March 2014

The Use of Next Generation Sequencing Technologies to Dissect the Aetiologies of Parkinson's disease and Dystonia

---

Una-Marie Sheerin, BSc, MRCP

Institute of Neurology
University College London

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

Declaration:

I, Una-Marie Sheerin, 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

Whole exome sequencing (WES) – the targeted sequencing of the subset of the human genome that is protein coding – is a powerful and cost-effective new tool for dissecting the genetic basis of diseases and traits, some of which have proved to be intractable to conventional gene-discovery strategies.

My PhD thesis focuses on the use of whole exome sequencing to dissect the genetic aetiologies of families with Mendelian forms of Parkinson's disease and Dystonia. First I present a project where next generation sequencing played an important role in the identification of a novel Parkinson's disease gene (*VPS35*). I then describe the use of WES in i) an autosomal dominant PD kindred, where a novel *DCTN1* mutation was identified; and show a number of examples of successes and failures of WES in ii) autosomal recessive Parkinson's disease and iii) autosomal recessive generalised dystonia.

ACKNOWLEDGEMENTS

I am extremely grateful to my supervisors Professor Nicholas Wood and Professor John Hardy for giving me the opportunity to work in the genetics lab at such an exciting time in genetics, for their excellent mentorship, support, advice, encouragement and most importantly their boundless enthusiasm. Thanks to Professor Wood for his clinical teaching in neurogenetics and movement disorders, which I found invaluable and identifying many interesting families to work on. I would like to particularly thank Dr. Vincent Plagnol for being approachable and his tireless help with the bioinformatics aspect of this work. I am indebted to Professor Kailash Bhatia, for providing me with so many opportunities in my PhD by putting me in touch with families, for his clinical teaching from which I have learnt so much and for his optimism and support. I would also like to thank Robert Kleta, Horia Stanescu and Mehmet Tekman who helped with linkage analysis.

Thanks also to the members of the Neurogenetics laboratory for help and support, in particular Gavin Charlesworth for his help, advice and laughter. I would also like to thank the following individuals: Alan Pittman, Arianna Tucci, Hallgeir Jonvik, Deborah Hughes, Boniface Mok, Mina Ryten, Helene Plun-Favreau, Emma Deas, Sonia Gandhi, Andrey Abramov, Nicole Gurunlian, Kerra Pearce, David Nicholl, Mark Gaskin, Daniah Trabzuni, Paola Giunti, Josh Hersheson, Niccolo Mencacci, Mike Parkinson, Jenny Mcgowan, Lee Stanyer and June Smalley. A special thanks to those working in the diagnostic laboratory for their advice and help: Mary Sweeney, Ese Mudanohwo, Jason Heir, James Polke, Vaneesha Gibbons, Robyn Labrum, Mohammad Ullah, Liz Redmond, and Nana Boateng.

I am also grateful to those outside of work who gave me help, support and encouragement including Emma Baple, Sheeba Irshad, Aadil Khan, Shenaz Nasim, Ann-Marie Eze, Mitun Majumdar, Chantal Misquitta and Nino Foti.

Most heartfelt thanks to my family Thomas, Mary, Fiona, Danielle and little Eleanor for their unwavering support in everything I do, particularly to my parents, who strive to give me opportunities they did not have. And finally, to Ippokratis, for his unwavering support and encouragement in everything I do, through the good and the bad times, with practical advice, laughter, infectious enthusiasm and love.

TABLE OF CONTENTS

ABBREVIATIONS

1000G	Thousand Genomes
ACMSD	aminocarboxymuconate semialdehyde decarboxylase
AD	Alzheimer's Disease
ALS	Amyotrophic Lateral Sclerosis
ALS	Amyotrophic lateral sclerosis
ANO3	Anoctamin 3
ATP13A2	ATPase 13A2
ATP1A3	ATPase, Na+/K+ Transporting, Alpha 3 Polypeptide
bp	base pair
BST1	bone marrow stromal cell antigen 1
bvFTD	behavioral variant frontotemporal dementia
C9orf72	chromosome 9 open reading frame 72
CAP-Gly	N-terminal cytoskeleton-associated protein, glycine-rich
CBD	Corticobasal degeneration
CCDC62	coiled-coil domain containing 62
CCDS	Consensus Coding Sequence Project
CIZ1	CDKN1A Interacting Zinc Finger Protein
cM	Centimorgan
dbSNP	The Single Nucleotide Polymorphism Database
DCTN1	Dynactin 1
DJ-1	*Oncogene DJ1*
DNA	Deoxyribonucleic acid
DNAJC6	Auxilin
dNTPs	Deoxynucleotide Triphosphates
DRPLA	Dentatorubro Pallido-Luysian Atrophy
DYT1	*TOR1A*
EDTA	Ethylenediaminetetraacetic acid
EIF4G	Eukaryotic translation initiation factor 4 gamma 1
EOPD	Early onset Parkinson's Disease
ER	endoplasmic reticulum
EVS	Exome variant server
FBX07	ATPase 13A2

FGF20,	fibroblast growth factor 20
FPD	Familial Parkinson's Disease
GA-1	Glutaric Aciduria Type 1
GAK	cyclin G associated kinase
GBA	glucosidase, beta, acid gene
GCH1	GTP cyclohydrolase I
GEF	guanine nucleotide exchange factor
GERP	Genomic Evolutionary Rate Profiling
GNAL	Guanine Nucleotide Binding Protein (G Protein), Alpha Activating Activity Polypeptide, Olfactory Type
GPNMB	glycoprotein (transmembrane) nmb
GWAS	Genome-wide association studies
HIP1R	huntingtin interacting protein 1 related
HLA-DRB5	major histocompatibility complex, class II, DR beta 5
HTRA2	HtrA serine peptidase 2
IBD	Identical by descent
IBS	identical by state
IPDGC	International Parkinson's Disease Genomics Consortium
KRS	Kufor-Rakeb syndrome
L-dopa	Levodopa
LAMP3	lysosomal-associated membrane protein 3
LOD	logarithm of the odds
LRRK2	leucine-rich repeat kinase 2
MAF	Minor Allele Frequency
MAPP	Multivariate Analysis of Protein Polymorphism
MAPT	microtubule-associated protein tau
Mb	Megabase
MCCC1	methylcrotonoyl-CoA carboxylase 1 (alpha)
MMSE	Mini Mental State Examination
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MR-1	Myofibrillogenesis Regulator 1
MREI	methionine-arginine-glutamic acid-isoleucine
MSA	Multiple System Atrophy
MSD	multiple sulfatase deficiency

mV	millivolt
NCL	neuronal ceroid lipofuscinosis
NGS	Next Generation Sequencing
NHLBI	National Heart, Lung, and Blood Institute
NHNN	National hospital for Neurology and Neurosurgery
nm	nanometre
PRKN	*Parkin*
PCR	Polymerase chain reaction
PD	Parkinson's disease
PED	paroxysmal exercise-induced dyskinesia
PET	positron emission tomography
PINK1	*Phosphatase and tensin homolog (PTEN)-induced putative kinase 1*
PLA2G6	Phospholipase A2, group VI (cytosolic, calcium-independent)
pmol	picomole
PNKD	paroxysmal non-kinesigenic dyskinesia
PRRT2	Proline-Rich Transmembrane Protein 2
PSP	progressive supranuclear palsy
RNA	Ribonucleic acid
rpm	Revolutions per minute
RXN	Reactions
SGCE	Sarcoglycan, Epsilon
siRNA	Small interfering RNA
SLC2A1	Solute Carrier Family 2 (Facilitated Glucose Transporter), Member 1
SNCA	alpha synuclein
SNV	Single Nucleotide Variant
ssDNA	single-stranded DNA
STBD1	starch binding domain 1
STX1B	syntaxin 1B
SYNJ1	Synaptojanin 1
SYT11	synaptotagmin XI
TAF1	TAF1 RNA Polymerase II, TATA Box Binding Protein

	(TBP)-Associated Factor
TBE	tris-borate-EDTA solution
TDP-43	TAR DNA-binding protein 43
TH	Tyrosine Hydroxylase
THAP1	Thanatos-associated protein domain-containing apoptosis-associated protein 1
TREM2	triggering receptor expressed on myeloid cells 2
TUBB4A	Tubulin beta 4A class IVa
UCHL1	Ubiquitin carboxyl-terminal hydrolase L1
UPSIT	University of Pennsylvania Smell Identification Test
UV	Ultraviolet light
VCP	valosin containing protein
VPS35	Vacuolar protein sorting-35
WDR62	WD repeat domain 62
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
xg	centifugal force
YOPD	Young-onset Parkinson's Disease
µl	microliter

PUBLICATIONS

Significant Contribution

Sheerin UM*, Schneider SA*, Carr L, Deuschl G, Hopfner F, Stamelou M, Wood NW, Bhatia KP. ALS2 mutations: Juvenile amyotrophic lateral sclerosis and generalized dystonia. Neurology. 2014 Feb 21. [Epub ahead of print]

Kojovic M, **Sheerin UM**, Rubio-Agusti I, Saha A, Bras J, Gibbons V, Palmer R, Houlden H, Hardy J, Wood NW, Bhatia KP. Young-onset parkinsonism due to homozygous duplication of α-synuclein in a consanguineous family. Mov Disord. 2012 Dec;27(14):1827-9.

Sheerin UM, Stamelou M, Charlesworth G, Shiner T, Spacey S, Valente EM, Wood NW, Bhatia KP. Migraine with aura as the predominant phenotype in a family with a PRRT2 mutation. J Neurol. 2013 Feb;260(2):656-60.

Tucci A, Charlesworth G, **Sheerin UM**, Plagnol V, Wood NW, Hardy J. Study of the genetic variability in a Parkinson's Disease gene: EIF4G1. Neurosci Lett. 2012 Jun 14;518(1):19-22.

Sadnicka A, **Sheerin UM**, Kaplan C, Molloy S, Muraro PA. Primary progressive multiple sclerosis developing in the context of young onset Parkinson's disease. Mult Scler. 2013 Jan;19(1):123-5.

Sheerin UM*, Charlesworth G*, Bras J, Guerreiro R, Bhatia K, Foltynie T, Limousin P, Silveira-Moriyama L, Lees A, Wood N. Screening for VPS35 mutations in Parkinson's disease. Neurobiol Aging. 2012 Apr;33(4):838.e1-5.

Others

Charlesworth G, Plagnol V, Holmström KM, Bras J, **Sheerin UM**, Preza E, Rubio-Agusti I, Ryten M, Schneider SA, Stamelou M, Trabzuni D, Abramov AY, Bhatia KP, Wood NW. Mutations in ANO3 cause dominant craniocervical dystonia: ion channel implicated in pathogenesis. Am J Hum Genet. 2012 Dec 7;91(6):1041-50.

Charlesworth G, Gandhi S, Bras JM, Barker RA, Burn DJ, Chinnery PF, Gentleman SM, Guerreiro R, Hardy J, Holton JL, Lees A, Morrison K, **Sheerin UM**, Williams N, Morris H, Revesz T, Wood NW. Tau acts as an independent genetic risk factor in pathologically proven PD. Neurobiol Aging. 2012 Apr;33(4):838.e7-11.

Stamelou M, Charlesworth G, Cordivari C, Schneider SA, Kägi G, **Sheerin UM**, Rubio-Agusti I, Batla A, Houlden H, Wood NW, Bhatia KP. The phenotypic spectrum of DYT24 due to ANO3 mutations. Mov Disord. 2014 Jan 17.

Rubio-Agusti I, Pareés I, Kojovic M, Stamelou M, Saifee TA, Charlesworth G, **Sheerin UM**, Edwards MJ, Bhatia KP. Tremulous cervical dystonia is likely to be familial: clinical characteristics of a large cohort. Parkinsonism Relat Disord. 2013 Jun;19(6):634-8.

Wray S, Self M; NINDS Parkinson's Disease iPSC Consortium; NINDS Huntington's Disease iPSC Consortium; NINDS ALS iPSC Consortium, Lewis PA, Taanman JW, Ryan NS, Mahoney CJ, Liang Y, Devine MJ, **Sheerin UM**, Houlden H, Morris HR, Healy D, Marti-Masso JF, Preza E, Barker S, Sutherland M, Corriveau RA, D'Andrea M, Schapira AH, Uitti RJ, Guttman M, Opala G, Jasinska-Myga B, Puschmann A, Nilsson C, Espay AJ, Slawek J, Gutmann L, Boeve BF, Boylan K, Stoessl AJ, Ross OA, Maragakis NJ, Van Gerpen J, Gerstenhaber M, Gwinn K, Dawson TM, Isacson O, Marder KS, Clark LN, Przedborski SE, Finkbeiner S, Rothstein JD, Wszolek ZK, Rossor MN, Hardy J. Creation of an open-access, mutation-defined fibroblast resource for neurological disease research. PLoS One. 2012;7(8):e43099.

Nalls MA, Saad M, Noyce AJ, Keller MF, Schrag A, Bestwick JP, Traynor BJ, Gibbs JR, Hernandez DG, Cookson MR, Morris HR, Williams N, Gasser T, Heutink P, Wood N, Hardy J, Martinez M, Singleton AB; International Parkinson's Disease Genomics Consortium (IPDGC); Wellcome Trust Case Control Consortium 2 (WTCCC2); North American Brain Expression Consortium (NABEC); United Kingdom Brain Expression Consortium (UKBEC). Genetic comorbidities in Parkinson's disease. Hum Mol Genet. 2014 Feb 1;23(3):831-41.

Beilina A, Rudenko IN, Kaganovich A, Civiero L, Chau H, Kalia SK, Kalia LV, Lobbestael E, Chia R, Ndukwe K, Ding J, Nalls MA; International Parkinson's Disease Genomics Consortium; North American Brain Expression Consortium, Olszewski M, Hauser DN, Kumaran R, Lozano AM, Baekelandt V, Greene LE, Taymans JM, Greggio E, Cookson MR. Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. Proc Natl Acad Sci U S A. 2014 Feb 18;111(7):2626-31

Moskvina V, Harold D, Russo G, Vedernikov A, Sharma M, Saad M, Holmans P, Bras JM, Bettella F, Keller MF, Nicolaou N, Simón-Sánchez J, Gibbs JR, Schulte C, Durr A, Guerreiro R, Hernandez D, Brice A, Stefánsson H, Majamaa K, Gasser T, Heutink P, Wood N, Martinez M, Singleton AB, Nalls MA, Hardy J, Owen MJ, O'Donovan MC, Williams J, Morris HR, Williams NM; IPDGC and GERAD Investigators. Analysis of genome-wide association studies of Alzheimer disease and of Parkinson disease to determine if these 2 diseases share a common genetic risk. JAMA Neurol. 2013 Oct;70(10):1268-76. PubMed PMID: 23921447.

Pichler I, Del Greco M F, Gögele M, Lill CM, Bertram L, Do CB, Eriksson N, Foroud T, Myers RH; PD GWAS Consortium, Nalls M, Keller MF; International Parkinson's Disease Genomics Consortium; Wellcome Trust Case Control Consortium 2, Benyamin B, Whitfield JB; Genetics of Iron Status Consortium, Pramstaller PP, Hicks AA, Thompson JR, Minelli C. Serum iron levels and the risk of Parkinson disease: a mendelian randomization study. PLoS Med. 2013;10(6):e1001462.

Nalls MA, Duran R, Lopez G, Kurzawa-Akanbi M, McKeith IG, Chinnery PF, Morris CM, Theuns J, Crosiers D, Cras P, Engelborghs S, De Deyn PP, Van Broeckhoven C, Mann DM, Snowden J, Pickering-Brown S, Halliwell N, Davidson Y, Gibbons L, Harris J, Sheerin UM, Bras J, Hardy J, Clark L, Marder K, Honig LS, Berg D, Maetzler W, Brockmann K, Gasser T, Novellino F, Quattrone A, Annesi G, De Marco EV, Rogaeva E, Masellis M, Black SE, Bilbao JM, Foroud T, Ghetti B, Nichols WC, Pankratz N, Halliday G, Lesage S, Klebe S, Durr A, Duyckaerts C, Brice A, Giasson BI, Trojanowski JQ, Hurtig HI, Tayebi N, Landazabal C, Knight MA, Keller M, Singleton AB, Wolfsberg TG, Sidransky E. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. JAMA Neurol. 2013 Jun;70(6):727-35

Klebe S, Golmard JL, Nalls MA, Saad M, Singleton AB, Bras JM, Hardy J, Simon-Sanchez J, Heutink P, Kuhlenbäumer G, Charfi R, Klein C, Hagenah J, Gasser T, Wurster I, Lesage S, Lorenz D, Deuschl G, Durif F, Pollak P, Damier P, Tison F, Durr A, Amouyel P, Lambert JC, Tzourio C, Maubaret C, Charbonnier-Beaupel F, Tahiri K, Vidailhet M, Martinez M, Brice A, Corvol JC; French Parkinson's Disease Genetics Study Group; International Parkinson's Disease Genomics Consortium (IPDGC). The Val158Met COMT polymorphism is a modifier of the age at onset in Parkinson's disease with a sexual dimorphism. J Neurol Neurosurg Psychiatry. 2013 Jun;84(6):666-73

Holmans P, Moskvina V, Jones L, Sharma M; International Parkinson's Disease Genomics Consortium, Vedernikov A, Buchel F, Saad M, Bras JM, Bettella F, Nicolaou N, Simón-Sánchez J, Mittag F, Gibbs JR, Schulte C, Durr A, Guerreiro R, Hernandez D, Brice A, Stefánsson H, Majamaa K, Gasser T, Heutink P, Wood NW, Martinez M, Singleton AB, Nalls MA, Hardy J, Morris HR, Williams NM. A pathway-based analysis provides additional support for an immune-related genetic susceptibility to Parkinson's disease. Hum Mol Genet. 2013 Mar 1;22(5):1039-49.

Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simón-Sánchez J, Mittag F, Büchel F, Sharma M, Gibbs JR, Schulte C, Moskvina V, Durr A, Holmans P, Kilarski LL, Guerreiro R, Hernandez DG, Brice A, Ylikotila P, Stefánsson H, Majamaa K, Morris HR, Williams N, Gasser T, Heutink P, Wood NW, Hardy J, Martinez M, Singleton AB, Nalls MA; International Parkinson's Disease Genomics Consortium (IPDGC); Wellcome Trust Case Control Consortium 2 (WTCCC2). Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. Hum Mol Genet. 2012 Nov 15;21(22):4996-5009.

Mittag F, Büchel F, Saad M, Jahn A, Schulte C, Bochdanovits Z, Simón-Sánchez J, Nalls MA, Keller M, Hernandez DG, Gibbs JR, Lesage S, Brice A, Heutink P, Martinez M, Wood NW, Hardy J, Singleton AB, Zell A, Gasser T, Sharma M; International Parkinson's Disease Genomics Consortium. Use of support vector machines for disease risk prediction in genome-wide association studies: concerns and opportunities. Hum Mutat. 2012 Dec;33(12):1708-18.

FIGURES

TABLES

CHAPTER 1: INTRODUCTION

1.1 Specific Aims of this Thesis

Since 2005, next-generation DNA sequencing (NGS) platforms have become widely available, reducing the cost of DNA sequencing by several orders of magnitude relative to Sanger sequencing. The development of methods for coupling targeted capture and massively parallel DNA sequencing has made it possible to determine cost-effectively nearly all of the coding variation present in an individual human genome, a process termed 'whole exome sequencing' (WES). This technique has become a powerful new approach for identifying genes that underlie Mendelian disorders in circumstances in which conventional approaches have failed. Even where conventional approaches are eventually expected to succeed, for example autozygosity mapping, WES provides a method for accelerating discovery.

My thesis is designed to study the use of next-generation sequencing technologies, specifically whole-exome sequencing, to try and identify the genetic aetiology of families with Mendelian forms of Parkinson's disease and Dystonia in whom the molecular cause has not been determined.

1.2 The Burden of Neurodegenerative diseases and Insights from Genetic Analysis

Neurodegenerative diseases represent a significant burden on patients and carers, as well to wider society and the economy. As the elderly population increases worldwide, this burden is set to increase further. Although treatment options are already available for some conditions, these are generally of very limited

effectiveness and treat the symptoms rather than preventing onset. Parkinson's disease (PD) is common neurodegenerative disease. It affects >2% of those over the age of 75 years.[1] In the UK, there are over 100,000 people with the disease and with an aging population this is only set to increase. The annual cost in nursing home care for PD alone in the UK is estimated to be ~600-800 million.[2] The development of new therapeutic approaches is therefore essential. In the past dozen years, genes have been identified for the familial forms of neurodegenerative disorders, including, Parkinson's disease, Alzheimer's disease, amyotrophic lateral sclerosis, and fronto-temporal dementia.

Identification of these genes has been important for a number of reasons. Firstly, much of the recent progress in our understanding of the pathogenesis of neurodegenerative disease has been based on genetic analysis. This is particularly true of Parkinson's Disease (PD), where much of the latest advances in our understanding of the pathogenesis of neurodegenerative disease has been based on genetic analysis, identification of genes causing Mendelian forms of PD, highlight pathways important in the development of PD. *PINK1*, *PARKIN*, and *DJ-1* map to the mitochondrial damage repair pathway, whilst variability at the *HLA* locus indicates that immune response also plays a role. Secondly, these discoveries have given us the opportunity to re-create and study the mechanisms of neurodegenerative diseases in cell-culture and animal models and to use the findings to point the way towards developing pharmacologic and biologic therapies. For example, in Parkinson's disease, a *LRRK2* inhibitor is in pre-clinical development for potential use. Thirdly, genes identified as causing Mendelian forms of neurodegenerative disorders, are good candidates in which genetic variability may contribute to the risk of developing the sporadic form of the disease in the general population, for example *LRRK2* and *SNCA* in Parkinson's disease, Lastly, one of the criticisms of trialled novel therapeutic agents for Alzheimer's disease, is that they were not instituted early enough. Identifying cohorts of pre-symptomatic genetically defined individuals will be important to offer mechanistic therapies, very early in the disease course, and also to study such cohorts to identify biomarkers and characterise the earliest consequences of disease.

We have DNA from many families with Mendelian neurological disorders at the UCL Institute of Neurology, ascertained from the highly specialised clinics at the

National Hospital for Neurology and Neurosurgery (NHNN), in whom the causal gene has not been identified, either because the families were not suitable for linkage analysis or autozygosity mapping has revealed very large regions of homozygosity not amenable to Sanger sequencing. This thesis focuses on a new technology, called whole-exome sequencing, which can be applied to such families in order to identify the molecular cause.

1.3 Next Generation Sequencing Technology

First described by Sanger in 1977,[3] dideoxynucleotide sequencing of DNA has evolved into a large-scale production. Using Sanger sequencing, the cost per reaction of DNA sequencing fell in line with Moore's Law until January 2008, at which point, introduction of next-generation sequencing (NGS) resulted in a sudden and profound out-pacing of Moore's law.[4] NGS sets itself apart from conventional capillary-based sequencing, by the ability to process millions of sequence reads in parallel rather than 96 at a time, in a cost-effective manner.

Different NGS technologies share general processing steps, as shown in figure 1.1, while differing in specific technical details. UCL in-house exome sequencing uses Illumina Truseq capture kit version 3.0 (62 Mb) and sequencing takes place on a HiSeq 1000 (Illumina). The first step is to prepare a "library" comprising DNA fragments ligated to platform specific oligonucleotide adapters. The DNA is fragmented, and terminal overhangs are repaired, following which there is ligation to platform specific oligonucleotide adapters.

Figure 1.1 Next generation sequencing process steps for platforms requiring clonally amplified templates (Illumina, Roche 454 and Life Technologies)

Figure 1.1: Next generation sequencing process steps for platforms requiring clonally amplified templates (Roche 454, Illumina and life Technologies). Input DNA is converted to a sequencing library by fragmentation, end repair, and ligation to platform specific oligonucleotide adapters. Individual library fragments are clonally amplified by either (1) water in oil bead–based emulsion PCR (Roche 454 and Life Technologies) or (2) solid surface bridge amplification (Illumina). Flow cell sequencing of clonal templates generates luminescent or fluorescent images that are algorithmically processed into sequence reads. Figure adapted from Voelkerding et al., Journal of Molecular Diagnostics, 2010, 12: 539-551 (see reference [5])

The next major step is to prepare the "library" for massively parallel sequencing. For the Illumina platform, adapter modified library fragments are automatically dispensed onto a glass slide flow cell that displays oligonucleotides complementary to Illumina adapter sequences.[6] Illumina technology then uses a process called bridge amplification to generate clonal "clusters" of approximately 1000 identical molecules per cluster. Single-stranded, adapter-ligated fragments are bound to the surface of the flow cell exposed to reagents for polymerase-based extension. Priming occurs as the free/distal end of a ligated fragment "bridges" to a complementary oligo on the surface. Repeated denaturation and extension result in localised amplification of single molecules in "clusters". The Illumina sequencing platform,

utilises a sequencing-by-synthesis approach, in which all four nucleotides are added simultaneously to the flow cell channels, along with DNA polymerase, for incorporation into the oligo-primed cluster fragments. The nucleotides carry a base-unique fluorescent label and the 3'-OH group is chemically blocked, so that each incorporation is a unique event. An imaging step follows each base incorporation step, during which the flow cell is imaged. Subsequently, the 3' blocking group is chemically removed to prepare each strand for the next incorporation. The cycle is repeated, one base at a time, generating a series of images each representing a single base extension at a specific cluster (figure 1.2).

Above: The Illumina sequencing-by-synthesis approach. Cluster strands created by bridge amplification are primed and all four fluorescently labeled, 3′-OH blocked nucleotides are added to the flow cell with DNA polymerase. The cluster strands are extended by one nucleotide. Following the incorporation step, the unused nucleotides and DNA polymerase molecules are washed away, a scan buffer is added to the flow cell, and the optics system scans each lane of the flow cell by imaging units called tiles. Once imaging is completed, chemicals that effect cleavage of the fluorescent labels and the 3′-OH blocking groups are added to the flow cell, which prepares the cluster strands for another round of fluorescent nucleotide incorporation. Figure adapted from Mardis et al., Annu Rev Genomi Human Genet, 2008, 9:387-402 (see reference [7])

1.4 The Promise of Whole-Exome Sequencing

Many loci for Mendelian diseases have been identified by positional cloning in the past 20 years.[8-10] Indeed, all Mendelian forms of PD, up until 2011, were identified using this strategy, bar one, mutations in *GBA* were recognized as a risk factor for PD through astute clinical observation. Such approaches usually require large families, with affected and unaffected individuals. However, positional cloning methods are not suitable for all diseases. Some families may not be genetically informative, being small in size, and therefore not suitable for linkage, sometimes because the causal mutation is under negative selection, and therefore not transmitted through many generations, and therefore not suitable for linkage, or consanguineous families with very large regions of homozygosity, make Sanger-sequencing a costly, time consuming process. Additionally, mutations under strong negative selection, are likely to be *de novo* events, which cannot be ascertained at all by linkage analysis. Of

the nearly 7,000 known or suspected Mendelian disorders identified based on clinical features, less than half have been linked to a gene.[11]

The introduction and widespread use of massively parallel or 'next generation' sequencing, has made it increasingly practical to generate large amounts of sequence data cost-effectively. However, although this has made it possible for individual laboratories to sequence a whole human genome, the cost and capacity required are still significant, and interpreting variants in the non-protein-coding portion of the genome, is extremely challenging. It is estimated that 85% of disease-causing mutations are exonic, however, it is likely that this is inflated through ascertainment bias, since failed protein-centric disease studies are rarely published. Nevertheless, since protein-coding genes constitute approximately 1% of the human genome (the 'exome'), and harbor the majority of disease-causing mutations, it was clear that the development of viable methods for exome sequencing[12] would provide a powerful alternative to positional cloning with some notable advantages. Firstly, because potentially causal variants are identified, this method can be applied in families that are too small to provide meaningful information using linkage, effectively allowing small families and even single probands to be analyzed jointly, irrespective of allelic heterogeneity.[13] Secondly, this method can be incredibly fast, moving from well-defined trait to mutation within weeks rather than years.

1.5 Defining the Exome

A particular challenge for applying exome sequencing has been how to define the set of targets that constitute the exome. Considerable uncertainty remains regarding which sequences of the human genome are truly protein coding. Initial exome-capture kits used the CCDS (Consensus Coding Sequence Project) definition,[14] which is subset of genes determined to be coding with high confidence. However, most currently available commercial kits now target, at a minimum, all of the Refseq collection of genes and an increasingly large number of hypothetical proteins. In at least one example, identification of a novel disease gene would have been missed, if the CCDS definition was not expanded to the Refseq database.[15] Reflecting this, initial exome capture kits had a target of ~30Mb, whilst more recent kits have a target up ~62Mb.

1.6 Identifying causal alleles

A key challenge of using exome sequencing to identify novel disease genes for Mendelian disorders, is how to identify disease-related alleles among the background of non-pathogenic polymorphisms and sequencing errors. Exome sequencing on average will identify 24,000 single nucleotide variants (SNVs) in African American samples and ~20,000 in European American samples.[16] More than 95% of these variants are already known as polymorphisms in human populations. ~10,000 variants are non-synonymous (lead to differences in protein sequence) and ~11,000 are synonymous. A number of variants are likely to have greater functional impact: 80-100 nonsense variants (premature stop codons), 40-50 splice site and 200 inframe indels (see Table 1.1).

Table 1.1: Mean number of coding variant per exome

Variant Type	Mean number of variants in African Americans	Mean number of variants in European Americans
Novel Variants		
Missense	303	192
Nonsense	5	2
Synonymous	209	109
Splice	2	2
Total	520	307
Non-novel Variants		
Missense	10,828	9,319

Nonsense	98	89
Splice	36	32
Total	23,529	19,976
Total Variants		
Missense	11,131	9,511
Nonsense	103	93
Synonymous	12,776	10,645
Splice	38	34
Total	24,049	20,283

This table has been published elsewhere.[16] This table lists the mean number of coding single nucleotide variants from 100 sampled African Americans and 100 European Americans.

