADD score[b]	Mutation Taster
19	54403577	ENST00000263431	c.1372G>A, p.Ala458Thr	absent	3.4	22.3	Disease causing

Figure 5-3 A) Pedigree showing family of patient 12 who is indicated with black arrow B) Sagittal MRI brain showing cerebellar atrophy C) Sequence chromatogram of the c.1372G>A substitution D) multi-species alignment of PRKCG showing conservation of the Ala458 residue E) updated variant annotation showing absence in multiple variant databases and in-silico pathogenicity prediction suggesting the variant is disease-causing - reproduced from (Chelban et al.*, 2018)*

5.5 Discussion

This project reports the development of a targeted next generation sequencing panel to comprehensively screen the known ataxia genes excluding repeat expansions in a UK ataxia cohort.

The coverage statistics indicate that the majority of exons (89%) have sufficient sequencing depth to enable mutation detection. Certain exons were missed however, and this tended to

Deleted: however

be in areas of high GC content. Because of the PCR step in the library preparation protocol, exons with high GC content can be difficult to amplify and may therefore fail enrichment. One positive control was missed as a result of incomplete coverage.

The P102L mutation in PRNP is well described as a cause of GSS and is consistent with the phenotype of a late onset ataxia and dementia. The H101R mutation in SCA14 has been previously shown to be pathogenic and multiple different missense mutations at the amino acid residue have been described. The F448L mutation in *KCNC3* was identified first in a large French pedigree and has been shown by functional analysis to be pathogenic. It is not clear as yet whether this patient is related to the French pedigree or whether this was a second family with an identical mutation.

6/87 (7%) of cases had mutations that were likely to be pathogenic based on a range of criteria including in-silico predictions, family segregation and the location relative to previously reported mutations. Where possible, family segregation was performed however this is not always possible and limits the extent to which some mutations can confidently be determined to be pathogenic. 5/87 (6%) of cases had a novel mutation in a known ataxia gene however in these cases there was inadequate information to confidently identify these as pathogenic. Although several were predicted by in-silico methods to be pathogenic and may have been conserved, it can be difficult to know whether these are true pathogenic mutations or whether these are merely rare neutral variants. Checking for the variant in other affected family members can be helpful however the most reliable strategy is to assess the functional impact of the mutations in vitro although this is not often practicable.

Screening for known genes using these relatively low-cost techniques will be important both for research and increasingly in clinical diagnostics. Issues remain with regard to inconsistent coverage and in certain cases Sanger sequencing will be necessary to fill in any gaps in parts of the genes not well covered by these enrichment techniques. As is the case with exome

Deleted: low cost

sequencing, interpretation of the pathogenicity of rare and novel variants in Mendelian disease remains a significant challenge.

6 Identification of two families with cathepsin-D deficiency associated neuronal ceroid lipofuscinosis

6.1 Statement of contribution

I undertook a clinical evaluation of the affected subjects in Families 1 and 2 and performed the skin biopsies for fibroblast culture. Fibroblast culture was undertaken by the cell culture laboratory team at ION. Exome sequencing was performed by Dr Deborah Hughes and subsequent analysis of homozygosity mapping and exome sequencing were performed by me. I performed all cathepsin D activity assays with the assistance of Derek Burke in the Institute for Child Health. Histological analysis and electron microscopy of the muscle biopsy samples was undertaken by Dr Thomas Jacques in the Histopathology department at Great Ormond St Hospital for Children.

Publications arising from this section:

- Cathepsin D deficiency causes juvenile-onset ataxia and distinctive muscle pathology**. Hersheson J**, Burke D, Clayton R, Anderson G, Jacques TS, Mills P, Wood NW, Gissen P, Clayton P, Fearnley J, Mole SE, Houlden H. Neurology. 2014 Nov 11;83(20):1873-5

6.2 Introduction

The recessive cerebellar ataxias are a complex group of neurodegenerative disorders with significant genetic and phenotypic heterogeneity. These are frequently early-onset multisystem disorders that present with a range of neurological manifestations including motor and cognitive deterioration, visual disturbance, epilepsy and neuropathy (Fogel and Perlman, 2007). Several groups of neurogenetics disorders exhibit clinical and pathological overlap with the recessive spinocerebellar ataxias including the complex spastic parapareses, mitochondrial and metabolic disorders, which can complicate the clinical evaluation of these disorders and also in the selection of appropriate diagnostic tests. The neuronal ceroid lipofuscinoses (NCLs) are themselves a genetically heterogeneous group of disorders that predominantly affect children and typically show autosomal recessive inheritance although adult variants have been described with dominant inheritance (Warrier, Vieira and Mole, 2013). The most common phenotype is of progressive motor deterioration cognitive decline, visual failure, epilepsy, cerebellar ataxia and premature death (Williams and Mole, 2012).

A defining pathological feature of these disorders is the accumulation of autofluorescent lipopigment storage material within neuronal lysosomes that has a characteristic ultrastructural appearance on electron microscopy (Anderson, Goebel and Simonati, 2013). The age of onset is variable, and the disorders are often classed as one of six subtypes: congenital, infantile, classic or variant late infantile, juvenile and adult. This is a clinical distinction and most genetic subtypes of NCL have been reported with a variable age of onset. To date 13 genes have been associated with NCL (Kousi, Lehesjoki and Mole, 2012; Warrier, Vieira and Mole, 2013): *PPT1* (CLN1), *TPP1* (CLN2), *CLN3*, *DNAJC5* (CLN4), *CLN5, CLN6, MFSD8* (CLN7), *CLN8*, *CTSD* (CLN10), *GRN* (CLN11), *ATP13A2* (CLN12), *CTSF* (CLN13), *KCTD7* (CLN14), with further genes causing NCL in mouse models. Five of these genes (*PPT1, TPP1, CLN5, CTSD*, *CTSF*) encode soluble lysosomal proteins, and five genes (*CLN3, MFSD8, CLN6, CLN8, ATP13A2*)

Deleted: variable

Deleted:

encode intracellular transmembrane proteins that localise to lysosomes, the Golgi or the endoplasmic reticulum. *DNAJC5* and *KCTD7* encode soluble cytoplasmic proteins and *GRN* encodes a secreted protein.

In this section I will report the identification of novel mutations in *CTSD*, which encodes the lysosomal protein cathepsin D which is the gene underlying neuronal ceroid lipofuscinosis type 10 (CLN10), in patients that had undergone extensive neurological investigation for many years without a genetic diagnosis having been established. Pathological examination of muscle biopsies from these patients revealed a characteristic muscle pathology not previously seen in other NCL subtypes.

Deleted: and

6.3 Methods

6.3.1 Sample collection

Genomic DNA was extracted, with informed consent, from whole blood of all members of family A and 84 genetically undiagnosed patients with a clinical diagnosis of retinitis pigmentosa with cerebellar ataxia seen in the Neurogenetics clinic at the National Hospital for Neurology and Neurosurgery. Skin biopsies for fibroblast samples were collected as per Methods 4.4 for patients A:3 and B:5.

Ethical approval to undertake this study was given by the NHNN/ION ethics committees.

6.3.2 Homozygosity mapping

DNA from the two living affected siblings (A:5, A:8) and one unaffected sibling (A:9) were genotyped using a CytoSNP12 array (Illumina, USA) containing 301,232 genome wide single nucleotide polymorphism (SNP) markers. Homozygosity mapping was performed using GenomeStudio v2011.1 software with the homozygosity detector plugin. Shared regions of homozygosity >1 Mb were subsequently used to filter variants from exome sequencing data.

6.3.3 Exome sequencing

3 micrograms of genomic DNA from A:5 was sent for exome sequencing to AROS Applied Biotechnology, Denmark. This generated 91,544,158 unique reads with a mean depth of 67. 90.6% of target was covered by at least 10 sequencing reads. Alignment to the hg19 reference genome was performed using Novoalign and variant calling using Samtools. 22,098 variants were called and annotated using Annovar. The variant list was filtered to exclude synonymous variants, heterozygous variants and variants with a mean allele frequency (MAF) of greater than 0.005 in 1000genomes and Exome Variant Server databases. Finally, the list was filtered to include only regions of homozygosity shared by A:5 and A:8 but not A:9.

Deleted: Finally

6.3.4 Sanger sequencing

To validate the results of exome sequencing and to sequence other individuals with a similar phenotype, primers for the coding exons and intron/exon boundaries of *CTSD* were designed using Primer3. PCR products were cleaned with Millipore filter plates prior to sequencing using BigDye Terminator 3.1 chemistry (Life Technologies, Carlsbad, CA, USA). Sequencing products were analysed through capillary electrophoresis on an ABI 3130XL Genetic Analyzer (ABI Biosystems, Foster City, CA, USA). Sequencing data was visualized using Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA).

Deleted: In order to

6.3.5 Immunohistochemistry

Frozen sections of muscle from patients A:3 and B:3 were taken at 10 micrometres and stained using a panel of histochemical and immunohistochemical stains in routine diagnostic use at the respective institutions and in keeping with standard diagnostic practice.

6.3.6 Electron microscopy

Electron microscopy was performed on fresh muscle samples and on cultured fibroblasts from B:3 Selected samples of muscle biopsy were fixed in 2.5% glutaraldehyde in 0.1 M cacodylate buffer followed by secondary fixation in 1.0% osmium tetroxide. Tissues were dehydrated in

graded ethanol, transferred to propylene oxide and then infiltrated and embedded in Agar 100 epoxy resin. Polymerisation was at 60°C for 48 hours. 90 nanometre ultrathin sections were cut using a Diatome diamond knife on a Leica Ultracut UCT ultramicrotome. Sections were picked up on copper grids and stained with alcoholic uranyl acetate and Reynold's lead citrate. The samples were examined in a JEOL 1400 transmission electron microscope, and images recorded using AMT XR80 digital camera and software.

Deleted: microscope

6.3.7 Cathepsin D (CatD) enzyme analysis

6.3.7.1 Fibroblast collection and culture:

Patients A:8 and B:3 underwent a skin biopsy from which dermal fibroblasts were cultured. Control fibroblasts were obtained from the Chemical Pathology department, Great Ormond St Hospital. Fibroblasts were cultured in Dulbecco's modified Eagle's medium GlutaMAX supplemented with 10% (v/v) heat-inactivated foetal bovine serum and 1% penicillin streptomycin. They were maintained at 37°C in a humidified atmosphere of 5% CO_2 and 95% air. Fibroblast samples were lysed by sonication and the protein concentration measured using a bicinchoninic acid (BCA) assay (Smith *et al.*, 1985).

Deleted: fetal

6.3.7.2 Cat D Kinetic Assay

Kinetic studies were performed on 0.5 µg fibroblast cell lysates from the patients A:8, B:3 and pooled control fibroblasts using the Sigma CatD Assay Kit (Sigma-Aldrich, USA). The fluorogenic peptide MOCAc-GKPIIFFRLK(Dnp)-R-NH2 is a substrate for Cat-D which cleaves the phenylalanine-phenylalanine bond to yield a cleavage product with specific emission/excitation characteristics and so can determine the rate of Cat-D mediated hydrolysis. The samples were incubated with the fluorogenic substrate and fluorescence readings recorded over a 15-minute period at 1-minute intervals using a Tecan Infinite 200 PRO microplate reader (Tecan, Switzerland). Reaction velocities are reported as picomoles of

Deleted: 15 minute

Deleted: 1 minute

MOCAc-peptide released per second per mg of protein. All experiments were performed in triplicate.

Fluorescence values obtained from blank reactions without cell lysates were subtracted from the sample readings. The linear phase was determined, and the velocity calculated with reference to a 7-methoxycoumarin-4-acetic acid (MCA) standard curve.

Deleted: determined

6.3.7.3 CatD end-point activity measurement

End-point CatD activity assays were also performed in sonicated leucocyte pellets from A:8, B:3, the mother of B:3 and control samples using a modification of the above protocol. Samples were incubated at 37°C for 10 minutes with 20uM of the fluorogenic substrate. 5% trichloroacetic acid was used as a stopping agent and end-point fluorescence measured with a fluorimeter (Perkin Elmer LS 50B, excitation 328 nm, emission 393 nm). Units of activity are reported as pmol/hr/mg protein.

Samples were assayed in a reaction volume of 100µl. 80µl of 50mM buffered sodium acetate (pH 4.0) was used in each reaction. For the inhibited samples, 10µl of 2mg/ml pepstatin A solution was added. 5µl of leucocyte sonicate at protein concentration of 5mg/ml was added to each reaction volume with each reaction was made up to 95µl with water.

Reaction velocities were plotted against increasing substrate concentrations. K_m and V_{max} were calculated by nonlinear regression using the equation $V=V_{max} \times [S]/ (K_m+[S])$ where V is the linear phase velocity, [S] is the substrate concentration and K_m is the Michaelis-Menten constant. Calculations were performed using GraphPad Prism software and both means, and standard deviations were plotted.

Deleted: means

6.4 Results

A summary of the clinical features of both pedigrees is shown in table 6-1 and pedigrees are shown in figure 6-1.

Deleted: 8

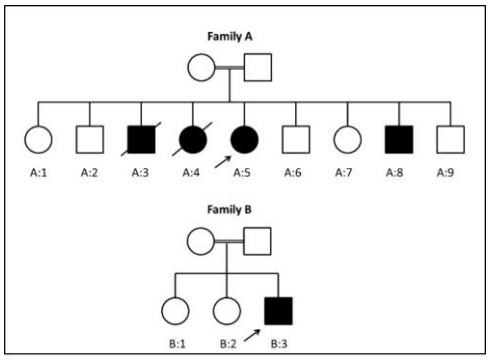


Figure 6-1 Pedigree of Families A & B. Affected individuals are in black. Index cases for each family are marked with an arrow

Patient	Ethnicity	Age of onset	Clinical details					CTSD mutation	CatD activity
			Retinitis	Ataxia	Dementia	Neuropathy	Myopathy		
A:3	Somalian	7yrs - retinitis	yes	yes	yes	no	no	p.G149V	
A:4	Somalian	6yrs - retinitis	yes	yes	yes	no	cardiomyopathy	p.G149V	
A:5	Somalian	6yrs - retinitis	yes	yes	yes	axonal	no	p.G149V	
A:8	Somalian	8yrs - retinitis	yes	yes	yes	no	no	p.G149V	Vmax 42.6 pmol/s/mg (11% control Vmax
B:3	Somalian	4yrs - retinitis	yes	yes	yes	axonal	myofibrillary myopathy	p.R399H	Vmax: 48.2 pmol/s/mg (11% of control Vmax)

Table 6-1 Summary of clinical features for the affected individuals in family A & B

6.4.1 Family A

6.4.1.1 Clinical features and investigations

The index patient (A:5) is one of 9 siblings from Somalia and the product of a 1st cousin marriage. This patient exhibited delayed motor development and had problems with balance in infancy. She developed progressive visual impairment from age 15 due to retinitis pigmentosa and is now blind. Four of the 9 siblings were affected by a similar disorder

characterised by retinitis pigmentosa, progressive cerebellar ataxia and cognitive decline. All of her affected siblings developed retinitis and progressive motor impairment from around age 15 years.

Nerve conduction studies showed a sensory axonal neuropathy in this patient; however, this was not found in the other three affected siblings. Her sister (A:4) underwent cardiac investigations at age 32 years after developing a peripheral venous embolism and was found to have dilated cardiomyopathy. There was no clinically relevant muscle weakness. Serum creatine kinase was slightly above the normal range - 288U/L (range 25-200 U/L).

Deleted: however

Both had gross cerebellar atrophy on MRI brain. Patient A:4 died age 33 years from complications relating to her cardiac disease. Patient A:3 died age 34 years from an aspiration pneumonia. The two affected, surviving siblings (A:5 – 36 yrs., A:8 – 26 yrs) both have significant disability. They are wheelchair bound and both have significant motor and functional impairment due to cerebellar ataxia, retinitis pigmentosa and cognitive decline.

6.4.1.2 Exome sequencing, homozygosity mapping and sequence analysis

Multiple regions of homozygosity were detected in patient A:5. These were used to refine the 22,098 variants identified on exome sequencing to a tractable list of 7 rare homozygous variants, the most plausible candidate being the c.446G>T, p.G149V variant in exon 4 of *CTSD* (Table 6-2)

The variant results in a glycine to valine missense substitution within the heavy chain of the CatD protein. The amino-acid site is conserved in mammalian species and the variant was predicted to be pathogenic using in-silico predictions of pathogenicity (SIFT, PolyPhen2 and MutationTaster). The variant was novel and had not previously been reported in dbSNP, 1000genomes or the Exome Variant Server. It was present in the homozygous state in all affected siblings and was heterozygous in both parents and the available unaffected siblings.

Deleted: 2 and

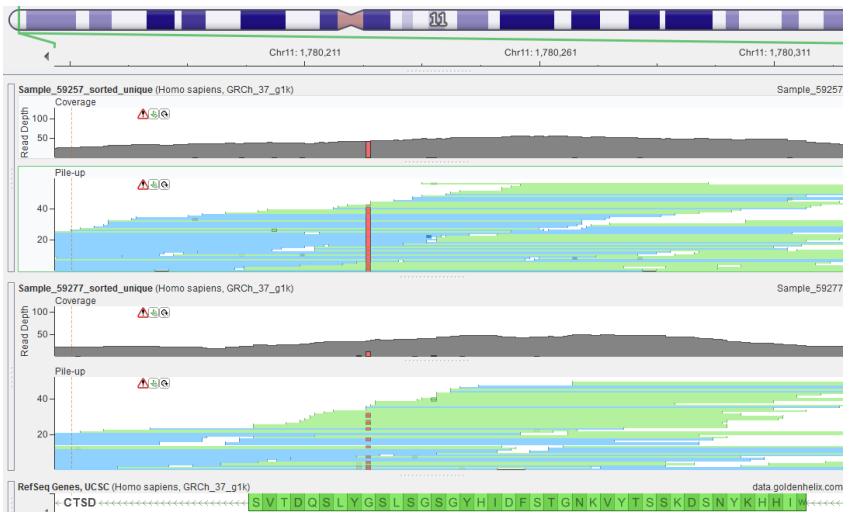


Figure 6-2 BAM alignment of exome sequencing data showing homozygous G149V mutation in patient A:5

Gene	Amino acid change	Exonic Function	Depth	1000g	EVS	SIFT	PolyPhen2	Mutation Taster	GERP
NLRP6	c.G2300C	nonsyn SNV	30	0.0018	0.002603	T	NA	N	-0.0477
C11orf35	c.G463A:p.E155K	nonsyn SNV	34	0.0023	0.002605	D	D	N	1.61
PHRF1	c.C4913T:p.A1638V	nonsyn SNV	36	5.00E-04	0.001984	NA	NA	NA	NA
CTSD	c.G446T:p.G149V	nonsyn SNV	29	NA	NA	D	D	D	2.64
OR52I2	c.C88A:p.P30T	nonsyn SNV	241	0.0014	0.00251	T	B	N	-0.818
OR52E8	:c.G685A:p.A229T	nonsyn SNV	70	NA	NA	T	B	N	0.36

Table 6-2 List of exonic variants within region of shared homozygosity in Patient A:5

6.4.1.3 CTSD screening in complex ataxia cohort

84 patients with a genetically undiagnosed ataxia complicated by retinitis pigmentosa were screened subsequently for mutations in *CTSD*. One additional case was identified (B:3). The clinical features and genetic results are described below.

6.4.2 Family B

6.4.2.1 Clinical features and investigations

The index patient (B:3) is also of Somali descent and was the product of a second-cousin marriage and is the only affected individual in the family (6-1). He had a normal early

psychomotor development and first presented with visual problems at age five years due to retinitis pigmentosa. From 8 yrs. he developed progressive ataxia. Cognitively, he started to deteriorate from 14 yrs. and now has significant disability due to cerebellar ataxia, visual and cognitive impairment.

Clinical features were similar to those of patient A:5 and included pigmentary retinopathy, cerebellar ataxia with broken pursuit eye movements, horizontal nystagmus, cerebellar dysarthria and truncal ataxia and peripheral neuropathy. There was no clinical evidence of myopathy.

MRI examination showed moderate cerebellar volume loss with diffuse signal change within the cerebral hemisphere white matter including the posterior limbs of the internal capsules. Nerve conduction studies showed evidence of a mild sensory axonal neuropathy affecting both large and small diameter sensory fibres. A homozygous variant c.1196G>A, p.R399H in exon 9 of *CTSD* was identified in the index case of a second consanguineous pedigree (patient B:3). As with p.G149V, this missense mutation is evolutionarily conserved and is predicted to be pathogenic using the same in-silico tools. It was absent in all publicly available exome sequencing datasets. The variant segregated in a heterozygous state in both parents and unaffected siblings. Both variants were highly conserved across multiple vertebrate species.