Strategies for finding causal alleles vary, depending of factors such as the mode of inheritance of a trait, the pedigree structure, whether a phenotype arises owing to de-novo or inherited variants; and the extent of locus heterogeneity for a trait. Such factors also influence the sample size needed to provide adequate power to detect trait-associated alleles.

Exome sequencing as a method to find causal mutations has already shown considerable promise, particularly in very rare diseases.[13, 15, 17, 18] Most of these studies have relied on comparisons of exonic variants found in a small number of unrelated or closely related affected individuals, to find rare alleles or novel alleles in the same gene shared among affected individuals.

1.7 Filtering for Rare Variants

Novelty of variants is assessed by filtering variants against a set of polymorphisms that are available in publically available databases (for example, dbSNP,[19] 1000 Genome project[20] and Washington Exome Server[21]). This approach is powerful because only a small fraction (2% on average) of the SNVs identified in an individual by exome sequencing is novel. Thus sequencing of only a modest number of affected individuals, then applying discrete filtering to the data, can be exceptionally

powerful for identifying new genes for Mendelian disorders.[13] However, this filtering approach can be problematic for a number of reasons, publically available databases such as dbSNP are 'contaminated' with a small but appreciable number of pathogenic alleles (e.g. the common p.G2019S mutation in *LRRK2* was present in dbSNP), filtering of observed alleles in a manner that is independent of their minor allele frequency (MAF) runs the risk of eliminating truly pathogenic alleles that are segregating in the general population at low but appreciable frequencies. This is particularly relevant for recessive disorders, in which the heterozygote state will not result in result in a phenotype that might otherwise exclude an individual from a 'control' population. However, analysis of rare recessive and dominant disorders in which one sets the maximum minor allele frequency (MAF) to 1% and 0.1% respectively, are still thought to be well powered.[16]

1.8 Deleteriousness of Variants

Further stratification of variants can be undertaken based on predictions of their deleteriousness. A greater weight may be given to nonsense and frameshift mutations, as they are predicted to result in a loss of protein function and are heavily enriched among disease-causal variation.[15, 22] However, this class of variation is not unambiguously deleterious, in some cases allowing functional protein production or resulting in loss of a protein that is apparently not harmful.[23] Alternatively, candidate variants can be stratified using existing biological or functional information about a gene: for example, a predicted role in a biological pathway or its interactions with genes or proteins that are known to cause a similar phenotype. Another approach for stratifying candidate alleles is to use quantitative estimates of mammalian evolution at the nucleotide level, which exploit the observation that regions of genes and genomes in which mutations are deleterious tend to show high sequence conservation as a result of purifying selection, examples of this include tools such as phastCONS[24], phyloP and Genetic Evolutionary Rate Profiling (GERP).[25] Several computational tools have been developed to predict the impact of a nonsynonymous SNV on protein function and hence distinguish pathogenic from neutral variants. These tools include SIFT[26], Polyphen2[27], MutationTaster[28] and Multivariate Analysis of Protein Polymorphism (MAPP).[29] The predictions are mostly based on the constraints imposed on amino acid changes in different regions

of a protein by checking the extent of sequence conservation across species. Each method has an estimated 80-90% sensitivity and 70-85% specificity in distinguishing mutations known to be pathogenic from those well-established to have no effects.[30, 31] Particular caution will be needed in late-onset diseases and in families with incomplete penetrance, however, as pathogenic variants often do not score highly in these programs.

1.9 Inheritance pattern

The pattern of inheritance of a monogenic disorder influences both the experimental design (for example, the number of cases to sequence, and selection of the most informative cases for sequencing) and the analytical approach. In recessive disorders, fewer cases need to be sequenced, and filtering of variants is likely to leave fewer candidate variants than dominant disorders, because the genome of any given individuals has around 50-fold fewer genes with two, rather than one, novel protein-altering alleles per gene. This is also supported by the greater number of genes for recessive disorders being identified, through exome sequencing.

1.10 Use of Pedigree Information

For Mendelian disorders, the use of pedigree information can substantially narrow the genomic search space for candidate causal alleles. Exactly which individuals are the most informative to sequence depends on the frequency of a disease-causing allele and the nature of the relationship between the individuals. For example, two first cousins share a rare allele that is identical-by-descent (that is they are inherited from a shared common ancestor) in approximately one-eighth of the genome. In the absence of mapping data, sequencing the two most distantly related individuals with the phenotype of interest can substantially restrict the genomic search space. When mapping data is available, the most efficient strategy is to sequence a pair of individuals whose overlapping haplotype (a combination of alleles on a single chromosome) produces the smallest genomic region. For consanguineous pedigrees in which a recessive mode of inheritance is suspected, sequencing the individual with the smallest region(s) of homozygosity, as determined by the genome-wide genotyping data, should be sufficient. Exome sequencing of parent-child trios is a

highly effective approach for identifying de novo coding mutations, as multiple de novo events occurring within a specific gene (or within a gene family or pathway) is an extremely unlikely event. [32]

1.11 How whole-exome sequencing is changing the field of Clinical Genetics and Neurogenetics

Genetic diagnosis and screening.

Many Mendelian neurological diseases, such as Parkinson's disease, dystonia, ataxia and dementias are genetically heterogeneous. Current screening is often designed to detect mutations in common mutational hot-spots. The utilization of WES to sequence several genes simultaneously for a genetically heterogeneous condition is more cost effective and quicker than by Sanger sequencing. A recent paper showed the utility of WES approach in the diagnosis of Mendelian disorders. Yang et al., applied WES to the diagnoses of 250 unselected, consecutive patients (80% of patients were children with neurological phenotypes) and observed a molecular diagnostic yield of 25%, which is higher than the positive rates of other genetic tests (karyotype, chromosomal microarray and Sanger sequencing).[33]

Expanding the phenotype associated with mutations in genes

WES has provided researchers with a powerful tool to identify mutations in genes previously associated to different disease phenotype or pathology. For example mutations in the *VCP* gene, previously linked to Paget's disease, inclusion body myopathy and frontotemporal dementia have been shown also to cause amyotrophic lateral sclerosis (ALS) in one of the first studies involving WES in neurodegenerative diseases.[34] Of note, this group showed that *VCP* mutations substantially contribute to the cause of familial ALS, being responsible for ~2% of cases. This finding broadened the clinical and pathological phenotype of *VCP* mutations to include ALS.

Recently, mutations in *ATP13A2*, a gene known to cause a form of dystonia-parkinsonism (Kufor-Rakeb syndrome, KRS), were found in a family with neuronal ceroid lipofuscinosis (NCL).[35] NCL is part of a heterogeneous group of inherited progressive degenerative diseases of the brain and sometimes the retina, that are characterized by lysosomal accumulation of auto fluorescent lipopigment. The

relationship between the diseases was not obvious, as the clinical features do not appear to overlap significantly. KRS typically presents with rigidity, bradykinesia, spasticity, supranuclear upgaze paresis and dementia. NCL disease varies according to the underlying gene defect and severity of mutation, but typically includes seizures, a progressive intellectual and motor deterioration, and in children but usually not adult onset cases, visual failure. These results indicate that broadening the phenotype associated with mutations provides information on the aetiological basis of disorders by uniting what is known about the biological underpinnings of apparently unrelated disorders into a single model. This finding shows that KRS is indeed linked to the lysosomal pathway, a pathway that was already hypothesized for a variety of parkinsonian phenotypes, but was not previously shown for KRS.

Gene identification in Mendelian Disorders

Large pedigrees are not available in many cases of late onset Neurogenetic disorders, in which older generations have died (DNA may not have been stored for all affected individuals) and younger generations have not yet reached the age of onset of disease onset. Previously, such pedigrees would have been intractable to typical gene mapping strategies such as linkage analysis. WES offers a technique that may be able to identify the causal variant in such families. One example of this is the discovery that recessive mutations in *WDR62* are a cause of a wide spectrum of cerebral cortical malformations. This study was carried out in a small kindred and would not have been amenable to traditional gene identification techniques.[18]

There have been a number of successes in novel gene discovery for Mendelian Neurogenetic conditions. Two groups used WES to identify the p.D620N mutation in *VPS35,* as a cause of autosomal dominant Parkinson's disease,[36, 37] whilst WES was used to identify mutations in *ANO3* and *GNAL* as a cause of primary dystonia[38, 39] and *PRRT2* as a cause of paroxysmal kinesigenic dyskinesia.[40]

De novo mutations in Neurological disease

De novo mutations represent the most extreme form of rare genetic variation: they are more deleterious, on average, than inherited variation because they have been subjected to less stringent evolutionary selection.[41, 42] This makes these mutations prime candidates for causing genetic diseases that occur sporadically. Indeed, recent WES studies have revealed de novo germline SNVs in single genes as the major

cause of rare sporadic malformation syndromes such as Schinzel-Giedion syndrome,[43] Kabuki syndrome[15] and Bohring-Opitz syndrome.[44] In addition, WES of affected and unaffected tissues has recently revealed de novo somatic SNVs as the cause of overgrowth syndromes such as Proteus syndrome.[45] Because de novo mutations are not rare events collectively it is possible that they are responsible for an important fraction of more commonly occurring diseases through disruption of any one of a large number of genes. Several pilot studies recently revealed that de novo mutations affecting may different genes in different individuals together might explain a proportion of common neurodevelopmental diseases such as intellectual disability,[46] autistic-spectrum disorders[47-51] and Schizophrenia.[32, 52] The realization that de novo mutations are potentially important in complex genetic diseases has major implications for our thinking about the causes, mechanisms and preventative strategies for these diseases.[53]

Complex diseases

In Alzheimer's Disease (AD), two groups have recently identified a rare variant in the *TREM2* gene associated with susceptibility to the disease, with a odds ratio of ~3.[54] *TREM2* was first nominated as a candidate gene, following the discovery of homozygous *TREM2* mutations as a cause of Nasu-Hakola disease, a rare recessive form of dementia with leukoencephalopathy and bone cysts.[55] Researchers generated exome sequence data sets and identified the R47H variant, which associated with the disease in cohorts from North America and Europe. This work was confirmed by researchers at deCODE Genetics,[56] who separately identified the R47H variant in a GWAS using the Icelandic population and replicated the association with AD in North American and European cohorts. TREM2 is an immune phagocytic receptor expressed in brain microglia. These studies suggest that reduced function of TREM2 causes reduced phagocytic clearance of amyloid proteins or cellular debris and thus impairs a protective mechanism in the brain, assuming that the risk variants impair TREM2 function.

1.12 Technical and Analytical Limitations

The most successful reports of the identification of a novel disease gene by exome sequencing have relied on discrete filtering, often with the aid of mapping data. However, it is difficult to know how often this approach has failed, as negative

results are rarely reported. Failure can result for many reasons, most of which can be broadly considered as either technical or analytical.

Technical failures

1. *Part or all, of the causative gene is not in the target definition*. The probes in sequence capture methods are designed based on the sequence information from gene annotation databases such as the consensus coding sequence (CCDS) database and Refseq database; therefore, unknown or yet-to-annotate exons cannot be captured. There may also be a failure in bait design so that an exonic region is not captured, for example, in GC-rich regions. Selectively sequencing the exome- which is, to our knowledge, the most likely region of the genome to contain pathogenic mutations – excludes noncoding regions. The contribution of mutations in non-coding regions to Mendelian disease has yet to be determined. For example, an intronic hexanucleotide repeat in *C9orf72* was recently identified as the cause of amyotrophic lateral sclerosis and frontotemporal dementia.[57, 58] It was missed using WES alone, as it was not part of the target definition and even deep resequencing of the entire region failed in the first instance. Finally, it is recognized that microRNAs, promotors and ultra-conserved elements may be associated with disease, but are not fully covered in WES capture kits.

2. *Inadequate coverage of the region that contains a causal variant*. A certain minimum depth of coverage is required for sufficient accuracy of variant detection; that is, positions or regions in the genome of the individual that are different from the reference human genome sequence. Typically, a minimum coverage of 8-10 reads per base is required for high-confidence detection of a heterozygous single nucleotide variant. Regions with repetitive sequences are more poorly characterized, as repetitive sequences may have prevented inclusion of a probe, or the reads originating from these regions cannot easily be mapped to a single position in the reference genome. Additionally probes may be poorly performing in GC rich regions.

3. *The causal variant is covered but not accurately called*. Frameshift indels in two individuals with Kabuki syndrome were undetected by exome sequencing, but were successfully identified by Sanger sequencing.[15] Mutations in the *MUC1* coding variable-number tandem repeat sequence in families with medullary cystic kidney disease type 1, were missed owing to poor sequence coverage because the region was excluded from whole-exome and regional capture probes owing to its low

complexity, extreme sequence composition and it was under-represented in quality-filtered data form the whole-genome sequence, owing to its high GC content and homopolymer content. Additionally, exome sequencing is unable to detect structural variants or chromosomal rearrangements, which are believed to be important for Mendelian disorders.

4. For analyses across families, true novel variants in the same gene are repeatedly identified but only because of the large size of the gene.

5. False variants in a gene are called because of mismapped reads or errors in the alignment or systemic artefacts that are specific to the peculiarities of a production pipeline.

Analytical failures

1. Analytical failures will result if there is analysis across several families/cases and there is genetic heterogeneity, or if an individual chosen for exome sequencing is a phenocopy. In such cases, detailed phenotyping may aid recognition of more than one causal variant.[15]

2. Additionally, false-positive calls are frequently observed in segmental duplications and processed pseudogenes. For example, repetitive regions and homologous sequences may mismap to the reference genome, generating false variants. This is the case of the glucosidase, beta, acid gene (*GBA*), that has a pseudogene with ~96% homology in the same genomic region. The existence of this similarity complicates the determination of the source of DNA sequenced fragments during alignment. Similarly, polyglutamine-type diseases are difficult to study by WES as the underlying defect is a repetitive sequence.

3. Pathogenic mutations may be present in publically available 'control' databases, and may therefore be erroneously filtered out. Currently, more than 17 million SNPs in the human genome have been documented in dbSNP with a false positive rate of ~15-17%.[59] Using an appropriate MAF for the mode of inheritance and the curating of databases such as 1000 genomes and Washington exome server will help to reduce this type of error.

Improvements in the next-generation sequencing technology and a wider definition of the exome will overcome some of these limitations in the future.

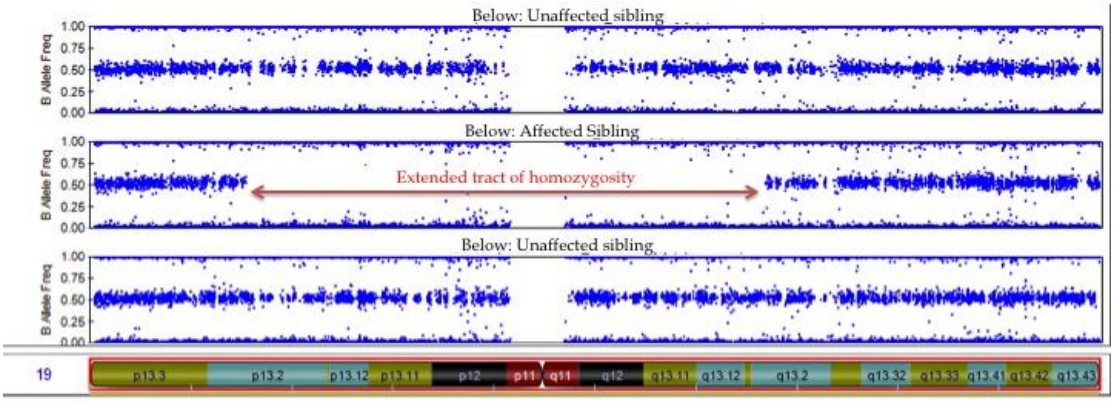
1.13 Mapping Strategies

1.13.1 Autozygosity Mapping

In a landmark paper in 1987, it was proposed that affected children born to consanguineous parents offered a powerful approach to disease gene mapping and identification.[60] This is because in such families, there is a high probability that an affected individual has inherited both copies (paternal and maternal) of the mutated gene from a common, comparatively recent ancestor. Consequently, the chromosomal region surrounding the mutation is expected to be homozygous; such a chromosomal region is said to be 'identical by descent' (IBD) or 'autozygous'. Of course it could be because a second, independent example of the same allele has entered the family at some stage (these alleles can be described as 'identical by state' IBS). Mathematically, it can be predicted that the longer the segment of homozygosity, the lower the probability that the markers tagging that segment have the same calls by chance alone.

Autozygosity mapping utilizes the fact that the disease causing mutation is the same in affected family members and will be embedded in a region of homozygosity in an inbred family. Each child will have multiple homozygous tracts in the genome, but only the tract(s) shared by all affected children would be the presumptive location of disease causing mutation. Furthermore, regions could be excluded using regions of homozygosity (for the same alleles) in unaffected siblings. Figure 1.2 shows a large region of homozygosity in chromosome 19 in a patient with generalised dystonia, her unaffected siblings are shown to be heterozygous for this region. A homozygous mutation in *GCDH* within this large tract of homozygosity was subsequently identified as the cause of dystonia. The patient is described in further detail in chapter 6.

Figure 1.2: Large tract of homozygosity in chromosome 19 in a patient with a *GCDH* mutation (see chapter 6.5)



Initially microsatellites were used for genome-wide genotyping however this was very time consuming. The development of high throughput genome-wide SNP genotyping revolutionized these projects allowing identification of regions of extended homozygosity with high resolution, with essentially complete genomic coverage and in a short amount of time. The data lends itself to immediate visualisation, with all tracts of disease segregating homozygosity being identified and all heterozygous regions/non-segregating homozygous tracts excluded, one can be confident that the region harboring the disease causing mutation has been identified making the generation of lod scores for these types of analysis effectively redundant.[61] Furthermore, this technique allows the direct visualisation of structural genetic variations, such as genomic deletions or duplications.

Up until recently, after the homozygous regions segregating with the disease have been identified, a candidate gene approach was employed. Genes within the homozygous regions were prioritised based on putative function, expression patterns and other data. Candidate genes were then Sanger-sequenced to pinpoint the causal mutation. Sanger sequencing of candidate genes was often the rate-limiting step in this process. The advent of WES has allowed the process of autozygosity mapping to be performed in a timelier manner, particularly in families who are highly inbred or have few affected siblings, in both instances this would

result in large amounts of autozygous regions, which would not have been amenable to a candidate gene approach.

Whether this technique identifies single or multiple regions of interest and the size of these regions relies on several factors; the degree of parental consanguinity, the number of informative family members, and the relatively stochastic nature of recombination. In families where affected family members exhibit a low level of inbreeding, or where there is a high degree of separation between affected family members, the size of a potential disease-segregating region is likely to be small.

There are however several pitfalls in autozygosity mapping. Although unlikely, it is possible for an extended consanguineous family to harbor mutations in two or more different genes giving rise to the same phenotype, particularly when the phenotype is known to display genetic heterogeneity.[62-68] Secondly, apparently shared autozygous blocks may in fact be Identical by state (IBS), which is particularly problematic when dealing with smaller intervals because the probability of sharing two haplotypes by chance is inversely correlated with their lengths. Finally, the number of shared autozygous blocks between different members of a given family is a function of the randomness of the crossing over events and their frequency. Although their randomness may not be predicted, the number of crossing over events correlates with number of meiotic events separating the patient from the shared parental ancestor.[69] Examples of the successful utilization of autozygosity mapping in Mendelian forms of parkinsonism include the mapping of *PLA2G6*[70], *SYNJ1*[71, 72] and *DNAJC6*.[73]

1.13.2 Genetic Linkage Analysis

Linkage analysis makes use of the exception to Mendel's Law of Independent Assortment which states that alleles at different genetic loci assort at random during meiosis; homologous chromosomes cross over and exchange genetic material during recombination, such that 50% of chromosomes will be recombinant, and 50% non-recombinant for these loci. Loci in close physical proximity on the same chromosome, however, tend to be inherited together and are said to be *linked* and alleles on the same small chromosomal segment tend to be transmitted as a block through a pedigree as a haplotype. Haplotypes mark chromosomal segments, which

can be tracked through pedigrees and through populations. Hence during a genome wide search, when DNA markers (with a known location) on the human genetic map, co-segregate with the disease (only affected subjects in a pedigree), linkage exists, and the DNA marker lies in close proximity to the disease gene.

The proportion of offspring in which two parental alleles are separated by recombination is the recombination fraction (θ). Θ is the probability that a parent will produce a recombinant offspring. The recombination fraction varies from 0 (for adjacent loci) to 0.5 (for distant loci) and may serve as a measure of the distance between the loci. For closely linked loci (where θ <0.05-0.1), it is reasonable to assume that the probability of more than one recombination occurring between the loci is small. In these circumstances the recombination fraction is equal to the genetic map distance between the loci, thus two loci showing recombination in 1% of meiosis (θ =0.01) are approximately 1 cM apart. Small values of θ are equivalent to the actual map distance (w) between loci, and thus recombination fractions are additive over small distances. The simplest case relating θ to w occurs when it can be assumed that multiple crossovers between two loci do not occur when the distance is very small, then θ=w.

Genetic mapping using linkage analysis has essential requirements including monogenic mode of inheritance that can be established by segregation analysis, correct phenotypic designation of affected and unaffected status. The LOD score method, [74] a maximum likelihood analysis, calculates the probability that two loci are linked, expressed as a LOD score, which is $\log_{10}$ of the odds ratio favouring linkage. Convention dictates that a LOD score >3, which indicates a probability in favour of linkage of 1000 to 1, is enough to establish linkage, and conversely a LOD score of -2 indicating a probability against linkage of 100 to 1 excludes linkage between the two loci being tested. Parametric LOD score analysis requires a precise genetic model, detailing the mode of inheritance, gene frequencies and penetrance of each genotype. The LOD score is calculated for various values of the recombination fraction (θ) using computer programs such as Merlin, to obtain the value of θ associated with the highest LOD score. This provides an estimate of the genetic distance between the two loci studied. Genetic studies of 'complex traits' such as PD and Dystonia face difficulties arising from uncertainties in diagnosis, disease definition and lack of understanding of genetic transmission. In addition in Mendelian disease, especially

with autosomal dominant inheritance, linkage analysis can be impaired by incomplete penetrance, variable phenotypic expression, genetic heterogeneity, and phenocopies.

1.14 Genetics of Selected Movement Disorders

1.14.1 Parkinson's Disease

Over the last 25 years, genetic findings have profoundly changed our views on the aetiology of Parkinson's disease (PD). Prior to the identification of the α-synuclein (*SNCA*) locus, epidemiologic studies consistently suggested that genes were unimportant in disease aetiology. To date, mutations in over 10 genes have been shown to cause monogenic forms of PD. Genome-wide association studies (GWAS) have provided convincing evidence that low-penetrance variants in at least some of these, but also in several other genes, play a direct role in the aetiology of the common sporadic form of the disease as well. In addition, rare variants with intermediate-effect strengths have been identified as important risk factors. Thus, an increasingly complex network of genes contributing in different ways to disease risk and progression is emerging. Functional work modeling PD disease-causing mutations in cell and animal models has greatly advanced our understanding of PD pathogenesis.

Idiopathic PD

The term parkinsonism defines the combination of two or more of four cardinal motor signs: bradykinesia, resting tremor, muscular rigidity, and postural instability.[75-77] Parkinson's disease is the most common cause of parkinsonism. The modified Queen Square Brain Bank criteria are the most frequently used diagnostic criteria, which rely on the presence of three cardinal signs (bradykinesia, rigidity, and rest tremor), responsiveness to dopaminergic therapy, and the absence of exclusion criteria, which if present, usually mean the diagnosis is that of Parkinson-plus syndrome or parkinsonism.[78] A constellation of non-motor symptoms often precede or accompany these features.

The pathological hallmark of PD is a region-specific selective loss of dopaminergic, neuromelanin-containing neurons from the pars compacta of the substantia nigra. This nerve cell loss is accompanied by three distinctive intraneuronal inclusions: the

Lewy body, the pale body, and the Lewy neurite. An abnormal, post-translationally modified, and aggregated form of the presynaptic protein, α-synuclein, is the main component of Lewy Bodies.[79]

Idiopathic PD is a common neurodegenerative disease, affecting >2% of those over 75 years.[1] The aetiology of PD is incompletely understood. Similar to other neurodegenerative diseases, ageing is the major risk factor. There is a negative association between PD and smoking, which is not accounted for by smokers dying younger, and therefore being less likely to develop a condition that is more common in old age.[80] Weak associations between PD and head injury, rural living, middle-age obesity, lack of exercise, well-water ingestion, and herbicide and insecticide exposure (paraquat, organophosphates, and rotenone) have also been reported.[81, 82] Environmental toxins (e.g. 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine MPTP, cyanine, carbon disulphide, and toluene) can produce a similar but not identical clinical picture.[83]

Twin studies have been used to estimate the genetic contribution to the pathogenesis and several of them showed low concordance rates in monozygotic and dizygotic twins.[84-86] A major criticism of these studies was that the cross- sectional study designs used did not exclude the possibility of a later disease onset in unaffected siblings. This obstacle has been overcome by using positron emission tomography studies (PET), which is sufficiently sensitive to identify pre- symptomatic subjects by detecting decreased striatal (18F)6- fluorodopa as a metric for decreased striatonigral dopaminergic function.[87, 88] Based on PET scan data, the concordance rate was significantly higher for monozygotic twins, than for dizygotic twins (55% versus 18%), suggesting a substantial genetic contribution to the PD pathogenesis.[88]

Mendelian Parkinson's disease and Parkinsonism

Parkinson's disease occurs as a sporadic disorder the vast majority of patients. However in approximately 5%-10% of cases, the disease occurs as a Mendelian disorder. As genes were identified to cause monogenic forms of PD, they were assigned *PARK* loci and numbered in chronological order of their identification. However, the *PARK* loci not only contain genes which cause monogenic forms of PD, but also loci identified from genome-wide linkage screens, some of which have not

been replicated in subsequent studies. To date, 18 PD-associated loci (*PARK1-18*) have been described (Table 1.2).

Table 1.2. *PARK*-designated Loci

Symbol	Locus	Features	Inheritance	Gene	Remarks
*PARK1**	4q21-22	Parkinsonism, dementia, autonomic dysfunction	AD	*SNCA* (Missense & Multiplications)	
PARK2	6q25.2-q27	Juvenile- or young-onset parkinsonism, often with foot dystonia, slow progression, good response to L-dopa	AR	*Parkin*	
PARK5	4p13	Classical PD	AD	*UCHL1*	Unconfirmed
PARK6	1p35-p36	early- onset parkinsonism, hyperreflexia, dystonia, slow progression	AR	*PINK1*	
PARK7	1p36	early- onset parkinsonism, psychiatric features, slow progression	AR	*DJ-1*	
PARK8	12q12	mid-late onset parkinsonism, good response to L-dopa	AD	*LRRK2*	
PARK9	1p36	Causes Kufor-Rakeb syndrome; juvenile-onset parkinsonism, spasticity, hallucinations, dementia, supranuclear gaze, paraparesis	AR	*ATP13A2*	
PARK13	2p12	L-dopa responsive Parkinson's disease	AD or risk factor	*HTRA2*	Unconfirmed
PARK14	22q13.1	L- dopa responsive dystonia-parkinsonism	AR	*PLA2G6*	
PARK15	22q12-13	Early-onset parkinsonian-pyramidal syndrome	AR	*FBX07*	
PARK17	16q11.2	mid-late onset parkinsonism, good response to L-dopa	AD	*VPS35*	
PARK18	3q27.1	mid-late onset parkinsonism, good response to L-dopa	AD	*EIF4G1*	Unconfirmed

*PARK4 is an erroneous locus, the family was subsequently proven to have a *SNCA* triplication (i.e. PARK1), AD, autosomal dominant; AR, autosomal recessive

Autosomal Dominant Parkinson's Disease Genes/Loci

Alpha synuclein, *SNCA* Mutations (PARK1)

In 1996, the mapping and subsequent identification of the first mutations responsible for PD indisputably showed that PD may be hereditary.[89, 90] The locus was mapped to chromosome 4q21-q22 in 1996, in an Italian-American kindred originating from Contursi, near Salerno in Southern Italy. This is the largest PD kindred characterized to date, with at least sixty affected members spanning four generations. A G>A transition at position 209 in exon 4 of the *SNCA* gene, resulting in an alanine to threonine substitution (p.Ala53Thr) was subsequently identified as the causative mutation in this kindred. Overall, point mutations in *SNCA* are rare, two other mutations p.A30P and p.E46K have been identified, in one family each.[91, 92]

Patients with *SNCA* missense mutations usually have early-onset PD (EOPD, age of onset <45 years), with an initially good response to levodopa treatment. However, the disease has a rapid progression and may be complicated by dementia or rarer features such as central hypoventilation and myoclonus.[93] Post-mortem studies in such patients show the presence of Lewy bodies, with spread through the substantia nigra, locus ceruleus, hypothalamus, and cerebral cortex.[89]

Recently, two additional *SNCA* point mutations (p.H50Q and p.G51D) have been identified. The p.H50Q mutation was reported in two patients of British ancestry, both had levodopa responsive, late-onset PD associated with cognitive impairment. In one case, there was no family history of the disease.[94, 95] The p.G51D *SNCA* mutation has been reported in association with a parkinsonian-pyramidal phenotype, with early-onset, moderate response to levodopa and rapid progression leading to death within a few years of onset.[96]

Further important insights into the link between *SNCA* and PD came from the discovery of gene multiplications. A triplication of the entire locus containing the *SNCA* gene was identified as a cause of autosomal dominant PD in a family.[97] This finding proved that not only structurally altered α-synuclein, but even the wild-type α-synuclein is pathogenic, if overexpressed. Additional families with

SNCA triplications and duplications, are described.[98] A dose dependence of the pathogenic effect has been observed: *SNCA* duplications lead to late-onset dopa-responsive parkinsonism resembling typical PD, whereas triplications cause earlier disease onset with more severe disease with dementia, rapid progression sometimes with additional features such as autonomic instability.[99] Penetrance of the missense mutations appears to be high (as high as 85% for the p.A53T mutation).[89] Data from rare kindred's with *SNCA* triplications suggests they are fully penetrant,[100] whereas *SNCA* duplications are not. [101, 102]

The identification of the first *SNCA* mutations soon led to the discovery that the encoded protein (aSYN) is the major fibrillar component of Lewy bodies and Lewy neurites,[103] the protein aggregates that had long been recognized as the pathological hallmark in familial as well as sporadic cases of the disease. The *SNCA* gene has six exons encoding an abundant 140-amino acid cytosolic protein, α-synuclein, which is natively unfolded, once bound to phospholipid membranes, through its amino-terminal repeats, it adopts structures rich in α-helical character. All three *SNCA* missense mutations impair the amino-terminal domain, and tend to produce mutant protein, which forms stable β-sheets, and thus exacerbate the formation of toxic oligomers, protofibrils, and fibrils.[104] Therefore, it is believed that the missense *SNCA* mutations cause PD through a toxic gain of function, and Lewy bodies may represent the attempt to purge the cell of toxic damaged α-synuclein.[105] Wild-type α-synuclein is selectively translocated into lysosomes for degradation,[106] and inhibitors of the lysosomal enzyme β-glucocerebrosidase, mutations in which represent a recognized risk factor for PD, modulate α-synuclein levels.[107]

Multiple genome-wide association studies have established that variation at the *SNCA* locus is the most significant genetic risk factor for sporadic PD.[108, 109] A meta-analysis of this data has found further association 5' of SNCA, suggesting variation in promotor/enhancer regions may modulate expression of *SNCA*.[110]

Leucine-rich repeat kinase 2, *LRRK2 Mutations (PARK8)*

Mutations in the *LRRK2* gene are the most frequent known cause of late-onset autosomal dominant and sporadic PD, with a mutation frequency ranging from 2% to 40% in different populations.[111-113] The clinical presentation of most

patients with *LRRK2* mutations is that of late-onset asymmetric levodopa-responsive tremor predominant PD, that is clinically indistinguishable from idiopathic PD.[114, 115] However, early onset and atypical features such as autonomic dysfunction, motor neuron disease and choreoathetosis and pyramidal tract signs are reported.[116, 117,118, 119]

LRRK2 is a large gene that consists of 51 exons. It encodes the 2527-amino acid cytoplasmic protein leucine-rich repeat kinase 2 (LRRK2) that consists of a leucine-rich repeat toward the amino terminus of the protein and a kinase domain toward the carboxyl terminus with various conserved domains in-between.[120] There are more than 100 *LRRK2* variants have been reported in association with PD (http://www.molgen.vib-ua.be/PDMutDB),[120, 121] however of these only 7 are unequivocally pathogenic based on co-segregation with disease in families and absence in controls: p.R1441G,[122] p.R1441C,[118] p.R1441H,[123] p.Y1699C,[118, 122] p.G2019S,[124, 125] p.I2020T,[118] and p.N1437H.[126] Several pathogenic *LRRK2* mutations cluster in functionally important domains and have been shown to alter their function: p.R1441C, p.R1441G and p.R1441H reduce the proteins GTPase activity, whilst the p.G2019S and p.I2020T mutations in the kinase domain have been shown to increase its activity and are thus a highly attractive therapeutic target (see [127] for review).

LRRK2 mutations are found in ~10% of autosomal dominant PD kindreds.[114] By far the most frequent and best-studied mutation is c.6055G>A (p.G2019S) that accounts for as many as 40% of cases of PD in Arab descent,[112] about 20% of Ashkenazi Jewish patients[113] and 1-7% of PD patients of European origin[128, 129]. Owing to a founder effect, the p.R1441G is very frequent in Basques[130] and p.I2020T in Japanese patients.[131] An incomplete, age-related penetrance (ranging from approximately 30%-70% by the age of 80 years in different studies) has been estimated for the G2019S mutation.[115, 132-134]

Patients with *LRRK2* mutations show pleiomorphic pathologies, which overlap with other neurodegenerative diseases. The commonest pathology is Lewy bodies with dopaminergic cell loss and gliosis in the substantia nigra. However, a minority of cases demonstrate nigral loss without Lewy bodies,[118, 135-138] whilst

others cases have been reported with neurofibrillary tangle pathology,[139] glial cytoplasmic inclusions similar to those seen in MSA[140] and some with anterior horn cell degeneration associated with axonal spheroids.[118]

A variety of functions and mechanisms have been ascribed to LRRK2 including neurite outgrowth,[141] protein translation through regulation of microRNA processing[142] and vesicle storage and mobilization within the recycling pool.[143] The physiological and pathological functions of LRRK2 have not yet been fully characterized.

Vacuolar protein sorting-35, *VPS35 Mutations* (PARK17)

Mutations in *VPS35* as a cause of familial PD were identified by exome sequencing by 2 independent groups, demonstrating the utility of this technique.[36, 37] To date, only 1 pathogenic mutation has been identified: p.D620N and it accounts for ~1% of familial PD cases, across several populations.[144, 145 146, 147] To-date the clinical phenotype associated with *VPS35* mutations closely resembles that of idiopathic PD. Reduced penetrance has been observed.[146, 147] Dementia although reported is not a prominent feature of presentation, the age of onset varies with some patients presenting with early onset PD.[144, 146, 147] So far, very little is known about the specific role of *VPS35* in PD, except that is a highly-conserved component of the retromer, a heteromeric complex that mediates retrograde transport of the transmembrane cargo from endosomes back to the trans-Golgi network.[148, 149] The pathology associated with mutations in *VPS35* has not been reported.

Ubiquitin carboxyl-terminal hydrolase L1, *UCH-L1* Mutations (PARK5)

Using a candidate gene approach, Leroy et al. described a small German autosomal dominant PD kindred, with a disease- segregating missense mutation (p.I93M) in *ubiquitin carboxyl-terminal hydrolase L1* (*UCH-L1*). Subsequent replication efforts have failed to confirm *UCH-L1* as a disease- causing gene in familial PD.[150, 151] The small size of the original family and the lack of replication indicate that this *UCH-L1* mutation is either a rare cause of PD or is not a disease- causing mutation.

Eukaryotic translation initiation factor 4 gamma 1, *EIF4G Mutations* (PARK18)

Mutations in *EIF4G* were described in a French family with autosomal dominant PD and Lewy body pathology. Linkage analysis showed a lod score of 3.01 for a region on 3q26-q28, a mutation in *EIF4G* was the only novel variant segregating in 10 affected family members. Several other variants in *EIF4G* were identified in other individuals with PD, but none came from large enough families to demonstrate segregation.[152] However, subsequent studies have not been able to convincingly replicate the association of autosomal dominant PD and *EIF4G* mutations, additionally, reportedly pathogenic variants have been identified in controls.[153-157] EIF4G is the core scaffold of a multisubunit translation initiation complex that regulates the translation initiation of mRNAs encoding mitochondrial, cell survival and growth genes in response to different stresses.[158, 159]

Dynactin 1, *DCTN1*

Mutations in *DCTN1* cause Perry Syndrome, an autosomal dominant, rapidly progressive form of parkinsonism associated with depression, weight loss and hypoventilation. Dynactin is a multisubunit protein complex that is required for most, if not all, long-distance trafficking of protein complexes and membrane organelles. All reported mutations cluster in exon 2, the P150^{glued} highly conserved N-terminal cytoskeleton-associated protein, glycine-rich (CAP-Gly) domain, which serves as a parking brake of the dynein motor. It is a very rare cause of familial parkinsonism with only 6 missense mutations reported (p.G67D, p.G71A/E/R, p.T72P, and p.Q74P) reported in 12 families worldwide.[160-171] Histology shows severe neuronal loss in the substantia nigra without Lewy bodies and a unique distribution of TDP-43-positive pathology in neurons and glial cells in a pallidonigral distribution.[172]

<u>Autosomal Recessive Parkinson's Disease: Genes and Loci</u>

***Parkin* Mutations (PARK2)**

Autosomal recessive early-onset PD with linkage to chromosome 6q was first recognized in Japanese families and subsequently mutations in *parkin* were identified by positional cloning.[173, 174] The disease usually starts in the third or fourth decade of the patient's life, and is usually slowly progressive with an

excellent response to dopaminergic treatment. However, some *Parkin*-mutation carriers have onset in childhood, and homozygous *Parkin* mutations are the most frequent cause of juvenile PD (age of onset <21years). Reported post-mortem examinations indicate that the substantia nigra shows neuronal loss and gliosis, however, it is frequently lacking Lewy bodies.[175, 176] Mutations in parkin explain up to 50% of cases with a clinical diagnosis of familial PD compatible with recessive inheritance and disease onset before the age of 45 years, and also ~15% of sporadic cases with onset before 45 years.[177] The Parkin protein functions as an E3 ubiquitin ligase in the process of ubiquitination, a form of posttranslational modification that conjugates ubiquitin proteins(s) to lysine residue(s) of target proteins, which in turn, determines their cellular fate. Disturbed elimination of damaged mitochondria, which is an important cellular process called mitophagy, appears to be one of the most important mechanisms for neuronal death in parkin deficiency. *Parkin* and *PINK1* (see below) appear to act together in this pathway[178, 179]. Parkin is recruited to damaged mitochondria, characterized by their low membrane potential and mediates the engulfment of mitochondria by autophagosomes, leading to their elimination.[180] Intact PINK1 seems to be necessary for this translocation of parkin.[180]

Phosphatase and tensin homolog (PTEN)-induced putative kinase 1, PINK1, (PARK6)

Mutations in *PINK1* are the second commonest cause of autosomal recessive early onset PD. The frequency of *PINK1* mutations is in the range of 1-9%, with considerable variation across difference ethnic groups.[181-184] An exception is in the Philippines, where one founder mutation has a high carrier frequency.[184] Interestingly, in contrast to *Parkin*, the majority of *PINK1* mutations reported are either missense or nonsense mutations, and to date only families with whole-exon deletions have been reported.[185-187] The pathology of *PINK1*-linked PD has been reported in a single patient and was characterized by neuronal loss in the substantia nigra pars compacta, Lewy bodies and sparing of the locus ceruleus.[188] PINK1 is a 581 amino acid ubiquitously expressed protein kinase. It consists of an amino-terminal 34 amino acid mitochondrial targeting motif, a conserved serine-theonine kinase domain (amino acids 156-509; exons 2-8), and a carboxy-terminal autoregulatory domain. Two-thirds of the reported mutations in PINK1 are loss-of-function mutations affected the kinase domain, demonstrating the importance

of PINK1's enzymatic activity in the pathogenesis of PD. PINK1 and Parkin function in a common pathway for sensing and selectively eliminating damaged mitochondria from the mitochondrial network. PINK1 is stabilized on mitochondria with lower membrane potential and recruits Parkin from the cytosol. Once recruited to mitochondria, Parkin becomes enzymatically active and initiates the autophagic clearance of mitochondria by lysosomes. i.e., mitophagy.[189]

Oncogene DJ-1, *DJ-1* Mutations (PARK7)

DJ-1 is the third gene associated with autosomal recessive PD, and it is mutated in about 1-2% of EOPD cases.[190] About 10 different point mutations and exonic deletions have been described, in both the compound heterozygous and homozygous state. The seven coding exons of the *DJ-1* gene code for a 189-amino acid-long protein that is ubiquitously expressed and functions as a cellular sensor of oxidative stress. [191, 192] The DJ-1 protein forms a dimeric structure under physiological conditions,[193] and some of the pathogenic missense mutations, particularly p.L166P prevent dimerization and lead to the degradation of protein.[194] No neuropathological examination of a patient with a homozygous pathogenic *DJ-1* mutation has been reported, and so it is not clear whether this is a Lewy body disorder. One possible hint that the disease caused by *DJ-1* mutations is not a Lewy body disease is the fact that, in patients with *DJ-1* related PD (as with *Parkin*-related PD) there is no olfactory involvement. Atypical phenotypes have been reported, including one family with early-onset parkinsonism, dementia and amyotrophic lateral sclerosis.[195]

Complex forms of Autosomal Recessive Parkinsonism

ATPase 13A2, *ATP13A2 Mutations* (PARK9)

ATP13A2 mutations underlie the autosomal recessive neurodegenerative disorder Kufor-Rakeb syndrome, a rare, juvenile-onset disorder characterized by parkinsonism, dementia, pyramidal tract signs, myoclonus and supranuclear gaze palsy.[196] Brain imaging often reveals moderate cerebral and cerebellar atrophy sometimes with iron deposition in the basal ganglia.[197] *ATP13A2* is a large gene, comprised of 29 exons coding for an 1180-amino acid protein. The

ATP13A2 protein is normally located in the lysosomal membrane and has 10 transmembrane domains and an ATPase domain.[196] Most mutations produce truncated proteins that are unstable and are retained in the endoplasmic reticulum and subsequently degraded by the proteasome. No disease causing copy number variants have been described. Mutations in *ATP13A2* are likely to have a role in lysosomal degradation based on the putative gene function, sural nerve pathology and the recent identification of a homozygous *ATP13A2* mutation in a kindred with a clinical and pathological diagnosis of Neuronal Ceroid Lipofuscinosis (NCL)[35] and that *ATP13A2* mutations cause a form of canine NCL.[198, 199]

F-box only protein, *FBXO7*, (*PARK15)*

Mutations in *FBXO7* present with juvenile-onset, severe levodopa-responsive parkinsonism and pyramidal signs.[200] The pathological features of this condition are not currently known. Recently, FBXO7 was shown to act in a common pathway with Parkin and PINK1 to induce mitophagy in response to mitochondrial depolarisation and that disease-associated mutations in *FBXO7* interfere with this pathway.[201]

Phospholipase A2, group VI (cytosolic, calcium-independent), *PLA2G6* Mutations (PARK14)

Recessive mutations in the phospholipase A2, group VI (*PLA2G6*) gene were initially described in as the cause of infantile neuroaxonal dystrophy and neurodegeneration associated with brain iron accumulation. However, *PLA2G6* mutations were later identified in patients with a different phenotype, consisting of dystonia predominant parkinsonism often accompanied by cognitive decline, eyelid opening apraxia, and supranuclear gaze palsy.[70] Brain MRI may show iron accumulation in the globus pallidus, substantia nigra and/or striatum. Pathologically, both early infantile and late onset *PLA2G6* is characterized by Lewy bodies and synuclein-positive dystrophic neuritis in the substantia nigra and cortex as well as tau immunoreactive cortical neurofibrillary tangles.[202]

Synaptojanin 1, *SYNJ1*

A homozygous mutation in *SYNJ1* (p.R258Q) was recently identified by exome sequencing, in two unrelated consanguineous kindred's, as a cause of a complex

autosomal recessive parkinsonian syndrome.[71, 72] The clinical phenotype includes early-onset, rapidly progressive parkinsonism with severe levodopa-induced dystonia. Additional features may be present, including: generalized seizures, cognitive decline, dysarthria, eyelid apraxia and supranuclear gaze palsy. Brain MRI demonstrated diffuse cortical atrophy with T2 hyperintensities. Synaptojanin 1 is a phosphoinositide phosphatase protein involved in clathrin-mediated endocytosis in the adult brain.[203]

Auxilin, *DNAJC6*

Recently, two recessive pathogenic mutations (splicing mutation c.801 -2 A>G and p.Q734X) have been reported in *DNAJC6* in association with juvenile onset parkinsonism with rapid progression to dependence 10-15 years after onset. In one family the parkinsonism was not levodopa responsive. Associated features, which may be present, include dysarthria, generalised seizures and brain atrophy on MRI.[73, 204] *DNAJC6* encodes auxilin, which is a clathrin-associated protein enriched in nerve terminals and may play a role in synaptic vesicle recycling.[205]

Identification of further multiplex kindreds with *DNAJC6* and *SYNJ1* mutations will confirm these as PD genes.

<u>High Risk Susceptibility Loci for PD</u>

Glucocerebrosidase, *GBA*

The astute clinical observation that patients and relatives of patients with Gaucher's disease, seem to present with PD more often than expected, led to studies which confirmed mutations in *GBA* are a risk factor for PD.[206-210] The odds ratio for developing PD with a single *GBA* mutation is estimated to be 5.4 (largely reflecting the common N370S mutation), with mutations in 15% of Ashkenazi Jewish patients and 3% of non-Ashkenazi Jewish patients, compared to 3% and <1% of matched controls.[211] Recently, mutations in *GBA* have also been confirmed as a significant risk factor for dementia with Lewy bodies (estimated odds ratio of 8.28) and PD with dementia (estimated odds ratio 6.48), confirming that PD, PD-dementia and Lewy body dementia form a continuum within the spectrum of Lewy body diseases.[212] Evidence suggests that impairment of the

aging lysosome, enhanced by deficient or mutant glucocerebrosidase, can affect α-synuclein degradation.[213] Functional loss of glucocerebrosidase causes accumulation and aggregation of α-synuclein, which in turn inhibits the trafficking and lysosomal activity of glucocerebrosidase leading to a self-propagating disease.[214]

Genome Wide Association Study (GWAS) identified Risk Factors For PD

Candidate gene association studies have been used to try and find loci in which common variability contributes to the risk of disease. Out of such studies genetic variability at the *SNCA*,[215] and the *MAPT (*microtubule-associated protein tau) locus,[216] were shown to contribute to the risk of disease. Over the last few years, there has been a revolution in the power of association studies through the development of platforms that allow the assessment of genetic variability on the order of 500,000 to 1 million variants (usually single-nucleotide polymorphisms) across the genome with the disease.[217] These analyses have consistently confirmed association of *SNCA* and *MAPT*. [109, 110, 218-220] Additional loci have been more recently identified and replicated, including: *HLA-DRB5*, *BST1*, *LRRK2*, *GAK*.[108-110, 218, 220-222] Imputation and a recent meta-analysis identified ten new loci, the nearest gene to these loci being *ACMSD*, *MCCC1/LAMP3*, *SYT11*, and *CCDC62/HIP1R*[222] and *PARK16*/1q32, *STX1B*, *FGF20*, *STBD1*, and *GPNMB*.[223]

Variability in both in the promotor and 3′ end of *SNCA* are the most consistently reported associations with PD. The REP1 microsatellite in *SNCA*, located ~10Kb upstream of the translational start of *SNCA*, has been linked to PD, and its pathological effect has been suggested to be mediated by gene expression. Interestingly, it is in linkage disequilibrium with PD associated 3′ polymorphisms ascertained using GWAS studies. This supports the notion that rare familial disease is aetiologically related to typical sporadic PD and suggests that genes that contain common risk variants are excellent candidates to also contain rare disease-causing mutations. It is more surprising that genetic variability in the *MAPT* locus, is associated with PD. Predictably, a number of tauopathies are associated with common or rare variants in *MAPT*, including progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and frontotemporal dementia with parkinsonism linked to chromosome 17.[224-226]

MAPT is part of a large conserved region on the long arm of chromosome 17. Two main extended haplotypes (H1 and H2) that diverged about 3 million years ago and stretch about 900kb in size have been described at this locus. These two haplotypes are inverted relative to each other and do not recombine.[227, 228] The H1 haplotype, confers risk with an odds ratio (OR) of 1.5. It is most likely that variants in *MAPT* act to increase expression.[222] A recent analysis, using only neuropathologically proven PD, shows that the *MAPT* association is not a result of unrecognized contamination of PD case cohorts with PSP.[229]

Currently known PD risk alleles each confer only a ~1.3 fold increase in risk. Even when considering cumulative risk burden for those in the highest quintile of disease risk versus the lowest quintile, the odds ratio is estimated to be 2.5.[222]

1.14.2 Genetics of Dystonia

The dystonias are a heterogeneous group of hyperkinetic movement disorders, characterized by involuntary sustained muscle contractions that lead to abnormal postures of the affected body part. Three basic parallel approaches are used to classify dystonia: age of onset, distribution of affected body parts, and cause. Clinically the most useful way to categorize dystonia is by aetiology:

- Primary Dystonia: dystonia +/- tremor without a secondary cause and no neurodegeneration
- Dystonia-plus syndromes: dystonia occurs with other clinical signs, however there is no secondary cause or neurodegeneration
- Secondary Dystonia: results from a secondary cause e.g. brain injury
- Heredo-degenerative dystonia: dystonia occurs in the context of a wider neurodegenerative syndrome
- Paroxysmal Dystonia: episodic attacks of dystonia
- Psychogenic Dystonia

The pathogenesis of primary dystonia is poorly understood, although family studies have led to the identification of 14 disease genes, no association studies for dystonia have been published.

Primary Torsion Dystonia

TOR1A Mutations (DYT1)

A 3bp (CAG) deletion in exon 5 of the *TOR1A* gene, is a major cause of early onset, generalized dystonia. The delGAG mutation removes glutamic acid form the C-terminal region of the encoded protein, torsin A. Two further pathogenic variants have been described: an 18bp deletion,[230] and a 4bp deletion.[231] DYT1 dystonia is inherited in an autosomal dominant fashion, but has a markedly reduced penetrance: of 30%.[232] The delGAG mutation is present across ethnicities, however, because of a founder mutation, it accounts for 80% of early onset dystonia in the Ashkenazi Jewish population. Penetrance is lower (~3%) in those with histidine at amino acid 216 of torsin A, suggesting that this single nucleotide polymorphism acts as a genetic modifier of DYT1.[233] Torsin A is a member of the AAA+ superfamily (ATPases associated with a variety of cellular activities). This family of proteins has many functions that are critical to assembly, disassembly, and operation of protein complexes. Torsin A is widely distributed in the brain; it is restricted to neurons, and normally associated with the endoplasmic reticulum (ER). In cellular models, mutated torsin A relocates from the ER to the nuclear envelope.[234] Its aberrant localization and interactions may result in stress-induced abnormalities, including impaired dopamine release. [235, 236] It has also been shown that mutant torsin A interferes with cytoskeletal events, which may affect development or neuronal pathways in the brain.[237]

Thanatos-associated protein domain-containing apoptosis-associated protein 1, *THAP1 (DYT6)*

Mutations in *THAP1* cause autosomal dominant dystonia with reduced penetrance (~ 60%), typically involving cranial or cervical muscles; laryngeal involvement is also frequent.[238] THAP1 is a member of a family of cellular factors sharing a highly conserved THAP DNA binding domain, which is an atypical zinc finger.[239] The mechanism by which mutant THAP1 protein causes dystonia is unclear, although recent work shows that THAP1 is a transcription factor and that mutant THAP1 leads to altered transcription of *TOR1A*, implying a linked mechanism.[240, 241]

CDKN1A Interacting Zinc Finger Protein, *CIZ1*

Linkage analysis combined with exome sequencing were recently used to identify a mutation in *CIZ1* (c.790A>G p.S264G) as the likely causal variant in a large Caucasian kindred with primary cervical dystonia inherited as an autosomal dominant trait. *CIZ1* is a $p21^{Cip1/Waf1}$-interacting zinc finger protein, expressed in brain and involved in DNA synthesis and cell-cycle control. The c.790A>G mutation is thought to alter splicing and subnuclear localization of CIZ1. Screening in a cohort of patients with adult-onset dystonia identified 2 additional missense mutations (p.P47S and p.R672M), however these were not found in multiplex kindred so segregation analysis was not possible. To date mutations in *CIZ1* have not been replicated as a cause of autosomal dominant cervical dystonia.

Anoctamin 3, *ANO3* (DYT23)

Mutations in *ANO3* as a cause of autosomal dominant cranio-cervical dystonia were identified as a result of WES and linkage analysis.[38] To date 3 mutations have been shown to segregated with the disease in unrelated kindreds (p.Arg494Trp, p.Trp490Cys and p.Ser685Gly). The phenotype is characterized by focal or segmental dystonia, variably affecting the cranio-cervical, laryngeal or brachial regions. There was often dystonic tremor with a jerky quality affecting the head, voice, or upper limbs. The age of onset ranges between childhood and 40 years. Little is known about the protein anoctamin 3 other than it is very highly expressed in the striatum. Functional work in patient fibroblasts with the p.trp490Cys mutation showed evidence of a potential defect in endoplasmic reticulum related Ca^{2+} handling. It is believed that Ca^{2+}-activated chloride channels have a role to play in the modulation of neuronal excitability.[242, 243] Since ANO3 is highly expressed in the striatum it is possible that mutations lead to abnormal excitability of striatal neurons.

Guanine Nucleotide Binding Protein (G Protein), Alpha Activating Activity Polypeptide, Olfactory Type, *GNAL*

A combination of WES and linkage analysis was also used to identify mutations in *GNAL* as a cause of autosomal dominant dystonia.[39] More than 8 mutations have been reported. Patients predominately have cranio-cervical onset of dystonia with generalisation in ~10%. *GNAL* encodes $G\alpha_{olf,}$ the alpha subunit of trihetrometric G protein $G_{olf,}$ which is involved in dopamine signalling. There is

evidence to show that $G\alpha_{olf}$ is mostly responsible for the coupling of D1 receptors to adenylyl cylase in striatal neurons and that $G\alpha_{olf}$ is required for D1-mediated behaviour and biochemical effects.[244-246]

Tubulin beta 4A class IVa, *TUBB4A* (DYT4)

Mutations in *TUBB4A* were identified as a cause of autosomal dominant inherited 'whispering dysphonia' and generalised dystonia with a peculiar 'hobby horse' gait using a combination of WES and linkage analysis in a very large pedigree, independently by two groups.[247, 248] The gene is expressed throughout the brain, but most highly in the cerebellum, which has been linked to dystonia.[247] The mutation identified in the family (p.Arg2Gly) occurs at a highly conserved residue in the autoregulatory MREI (methionine-arginine-glutamic acid-isoleucine) domain. Mutations in this domain abrogate the autoregulatory capability of TUBB4A, which may affect the balance of tubulin subunits and interfere with proper assembly.[249] The findings suggest a role for the cytoskeleton in dystonia pathogenesis.

<u>Dystonia Plus Disorders</u>

TAF1 RNA Polymerase II, TATA Box Binding Protein (TBP)-Associated Factor, *TAF1 (DYT3)*

DYT3 related dystonia is primarily found in Filipino males, due to a founder mutation, and is inherited in an X-linked recessive fashion. It is due to a 2.6kb retrotransposon insertion in intron 32 of *TATA box-binding protein-associated factor 1 (TAF1),*[250] which appears to reduce neuron-specific expression of TAF1 and the dopamine receptor D2 gene in the caudate nucleus.[250, 251] Symptoms start as focal dystonia and progress to multifocal/generalized dystonia, sometimes with parkinsonism. Neuronal degeneration on postmortem analysis is described in association with DYT3.[252, 253]

GTP cyclohydrolase I, *GCH1*, (DYT5a)

Mutations in *GCH1* lead to autosomal dominantly inherited, L- DOPA-responsive dystonia (DRD). Typically, it presents in childhood, with limb onset dystonia, and a dramatic response to L-dopa. Parkinsonism may occur in the later stages of the disease. There is a female preponderance, which may be explained by a higher GCH1 enzyme activity in males.[254]

Tyrosine Hydroxylase, *TH* (DYT5b)

Dopa-responsive dystonia can also be inherited as an autosomal recessive disorder associated with mutations in *TH*. Typically this presents with a more severe disease, reflecting deficiencies of serotonin and norephinephrine as well as dopamine, with onset of parkinsonism, ptosis, myoclonic jerks, seizures and truncal hypotonia in infancy.

ATPase, Na+/K+ Transporting, Alpha 3 Polypeptide, *ATP1A3 (DYT12)*

Mutations in *ATP1A3* cause rapid onset dystonia with parkinsonism, which is inherited in an autosomal dominant trait with reduced penetrance (90%). The phenotype is characterized by abrupt onset and typically follows a rostro-caudal gradient of spread. [255] *ATP1A3* encodes the catalytic unit of the sodium pump that uses ATP hydrolysis to exchange Na+ and K+ across the cell membrane to maintain ionic gradients.

Dystonia Plus Myoclonus

Sarcoglycan, Epsilon, *SGCE* (DYT11)

Myoclonus-dystonia usually begins in childhood and is characterized by a combination of brief lightning-like myoclonic jerks most often affecting the neck, trunk and upper limbs, and focal or segmental dystonia in the neck or arms in about half of the cases.[256] Inheritance is autosomal dominant, and caused by mutations in the epsilon-sarcoglycan gene, penetrance is reduced and related to maternal imprinting.[257] *SGCE* is a member of the sarcoglycan gene family. In the brain *SGCE* mutations are thought to lead to mislocalisation from the plasma membrane to the endoplasmic reticulum and to promotion of its degradation by the proteasome.[258]

Paroxysmal Dystonia

Myofibrillogenesis Regulator 1, *MR-1 (DYT8)*

Symptoms of paroxysmal non-kinesigenic dyskinesia (PNKD) typically begin in childhood or adolescence and include dystonic and choreatic dyskinesias, lasting from minutes to hours with a frequency of 20 per day to twice per year, often precipitated by alcohol or caffeine.[259] It results from mutations in the *myofibrillogenesis regulator* (*MR-1*) gene and is inherited as an autosomal

dominant trait. The function of MR-1 is not fully understood, but it is homologous to glyoxalase hydroxyacylglutathione hydrolase, known to detoxify methylglyoxal, a compound found in coffee and alcohol, and a by-product of oxidative stress.[260]

Proline-Rich Transmembrane Protein 2, *PRRT2 (DYT10)*

Mutations in *PRRT2* cause paroxysmal kinesigenic dyskinesia. Affected individuals have short (seconds to minutes), frequent (up to 100 times per day) attacks of dystonic or choreiform movements, precipitated by sudden unexpected movements or startle, which are responsive to anticonvulsant therapy and usually begin in childhood. PRRT2 is highly expressed in the developing nervous system and truncating mutations, produce a protein, lacking the transmembrane domain, resulting in altered subcellular localization.[40]

Solute Carrier Family 2 (Facilitated Glucose Transporter), Member 1, *SLC2A1 (DYT18)*

Mutations in the *SLC2A1* gene encoding the glucose transporter GLUT1 cause paroxysmal exercise-induced dyskinesia (PED), which is characterized by attacks of combined chorea, athetosis, and dystonia in excessively exercised body regions which last from a few minutes to an hour.[261] The disease usually begins in childhood and can have other disease manifestations, including epilepsy and haemolytic anemia.[262, 263] GLUT1 is the main glucose transporter in the brain and PED is thought to be caused by reduced glucose transport into the brain, particularly when energy demand is high after prolonged exercise. The symptoms improve with intravenously administered glucose and with permanent ketogenic diet.