Deleted: showed

Deleted: Similar to

```
                                                149
                                          *     ↓
CTSD_Homo sapiens              HKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLS-QDTVSVPCQSASSASALGGVKVERQVF  179
CTSD_Sus scrofa                HKYNSGKSSTYVKNGTTFAIHYGSGSLSGYLSSQDTVSVPCNSAL--SGVGGIKVERQTF  114
CTSD_Bos taurus                RKYNSDKSSTYVKNGTTFDIHYGSGSLSGYLS-QDTVSVPCNPSS--SSPGGVTVQRQTF  157
CTSD_Rattus rattus             HKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLS-QDTVSVPCKS-----DLGGIKVEKQIF  174
CTSD_Gallus gallus             HKYDASKSSTYVENGTEFAIHYGTGSLSGFLSQ-DT-----------VTLGNLKIKNQIF  167
CTSD_Chionodraco hamatus       HKYNSGKSSTYVKNGTAFAIQYGSGSLSGYLSQDT-----------CTIGDLAIDSQLF  165
CTSD_Dictyostelium discoideum  NKYNSGASSTYVANGTDFTIQYGSGAMSGFVSQDS-----------VTVGSLTVKDQLF  152
                                                                  399
                                                                  ↓
CTSD_Homo sapiens              SQAGKTLCLSGFMGMDIPPPSGPLWILGDVFIGRYYTVFDRDNNRVGFAEAARL        412
CTSD_Sus scrofa                SQAGQTICLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRDLNRVGLAEAA--        345
CTSD_Bos taurus                SQAETTVCLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRDQNRVGLAEAARL        390
CTSD_Rattus rattus             SQAGKTICLSGFMGMDIPPPSGPLWILGDVFIGCYYTVFDREYNRVGFAKAATL        407
CTSD_Gallus gallus             SAQGETICLSGFSGLDVPPPGGPLWILGDVFIGPYYTVFDRDNDSVGFAKCV--        398
CTSD_Chionodraco hamatus       TQAGKTMCLSGFMGLDIPAPAGPLWILGDVFMGQYYTVFDRDANRVGFAKAK--        396
CTSD_Dictyostelium discoideum  TEFGKTECLSGFMGIELN--MGNFWILGDVFISAYYTVFDFGNKQVGFATAIQG        383
```

Figure 6-4 Multi-species protein sequence alignment of CatD. Note location of residue 149 within peptidase flap, close to the Y149 residue

```
                                                *       ↓
CATD_HUMAN     VPSIHCKLLDIACWIHHKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLSQDTVSVPCQSAS 164
BACE1_HUMAN    VGAAPHP------FLHRYYQRQLSSTYRDLRKGVYVPYTQGKWEGELGTDLVSIPHGPNV 154
BACE2_HUMAN    VAGTPHS------YIDTYFDTERSSTYRSKGFDVTVKYTQGSWTGFVGEDLVTIPKGFNT 171
RENI_HUMAN     VPSSKCSRLYTACVYHKLFDASDSSSYKHNGTELTLRYSTGTVSGFLSQDII-------- 163
PEPC_HUMAN     VPSVYCQ--SQACTSHSRFNPSESSTYSTNGQTFSLQYGSGSLTGFFGYDTL-------- 148
CATE_HUMAN     VPSVYCT--SPACKTHSRFQPSQSSTYSQPGQSFSIQYGTGSLSGIIGADQVSAF----- 156
PEPA5_HUMAN    VPSVYCS--SLACTNHNRFNPEDSSTYQSTSETVSITYGTGSMTGILGYDTV-------- 151
PEPA3_HUMAN    VPSVYCS--SLACTNHNRFNPEDSSTYQSTSETVSITYGTGSMTGILGYDTV-------- 151
PEPA4_HUMAN    VPSVYCS--SLACTNHNRFNPEDSSTYQSTSETVSITYGTGSMTGILGYDTV-------- 151
NAPSA_HUMAN    VPSRRCHFFSVPCWLHHRFDPKASSSFQANGTKFAIQYGTGRVDGILSEDKL-------- 155

CATD_HUMAN     -GMDIPPP-SGPLWILGDVFIGRYYTVFDRDNNR----VGFAEAARL-------------- 412
BACE1_HUMAN    YKFAISQ--SSTGTVMGAVIMEGFYVVFDRARKR----IGFAVSACHVHDEFRTAAVEGP 434
BACE2_HUMAN    YRFGISP--STNALVIGATVMEGFYVIFDRAQKR----VGFAASPCAEIAGAAVSEISGP 447
RENI_HUMAN     -AMDIPPP-TGPTWALGATFIRKFYTEFDRRNNR----IGFALAR---------------- 406
PEPC_HUMAN     -PTYLSSQNGQPLWILGDVFLRSYYSVYDLGNNR----VGFATAA---------------- 388
CATE_HUMAN     -GLDIHPP-AGPLWILGDVFIRQFYSVFDRGNNR----VGLAPAVP--------------- 401
PEPA5_HUMAN    -GMNVPTE-SGELWILGDVFIRQYFTVFDRANNQ----VGLAPVA---------------- 388
PEPA3_HUMAN    -GMNLPTE-SGELWILGDVFIRQYFTVFDRANNQ----VGLAPVA---------------- 388
PEPA4_HUMAN    -GMNLPTE-SGELWILGDVFIRQYFTVFDRANNQ----VGLAPVA---------------- 388
NAPSA_HUMAN    -ALDVPPP-AGPFWILGDVFLGTYVAVFDRGDMKSSARVGLARARTRGADLGWGETAQAQ 417
```

Figure 6-3 Protein sequence alignment of members of peptidase A1 family showing conservation of residue 149 in all members and conservation of residue 399 in all members except PEPC (gastricisin)

6.4.3 Histopathology and electron microscopy findings

Both biopsies from patients A:3 and B:3 showed a distinctive pattern of pathology characterised by granulovacuolar material in angular atrophic fibres. The biopsy from patient B:3, taken at 15 years of age, showed increased variation in fibre size (5 to 50 µm, normal range for age approx. 38 to 50) due to angular atrophic fibres that were present throughout the fascicles (Figure 6-5 A-E). Many of the fibres, particularly the atrophic ones, contained vacuoles associated with granular material. Many of these resembled rimmed vacuoles. The material was stained positive for Gömöri's trichrome, Periodic acid Schiff (for polysaccharides; PAS), oil red O (for lipid) and acid phosphatase (a lysosomal enzyme). In all fibres, there were prominent puncta of acid phosphatase staining and there was excess lipid deposition for the age of the patient. There were cytoplasmic bodies. There was type 2 fibre predominance. The small fibres were mostly of type 1 as were many, but not all, of the fibres containing vacuoles. Several fibres stained for developmental myosin.

There was no excess endomysial connective tissue. Staining with menadione NBT for reducing bodies was negative. Oxidative stains (SDH, NADH-TR, COX) were unremarkable.

The granulovacuolar material showed immunoreactivity for desmin and p62. Immunohistochemistry for a panel of dystrophy-associated molecules (dystrophin, sarcoglycans α, β, γ, δ or ε), dystroglycan (α and β), merosin, dysferlin, emerin and collagen VI) was normal. These membrane proteins did not stain the vacuoles. Utrophin showed a patchy pattern of up-regulation at the sarcolemma.

On electron microscopy (Figure 6-5 F-I), there was marked lipid droplet deposition in many myofibres. Some areas contained numerous, large vacuoles with electron dense myelinoid whorls, up to 6 microns in diameter and associated with filaments. There were rare foci of membrane-bound granular osmiophilic deposits material in endothelium, smooth muscle and skeletal muscle fibres. Many of these deposits were membrane bound.

Limited material was available to review from an archival biopsy from patient A:3), taken at age 27 yrs. (Figure 6-5 J-L). However, this showed similar features to the biopsy from patient B:3. There was increased variation in fibre size (11 to 120 um, normal range for age is approximately 38 - 55 micrometres) mostly due to scattered angular atrophic fibres. Some fibres contained vacuoles associated with eosinophilic globules and Gömöri-positive granular material. Most, but not all, of the vacuoles were in the atrophic fibres.

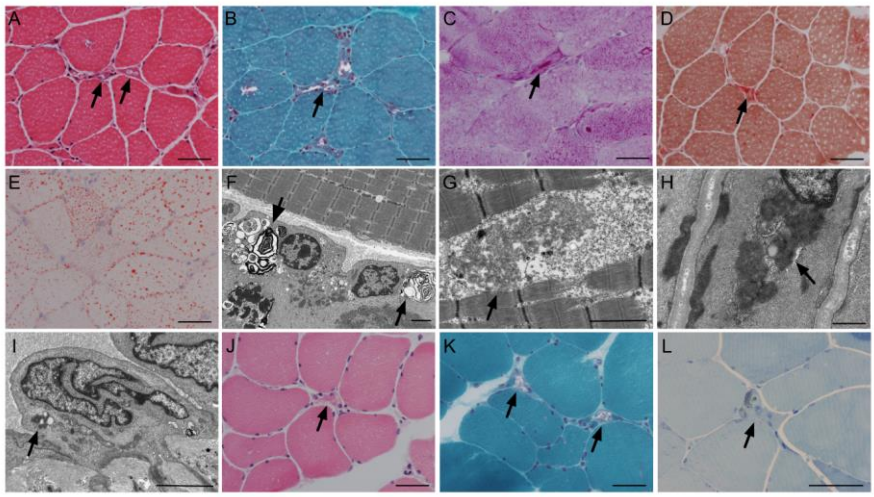


Figure 6-5 Histopathology of patients A:3 and B:3

Muscle biopsies from patients A:3 and B:3 showed similar features. The biopsy from (B:3) contained angular atrophic fibres, many of which contained vacuoles associated with granular material, reminiscent of rimmed vacuoles (A-H&E B-Gomori trichrome; arrows indicate the atrophic vacuolated fibres). Occasional larger fibres contained similar vacuoles. These vacuoles contained PAS-positive material (C; an arrow indicates an example of the accumulated material), acid phosphatase (D-acid phosphatase; an arrow indicates an example of the accumulated material) and lipid (E-oil red O). In addition, in all fibres, there were prominent puncta of acid phosphatase staining (D) and there was excess lipid for age (E-oil red O). Electron microscopy showed myelinoid whorls in muscle fibres (F-arrows) and granular osmiophilic deposits in skeletal muscle (G-arrow), smooth muscle (H-arrow) and endothelium (I-arrow). The biopsy from patient A:3 showed similar scattered angular atrophic fibre containing granular vacuolar material sometimes associated with eosinophilic material (J-H&E, K-Gomori trichrome, L-toluidine blue-stained resin section; arrows indicate the atrophic vacuolated fibres). Scale bars: A-E, J&K-50 μm, F&I -2μm, G&H-500nm

A skin biopsy was not available for electron microscopy for either patient; however, the ultrastructural features of dermal fibroblasts from patient B:3 were examined. Figure 6-6 shows frequent membrane bound vacuoles without any evidence of osmiophilic storage

Deleted: however

material. Whilst this is abnormal, it is a non-specific finding and has been described in a variety of neurometabolic disorders including NCL, oligosaccharidoses, mucopolysaccharidoses, lysosomal storage disorders and peroxisomal disorders (Ceuterick and Martin, 1992).

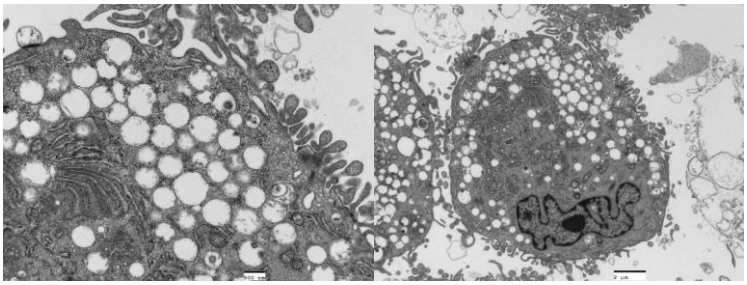


Figure 6-6 Electron micrograph of cultured fibroblasts of patient B:3 showing abnormal membrane bound vacuoles

6.4.4 CatD kinetic analysis

Fibroblasts homozygous for either p.R399H or p.G149V mutations showed significantly reduced enzyme activity compared to that of pooled control fibroblasts (Figure 6-7 - A). The kinetic properties of CatD in control fibroblasts were: K_m = 6.5 and V_{max} = 408. Both mutations resulted in similar values for K_m compared to the pooled controls: 6.94 (p.G149V) and 5.95 (p.R399H). The V_{max} for both mutations was approximately 11% of the value calculated for controls – 42.6 pmol/s/µg (p.G149V) and 48.2 pmol/s/µg (p.R399H).

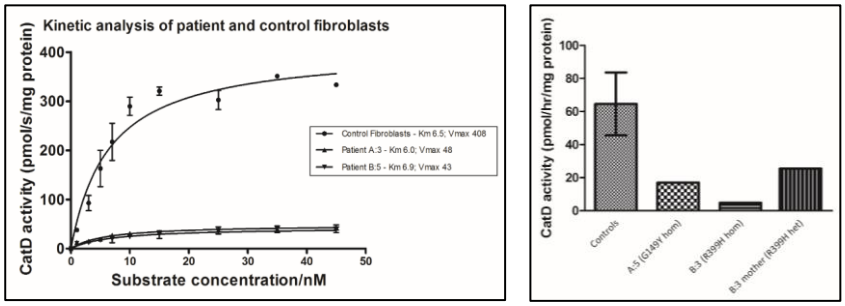


Figure 6-7 A) Kinetic analysis of CatD activity in patient A:3 and B:5 and pooled control fibroblasts B) Endpoint CatD activity in leucocyte preparations

Mean cathepsin D end-point activity in control leucocyte preparations (n=26) was 64.6 pmol/hr/mg (see figure 6-7 – B). Activity in the G149Y homozygote was 14.5 pmol/hr/mg, p.R399H homozygote was 4.8 pmol/hr/mg and p.R399H heterozygote was 26 pmol/hr/mg.

6.5 Discussion

I have reported the identification of two novel mutations in CTSD in two consanguineous pedigrees. These were identified through a strategy of exome sequencing, homozygosity mapping and candidate gene sequencing with functional analysis of CatD activity in patient fibroblasts and leucocytes, confirming the pathogenicity of the two mutations.

6.5.1 Diagnostic challenges

The phenotype of all the patients in this study is of a complex recessive ataxia with associated pigmentary retinopathy and cognitive decline. These patients had been extensively investigated previously in both paediatric and adult clinics without successful diagnosis. Both had been screened for a range of metabolic defects including the most common lysosomal disorders, recessive ataxias such as FRDA, AOA1, AOA2, ATM and mitochondrial disorders. Difficulty in establishing a diagnosis in these patients was unsurprising given the phenotypic overlap with other disorders including mitochondrial disorders, complex hereditary spastic parapareses, neurodegeneration with brain iron accumulation (NBIA) and disorders of peroxisomal and lipid metabolism. Many metabolic disorders can be screened enzymatically, however for many others, including CLN10, assays are not routinely available on a diagnostic basis. This highlights the importance of genetic techniques such as exome sequencing and homozygosity mapping in diagnosing such disorders. These approaches will not however obviate the requirement for confirmatory functional analysis and enzyme assays will remain an essential component in the diagnosis in metabolic diseases.

Deleted:

Deleted: approaches

Deleted: of

6.5.2 Neuronal ceroid lipofuscinosis type 10/CLN10

Prior to their discovery as a cause of human disease, mutations in *CTSD* had been associated with an NCL phenotype in a sheep model for congenital NCL. In humans, congenital NCL is a severe neurodegenerative disorder characterised by microcephaly, post-natal respiratory failure, status epilepticus and death within hours to weeks of birth (Barohn, Dowd and Kagan-Hallet, 1992). Only two families have been described with congenital NCL due to proven *CTSD* mutations (CLN10) – one Pakistani kindred with a homozygous c.764dupA, p.Y255* nonsense mutation 10 and one US patient with a homozygous c.299C>T, p.S100F missense mutation (Fritchie *et al.*, 2009; Warrier, Vieira and Mole, 2013).

Cathepsin D (CatD) is a ubiquitously expressed lysosomal aspartic protease belonging to the pepsin family. It is involved in a diverse range of physiological processes including cell proliferation, apoptosis, senescence and tissue homeostasis as well as mediating protein degradation, protease activation and protease inhibitor inactivation. It also been implicated in several pathological processes including cancer cell invasion and proliferation, ischaemic heart disease, inflammatory bowel disease and Alzheimer's disease (Hausmann *et al.*, 2004; Masson *et al.*, 2010; Schuur *et al.*, 2011).

Mutations in *CTSD* were first associated with a very severe, congenital form of NCL resulting in death early in life, with several patients later described presenting in late infancy or early school age (Mole, no date). Table 6-3 shows a summary of the clinicogenetic characteristics of known CLN10 patients.

Only one confirmed case of juvenile onset CLN10 has been previously reported (Steinfeld *et al.*, 2006). Compound heterozygous missense mutations c.685T>A, p.F229I and c.1149G>C, p.W383C were identified in a German patient who exhibited a very similar phenotype to the cases we report in this study. Functional analysis of CatD activity in patient fibroblasts showed that the maximal enzyme velocity (V_{max}) was 14% of the level found in control fibroblasts.

In a naturally occurring American Bulldog pedigree, a homozygous missense mutation c.597G>A, p.Met199Ile in exon 5 of *CTSD* was associated with a young-adult onset of NCL (Awano *et al.*, 2006). Residual CatD activity in brain samples from affected dogs was 36% of the level found in control dog brains.

I have shown that both mutations identified result in a significantly reduced CatD enzyme activity in patient fibroblasts and leucocytes when compared to controls. The juvenile onset in both families and the residual enzyme activity detected is consistent with previous reports in both humans and animals that residual CatD activity results in a milder phenotype and delayed age of onset (Awano *et al.*, 2006)

Family	Patient	Country of origin	CTSD mutation	CTSD activity (% of control)	Phenotype	Age at onset	Age at death	Clinical features	Cerebellar atrophy	Neurophysiology	Histopathology	Reference
A	A:3	Somalia	p.Gly149Val - hom	-	juvenile NCL	~ 15yrs	33 yrs.	ataxia; RP; cognitive decline	yes	normal	muscle: granulovacuolar material in angular atrophic fibres, GRODs	(Hersheson et al., 2014)
	A:4	Somalia	p.Gly149Val - hom	-	juvenile NCL	~ 15yrs	34 yrs.	ataxia; RP; cognitive decline; cardiomyopathy	yes	normal	-	(Hersheson et al., 2014)
	A:5	Somalia	p.Gly149Val - hom	-	juvenile NCL	~ 15yrs	-	ataxia; RP; cognitive decline	yes	sensory axonal neuropathy	-	(Hersheson et al., 2014)
	A:8	Somalia	p.Gly149Val - hom	11%	juvenile NCL	~ 15yrs	-	ataxia; RP; cognitive decline	yes	normal	-	(Hersheson et al., 2014)
B	B:3	Somalia	p.Arg399His - hom	11%	juvenile NCL	~ 8yrs	-	ataxia; RP; cognitive decline	yes	sensory axonal neuropathy	muscle: granulovacuolar material in angular atrophic fibres, GRODs; fibroblasts: membrane bound vacuoles	(Hersheson et al., 2014)
C	-	Germany	p.Phe229Ile; p.Trp383Cys	14%	juvenile NCL	early school-age	-	ataxia; RP; cognitive decline	yes	-	skin: GRODs, myelin-like lamellar structures	(Steinfeld et al., 2006)
D	-	USA	p.Ser100Phe - hom	~ 0%	congenital NCL	birth	2 days	microcephaly, hypertonia	yes	-	brain: hypercellular cortex, autofluorescent storage material in neurons, astrocytes and macrophages	(Fritchie et al., 2009)
E	-	Pakistan	p.Tyr255X - hom	~ 0%	congenital NCL	Birth	1 day	microcephaly, hypertonia, seizures	yes	-	brain: extensive neuronal loss and glial activation, autofluorescent storage material in all cell types	(Steinfeld et al., 2006)
F	-	Germany	c.268_269insC - hom	~0%-	Congenital NCL	Birth	4 weeks	severe global cerebral and cerebellar hypoplasia, ventriculomegaly, pachygyria	yes	-	brain: extensive neuronal loss, autofluorescent material in neurones and astrocyte, absent myelin	(Meyer et al., 2015)
G	-	Italy	p.Glu69Lys - hom	near absent	Early infantile NCL	By 8 months	-	Microcephaly, progressive epilepsy, hypertrophic cardiomyopathy	yes	'vanishing' EEG pattern	GRODS in skin biopsy	(Doccini et al., 2016)

Table 6-3 Summary of the clinical, genetic and histological findings in the reported cases of CLN10

6.5.2.1 Analysis of CTSD p.G149V and p.R399H mutations identified in this study

The p.G149V variant is in a region conserved both throughout vertebrate species but also conserved throughout all human A1 peptidases. Residue 149 is located within a flexible loop structure (residues F138-G149 in human CatD sequence), commonly referred to as a flap, around the Y142 active site residue (Sielecki *et al.*, 1990). Flap structures are known to contribute to enzyme specificities and are essential for efficient catalytic activity (Okoniewska, Tanaka and Yada, 1999). The reduction in enzyme activity seen in patient fibroblasts and leucocytes may be the result of disturbed flap dynamics caused by the amino-acid substitution at this position.

The p.R399H mutation also affects an amino acid that is conserved throughout vertebrate species and most human peptidases of the A1 peptidase family (Figure 8-3). This mutation is in close proximity to the reported p.W383C mutant which was investigated functionally by Steinfeld et al (Steinfeld *et al.*, 2006). This mutant allele resulted in impaired processing and intracellular mistargeting of the 53kDa precursor pro-enzyme to the mature 33-kDa peptidase. A similar mechanism may underlie the pathogenicity of the p.R399H mutation.

Subsequent to our published reports on these cases, functional work by other groups has been undertaken which confirms the pathogenicity of the G149V and R399H mutations. Expression of these and other CLN10 associated variants in an SH-SY5Y *CTSD* knockout cell model showed that these variants were associated with significantly impaired Cat-D protein maturation, significantly reduced CatD activity and abnormal lysosomal localisation (see figure 6-8) (Bunk *et al.*, 2021).

CTSD deficiency may yet be proven treatable with enzyme-replacement therapies. Several enzyme replacement therapies for other lysosomal storage disease have been

approved (Gaucher disease, Fabry disease, Pompe disease and mucopolysaccharidosis types 1, 2 and 6. Enzyme replacement for CLN2 (cerliponase alfa) was licensed and authorised by NICE for use in the UK in 2019. Cerliponase alfa is infused directly into the brain using an implanted device and has been shown to restore enzyme activity and significantly slow the progressive decline in motor and language function by 80% (Schulz *et al.*, 2018). The current cost for this effective intervention is estimated at around £500,000 per patient per annum.

A proof-of principle study in a mouse model of CLN10 has shown that recombinant human pro-CTSD is able to ameliorate impaired lysosomal function in hippocampal slice cultures in vitro and retinal cells in vivo.

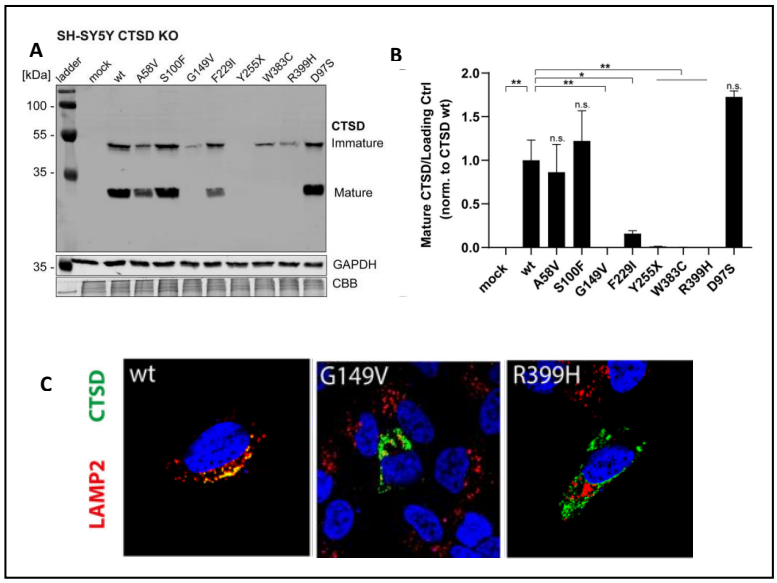


Figure 6-8 A) Immunoblot of transiently overexpressed CatD wildtype and CLN10 associated CTSD variants including G149V and R399H (red arrow) B) Quantification of western blot showing undetectable expression of G149V and R399H CatD variants C) Immunofluorescence studies showing abnormal lysosomal localisation of G149V and R399H variants – adapted from (Bunk et al.*, 2021).*

6.5.2.2 Muscle abnormalities associated with CatD deficiency:

The abnormal muscle biopsy from patient B:3 was interesting. This showed a very distinctive pattern of muscle pathology characterised by granular vacuolar material in angular atrophic fibres. This pattern of morphological abnormalities caused significant diagnostic difficulty in this patient prior to identification of the gene but may be sufficiently distinctive to permit a provisional diagnosis in future cases based on muscle biopsy. Similar findings were seen on the more limited biopsy material which was available for patient A:3 who also had no clinical evidence of myopathy. Her sister however died as a result of complications from cardiomyopathy.

Deleted: on the basis of

Deleted: the

Although not used diagnostically, skeletal muscle is frequently involved in other NCL subtypes, with ultrastructural characteristics of NCL often seen on muscle biopsy (Rapola and Haltia, 1973) (Carpenter S, Karpati G, 1972). Clinical myopathy is however less frequently reported. Dilated cardiomyopathy has been reported at autopsy in a Japanese CLN3 case (Tomiyasu H, Takahashi W, Ohta T, Yoshii F, Shibuya M, 2000). Recently it has been demonstrated in a zebrafish model that knockdown of CatD in zebrafish fertilised embryos resulted in a congenital myopathy (Follo *et al.*, 2013).

6.5.3 Conclusion

Although many of the NCLs share some characteristic phenotypes, genetic and phenotypic heterogeneity often results in difficulty establishing a genetic diagnosis. Currently enzyme analysis for NCL is only routinely performed for TPP1 (CLN2) and PPT1 (CLN1). However, *CTSD* mutations remain a very rare cause of NCL and next-generation approaches such as exome sequencing and targeted gene sequencing are increasingly useful in establishing diagnoses for these rare neurodegenerative conditions. Furthermore, the characteristic muscle pathology seen in CTSD mutations may help to

guide clinical assessment, aid molecular diagnosis in NCL patients and help support or confirm pathogenicity of novel CTSD mutations.

7 Identification of SYNE1 associated cerebellar ataxia cases – genetic and clinical characterisation

7.1 Statement of contribution

I performed the MiSeq design and sequencing of the *SYNE1* gene to identify the 3 patients reported in this study. Genetic analysis was jointly performed with Dr Sarah Wiethoff who performed the Sanger sequencing. Clinical evaluation was undertaken by me and Prof Houlden.

The results from this study were published in a paper co-authored by myself and Dr Wiethoff (Wiethoff *et al.*, 2016)

Publications arising from this section:

- Heterogeneity in clinical features and disease severity in ataxia-associated SYNE1 mutations. Wiethoff S**, Hersheson J,** Bettencourt C, Wood NW, Houlden H. J Neurol. 2016 Aug;263(8):1503-10

Deleted: Syne

Deleted: -

Deleted: myself

7.2 Introduction

The recessive cerebellar ataxias are a complex group of neurodegenerative disorders with significant genetic and phenotypic heterogeneity. These are frequently early-onset multisystem disorders that present with a range of neurological manifestations including motor and cognitive deterioration, visual disturbance, epilepsy and neuropathy. SYNE1 was the first recessive ataxia associated with a pure cerebellar phenotype, the recessive ataxia of Beauce (SCAR8), and was first identified in a number French-Canadian families originating from the Beauce and Bas-St-Laurent regions of Quebec (Gros-Louis *et al.*, 2007).

In this study I examined a UK cohort of recessive cerebellar ataxia patients, identified through a combination of targeted next generation sequencing and exome sequencing. Novel truncating mutations in SYNE1 were identified in a further four cases from three British, Turkish and Sri Lankan pedigrees.

7.3 Methods

7.3.1 Patients

All patients were recruited through the Neurogenetics clinic at the National Hospital for Neurology and Neurosurgery, Queen Square, London. Patients all had a diagnosis of progressive cerebellar ataxia with either known autosomal recessive inheritance or sporadic cases with an early onset of disease. In total 196 patients were screened for SYNE1 mutations through either exome sequencing (110 patients) or targeted next generation sequencing (86 patients).

7.3.2 Genetic analysis

DNA was extracted from peripheral leucocytes of all patients using standard procedures. Additional samples were taken from affected or unaffected relatives to check for mutation segregation where appropriate.

Exome sequencing libraries were prepared using Illumina Nextera Focussed exome kits according to standard protocols. Libraries were indexed and sequenced on an Illumina HiSeq 2500 machine.

A custom sequencing panel was designed to amplify the coding exons of SYNE1 using the Illumina TruSeq custom amplicon kit. Libraries were prepared according to standard protocols and then sequenced on an Illumina MiSeq platform.

Bioinformatic analysis was the same for both exome and targeted NGS. Reads were aligned to the hg19 genome build using Novoalign with variant calling performed using SAMTools. Variant annotation was performed with ANNOVAR and coverage metrics using a modified Bedtools coverageBed script. All SYNE1 annotations are for the RefSeq NM_0330714 transcript.

Due to the large size of the SYNE1 gene which comprises >140 exons, a large number of private missense mutations are often observed on sequencing. The list of variants was filtered according to the following criteria: Qual > 30, Depth > 10, minor allele frequency in Exome Variant Server and 1000g <0.005, not synonymous and present in a homozygous or compound heterozygous state.

Deleted: and if

Variants were confirmed using Sanger sequencing in affected cases and in parents or unaffected siblings to confirm the mutation phase in compound heterozygous mutations.

Deleted: in

7.4 Results

7.4.1 Family 1

Patient I:1 and I:2 are siblings from non-consanguineous English family. There was no reported motor or cognitive delay however at school their handwriting was noted to be poor, and were poor at sports, unlike their siblings. Patient I:I suffered a major head injury at the age of 21 and was noted to have dysarthric speech following his accident, initially thought to be due to multiple jaw fractures. His gait and mobility deteriorated from age 40 when he noticed difficulty rising from a chair and sustained several falls when walking. At the time of his last clinical assessment, he was 65 years old and had significant gait ataxia and dysarthria. His sister started experiencing progressive gait and speech difficulties from age 32 following a road traffic accident after she also was involved in a road traffic accident.

Examination of both siblings revealed clinical signs of cerebellar ataxia with broken ocular pursuits, dysarthria with limb and gait ataxia. There were no pyramidal signs or evidence of neuropathy or myopathy. Electromyogram and neurophysiology were normal. Both patients exhibited cerebellar atrophy on MRI.

Both siblings shared novel compound heterozygous truncating mutations SYNE1 in exon 18 (c.G1849T:p.E617X) and exon 99 (c.G18431A:p.W6144X). Mutation screening was undertaken in one unaffected sibling who was heterozygous for only the p.W6144X thus confirming the mutation phase to be trans.

Deleted: poor

Deleted: they

Deleted: now

Deleted: s

Deleted:

Deleted:

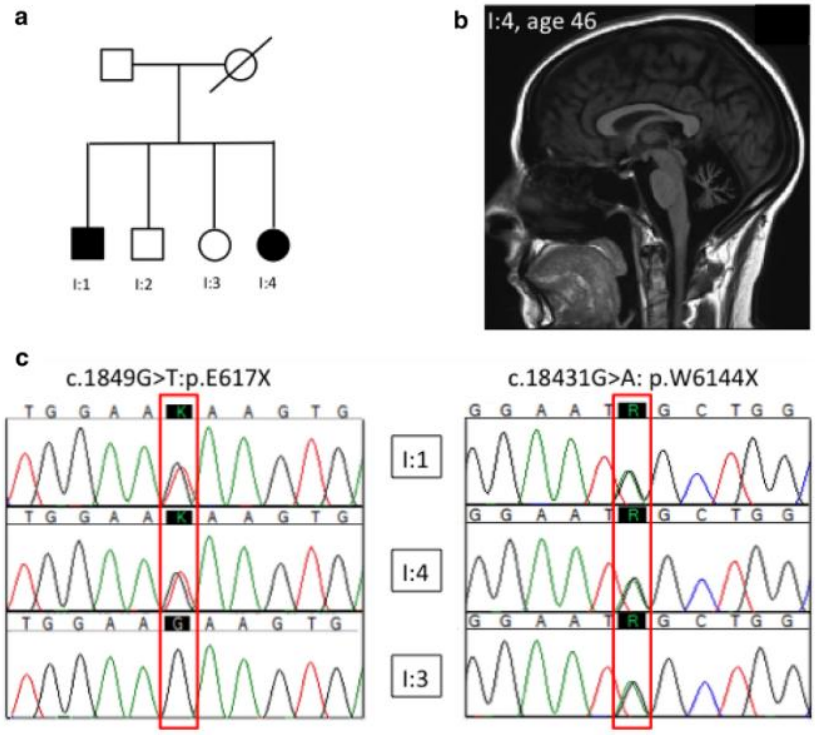


Figure 7-1 A) Pedigree of family 1 - affected cases are in black B) Sagittal MRI brain of patient I:4 showing cerebellar atrophy C) Sequence chromatograms showing compound heterozygous mutations in I:1 and I:4. The unaffected sibling carries only the W6144X mutation (reproduced from (Wiethoff et al*., 2016))*

7.4.2 Family 2

Patient 2:1 is from consanguineous parents (1st cousins) of Turkish origin. She is one of 11 siblings, 4 of whom were also affected. She developed progressive gait ataxia and dysarthria from age 18. At age 23 she was still mobilising independently however had a high frequency of falls and required minor adaptations to her home. Her four affected siblings all developed symptom onset in their late teens and were reported to be very similar to the index case. Further clinical details about her affected siblings were not available.

Deleted: 4

Clinical examination of the index patient showed broken pursuit eye movements, moderate limb ataxia and significant gait ataxia. Lower limb reflexes were brisk with sustained clonus at both ankles and extensor plantar reflexes.

Brain MRI of the index case showed marked cerebellar atrophy. Nerve conduction studies was normal.

A novel homozygous variant was identified in exon 108 of SYNE1 (c.19879C>T p.Q6633X). DNA for the parents and her siblings was not available.

7.4.3 Family 3

The proband is a Sri Lankan man who developed gait and balance problems at 22 years. His siblings were unaffected. There was no known motor or cognitive delay and was active in sports as a child. His parents were not known to be related. Clinical examination showed hypermetric saccades, limb ataxia and extensor plantar reflexes. There was no hyperreflexia or limb spasticity.

Brain MRI showed cerebellar atrophy particularly in the vermis. Nerve conduction studies and electromyography were normal.

A novel homozygous variant in SYNE1 c.C13429T:p.Q4477X was identified.

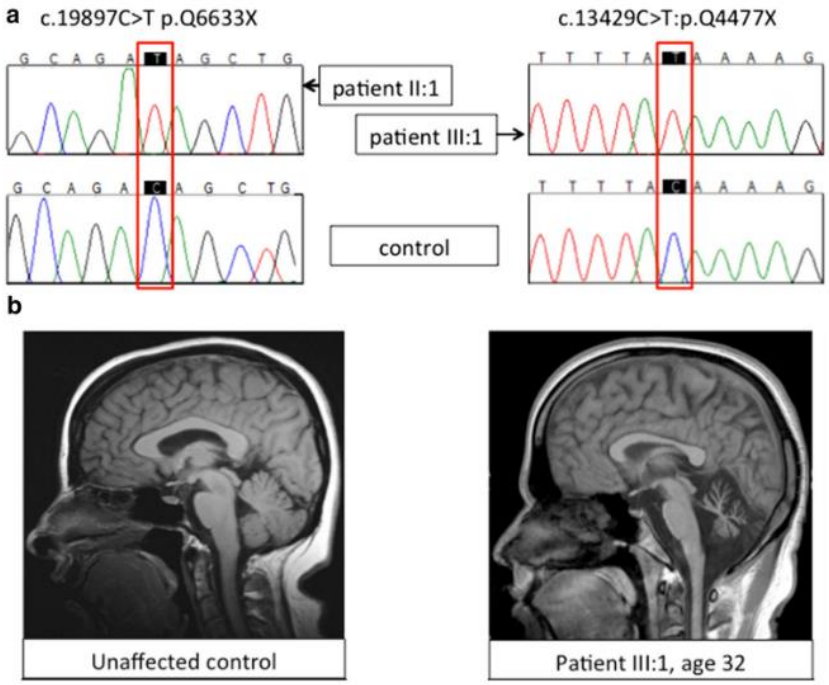


Figure 7-2 A) Sequence chromatograms showing homozygous Q6633X mutation in patient II:1 and homozygous Q4477X mutation in patient III:1 B) sagittal MRI brain showing cerebellar atrophy in patient III:1 – reproduced from (Wiethoff et al.*, 2016)*

7.4.4 Discussion

SYNE1 encodes the Spectrin repeat containing nuclear envelope protein 1, an exceptionally large ~1 MDa protein which is highly expressed in human cerebellum and is a structural protein thought to link organelles to the actin cytoskeleton.

A large multicentre study determined that *SYNE1* mutations account for 5.3% of early-onset cerebellar ataxia after Friedreich's ataxia had been excluded (Synofzik et al., 2016). The study further expanded the clinical phenotype associated with *SYNE1* mutations. Whilst the first reported cases demonstrated a relatively pure ataxic phenotype, *SYNE1* mutations have been associated with a range of non-cerebellar

Formatted: Font: Italic

features including motor neurone disease and mental retardation. The study showed the isolated ataxic phenotype to be relatively uncommon, accounting for 19% of cases with 81% showing additional non-cerebellar features. Variable combinations of upper and lower motor neurone dysfunction (58% of cases) were observed across what can be considered a spectrum of severity ranging from pure cerebellar ataxia, to ataxia with motor neurone features, to a complex childhood onset multisystemic syndrome comprising ataxia, spasticity, mental retardation and a range of skeletal and soft tissue abnormalities. No clear genotype/phenotype correlations have been determined. Muscle biopsies showed significantly reduced or absent staining for SYNE1 although a myopathic phenotype was not typically seen. Two families with a rare form of arthrogryposis, arthrogryposis multiplex congenita were found to have biallelic mutations in *SYNE1* – the affected patients exhibited abnormal muscle pathology with fatty replacement of muscle tissue, fibre size variation and abnormalities of the nuclear membrane (Attali et al., 2009).

Homozygous loss of function mutations in *SYNE1* have been reported in French-Canadian, French, Japanese and Brazilian individuals (Synofzik *et al.*, 2016).

In this study, all the cases reported are associated with mutations likely to result in a truncated protein or loss of protein production. There appears to be little correlation between the position of the stop codon in the reading frame and the severity or age at onset although overall *SYNE1*-associated cerebellar ataxia is relatively mild and slowly progressive disease. The affected cases in families 1 and 3 are typical of previously reported cases with a pure cerebellar phenotype with no neuropathy or significant pyramidal signs.

Deleted: all of

The proband from family 2 however exhibited marked lower limb pyramidal signs with hyperreflexia and ankle clonus. Of note there were no variants in any other ataxia or

hereditary spastic paraparesis genes identified in the exome sequence data of this individual. Recently, two other Turkish families have been described with affected members manifesting a spastic ataxia phenotype associated with two novel homozygous *SYNE1* nonsense mutations (R7842X and Q7644X). As with the age of onset and disease severity, there does not appear to be a correlation between the location of the truncating mutation and the presence of significant pyramidal signs. It is possible that genetic or epigenetic modifiers are contribute to the phenotypic variability and it is recommended therefore that *SYNE1* mutations be considered also in the aetiology of both complex and pure recessive ataxia.

Heterozygous mutations in *SYNE1* have also associated with Emery-Dreifuss muscular dystrophy type 4 (EDMD4), an inherited muscle disease associated with progressive muscle weakness and atrophy, contractures and cardiac abnormalities (Zhang et al., 2007). There appears however to be little phenotypic overlap with the reported muscle abnormalities seen in recessive *SYNE1*-ataxia cases and EDMD4 cases. It is possible that a different pathogenetic mechanism underlies the two disorders however it has been suggested that a re-interpretation of the significance of heterozygous *SYNE1* mutations in EDMD4 cases is warranted and whether the reported mutations are truly pathogenic.

Further development of next-generation techniques in clinical diagnostics will enable *SYNE1*-associated cerebellar ataxia to be more readily identified in future and at lower cost.

Deleted: -

Deleted: mutations

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: is warranted

Formatted: Font: Italic

8 A broad spectrum of clinical phenotypes is associated with mutations in *PNPLA6*

8.1 Statement of contribution

Oliver-Mcfarlane/Laurence-Moon syndromes study: I undertook a clinical re-evaluation of the index case in family A, performed exome sequencing analysis and mutation confirmation. I facilitated the re-evaluation of the stored muscle biopsy for the patient. The muscle histology analysis was performed by Prof Tamas Refesz. The case was published as part of a larger series of Oliver-McFarlane and Laurence Moon syndrome patients by Robert Hufnagel (Hufnagel *et al.*, 2015). Homozygosity mapping and exome sequencing and Sanger confirmation for the other reported cases were undertaken by collaborators.

Cerebellar ataxia case: I performed the exome sequencing analysis jointly with Sarah Wiethoff. Homozygosity mapping was performed by Reema Paudel.

Publications arising from this section:

Neuropathy target esterase impairments cause Oliver-McFarlane and Laurence-Moon syndromes. Hufnagel RB, Arno G, Hein ND, **Hersheson J**, Prasad M, Anderson Y, Krueger LA, Gregory LC, Stoetzel C, Jaworek TJ, Hull S, Li A, Plagnol V, Willen CM, Morgan TM, Prows CA, Hegde RS, Riazuddin S, Grabowski GA, Richardson RJ, Dieterich K, Huang T, Revesz T, Martinez-Barbera JP, Sisk RA, Jefferies C, Houlden H, Dattani MT, Fink JK, Dollfus H, Moore AT, Ahmed ZM. J Med Genet. 2015 Feb;52(2):85-94. doi: 10.1136/jmedgenet-2014-102856. Epub 2014 Dec 5

Pure Cerebellar Ataxia with Homozygous Mutations in the PNPLA6 Gene. Wiethoff S, Bettencourt C, Paudel R, Madon P, Liu YT, **Hersheson J**, Wadia N, Desai J, Houlden H. Cerebellum. 2017 Feb;16(1):262-26

8.2 Introduction

Figure 8-1 Composite newspaper articles showing headlines related to 'Jamaica Ginger' outbreaks in Prohibition-era USA. Bottles of original 'Jamaica Ginger' extract on display - source Wikipedia/public domain

Mutations in patatin-like phospholipase domain-containing 6 (*PNPLA6*) have been associated with a particularly diverse range of neurodegenerative phenotypes (Synofzik *et al.*, 2014). *PNPLA6* was previously known Neuropathy Target Esterase (NTE), and was prior to 2008, mainly associated with a syndrome caused by exposure to organophosphates called organophosphorus compound-induced delayed neuropathy (OPIDN) (Melentev, Agranovich and Sarantseva, 2020). The first reports of this syndrome came in the USA in the 1930s during Prohibition. At the time, a popular 'medicinal' beverage Jamaica Ginger was frequently adulterated by bootleggers who added toxic triorthocresil phosphate (TOCP) instead of bitter Jamaica Ginger extract (Aring, 1942). This resulted in a syndrome characterised by severe length dependant axonal neuropathy and spasticity affecting an estimated 30-50000 individuals at the time (Melentev, Agranovich and Sarantseva, 2020). PNPLA6 (then NTE) was first characterised as the key enzyme in the aetiology of OPIDN in 1969 (Johnson, 1969) and remains important in the assessment of delayed toxicity of organophosphates including pesticides and chemical weapons.