CHAPTER 2. MATERIALS AND METHODS

2.1 DNA EXTRACTION

2.1.1 DNA extraction from saliva

Genomic DNA was extracted from saliva samples using the Oragene© DNA kit according to the manufacturers instructions. Essentially, the saliva sample and Oragene solution was incubated at 50°C in an air incubator for 2 hours. Then, 500µl of the solution was transferred to a 1.5mL microcentrifuge tube. 20µl of Oragene DNA purifier was added and the tube was vortexed for a few seconds. The sample was then incubated on ice for 10 minutes. Next, the sample was centrifuged at room temperature for 15 minutes at 13,000 rpm. The clear supernatant was transferred into a fresh microcentrifuge tube. The pellet containing impurities was discarded. 500µl of 100% ethanol was added to the solution and inverted 10 times. The sample was left at room temperature for 10 minutes allowing the DNA to fully precipitate. The sample was then centrifuged at room temperature for 2 minutes at 13,000 rpm. The supernatant was removed and discarded. 250µl of 70% ethanol was added. The sample was allowed to stand at room temperature for 1 minute. The ethanol was then removed, taking care not to disturb the pellet. 100µl of DNA buffer was then added to dissolve the DNA pellet.

2.1.2 DNA extraction from blood

Blood from families was taken in EDTA (ethylenediaminetetraacetic acid) bottles and genomic DNA was extracted from whole blood in the Diagnostic Genetics Laboratory in the UCL Institute of Neurology using a FlexiGene© kit (Quiagen) according to manufacturers instructions. In brief, 300ul of whole blood is mixed with 750ul of buffer (FG1), which lyses the cells. The sample is centrifuged for 20 seconds at 10,000xg. The supernatant is discarded leaving only the pellet. 150ul of Buffer (FG2 – which contains a protease) is added and the tube vortexed until the pellet is homogenized. The sample is centrifuged for 5 seconds and then placed

in a heating block at 65°C for 5 minutes. 150ul of isopropanolol (100%) is then added and the tube inverted several times. This step precipitates the DNA. The sample is centrifuged for 3 minutes at 10,000xg. The supernatant is discarded. 150ul of 70% ethanol is added and the sample vortexed for 5 seconds and centrifuged for 3 minutes at 10,000xg. The supernatant is discarded and the pellet air dried. 200ul of Buffer (FG3) is added and the sample vortexed at 5 seconds. The DNA is then dissolved by heating the sample at 65°C in a heating block.

2.2 DNA Quantification

Genomic DNA is measured using the NanoDrop Spectrophotometer equipped with the software ND-1000. Absorbance measurements will measure any molecules absorbing at a specific wavelength, DNA will absorb at 260nm and will contribute to the total absorbance. For quality control, the ratios of absorbance at 260/280 nm and 260/230nm are used to assess the purity of DNA. Ratios of 260/280 + ~1.8 and 260/230 = 2.0-2.2 denote pure DNA. Ratios significantly lower than expected may indicate the presence of contaminants. The NanoDrop spectrophotometer is loaded with 1µl of solution and reads the DNA concentration that will be expressed in ng/µl.

2.3 Polymerase chain reaction (PCR)

PCR is performed using Roche Faststart PCR Master (400RXN/10ml) (Roche Applied Sciences), primers designed by ExonPrimer (http://ihg.gsf.de/ihg/ExonPrimer.html) or Primer 3 software (http://primer3.ut.ee/) at a dilution of 10 pmol/µl and genomic DNA is added to the reaction at a final concentration of 15-20ng. PCR amplification is performed using a program that has been shown to optimally amplify the product.

2.4 Agarose gel electrophoresis

10x TBE (tris-borate-EDTA solution) solution was prepared using 121.1g of Trizma base (Sigma), 61.8g Boric Acid (Sigma), 7.4g of Ethylenediaminetetraacetic acid (EDTA) (Sigma) and dissolved in 1 litre of

distilled water. Agarose gel electrophoresis is used to verify the quality and size of PCR amplification product. A 1.5% gel is prepared using Ultrapure Agarose (Invitrogen) and TBE 1X buffer and stained with gel red (Cambridge bioscience). 3µl of PCR product and 3µl of X6 Orange DNA loading dye (Thermo scientific) was loaded into each gel well. A DNA ladder (Midrange 100-2000 bp) (Qiagen) was run alongside the samples and a run at 100 mV for 30 minutes. DNA fragments are visualized using UV transilluminator.

2.5 PCR Cleanup

PCR cleanup can be performed either using an enzymatic or filtration method.

2.5.1 Enzymatic method

The enzymatic cleanup solution is a mixture of Fast-Alkaline phosphatase (Thermo scientific), which removes unused dNTPs, and Exonuclease I (Thermo scientific), which removes ssDNA from PCR products. The enzyme mix is prepared 1ml at a time, to minimize freeze thaw, into an eppendorf tube combining: 50µl Exonuclease I, 200µl Fast-Alkaline phosphatase and 750µl of purified water. This enzymatic cleanup solution is stored at -20°C.

5µl of PCR product is added to 2µl of the enzymatic cleanup solution. It is then run on a thermal cycler at 37°C for 30 minutes followed by 80°C for 15 minutes.

2.5.2 Filtration method

Alternatively, PCR product can be purified using a filtration method using 96 wells PCR filter plates (Millipore, Co. Cork, Ireland). 80µl of distilled/purified water was added to the PCR product. It was then transferred to the PCR filter plate and was placed on the vacuum for ~5 minutes until the membrane was dry. Subsequently PCR product is re-suspended in 25µl of distilled/purified water and shaken on the thermo-shaker (Eppendorf, Hauppauge, NY, USA) at room temperature at 300 rpm for 25 minutes. Purified PCR product is used for sequencing reaction.

2.6 Sequencing reaction

Sequencing Reaction is performed by means of 0.6µl BigDye Terminator v3.1, 2µl 5X Sequencing Buffer (Applied Biosystems), 0.6µl forward or reverse primers at 10 pmol/µl and purified PCR product (3.8µl of PCR product purified by filtration method or 3µl of PCR product purified by enzymatic cleanup) and 2.15µl of PCR grade water. The standard program recommended by Applied Biosystems is used to run the Sequencing reaction in the thermo-cycler.

2.7 Sequencing purification

Sequencing purification was performed via a filtration method using Millipore filtration plates or Sephadex plates.

2.7.1 Filtration method

Sequencing product can be purified using a filtering method using 96 wells PCR filter plates (Millipore, Co. Cork, Ireland). 80µl of distilled/purified water was added to the PCR product. It was then transferred to the PCR filter plate and was placed on the vacuum for ~5 minutes until the membrane was dry. Subsequently, the sequencing product is re-suspended in 25µl of distilled/purified water and shaken on the thermo-shaker (Eppendorf, Hauppauge, NY, USA) at room temperature at 300 rpm for 25 minutes.

2.7.2 Sephadex purification

A hydrated solution of Sephadex© was prepared using 2.9 grams of Sephadex G-50 powder (Sigma-Aldrich) and 40ml of autoclaved water. It was mixed well and allowed to hydrate for at least 30 minutes at room temperature. Next, 350µl of the well-mixed Sephadex solution was transferred to each well of a Corning FiltrEX™ 96 well filter plates (0.66 mm glass fibre filter). The Corning FiltrEX™ filter plate was placed on top of an empty collection plate and centrifuged for 3 minutes at 700xg. The Corning FiltrEX™ was then placed on top of a new plate.

The entire contents of the sequencing reaction were pipetted onto the Sephadex columns. The plates were then centrifuged for 5 minutes at 910xg.

2.8 Sanger sequencing and analysis

Sanger sequencing was performed on a 3730 DNA analyser (Applied Biosystems, Foster City, CA, USA) and electropherograms were visualised using Sequencher software (Gene Codes Corporation, MI, USA, version 5.0.1).

2.9 DNA array SNP analysis

Two types of DNA array SNP beadchips were used for mapping in conjunction with WES data. Human CytoSNP beadchips (Illumina) have ~220,000 markers and were used for whole genome linkage analysis. HumanOmniExpress (Illumina) beadchips have ~715,000 markers and were utilized in families where autozygosity mapping was performed. Samples were processed, hybridized and scanned in accordance with the manufacturers instructions at UCL Genomics. Clustering, normalization and genotype calls were performed using Genome Studio 2010.3 Genotyping module (Illumina).

2.10 Autozygosity Mapping

Autozygosity mapping was performed using the Homozygosity Detector plug-in software within the BeadStudio suite. Regions of shared homozygosity over 0.5Mb, that segregated with disease, were visually identified using the Illumina Genome Viewer tool with the Beadstudio suite.

2.11 Genome wide Parametric Linkage Analysis

Individuals from kindred's that were analysed by parametric linkage analysis were run on Human CytoSNP beadchips (Illumina). Clustering, normalization and genotype calls were performed using Genome Studio 2010.3 Genotyping module (Illumina). Plink output files were created in Genome Studio using the Plink report plug-in. Plink (http://pngu.mgh.harvard.edu/~purcell/plink/)

was then used to removed SNPs that have low call rates (<90%) and monomorphic SNPs (MAF <0.5%) using the commands –-maf 0.05 –-geno 0.1. A subset of markers were extracted to use for linkage analysis typically in the region of 5,000 using the command –-thin. The .ped, .dat, .map and .model files were then created as per instructions on the Merlin website (http://www.sph.umich.edu/csg/abecasis/merlin/tour/linkage.html).[264]

For the .ped file, affection status was indicated by a number: 1=unaffected and 2=affected 0=unknown, the sex of the individual was indicated: 1=male, 2=female. The structure of the pedigree was indicated by assigning every individual a unique number. In the model file the mode of inheritance was specified together with the estimated disease allele frequency and penetrance.

A program called Pedstats (http://www.sph.umich.edu/csg/abecasis/PedStats/download/) was used to check the input formats and pedigree consistency and genotyping errors using the command: pedstats –d file.dat –p file.ped.

Merlin was then used to perform genome wide parametric linkage analysis using the command: merlin –d file.dat –p file.ped –m file.map –model file.model –markernames –-pdf –-tabulate. This created output files of LOD scores by marker and a PDF graphical representation of the results. Linkage analysis was performed on at least 3 different randomly generated marker sets.

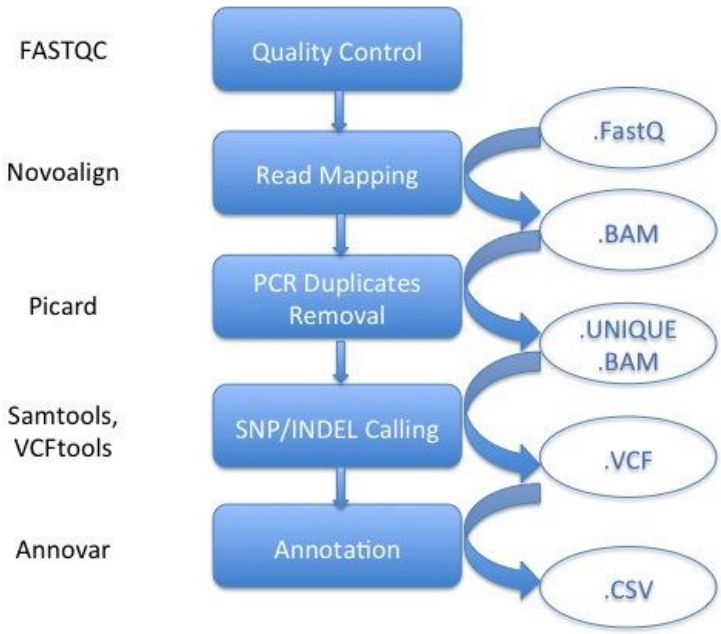
2.12 Whole exome sequencing

WES was either performed at UCL Institute of Neurology or at AROS Applied Biotechnology using the Truseq Exome Enrichment Kit, which targets 62Mb of exonic sequence and run on an Illumina Hiseq 2000. The Illumina truseq kit has more than 90% coverage of the exons or genes in the latest version of the CCDS and Refseq database. All sequencing reads were aligned to the hg19 build of the human genome. The bioinformatics pipeline for data analysis is described below.

2.13 Bioinformatic pipeline for next generation sequencing data

The schematic representation of WES data analysis is represented in figure 3.1. The whole exome sequencing data for the families described in this thesis was run through a bioinformatics pipeline, utilizing various programs, made by Dr. Vincent Plagnol. The components of the tools used to perform various steps of the data analysis are described below.

Figure 2.1: Analytical pipeline for WES.
Left panel indicates program used for each step. Middle panel indicates analysis steps. Right panel indicates data formats.



1. Quality control

The sequenced data generated by the HiSeq 2000 is stored in a FASTQ file, a text-based format for storing nucleotide sequence and it's corresponding quality scores. The certainty of each base call is recorded as a 'Phred' quality score, which measures the probability that a base is called incorrectly. The quality score of a given base, Q, is defined by the equation: $Q=-10\log_{10}(e)$ where 'e' is the

estimated probability of the base call being wrong. Thus, a higher quality score indicates a smaller probability of error. A Q20 value, for example, corresponds to a 1 in 100 error probability, and a Q30 value to a 1 in 1000 error rate (see table 2.1). NGS platforms have different error profiles and thus, quality values need to be derived accordingly.

Table 2.1: Phred quality scores

Quality Score	**Probability of Incorrect Base Call**	**Inferred Base Call Accuracy**
10 (Q10)	1 in 10	90%
20 (Q20)	1 in 100	99%
30 (Q30)	1 in 1000	99.9%

In Illumina technology the per-base quality score is determined by background noise during imaging[265]. In reality, Phred scores over 40 are uncommon and a cut-off Phred score of 20 is commonly used.[266]

Raw reads were analysed using FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FastQC runs a series of tests on a fastq file to generate a comprehensive quality control report. FastQC assesses data quality by evaluating: read length, per base quality score, per sequence quality scores, GC content, nucleotide content, sequence duplication and overrepresented sequences.

2. Alignment/Mapping

The first and arguably most crucial step of most NGS analysis pipelines is to map reads to sequences to the reference genome. The HiSeq 2000 can produce up to 6 billion paired-end reads, with a maximum read length of 2x100bp, equivalent to a maximum of 600Gb of data. Because of the large volume of reads and the huge size of the whole reference genome, alignment algorithms have been optimized for speed and memory usage. Therefore different alignment tools are designed with different approaches to trading off speed and accuracy to optimize detection of different types of variations in donor genomes. Alignment algorithms usually

follow a multistep procedure to map a sequence. First they quickly identify a small set of places in the reference sequence where the sequence is most likely to accurately align to. Then slower and more accurate alignment algorithms are run on the limited subset of possible mapping location identified in the first step. To speed up the process most alignment algorithms construct auxiliary data structures, called indices, for the reference sequence and/or read sequences. In addition to the selection of the alignment program, three issues are noteworthy. First, to overcome the problem of ambiguity when mapping short reads to a reference genome, paired-end reads have proven to be a valuable solution and are highly recommended, if not a requirement for WES.[267] Second, reads that can only be mapped with many mismatches should not be considered and as a consequence, mutations that are only backed by such reads should be discarded from further analysis. Thirdly, as current NGS technologies incorporate PCR steps in their library preparations, multiple reads originating from only one template might be sequenced, thereby interfering with variant calling statistics. For that reason, it is common practice to remove PCR duplicates after alignment in WES studies.

The analytical pipeline utilizes Novoalign (www.novocraft.com), an aligner designed for single and paired-end reads. The reference assembly (downloaded from UCSC: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/) was indexed; then the paired-end reads were aligned to the indexed set of reference sequence. Novoalign finds global optimum alignments using an index called fast-hashing Needleman-Wunsch algorithm, and uses affine gap penalties to allow consecutive run of spaces in a single string of a given alignment. Novoalign allows gaps up to 7bp on single end reads, even longer on paired end reads. Novoalign also supports paired-end mapping: it first finds the positions of all the good hits, sorts them according to the chromosomal coordinates and then does a scan through all the potential hits to pair the two ends. The expected size distribution for these sequencing runs were on average 300 base pairs as determined in the library preparation. Importantly, Novoalign calculates a mapping quality score, which is Phred-scaled probability of the alignment for the whole read being incorrect. This probability depends on the length of alignment, on the number of mismatches and gaps and on the uniqueness of the aligned region on the genome. It is important in the downstream analysis to distinguish

the real SNPs from the mismatches between, for instance repeated homologous genomic regions.

3. PCR duplicate removal; Samtools and Picard

Samtools (http://samtools.sourceforge.net/) is a suite of tools designed primarily to manipulate SAM files and the binary transformed BAM files, in preparation for downstream analysis. The SAM format was created in 2009[268] to define a generic nucleotide alignment format that described the alignment of large nucleotide sequences to a reference sequence. SAM files are tab-delimitated text files that contain an header section, which carries information on the project and the genome, and an alignment section, that contain alignment information such as such as mapping position and information on mismatches. This information is used for downstream analysis. Samtools was used for converting SAM and BAM files, for sorting (arranging the file according to left-most coordinates) and indexing (generating a complementary file '.bai' which aids fast access to BAM file) the alignment files generated by Novoalign. Samtools was also used for calling SNPs and short indel variants (see paragraph on SNP calling).

One technical artifact of capture-sequencing procedures is the generation of duplicate DNA sequencing reads (defined as reads with the same start point and direction) that are due to PCR-induced duplication. They share the same sequence and have the same alignment position. It is essential to identify and 'mark' duplicate reads so that they do not influence variant calling. PCR artifacts were removed with Picard (http://sourceforge.net/projects/picard/), which comprises Java-based command-line utilities that manipulate SAM or BAM files. Picard removes all read pairs with identical coordinates, only retaining the pair with the highest mapping quality. Picard examines aligned records in the supplied SAM file to locate duplicate molecules and generates a SAM output file that includes all aligned reads, without the duplicate records. Picard also generates a file that contains information on the percentage of PCR duplicates found in the original aligned file.

4.SNP Calling

Once an alignment is generated, SNP calling can be performed by comparing the aligned SAM/BAM file to the reference genome. The end result of a SNP calling

analysis is a collection of SNPs, each associated to a Phred-like quality score that takes into account base calling as well as mapping scores. A standard file has been created to hold these SNPs and related information, the Variant Call Format (VCF). BCF is the binary version of VCF: it keeps the same information in VCF, while much more efficient to process especially for many samples. The analytical pipeline utilizes Samtools and VCFtools for SNP calling,[269] a software suite that implements various utilities for processing VCF files, including validation, merging and comparing. Samtools collects summary information in the input BAMs, such as the number of different reads that share a mismatch from the reference, the cloning process artifacts (e.g. PCR induced mutations), the error rate associated with the sequence reads (e.g. the Phred score associated to every base in the read), the error rate associated with the mapping (mapping quality) and computes the likelihood of data given each possible genotype; then it stores the likelihoods in the BCF format. Bcftools performs the actual SNP calling. Then the .bcf was converted to .vcf, and only the SNPs on the set of exons targeted by Nimblegen (i.e., CCDS) were included. Lastly, the varFilter script was used to rule out error-prone variant calls.

5.Indels Calling

The analytical pipeline utilizes Dindel,[270] a program specifically designed for calling small indels from next-generation sequence data by realigning reads to candidate haplotypes. Dindel considers all candidate indels in a BAM file, and tests whether each of these is a real indel or a sequencing error or mapping error. In stage I) Dindel extracts all indels from the read-alignments in the BAM file. These indels are the candidate indels around which the reads will be realigned in stage III). In this stage, Dindel also infers the library insert size distributions. These will be used in stage III) for paired-end reads. In stage II) the candidate indels obtained in stage I) are grouped into windows of ~120 bp, into a realign-window-file. In stage III) for every window, Dindel generates candidate haplotypes from the candidate indels it detects in the BAM file, and realign the reads to these candidate haplotypes. Realignment is the computationally most intensive step. In stage IV) indel calls and qualities are produced in the VCF4 format. This step integrates the results from all windows into a single VCF4 file.

6. Annotation

The ANNOVAR tool[271] is a command-line tool for up to date functional annotation of various genomes, supporting SNPs, INDELs, block substitutions as well as CNVs. It provides a wide variety of different annotation techniques, organized in the categories gene-based, region-based and filter-based annotation. The tool relies on several databases, which need to be downloaded individually. This approach ensures that the correct database version is used and the download of large unnecessary data sets is avoided. ANNOVAR uses six different scores to estimate deleterious impact: GERP++,[272] LRT,[273] MutationTaster,[28] PolyPhen,[27] PhyloP conservation,[274] and SIFT.[26]

2.13.1 Copy Number Variant Analytical Pipeline for Next Generation Sequencing Data

A custom package called Exome Depth[275] was used to examine copy number variants in next generation sequencing data. This package used a read depth analysis approach to identifying copy number variants[276-278], i.e., it compares the number of reads mapping to a chromosomal window with an expectation under a statistical model. Deviations from this expectation are indicative of CNV calls. Reference samples will be those samples run on the same Illumina Hiseq run.

CHAPTER 3. *VPS35* SCREENING IN A PARKINSON'S DISEASE COHORT

3.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

Dr. Gavin Charlesworth and I both performed Sanger sequencing of *VPS35* in a large series of PD patients. I collected the clinical data relating to the family with the *VPS35* c.1858G>A; p.Asp620Asn mutation. Dr Laura Moriyama performed and interpreted the olfactory testing. Dr. Gavin Charlesworth and I jointly wrote the manuscript detailing this work, which is published in Neurobiology of aging.[144]

3.2 BACKGROUND

In 2011, using an whole-exome sequencing approach, two independent groups identified a missense mutation in vacuolar protein sorting 35 homolog (*VPS35* c.1858G>A; p.Asp620Asn) as the cause of autosomal dominant late-onset PD in a number of kindreds.[36] Pathogenicity is supported by segregation, evolutionary conservation at that base, and software predictions that the variant is likely to be damaging. However, at the time of this work, no other pathogenic mutations had been identified with certainty in *VPS35*, and the c.1858G>A mutation, had only been found in Caucasian individuals.

In this study, we screened a large series of Parkinson's disease patients for variants in *VPS35*, both in order to estimate the frequency of the published mutations and in order to search for novel mutations that may be disease-causing.

3.3 MATERIALS AND METHODS

3.3.1 Samples

The series included 160 familial PD cases, 175 young-onset PD cases, and 262 sporadic, neuropathologically confirmed PD cases (total number screened, 501). Neuropathologically confirmed cases were selected from the Queen Square and Parkinson's UK brain bank. The study was approved by the East Central London Research Ethics Committee 1 and informed consent was obtained as per its guidelines.

All living patients fulfilled criteria for clinical diagnosis of PD at the time of the study with at least 2 of 3 cardinal signs of tremor, rigidity, and bradykinesia, as well as a positive response to levodopa therapy. Familial cases were defined as those reporting 1 or more first degree relatives with PD with a pedigree consistent with autosomal dominant inheritance, who had tested negative for mutations in *SNCA* and *LRRK2*. Early-onset PD cases were defined as those who had developed their first signs of the disease at age 45 or younger (mean age of onset in young-onset PD was 37 ± 6, range 14–45 with a male to female ratio of 1.63:1). A family history of PD (first or second degree relative) was noted in 31 early-onset PD cases (17.7%). The neuropathological diagnosis for brain bank cases was made by an experienced neuropathologist and based on accepted morphological criteria.[279] This group included 166 males and 96 females, with an average age at onset of 64 ± 11 years (range, 29–85) and a family history recorded in 7.6% of cases.

3.3.2 PCR and Sanger sequencing

PCR and sequencing of the 17 coding exons of *VPS35* was performed as described in the Materials and Methods (chapter 2). PCR and sequencing purification was performed using the filtration method. To amplify exon 9 and 10, we added 1ul of 5% dimethylsulfoxide (DMSO; American Bioanalytical, MA, and USA) to the PCR mix.

All 17 exons and exon-intron boundaries of *VPS35* were sequenced in 96 familial PD cases (FPD), and exon 15 (in which the c.1858G>A; p.Asp620Asn mutation is found) in an additional 64 familial PD cases, 175 young-onset PD cases, and 262 sporadic, neuropathologically confirmed PD cases (total number screened, 501).

3.3.3 Variant Definitions

Any deviation from the reference sequence, ENST00000299138, was called a variant or a mutation. Variants were described as novel, if they were absent from dbSNP and 1000genomes data.

3.3.4 Clinical Characterisation of VPS35-related PD

Clinical details were collated were possible from living family members for those affected relatives that were deceased. Where possible affected individuals were clinical examined and olfactory testing undertaken using UPSIT.[280]

3.4 RESULTS

3.4.1 Sanger sequencing

Sanger sequencing revealed a single case from our familial PD series who harboured the previously described p.Asp620Asn mutation. No other nonsynonymous variants were detected in any other exon of any sample. In addition to this we observed 7 other sequence variants, including 2 novel and 5 previously described variants (Table 3.1)

Table 3.1: Summary of Variants found in *VPS35*.

This table has been published elsewhere (see reference [144]).

Variant	Nucleotide change	Amino Acid Change	Rs Number	Exon	Number of Cases
Coding					
Non-synonymous	c.1858G>A	p.Asp620Asn	Recently published	15	1
Synonymous	c.231T>C	p.Leu77Leu	rs11550462	4	1
	c.1842T>C	p.Tyr615Tyr	Novel	15	1
	c.1938C>T	p.His646His	rs168745	15	1
Non-coding					
UTR	c.1-34G>A		rs3743928		1
Intronic	c.3+25A>G		Novel	1	1
	c.1524+42G>C		rs4966616	12	27 (2 hom)
	c.1648-26G>A		rs2304492	14	58 (10 hom)

3.4.2 Clinical Characterisation of VPS35-related PD

In order to determine whether the p.Asp620Asn mutation segregated with PD in the family of the individual from our FPD series, we contacted other affected family members: a sister, and two cousins, they subsequently were found also to have the p.Asp620Asn mutation. The family is of European ancestry, the

pedigree (figure 3.1) is consistent with autosomal dominant mode of inheritance and revealed 9 further affected individuals across 3 generations, all of whom are deceased.

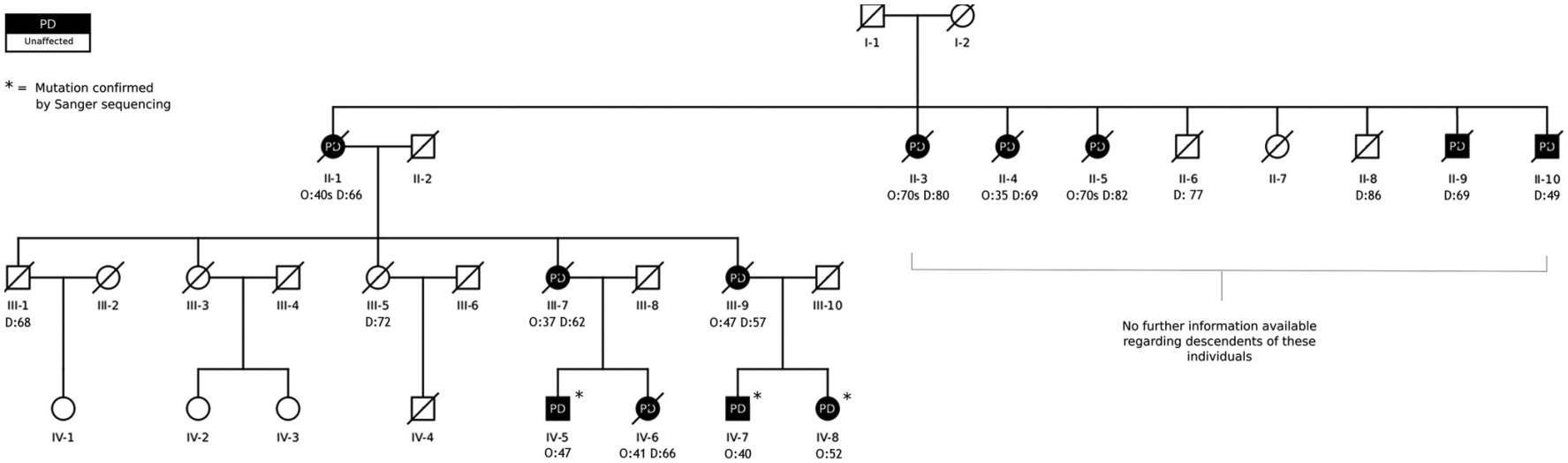


Figure 3.1: Pedigree of a family showing autosomal dominant inheritance of Parkinson's disease.

Age of onset (O:) and age at death (D:) are indicated where known for all descendants of I-1 and I-2. The p.Asp620Asn mutation in *VPS35* was confirmed by Sanger sequencing in the 3 living individuals affected by the disease.

The index case (IV-7) noticed his first symptoms in 1993 at the age of 40, consisting of stiffness in the left arm on waking followed, a few months later, by gradually worsening stiffness in the left leg. He was diagnosed as suffering from PD in 1995. At that time, examination showed unilateral bradykinesia and reduced arm swing. L-dopa was commenced with a good response, although he continued to gradually worsen. He began to notice a mild intermittent tremor of the left upper limb 13 years after his initial symptoms, followed by the spread of the stiffness to his right upper limb. Dyskinesias started after 9 years of L-dopa therapy, but were not severe. He complains of difficulty with balance, but has not had any falls. There is no history of mood disturbance or other psychiatric features, including visual hallucinations. Cognitively, he feels that his concentration has deteriorated somewhat over the years, but he has only recently retired from a demanding job. A Mini Mental State Examination (MMSE) performed 18 years into the disease revealed a score of 28/30. He has recently had deep brain stimulation, which has significantly improved his symptoms. Olfactory testing was undertaken 18 years into the disease, using the 40-item University of Pennsylvania Smell Identification Test (UPSIT)[280] and results were interpreted in comparison with 55 control subjects (27 males and 28 females) and 46 PD patients (30 males and 16 females) who had been previously tested for a study about olfaction in the UK.[281] The patient correctly identified 29/40 items of the UPSIT, scoring on the 37th percentile of 55 controls aged between 55 and 65 from the UK, and the 87th percentile of 46 PD subjects aged between 55 and 65 from the UK. His sibling (IV-8) developed symptoms aged 52 years of age and was formally diagnosed 2 years later. Initial symptoms included cramping in the left foot and tripping when walking. Within 6 months she developed a unilateral leg tremor. Further symptoms included micrographia, and freezing. Dopamine agonists were tried initially but within 3 years she required L-dopa. She has not developed any dyskinesias. She does not report any cognitive symptoms and a Mini Mental State Examination was 30/30. Olfactory testing was undertaken 11 years into the disease as described above with the patient also identifying 29/40 items of the UPSIT. Individual IV-5 presented with a stiffness and rigidity of the left leg at the age, followed by walking difficulties, bradykinesia, loss of postural reflexes and micrographia aged 47. There was no rest tremor. He was diagnosed shortly afterward. He responded well to L-

dopa, but developed dyskinesias after 10 to 15 years of treatment and has now had a deep brain stimulator implanted to good effect. There have been no hallucinations or psychiatric complications. The clinical history for individual IV-6, was provided by her brother. She developed PD at the age of 41 with noticeable tremor and generalized bradykinesia and rigidity. She responded well to L-dopa, developing only mild dyskinesias after a reasonable period of treatment. She underwent a pallidotomy at around 60 years of age and died at age 66 as a consequence of a nosocomial infection after a hip replacement operation.

Limited information is available for other affected family members. Individual II-1 presented with akinetic-rigid parkinsonism in her late fourth decade, family members noted that she never developed tremor, but suffered with depression late in the disease course. She died aged 66 years. Individuals II-3 and II-5 developed parkinsonism in their seventh decade and died in their eighth decade. II-4 developed tremor-predominant Parkinson's disease aged 35 years and had a slowly progressive disease course, dying at the age of 69. Individuals II-9 and II-10 are known to have been affected with parkinsonism, but ages of onset are not known. They died in their sixth and fourth decade respectively. Individual III-7 developed parkinsonism at age 37. Medical records and family members indicate left-sided rigidity, bradykinesia, hypomimia, falls, and freezing. Tremor was not recorded. Surgical interventions, including pallidotomy (right then left) and thalamotomy, were performed before L-dopa was commenced. No dyskinesias developed despite 10 years of treatment and she died at age 62. Individual III-9 developed parkinsonism aged 47 years presenting with a unilateral rest tremor. Family members report early severe muscular spasms, difficulty walking, and falls. Later in the disease, she developed depression and cognitive impairment. She died aged 57 years.

3.5 DISCUSSION

Two independent research groups used WES to identified a single c.1858G>A mutation in *VPS35* as the cause of Parkinson's disease in several kindreds. This study was an attempt to estimate the frequency of the published mutations and in order to search for novel mutations that may be disease-causing. We included PD

patients with a positive family history, as well as those with early onset and late-onset sporadic disease in order to maximize our chances of identifying mutations.

We identified one individual with the published disease-causing mutation (c.1858G>A; p.Asp620Asn). This individual had a family history consistent with highly penetrant, autosomal dominant Parkinson's disease. Based on clinical examination, clinical records and family reports, the disease in this family appears clinically similar to idiopathic PD. Onset was generally unilateral, with slow progression and variable tremor. Cognitive or psychiatric features are not prominent. Response to L-dopa was generally good and was not associated with early or unusually severe dyskinesias. The 2 patients in which olfaction was tested here show mild to moderate olfactory dysfunction when compared with controls in the UK, but still performed better than the vast majority of PD patients. Olfaction has been demonstrated to be unimpaired in homozygous and compound heterozygous carriers of Parkin mutations[282, 283] but impaired in *LRRK2* parkinsonian carriers,[115, 284-286] although not to the same extent as in sporadic PD. The most notable feature of our kindred is the relatively early age at onset. Six individuals developed PD in their third or fourth decades, of which 4 were younger than the age of 45. No other potentially disease-causing mutations were found in exon 15 (597 cases screened) or in any other exon (96 cases screened). We concluded that the c.1858G>A *VPS35* mutation, is not a common cause of PD, at least in our series of patients.

To date, robust evidence for pathogenicity has only been found for the c.1858G>A; p.Asp620Asn *VPS35* mutation. This mutation is located in the C-terminal region of the VPS35 protein, suggesting that subtle structural changes might influence the disease pathogenesis.[36] Subsequent to our screening study, a much larger multi-centre screening study also found that the c.1858G>A; p.Asp620Asn *VPS35* mutation was a rare cause of PD.[146] In this study 8,870 PD patients were screened for mutations in *VPS35*. Seven patients were found to have the c.1858G>A; p.Asp620Asn mutation in *VPS35*. Two of these patients were of Asian descent (the remainder were Caucasian) thus confirming a pathogenic relevance for this mutation in different populations. Two patients with this mutation did not have a

family history of PD confirming that this mutation is not likely to be fully penetrant.[146]

CHAPTER 4. WHOLE EXOME SEQUENCING IN AUTOSOMAL DOMINANT PARKINSONISM

4.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

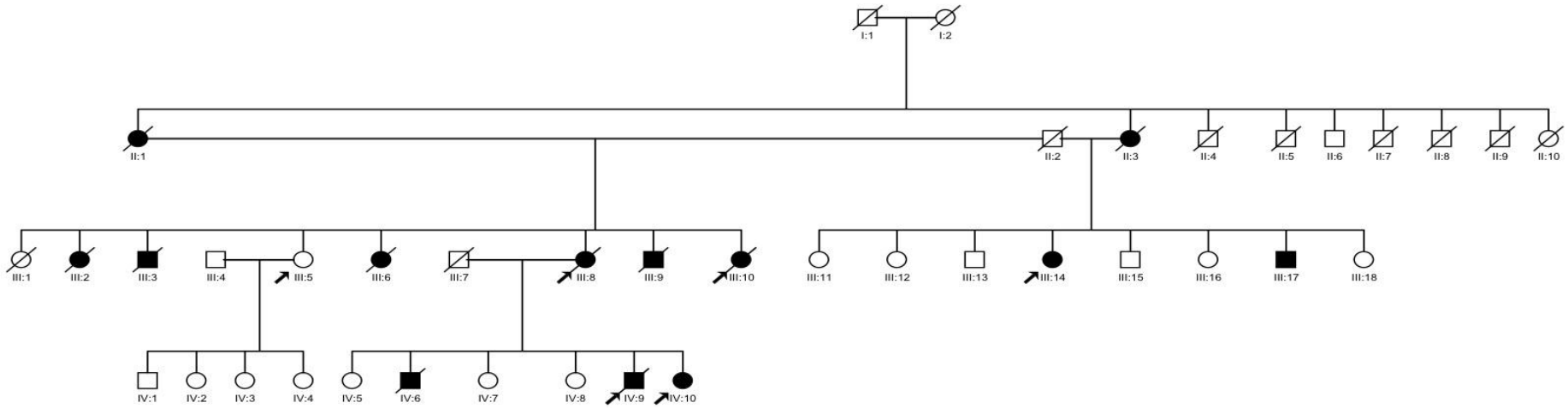
I examined affected and unaffected living family members and gathered medical records and DNA on deceased individuals. I analysed the clinical and WES data and performed the confirmatory Sanger sequencing. Prof. Tamas Revesz performed the histological examination of brain tissue from an affected family member.

4.2 BACKGROUND

We identified a large UK Caucasian family with atypical autosomal dominant Parkinson's disease (see Figure 4.1 for pedigree) through the Neurogenetics clinic at the National Hospital of Neurology and Neurosurgery. The index case (III:14) had been extensively investigated in multiple hospitals, including genetic testing for PD.

The family were keen to pursue genetic research studies to try and identify the molecular cause as family members wished to have predictive genetic testing and make reproductive decisions. DNA from four affected individuals was available at the beginning of the study, however two samples had only a very small amount of DNA (enough for a couple of PCR reactions), therefore precluding linkage analysis. In view of the limited samples available a WES approach was taken without linkage analysis to try and identify the causal variant.

Figure 4.1: A truncated pedigree of the family with atypical autosomal dominant parkinsonism.



Above: An arrow indicates those individuals for whom DNA was available. Individual IV:10 only became clinically affected after the causal variant was identified in the family. Individuals IV:9 and III:14 were selected for whole exome sequencing.

4.3 MATERIALS AND METHODS

4.3.1 Samples

DNA samples were available from 3 affected deceased individuals (III:8, III:10 and IV:9), 1 affected living individual (III:14) and 1 unaffected individual who had passed the age at onset of the disease (III:5). Participants or their named next of kin gave written informed consent. The local ethics board approved the study.

4.3.2 Exome Sequencing

Two individuals (individuals IV:9 and III:14) were selected for WES using Illumina's TruSeq (62Mb) DNA Sample Prep Kit and Exome Enrichment Kit, based on the fact that they were distantly related and sufficient DNA was available for WES.

4.3.3 PCR and Sanger Sequencing

This was performed as described in Materials and Methods (chapter 2).

4.4 RESULTS

4.4.1 Phenotypic Characterization

Clinical features for affected individuals are summarized in table 4.1.

II:1 Medical records are not available for individual II:1, but the family report she developed Parkinsonism in her 30's, and died age 37 years.

II:3 developed parkinsonism aged 65 years according to her family and died 7 years later. No further information is available as medical records are not available.

III:5 This individual was examined by myself at the age of 80 years. She had no neurological symptoms and her clinical examination was normal. She had 4 children who were in their 50's and 60's, none had neurological abnormalities. For the purposes of this study, given that the disease appeared to be highly penetrant in this family, I assumed this individual was unaffected.

III:8 presented aged 49 years with tremulous parkinsonism and cognitive decline. Over 4 years she developed a supranuclear gaze palsy, bulbar palsy and diffuse muscle wasting, weakness in all four limbs and a severe peripheral neuropathy. She had an initial good clinical response to low-dose levodopa. Electrophysiology indicated mild denervation changes distally in the right lower limb and a severe mixed demyelinating/axonal sensory motor peripheral neuropathy. III:8 died aged 53 years.

III:10 presented aged 66 years with bradykinesia which was quickly followed by recurrent falls, memory problems, weight loss, hearing problems, depression and social withdrawal. Reversal of the sleep-wake cycle and frequent episodes of urinary and faecal incontinence were noted. The patient had several 'vacant' episodes, lasting 1-2 minutes, with no recall of the event. Motor symptoms improved with low-dose levodopa. Medical records documented hypophonia, hypomimia, slow horizontal saccades and a fine rest tremor of the hands. Reflexes were pathologically brisk but with flexor plantar responses. MMSE aged 68 years was 24/30. Formal neuropsychometry showed global cognitive dysfunction, with a frontotemporal predominance. An EEG was not performed. The patient died aged 68 years of ischaemic heart disease.

III:14 presented with a more slowly progressive form of the disease. Aged 52 years she presented with 3 stone weight loss. Investigations revealed a renal carcinoma and following a curative resection her weight stabilized. Aged 54, she developed tremulous parkinsonism, and was noted to be apathetic, irritable, depressed and required prompting to carry out activities of daily living. III:14 started shoplifting and developed obsessional tendencies and a preference for sweet foods. She developed daily episodes of urinary and faecal incontinence, for which no secondary cause was identified. Aged 59 years she had several

unprovoked generalised seizures. Treatment with low dose levodopa improved both motor and behavioral symptoms. Examination revealed frontalis over activity, hypomimia, restriction of upgaze, and slow, hypometric saccades particularly in the vertical plane. The jaw jerk was brisk and a left palmomental reflex was present. There was moderate bradykinesia without cogwheeling or rigidity. There was moderate spasticity in all limbs and mild pyramidal weakness in the legs. Reflexes were pathologically brisk. Plantar responses were extensor on the right and flexor on the left. Arm swing was reduced bilaterally with loss of postural reflexes. MRI brain showed small vessel disease but no atrophy. Interictal EEG was normal. Cognitive assessment demonstrated impaired visual memory, executive function, cognitive speed and poor attention, indicating fronto-temporal compromise, more marked in the dominant hemisphere and with subcortical features. Vital capacity was normal.

III:17 is affected with parkinsonism, however no further details are known as he declined to participate in the research study.

IV:6 presented aged 51 with one month history of shortness of breath, nocturnal dyspnoea, recurrent chest infections and weight loss. He required intubation for type 1 respiratory failure and frequent nocturnal dyspnoeic episodes. On examination he had generalised wasting with a mild akinetic-rigid parkinsonian syndrome, vertical gaze palsy, hypometric saccades, pathologically brisk reflexes and frontal disinhibition. Despite nasogastric feeding he continued to lose weight and died of aspiration pneumonia five months after admission.

IV:9 presented aged 43 years with personality change. He had lost confidence at work and was prone to aggressive outbursts, emotional lability and reported motor and cognitive slowing. Over the ensuing year he developed tremulous parkinsonism. Lower limb reflexes were diminished. He responded well to low-dose levodopa. He had severe recurrent chest infections and died aged 46 years of aspiration pneumonia. An MRI brain showed an incidental venous abnormality but no atrophy.

IV:10 developed rapidly progressive symptoms including dramatic weight loss, memory problems and tremor aged 56 years. Within a few months she developed recurrent apnoeic episodes. Examination revealed a fine rest and postural tremor of the upper limbs with cogwheeling. Reflexes were pathologically brisk with equivocal plantar responses, brisk jaw jerk and grasp reflex. Dysdiadochokinesis was present in the upper limbs. Gait was narrow based. MMSE was 22/30.

Table 4.1: Summary of clinical features of affected family members (see pedigree figure 4.1)

Individual	II:1	II:3	III:3	III:8	III:10	III:14	III:17	IV:6	IV:9	IV:10
Age at onset	~35	~65	n/k	49	66	52	nk	51	43	56
Duration (yrs)	7	7	n/k	4	2		nk	0.4	3	
Age at death	~42	72	53	53	68	alive	alive	51	46	alive
Cause of death	nk	nk	Respiratory failure	Bronchial pneumonia	Ischaemic heart disease	n/a	n/a	Aspiration pneumonia	Aspiration pneumonia	n/a
Presenting clinical symptom	Parkinsonism	Parkinson-ism	Parkinsonism	Parkinsonism	Parkinson-ism	bvFTD	Parkinson-ism	Respiratory failure	Personality change	Parkinsonism
Parkinsonism	+	+	+	+	+	+	+	+	+	+
Hypo-ventilation	nk	nk	+	+	-	-	nk	+	+	+
Psychiatric Symptoms	nk	nk	nk	-	Depression, social withdrawal	Apathy, depression, obsessions	nk	Frontal disinhibition Apathy	Frontal disinhibition Aggression	-
Weight loss	nk	nk	nk	-	+	+	nk	+	-	+
Response to Levodopa	nk	nk	nk	+	+	+	nk	Not tried	+	+
Cognitive decline	nk	nk	nk	+	Fronto-temporal compromise	Fronto-temporal compromise	nk	-	+	+
Spasticity	nk	nk	nk	-	-	+	nk	-	-	-
Hyperreflexia	nk	nk	nk	-	+	+	nk	+	-	+

Extensor plantar responses	nk	nk	nk	-	-	+	nk	+	-	Equivocal
Eye movement abnormality	nk	nk	nk	Supranuclear gaze palsy	Slow horizontal saccades	Restricted upgaze, hypometric saccades	nk	Supranuclear gaze palsy, hypometric saccades	Normal	Normal
Bulbar palsy	nk	nk	nk	+	-	-	nk	-	-	-
Peripheral neuropathy	nk	nk	nk	Severe demyelinating & axonal sensory motor neuropathy	-	-	nk	-	Possibly – LL reflexes diminished	-
Urinary and faecal incontinence	nk	nk	nk	-	+	+	nk	-	-	-
Muscle wasting	nk	nk	nk	+	-	-	nk	+	-	-
Muscle weakness	nk	nk	nk	+	-	Pyramidal weakness LL	nk	-	-	-
Seizures	nk	nk	nk	-	Possible seizures	Several GTC seizures	nk	-	-	-
MRI Brain	nk	nk	nk	nk	Generalised atrophy / SVD	SVD	nk	Normal	Normal	Normal

Key: bvFTD=behavioural variant frontotemporal dementia, SVD=small vessel disease, nk=not known, n/a not applicable, += present, -= absent, GTC generalised tonic-clonic, LL=lower limb

4.4.2 Pathological examination in an affected family member

A brain post mortem was performed on III:8 in 1991. It showed generalised atrophy, with slight preponderance of the temporal lobes and depigmentation of the substantia nigra with occasional neurons containing lewy bodies. Studies of the median nerve showed axonopathy.

4.4.3 Candidate gene screening

Mutations in *SNCA*, *LRRK2*, *VPS35*, *MAPT*, *GRN* and repeat expansions causing spinocerebellar ataxia and Huntington's disease had been excluded in III:14.

4.4.4 Whole exome sequencing

Alignment metrics on both individuals WES data showed good coverage in both individuals (Table 4.2).

Table 4.2: summary of metrics of WES in individual III:16 and IV:9

	III:14	IV:9
Total number of reads	136,078,746	127,297,302
Non-duplicated reads	106,855,818	97,105,484
Reads aligned to target	92,043,322	83,578,035
Mean Target Coverage	88.123837	79.879976
% Target covered by 2 reads	96	96
% Target covered by 10 reads	92	92
% Target covered by 20 reads	87	86
Total number of variants	22,308	22,566

4.4.5 Variant filtering

I assumed that both III:14 and IV:9 had the same genetic cause of disease and that the mode of inheritance was autosomal dominant. Based on these assumptions

synonymous and homozygous variants were excluded. Next, variants were removed if they were present in publically available databases (National Heart, Lung, and Blood Institute (NHLBI) Exome Variant Server (EVS), 1000 genomes and dbSNP) at a frequency above 0.1%. Variants were also filtered out if they were present in an in-house set of exomes (from an unrelated non-neurological disease) more than 3 times. After filtering, 12 variants remained (Table 4.3).

Table 4.3 Rare shared variants between III:14 and IV:9 after variant filtering.

Gene	Full Gene Name	Variant Annotation	Variant Type	Frequency in public databases	Sift Score	Poly phen 2 score	GERP Score	Comment
ARHGEF19	Rho guanine nucleotide exchange factor 19	*ARHGEF19* ENST00000270747.3 exon10 c.A1616G p.D539G	nonsynonymous SNV	Novel	0.16	0.99431	3.07	Acts as guanine nucleotide exchange factor for RhoA GTPase
DCTN1	dynactin 1	*DCTN1* ENST00000361874.3 exon2 c.A202G p.K68E	nonsynonymous SNV	Novel	0	0.95502	3.97	Required for the cytoplasmic dynein-driven retrograde movement of vesicles and organelles along microtubules. Dynein-dynactin interaction is a key component of the mechanism of axonal transport of vesicles and organelles. *DCTN1* mutations cause Perry Syndrome, a rare autosomal dominant form of parkinsonism

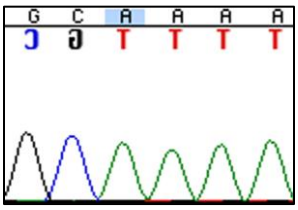
VARS	valyl-tRNA synthetase 2, mitochondrial	*VARS* ENST00000375663.3 exon2 c.C136G p.P46A	nonsynonymous SNV	Novel	0.03	0.99708	3.71	Aminoacyl-tRNA synthetases catalyze the aminoacylation of tRNA by their cognate amino acid
MOB3B	MOB kinase activator 3B	*MOB3B* ENST00000262244.5 exon2 c.C253T p.R85W	nonsynonymous SNV	Novel	0.01	0.97534	1.84	May regulate the activity of kinases
ZNF485	zinc finger protein 485	*ZNF485* ENST00000374437.2 exon3 c.669_670insT p.I223fs	frameshift insertion	Novel	N/A	N/A	N/A	May be involved in transcriptional regulation
VWA2	von Willebrand factor A domain containing 2	*VWA2* ENST00000298715.4 exon12 c.G1711A p.G571S	nonsynonymous SNV	Novel	0.02	0.99945	4.96	The encoded protein is localized to the extracellular matrix and may serve as a structural component in basement membranes or in anchoring structures on scaffolds of collagen VII or fibrillin.

C10orf82	chromosome 10 open reading frame 82	*C10orf82* ENST00000369210.3 exon3 c.A247C p.N83H	nonsynonymous SNV	Novel	0.55	0.01972	-2.12	Function unknown. Poorly conserved therefore unlikely to be candidate variant
MAP1A	microtubule-associated protein 1A	*MAP1A* ENST00000300231.5 exon4 c.C4257G p.D1419E	nonsynonymous SNV	Novel	0.28	0.19407	-2.5	Structural protein involved in the filamentous cross-bridging between microtubules and other skeletal elements. Poorly conserved therefore unlikely to be candidate variant
LRRC8E	leucine rich repeat containing 8 family, member E	*LRRC8E* ENST00000306708.5 exon3 c.2113delG p.A705fs	frameshift deletion	Novel	N/A	N/A	N/A	This gene encodes a member of a small, conserved family of proteins with similar structure, including a string of extracellular leucine-rich repeats. A related protein was shown to be involved in B-cell development.
EML2	triggering receptor expressed on myeloid cells-like 2	*EML2* ENST00000245925.2 exon9 c.C803T p.T268M	nonsynonymous SNV	Novel	0	0.99859	3.74	May modify the assembly dynamics of microtubules, such that microtubules are slightly longer, but more dynamic

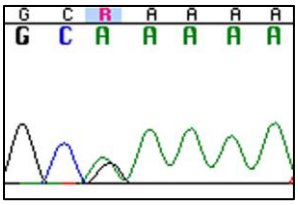
CEP250	centrosomal protein 250kDa	CEP250 ENST00000342580.4 exon34 c.G7001A p.S2334N	nonsynonymous SNV	Novel	0.29	0.98298	2.71	Probably plays an important role in centrosome cohesion during interphase
TSHZ2	teashirt zinc finger homeobox 2	TSHZ2 ENST00000329613.5 exon2 c.73_75del p.25_25del	nonframeshift deletion	Novel	N/A	N/A	N/A	Probable transcriptional regulator involved in developmental processes. May act as a transcriptional repressor

The variant in the *DCTN1* gene (Ensemble transcript ID ENST000003618750 c.A202G p.K68E), was of immediate interest as mutations in this gene are an extremely rare cause of atypical autosomal dominant parkinsonism. The variant occurred at a residue that was highly conserved (GERP score 3.97), and was highly likely to be deleterious (SIFT score 0; Polyphen2 score 0.955). The mutation was confirmed by Sanger sequencing in all affected individuals for whom DNA was available (III:8, III:10, III:14, IV:9 and IV:10). It was absent in one unaffected family member who is 82 years old (III:5) (figure 4.2).

Figure 4.2: Chromatograms showing the novel heterozygous *DCTN1* mutation



Left: wild type sequence (III:5)



Left: chromatogram for individual III:14 showing c.A151G:pK68E in exon 2 of *DCTN1*. Individuals III:8, III:10, III:14, and IV:9 were also found to have the same mutation.

Subsequently, a further family member (IV:10) developed symptoms consistent with the phenotype in the family. Sanger sequencing also confirmed this individual had the *DCTN1* p.K68E variant.

4.4.6 Re-examination of the brain histopathology for individual III:8

At this point in the study it seemed likely that the variant in *DCTN1* was causal. *DCTN1* mutations are known to cause a rare autosomal dominant Parkinson syndrome with clinical features that overlap with those seen in this family. Secondly, the putative causal variant segregates with disease in family members

for whom DNA was available. However in the absence of genetic linkage data, further evidence would be needed to determine whether this variant was pathogenic. Brain tissue in Perry Syndrome has been described in association with positive TDP-43 immunohistochemical staining.[172] Brain tissue of III:8 was re-examined in 2013 by Prof Tamas Revesz. This demonstrated abnormal TDP-43 immunohistochemistry, abnormal neurites and axonal swellings in the globus pallidus and midbrain tegmentum. There were no TDP-43 positive neuronal cytoplasmic inclusions in the substantia nigra, however, tissue was not available for review from all of the nigra. It is likely that the cell loss in the nigra was very severe, which may explain the absence of TDP-43 positive neuronal cytoplasmic inclusions in the nigra. No neocortical plaques, tau or alpha-synuclein staining was detected. Overall, these histopathological findings were felt to be highly supportive of a diagnosis of Perry Syndrome.

4.5 DISCUSSION

Mutations in *DCTN1* cause Perry Syndrome,[160] which has traditionally been described as an autosomal dominant, rapidly progressive form of parkinsonism associated with depression, weight loss, and hypoventilation. It is an extremely rare form of familial parkinsonism with only six mutations (p.G67D, p.G71A/E/R, p.T72P, and p.Q74P) reported in 12 families worldwide.[161-171] All mutations are located within the p150^{glued} highly conserved N-terminal cytoskeleton-associated protein, glycine-rich (CAP-Gly) domain within a hydrophobic pocket upon the surface. p150^{glued} is the major subunit of the dynactin protein complex, which plays a crucial role in retrograde axonal and cytoplasmic transport of various cargoes. The CAP-Gly domain is critical for microtubule binding and in *vitro assays* have shown mutations result in a reduced affinity of dynactin for microtubules.[160 287, 288]

In this family, WES of two distantly affected family members, without linkage analysis, has identified a novel missense mutation (p.K68E) in *DCTN1* as the cause of disease. There is strong evidence to support its causality. The mutation is absent from publically available databases, it is located in the CAP-Gly domain, where all reported mutations, which cause Perry Syndrome occur. The mutation segregates

with disease in the family, it is present in 5 affected family members and is absent from 1 unaffected family member, who is past the age of onset of the disease. The mutation is predicted to be highly damaging using in-silico prediction software (SIFT score 0,[289] Polyphen2 score 0.955[289]) and occurs at a highly conserved residue (GERP score of 3.97). Furthermore, the diagnosis of Perry Syndrome is supported by the detection of TDP-43 positive inclusions, predominantly in a pallidonigral distribution and nigral cell loss in an affected individual.

For individuals whom medical records were available, the average age at onset was 52 years (range 35-66 years) and average duration of disease was 3 years 9 months (range 5 months-7 years). To-date, clinical features of Perry Syndrome are relatively consistent between families. Notable exceptions include a family who did not have weight loss or hypoventilation[290] and a patient with a vertical gaze palsy and features of behavioural variant frontotemporal dementia (bvFTD).[162] Clinical and pathological criteria have been proposed for the diagnosis of Perry Syndrome. A 'definite' diagnosis requires the presence of five cardinal features (autosomal dominant inheritance, hypoventilation, parkinsonism, weight loss and psychiatric features) and characteristic pathological findings of severe substantia nigra neuronal loss and few or no Lewy bodies.[290]

The commonest presenting symptom in this family was parkinsonism, however atypical presentations included bvFTD, personality change and fulminant type 1 respiratory failure. Only one individual had all five cardinal features required for a definite diagnosis of Perry syndrome (autosomal dominant, hypoventilation, parkinsonism, weight loss and psychiatric features), suggesting these criteria may not identify all patients with Perry syndrome.

Unusually, several individuals had upper or lower motor neuron signs or frontotemporal compromise, features more commonly seen in FTD/Amyotrophic lateral sclerosis. Upper motor neuron signs observed in affected individuals included pathologically brisk reflexes, extensor plantar responses and pyramidal pattern of weakness. One patient had a bulbar palsy. A frontotemporal pattern of cognitive impairment and frontal disinhibition was prominent in several individuals. Motor neuron involvement has not been described in association with

Perry syndrome, however the p.G59S *DCTN1* mutation, which is buried within the center of the CAP-Gly domain, has been identified in a family with distal hereditary motor neuropathy with vocal cord paresis.[291] TDP-43 pathology is common to Perry Syndrome, amyotrophic lateral sclerosis and some forms of frontotemporal lobar degeneration, indicating pathological overlap between these conditions. Four individuals had eye movement abnormalities (slow saccadic eye movements, restricted vertical gaze and supranuclear gaze palsy), confirming this as a clinical manifestation of Perry Syndrome. Additionally, incontinence, seizures and peripheral neuropathy may be part of the extended phenotype. Epileptiform discharges on EEG and neuropathy have previously been described in a single patient each.[167, 292]

Perry Syndrome is probably under diagnosed, as it is rarely tested for. It is likely that the phenotype is broader than that described in the literature to date and will continue to widen with identification of additional families. In view of the fact that this family widen the phenotypic spectrum of disease associated with mutations in *DCTN1*, we suggest genetic testing for Perry Syndrome be considered in any patient with autosomal dominant parkinsonism and one other of the cardinal features as in the proposed diagnostic criteria[290] (hypoventilation, weight loss and psychiatric features), keeping in mind that a eye movement abnormalities, neuropathy, upper or lower motor neuron features, frontotemporal cognitive impairment, urinary and faecal incontinence, may be prominent features of the clinical presentation.

The rapid identification of a novel *DCTN1* mutation in this family demonstrates the utility of next-generation sequencing technologies applied to families with undiagnosed neurological disorders, in whom linkage analysis is not possible. This is a common scenario facing clinicians, particularly those families with late-onset neurodegenerative disorders, where DNA samples may not be available for the preceding generations.

CHAPTER 5. WHOLE EXOME SEQUENCING IN AUTOSOMAL RECESSIVE PARKINSONISM DISORDERS

5.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

Unless otherwise indicated I performed autozygosity mapping and analysed the WES data. I performed confirmatory Sanger sequencing and segregation analysis. I examined all family members apart from family 2, who were examined by Dr. Uday Muthane and Prof Kailash Bhatia.