In 2008, mutations in PNPLA6 were first associated with a human neurodegenerative syndrome (Rainier *et al.*, 2008). Rainier et al investigated a consanguineous Ashkenazi pedigree with affected individuals manifesting a childhood onset motor neuronopathy with distal myotrophy and pyramidal signs. Homozygosity mapping identified a single ~4Mb region of shared homozygosity containing 127 genes including PNPLA6. A candidate gene approach prioritised this for sequencing due to similarities between the clinical phenotype in the family and OPIDN.

Subsequently, PNPLA6 mutations were identified in 12 patients with either Boucher-Neuhauser syndrome (BNS), Gordon Holmes syndrome (GHS) or spastic ataxia (Synofzik *et al.*, 2014). Boucher-Neuhauser and Gordon Holmes syndromes are both characterised by early onset cerebellar ataxia and hypogonadotropic hypogonadism however Boucher-Neuhauser Syndrome also includes chorioretinal dystrophy as a clinical feature.

Deleted: and

Oliver-Mcfarlane Syndrome (OMS) was first described in case report, published in 1965, of a child with congenital hypopituitarism, chorioretinal dystrophy and trichomegaly (long eyelashes) (Oliver and Mcfarlane, 1965). The child had been examined at 30 months of age and it was not documented (or possible to assess) whether there was cerebellar ataxia however this has been confirmed as a feature in subsequent case reports of OMS.

Nearly a century before this in 1866, Laurence and Moon described a family with 4 affected children with pigmentary retinopathy, hypopituitarism and either cerebellar ataxia or spastic paraplegia (Laurence and Moon, 1995). Ataxia was a feature of most subsequent documented cases. (Andrew P. Schachat and Maumenee, 1982). Unfortunately for many years, Laurence-Moon syndrome (LMS) was erroneously conflated with Bardet-Biedl syndrome (BBS) (historically termed Laurence-Moon-Bardet-Biedl syndrome) which comprises pigmentary retinopathy, hypogonadism and digital

abnormalities such as syndactyly. Ataxia or spastic paraplegia were not reported associations of BBS and it is generally now considered that , Laurence-Moon syndrome and Bardet-Biedl syndromes are distinct entities (Andrew P Schachat and Maumenee, 1982). Moreover, Bardet-Biedl syndrome exhibits significant genotypic heterogeneity with >20 genes known to be associated with the disorder.

In this chapter I report the clinical and genetic reappraisal of a case of Oliver-Mcfarlane Syndrome, identifying pathogenic compound heterozygous mutations in PNPLA6 as the underlying cause. This work was a collaboration with colleagues in the UK, USA, France and New Zealand with a total of total 6 Oliver-Mcfarlane Syndrome cases and 4 Laurence-Moon syndrome cases all of whom were found to have mutations in PNPLA6. I further report a patient with a pure cerebellar ataxic phenotype found to have a presumed pathogenic homozygous mutation in PNPLA6, expanding the phenotypic spectrum to encompass a pure cerebellar phenotype, complex ataxia with pigmentary retinopathy and hypogonadism, complex spastic paraplegia with motor neuronopathy and uncomplicated HSP.

8.3 Materials and Methods

8.3.1 Patient identification and clinical characterisation

8.3.1.1 OMS/LMS cases

Patient E:1 in family 1, who had been clinically diagnosed with Oliver-McFarlane syndrome, was identified from clinical information submitted to the clinical diagnostic neurogenetics database at NHNN. He was deceased at the time of the research but had previously been clinically assessed by Michael Baraitser and Anita Harding at the UCL Institute for Child Health, Great Ormond St. I reviewed his clinical records with additional information obtained from the published case report (Patton, Harding and Baraitser, 1986).

Additional cases of Oliver McFarlane syndrome (5 patients in 4 families) and Laurence Moon Syndrome (4 patients in 1 family) were identified by collaborators in Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, Ohio, Laboratoire de génétique Médicale, Université de Strasbourg, FMTS, Strasbourg, France and Department of Paediatrics, Taranaki Base Hospital, New Plymouth, New Zealand.

8.3.1.2 Pure cerebellar ataxia case

Patient B:1 from a multigenerational Parsi family, is an 80 yr. old male, under the care of Dr N Wadia (Director Emeritus at the Department of Neurology, Jaslok Hospital, Mumbai). He was first assessed in his 20s and remained under periodic clinical review. At age 57 he was assessed, with his affected 1st cousin (B:2), by Dr Desai at the Department of Assisted Reproduction and Genetics. Dr Desai undertook detailed clinical neurological and genetic assessment of the affected individuals. Both patients were screened for SCA1, 2,3, 6, 7, 11, 15, FRDA, HD and DRPLA.

8.3.2 DNA and RNA sample collection and preparation

DNA was extracted from whole blood from A:1 and his unaffected sister A:2 for use in subsequent exome sequencing and variant screening.

DNA was extracted from the two affected cousins in family B and 20 unaffected relatives. RNA was extracted from whole blood from B:1 to be used for cDNA synthesis to assess the effect of any variants on splicing.

8.3.3 Exome sequencing

Exome sequencing and bioinformatic analysis for A:1 was performed as described in the Materials and Methods section. Exome sequencing for B:1 was performed using NimbleGen SeqCap Target Enrichment EZ-system (Roche Sequencing) and Illumina 76 bp paired end sequencing on an Illumina GAII platform as a commercial service.

Comparable methods were employed by the teams investigating the other Oliver-McFarlane Syndrome and Laurence-Moon syndrome cases.

8.3.4 Homozygosity mapping

Patients B:1 and B:2 underwent whole genome genotyping using an HumanOmniExpress BeadChip Kit (Illumina). Homozygosity mapping was performed using the Homozygosity Detector plugin within the GenomeStudio Software package (Illumina). Shared regions of homozygosity were visualised, and the region used to refine the list of exome sequencing variants.

Deleted: visualised

8.3.5 Sanger sequencing

Primers for the candidate variants in PNPLA6 in patient A:1 and B:1 were designed as per the Materials and Methods section. The PNPLA6 mutations identified in A:1 were confirmed using Sanger sequencing and also screened in his unaffected sister to assess the phase of the variants.

Mutation genotyping was performed in B:1, B:2 and 20 other unaffected relatives in family B were screened for the putative pathogenic PNPLA6 variant. The sequence around the variants was checked in B:1 to assess for effect on splicing.

8.3.6 Neuropathological assessment of post-mortem brain of A:1

Post-mortem brain tissue was kindly donated and obtained from Queen Square Brain Bank, University College London. This project was approved by the Joint Local Research Ethics Committee of the National Hospital for Neurology and Neurosurgery and the UCL Institute of Neurology.

Paraffin blocks from representative areas of the brain were available for histological and immunohistochemical investigation. In brief, 8µm thick sections of paraffin embedded tissue blocks were deparaffinised and rehydrated through graded alcohols. Sections

were placed in 0.3% H202/methanol to block endogenous peroxidase activity. Antigen retrieval was undertaken to unmasked antigen epitopes and was achieved either by pressure cooking in citrate buffer (ph. 6.0) for 10 minutes, incubation in 99% formic acid for 10 minutes or incubation in Proteinase K solution (DAKO) for 10 minutes. The sections were washed with tris-buffered saline and incubated in 10% non-fat milk solution to prevent non-specific antibody binding. The appropriate primary antibody was applied (see Supplementary Table), followed by an anti-mouse or anti-rabbit secondary antibody (Dako, Ely, UK) at 1/200 dilution, as appropriate. Subsequently avidin-biotin complex Elite kit (Vector, Peterborough, UK) was applied, and colour was developed by diaminobenzidine/H2O2. Sections were finally counterstained with Mayer's haematoxylin. Routine haematoxylin and eosin histological staining was also carried out.

Deleted: applied

8.4 Results

8.4.1 OMS/LMS cases: Clinical phenotype of A:1

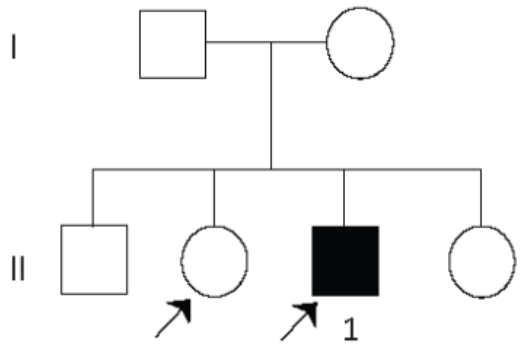


Figure 8-2 Pedigree of patient A:1 (highlighted in black) and his unaffected sister

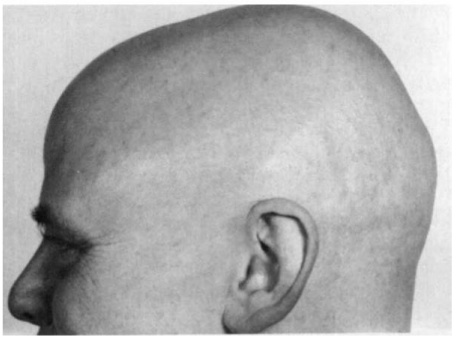
The pedigree is shown in figure 8-2. A:1 is a British Caucasian male. His parents and 3 siblings were unaffected. There was no parental consanguinity. He was initially examined at 2 years due to failure to thrive and later at age 37 years. He had a low birth weight, delayed motor milestones and was noted to have night-blindness by the age of 2 ½ years. He was noted to have very long eyelashes (20mm) with bushy eyebrows and sparse scalp hair. He had a small penis, and testes were not palpable. His vision subsequently deteriorated and when assessed at age 37 years was only able to see shapes.

Deleted: 10

Deleted: penis

At age 10 he was investigated for obesity and gynecomastia and was treated from age 14 with thyroxine and testosterone. Secondary sexual characteristics never developed. Ophthalmological assessment as a child showed severe pigmentary retinopathy with relatively normal optic discs.

He attended a school for visually impaired children and IQ on Wechsler Adult Intelligence Scale (WAIS) was assessed at 93 in his teens. His coordination was never normal but deteriorated in his 30s.



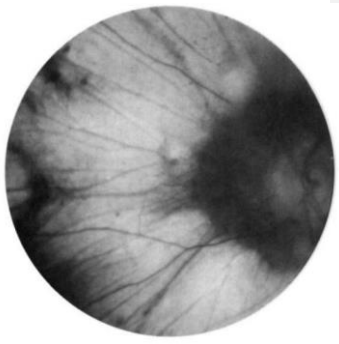


Figure 8-3 A) Photograph of patient A:1 taken at 37yrs of age. Bushy eyebrows and long eyelashes can be seen in addition to scalp alopecia. B) Retinal photograph showing pigmentary retinopathy – reproduced from (Patton, Harding and Baraitser, 1986).

When examined at age 37, there was alopecia of the scalp with evidence of trichomegaly. He was short and obese with evidence of hypogenitalism. Visual acuity of perception of light bilaterally with marked pigmentary degeneration of the retina. He had severe myopia and nystagmus. He had a broad-based ataxic gait. Power and tone were normal in the limbs but there was mild lower limb ataxia. Sensory examination was normal. Tendon reflexes were absent and both plantar reflexes were extensor.

Investigations as a child showed hypothyroidism, growth hormone deficiency and hypogonadotropic hypogonadism. CT brain showed evidence of cerebellar atrophy and an empty sella. Nerve conduction studies showed a sensory axonal neuropathy.

At 46 years he was treated for nodular sclerosing Hodgkin's lymphoma. He died at age 67 from disseminated Hodgkin's lymphoma.

A detailed clinical assessment of the other Oliver-McFarlane Syndrome and Laurence-Moon syndrome cases can be found in the supplementary information in (Hufnagel *et al.*, 2015).

8.4.2 Analysis of Family A, 4 other OMS families and 1 LMS family

8.4.2.1 Genetic analysis

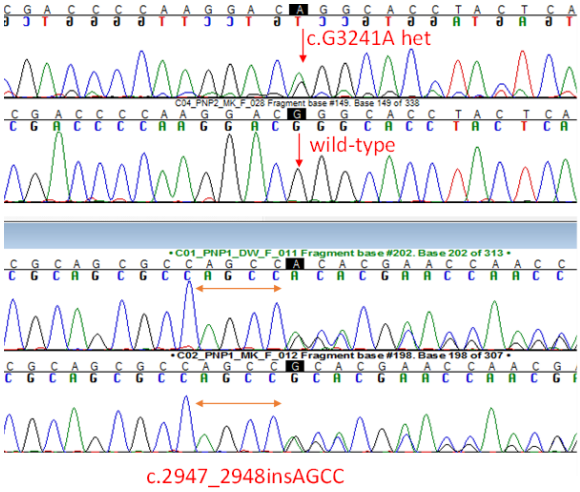


Figure 8-4 Sequence chromatograms showing the c.G3241A heterozygous mutation and c.2947_2948insAGCC in A:1. His unaffected sister carried only the c.2947_2948insAGCC mutation

In A:1, there was a 4bp frame-shift insertion c.2947_2948insAGCC:p.R983fs and a missense mutation c.G3241A:p.G1081R in PNPLA6. Sanger sequencing confirmed both mutations in A:1 and only the frame-shift mutation in A:2, thus confirming that the mutation phase was trans. The missense mutation was predicted to be deleterious using SIFT, Polyphen2 and MutationTaster. The amino acid site is highly conserved across species. The p.R983fs mutation had previously been identified as pathogenic in a Boucher-Neuhauser Syndrome case (Synofzik *et al.*, 2014) and at the time of our study the p.G1081R mutation was not present in EVS or 1000G however it has subsequently been reported in 3 other unrelated Oliver-McFarlane Syndrome patients (Kmoch *et al.*, 2015). The mutation is present in the gnomAD database with an MAF of 0.000032.

The table below describes the filtering strategy employed in the analysis of A:1. Based on the documented clinical phenotype and the clear overlap with that of Boucher-Neuhauser syndrome, PNPLA6 was prioritised as the candidate gene.

Filtering strategy:	**Variants**	**Genes**
Exonic	16299	7768
Quality >20, depth >10, segdup <0.97	9605	4969
Nonsense, frameshift, splice, missense	4453	2804
Frequency: MAF <0.05 1000g, EVS, CG69	409	379
Inheritance:		
A: homozygous	20	20
B: compound het	54	26

Table 8-1 Filtering strategy used for exome sequencing analysis for patient A:1

Exome sequencing identified compound heterozygous variants in PNPLA6 in all 10 patients with Oliver-McFarlane Syndrome and Laurence-Moon syndrome. In the published results, the variants are coded according to the NM_001166111.1 transcript however the Human Genome Mutation Database, and subsequent publications code the variants according to the NM_006702.5 transcript which is used here.

In total 8 different mutations were identified in PNPLA6 in the affected families – five missense, one splice site, one frame-shift insertion and a multi-exon duplication leading to a frameshift. All but one family/individual had compound heterozygous mutations with one LOF mutation (frame-shift, splice mutation) and one missense mutation. All remaining 5 missense mutations were novel and at highly conserved amino acid sites.

Deleted: s

8.4.2.2 Muscle histopathology in patient A:1

The major neuropathological findings include an overall reduction in size of the brain (brain weight: 981g), slight reduction in bulk of the frontal white matter, and midbrain. The substantia nigra and locus coeruleus were well pigmented. There was severe cerebellar cortical atrophy involving both the cerebellar vermis and hemisphere. Microscopy demonstrated a previous ischaemic episode affecting the lateral aspect of the right caudate, internal capsule and putamen and also severe cerebellar cortical degeneration with loss of Purkinje and granule cells, numerous empty baskets and prominent Bergmann gliosis.

Formatted: Font: 11 pt

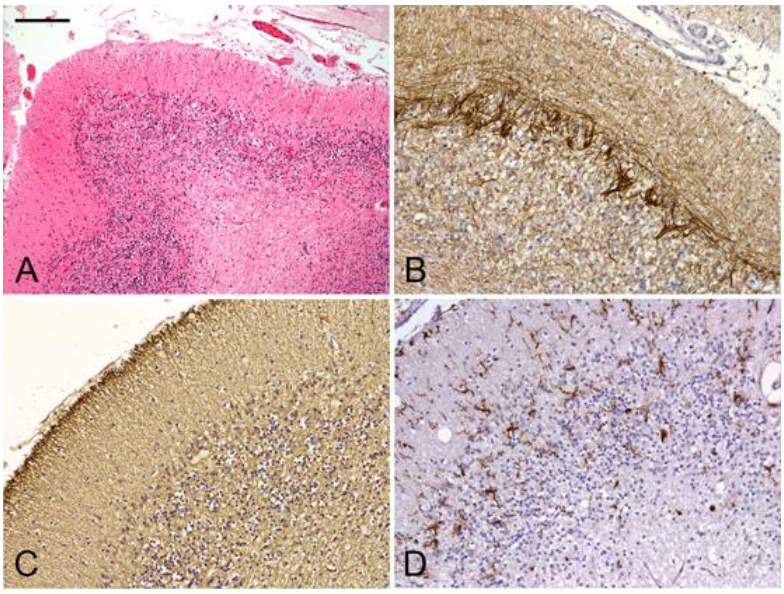


Figure 8-5 Muscle histopathology from patient A:1 - A) Cerebellar cortex showing severe depletion of Purkinje and granule cells.B) Numerous empty baskets, C) severe astrogliosis and D) increase in number of activated microglial cells. (A: Haematoxylin and eosin; B: SMI31(neurofilament) immunohistochemistry; C: GFAP (glial cell marker) immunohistochemistry; D: IBA1 (activated microglial cell marker) immunohistochemistry)

8.4.3 Pure cerebellar ataxia case: Clinical phenotypes of B:1 and B:2

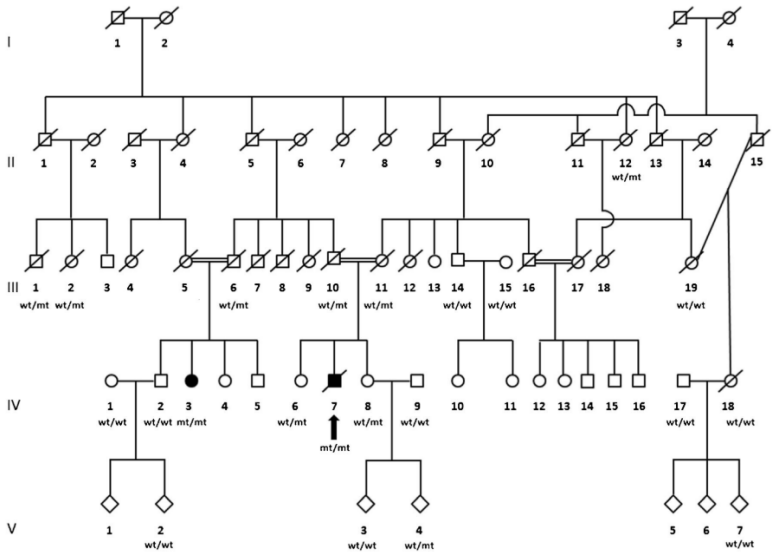


Figure 8-6 Extended pedigree of family of index case B:1 (on this pedigree IV-7, highlighted in black) B:2 is indicated on this pedigree as IV-3 – reproduced from (Wiethoff et al.*, 2017)*

The extended pedigree of family B is shown above. B:1 developed cerebellar symptoms from age 12 although he was noted to have difficulties with handwriting and nystagmus as an infant. There was no cognitive impairment. He had brisk lower limb reflexes and an ataxic gait. There was no evidence of pigmentary retinopathy or hypogonadotropic hypogonadism. He suffered a stroke due to cerebral venous sinus thrombosis at age 60 which resulted in significant disability. MRI at the time of the stroke showed severe cerebellar atrophy. There was no neurophysiological evidence of neuropathy.

Deleted: electro

B:2 developed cerebellar ataxia around puberty and was wheelchair bound when assessed at age 68. Reflexes were brisk with an ataxic gait. There was no neuropathy. Overall B:2 was more mildly affected with respect to cerebellar syndrome compared to her affected cousin although his mobility impairment was almost certainly exacerbated by his unrelated stroke.

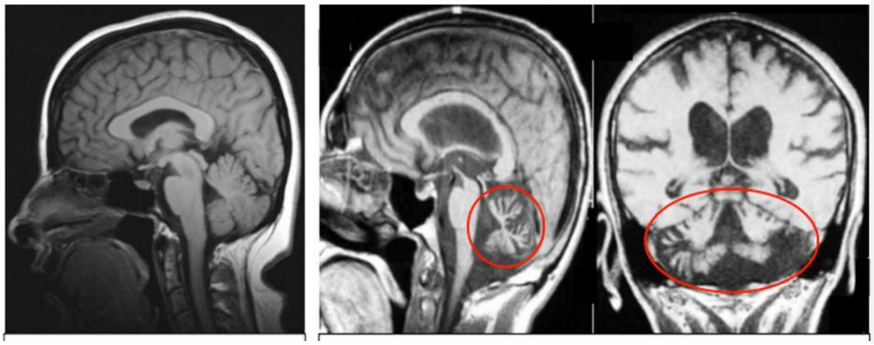


Figure 8-8 A) Sagittal brain MRI from normal control B) Sagittal and coronal MRI brain of B:2 showing severe cerebellar atrophy – adapted from (Wiethoff et al., 2017)

8.4.4 Analysis of Family B

8.4.4.1 Homozygosity mapping

Homozygosity mapping identified a single ~4Mb segment of homozygosity shared between B:1 and B:2 on chromosome 19 (Chr19 3630740 – 7759053). These genomic coordinated were used to filter the exome data obtained for B:1.

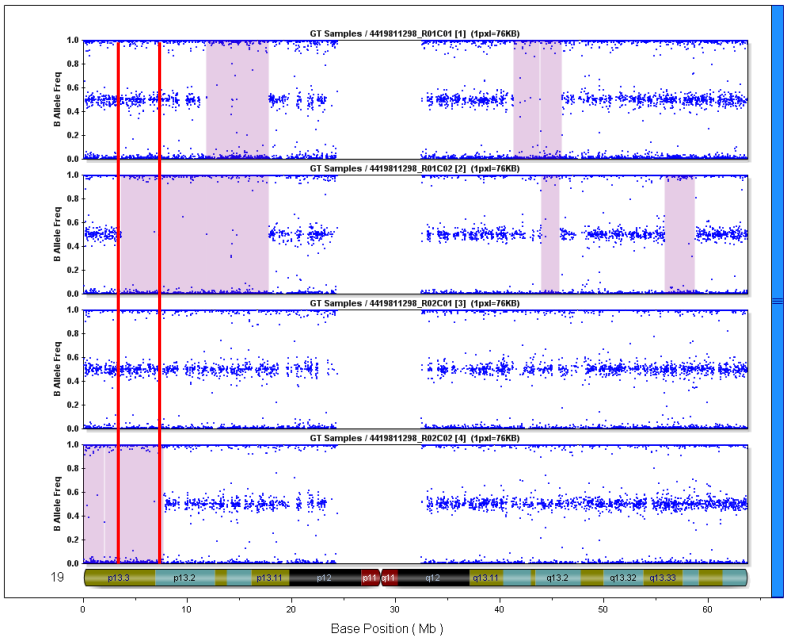


Figure 8-9 Homozygosity mapping of B:1 (upper two plots) and B:3 (lower 2 plots) showing region of shared homozygosity within the 2 red lines

8.4.4.2 Exome sequencing and segregation analysis

The total variants (n=50,279) from exome sequencing in B:1 were filtered to include only homozygous (n=10399), rare or novel nonsynonymous variants (n=91) within the region of shared homozygosity (n=2). The two remaining variants were both homozygous missense mutations in PNPLA6 – p.V1262M and p.D1283V.

Segregation analysis showed that only B:1 and B:2 were homozygous for the 2 variants. 9/20 unaffected family members were heterozygous carriers of the two mutations and 11/20 carried neither variant. cDNA sequencing of the p.V1262M did not identify any impact on splicing.

Neither variant was present in the EVS, dbSNP, cg69 or 1000genomes databases. The p.V1262M variant was predicted to be benign using in-silico pathogenicity prediction and was not at a highly conserved amino acid site. The residue is located in the 1st codon of the exon however there was no predicted splicing impact and cDNA sequencing across the intron/exon boundary did not show any impact on splicing. The p.D1283V however was predicted to be pathogenic and was at a highly conserved amino acid which is supportive of this being a pathogenic variant. Both variants were located within an intrinsically disordered domain within PNPLA6 whereas the majority of the known pathogenic mutations seen in this study and other reported phenotypes (Boucher-Neuhauser Syndrome, Gordon-Holmes syndrome, hereditary spastic paraplegia, isolated retinitis) have clustered in the patatin-like esterase domain containing the active enzymatic site residues, and to a lesser extent the within the cyclic nucleotide monophosphate (cNMP) binding domains.

Deleted: was

Deleted: the

8.5 Discussion

At the time that this study was first published, PNPLA6 mutations had been reported to be associated with three recessively inherited neurodegenerative disorders with overlapping phenotypes – SPG38 (spastic paraplegia and motor neuropathy), Boucher-Neuhauser Syndrome (cerebellar atrophy, chorioretinal dystrophy and hypogonadotropic hypogonadism) and Gordon-Holmes syndrome (cerebellar atrophy and hypogonadotropic hypogonadism). Historically, Oliver-McFarlane Syndrome and Laurence-Moon Syndrome had been considered distinct clinical phenotypes although reviewing the clinical features reported for these disorders highlights the marked similarities between Oliver-McFarlane Syndrome, Laurence-Moon syndrome and Boucher-Neuhauser Syndrome. When Boucher-Neuhauser Syndrome was first defined as a syndrome in 1989, similarities to Oliver-McFarlane Syndrome (reported in 1965) and Laurence-Moon Syndrome (reported in 1866) were not considered. Oliver-McFarlane Syndrome was first described in a 2yr old with strikingly long eyelashes, growth retardation and pigmentary retinopathy. It is perhaps less surprising that a phenotypic overlap with the previously reported Laurence-Moon Syndrome and Gordon-Holmes syndromes was not considered. The main phenotypic difference between Laurence-Moon Syndrome and Oliver-McFarlane Syndrome arguably is the presence of trichomegaly although interestingly one of the individuals from the Laurence-Moon Syndrome family assessed in our study (F:3) was noted, unlike her siblings, to have long eyelashes and bushy eyebrows.

In this study investigation of an individual with Oliver-McFarlane Syndrome identified compound heterozygous mutations in PNPLA6, which together with similar results in 5 other affected individuals from 4 families with Oliver-McFarlane Syndrome and 4 affected individuals from a family with Laurence-Moon Syndrome, confirmed the

Deleted: and

Deleted: have

Deleted: i

Deleted: has

Deleted: has

hypothesis that not only are these disorders related but are better considered to reflect the spectrum of complex phenotypes associated with PNPLA6 mutations.

In 4/5 of the families reported here, affected individuals had one missense mutation in trans with a LOF mutation (frameshift insertion, splice mutation or in-frame duplication which putatively results in downstream truncation of the peptide sequence). No patients have been reported with two loss of function alleles, which is consistent with the finding that biallelic PNPLA6 loss in a knock-out mouse model is embryologically lethal (Moser *et al.*, 2004). 4/5 of the missense mutations identified were clustered in the patatin-like esterase domain (EST) and 1/5 in the cyclic nucleotide monophosphate (cNMP) binding domains. An analysis of all mutations reported in PNPLA6 across the range of known phenotypes confirms clustering of missense mutations in the EST domain (Nanetti *et al.*, 2022).

Interestingly patients B:1 and E:1 both had identical genotypes (p.Arg1031fs*38; p.Gly1129Arg) . The p.Arg1031fs*38 mutation had previously been reported in 2 individuals with spastic ataxia and GH syndrome (Synofzik *et al.*, 2014) however a recent analysis of all reported PNPLA6 mutations shows that this is now known to be the commonest mutation reported across all PNPLA6 phenotypes (14/72 index cases ~20%) (Nanetti *et al.*, 2022). The p.Gly1129Arg mutation has subsequently been reported in 3 other individuals (in trans with a severe complex phenotypes including pigmentary retinopathy, endocrinopathy and neuropathy or ataxia (Kmoch *et al.*, 2015). A mutation at the same amino acid site (p.Gly1129Val) has been reported in a patient with Boucher-Neuhauser Syndrome (Zheng *et al.*, 2018).

Further evidence of the pathogenicity of the Oliver-McFarlane Syndrome/Laurence-Moon Syndrome associated missense mutations come from studies in a zebrafish morpholino PNPLA6 knockdown model in which mRNA encoding the 4 EST domain

Deleted: part of

Deleted:

Deleted:

Deleted: s

mutations (p.Arg1099Gln, p.Gly1129Arg, p.Gly1176Ser, p.Val1215Ala) failed to rescue the observed abnormal phenotype compared to the wt mRNA which *was* successful in restoring normal larvae development and morphology (Hufnagel *et al.*, 2015).

Formatted: Font: Italic

8.5.1 Clinical phenotypes associated with PNPLA6 mutations

As discussed, PNPLA6 mutations are associated with a range of overlapping phenotypes upper and lower motor neurone/SPG38, pure upper motor neurone/HSP, isolated pigmentary retinopathy, cerebellar ataxia +/- spasticity, cerebellar ataxia with hypogonadotropic hypogonadism (Gordon-Holmes syndrome) and Boucher-Neuhauser Syndrome, Oliver-McFarlane Syndrome and Laurence-Moon Syndrome comprising cerebellar ataxia, pigmentary retinopathy, hypogonadotropic hypogonadism, neuropathy and spasticity to varying degrees and, in the case of Oliver-McFarlane Syndrome, trichomegaly. Attempts at describing the genotype/phenotype correlation have to an extent been hampered by continued reference to syndromic classifications (i.e. Boucher-Neuhauser Syndrome, Laurence-Moon Syndrome, Oliver-McFarlane Syndrome) rather than clear delineation of relevant key phenotypes (motor neurone involvement, retinal, cerebellar, pituitary).

Deleted: (

Based on a review of all patients reported to date who harbour pathogenic PNPLA6 mutations (76 patient in 49 families), hypogonadotropic hypogonadism was the most common clinical phenotype (74% of cases) followed by pigmentary retinopathy (65% of cases) (Nanetti *et al.*, 2022). A pure motor phenotype was seen in 12% of cases and pure cerebellar ataxia in 8%. Isolated retinopathy was seen in 5%. A syndromic diagnosis was made in 73% of cases (Boucher-Neuhauser Syndrome – 26%, Oliver-McFarlane Syndrome – 21%, Gordon Holmes syndrome – 21%, Laurence-Moon Syndrome – 5%). The age at onset was variable however the majority (70%) presented at <8yrs, 27% presented at 9-18yrs and 6% presenting in adulthood.

The majority of known PNPLA6 mutations are located in the EST domain and the presence of one or more mutations in this domain is highly correlated with pigmentary retinopathy. Analysis of NTE activity in patient fibroblasts with SPG38 and Oliver-McFarlane Syndrome show significantly reduced enzyme activity in two Oliver-McFarlane Syndrome patients compared to that seen in the SPG38 fibroblasts which suggests that differential residual activity in NTE/PNPLA6 underlies some of the phenotypic variation seen with more severe complex phenotypes associated with more deleterious mutations in PNPLA6 (Hufnagel *et al.*, 2015).

A precise explanation of the phenotypic variation seen remains elusive although studies of PNPLA6 expression in human embryos demonstrate expression in the retina, anterior and posterior pituitary, cerebellum and ventricular zones of the developing brain (Hufnagel *et al.*, 2015). Disruption of phospholipid metabolism appears to be one of the potential mechanisms underlying the pathogenicity of impaired PNPLA6 function. In a Drosophila model, the loss of the PNPLA6 ortholog Swiss-Cheese protein (sws) was associated, in affected flies, with increased levels of lyophosphatidylcholine (LPC), which is thought to be the preferred biological substrate for PNPLA6 (Sunderhaus, Law and Kretzschmar, 2019). Impaired complex lipid synthesis is important in the pathophysiology of a range of inherited neurological diseases including several complex HSP syndromes, neurodegeneration with brain-iron accumulation (NBIA) and complex neuropathies (Garcia-Cazorla *et al.*, 2015).

Deleted: , in the affected flies

8.5.2 Conclusion

PNPLA6-related disorders manifest as a range of isolated and complex neurodegenerative phenotypes, the discovery and characterisation of which describes a journey which started over 150 years ago with Laurence and Moon's first case report, through Prohibition-era boot-legged Jamaica Ginger and finally to the present day of

modern genomic research. These disorders are still extremely rare but offer insights into some of the molecular mechanisms underlying a range of range of complex neurodegenerative disorders.

9 Conclusions

My research was focussed on investigating the genetic aetiology of several rare hereditary neurodegenerative conditions through the use of whole exome sequencing and other next generation sequencing technologies. At the time that I started my PhD, NGS had been established as a powerful research tool and was still a nascent technology. Through this work, I have been able to consider and appreciate the huge potential of NGS in understanding the genetic basis for disease but also the challenges and limitations.

I was able to use NGS to develop a panel for large scale screening of cerebellar ataxia genes (chapter 5), an approach that today is now firmly established in diagnostic genetic sequencing in ataxia and most genetic disorders. I have identified a novel gene in a large autosomal dominant dystonia family (chapter 3) and a (nearly) novel gene in a recessive spastic ataxia family (chapter 4. I have used NGS to identify the causal gene in autosomal recessive ataxia cases and used functional studies to confirm the pathogenicity of novel presumed pathogenic variants (Chapters 6 & 7).

Identification of TUBB4A as the cause of DYT4 dystonia (Chapter 3)

I was able to show, using a combination of exome sequencing and linkage analysis, that the gene causing DYT4, was TUBB4A which was the first time that this gene had been associated with a human Mendelian disorder. Subsequently, mutations in TUBB4A were identified to be the cause of a rare hypomyelinating leukodystrophy (H-ABC syndrome) and the phenotype of TUBB4A associated human disease was further expanded.

The family in whom I undertook the genetic analysis was the same as that first reported by Parker in 1981 and had previously been subject to the genetic techniques available in the 1980s and 1990s which however failed to identify the causal gene. This project was

Deleted: s

Deleted: but

Deleted: relative to today,

Deleted: that causes

Deleted: and elucidated.

one of the first I worked on after starting my PhD. Colleagues in our department had previously attempted linkage analysis on the large pedigree however this was unsuccessful due to errors in the reconstruction of the large multigenerational pedigree which impeded the linkage analysis as this relied on an accurate pedigree and accurate designation of affected individuals. My reappraisal of the pedigree construction and subsequent linkage analysis ultimately yielded a highly significant single linkage peak on chromosome 19 which enabled the exome sequencing data to be constrained to a very short list of candidate genes. The p.R2G TUBB4A was the most plausible candidate since the variant was located at the highly conserved and important MREI domain which had previously been shown to have a key role in the autoregulation of the gene. TUBB4A was highly expressed in the brain, in particular the cerebellum known to have a key role in dystonia pathogenesis. Subsequent functional studies by other groups showed clear morphological differences in cerebellar granule cells confirming the pathogenic effect of the R2G mutation on beta-tubulin structure and function.

Shortly after our paper was published, de novo mutations in TUBB4A were shown to underlie a rare hypomyelinating leukodystrophy (H-ABC syndrome) which was distinct from DYT4 in that it was a childhood onset neurodegenerative leukodystrophy. However, there were overlapping features and dystonia was a common clinical feature. Functional work has to an extent explained the divergent phenotypes with the H-ABC associated mutations showing significantly impaired microtubule polymerisation affecting both oligodendrocyte and neuronal cell populations whereas the R2G mutation resulted in abnormal neuronal morphology but normal oligodendrocytes and no observed change in microtubule polymerisation dynamics.

Deleted: limited

Deleted: has

Deleted: which

Deleted: has

Deleted: h

TUBB4A is now established as an important gene in dystonia pathogenesis further studies into the role of microtubule structure and function may yet yield further insights into the neurobiology of dystonia and other neurodegenerative diseases.

Identification of CAPN1 as a cause of recessive spastic ataxia (Chapter 4)

In an international collaboration with colleagues in the US and France, I identified, using a combination of homozygosity mapping and exome sequencing, a homozygous mutation in CAPN1 as the cause of an autosomal recessive spastic ataxia. The gene had already been shown to cause cerebellar ataxia in a canine (Parson's Russell Terrier) and our work confirmed that CAPN1 was associated with a neurodegenerative phenotype in humans. The splice mutation I identified in Family 1 was confirmed to have a significant impact on calpain 1 protein expression and calpain-1 enzyme activity in functional studies using cultured patient fibroblasts that I had collected.

Calpain-1 had previously been investigated by my US collaborators (Prof Baudry and Wang et al) as having an important role in long-term potentiation (LTP) in neuronal cells. Baudry's team had previously shown that calpain-1 activity was important in inactivating the PHLPP1/Akt pathway which has a role in neuroprotection. Prof Baudry and his team had previously investigated defects in LTP in a CAPN1 knockout mouse model but had not previously investigated whether there was a cerebellar phenotype in the mice. After sharing our genetic results, the CAPN1 mice were re-evaluated and indeed they were found to exhibit a cerebellar phenotype demonstrated by impaired roto-rod performance and a reduction in cerebellar granule cell density on histological analysis. The patient fibroblasts showed similar reductions in the knockout mice to activation of the PHLPP1/Akt pathway confirming that the knockout mouse was a good model for CAPN1-related neurodegeneration.

Deleted:

Deleted: a an

Deleted: He

Mutations in CAPN1 have now been established as a cause of cerebellar neurodegeneration and have been reported in over 80 patients exhibiting a broad range of additional phenotypes including spasticity and peripheral neuropathy. Further studies into patients with CAPN1 may yet expand the clinical phenotype and provide insights into the role of the calpain pathway in human neurodegenerative disease.

Targeted amplicon sequencing in a UK autosomal dominant ataxia cohort (Chapter 5)

The development of next generation DNA sequencing led to a dramatic reduction in the cost and speed of DNA sequencing. Diagnostic sequencing in cerebellar ataxia had at the time of my research been limited in practice to analysis of the most common genetic causes of cerebellar ataxia – the trinucleotide repeat expansions. Conventional mutations in a large number of other genes known to be associated with cerebellar ataxia were difficult to screen for due to the cost and time required for Sanger sequencing. Whilst exome sequencing had shown significant promise in both diagnostic and research settings, other applications of NGS such as targeted panel sequencing were being developed to expand the range of genes sequencing in clinical diagnostics.

I undertook a project to develop a NGS based sequencing approach to sequence, in a cohort of patients identified via the neurogenetics clinic at NHNN, all of the genes known at the time to be associated with autosomal dominant cerebellar ataxia. Using the Illumina Miseq platform, I designed a panel to sequence 15 genes associated with ADCA. Using this approach, I was able to identify novel, or possibly pathogenic heterozygous mutations were identified in 14/87 (15%) of cases.

3/87 patients had mutations in PRKCG (SCA14), and these patients were subsequently included in a published case series of SCA14 patients.

Deleted: has

Deleted: all of

Deleted: known

Deleted: in a cohort of patients identified via the neurogenetics clinic at NHNN

Deleted: novel

Deleted:)

Diagnostic panel sequencing is now established as a key tool in clinical genetic diagnostics. although there remain challenges however with the interpretation of the pathogenicity of rare and novel variants.

Identification of two families with cathepsin-d deficiency associated neuronal ceroid lipofuscinosis (Chapter 6)

I conducted exome sequencing in a group of patients with clinically recessive complex spinocerebellar ataxia. In one recessive pedigree I identified a homozygous missense mutation in cathepsin D (CTSD/CatD) consistent with a diagnosis of neuronal ceroid lipofuscinosis, specifically juvenile CLN10 disease. The phenotype of this family was of progressive ataxia, pigmentary retinopathy and cognitive problems. Through screening a group of complex ataxia cases by Sanger sequencing, I identified a further individual with another missense mutation in CTSD and a very similar phenotype.

I was able to demonstrate, through functional analysis of CatD activity in affected patient fibroblasts, that both mutations were associated with a significant reduction in CatD activity compared to control samples.

Neuropathological examination of muscle biopsies in affected individuals from both families revealed a distinctive pattern of pathology characterised by granular vacuolar material in angular atrophic fibres. Electron microscopy of patient fibroblasts also showed abnormal membrane-bound rimmed vacuoles.

Identification of SYNE1 associated cerebellar ataxia cases – genetic and clinical characterisation (Chapter 7)

SYNE1 mutations are a is a relatively common cause of autosomal recessive cerebellar ataxia. The SYNE1 gene however is one of the largest genes in the human genome and

Deleted: *Syne*

Deleted: -

Deleted: -

screening for pathogenic mutations using traditional Sanger sequencing approaches is prohibitive. In this study I employed a NGS sequencing approach using the Illumina MiSeq platform to screen a cohort of 84 autosomal recessive cerebellar ataxia patients for SYNE1 mutations. An additional 110 exomes sequenced in ARCA patients were also examined yielding a total of 3 patients with presumed pathogenic mutations.

In this study I examined an ethnically diverse UK cohort of recessive cerebellar ataxia patients, identified through a combination of targeted next generation sequencing and exome sequencing. Novel truncating mutations in SYNE1 were identified in a further four cases from three British, Turkish and Sri Lankan pedigrees.

A broad spectrum of clinical phenotypes is associated with mutations in PNPLA6 (Chapter 8)

In this section I undertook a genetic analysis of a patient with Oliver-McFarlane syndrome, a rare neurodegenerative disorder characterised by cerebellar ataxia, pigmentary retinopathy and hypogonadotropic hypogonadism. I identified compound heterozygous mutations in PNPLA6 using exome sequencing. PNPLA6 had already been associated with a recessive form of hereditary spastic paraplegia. The data from the affected case formed part of a series of patients with either Oliver-McFarlane Syndrome and a phenotypically similar condition Laurence-Moon syndrome and we determined that PNPLA6 mutations were the cause of both syndromes.

Deleted: or

With colleagues, I also identified a further PNPLA6 homozygous mutation in a patient with isolated cerebellar ataxia using exome sequencing and homozygosity mapping, confirming the broad phenotypic spectrum associated with PNPLA6 mutations.

9.1 Future Directions

The last decade has seen huge advances in the understanding of the causes of neurogenerative disease through the convergence and maturation of sequencing technology and a much greater understanding of the complexity of the human genome. Whilst the majority of the 'low-hanging fruit' of gene discovery may have seemingly been plucked - that is to say, the discovery of novel genes in large families with segregating highly penetrant mutations – a great deal of the heredity underlying neurogenerative disease remains to be discovered. Some of the most notable discoveries, such as the identification of complex polynucleotide repeats underlying C9orf72 associated frontotemporal dementia or RFC-1 associated cerebellar ataxia, have led to a reappraisal of where and how future gene discoveries may be found. Such complex repeat disorders have been challenging to identify using (now-traditional) techniques such as exome sequencing due to technical limitations however novel bioinformatic approaches in the analysis of whole genome data may facilitate the discovery of other complex repeat disorders in other neurodegenerative disease. This is important because whilst both aforementioned complex expansions are now likely to represent the commonest genetic cause of familial FTD/MND and in the case of RFC1 expansions, the most common cause of sporadic late-onset ataxia. It seems plausible therefore that other complex expansions, yet to be discovered, will be crucial in the understanding of a significant proportion of neurodegenerative disease. Technology such as long-read sequencing, which whilst expensive and not currently widely available, will facilitate this ongoing work.