5.2 BACKGROUND

Large scale screening studies have shown that of patients with autosomal recessive, early-onset (<45 years) typical parkinsonism, ~50% of patients will have mutations in *Parkin*,[177] ~1-8% will have mutations in *PINK1*,[293] and ~1% will have mutations in *DJ-1*.[294] This suggests that there are further autosomal recessive Parkinson's disease genes yet to be identified. The NHNN Neurogenetics database contains a large number of samples from patients with suspected genetic forms of Parkinson's disease. A search of this database revealed several families with

suspected autosomal recessive parkinsonism in whom the genetic cause had not been identified.

Where parental consanguinity is present in the context of a suspected autosomal recessive disorder, the initial assumption is that the disease is likely to be caused by a homozygous variant inherited from both parents (identical by descent) and this variant resides within a large stretch of homozygosity. Such families are suitable for autozygosity mapping, which will reduce the proportion of the genome which needs to be considered. Homozygous variants which lie in large homozygous regions in affected family members only can therefore be prioritized. The use of autozygosity mapping and WES reduces the amount of variants for follow-up and WES data is only required from a single affected individual, reducing costs.

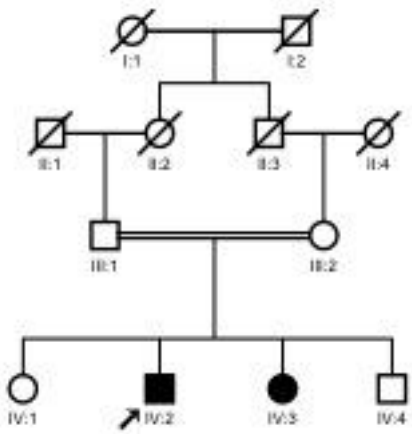
This chapter details the work done on 3 families with suspected autosomal recessive Parkinsonism in whom the genetic cause was not known.

5.3 FAMILY 1: WHOLE EXOME SEQUENCING AND AUTOZYGOSITY MAPPING IN AUTOSOMAL RECESSIVE EARLY-ONSET PARKINSON'S DISEASE

Family 1 is a consanguineous Pakistani family with four offspring, two of whom were affected with young-onset parkinsonism. The parents were first cousins (figure 5.1). In this family a genetic approach of WES and autozygosity mapping resulted in a candidate variant list. However, the causal variant could not be identified as none of the candidate variants were found in a series of early-onset PD. There are many reports in the literature where a candidate variant for a Mendelian disorder has been identified from exome sequencing data and then confirmed as pathogenic after identification of the same variant or a different

variant in the same gene in a 'replication cohort' of individuals with the same disease. Demonstrating that the variant segregates with disease in the family of the patient from the 'replication cohort' provides further evidence for pathogenicity.[38] [15, 36, 37]

Figure 5.1: Pedigree of family 1



5.3.1 MATERIALS AND METHODS

5.3.1.1 Samples

Genomic DNA samples were obtained from peripheral blood lymphocytes, for individuals III:1, III;2, IV:1, IV:2 and IV:3. Individual IV:4 was not available for participation. Participants gave written consent and the study met with the local ethic boards approval.

5.3.1.2 Whole exome sequencing

Individual IV:2 was selected for WES using Illumina Truseq kit. This was performed as described in the Materials and Methods chapter (chapter 2).

5.3.1.3 Autozygosity Mapping

Autozygosity mapping was performed in individuals IV:1, IV:2 and IV:3 as per Materials and Methods chapter 2. Regions of homozygosity larger than 0.5Mb shared by the affected individuals (IV:2 and IV:3) were identified.

5.3.1.4 Variant Filtering

Variants within the regions of homozygosity, only shared by the affected individuals were selected. Heterozygous variants, synonymous variants and those above a frequency of above 1% were filtered out.

5.3.1.5 PCR and Sanger Sequencing

This was performed as described in the materials and methods (chapter 2) for rare variants that remained in the autozygous regions.

5.3.1.6' Replication cohort' of PD patients for screening candidate variants

The exon in which a candidate variant was found was screened in a series of early-onset PD cases (age of onset <45 years), which was collected through specialized Neurogenetics and Movements disorder clinics at the National Hospital of Neurology and Neurosurgery (NHNN). This series totaled 364 patients with early-onset PD, 70% had a family history of PD. Patients in this cohort were known to be negative for the common *SNCA* mutations, the p.G2019S *LRRK2* mutation and mutations in *Parkin* and *PINK1*.

5.3.2 RESULTS

5.3.2.1 Clinical Phenotyping of affected individuals

The index case (IV:2) presented at the age of 31 years with a right sided thumb tremor and right arm bradykinesia. Shortly after a diagnosis of Parkinson's disease was confirmed. He was commenced on Rasagiline, which provided no clinical benefit. Ropinirole was withdrawn due to gambling behavior. He had a good clinical response to Trihexyphenidyl and Sinemet. Examination aged 34 years revealed hypomimia and mild hypophonia. Eye movements were normal. He had a severe tremor at rest in the right arms and leg with more moderate tremor

affecting the left side. He had increased tone in all four limbs and bradykinesia and loss of postural reflexes. He has not subjectively noticed any change in his sense of smell.

IV:2 has been extensively investigated. A DATScan showed bilateral (worse on the left) pre-synaptic dopaminergic deficit. Diagnostic genetic testing was negative for the p.G2019S mutation in *LRRK2*. Sanger sequencing and *MLPA* (for copy number variants) of *Parkin* and *PINK1* was also negative. Furthermore sequencing of *DJ-1* and selected exonic MLPA for this gene was also negative. Screening for Wilson's disease was negative and included formal ophthalmology, 24-hour urinary copper studies and serum ceruloplasmin. Formal neuropsychometry revealed dysfunction in anterior and sub-cortical regions, with some involvement of the dominant temporal lobe.

IV:3 was clinically assessed aged 33 years old. Subjectively she had no symptoms however there was evidence of hypomimia, left-sided bradykinesia and reduced arm-swing on the left. Postural reflexes were retained.

IV:1 was clinically assessed aged 38 years and had an entirely normal neurological examination. III:1 and III:2 were examined, both had normal neurological examinations. IV:4 not available to participate in the study as he lived abroad and was not contactable. He was reported by the family to have no symptoms of PD.

Autosomal recessive inheritance was felt to be the most likely mode of inheritance firstly, because of parental consanguinity and due to the young age of onset of symptoms in IV:2.

5.3.2.2 Autozygosity Mapping

Only the two affected siblings were used to derive the regions of homozygosity as it was felt that IV:1 could yet develop the disease. 23 homozygous regions were identified that were >0.5Mb and shared between IV:2 and IV:3 (table 5.1) encompassing >133Mb.

Table 5.1: Homozygous regions >0.5Mb in individual IV:2 and IV:3.

Chromosome	Position from	Position To	Size (bp)
1	8,168,564	8,960,153	791,589
1	23,145,239	23,902,416	757,177
1	35,268,368	45,035,621	9,767,253
1	46,954,587	55,136,529	8,181,942
1	73,227,348	74,170,796	943,448
1	84,058,458	89,459,475	5,401,017
1	224,021,459	224,794,237	772,778
3	0	9,836,386	9,836,386
3	74,831,395	75,974,251	1,142,856
3	82,602,965	83,558,368	955,403
3	100,864,107	101,712,255	848,148
3	159,686,068	160,398,885	712,817
4	33,628,239	34,910,052	1,281,813
5	36,678,275	37,671,746	993,471
5	136,972,187	138,668,704	1,696,517
7	138,742,981	142,041,961	3,298,980
8	0	1,469,213	1,469,213
10	74,482,123	75,669,319	1,187,196
12	63,158,162	116,977,528	53,819,366
14	62,366,843	72,334,903	9,968,060
15	38,030,147	54,353,991	16,323,844
19	3,278,206	5,841,356	2,563,150
20	13,559,616	14,337,539	777,923

5.3.2.3 Exome Sequencing

WES summary metrics for exome sequencing results for individual IV:2 indicated good coverage (see table 5.2).

Table 5.2: Showing coverage statistics for the WES in individual IV:2

	IV:2 Family 1
Total number of reads	104,750,902
Non-duplicated reads	85,652771
Reads aligned to target	74,665,663
Mean Target Coverage	71
% Target covered by 2 reads	96
% Target covered by 10 reads	91
% Target covered by 20 reads	85
Total number of variants	22,521

5.3.2.4 Variant Filtering

5 novel or rare variants were present in the homozygous regions shared between the two affected individuals (table 5.3). All the variants were confirmed with Sanger-sequencing, which demonstrated that both parents were heterozygous carriers, and that both siblings with PD were homozygous for the variants, but the unaffected sibling was not.

In order to try and investigate these candidate variants further I Sanger sequenced the exon in which the variant occurred, in a 'replication cohort' of early-onset PD patients, hoping to find another patient with a rare/novel variant in that exon also. I prioritised the variants so that only the best candidate variants would be tested in the replication cohort, as discussed in chapter 1. All of the variants occurred at highly conserved amino acids, so it was not possibly to exclude variants on this basis. Instead I focused on the functional class of variant, putative gene function and expression patterns. I did not rely heavily on the in-silico prediction of deleteriousness as this is not highly specific or sensitive.[30, 31]

Table 5.3: Variants within the regions of homozygosity shared between IV:2 and IV:3

Gene	Full Gene Name	Variant annotation	Variant Type	Frequency in publically available databases	Sift score	Polyphen 2 score	GERP score	Comment
ECHDC2	enoyl CoA hydratase domain containing 2	*ECHDC2* ENST00000358358.5 exon5 c.C398Tp.T133M	nonsynonymous SNV	Novel	0.02	0.999	2.31	Fatty acid metabolic process. Likely located in the mitochondria. Widely expressed. Most highly expressed in the liver. Expressed in brain, most highly in the cerebellum
SUMF1	sulfatase modifying factor 1	*SUMF1* ENST00000272902.5 exon2 c.C287T p.A96V	nonsynonymous SNV	Novel	0	0.209	4.59	Encodes an enzyme that catalyzes the hydrolysis of sulfate esters by oxidizing a cysteine residue in the substrate sulfatase to an active site 3-oxoalanine residue. Mutations cause multiple sulfatase deficiency, an infantile onset autosomal recessive fatal lysosomal storage disorder characterised by coarsened facial features, deafness, icthyosis, organomegaly and skeletal abnormalities.

HAL	histidine ammonia-lyase	*HAL* ENST00000541929.1 exon12 c.G469A p.A157T	nonsynonymous SNV	Novel	0	0.997	4.06	Histidine ammonia-lyase is a cytosolic enzyme catalyzing the first reaction in histidine catabolism, the nonoxidative deamination of L-histidine to trans-urocanic acid. Histidine ammonia-lyase defects cause histidinemia, which is characterized by increased histidine and histamine and decreased urocanic acid in body fluids. Although histidinaemia was reported initially in association with speech difficulties and/or mental retardation, more recently it is suggested that histidinemia does not cause central nervous abnormalities or other forms of disease. Low brain expression profile.
THBS1	thrombospondin 1	*THBS1* ENST00000397593.3 exon3 c.C166A p.H56N	nonsynonymous SNV	Novel	0.02	0.92	5.21	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. Binds heparin. May play a role in dentinogenesis and/or maintenance of dentin and dental pulp, platelet aggregation, angiogenesis, and tumorigenesis. Highly expressed in blood, bone and smooth muscle. Minimal brain expression.

TMBIM4	Transmembrane BAX Inhibitor Motif Containing 4	*TMBIM4* ENST00000358230.3 exon2 c.113delT p.V38fs	frameshift deletion	Novel	NA	NA	NA	Anti-apoptotic protein which can inhibit apoptosis induced by intrinsic and extrinsic apoptotic stimuli. Can modulate both capacitative Ca2+ entry and inositol 1,4,5-trisphosphate (IP3)-mediated Ca2+ release. Widely expressed.

THBS1, Thrombospondin-1

Thrombospondin-1 is a matricellular, calcium-binding protein that participates in cellular responses to growth factors, cytokines and injury. It regulates cell proliferation, migration and apoptosis in a variety of physiological and pathological settings, including wound healing, inflammation, angiogenesis and neoplasia. These processes involve the formation of multi-protein complexes on the cell surface and the clustering of receptors that initiate signal transduction. Thrombospondin-1 dictates the composition of these multi-protein complexes through specific interactions with growth factors, cytokines, other matrix components and membrane proteins.[295] In the central nervous system *THBS1* secreted by astrocytes promotes synaptogenesis and neurite outgrowth.[296, 297] Thrombospondin-1 is expressed in many tissues during embryonic development but has limited expression in the healthy adult. *THBS1* is the most abundant protein in alpha granules of platelets. Expression in other cell types is induced by wounding, during tissue remodeling, in atherosclerotic lesions, rheumatoid synovium, glomerulosclerosis, and in stroma of many tumours. *THBS1* null mice have impaired wound repair, increased retinal angiogenesis, and are hyper-responsive to several inflammatory stimuli. *THBS1* has been implicated in myocardial infarction,[298] stroke[299] and tumours invasive behavior, tumour neovascularization and metastasis.[300, 301] Both SIFT and Polyphen2 scores predict the variant in *THBS1* to be damaging.

The variant in *THBS1* was prioritised for screening in the 'replication cohort' of PD patients because of its role in synaptogenesis in the central nervous system. However, factors suggesting that it may not be causal include the fact it is not highly expressed in the adult human brain.

HAL, Histidine ammonia-lyase

Histidine ammonia-lyase is a cytosolic enzyme catalyzing the first reaction in histidine catabolism, the nonoxidative deamination of L-histidine to trans-urocanic acid. Histidine is an α-amino acid with a imidazole functional group. It is one of the amino acids that are precursors to proteins. Histidine ammonia-lyase enzyme is expressed in a tissue-specific manner in the liver and epidermis in the rat.[302] Histidinaemia results from histidine ammonia-lyase deficiency.[303] It is inherited as

an autosomal recessive disorder that affects the liver,[304] with several mutations identified. It was thought to be associated with mental retardation, infantile spasms, epilepsy, speech defects, and motor incoordination.[305] However, the long-term outcome of patients diagnosed by newborn screening in the north-west of England between 1966 and 1990 showed that in most cases the disorder is benign: IQ did not correlate with plasma histidine at diagnosis or with the mean plasma histidine throughout life.[306] Growth has been normal in all patients. In contrast to some earlier studies, there was no excess of clinical symptoms. It has been concluded that histidinaemia in the typical form (due to impairment of histidine ammonia-lyase activity) is not a disease in humans, because retrospective and prospective studies together indicate that the prevalence of disadaptive phenotypes (e.g., impaired intellectual or speech development, seizures, behavioural or learning disabilities) in histidinaemia population is not higher than the frequency of these functional disorders in the nonhistidinaemia population. The possibility remains that histidinaemia is a risk factor for development of an unfavourable central nervous system phenotype, in particular in individuals under specific circumstances. Both SIFT and Polyphen2 scores predict the variant in *HAL* to be damaging.

I did not prioritise the variant in *HAL* for screening in the replication cohort of early-onset PD as it is not highly expressed in human brain and loss of activity has been investigated in epidemiological studies and has not been shown to be associated with a neurological phenotype.

ECHDC2, enoyl CoA hydratase domain containing 2

Little is known about the function of *ECHDC2.* It is thought to be a mitochondrion protein and is known to be highly expressed in the liver. It is thought to be involved in fatty acid and lipid metabolism. Enoyl-CoA hydratases are enzymes that hydrate the double bond between the second and third carbons on acyl-CoA. They are essential in metabolizing fatty acids to produce acetyl CoA and energy. Many inborn errors of metabolism are due to defects in mitochondrial fatty acid oxidation (e.g., Medium-chain acyl-coenzyme A dehydrogenase deficiency MCAD deficiency) and very long chain fatty acids (e.g., Refsums disease) and many of

these disorders have neurological problems as part of their phenotype. Both SIFT and Polyphen2 scores predict the variant in *ECHDC2* to be damaging.

The variant in *ECHDC2* was prioritised as it is expressed in brain, involved in processes in which defects can result in human disease. Of note it is a mitochondrial protein and four autosomal recessive PD disorders caused by mutations in *PINK1*, *PARKIN*, *DJ-1* and *FBXO7* are linked to mitochondrial maintenance.

SUMF1, sulfatase modifying factor 1

SUMF1 encodes the enzyme sulfatase modifying factor 1, which is highly conserved among eukaryotes and localised in the lumen of the endoplasmic reticulum. The profound deficiency but not complete absence of all known sulfatase activities of sulfatase modifying factor 1 results in 'multiple sulfatase deficiency' (MSD), which is an autosomal recessive lysosomal disorder with storage of sulphated lipids and sulphated glycans and a profound deficiency of all lysosomal and nonlysosomal sulfatases.[307-309] Deletions, nonsense and missense mutations have been described in association with MSD. The clinical symptoms of MSD represent a composite of the symptoms found in disorders caused by deficiency of single sulfatases such as metachromatic leukodystrophy, mucopolysaccharidosis, chondrodysplasia punctate, hydrocephalus, ichthyosis, neurological deterioration and developmental delay. The leukodystrophy-like (deficiency of arylsulfatase A) and mucopolysaccharidosis-like (deficiency of glycosaminoglycan degrading sulfatases) features prevail in MSD. The variant p.A96V has not been reported in association with MSD previously, there are no mutations in adjacent codons which cause MSD. Mutations in *SUMF1* causing MSD are spread throughout the length of the gene.

SUMF1 was also prioritised, as mutations in this gene are known to cause a lysosomal storage disorder: multiple sulfatase deficiency. Lysosomal storage disorders are linked to the pathogenesis of PD as either compound heterozygous or homozygous mutations in *GBA* cause the storage disorder Gaucher's disease in which parkinsonism can be a feature and heterozygous mutations in *GBA* are a risk

factor for developing PD. SIFT but not Polyphen2 predicts the variant in *SUMF1* to be damaging.

TMBIM4, Transmembrane BAX Inhibitor Motif Containing 4

TMBIM4 is a novel regulator of cell death that is highly conserved in eukaryotes and present in some poxviruses. Its predominant localization is in the Golgi and it is known to be able to suppress apoptosis.[310] *TMBIM's* are highly conserved in a broad range of organisms including human, orangutan, dog, mouse, rat, *Xenopus laevis* and zebrafish, and related proteins are present in *Drosophila* and *Arabidopsis*. Human TMBIM4 is expressed ubiquitously in human tissues. Stable expression of human TMBIM4 is associated with suppression of cell death induced by a broad variety of intrinsic and extrinsic apoptotic stimuli. Conversely, knockdown of *TMBIM4* in tissue culture cells by siRNA resulted in cell death. Alterations in the finely tuned intracellular Ca^{2+} homeostasis and compartmentalization contribute to the induction of apoptosis. The switch from the control of physiological functions to the involvement in apopotosis most likely entails changes in the tightly regulated spatiotemporal Ca^{2+} signaling pattern affecting cytosolic effector proteins and effort organelles.[311] Ca^{2+} signaling between storage organelles and mitochondria plays an important role in sensitizing cells to apoptosis.[312] TMBIM4 has been shown to modulate both capacitive Ca^{2+} entry and inositol-1,4,5-trisphosphate-mediated Ca^{2+} release, suggesting it has an important role in the cross-talk between the intracellular Ca^{2}+ stores, the cytosol and the mitochondria, and this may explain how it plays a decisive role in regulating cell death by apoptosis.

TMBIM4 was prioritised as it is a frameshift deletion and therefore predicted to result in loss of protein function and this type of mutation is enriched among known disease-causing variation.[22] *TMBIM4* function in anti-apoptosis and calcium signaling in apoptosis and could therefore link in with known functions of *PARKIN* and *PINK1* in mitochondrial biogenesis and destruction by autophagy.

5.3.2.5 Sanger sequencing of candidate variants in a replication cohort of EOPD

THBS1, SUMF1, ECHDC2, TMBIM4 were selected as potential disease causing candidate variants. The exon in which the candidate variant in each of these genes was Sanger sequenced in a cohort of 364 early-onset PD patients. However, no variants were found in the exon in which they occurred in the cohort of early-onset PD patients (table 5.4).

Table 5.4: Sanger Sequencing of potential causal variants in the replication cohort

Gene	Transcript and exon Sanger sequenced	Sanger sequencing result in 364 EOPD patients
THBS1	ENST00000397593 exon 3	No variants in this gene were found
SUMF1	ENST00000272902 exon 2	No variants in this gene were found
ECHDC2	ENST00000358358 Exon 5	No variants in this gene were found
TMBIM4	ENST00000358230 Exon 2	No variants in this gene were found

5.3.3 DISCUSSION

In this small consanguineous family with two affected individuals with early onset PD, WES and autozygosity mapping was used to try and identify the causal variant. Five homozygous, rare variants occurred at codons that are highly conserved. Of these, the exon in which 4 variants occurred was selected for Sanger sequencing in a cohort of early-onset PD patients (n=364). No variants were found in the exon in which the candidate variants occurred in this cohort of patients. Consequently, it is not possible to assign causality to any of the variants that remain in the candidate list of variants.

There are two possibilities regarding the candidate variants. Firstly, the causal variant may be one of the candidate variants listed in table 5.3, however this form

of early-onset PD is very rare and a much larger sequencing project needs to be performed in order to identify a second patient with a mutation in that gene. Such a sequencing project might include many more patients and sequence all exons of all the candidate genes. Due to the cost of Sanger-sequencing, such a project would be better performed by a targeted next generation sequencing panel of candidate genes. Data sharing between PD research groups of WES data from EOPD patients may also be of help in determining if any of the candidate variants are pathogenic, particularly if co-segregation can be shown in another multiplex kindred. Secondly, the causal variant may not be present in the candidate list of disease causing variants due to the technical or analytic limitations of WES:

Technical Limitations:

1. Part or all, of the causative gene is not in the target definition (for example, it is not a known gene, or there is a failure in the bait design). The probes in sequence capture methods are designed based on the sequence information from gene annotation databases such as the consensus coding sequence (CCDS) database and Refseq database; therefore, unknown or yet-to-annotate exons cannot be captured. There may also be a failure in bait design so that an exonic region is not captured. Selectively sequencing the exome- which is, to our knowledge, the most likely region of the genome to contain pathogenic mutations – also excludes noncoding regions. The contribution of mutations in non-coding regions to Mendelian disease has yet to be determined. MicroRNAs, promotors and ultra-conserved elements may be associated with disease, however these are not fully covered in current exome capture kits.
2. Inadequate coverage of the region that contains a causal variant. A certain minimum depth of coverage is required for sufficient accuracy of variant detection; that is, positions or regions in the genome of the individual that are different from the reference human genome sequence. Regions with repetitive sequences are more poorly characterized, as repetitive sequences may have prevented inclusion of a probe, or the reads originating from these regions cannot easily be mapped to a single position in the reference genome. Additionally probes may be poorly performing in GC rich regions. The causal variant may be covered but not accurately called (for example, in the presence of a small indel)

3. The causal variant is covered but not accurately called, for example, in the presence of a small but complex indel. WES is unable to detect structural variants or chromosomal rearrangements, which are believed to be important for Mendelian disorders.

Analytical Limitations:

4. The mode of inheritance was assumed to be autosomal recessive in this kindred, as there was an early age at onset of symptoms and parental consanguinity. It is possible that the mode of inheritance is autosomal dominant with reduced penetrance in either III:1 or III:2.

5. Pathogenic mutations may be present in publically available 'control' databases, and may therefore be erroneously filtered out. Currently, more than 17 million SNPs in the human genome have been documented in dbSNP with a false positive rate of ~15-17%.[59] Using an appropriate MAF for the mode of inheritance and the curating of databases such as 1000 genomes and Washington exome server will help to reduce this type of error. Alternatively, the causal variant could be synonymous and have been excluded in the variant filtering process.

5.4 FAMILY 2: WHOLE EXOME SEQUENCING AND AUTOZYGOSITY MAPPING IN A SINGLETON CASE WITH COMPLICATED PARKINSONISM

5.4.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

I performed autozygosity mapping, WES analysis and confirmatory Sanger sequencing.

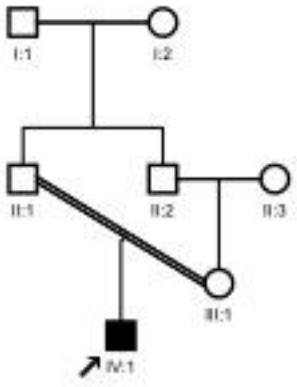
5.4.2 BACKGROUND

This patient was brought to my attention by Prof. Kailash Bhatia who is a movement disorders specialist at the National hospital for Neurology and

Neurosurgery and Dr. Uday Muthane who is a movement disorders specialist in Bangalore, India. The pedigree of the family is shown in figure 5.2. Parental consanguinity suggested a genetic aetiology with an autosomal recessive mode of inheritance. However, the patient had been treated with antipsychotic drugs, a recognized side effect of which is drug-induced parkinsonism, leading to diagnostic uncertainty.

In family 2, an approach of WES and autozygosity mapping in a single individual identified a novel homozygous missense mutation: c.T2525C p.L842P in the hydrolase domain of the *ATP13A2* gene confirming the diagnosis of Kufor-Rakeb Syndrome.

Figure 5.2: Pedigree of family 2



5.4.3 MATERIALS AND METHODS

5.4.3.1 Samples

The DNA samples for the index case (IV:1) and the parents (II:1 and III:1) were collected by Dr. Uday Muthane at the Movement disorders Clinic, Bangalore, India. All participants gave written informed consent. The study was approved by the local ethics committee.

5.4.3.2 Phenotypic characterization

The index case (Figure 1: IV:1) is a 34 year old Indian gentleman, the only child born to consanguineous parents (uncle-niece relationship) with a juvenile

complicated parkinsonian syndrome. His first symptoms were at the age of 13 years with falls and dysarthria. A Neurologist formally assessed him at the age of 22; he was found to have signs of parkinsonism and was commenced on half a tablet twice a day of controlled release Carbidopa/Levodopa, 250 mg (Syndopa® CR). Four years later he developed L-Dopa induced dyskinesias. At this point Syndopa® was stopped and Ropinirole was commenced. One year later, aged 27 years, he developed acute psychosis with both visual and auditory hallucinations. Initially, he was treated with Olanzapine and later Risperidone and Levosulpiride. However, these led to a worsening of his parkinsonism. Recently the patient has been restarted on Syndopa® and the aforementioned antipsychotic medications have been discontinued and substituted with Quetiapine, which has resulted in some improvement in the patients parkinsonism. The patient is now 34 years of age, he is wheelchair bound, has bowel and bladder incontinence, cognitive impairment and dependent for all his activities of daily living. There is no family history of parkinsonism.

Examination of the patient's parents was entirely normal. Clinical examination of IV:1 revealed normal pursuit eye movements but impaired saccadic eye movements with a supranuclear gaze palsy. His mouth was partially open due to jaw opening dystonia and his speech was dysarthric. Tongue movements were normal. He had intermittent myoclonic jerks at rest in both hands and an intention tremor in the left hand. Finger-nose testing was normal. During rapid alternate movements of the hands, the jaw opening dystonia and laterocollis and eversion of the lower lip became more apparent. There was asymmetrical bradykinesia present in all four limbs. There was mild weakness of left hand grip, plantar and dorsiflexor weakness of both feet. There was spasticity in the lower limbs. Reflexes were pathologically brisk in the upper limbs and at the knee but plantar responses were flexor. Bedside examination revealed significant cognitive impairment.

MRI brain age 32 years showed pontine, midbrain and caudate atrophy with no cerebellar atrophy. No iron deposition was seen on T2* images. A peripheral blood smear for acanthocytes, axillary skin biopsy for adult polyglucosan body disease and neuronal ceroid lipofuscinosis, nerve conduction studies and genetic testing

for Huntington's disease and Dentatorubro Pallido-Luysian Atrophy (DRPLA) were all normal or negative.

There was diagnostic uncertainty in this case as it was initially suspected that the patient had drug-induced parkinsonism. However, parental consanguinity suggested the possibility of autosomal recessive PD. WES and autozygosity mapping was used to investigate the possibility of an autosomal recessive juvenile parkinsonian syndrome.

5.4.4 METHODS

5.4.4.1 Autozygosity Mapping

Autozygosity mapping was performed as described in chapter 2 identifying homozygous regions in IV:1 (figure 5.2) larger than 0.5Mb.

5.4.4.2 Whole Exome Sequencing

WES was performed in IV:1 (figure 5.2) as described in the Materials and methods chapter (chapter 2).

5.4.4.3 Variant Filtering

Only variants within regions of homozygosity were used for filtering. Synonymous variants and any variant present at a global minor allele frequency of more than 1% in a range of publically available databases of sequence variation (1000 Genomes, Complete Genomic 69 Database and NHLBI Exome Sequencing Project database) was filtered out, as well as those found in our in-house exomes from individuals (n=200) with unrelated diseases.

5.4.4.4 Sanger Sequencing

Sanger Sequencing was used to confirm the variant in *ATP13A2* in the index case (IV:1) and both parents (II:1 and III:1).

5.4.5 RESULTS

5.4.5.1 Whole exome sequencing

WES metrics revealed good coverage of the WES in individual IV:1 (Table 5.5).

Table 5.5: WES metrics of individual IV:1 (family 2)

	Family 2 IV:1
Total number of reads	43,413,599
Reads aligned to target	34,945,383
Mean Target Coverage	33
% Target covered by 2 reads	94%
% Target covered by 10 reads	85%
% Target covered by 20 reads	68%
Total number of variants	21,308

5.4.5.2 Variant filtering

Following variant filtering, 29 homozygous, rare, variants remained (Table 5.6). Of immediate interest was the homozygous missense variant in the *ATP13A2* gene. Homozygous or compound heterozygous mutations in *ATP13A2* are associated with a juvenile onset complicated parkinsonian syndrome: Kufor-Rakeb Syndrome.[196] The residue at which the mutation occurs is highly conserved, with a GERP (Genomic Evolutionary Rate Profiling) score of 3.46, and is predicted with a high probability to be deleterious using in-silico prediction software (SIFT score of 0, and a Polyphen2 score of 0.96).