That said, discoveries of conventional mutations in novel genes continues apace and challenges remain in the interpretation of sequence variation. Databases of normal

human polymorphisms obtained through large-scale genome sequencing projects will provide a key resource for the investigation and analysis of clinically relevant mutations.

.

10 References

Abreu, N.J. and Waldrop, M.A. (2021) 'Overview of gene therapy in spinal muscular atrophy and Duchenne muscular dystrophy', *Pediatric Pulmonology*, pp. 710–720. Available at: https://doi.org/10.1002/ppul.25055.

Adzhubei, I. a *et al.* (2010) 'A method and server for predicting damaging missense mutations.', *Nature methods*, 7(4), pp. 248–9. Available at: https://doi.org/10.1038/nmeth0410-248.

Albanese, A. *et al.* (2013) 'Phenomenology and classification of dystonia: A consensus update', *Movement Disorders*, 28(7), pp. 863–873. Available at: https://doi.org/10.1002/mds.25475.

Amirifar, P. *et al.* (2020) 'Ataxia-telangiectasia: epidemiology, pathogenesis, clinical phenotype, diagnosis, prognosis and management', *Expert Review of Clinical Immunology*. Taylor & Francis, pp. 859–871. Available at: https://doi.org/10.1080/1744666X.2020.1810570.

Andersen, B.B. *et al.* (1992) *A Quantitative Study of the Human Cerebellum With Unbiased Stereological Techniques*, *THE JOURNAL OF COMPARATIVE NEUROLOGY*.

Anderson, G.W., Goebel, H.H. and Simonati, A. (2013) 'Human pathology in NCL.', *Biochimica et biophysica acta*, 1832(11), pp. 1807–1826. Available at: https://doi.org/10.1016/j.bbadis.2012.11.014.

Aring, C.D. (1942) 'The systemic nervous affinity of triorthocresyl phosphate (jamaica ginger palsy)', *Brain*, 65(1), pp. 34–47. Available at: https://doi.org/10.1093/brain/65.1.34.

Atai, N. a *et al.* (2012) 'Untethering the nuclear envelope and cytoskeleton: biologically distinct dystonias arising from a common cellular dysfunction.', *International journal of cell biology*, 2012, p. 634214. Available at: https://doi.org/10.1155/2012/634214.

Awano, T. *et al.* (2006) 'A mutation in the cathepsin D gene (CTSD) in American Bulldogs with neuronal ceroid lipofuscinosis.', *Molecular genetics and metabolism*, 87(4), pp. 341–8. Available at: https://doi.org/10.1016/j.ymgme.2005.11.005.

Bagaria, J., Bagyinszky, E. and An, S.S.A. (2022) 'Genetics of Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS) and Role of Sacsin in Neurodegeneration', *International Journal of Molecular Sciences*, p. 552. Available at: https://doi.org/10.3390/ijms23010552.

Bahl, S. *et al.* (2005) 'Evidence of a common founder for SCA12 in the Indian population', *Annals of Human Genetics*, 69(5), pp. 528–534. Available at: https://doi.org/10.1046/j.1529-8817.2005.00173.x.

Bahlo, M. *et al.* (2018) 'Open Peer Review Recent advances in the detection of repeat expansions with short-read next-generation sequencing [version 1; referees: 3 approved]'. Available at: https://doi.org/10.12688/f1000research.13980.1.

Bakalkin, G. *et al.* (2010) 'Prodynorphin mutations cause the neurodegenerative disorder

spinocerebellar ataxia type 23.', *American journal of human genetics*, 87(5), pp. 593–603. Available at: https://doi.org/10.1016/j.ajhg.2010.10.001.

Balint, B. and Bhatia, K.P. (2015) 'Isolated and combined dystonia syndromes - an update on new genes and their phenotypes', *European Journal of Neurology*, 22(4), pp. 610–617. Available at: https://doi.org/10.1111/ene.12650.

Bally, J.F. *et al.* (2021) 'DYT-TUBB4A (DYT4 Dystonia): New clinical and genetic observations', *Neurology*, 95(14), pp. E1887–E1897. Available at: https://doi.org/10.1212/WNL.0000000000010882.

Bally, J.F. *et al.* (2022) 'DYT-TUBB4A (DYT4 Dystonia): Clinical Anthology of 11 Cases and Systematized Review', *Movement Disorders Clinical Practice* [Preprint]. Available at: https://doi.org/10.1002/mdc3.13452.

Bamshad, M.J. *et al.* (2011) 'Exome sequencing as a tool for Mendelian disease gene discovery', *Nature Reviews Genetics*, 12(11), pp. 745–755. Available at: https://doi.org/10.1038/nrg3031.

Bamshad, M.J., Nickerson, D.A. and Chong, J.X. (2019) 'Mendelian Gene Discovery: Fast and Furious with No End in Sight', *American Journal of Human Genetics*. Cell Press, pp. 448–455. Available at: https://doi.org/10.1016/j.ajhg.2019.07.011.

Barohn, R.J., Dowd, D.C. and Kagan-Hallet, K.S. (1992) 'Congenital ceroid-lipofuscinosis', *Pediatric Neurology*, 8(1), pp. 54–59. Available at: https://doi.org/10.1016/0887-8994(92)90054-3.

Bates, G. and Lehrach, H. (1994) 'Trinucleotide repeat expansions and human genetic disease', *BioEssays*, 16(4).

Bauer, P. *et al.* (2010) 'Spinocerebellar ataxia type 11 (SCA11) is an uncommon cause of dominant ataxia among French and German kindreds.', *Journal of neurology, neurosurgery, and psychiatry*, 81(11), pp. 1229–32. Available at: https://doi.org/10.1136/jnnp.2009.202150.

Di Bella, D. *et al.* (2010) 'Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28.', *Nature genetics*, 42(4), pp. 313–21. Available at: https://doi.org/10.1038/ng.544.

Blumkin, L. *et al.* (2014) 'Expansion of the spectrum of TUBB4A-related disorders: a new phenotype associated with a novel mutation in the TUBB4A gene.', *Neurogenetics* [Preprint]. Available at: https://doi.org/10.1007/s10048-014-0392-2.

Botstein, D. and Risch, N. (2003) 'Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease', *Nature Genetics*, 33(3S), pp. 228–237. Available at: https://doi.org/10.1038/ng1090.

Bressman, S.B. *et al.* (2009) 'Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study.', *Lancet neurology*, 8(5), pp. 441–6. Available at: https://doi.org/10.1016/S1474-4422(09)70081-X.

Browne, D. (1994) 'Episodic ataxia/myokymia syndrome is associated with point mutations in the human potassium channel gene, KCNA1', *Nature genetics*, 8.

Brusse, E. *et al.* (2006) 'Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene (SCA27): A new phenotype.', *Movement disorders : official journal of the Movement Disorder Society*, 21(3), pp. 396–401. Available at: https://doi.org/10.1002/mds.20708.

Budd Haeberlein, S. *et al.* (2022) 'Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease', *Journal of Prevention of Alzheimer's Disease*, 9(2), pp. 197–210. Available at: https://doi.org/10.14283/jpad.2022.30.

Bunk, J. *et al.* (2021) 'Cathepsin D Variants Associated With Neurodegenerative Diseases Show Dysregulated Functionality and Modified α-Synuclein Degradation Properties', *Frontiers in Cell and Developmental Biology*, 9. Available at: https://doi.org/10.3389/fcell.2021.581805.

Cagnoli, C. *et al.* (2010) 'Missense mutations in the AFG3L2 proteolytic domain account for ∼1.5% of European autosomal dominant cerebellar ataxias.', *Human mutation*, 31(10), pp. 1117–24. Available at: https://doi.org/10.1002/humu.21342.

Carpenter S, Karpati G, A.F. (1972) 'Specific involvement of muscle, nerve, and skin in late infantile and juvenile amaurotic idiocy', *Neurology*, 22, pp. 170–86.

Casjens, S.R. and Hendrix, R.W. (2015) 'Bacteriophage lambda: Early pioneer and still relevant', *Virology*. Academic Press Inc., pp. 310–330. Available at: https://doi.org/10.1016/j.virol.2015.02.010.

Cecilia Marelli, MD; Joyce van de Leemput, PhD; Janel O. Johnson, MS; Francois Tison, MD; Christel Thauvin-Robinet, MD, PhD; Fabienne Picard, MD; Christine Tranchant, MD; Dena G. Hernandez, MS; Bernard Huttin, MD; Jacques Boulliat, MD; Iban Sangla, MD; Ch, M. (2011) 'SCA15 Due to Large ITPR1 Deletions in a Cohort of 333 White Families With Dominant Ataxia', *Archives of Neurology*, 68(5), pp. 637–643.

Ceuterick, C. and Martin, J. (1992) 'Electron microscopic features of skin in neurometabolic disorders', *Journal of the Neurological Sciences*, 112, pp. 15–29.

Charlesworth, G. *et al.* (2012) 'Mutations in ANO3 cause dominant craniocervical dystonia: Ion channel implicated in pathogenesis', *American Journal of Human Genetics*, 91(6), pp. 1041–1050. Available at: https://doi.org/10.1016/J.AJHG.2012.10.024.

Charzewska, A. *et al.* (2016) 'Hypomyelinating leukodystrophies — a molecular insight into the white matter pathology', *Clinical Genetics*. Blackwell Publishing Ltd, pp. 293–304. Available at: https://doi.org/10.1111/cge.12811.

Chelban, V. *et al.* (2018) 'Genotype-phenotype correlations, dystonia and disease progression in spinocerebellar ataxia type 14', *Movement Disorders*, 00(00), pp. 1–11. Available at: https://doi.org/10.1002/mds.27334.

Claussnitzer, M. *et al.* (2020) 'A brief history of human disease genetics', *Nature*, pp. 179–189. Available at: https://doi.org/10.1038/s41586-019-1879-7.

Cleveland, D., Pittenger, M. and Feramisco, J. (1983) 'Elevation of Tubulin levels by microinjection suppresses new tubulin synthesis', *Nature*, 305.

Cleveland, D.W. (1988) 'Reviews Autoregulated instability of tubulin mRNAs ':, *Cell*,

(September), pp. 339–343.

Corral-Juan, M. *et al.* (2018) 'Clinical, genetic and neuropathological characterization of spinocerebellar ataxia type 37', *Brain*, 141(7), pp. 1981–1997. Available at: https://doi.org/10.1093/brain/awy137.

Cortese, A. (2020) 'CANVAS with cerebellar/sensory/vestibular dysfunction from RFC1 intronic pentanucleotide expansion', *Brain*, pp. 386–390. Available at: https://doi.org/10.1093/brain/awaa015.

Curiel, J. *et al.* (2017) 'TUBB4A mutations result in specific neuronal and oligodendrocytic defects that closely match clinically distinct phenotypes', *Human Molecular Genetics*, 26(22), pp. 4506–4518. Available at: https://doi.org/10.1093/hmg/ddx338.

Darwin, C. (1859) *On the origin of species: By means of natural selection, or the preservation of favoured races in the struggle for life*, *On the Origin of Species: By Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life*. Br Foreign Med Chir Rev. Available at: https://doi.org/10.4324/9781912281244.

Delorme, C. *et al.* (2021) 'Whispering dysphonia in TUBB4A-related disorders responsive to bipallidal deep brain stimulation', *European Journal of Neurology*, 28(3), pp. 1082–1083. Available at: https://doi.org/10.1111/ene.14602.

Depienne, C. and Mandel, J.-L. (2021) *30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? | Elsevier Enhanced Reader*, *American journal of human genetics*. Available at: https://reader.elsevier.com/reader/sd/pii/S0002929721000951?token=E0BB2C3F5696FF73E2B2B0C0AC899DFAD144EE1F95DFACF6901BF726E3ADC5E90FA23761AFC2E3A3AA3C743CCF470BB7&originRegion=eu-west-1&originCreation=20211201133146 (Accessed: 19 January 2022).

Dueñas, A.M., Goold, R. and Giunti, P. (2006) 'Molecular pathogenesis of spinocerebellar ataxias.', *Brain : a journal of neurology*, 129(Pt 6), pp. 1357–70. Available at: https://doi.org/10.1093/brain/awl081.

Duncan, I.D. *et al.* (2017) 'A mutation in the Tubb4a gene leads to microtubule accumulation with hypomyelination and demyelination', *Annals of Neurology*, 81(5), pp. 690–702. Available at: https://doi.org/10.1002/ana.24930.

Ellard, S. *et al.* (2020) 'ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020 Recommendations ratified by ACGS Quality Subcommittee on 4 th'. Available at: https://doi.org/10.1101/531210.

Engert, J.C. *et al.* (2000) 'ARSACS, a spastic ataxia common in northeastern Québec, is caused by mutations in a new gene encoding an 11.5-kb ORF.', *Nature genetics*, 24(2), pp. 120–5. Available at: https://doi.org/10.1038/72769.

Erro, R. *et al.* (2015) 'H-ABC syndrome and DYT4: Variable expressivity or pleiotropy of TUBB4 mutations?', *Movement Disorders*, 30(6), pp. 828–833. Available at: https://doi.org/10.1002/mds.26129.

Escayg, a *et al.* (2000) 'Coding and noncoding variation of the human calcium-channel

beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia.', *American journal of human genetics*, 66(5), pp. 1531–9. Available at: https://doi.org/10.1086/302909.

Fa, M. *et al.* (2015) 'Novel selective calpain 1 inhibitors as potential therapeutics in alzheimer's disease', *Journal of Alzheimer's Disease*, 49(3), pp. 707–721. Available at: https://doi.org/10.3233/JAD-150618.

Fan, H. and Chu, J.Y. (2007) 'A Brief Review of Short Tandem Repeat Mutation', *Genomics, Proteomics and Bioinformatics*. Elsevier, pp. 7–14. Available at: https://doi.org/10.1016/S1672-0229(07)60009-6.

Feigin, V.L. *et al.* (2020) 'The global burden of neurological disorders: translating evidence into policy', *The Lancet Neurology*. Lancet Publishing Group, pp. 255–265. Available at: https://doi.org/10.1016/S1474-4422(19)30411-9.

Figueroa, K.P. *et al.* (2010) 'KCNC3: phenotype, mutations, channel biophysics-a study of 260 familial ataxia patients.', *Human mutation*, 31(2), pp. 191–6. Available at: https://doi.org/10.1002/humu.21165.

Figueroa, K.P. *et al.* (2011) 'Frequency of KCNC3 DNA variants as causes of spinocerebellar ataxia 13 (SCA13).', *PloS one*, 6(3), p. e17811. Available at: https://doi.org/10.1371/journal.pone.0017811.

Fineberg, N.A. *et al.* (2013) 'The size, burden and cost of disorders of the brain in the UK', *Journal of Psychopharmacology*, 27(9), pp. 761–770. Available at: https://doi.org/10.1177/0269881113495118.

Flanagan, S.E., Patch, A.M. and Ellard, S. (2010) 'Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations', *Genetic Testing and Molecular Biomarkers*, 14(4), pp. 533–537. Available at: https://doi.org/10.1089/gtmb.2010.0036.

Fogel, B.L. and Perlman, S. (2007) 'Clinical features and molecular genetics of autosomal recessive cerebellar ataxias.', *Lancet neurology*, 6(3), pp. 245–57. Available at: https://doi.org/10.1016/S1474-4422(07)70054-6.

Follo, C. *et al.* (2013) 'Knock-down of Cathepsin D in zebrafish fertilized eggs determines congenital myopathy.', *Bioscience reports*, pp. 371–378. Available at: https://doi.org/10.1042/BSR20120100.

Forman, O.P., De Risio, L. and Mellersh, C.S. (2013) 'Missense Mutation in CAPN1 Is Associated with Spinocerebellar Ataxia in the Parson Russell Terrier Dog Breed', *PLoS ONE*, 8(5), p. 64627. Available at: https://doi.org/10.1371/journal.pone.0064627.

Fritchie, K. *et al.* (2009) 'Novel mutation and the first prenatal screening of cathepsin D deficiency (CLN10).', *Acta neuropathologica*, 117(2), pp. 201–8. Available at: https://doi.org/10.1007/s00401-008-0426-7.

Frozza, R.L., Lourenco, M. V and de Felice, F.G. (2018) 'Challenges for Alzheimer's disease therapy: Insights from novel mechanisms beyond memory defects', *Frontiers in Neuroscience*, p. 37. Available at: https://doi.org/10.3389/fnins.2018.00037.

Fuchs, T. *et al.* (2009) 'Mutations in the THAP1 gene are responsible for DYT6 primary

torsion dystonia.', *Nature genetics*, 41(3), pp. 286–8. Available at: https://doi.org/10.1038/ng.304.

Fuchs, T. *et al.* (2012) 'Mutations in GNAL cause primary torsion dystonia', *Nature Genetics* [Preprint]. Available at: https://doi.org/10.1038/ng.2496.

van Gaalen, J., Giunti, P. and Van de Warrenburg, B.P. (2011) 'Movement disorders in spinocerebellar ataxias', *Movement Disorders*. John Wiley & Sons, Ltd, pp. 792–800. Available at: https://doi.org/10.1002/mds.23584.

Gan-Or, Z. *et al.* (2016) 'Mutations in CAPN1 Cause Autosomal-Recessive Hereditary Spastic Paraplegia', *American Journal of Human Genetics*, 98(5), pp. 1038–1046. Available at: https://doi.org/10.1016/j.ajhg.2016.04.002.

Garcia-Cazorla, À. *et al.* (2015) 'The clinical spectrum of inherited diseases involved in the synthesis and remodeling of complex lipids. A tentative overview', *Journal of Inherited Metabolic Disease*, 38(1), pp. 19–40. Available at: https://doi.org/10.1007/s10545-014-9776-6.

García-Murias, M. *et al.* (2012) '"Costa da Morte" ataxia is spinocerebellar ataxia 36: Clinical and genetic characterization', *Brain*, 135(5), pp. 1423–1435. Available at: https://doi.org/10.1093/brain/aws069.

Garden, G.A. and La Spada, A.R. (2008) 'Molecular pathogenesis and cellular pathology of spinocerebellar ataxia type 7 neurodegeneration', *Cerebellum*, 7(2), pp. 138–149. Available at: https://doi.org/10.1007/s12311-008-0027-y.

Gavarini, S. *et al.* (2010) 'Direct Interaction between Causative Genes of DYT1 and DYT6 Primary Dystonia', *ANN NEUROL*, 68, pp. 549–553. Available at: https://doi.org/10.1002/ana.22138.

Gilissen, C. *et al.* (2012) 'Disease gene identification strategies for exome sequencing', *European Journal of Human Genetics*, 20(5), pp. 490–497. Available at: https://doi.org/10.1038/ejhg.2011.258.

Giudicessi, J.R. *et al.* (2011) 'Transient outward current (I to) gain-of-function mutations in the KCND3-encoded Kv4.3 potassium channel and Brugada syndrome', *Heart Rhythm*, 8(7), pp. 1024–1032. Available at: https://doi.org/10.1016/J.HRTHM.2011.02.021.

Goll, D.E. *et al.* (2003) 'The calpain system', *Physiological Reviews*. American Physiological Society, pp. 731–801. Available at: https://doi.org/10.1152/physrev.00029.2002.

Greenfield, J.G. (1954) *A Monograph in American Lectures in Neurology. Price, 17s. 6d. Pp. 112*, *Blackwell Scientific Publications*. The Ryerson Press. Available at: https://jamanetwork.com/.

Gros-Louis, F. *et al.* (2007) 'Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia', *Nature Genetics*, 39(1), pp. 80–85. Available at: https://doi.org/10.1038/ng1927.

Guergueltcheva, V. *et al.* (2012) 'Autosomal-recessive congenital cerebellar ataxia is caused by mutations in metabotropic glutamate receptor 1', *American Journal of Human*

Genetics, 91(3), pp. 553–564. Available at: https://doi.org/10.1016/j.ajhg.2012.07.019.

Gusella, J. et al (1983) 'A polymorphic DNA marker genetically linked to Huntington's disease', *Nature*, 306. Available at: https://doi.org/10.1080/07345410.1982.11677818.

De Gusmão, C.M. *et al.* (2016) 'Dystonia-Causing Mutations as a Contribution to the Etiology of Spasmodic Dysphonia', in *Otolaryngology - Head and Neck Surgery (United States)*, pp. 624–628. Available at: https://doi.org/10.1177/0194599816648293.

Hamilton, E.M. *et al.* (2014) 'Hypomyelination with atrophy of the basal ganglia and cerebellum: Further delineation of the phenotype and genotype-phenotype correlation', *Brain*, 137(7), pp. 1921–1930. Available at: https://doi.org/10.1093/brain/awu110.

Harding, A.E. (1981a) 'Classification of the heriditary ataxias and paraplegias', *The Lancet*, pp. 1151–1155.

Harding, A.E. (1981b) 'Friedreich's ataxia: A clinical and genetic study of 90 families with an analysis of early diagnostic criteria and intrafamilial clustering of clinical features', *Brain*, 104(3), pp. 589–620. Available at: https://doi.org/10.1093/brain/104.3.589.

Hardy, J. and Selkoe, D.J. (2002) 'The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics', *Science*, pp. 353–356. Available at: https://doi.org/10.1126/science.1072994.

Hausmann, M. *et al.* (2004) 'Cathepsin D is up-regulated in inflammatory bowel disease macrophages.', *Clinical and experimental immunology*, 136(1), pp. 157–67. Available at: https://doi.org/10.1111/j.1365-2249.2004.02420.x.

Hegreness, M. and Meselson, M. (2007) 'What did Sutton see?: Thirty years of confusion over the chromosomal basis of mendelism', *Genetics*, pp. 1939–1944. Available at: https://doi.org/10.1534/genetics.104.79723.