5.4.5.3 Sanger sequencing

Sanger sequencing confirmed the index case (IV:1) was homozygous for the c.T2525C p.L842P mutation in *ATP13A2*, both parents (II:1 and III:1) were also confirmed to be heterozygous carriers of this mutation.

Table 5.6: Homozygous Variants Remaining in Regions of Homozygosity Following Exome Filtering

Gene	Full Gene Name	Annotation	Exonic Function	Call	Present in publically available databases	SIFT Score	Poly Phen2	GERP Score
ATP13A2	ATPase type 13A2	ATP13A2:uc001bac.2:exon23:c.T2525C:p.L842P	nonsynonymous SNV	hom	Novel	0	0.965088	3.46
CASZ1	castor zinc finger 1	CASZ1:uc009vmx.2:exon9:c.C2639T:p.P880L	nonsynonymous SNV	hom	Novel	0	0.651	3.85
NECAP2	NECAP endocytosis associated 2	NECAP2:uc001ayo.3:exon6:c.A502G:p.K168E	nonsynonymous SNV	hom	Novel	0	0.791	5.33
GJB3	gap junction protein, beta 3, 31kDa	GJB3:uc001bxz.4:exon1:c.342delC:p.A114fs	frameshift deletion	hom	Novel	NA	NA	NA
RPL5	ribosomal protein L5	RPL5:uc001dpd.3:exon2:c.A104G:p.D35G	nonsynonymous SNV	hom	Novel	0.11	0.616	5.01
DPH5	DPH5 homolog (S. cerevisiae)	DPH5:uc001dty.2:exon4:c.G108A:p.M36I	nonsynonymous SNV	hom	Novel	0.01	0.128	4.03

LCE1F	late cornified envelope 1F	LCE1F:uc010pdv.2:exon1:c.191_197GCTGCTGCAGCTCTGGGGGTGGTGG	nonframeshift substitution	hom	Novel	NA	NA	NA
KCNN3	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3	KCNN3:uc001ffp.3:exon1:c.199_200insCAGCAGCAGCAGCAG:p.Q67delinsQQQQQQ	nonframeshift insertion	hom	Novel	NA	NA	NA
EFNA4	ephrin-A4	EFNA4:uc001fhd.3:exon4:c.T590C:p.L197P	nonsynonymous SNV	hom	Novel	0	0.570624	4.76
PCNXL2	pecanex-like 2 (Drosophila)	PCNXL2:uc001hvk.1:exon2:c.G226T:p.D76Y	nonsynonymous SNV	hom	Novel	0.1	NA	NA
LYST	lysosomal trafficking regulator	LYST:uc001hxi.2:exon15:c.G2140A:p.G714R	nonsynonymous SNV	hom	9.30E-05	0.19	0.889	5.53
SMEK2	SMEK homolog 2, suppressor of mek1 (Dictyostelium)	SMEK2:uc002rzb.3:exon13:c.G1838A:p.R613H	nonsynonymous SNV	hom	Novel	0.31	0	5.1

TMEM131	transmembrane protein 131	TMEM131:uc002syh.4:exon5:c.C419T:p.T140M	nonsynonymous SNV	hom	Novel	0	NA	NA
RFX8	regulatory factor X, 8	RFX8:uc010yvx.1:exon8:c.C647T:p.A216V	nonsynonymous SNV	hom	Novel	0.06	NA	NA
FOXL2	forkhead box L2	FOXL2:uc003esw.3:exon1:c.C668A:p.A223D	nonsynonymous SNV	hom	Novel	0.01	0.509	2.94
MCCC1	methylcrotonoyl-CoA carboxylase 1 (alpha)	MCCC1:uc003flf.3:exon9:c.G911A:p.R304Q	nonsynonymous SNV	hom	Novel	0.03	0.666	5.1
G3BP2	GTPase activating protein binding protein 2	G3BP2:uc003hit.3:exon9:c.A871G:p.I291V	nonsynonymous SNV	hom	Novel	0.55	0.001	4.5
SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic	SEMA5A:uc003jek.2:exon9:c.G824A:p.R275H	nonsynonymous SNV	hom	Novel	0	0.986	5.25

	domain 5A							
REV3L	REV3-like, catalytic subunit of DNA polymerase zeta (yeast)	REV3L:uc003puy.4:exon13:c.G2704A:p.G902R	nonsynonymous SNV	hom	Novel	0	0.995	5.32
TTYH3	tweety homolog 3 (Drosophila)	TTYH3:uc003smp.3:exon2:c.C278A:p.A93D	nonsynonymous SNV	hom	Novel	0	0.668	4.61
RPA3	replication protein A3, 14kDa	RPA3:uc003sri.3:exon6:c.C110A:p.T37N	nonsynonymous SNV	hom	Novel	0.25	0.039	3.37
PCSK5	proprotein convertase subtilisin/kexin type 5	PCSK5:uc004ajy.2:exon14:c.2008_2017del:p.670_673del	frameshift deletion	hom	Novel	NA	NA	NA
PTRH1	peptidyl-tRNA hydrolase 1 homolog	PTRH1:uc004bro.3:exon2:c.176delG:p.R59fs	frameshift deletion	hom	Novel	NA	NA	NA
ANXA11	annexin A11	ANXA11:uc010qlx.1:exon6:c.	nonsynonymous	hom	0.000186	0	1	4.19

		G1205A:p.R402H	SNV					
SYT4	synaptotagmin IV	SYT4:uc002law.3:exon2:c.G505A:p.A169T,SYT4:uc010xcm.2:exon2:c.G451A:p.A151T	nonsynonymous SNV	hom	0.0005	0.05	0.045	3.11
MYLK2	myosin light chain kinase 2	MYLK2:uc002wws.2:exon4:c.C298A:p.L100M	nonsynonymous SNV	hom	Novel	0.04	0	3.12
CTNNBL1	catenin, beta like 1	CTNNBL1:uc002xhj.3:exon7:c.C694T:p.R232W	nonsynonymous SNV	hom	Novel	0	0.998	4.93
BC112340	NA	BC112340:uc002zoc.3:exon1:c.C329T:p.T110I	nonsynonymous SNV	hom	Novel	0	NA	NA
FER1L4	NA	FER1L4:uc002xcx.3:exon5:c.C622T:p.R208X	stopgain SNV	hom	Novel	1	NA	NA

5.4.6 DISCUSSION

Exome sequencing and autozygosity mapping in this family has revealed a novel homozygous missense mutation in *ATP13A2* (c.T2525C p.L842P) supporting a diagnosis of Kufor-Rakeb Syndrome in this patient. Kufor-Rakeb syndrome (KRS) (MIM: 606693) was originally reported in a consanguineous family originating from Kufor-Rakeb in Jordan.[313] KRS is an autosomal recessive inherited juvenile Parkinsonian syndrome caused by mutations in the *ATP13A2* gene.[196] The clinical features comprise akinetic-rigid parkinsonism, supranuclear gaze palsy, oculogyric dystonic spasms, facial-faucial-finger mini-myoclonus, and spasticity in some. Dementia and visual hallucinations are also recognised cognitive features. Brain imaging often reveals moderate cerebral and cerebellar atrophy sometimes with iron deposition in the basal ganglia.[197] Individual IV:1 (family 2) has a clinical presentation in keeping with that recognised with KRS, with juvenile onset symptoms including: parkinsonism, pyramidal tract signs, dystonia, eye movement abnormalities, supranuclear gaze palsy, cognitive impairment and myoclonus. IV:1 developed L-dopa-induced dyskinesias early in the course of the disease, as has been reported in other KRS cases.[314, 315] To date, only nine KRS families have been reported in the literature and they carry 11 different mutations.[196, 197, 314, 316-322] The patient (IV:1) is the second patient of Asian ancestry in the literature to be reported with KRS, confirming a relevance for this condition in this population.[322] Recently, a homozygous missense mutation in *ATP13A2* was identified in a family with pathologically confirmed neuronal ceroid-lipofuscinosis.[35] This finding suggests that KRS is linked to the lysosomal pathway, a pathway that was already hypothesized for a variety of parkinsonian phenotypes, but was not previously shown for KRS.

This novel *ATP13A2* mutation c.T2525C p.L842P, resides in the hydrolase domain of the protein, a domain in which other pathogenic frameshift and missense mutations have already been described.[196, 314, 316, 319, 320, 323-327] To-date, only three other homozygous missense variants have been reported in association with KRS,[314, 316, 321] the vast majority being frameshift and splice-site mutations. The mechanism(s) through which missense mutations cause KRS are not fully understood at present, however a recent study implicates that another missense

mutation in the hydrolase domain results in (p.G877R) impaired protein stability and enhanced degradation by the proteasome. Other missense mutations resulted in mislocalization of ATP13A2 to the endoplasmic reticulum.[328]

Family 2 demonstrate the utility of WES and autozygosity mapping as a cost-effective, timely way of simultaneous screening of several genes that can cause juvenile onset, complex parkinsonian syndromes, even in a singleton case.

5.5 FAMILY 3: EXOME SEQUENCING AND LINKAGE ANALYSIS IN A NON-CONSANGUINEOUS PARKINSONISM-DYSTONIA FAMILY

5.5.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

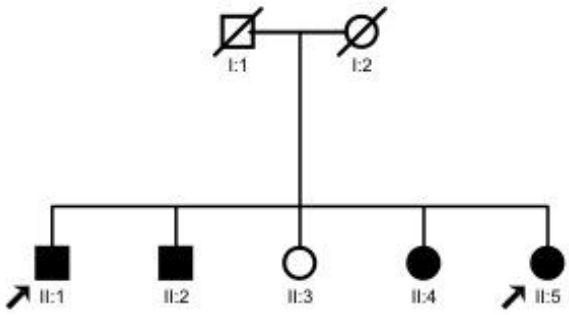
I examined all family members. I performed parts of the genome-wide linkage analysis. I analysed the WES data and performed confirmatory Sanger sequencing and segregation analysis.

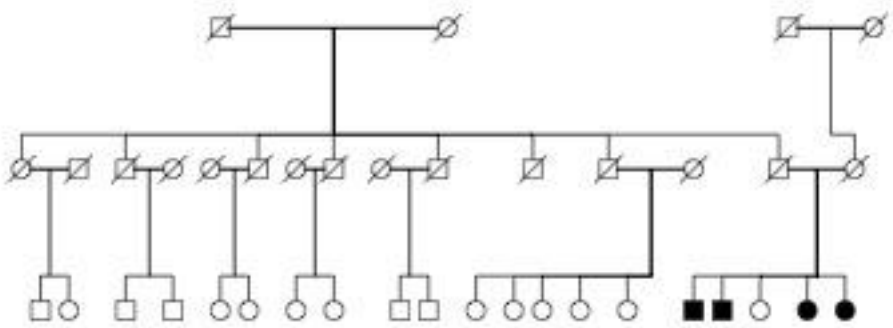
5.5.2 BACKGROUND

Dr. Nin Bajaj (a movements disorders expert in Birmingham, UK) identified this interesting family (figure 5.3) in whom there is no known parental consanguinity. In order to try and identify the causative variant in this family a combination of linkage analysis and exome sequencing was performed. The mode of inheritance was assumed to be autosomal recessive with a penetrance of 90%. Analysis of WES variants did not reveal any homozygous or compound heterozygous variants in the linkage regions. However, a novel, highly conserved variant was identified in

the *LRRK2* gene, a gene in which mutations are known to cause autosomal dominant PD. The variant segregates with PD in this family.

Figure 5.3: Pedigree of family 3. An arrow indicates the individuals selected for exome sequencing





Above: Full pedigree for family 3.

5.5.3 METHODS

5.5.3.1 Samples

DNA samples from II:1-II:5 were collected by myself. II:1-II:5 were examined by both Dr. Nin Bajaj and myself. Individuals I:1 and I:2 are deceased, DNA is not available on them, nor were they formally clinically examined. The wider family were not available to participate in the study.

5.5.3.2 Genome wide Linkage Analysis

Genome-wide multipoint parametric linkage utilizing ~30,000 equally spaced markers was performed using the computer program Allegro. This work was performed by Prof Robert Kleta, Dr. Dr. Horia Stanescu and Mehmet Tekman at UCL.

5.5.3.4 Whole Exome Sequencing

Individuals II:1 and II:5 were selected for WES. This was carried out as described in the materials and methods section (chapter 2).

5.5.3.5 Variant filtering

Only variants within regions of linkage were used for filtering. Synonymous variants and any variant present in a range of publically available databases of sequence variation (dbSNP, 1000 Genomes, Complete Genomic 69 Database and NHLBI Exome Sequencing Project database) with a frequency more than 1% were filtered out.

5.5.3.6 Sanger sequencing

The variant in the *LRRK2* gene was confirmed by Sanger Sequencing as described in Chapter 2).

5.5.4 RESULTS

5.5.4.1 Phenotypic Characterisation of patients

II:1 is a 49 years old gentleman who developed a tremor from very early childhood. He had learning difficulties and attended a specialist school. He had poor motor skills and struggled with sports at school. He also had many falls as a child. He left school aged 16 without any qualifications. He became aware of an arm tremor and dysarthria aged 28 years of age. He has a tendency to fall backwards.

Investigations showed normal copper and ceruloplasmin. Genetic testing was negative for DRPLA, Freidreich's ataxia and the common Spinocerebellar ataxias. DATscan showed a presynaptic dopaminergic deficit.

Examination revealed cognitive impairment especially in the frontal domain. He had a mild right-sided torticollis. Pursuit eye movements were broken. There was a right-sided rest tremor with a dystonic postural tremor and hyperextension of the thumb. Cogwheel rigidity and bradykinesia were present bilaterally but worse on the right. Gait examination revealed bilaterally reduced arms swing and dragging of the right leg. Postural reflexes were impaired. The patient is L-dopa responsive and on a low dose of Madopar twice daily. In summary there is evidence of learning disability, dystonia and parkinsonism.

II:2 developed a bilateral hand tremor as a child. He had normal motor skills as a child and excelled in sport at school. He developed psychiatric problems aged 25 years and was treated with neuroleptics including Dosulepin, Chlorpromazine and Olanzapine. He had a dystonic reaction to one of these drugs necessitating treatment with Chlorpromazine. He developed falls aged 35 years. He also has dysphagia and constipation, which has required hospital admission. A DATscan was positive indicating that the parkinsonism was not drug induced but due to a primary dopaminergic deficit. His parkinsonism symptoms respond to a low dose of Madopar.

His examination showed cogwheel rigidity, which was asymmetric being worse on the right, together with some axial rigidity. There was no rest or postural tremor. His gait was shuffling with reduced arm swing. He had retained postural reflexes.

II:3 has a normal examination.

II:4 also developed a bilateral arm tremor in the first decade of life, which has progressively become worse. Her mobility is not affected. She has no REM sleep behavioral disturbances or anosmia. She was found to have signs of parkinsonism incidentally on examining her as part of this study aged 46 years. Neurological examination reveals mild facial hypomimia with a bilateral rest tremor worse on

the right with an emergent postural tremor and dystonic posturing of the right hand. There was mild right torticollis. Repetitive movements showed bradykinesia more on the right. There was also bilateral cogwheel rigidity. Gait examination showed reduced arm swing on the right. There was also a mild degree of micrographia. The patient was L-dopa responsive and had a severe type III abnormality on DAT scan.

II:5 developed an arm tremor aged 26 years. Over the years this has worsened and her mobility has become affected resulting in her becoming wheelchair bound aged 48 years. There is no history of REM sleep disorder or anosmia. Memory is not impaired. Past medical history includes childhood polio affecting the left arm and leg and diabetes. Examination showed a bilateral rest and postural tremor. There was thumb hyperextension bilaterally indicating dystonic posturing. There was marked bradykinesia and cog-wheeling bilaterally.

Clinical examination for the wider family was not available as they were unavailable to participate in the research study, however none had a diagnosis of PD.

5.5.4.2 Clinical summary

This is a non-consanguineous family originating from Pakistan. 4 of the 5 siblings are affected by varying degrees of parkinsonism-dystonia which is confirmed on DAT scan. An autosomal recessive inheritance is most likely for a number of reasons. 1. There is an early age of onset with several siblings developing an arm tremor in the first decade. 2. Neither parent was known to be affected, they lived into their late 50's. 3. None of the affected sibling's children have developed symptoms, despite the early age of onset, some of the children are now in their 20's. However, there is a possibility that this is an early-onset autosomal dominant disorder, most likely inherited through the maternal side of the family. Their mother may not have been symptomatic due to reduced penetrance, or because she died before the disease onset. Transmission of the disorder via the maternal side of the family may be masked by the fact the mother has no siblings.

5.5.4.2 Genome Wide Parametric Linkage Analysis

Multipoint parametric linkage analysis resulted in a total of ~134Mb of linked regions (table 5.7), narrowing the region in which the candidate variant(s) are most likely to be found to 4.5% of the total genome.

Table 5.7: Linked regions for family 3

Chromosome	Position From	Position To	Lod Score
chr4	4,826,875	5,116,247	1.9311
chr4	5,576,140	5,827,079	1.9311
chr5	8,791,745	56,733,883	1.9311
chr5	150,760,282	151,351,067	0.9993
chr7	47,898,734	68,549,573	1.9311
chr9	93,874,879	109,989,135	1.846
chr10	1,948,064	2,394,389	1.5551
chr10	105,328,241	105,817,891	1.7334
chr11	86,093,568	86,621,300	1.9311
chr11	93,281,716	94,902,171	1.9311
chr11	110,023,164	110,833,266	1.0272
chr12	7,645,777	22,263,219	1.9311
chr12	30,938,922	50,306,148	1.9311
chr12	51,585,335	53,210,145	1.6074
chr12	67,880,042	68,485,984	1.0808
chr13	113,530,199	114,056,524	1.9311
chr15	22,916,162	23,877,411	1.9311
chr15	45,934,869	46,845,278	1.5562
chr19	2,032,148	2,930,806	1.9311
chr20	2,978,928	6,836,952	1.9311
chr21	16,139,324	16,576,783	1.5031

5.5.4.3 Whole Exome Sequencing

WES metrics indicated good coverage in both II:2 and II:5 (table 5.8)

Table 5.8: WES metrics

	II:1 Family 3	II:5 Family 3
Total number of reads	70,225,472	150,593,410
Non-duplicated reads	43,796,142	130,413,440
Reads aligned to target	85%	85%
Mean Target Coverage	35	87
% Target covered by 2 reads	91%	96%
% Target covered by 10 reads	81%	94%
% Target covered by 20 reads	67%	91%
Total number of variants	21,025	23,555

5.5.4.4 Variant Filtering

Tables 5.9 and 5.10 show the variants remaining after variant filtering as described above in individuals II:2 and II:5.

Table 5.9: Variants remaining after filtering in linkage regions in individual II:1

Gene	Gene Name	Amino acid change	Exonic Function	Call	Highest Frequency in Publically Available databases	SIFT Score	Poly phen2 Score	GERP Score	Comment
AGXT2	alanine-glyoxylate aminotransferase 2	ENST00000231420.6 exon12 c.T1269G p.D423E	nonsynonymous SNV	het	7.70E-05	0.15	0.122	2.67	Class III pyridoxal-phosphate-dependent mitochondrial aminotransferase. Catalyzes conversion of glyoxylate to glycine.
ATF7IP	activating transcription factor 7 interacting protein	ENST00000540793.1 exon1 c.C1505T p.S502L	nonsynonymous SNV	het	0.003695	0.25	0.673	4.5	Protein that can act as a transcriptional coactivator or corepressor depending upon its binding partners

CARD6	caspase recruitment domain family, member 6	ENST00000254691.5 exon3 c.C1016T p.T339I	nonsynonymous SNV	het	Novel	0	0.988	2.79	Microtubule-associated protein that interacts with receptor-interacting protein kinases and positively modulate signal transduction pathways converging on activation of the inducible transcription factor NF-kB. May be involved in apoptosis
CDKN1B	cyclin-dependent kinase inhibitor 1B	ENST00000228872.4 exon1 c.C277T p.R93W	nonsynonymous SNV	het	5.00E-04	0	1	5.05	Important regulator of cell cycle progression. Involved in G1 arrest.
CLEC4E	C-type lectin domain family 4, member E	ENST00000299663.3 exon2 c.G83T p.G28V	nonsynonymous SNV	het	7.70E-05	0.28	0.861	2.23	C-type lectin that functions as cell-surface receptor for a wide variety of ligands such as damaged cells, fungi and mycobacteria.
DNAJC22	DnaJ homolog, subfamily C, member 22	ENST00000395069.2 exon2 c.G742A p.G248R	nonsynonymous SNV	het	0.000769	0	0.995	3.52	May function as a co-chaperone

GZMK	granzyme K	ENST00000231009.2 exon2 c.G212A p.R71QGZMK;GZMK	Splicing	het	Novel	0.27	0.137	-3.56	Serine proteases from the cytoplasmic granules of cytotoxic lymphocyte. Poorly conserved.
LRRK2	leucine-rich repeat kinase 2	ENST00000298910.7 exon47 c.C6974T p.S2325F	nonsynonymous SNV	het	Novel	0	0.544	5.38	Gene in which mutations are known to cause autosomal dominant Parkinson's disease
NANOGP1	Nanog Homeobox Pseudogene 1	ENST00000530989.1 exon3 c.T305C p.L102P	nonsynonymous SNV	het	0.007	0.26	NA	NA	Probable transcriptional regulator
OR13C2	olfactory receptor, family 13, subfamily C, member 2	ENST00000542196.1 exon1 c.C88G p.L30V	nonsynonymous SNV	het	Novel	0.11	0	2.66	Odorant receptor
PKP2	plakophilin 2	ENST00000340811.4 exon11 c.A2233G p.I745V	nonsynonymous SNV	het	Novel	0.29	0.022	2.64	May play a role in junctional plaques

PLEKHA5	pleckstrin homology domain containing, family A member 5	ENST00000424268.1 exon8 c.G791A p.S264N	nonsynonymous SNV	het	Novel	0.77	0	2.55	Unknown function
RAI14	retinoic acid induced 14	ENST00000397449.1 exon2 c.C110T p.A37V	nonsynonymous SNV	het	Novel	0.17	0.003	5.37	Unknown function
RECQL	RecQ protein-like	ENST00000444129.2 exon6 c.G644T p.R215L	nonsynonymous SNV	het	Novel	0.03	0.37	4.21	DNA helicase that may play a role in the repair of DNA that is damaged by ultraviolet light or other mutagens
THOP1	thimet oligopeptidase 1	ENST00000307741.5 exon5 c.C557T p.T186M	nonsynonymous SNV	het	0.001461	0.04	0.001	2.81	Involved in the metabolism of neuropeptides <20 amino acid residues long. Involved in cytoplasmic peptide degradation. Able to degrade the beta-amyloid precursor protein

TMOD1	tropomodulin 1	ENST00000259365.3 exon2 c.G14A p.R5Q	nonsynonymous SNV	het	0.000846	0.07	0.027	4.91	Is an actin-capping protein that regulates tropomyosin, inhibiting depolymerization and elongation of the pointed end of actin filaments and thereby influencing the structure of the erythrocyte membrane skeleton
TTC23L	tetratricopeptide repeat domain 23-like	ENST00000505624.1 exon3 c.C218T p.S73F	nonsynonymous SNV	het	Novel	0.05	NA	NA	Not known
ZNF462	zinc finger protein 462	ENST00000441147.2 exon1 c.C1819T p.R607W	nonsynonymous SNV	het	Novel	0	0.997	3.05	May be involved in transcriptional regulation

Table 5.10: Variants remaining after filtering in linkage regions in individual II:5

Gene	Gene Name	Amino acid change	Exonic Function	Call	Highest Frequency in Publically Available databases	SIFT Score	Poly Phen2 Score	GERP Score	Comments
CARD6	caspase recruitment domain family, member 6	ENST00000254691.5 exon3 c.C1016T p.T339I	nonsynonymous SNV	het	Novel	0	0.988	2.79	Microtubule-associated protein that interacts with receptor-interacting protein kinases and positively modulate signal transduction pathways converging on activation of the inducible transcription factor NF-kB. May be involved in apoptosis
CDKN1B	cyclin-dependent kinase inhibitor 1B	ENST00000228872.4 exon1 c.C277T p.R93W	nonsynonymous SNV	het	5.00E-04	0	1	5.05	Important regulator of cell cycle progression. Involved in G1 arrest.

CLEC4E	C-type lectin domain family 4, member E	ENST00000299663.3 exon2 c.G83T p.G28V	nonsynonymous SNV	het	7.70E-05	0.28	0.861	2.23	C-type lectin that functions as cell-surface receptor for a wide variety of ligands such as damaged cells, fungi and mycobacteria.
DAP	death-associated protein	ENST00000230895.6 exon1 c.A44C p.H15P	nonsynonymous SNV	het	Novel	0	0.998	3.38	Negative regulator of autophagy. Involved in mediating interferon-gamma-induced cell death
DNAJC22	DnaJ (Hsp40) homolog, subfamily C, member 22	ENST00000395069.2 exon2 c.G742A p.G248R	nonsynonymous SNV	het	0.000769	0	0.995	3.52	May function as a co-chaperone
GOLGA6L2	golgin A6 family-like 2	ENST00000567107.1 exon8 c.1788_1790del p.596_597del	nonframeshift deletion	het	Novel	NA	NA	NA	Not known
GZMK	granzyme K (granzyme 3; tryptase II)	ENST00000231009.2 exon2 c.G212A p.R71Q	nonsynonymous SNV	het	Novel	0.27	0.137	-3.56	Serine protease from the cytoplasmic granules of cytotoxic lymphocyte. Poorly conserved.

KRT2	keratin 2	ENST00000309680.3 exon1 c.286_287insTTTGGAGGCGGCAGCGGC p.F96delinsFGGGSGF	nonframeshift insertion	het	Novel	NA	NA	NA	Probably contributes to terminal cornification. Associated with keratinocyte activation, proliferation and keratinization
KRT72	keratin 72	ENST00000354310.4 exon6 c.G1045C p.E349Q	nonsynonymous SNV	het	Novel	0.03	0.011	1.97	Has a role in hair formation
LRRK2	leucine-rich repeat kinase 2	ENST00000298910.7 exon47 c.C6974T p.S2325F	nonsynonymous SNV	het	Novel	0	0.544	5.38	Gene in which mutations are known to cause autosomal dominant Parkinson's disease
PKP2	plakophilin 2	ENST00000340811.4 exon11 c.A2233G p.I745V	nonsynonymous SNV	het	Novel	0.29	0.022	2.64	May play a role in junctional plaques

PLEKHA5	pleckstrin homology domain containing, family A member 5	ENST00000424268.1 exon8 c.G791A p.S264Nn11 c.G1133A p.S378N	nonsynonymous SNV	het	Novel	0.77	0	2.55	Unknown function
PRB3	proline-rich protein BstNI subfamily 3	ENST00000381842.3 exon5 c.605-2A>G	Splicing	hom	Novel	NA	NA	NA	Major constituent of parotid saliva. Proposed to act as a bacterial receptor.
PRDM9	PR domain containing 9	ENST00000296682.3 exon11 c.C2042G p.T681S	nonsynonymous SNV	hom	Novel	1	0.109952	-4.57	Likely to be involved in transcriptional regulation
RAI14	retinoic acid induced 14	ENST00000397449.1 exon2 c.C110T p.A37V	nonsynonymous SNV	het	Novel	0.17	0.003	5.37	Unknown function
RECQL	RecQ protein-like	ENST00000444129.2 exon6 c.G644T	nonsynonymous SNV	het	Novel	0.03	0.37	4.21	DNA helicase that may play a role in the repair of DNA that is damaged by ultraviolet light or other

		p.R215L							mutagens
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin	NASIGLEC1	splicing	het	Novel	NA	NA	NA	Acts as an endocytic receptor mediating clathrin dependent endocytosis.
TAS2R31	taste receptor, type 2, member 31	ENST00000390675.2 exon1 c.G259A p.V87I	nonsynonymous SNV	het	present in dbSNP but MAF not available	NA	NA	NA	Receptor that may play a role in the perception of bitterness. Unlikely candidate
TMOD1	tropomodulin 1	ENST00000259365.3 exon2 c.G14A p.R5Q	nonsynonymous SNV	het	0.000846	0.07	0.027	4.91	Actin-capping protein that regulates tropomyosin by binding to its N-terminus, inhibiting depolymerization and elongation of the pointed end of actin filaments and thereby influencing

									the structure of the erythrocyte membrane skeleton.
TTC23L	tetratricopeptide repeat domain 23-like	ENST00000505624.1 exon3 c.C218T p.S73F	nonsynonymous SNV	het	Novel	0.05	NA	NA	Not known
ZNF462	zinc finger protein 462	ENST00000441147.2 exon1 c.C1819T p.R607W	nonsynonymous SNV	het	Novel	0	0.997	3.05	May be involved in transcriptional regulation

Review of the variants in both individual II:1 and II:5 revealed there was no gene which had a compound heterozygous or homozygous variant in. This remained the case even when synonymous variants were included in the analysis. A copy number analysis was performed across the exome using 'Exome depth' (as described in chapter 2), however this did not reveal any homozygous copy number variants within the linkage region, nor did it reveal any heterozygous copy number variants in genes which already had a splicing, nonsynonymous or indel variant ascertained from exome sequencing. The coverage of ascertained variants within the linkage regions in II:1 was checked in II:5 (and vice versa) to confirm that variants that were not shared between the two individuals were sufficiently covered in the other individual.

Next, I considered only shared variants between individuals II:1 and II:5 (Table 5.11). Of interest, there is a novel, nonsynonymous variant in *LRRK2* (ENST00000298910.7 exon47 c.C6974T p.S2325F), at a highly conserved residue (GERP score 5.38), within the WD40 domain of the protein. Mutations in *LRRK2* are an established cause of autosomal dominant PD. Sanger sequencing confirmed that this variant was present in the four affected individuals and not in the unaffected individual.