Hersheson, J. *et al.* (2012) 'Mutations in the autoregulatory domain of β-tubulin 4a cause hereditary dystonia.', *Annals of neurology* [Preprint]. Available at: https://doi.org/10.1002/ana.23832.

Hirano, T. (2018) 'Purkinje Neurons: Development, Morphology, and Function', *Cerebellum*, 17(6), pp. 699–700. Available at: https://doi.org/10.1007/s12311-018-0985-7.

Holmes, S.E. *et al.* (1999) 'Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12', *Nature Genetics*, 23(4), pp. 391–392. Available at: https://doi.org/10.1038/70493.

Houlden, H. *et al.* (2007) 'Mutations in TTBK2, encoding a kinase implicated in tau phosphorylation, segregate with spinocerebellar ataxia type 11.', *Nature genetics*, 39(12), pp. 1434–6. Available at: https://doi.org/10.1038/ng.2007.43.

Hufnagel, R.B. *et al.* (2015) 'Neuropathy target esterase impairments cause Oliver-McFarlane and Laurence-Moon syndromes', *Journal of Medical Genetics*, 52(2), pp. 85–94. Available at: https://doi.org/10.1136/jmedgenet-2014-102856.

Ikeda, Y. *et al.* (2006) 'Spectrin mutations cause spinocerebellar ataxia type 5.', *Nature*

genetics, 38(2), pp. 184–90. Available at: https://doi.org/10.1038/ng1728.

Ingram, V.M. (1956) 'A SPECIFIC CHEMICAL DIFFERENCE BETWEEN THE GLOBINS OF NORMAL HUMAN AND SICKLE-CELL ANJEMIA HAEMOGLOBIN', *Nature*, 178(4537).

Jarman, P.R. *et al.* (1999) 'Primary torsion dystonia: the search for genes is not over.', *Journal of neurology, neurosurgery, and psychiatry*, 67(3), pp. 395–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1736535&tool=pmcentrez&rendertype=abstract.

Jinnah, H.A., Neychev, V. and Hess, E.J. (2017) 'The Anatomical Basis for Dystonia: The Motor Network Model', *Tremor and other hyperkinetic movements (New York, N.Y.)*, 7, p. 506. Available at: https://doi.org/10.7916/D8V69X3S.

Johnson, M.K. (1969) *The Delayed Neurotoxic Effect of some Organophosphorus Compounds IDENTIFICATION OF THE PHOSPHORYLATION SITE AS AN ESTERASE*, *Biochem. J*. Available at: http://portlandpress.com/biochemj/article-pdf/114/4/711/764904/bj1140711.pdf.

Kang, S. and Hong, S. (2009) 'Molecular pathogenesis of spinocerebellar ataxia type 1 disease', *Molecules and Cells*, pp. 621–627. Available at: https://doi.org/10.1007/s10059-009-0095-y.

Klebe, S. *et al.* (2005) 'New mutations in protein kinase Cgamma associated with spinocerebellar ataxia type 14.', *Annals of neurology*, 58(5), pp. 720–9. Available at: https://doi.org/10.1002/ana.20628.

Klein, C. and Fahn, S. (2013) 'Translation of Oppenheim's 1911 paper on dystonia', *Movement Disorders*, pp. 851–862. Available at: https://doi.org/10.1002/mds.25546.

Klockgether, T. (2011) 'Update on degenerative ataxias.', *Current opinion in neurology*, 24(4), pp. 339–45. Available at: https://doi.org/10.1097/WCO.0b013e32834875ba.

Kmoch, S. *et al.* (2015) 'Mutations in PNPLA6 are linked to photoreceptor degeneration and various forms of childhood blindness', *Nature Communications*, 6. Available at: https://doi.org/10.1038/ncomms6614.

Van Der Knaap, M.S. *et al.* (2007) 'Hypomyelination with atrophy of the basal ganglia and cerebellum: Follow-up and pathology', *Neurology*, 69(2), pp. 166–171. Available at: https://doi.org/10.1212/01.wnl.0000265592.74483.a6.

Knight, M. a *et al.* (2004) 'Dominantly inherited ataxia and dysphonia with dentate calcification: spinocerebellar ataxia type 20.', *Brain : a journal of neurology*, 127(Pt 5), pp. 1172–81. Available at: https://doi.org/10.1093/brain/awh139.

Kobayashi, H. *et al.* (2011) 'Expansion of Intronic GGCCTG Hexanucleotide Repeat in NOP56 Causes SCA36, a Type of Spinocerebellar Ataxia Accompanied by Motor Neuron Involvement.', *American journal of human genetics*, 89(1), pp. 121–130. Available at: https://doi.org/10.1016/j.ajhg.2011.05.015.

Koboldt, D.C. (2020) 'Best practices for variant calling in clinical sequencing', *Genome Medicine*. Available at: https://doi.org/10.1186/s13073-020-00791-w.

Kousi, M., Lehesjoki, A.-E. and Mole, S.E. (2012) 'Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses.', *Human mutation*, 33(1), pp. 42–63. Available at: https://doi.org/10.1002/humu.21624.

Krajka, V. *et al.* (2022) 'H-ABC– and dystonia-causing TUBB4A mutations show distinct pathogenic effects', *Science Advances*, 8(10), p. 9229. Available at: https://doi.org/10.1126/sciadv.abj9229.

Kumar, P., Henikoff, S. and Ng, P.C. (2009) 'Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.', *Nature protocols*, 4(7), pp. 1073–81. Available at: https://doi.org/10.1038/nprot.2009.86.

Kurbatskaya, K. *et al.* (2016) 'Upregulation of calpain activity precedes tau phosphorylation and loss of synaptic proteins in Alzheimer's disease brain', *Acta neuropathologica communications*, 4, p. 34. Available at: https://doi.org/10.1186/s40478-016-0299-2.

Lander, E.S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. Available at: https://doi.org/10.1038/35057062.

Lander, E.S. and Botstein, D. (1987) 'Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children', *Science*, 236(4808), pp. 1567–1570. Available at: https://doi.org/10.1126/science.2884728.

Laurence, J.Z. and Moon, R.C. (1995) 'Four cases of "retinitis pigmentosa" occurring in the same family, and accompanied by general imperfections of development. 1866.', *Obesity research*, 3(4), pp. 400–403. Available at: https://doi.org/10.1002/j.1550-8528.1995.tb00166.x.

Lee, Y.C. *et al.* (2012) 'Mutations in KCND3 cause spinocerebellar ataxia type 22', *Annals of Neurology*, 72(6), pp. 859–869. Available at: https://doi.org/10.1002/ana.23701.

van de Leemput, J. *et al.* (2007) 'Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans.', *PLoS genetics*, 3(6), p. e108. Available at: https://doi.org/10.1371/journal.pgen.0030108.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools.', *Bioinformatics (Oxford, England)*, 25(16), pp. 2078–9. Available at: https://doi.org/10.1093/bioinformatics/btp352.

Lohmann, K. *et al.* (2013) 'Whispering dysphonia (DYT4 dystonia) is caused by a mutation in the TUBB4 gene', *Annals of Neurology*, 73(4), pp. 537–545. Available at: https://doi.org/10.1002/ana.23829.

Lopata, M. and Cleveland, D. (1987) 'In Vivi microtubules are coplymers of available beta-tubulin isotypes: Localisation of each of six vertebrate beta-tubulin isotypes using polyclonal antibodies elicited by synthetic peptide antigens', *Journal of Cell Biology*, 105(October), pp. 1707–1720.

Lopez, S. and He, F. (2022) 'Spinocerebellar Ataxia 36: From Mutations Toward Therapies', *Frontiers in Genetics*, 13(March), pp. 1–7. Available at: https://doi.org/10.3389/fgene.2022.837690.

MacDonald, M.E. *et al.* (1993) 'A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes', *Cell*, 72(6), pp. 971–983. Available at: https://doi.org/10.1016/0092-8674(93)90585-E.

Magaña, J.J., Velázquez-Pérez, L. and Cisneros, B. (2013) 'Spinocerebellar ataxia type 2: Clinical presentation, molecular mechanisms, and therapeutic perspectives', *Molecular Neurobiology*, pp. 90–104. Available at: https://doi.org/10.1007/s12035-012-8348-8.

Mandelkow, E. and Mandelkow, E.-M. (1994) 'Microtubule structure', *Current Opinion in Structural Biology*, 4(2), pp. 171–179. Available at: https://doi.org/10.1016/S0959-440X(94)90305-0.

Mantuano, E. *et al.* (2003) 'Spinocerebellar ataxia type 6 and episodic ataxia type 2: Differences and similarities between two allelic disorders', *Cytogenetic and Genome Research*, pp. 147–153. Available at: https://doi.org/10.1159/000072849.

Marras, C. *et al.* (2016) 'Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force', *Movement Disorders*. John Wiley and Sons Inc., pp. 436–457. Available at: https://doi.org/10.1002/mds.26527.

Masson, O. *et al.* (2010) 'Pathophysiological functions of cathepsin D: Targeting its catalytic activity versus its protein binding activity?', *Biochimie*, 92(11), pp. 1635–43. Available at: https://doi.org/10.1016/j.biochi.2010.05.009.

Matsuura, T. *et al.* (2000) 'Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10.', *Nature genetics*, 26(2), pp. 191–4. Available at: https://doi.org/10.1038/79911.

McLoughlin, H.S., Moore, L.R. and Paulson, H.L. (2020) 'Pathogenesis of SCA3 and implications for other polyglutamine diseases', *Neurobiology of Disease*. Academic Press Inc. Available at: https://doi.org/10.1016/j.nbd.2019.104635.

Melentev, P.A., Agranovich, O.E. and Sarantseva, S. V. (2020) 'Human diseases associated with NTE gene', *Ecological Genetics*, 18(2), pp. 229–242. Available at: https://doi.org/10.17816/ecogen16327.

Mendel, G. (1866) 'EXPERIMENTS IN PLANT HYBRIDIZATION (1865)', *Verhandlungen des naturforschenden Ver-eines in Brünn*, 4, pp. 3–47. Available at: http://www.netspace.org./MendelWeb/ (Accessed: 21 July 2022).

Méreaux, J.L. *et al.* (2021) 'Increasing involvement of CAPN1 variants in spastic ataxias and phenotype-genotype correlations', *Neurogenetics*, 22(1), pp. 71–79. Available at: https://doi.org/10.1007/s10048-020-00633-2.

Mole, S.E. (no date) *No Title*. Available at: http://www.ucl.ac.uk/ncl/CLN10CTSDmutationtable.htm.

Moser, M. *et al.* (2004) 'Placental Failure and Impaired Vasculogenesis Result in Embryonic Lethality for Neuropathy Target Esterase-Deficient Mice', *Molecular and Cellular Biology*, 24(4), pp. 1667–1679. Available at: https://doi.org/10.1128/mcb.24.4.1667-1679.2004.

Moskowitz, P.F. and Oblinger, M.M. (1995) 'Transcriptional and post-transcriptional mechanisms regulating neurofilament and tubulin gene expression during normal development of the rat brain.', *Brain research. Molecular brain research*, 30(2), pp. 211–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7637572.

Nalls, M.A. *et al.* (2019) 'Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies', *The Lancet Neurology*, 18(12), pp. 1091–1102. Available at: https://doi.org/10.1016/S1474-4422(19)30320-5.

Nanetti, L. *et al.* (2022) 'Multifaceted and Age-Dependent Phenotypes Associated With Biallelic PNPLA6 Gene Variants: Eight Novel Cases and Review of the Literature', *Frontiers in Neurology*, 12, p. 793547. Available at: https://doi.org/10.3389/fneur.2021.793547.

Nery, F.C. *et al.* (2008) 'TorsinA binds the KASH domain of nesprins and participates in linkage between nuclear envelope and cytoskeleton.', *Journal of cell science*, 121(Pt 20), pp. 3476–86. Available at: https://doi.org/10.1242/jcs.029454.

Ng, P.C. and Henikoff, S. (2001) 'Predicting deleterious amino acid substitutions', *Genome Research*, 11(5), pp. 863–874. Available at: https://doi.org/10.1101/gr.176601.

Ng, S.B. *et al.* (2010) 'Exome sequencing identifies the cause of a mendelian disorder.', *Nature genetics*, 42(1), pp. 30–5. Available at: https://doi.org/10.1038/ng.499.

Nibbeling, E.A.R. *et al.* (2017) 'Using the shared genetics of dystonia and ataxia to unravel their pathogenesis', *Neuroscience and Biobehavioral Reviews*, 75, pp. 22–39. Available at: https://doi.org/10.1016/j.neubiorev.2017.01.033.

Nurk, S. *et al.* (2022) 'The complete sequence of a human genome', *Science*, 376(6588), pp. 44–53. Available at: https://doi.org/10.1126/science.abj6987.

O'Leary, N.A. *et al.* (2016) 'Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation', *Nucleic Acids Research*, 44(D1), pp. D733–D745. Available at: https://doi.org/10.1093/nar/gkv1189.

Okoniewska, M., Tanaka, T. and Yada, R.Y. (1999) 'The role of the flap residue, threonine 77, in the activation and catalytic activity of pepsin A.', *Protein engineering*, 12(1), pp. 55–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10065711.

Olgiati, S., Quadri, M. and Bonifati, V. (2016) 'Genetics of movement disorders in the next-generation sequencing era', *Movement Disorders*. John Wiley and Sons Inc., pp. 458–470. Available at: https://doi.org/10.1002/mds.26521.

Oliver, G.L. and Mcfarlane, D.C. (1965) 'Congenital Trichomegaly: With Associated Pigmentary Degeneration of the Retina, Dwarfism, and Mental Retardation', *Archives of Ophthalmology*, 74(2), pp. 169–171. Available at: https://doi.org/10.1001/archopht.1965.00970040171008.

Opal, P. *et al.* (2002) 'Intrafamilial phenotypic variability of the DYT1 dystonia: From asymptomatic TOR1A gene carrier status to dystonic storm', *Movement Disorders*, pp. 339–345. Available at: https://doi.org/10.1002/mds.10096.

Ophoff, R. a *et al.* (1996) 'Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4.', *Cell*, 87(3), pp. 543–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8898206.

Orr, H.T. *et al.* (1993) 'Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1', *Nature Genetics*, 4(3), pp. 221–226. Available at: https://doi.org/10.1038/ng0793-221.

Ott, J., Wang, J. and Leal, S.M. (2015) 'Genetic linkage analysis in the age of whole-genome sequencing', *Nature Reviews Genetics*, pp. 275–284. Available at: https://doi.org/10.1038/nrg3908.

Ozelius, L. *et al.* (1997) 'The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein', *Nature genetics*, 17, pp. 40–48.

Palau, F. and Espinós, C. (2006) 'Autosomal recessive cerebellar ataxias.', *Orphanet journal of rare diseases*, 1, p. 47. Available at: https://doi.org/10.1186/1750-1172-1-47.

Parker, N. (1985) 'Hereditary whispering dysphonia.', *Journal of neurology, neurosurgery, and psychiatry*, 48(3), pp. 218–24. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1028253&tool=pmcentrez&rendertype=abstract.

Parkinson, M.H. *et al.* (2018) 'Optical coherence tomography in autosomal recessive spastic ataxia of charlevoix-Saguenay', *Brain*, 141(4), pp. 989–999. Available at: https://doi.org/10.1093/brain/awy028.

Patton, M.A., Harding, A.E. and Baraitser, M. (1986) 'Congenital trichomegaly, pigmentary retinal degeneration, and short stature', *American Journal of Ophthalmology*, pp. 490–491. Available at: https://doi.org/10.1016/0002-9394(86)90656-2.

Perez-Siles, G. *et al.* (2022) 'A Compound Heterozygous Mutation in Calpain 1 Identifies a New Genetic Cause for Spinal Muscular Atrophy Type 4 (SMA4)', *Frontiers in Genetics*, 12, p. 2800. Available at: https://doi.org/10.3389/fgene.2021.801253.

Pericak-Vance, M.A. (2001) 'Analysis of genetic linkage data for Mendelian traits.', *Current protocols in human genetics*, Chapter 1(1), pp. 1.4.1-1.4.31. Available at: https://doi.org/10.1002/0471142905.hg0104s09.

Pizzino, A. *et al.* (2014) 'TUBB4A de novo mutations cause isolated hypomyelination', *Neurology*, 83(10), pp. 898–902. Available at: https://doi.org/10.1212/WNL.0000000000000754.

Pollard, K.S. *et al.* (2010) 'Detection of nonneutral substitution rates on mammalian phylogenies', *Genome Research*, 20(1), pp. 110–121. Available at: https://doi.org/10.1101/gr.097857.109.

Polymeropoulos, M.H. *et al.* (1997) 'Mutation in the α-synuclein gene identified in families with Parkinson's disease', *Science*, 276(5321), pp. 2045–2047. Available at: https://doi.org/10.1126/science.276.5321.2045.

Posey, J.E. (2019) 'Genome sequencing and implications for rare disorders', *Orphanet*

Journal of Rare Diseases. Available at: https://doi.org/10.1186/s13023-019-1127-0.

Pruitt, K.D. *et al.* (2009) 'The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes', *Genome Research*, 19(7), pp. 1316–1323. Available at: https://doi.org/10.1101/gr.080531.108.

Pulst, S.M. (1999) 'Genetic linkage analysis', *Archives of Neurology*, 56(6), pp. 667–672. Available at: https://doi.org/10.1001/archneur.56.6.667.

Putzel, G.G. *et al.* (2016) 'GNAL mutation in isolated laryngeal dystonia', *Movement Disorders*, 31(5), pp. 750–755. Available at: https://doi.org/10.1002/mds.26502.

Rai, S.N. *et al.* (2019) 'The Role of PI3K/Akt and ERK in Neurodegenerative Disorders', *Neurotoxicity Research*, pp. 775–795. Available at: https://doi.org/10.1007/s12640-019-0003-y.

Rainier, S. *et al.* (2008) 'Neuropathy Target Esterase Gene Mutations Cause Motor Neuron Disease', *American Journal of Human Genetics*, 82(3), pp. 780–785. Available at: https://doi.org/10.1016/j.ajhg.2007.12.018.

Rapola, J. and Haltia, M. (1973) 'CYTOPLASMIC INCLUSIONS IN THE VERMIFORM APPENDIX AND SKELETAL MUSCLE IN TWO TYPES OF SO-CALLED RECENTLY , an infantile type of so-called neuronal ceroid-lipofuscinosis has been described as an entity (Santavuori et ah , 1973 ; Haltia et ah , 1973). The', *Brain*, (96), pp. 833–840.

Renton, A.E. *et al.* (2011) 'A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD', *Neuron*, pp. 1–12. Available at: https://doi.org/10.1016/j.neuron.2011.09.010.

Roostaei, T. *et al.* (2014) 'The human cerebellum: A review of physiologic neuroanatomy', *Neurologic Clinics*. Elsevier, pp. 859–869. Available at: https://doi.org/10.1016/j.ncl.2014.07.013.

Ruano, L. *et al.* (2014) 'The global epidemiology of hereditary ataxia and spastic paraplegia: A systematic review of prevalence studies', *Neuroepidemiology*. S. Karger AG, pp. 174–183. Available at: https://doi.org/10.1159/000358801.

Rubio-Agusti, I. *et al.* (2013) 'Tremulous cervical dystonia is likely to be familial: Clinical characteristics of a large cohort', *Parkinsonism and Related Disorders*, 19(6), pp. 634–638. Available at: https://doi.org/10.1016/j.parkreldis.2013.02.017.

Sakai, H. *et al.* (2010) 'Analysis of an insertion mutation in a cohort of 94 patients with spinocerebellar ataxia type 31 from Nagano, Japan', *Neurogenetics*, 11(4), pp. 409–415. Available at: https://doi.org/10.1007/s10048-010-0245-6.

Sato, N. *et al.* (2009) 'Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n.', *American journal of human genetics*, 85(5), pp. 544–57. Available at: https://doi.org/10.1016/j.ajhg.2009.09.019.

Schachat, Andrew P. and Maumenee, I.H. (1982) 'Bardet-Biedl Syndrome and Related Disorders', *Archives of Ophthalmology*, 100(2), pp. 285–288. Available at: https://doi.org/10.1001/archopht.1982.01030030287011.

Schachat, Andrew P and Maumenee, I.H. (1982) 'Bardet-Biedl Syndrome and Related Disorders', *Archives of Ophthalmology*, 100(2), pp. 285–288. Available at: https://doi.org/10.1001/archopht.1982.01030030287011.

Schicks, J. *et al.* (2011) 'Mutations in the PDYN gene (SCA23) are not a frequent cause of dominant ataxia in Central Europe', *Clinical Genetics*, 80(5), pp. 503–504. Available at: https://doi.org/10.1111/j.1399-0004.2011.01651.x.

Schirinzi, T. *et al.* (2018) 'Dystonia as a network disorder: A concept in evolution', *Current Opinion in Neurology*, 31(4), pp. 498–503. Available at: https://doi.org/10.1097/WCO.0000000000000580.

Schöls, L. *et al.* (2004) 'Review Autosomal dominant cerebellar ataxias : clinical features , genetics , and pathogenesis', *The Lancet*, 3(May), pp. 291–304.

Schulz, A. *et al.* (2018) 'Study of Intraventricular Cerliponase Alfa for CLN2 Disease', *New England Journal of Medicine*, 378(20), pp. 1898–1907. Available at: https://doi.org/10.1056/nejmoa1712649.

Schuur, M. *et al.* (2011) 'Cathepsin D gene and the risk of Alzheimer's disease: a population-based study and meta-analysis.', *Neurobiology of aging*, 32(9), pp. 1607–14. Available at: https://doi.org/10.1016/j.neurobiolaging.2009.10.011.

Schwarz, J.M. *et al.* (2010) 'MutationTaster evaluates disease-causing potential of sequence alterations.', *Nature methods*, 7(8), pp. 575–6. Available at: https://doi.org/10.1038/nmeth0810-575.

Seelow, D. *et al.* (2009) 'HomozygosityMapper - An interactive approach to homozygosity mapping', *Nucleic Acids Research*, 37(SUPPL. 2), pp. 593–599. Available at: https://doi.org/10.1093/nar/gkp369.

Seixas, A.I. *et al.* (2017) 'A Pentanucleotide ATTTC Repeat Insertion in the Non-coding Region of DAB1, Mapping to SCA37, Causes Spinocerebellar Ataxia', *American Journal of Human Genetics*, 101(1), pp. 87–103. Available at: https://doi.org/10.1016/j.ajhg.2017.06.007.

Sielecki, a R. *et al.* (1990) 'Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 A resolution.', *Journal of molecular biology*, 214(1), pp. 143–70. Available at: https://doi.org/10.1016/0022-2836(90)90153-D.

Simons, C. *et al.* (2013) 'A de novo mutation in the β-tubulin gene TUBB4A results in the leukoencephalopathy hypomyelination with atrophy of the basal ganglia and cerebellum', *American Journal of Human Genetics*, 92(5), pp. 767–773. Available at: https://doi.org/10.1016/j.ajhg.2013.03.018.

Smedley, D. *et al.* (2021) '100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report', *New England Journal of Medicine*, 385(20), pp. 1868–1880. Available at: https://doi.org/10.1056/nejmoa2035790.

Smith, K.R. *et al.* (2011) 'Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes', *Genome Biology*, 12(9), p. R85. Available at: https://doi.org/10.1186/gb-2011-12-9-r85.

Smith, P. *et al.* (1985) 'Measurement of Protein Using Bicinchoninic Acid', *Analytical Biochemistry*, 85, pp. 76–85.

Stamelou, M. *et al.* (2014) 'The Phenotypic Spectrum of DYT24 Due to ANO3 Mutations', *Movement Disorders*, 29(7). Available at: https://doi.org/10.1002/mds.25802.

Steeves, T.D. *et al.* (2012) 'The prevalence of primary dystonia: A systematic review and meta-analysis', *Movement Disorders*, 27(14), pp. 1789–1796. Available at: https://doi.org/10.1002/mds.25244.

Steinfeld, R. *et al.* (2006) 'Cathepsin D deficiency is associated with a human neurodegenerative disorder.', *American journal of human genetics*, 78(6), pp. 988–98. Available at: https://doi.org/10.1086/504159.

Steward, C.A. *et al.* (2017) 'Genome annotation for clinical genomic diagnostics: Strengths and weaknesses', *Genome Medicine*. Available at: https://doi.org/10.1186/s13073-017-0441-1.

Stratton, M.R., Campbell, P.J. and Futreal, P.A. (2009) 'The cancer genome', *Nature*. Nature Publishing Group, pp. 719–724. Available at: https://doi.org/10.1038/nature07943.

Sullivan, K.F. and Cleveland, D.W. (1986) 'Identification of conserved isotype-defining variable region sequences for four vertebrate beta tubulin polypeptide classes.', *Proceedings of the National Academy of Sciences of the United States of America*, 83(12), pp. 4327–31. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=323725&tool=pmcentrez&rendertype=abstract.

Sunderhaus, E.R., Law, A.D. and Kretzschmar, D. (2019) 'Disease-Associated PNPLA6 Mutations Maintain Partial Functions When Analyzed in Drosophila', *Frontiers in Neuroscience*, 13. Available at: https://doi.org/10.3389/fnins.2019.01207.

Swinnen, B., Robberecht, W. and Van Den Bosch, L. (2020) 'RNA toxicity in non-coding repeat expansion disorders', *The EMBO Journal*, 39(1). Available at: https://doi.org/10.15252/embj.2018101112.

Synofzik, M. *et al.* (2014) 'PNPLA6 mutations cause Boucher-Neuhäuser and Gordon Holmes syndromes as part of a broad neurodegenerative spectrum', *Brain*, 137(1), pp. 69–77. Available at: https://doi.org/10.1093/brain/awt326.

Synofzik, M. *et al.* (2016) 'SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: A large multi-centre study', *Brain*, 139(5), pp. 1378–1393. Available at: https://doi.org/10.1093/brain/aww079.

Thiffault, I. *et al.* (2013) 'Diversity of ARSACS mutations in french-canadians', *Canadian Journal of Neurological Sciences*, 40(1), pp. 61–66. Available at: https://doi.org/10.1017/S0317167100012968.

THOMPSON, E.A. (1975) 'The estimation of pairwise relationships', *Annals of Human Genetics*, 39(2), pp. 173–188. Available at: https://doi.org/10.1111/j.1469-1809.1975.tb00120.x.

Tischfield, M. a *et al.* (2010) 'Human TUBB3 mutations perturb microtubule dynamics, kinesin interactions, and axon guidance.', *Cell*, 140(1), pp. 74–87. Available at: https://doi.org/10.1016/j.cell.2009.12.011.

Tischfield, M.A. *et al.* (2011) 'Phenotypic spectrum of the tubulin-related disorders and functional implications of disease-causing mutations.', *Current opinion in genetics & development*, 21(3), pp. 286–94. Available at: https://doi.org/10.1016/j.gde.2011.01.003.Phenotypic.

Tomiyasu H, Takahashi W, Ohta T, Yoshii F, Shibuya M, S.Y. (2000) 'No Title', *Rinsho Shinkeigaku*, 40(4), pp. 350–7.

Traschütz, A. *et al.* (2021) 'Natural History, Phenotypic Spectrum, and Discriminative Features of Multisystemic RFC1 Disease', *Neurology*, 96(9), pp. e1369–e1382. Available at: https://doi.org/10.1212/WNL.0000000000011528.

Untergasser, A. *et al.* (2012) 'Primer3--new capabilities and interfaces.', *Nucleic acids research*, 40(15), p. e115. Available at: https://doi.org/10.1093/nar/gks596.

Valante, E.M. *et al.* (1998) 'The role of DYT1 in primary torsion dystonia in Europe', *Brain*, 121(12), pp. 2335–2339. Available at: https://doi.org/10.1093/brain/121.12.2335.

Vankan, P. (2013) 'Prevalence gradients of Friedreich's Ataxia and R1b haplotype in Europe co-localize, suggesting a common Palaeolithic origin in the Franco-Cantabrian ice age refuge', *Journal of Neurochemistry*. Blackwell Publishing Ltd, pp. 11–20. Available at: https://doi.org/10.1111/jnc.12215.

Vemula, S.R. *et al.* (2013) 'Role of Gα(olf) in familial and sporadic adult-onset primary dystonia', *Human Molecular Genetics*, 22(12), pp. 2510–2519. Available at: https://doi.org/10.1093/hmg/ddt102.

de Vries, B. *et al.* (2009) 'Episodic ataxia associated with EAAT1 mutation C186S affecting glutamate reuptake.', *Archives of neurology*, 66(1), pp. 97–101. Available at: https://doi.org/10.1001/archneurol.2008.535.

Vulinovic, F. *et al.* (2017) 'Screening study of TUBB4A in isolated dystonia', *Parkinsonism and Related Disorders*, 41, pp. 118–120. Available at: https://doi.org/10.1016/j.parkreldis.2017.06.001.

Vulinovic, F. *et al.* (2018) 'Motor protein binding and mitochondrial transport are altered by pathogenic TUBB4A variants', *Human Mutation*, 39(12), pp. 1901–1915. Available at: https://doi.org/10.1002/humu.23602.

Wade, R.H. (2009) 'On and around microtubules: An overview', *Molecular Biotechnology*, pp. 177–191. Available at: https://doi.org/10.1007/s12033-009-9193-5.

Wakamiya, M. *et al.* (2006) 'The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10', *Neurology*, 67(4), pp. 607–613. Available at: https://doi.org/10.1212/01.wnl.0000231140.26253.eb.

Wang, J.L. *et al.* (2010) 'TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing.', *Brain : a journal of neurology*, 133(Pt 12), pp. 3510–8. Available at: https://doi.org/10.1093/brain/awq323.

Wang, K., Li, M. and Hakonarson, H. (2010) 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.', *Nucleic acids research*, 38(16), p. e164. Available at: https://doi.org/10.1093/nar/gkq603.

Wang, Y. *et al.* (2013) 'Distinct roles for μ-calpain and m-calpain in synaptic NMDAR-mediated neuroprotection and extrasynaptic NMDAR-mediated neurodegeneration', *Journal of Neuroscience*, 33(48), pp. 18880–18892. Available at: https://doi.org/10.1523/JNEUROSCI.3293-13.2013.

Wang, Y. *et al.* (2016) 'Defects in the CAPN1 Gene Result in Alterations in Cerebellar Development and Cerebellar Ataxia in Mice and Humans', *Cell Reports*, 16(1), pp. 79–91. Available at: https://doi.org/10.1016/j.celrep.2016.05.044.

Wang, Y. *et al.* (2018) 'Protection against TBI-Induced Neuronal Death with Post-Treatment with a Selective Calpain-2 Inhibitor in Mice', *Journal of Neurotrauma*, 35(1), pp. 105–117. Available at: https://doi.org/10.1089/neu.2017.5024.

Wang, Y. *et al.* (2020) 'Calpain-1 and Calpain-2 in the Brain: New Evidence for a Critical Role of Calpain-2 in Neuronal Death', *Cells*. NLM (Medline). Available at: https://doi.org/10.3390/cells9122698.

Warrier, V., Vieira, M. and Mole, S.E. (2013) 'Genetic basis and phenotypic correlations of the neuronal ceroid lipofusinoses.', *Biochimica et biophysica acta*, 1832(11), pp. 1827–30. Available at: https://doi.org/10.1016/j.bbadis.2013.03.017.

Washington Neuromuscular Website (no date). Available at: http://neuromuscular.wustl.edu/ataxia/recatax.html (Accessed: 7 February 2012).

Watanabe, N. *et al.* (2018) 'Dystonia-4 (DYT4)-associated TUBB4A mutants exhibit disorganized microtubule networks and inhibit neuronal process growth', *Biochemical and Biophysical Research Communications*, 495(1), pp. 346–352. Available at: https://doi.org/10.1016/j.bbrc.2017.11.038.

Wetterstrand KA (2021) *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program*. Available at: www.genome.gov/sequencingcostsdata (Accessed: 30 July 2022).

White, M. *et al.* (2012) 'Transgenic mice with SCA10 pentanucleotide repeats show motor phenotype and susceptibility to seizure: A toxic RNA gain-of-function model', *Journal of Neuroscience Research*, 90(3), pp. 706–714. Available at: https://doi.org/10.1002/jnr.22786.

Wiethoff, S. *et al.* (2016) 'Heterogeneity in clinical features and disease severity in ataxia-associated SYNE1 mutations', *Journal of Neurology*, 263(8), pp. 1503–1510. Available at: https://doi.org/10.1007/s00415-016-8148-6.

Wiethoff, S. *et al.* (2017) 'Pure Cerebellar Ataxia with Homozygous Mutations in the PNPLA6 Gene', *Cerebellum*, 16(1), pp. 262–267. Available at: https://doi.org/10.1007/s12311-016-0769-x.

Wilcox, R. a *et al.* (2011) 'Whispering dysphonia in an Australian family (DYT4): a clinical and genetic reappraisal.', *Movement disorders*, 26(13), pp. 2404–8. Available at: https://doi.org/10.1002/mds.23866.

Williams, R.E. and Mole, S.E. (2012) 'New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses.', *Neurology*, 79(2), pp. 183–91. Available at: https://doi.org/10.1212/WNL.0b013e31825f0547.

Yabe, I. *et al.* (2003) 'Spinocerebellar ataxia type 14 caused by a mutation in protein kinase C gamma.', *Archives of neurology*, 60(12), pp. 1749–51. Available at: https://doi.org/10.1001/archneur.60.12.1749.

Ye, J. *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.', *BMC bioinformatics*, 13, p. 134. Available at: https://doi.org/10.1186/1471-2105-13-134.

Yen, T., Machlin, P. and Cleveland, D. (1988) 'Autoregulated instability of beta-tubulin mRNAs by recognition of the nascent amino terminus of beta-tubulin', *Nature*, 334(18), pp. 580–585.

Yen, T.J. *et al.* (1988) 'Autoregulated changes in stability of polyribosome-bound beta-tubulin mRNAs are specified by the first 13 translated nucleotides.', *Molecular and cellular biology*, 8(3), pp. 1224–35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=363267&tool=pmcentrez&rendertype=abstract.

Zech, M. *et al.* (2015) 'Large-scale TUBB4A mutational screening in isolated dystonia and controls', *Parkinsonism and Related Disorders*, 21(10), pp. 1278–1281. Available at: https://doi.org/10.1016/j.parkreldis.2015.08.017.

Zheng, R. *et al.* (2018) 'A novel PNPLA6 compound heterozygous mutation identified in a Chinese patient with Boucher-Neuhäuser syndrome', *Molecular Medicine Reports*, 18(1), pp. 261–267. Available at: https://doi.org/10.3892/mmr.2018.8955.

Zhu, G. *et al.* (2017) 'Calpain-1 deletion impairs mGluR-dependent LTD and fear memory extinction', *Scientific Reports*, 7. Available at: https://doi.org/10.1038/srep42788.

Zhuchenko, O. *et al.* (1997) 'Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the α(1A)-voltage-dependent calcium channel', *Nature Genetics*, pp. 62–69. Available at: https://doi.org/10.1038/ng0197-62.

Zühlke, C. *et al.* (2007) 'Screening of the SPTBN2 (SCA5) gene in German SCA patients.', *Journal of neurology*, 254(12), pp. 1649–52. Available at: https://doi.org/10.1007/s00415-007-0600-1.

Research Declarations

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
a) Where was the work published? (e.g. journal name)	Human Mutation	
b) Who published the work? (e.g. Elsevier/Oxford University Press):	Wiley	
c) When was the work published?	11/09/2012	
d) Was the work subject to academic peer review?	Yes	
e) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
a) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
b) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
c) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
d) Stage of publication	Please select.	
3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)**:**		
I drafted the paper with additional review from the other co-authors.		

4. In which chapter(s) of your thesis can this material be found?			
Chapter 1.5			
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):			
Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

6. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
f) Where was the work published? (e.g. journal name)	Annals of Neurology	
g) Who published the work? (e.g. Elsevier/Oxford University Press):	Wiley	
h) When was the work published?	13/12/2012	
i) Was the work subject to academic peer review?	Yes	
j) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
7. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
e) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
f) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
g) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
h) Stage of publication	Please select.	
8. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		
I drafted the paper with additional review from the other co-authors. **Section 3.2** The kindred were originally investigated in 1992 by Anita Harding who supervised the extraction of DNA from affected members of the DYT4 kindred and stored in the Institute of Neurology sample library. I performed a clinical re-evaluation based on available records and videos of affected patients. I performed the linkage analysis based on genotype data obtained by Dr Katherine Fawcett. This		

was confirmed and repeated by Dr Robert Kleta. I performed the analysis of exome sequencing data and subsequent expression study of TUBB4a cDNA in healthy controls. Analysis of brain expression data was based on data obtained by Prof Mina Ryten and Dr Trabzuni.

9. In which chapter(s) of your thesis can this material be found?

Chapter 3

10. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

11. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
k) Where was the work published? (e.g. journal name)	Cell Reports	
l) Who published the work? (e.g. Elsevier/Oxford University Press):	Cell Press	
m) When was the work published?	28/06/2016	
n) Was the work subject to academic peer review?	Yes	
o) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
12. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
i) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
j) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
k) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
l) Stage of publication	Please select.	
13. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		
The article was drafted by the main author (Yubin Wang) with the relevant sections to my data added by me and Prof Houlden. The index family (Family 1) was under the care of the neurogenetics service at NHNN. I undertook a clinical revaluation of the index case, exome sequencing and homozygosity mapping based on genotype data obtained by Dr Katherine Fawcett. I collected fibroblasts from the index		

patient in order to undertake cDNA sequencing. I performed the protein sequence modelling of the mutant protein sequence. My initial findings of a probable causative homozygous mutation in the CAPN1 gene led to an international collaboration with Prof Michel Baudry and colleagues at Western University of Health Sciences who had been assessing learning and memory in a CAPN1 knockout mouse model. His team performed CAPN1 expression studies and activity assays in the patient fibroblasts and control samples.
Additional cases (Families 2-4) were identified from the Department of Molecular Neurobiology, National Institute of Neurology, Tunisia and sequenced at the National Institute of Health, Bethesda, USA (Monia Hammer, Faycal Hentati and Rim Amouri) and also by colleagues at Centre de Reference de Neurogenetique, Hopital de la Pitie-Salpetriere, Paris. (Alexis Brice, Giovanni Stevanin)

14. In which chapter(s) of your thesis can this material be found?

Chapter 4

15. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

16. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
p) Where was the work published? (e.g. journal name)	Movement Disorders	
q) Who published the work? (e.g. Elsevier/Oxford University Press):	Wiley	
r) When was the work published?	01/07/2018	
s) Was the work subject to academic peer review?	Yes	
t) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
17. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
m) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
n) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
o) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
p) Stage of publication	Please select.	
18. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		
The article was drafted by the main author (Viorica Chelban) including the data that I had obtained and presented in the thesis. The patients identified with PRKCG/SCA14 mutations were published as part of a larger case series of SCA14 patients (Chelban *et al.,* 2018). The clinical reappraisal of one of the SCA14 cases		

was undertaken by Dr Chelban (section 5.4.3). I undertook the genetic sequencing analysis of the SCA14 case.
19. In which chapter(s) of your thesis can this material be found?
Chapter 5
20. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

21. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
u) Where was the work published? (e.g. journal name)	Neurology	
v) Who published the work? (e.g. Elsevier/Oxford University Press):	Wolters Kluwer	
w) When was the work published?	11/11/2014	
x) Was the work subject to academic peer review?	Yes	
y) Have you retained the copyright for the work?	No	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
22. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
q) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
r) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
s) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
t) Stage of publication	Please select.	
23. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		
The article was drafted by the me with drafting of the muscle biopsy results drafted by Dr Jacques. I undertook a clinical evaluation of the affected subjects in Families 1 and 2 and performed the skin biopsies for fibroblast culture. Fibroblast culture was undertaken by the cell culture laboratory team at ION. Exome sequencing was performed by Dr Deborah Hughes and		

subsequent analysis of homozygosity mapping and exome sequencing were performed by me. I performed all cathepsin D activity assays with the assistance of Derek Burke in the Institute for Child Health. Histological analysis and electron microscopy of the muscle biopsy samples was undertaken by Dr Thomas Jacques in the Histopathology department at Great Ormond St Hospital for Children.
24. In which chapter(s) of your thesis can this material be found?
Chapter 6
25. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

26. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
z) Where was the work published? (e.g. journal name)	Journal of Neurology	
aa) Who published the work? (e.g. Elsevier/Oxford University Press):	Springer	
bb) When was the work published?	13/05/2016	
cc) Was the work subject to academic peer review?	Yes	
dd) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
27. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
u) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
v) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
w) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
x) Stage of publication	Please select.	
28. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		

I performed the MiSeq design and sequencing of the *SYNE1* gene to identify the 3 patients reported in this study. Genetic analysis was jointly performed with Dr Sarah Wiethoff who performed the Sanger sequencing. Clinical evaluation was undertaken by me and Prof Houlden.

The results from this study were published in a paper co-authored by myself and Dr Wiethoff (Wiethoff *et al.*, 2016)

Deleted: myself

29. In which chapter(s) of your thesis can this material be found?

Chapter 7

30. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

31. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
ee) Where was the work published? (e.g. journal name)	Journal of Medical Genetics	
ff) Who published the work? (e.g. Elsevier/Oxford University Press):	BMJ	
gg) When was the work published?	05/12/2014	
hh) Was the work subject to academic peer review?	Yes	
ii) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
32. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
y) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
z) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
aa) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
bb) Stage of publication	Please select.	
33. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		
The article was drafted by the main author (Robert Hufnagel) with additional data on the case presented in the thesis drafted by me.		

Oliver-Mcfarlane/Laurence-Moon syndromes study: I undertook a clinical re-evaluation of the index case in family A, performed exome sequencing analysis and mutation confirmation. I facilitated the re-evaluation of the stored muscle biopsy for the patient. The muscle histology analysis was performed by Prof Tamas Refesz. The case was published as part of a larger series of Oliver-McFarlane and Laurence Moon syndrome patients by Robert Hufnagel (Hufnagel *et al.*, 2015). Homozygosity mapping and exome sequencing and Sanger confirmation for the other reported cases were undertaken by collaborators.

34. In which chapter(s) of your thesis can this material be found?

Chapter 8

35. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

36. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
jj) Where was the work published? (e.g. journal name)	Cerebellum	
kk) Who published the work? (e.g. Elsevier/Oxford University Press):	Springer	
ll) When was the work published?	19/03/2016	
mm) Was the work subject to academic peer review?	Yes	
nn) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
37. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
cc) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
dd) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
ee) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
ff) Stage of publication	Please select.	
38. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**		

The article was drafted by the main author (Sarah Wiethoff) with additional data on the case presented in the thesis drafted by me. Cerebellar ataxia case: I performed the exome sequencing analysis jointly with Sarah Wiethoff. Homozygosity mapping was performed by Reema Paudel.
39. In which chapter(s) of your thesis can this material be found?
Chapter 8.4.3/8.4.4
40. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Joshua Hersheson	**Date:**	22/09/2022
Supervisor/ Senior Author (where appropriate)**:**	Henry Houlden	**Date:**	22/09/2022