Table 5.11: Shared rare variants between individual II:1 and II:5.

Gene	Gene Name	Amino acid change	Exonic Function	Call	Highest Frequency in Publically Available databases	SIFT Score	Poly Phen2 Score	GERP Score	Comments
CARD6	caspase recruitment domain family, member 6	ENST00000254691.5 exon3 c.C1016T p.T339I	nonsynonymous SNV	het	Novel	0	0.988	2.79	Microtubule-associated protein that interacts with receptor-interacting protein kinases and positively modulate signal transduction pathways converging on activation of the inducible transcription factor NF-kB. May be involved in apoptosis
CDKN1B	cyclin-dependent kinase inhibitor 1B	ENST00000228872.4 exon1 c.C277T p.R93W	nonsynonymous SNV	het	5.00E-04	0	1	5.05	Important regulator of cell cycle progression. Involved in G1 arrest.

CLEC4E	C-type lectin domain family 4, member E	ENST00000299663.3 exon2 c.G83T p.G28V	nonsynonymous SNV	het	7.70E-05	0.28	0.861	2.23	C-type lectin that functions as cell-surface receptor for a wide variety of ligands such as damaged cells, fungi and mycobacteria.
DNAJC22	DnaJ (Hsp40) homolog, subfamily C, member 22	ENST00000395069.2 exon2 c.G742A p.G248R	nonsynonymous SNV	het	0.000769	0	0.995	3.52	May function as a co-chaperone
GZMK	granzyme K (granzyme 3; tryptase II)	ENST00000231009.2 exon2 c.G212A p.R71Q	nonsynonymous SNV	het	Novel	0.27	0.137	-3.56	Serine protease from the cytoplasmic granules of cytotoxic lymphocyte. Poorly conserved.
LRRK2	leucine-rich repeat kinase 2	ENST00000298910.7 exon47 c.C6974T p.S2325F	nonsynonymous SNV	het	Novel	0	0.544	5.38	Gene in which mutations are known to cause autosomal dominant Parkinson's disease

PKP2	plakophilin 2	ENST00000340811.4 exon11 c.A2233G p.I745V	nonsynonymous SNV	het	Novel	0.29	0.022	2.64	May play a role in junctional plaques
PLEKHA5	pleckstrin homology domain containing, family A member 5	ENST00000424268.1 exon8 c.G791A p.S264Nn11 c.G1133A p.S378N	nonsynonymous SNV	het	Novel	0.77	0	2.55	Unknown function
RAI14	retinoic acid induced 14	ENST00000397449.1 exon2 c.C110T p.A37V	nonsynonymous SNV	het	Novel	0.17	0.003	5.37	Unknown function
RECQL	RecQ protein-like	ENST00000444129.2 exon6 c.G644T p.R215L	nonsynonymous SNV	het	Novel	0.03	0.37	4.21	DNA helicase that may play a role in the repair of DNA that is damaged by ultraviolet light or other mutagens

TTC23L	tetratricopeptide repeat domain 23-like	ENST00000505624.1 exon3 c.C218T p.S73F	nonsynonymous SNV	het	Novel	0.05	NA	NA	Not known
ZNF462	zinc finger protein 462	ENST00000441147.2 exon1 c.C1819T p.R607W	nonsynonymous SNV	het	Novel	0	0.997	3.05	May be involved in transcriptional regulation

Parametric genome-wide linkage analysis was also performed using Merlin with an autosomal dominant model of inheritance (see materials and methods, chapter 2). This revealed 71 linkage regions (Table 5.12). When variant filtering was repeated for these linkage regions only, after synonymous and homozygous variants were excluded, and those variants with a frequency of 0.1%, 38 shared variants between II:1 and II:5 remained (table 5.13). Once again, the novel, highly conserved, variant in the *LRRK2* gene is in one of the linkage regions with the highest Lod score of 1.12.

Table 5.12: Parametric Linkage regions and Lod scores for autosomal dominant inheritance.

Chromosome	Position from	Position to	Max lod score
1	20,056,631	22,674,370	0.8
1	117,845,032	180,543,539	0.82
1	201,883,199	205,418,134	0.82
1	244,930,726	245,474,078	0.82
2	6,418,557	11,753,697	0.83
2	21,586,117	29,229,806	0.82
2	30,860,007	30902939	0.82
2	34,672,921	40,341,388	0.82
2	45,905,946	65,892,430	0.82
2	69,126,125	73,857,002	0.82
2	78,307,412	78,755,686	0.82
2	82,410,462	86,376,474	0.82
2	101,911,385	115,730,919	0.82
2	123,818,973	137,179,433	0.82
2	173,369,231	210,330,754	0.82
2	231,976,460	238,294,434	0.82
3	9,877,481	15,935,841	0.77
3	25,212,878	29,258,177	0.82
3	34,822,886	35,278,145	0.82

3	45,828,586	59,625,877	0.82
3	151,507,750	172,321,434	0.82
4	3,454,304	23,962,061	1.12
4	159,876,379	180,472,339	0.82
5	1,529,358	14,776,897	1.1
5	73,030,942	73330562	0.82
5	76,171,063	77,284,789	0.82
5	90,858,258	100,567,453	0.82
5	110,708,854	116,080,067	0.82
5	135,978,129	152,797,679	0.82
5	173,684,581	178950814	0.82
6	20,274,321	33,753,673	0.82
6	41,941,980	46,914,760	0.82
6	53,489,402	68,921,177	0.82
6	109,410,569	119,903,523	0.82
6	126,029,043	152,624,699	0.82
7	39,285,620	96,519,876	1.12
7	152,162,209	157,600,780	0.82
8	176,818	6,471,526	0.82
8	13,174,945	14,489,024	0.82
8	141,727,310	144,163,378	0.82
9	2,267,871	8,215,609	0.82
9	19,386,565	38,214,047	0.82
10	1,276,186	14,332,431	0.82
10	89,175,636	95,101,480	0.82
10	124,931,515	133,441,851	0.82
11	20,989,375	22,475,813	0.82
11	27,401,190	32,226,445	0.82
11	78,698,287	101,470,545	1.12
11	128,371,441	134,617,321	0.82
12	3,647,268	59,850,416	1.12
12	68,617,219	75,500,790	0.81

12	80,580,111	96,447,847	0.82
13	21,745,977	23,900,652	0.82
13	44482204	57,064,916	0.3
13	63,644,364	83,837,607	0.82
13	108,403,984	109,865,139	0.82
14	23,596,320	34,199,638	0.82
14	88,275,924	93,796,411	0.82
15	22,759,438	34,574,330	0.82
15	46158407	47,246,651	0.82
15	61,124,220	101,862,970	1.12
16	442,805	6,144,648	0.82
16	13,803,064	17,722,115	0.82
16	23,888,840	30,154,740	0.82
16	73,421,001	81,427,483	0.82
17	1,943,888	6,733,672	0.82
18	6,653,590	8,546,911	0.61
18	72,845,315	5,426,449	0.82
20	191,797	18,889,562	1.12
21	15,659,254	40,935,773	0.82

Table 5.13: Showing variants shared between II:1 and II:5 within Linkage regions (autosomal dominant model of inheritance)

Gene	Gene Name	Amino acid change	Exonic Function	Call	Frequency in publically available databases	SIFT score	Poly Phen2 Score	GERP score	Comment
RECQL	RecQ protein-like	RECQL ENST00000444129.2 exon6 c.G644T p.R215L	nonsynonymous SNV	het	Novel	0.03	0.37	4.21	DNA repair
PKP2	plakophilin 2	PKP2 ENST00000340811.4 exon11 c.A2233G p.I745V	nonsynonymous SNV	het	Novel	0.29	0.022	2.64	May play a role in junctional plaques
LRRK2	leucine-rich repeat kinase 2	LRRK2 ENST00000298910.7 exon47 c.C6974T p.S2325F	nonsynonymous SNV	het	Novel	0	0.544	5.38	Gene in which mutations are known to cause autosomal dominant Parkinson's disease

ARHGEF11	Rho guanine nucleotide exchange factor 11	ARHGEF11 ENST00000361409.2 exon6 c.C416T p.A139V	nonsynonymous SNV	het	Novel	0.14	0.006	1.4	Play a fundamental role in cellular processes that are initiated by extracellular stimuli that work through G protein coupled receptors
TTC30A	tetratricopeptide repeat domain 30A	TTC30A ENST00000355689.4 exon1 c.G274A p.A92T	nonsynonymous SNV	het	Novel	0.01	0.503	5.68	Plays a role in anterograde intraflagellar transport, the process by which cilia precursors are transported from the base of the cilium to the site of their incorporation at the tip
RAPH1	Ras association and pleckstrin homology domains 1	RAPH1 ENST00000319170.5 exon14 c.2570_2578del p.857_860del	nonframeshift deletion	het	Novel	NA	NA	NA	Mediator of localized membrane signals. Implicated in the regulation of lamellipodial dynamics. Negatively regulates cell adhesion
PLXNB1	plexin B1	PLXNB1 ENST00000456774.1 exon19 c.G3265A p.V1089I	nonsynonymous SNV	het	Novel	0	0.984	4.8	Receptor for SEMA4D. Plays a role in RHOA activation and subsequent changes of the actin cytoskeleton. Plays a role in axon guidance, invasive growth and cell migration

ZNF518B	zinc finger protein 518B	ZNF518B ENST00000326756.3 exon3 c.G2758A p.A920T	nonsynonymous SNV	het	Novel	0	0.889	1.72	Transcriptional regulation
TKTL2	transketolase-like 2	TKTL2 ENST00000280605.3 exon1 c.G617C p.C206S	nonsynonymous SNV	het	Novel	0	0.907	2.66	Plays a pivotal role in carcinogenesis
PAPD7	PAP associated domain containing 7	PAPD7 ENST00000230859.6 exon13 c.A1621C p.S541R	nonsynonymous SNV	het	Novel	0	0.906	4.83	DNA polymerase that is likely involved in DNA repair.
ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	ST8SIA4 ENST00000231461.4 exon4 c.G619T p.D207Y	nonsynonymous SNV	het	Novel	0.02	0.418	4.97	Catalyzes the polycondensation of alpha-2,8-linked sialic acid required for the synthesis of polysialic acid, which is present on the embryonic neural cell adhesion molecule, necessary for plasticity of neural cells

APC	adenomatous polyposis coli	APC ENST00000257430.4 exon16 c.G4919A p.R1640Q	nonsynonymous SNV	het	Novel	0.31	1	5.81	Tumor suppressor gene. Mutations in this cause a familial cancer syndrome. Unlikely candidate
OR14J1	olfactory receptor, family 14, subfamily J, member 1	OR14J1 ENST00000377160.2 exon1 c.G56A p.R19H	nonsynonymous SNV	het	Novel	0.24	0.01	-4.29	Odorant receptor
VARS	valyl-tRNA synthetase	VARS ENST00000375663.3 exon7 c.G881C p.G294A	nonsynonymous SNV	het	submitted to dbSNP	0.34	0.038	2.91	Aminoacyl-tRNA synthetases catalyze the aminoacylation of tRNA by their cognate amino acid. Many tRNA synthetases are linkaed to neurological diseases
PRPH2	peripherin 2	PRPH2 ENST00000230381.5 exon2 c.G616T	nonsynonymous SNV	het	Novel	0.28	0.005	4.66	Involved in disk morphogenesis

		p.V206L							
HCRTR2	hypocretin (orexin) receptor 2	HCRTR2 ENST00000370862.3 exon7 c.G1211A p.R404Q	nonsynonymous SNV	het	Novel	0.07	0.293	4.99	G-protein coupled receptor involved in the regulation of feeding behaviour.
C7orf10	chromosome 7 open reading frame 10	C7orf10 ENST00000309930.5 exon3 c.G242A p.R81K	nonsynonymous SNV	het	Novel	0.36	NA	NA	Mutations in this gene are associated with glutaric aciduria type III, an autosomal recessive inborn error of metabolism
HTR5A	5-hydroxytryptamine receptor 5A	HTR5A ENST00000287907.2 exon2 c.G817T p.D273Y	nonsynonymous SNV	het	Novel	0	0.964	4.51	Receptors for serotonin, a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen
INSL6	insulin-like 6	INSL6 ENST00000381641.3 exon2 c.329_337del p.110_113del	nonframeshift deletion	het	Novel	NA	NA	NA	May have a role in sperm development and fertilization. Unlikely candidate gene

SEC61A2	Sec61 alpha 2 subunit (S. cerevisiae)	SEC61A2 ENST00000304267.8 exon12 c.A1249G p.K417E	nonsynonymous SNV	het	Novel	0.02	NA	NA	Appears to play a crucial role in the insertion of secretory and membrane polypeptides into the ER.
ESPL1	extra spindle pole bodies homolog 1	ESPL1 ENST00000257934.4 exon3 c.A937G p.K313E	nonsynonymous SNV	het	Novel	0.64	0	-2.85	Caspase-like protease, which plays a central role in the chromosome segregation. Unlikely candidate as poorly conserved
SGCG	sarcoglycan, gamma	SGCG ENST00000218867.3 exon5 c.T479C p.V160A	nonsynonymous SNV	het	Novel	0	0.551	5.06	Component of the sarcoglycan complex
TRIM13	tripartite motif containing 13	TRIM13 ENST00000298772.5 exon2 c.G137A p.R46Q	nonsynonymous SNV	het	Novel	0.04	0.999	5.17	E3 ubiquitin ligase. Candidate variant.
LMO7	LIM domain 7	LMO7 ENST00000341547.4 exon3 c.254_255del p.85_85del	frameshift deletion	het	Novel	NA	NA	NA	Involved in protein-protein interactions

AKAP6	A kinase (PRKA) anchor protein 6	AKAP6 ENST00000280979.4 exon4 c.G809A p.R270Q	nonsynonymous SNV	het	Novel	0.17	0.954	4.56	Binds to type II regulatory subunits of protein kinase A and anchors/targets them to the nuclear membrane or sarcoplasmic reticulum
BTBD7	BTB (POZ) domain containing 7	BTBD7 ENST00000393170.2 exon5 c.G811A p.G271S	nonsynonymous SNV	het	Novel	0.02	0.98	5.46	Unknown function. Highly conserved
PKM2	pyruvate kinase, muscle	PKM2 ENST00000449901.2 exon3 c.T62C p.I21T	nonsynonymous SNV	het	Novel	0	NA	NA	Involved in glycolysis
CHRNB4	cholinergic receptor, nicotinic, beta 4	CHRNB4 ENST00000412074.2 exon5 c.G595A p.D199N	nonsynonymous SNV	het	Novel	0	NA	NA	Subunit of Nicotinic receptor, ligand-gated ion channels that modulate cell membrane potentials
ZNF592	zinc finger protein 592	ZNF592 ENST00000299927.3 exon1 c.T1762A p.C588S	nonsynonymous SNV	het	Novel	0.3	0.408	4.8	Involved in cerebellar development. Mutations in this gene cause autosomal recessive spinocerebellar ataxia.

ACAN	aggrecan	ACAN ENST00000561243.1 exon11 c.G4150A p.A1384T	nonsynonymous SNV	het	submitted to dbSNP	0.24	NA	NA	Integral part of the extracellular matrix in cartilagenous tissue. Mutations in this gene may be involved in skeletal dysplasia and spinal degeneration.
FBXL16	F-box and leucine-rich repeat protein 16	FBXL16 ENST00000397621.1 exon2 c.G19A p.D7N	nonsynonymous SNV	het	Novel	0.66	0	-1.26	Substrate-recognition component of the SKP1-CUL1-F-box protein-type E3 ubiquitin ligase complex. Poorly conserved, unlikely candidate
IGFALS	insulin-like growth factor binding protein, acid labile subunit	IGFALS ENST00000215539.3 exon2 c.G1406A p.R469H	nonsynonymous SNV	het	Novel	0.08	0.003	-4.36	Involved in protein-protein interactions. Poorly conserved, unlikely candidate
NTN3	netrin 3	NTN3 ENST00000293973.1 exon1 c.G620A p.S207N	nonsynonymous SNV	het	Novel	1	0	-0.617	Control guidance of CNS commissural axons and peripheral motor axons. Unlikely candidate as poorly conserved
NTN3	netrin 3	NTN3 ENST00000293973.1	nonsynonymous SNV	het	Novel	0.01	0.288	2.28	Control guidance of CNS commissural axons and peripheral

		exon1 c.C621A p.S207R							motor axons. Unlikely candidate as poorly conserved
ADCY9	adenylate cyclase 9	ADCY9 ENST00000294016.3 exon2 c.A1609G p.K537E	nonsynonymous SNV	het	Novel	0.28	0.015	3.98	Adenylate cyclase is a membrane bound enzyme that catalyses the formation of cyclic AMP from ATP.
C17orf85	chromosome 17 open reading frame 85	C17orf85 ENST00000389005.4 exon10 c.G1192A p.D398N	nonsynonymous SNV	het	Novel	0.01	0.97	5.5	Unknown function
PCDHGA8	protocadherin gamma subfamily A, 8	PCDHGA8 ENST00000398604.2 exon1 c.G2063A p.S688N	nonsynonymous SNV	het	Novel	0.34	NA	NA	Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific neuronal connections in the brain

5.5.5 DISCUSSION

The genetic aetiology was investigated in a small non-consanguineous family with parkinsonism-dystonia using linkage analysis and WES of two affected individuals. The mode of inheritance was thought to be autosomal recessive, however, we could not exclude autosomal dominant inheritance.

For the analysis using an autosomal recessive model of inheritance, no compound heterozygous or homozygous variants were found within the linkage regions, even when widening filtering to include synonymous and copy number variants. However, a variant was found in the *LRRK2* gene (ENST00000298910.7 exon47 c.C6974T p.S2325F), at a very highly conserved residue in the WD40 domain (GERP score 5.38). The variant generates high scores for in-silico analysis of deleteriousness, SIFT predicts that the variant is damaging, whilst Polyphen2 predicts the variant is possibly damaging (SIFT score 0.05, Polyphen 2 score 0.57). The variant is absent from all current control databases. Mutations in *LRRK2* are a known cause of autosomal dominant PD. The p.S2325F *LRRK2* variant has not been reported previously in association with PD. As discussed in the introduction (chapter 1), no unequivocally pathogenic mutations have been reported in the WD40 domain to-date, however there at least nine variants in the WD40 domain have been described in association with PD (http://www.molgen.vib-ua.be/PDmutDB).[120, 121] In particular, the T2356I variant in the WD40 domain, has been reported several times in association with PD,[329-331] suggesting mutations in this domain might cause PD.

All affected individuals in this family developed an arm tremor in the first decade, however they were not examined at this time, therefore it is not clear whether the tremor is part of their Parkinson's disease or a separate condition. The age of onset of parkinsonism in this family ranges from 26-46 years. Early age of onset has been described in *LRRK2*-related PD,[332] however it is rare, a large study examining age related penetrance of *LRRK2* mutations, calculated the penetrance of non-G2019S *LRRK2* mutations to be less than 5% before the age of 35 years.[115] Many of the affected individuals in this family have dystonia as part of their clinical symptoms after L-dopa treatment. Dystonia is commoner in *LRRK2*-related PD than in

idiopathic parkinsonism and most often occurs after dopamine replacement.[115] Dystonia preceding L-dopa treatment is generally a feature of autosomal recessive parkinsonism (*Parkin*, *PINK1*, *DJ1*).

There are a number of factors that suggest that this *LRRK2* variant might be causal. The in-silico prediction suggests the variant is deleterious, the codon is highly conserved and other mutations in *LRRK2* are associated with genetic forms of PD. However, there are a number of considerations which are suggest that this may not be the causal variant. Asian individuals have more genetic variability than Caucasian individuals,[16] therefore this may represent a previously undescribed benign polymorphism in the *LRRK2* gene that is not present in control databases as genetic data from individuals of Asian ancestry are underrepresented in these databases. As the common mutations in *LRRK2* are clustered in the Roc, Cor and Kinase domain, typically most laboratories will only sequence these hot-spots in families with autosomal dominant PD. With the advent of next-generation sequencing all exons of the *LRRK2* gene will be screened in patients and it is likely to become clear whether mutations in other domains can cause PD. Additionally, other families may be identified with the p.S2325F *LRRK2* variant thus helping clarify whether or not it is pathogenic. Lastly, the mode of inheritance will become clearer as the affected individuals children reach adulthood.

The work on this family demonstrates the difficulties in WES data analysis in small families where the affected individuals are in a single generation. One of the simplest ways to determine a candidate variants pathogenicity is to show it segregates with disease in other families/patients. When the variant is novel and absent from both control cohort databases and patient replication cohorts functional work is needed to explore further the pathogenicity. If the causal variant in this family is not the p.S2325F *LRRK2* variant, it may have been ascertained through WES and be present in the candidate variants lists Table 5.9-5.11 and 5.13, or it may not have been ascertained due to the limitations of WES (discussed in section 5.3.3).

Chapter Conclusion

In summary the 3 families presented herein in this chapter illustrate a number of important points about the use of WES in autosomal recessive PD. Family 1 demonstrates that in small consanguineous families, WES and autozygosity mapping may produce candidate lists of variants/genes, but in order to confirm or refute the pathogenicity of such variants, replication studies in a large number of PD patients may be necessary. It is likely this will necessitate sharing of data amongst large research consortia. Family 2 demonstrate that WES and autozygosity mapping may be a very powerful tool even in singleton cases where the gene is already known to be associated with PD. Family 3 demonstrate that in small kindreds if the mode of inheritance is not be confidently known this can significantly hamper the data interpretation.

CHAPTER 6. WHOLE EXOME SEQUENCING IN AUTOSOMAL RECESSIVE GENERALISED DYSTONIA

6.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

Unless otherwise indicated I performed autozygosity mapping and analysed the WES data. I performed confirmatory Sanger sequencing and segregation analysis. I examined all family members.

6.2 BACKGROUND

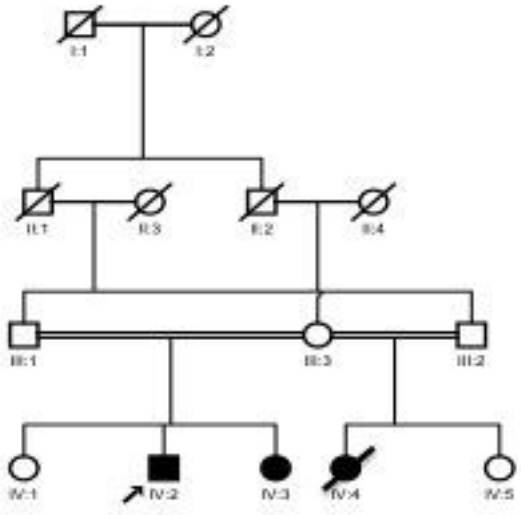
I identified 3 consanguineous families on the departmental neurogenetics database who had generalised dystonia where the genetic cause was unknown. Although these families were small I decided to perform WES in combination with autozygosity mapping or linkage analysis to try and determine the genetic cause.

6.3 FAMILY 4: AUTOZYGOSITY MAPPING AND EXOME SEQUENCING IN A CONSANGUINEOUS KINDRED WITH GENERALISED DYSTONIA AND SPASTIC PARAPARESIS

6.3.1 Background

A Pakistani family with generalised dystonia, parkinsonism, and a spastic paraparesis was identified by Prof Bhatia, though the movement disorders clinic at the NHNN. Parental consanguinity and the young age at onset suggested an autosomal recessive mode of inheritance. The pedigree for the family is shown in figure 6.1. WES and linkage analysis was used to try and identify the genetic cause.

Figure 6.1: Pedigree of family 4



6.3.2 MATERIALS AND METHODS

6.3.2.1 Samples

Genomic DNA samples were obtained from peripheral blood lymphocytes, for individuals III:1, III:3, IV:1, IV:2, IV:3, IV:4 and IV:5 using standard protocols. Participants gave written consent, and the study met with the local ethic boards approval.

6.3.2.2 Phenotypic characterization of the affected individuals

IV:2 was examined by both Prof. Kailash Bhatia (a Movement Disorders Specialist) and myself, IV:4 was examined by Dr. Lucinda Carr (Paediatric Neurologist) and was deceased at the time this research study commenced. IV:1, IV:3 and IV:5 were examined by myself.

6.3.2.3 Whole Exome Sequencing

Individual IV:2 was selected for WES, being the most severely affected living individual. WES was performed as described in the Materials and Methods chapter (chapter 2).

6.3.2.4 Genome wide Linkage Analysis

Multipoint parametric linkage utilizing ~30,000 equally spaced markers was performed using the computer program Allegro. This work was performed by Prof Robert Kleta, Dr. Dr. Horia Stanescu and Mehmet Tekman at UCL.

6.3.2.5 Variant filtering

Only variants within regions of linkage were used for filtering. Synonymous variants and any variant present in a range of publically available databases of sequence variation (dbSNP, 1000 Genomes, Complete Genomic 69 Database and NHLBI Exome Sequencing Project database) with a frequency greater than 1% were filtered out.

6.3.3 RESULTS

6.3.3.1 Phenotypic Characterisation of patients

Individual IV:4 was born at term by normal delivery. Motor milestones were normal up until the age of 12 months. She had been able to walk with support at the age of 12 months however from the age of 18 months she had become less steady on her feet and never achieved independent mobility. Her speech was delayed only being able to speak single words at the age of 2 years. Examination aged 2 revealed an unsteady high steppage gait requiring assistance. In the lower limbs there was pathologically brisk reflexes and extensor plantars. There was fine nystagmus in the primary position and slight pallor of the optic discs and fine pendular nystagmus on pursuit eye movement with poor optokinetic nystagmus. Over a period of 10 years her condition rapidly progressed. She developed spasticity in both upper and lower limbs with generalised dystonia and torticollis, which left her wheelchair bound. She also had severe learning difficulties. Surgery was required for a severe rapidly progressive scoliosis. She was dependent for all

activities of daily living and was fed via a gastrostomy. A trial of L-dopa was thought to be of some help. Baclofen and botulinum toxin injections were used for management of spasticity and dystonia. The patient died aged 20 years as a result of an intercurrent chest infection. Investigations showed normal routine bloods, karyotype, EEG, visual evoked potentials, auditory evoked potentials and nerve conduction studies. A CT brain showed mild cerebellar atrophy.

Individual IV:3 had normal motor milestones and early development. She first developed symptoms aged 17 with stiffness of the legs when walking. Examination aged 29 years revealed brisk upper limb reflexes. In the lower limbs there was a spastic paraparesis with pathologically brisk reflexes and extensor plantar responses. There were no features of dystonia present on examination.

Individual IV:2 developed mild walking difficulties and a tendency to fall from the age of 2 years at the time of a intercurrent viral illness. His speech development was very good up until the age of 4.5 years. At this time he had a dramatic worsening of his symptoms following an episode of hepatitis A and lost all movement including speech and swallowing. He made a partial recovery with regards to his walking. However later that year he had a second acute worsening during an intercurrent illness with tonsillitis. He was wheelchair bound thereafter. In the intervening period his clinical condition has remained stable. His examination aged 22 years revealed that he was significantly disabled due to generalised dystonia with bulbar involvement rendering him anarthric. He also has a spastic quadriparesis. Neuropsychometry demonstrated the patient has some degree of intellectual dysfunction. He communicates with an aid. He requires feeding via a percutaneous gastrostomy. His medications include Baclofen, Clonazepam, L-dopa and botulinum toxin injections. More recently he has had deep brain stimulation for management of his generalised dystonia with a good clinical response. MRI Brain scan showed normal basal ganglia and cerebellum with bifrontal atrophy – thought to be secondary to perinatal ischaemia. T2* MRI sequences did not reveal any mineralization.

Summary of investigations in Family

In order to rule out known genetic causes of dystonia or parkinsonism, affected family members were screened for mutations in the following genes as part of their diagnostic work-up: *DYT1*, *PANK2*, *PLA2G6*. Copper studies, CSF, white cell enzymes, amino acids, and lactate were normal in individual IV:1.

6.3.3.2 Genome Wide Linkage Analysis

Review of the genotyping data revealed clustered genotyping errors in 2 samples (IV:4 and IV:5), which remained when fresh samples were obtained and chipped for a second time. The additional genotype was highly suggestive of non-paternity, but from an individual who was very closely related to III:1 e.g., a brother. Multipoint parametric linkage analysis was performed which revealed only 2 regions of linkage (table 6.1). Autozygosity mapping confirmed there were no other regions of homozygosity outside of these 2 linkage regions shared only by affected siblings.

Table 6.1: regions of linkage in Family 4

Chromosome	SNP from	SNP To	Physical position (hg19) From	Physical position (hg19) To	LOD score
6	rs6912426	rs707542	25,409,785	36,924,102	2.958600
13	rs943782	rs7320510	27,575,037	30,997,831	2.958700

6.3.3.3 Exome Sequencing in Individual IV:2

WES metrics indicated good coverage in IV:2 (Table 6.2)

Table 6.2: WES metrics in IV:2

	IV:2 Family 4
Total number of reads	34,992,914
Reads aligned to target	27,150,170
Mean Target Coverage	26%
% Target covered by 2 reads	92%

% Target covered by 10 reads	80%
% Target covered by 20 reads	57%
Total number of variants	21,308

6.3.3.4 Variant filtering

First the linkage regions established from the analysis taking into account the non-paternity were used. Only variants within these regions were used for filtering. Variants were filtered out if they were present in publically available databases at a frequency of more than 1%. The variants that remained after filtering are detailed in table 6.3.

Table 6.3: Variants remaining within linkage regions following filtering

Gene	Gene Name	Amino acid Change	Exonic Function	Call	Allele frequency in publically available databases	SIFT	Polyphen 2	Comments
HLA-A	Major histocompatibility complex, class I, A	ENST00000376802.2: exon4:c.751delG: p.D251fs	frameshift deletion	het	recorded once in dbSNP	NA	NA	*HLA-A* is involved in the presentation of foreign antigens to the immune system unlikely candidate for a complex dystonia syndrome
TNXB	tenascin XB	ENST00000375244.3: exon3:c.T1388C: p.V463A	Nonsynonymous SNV	het	novel	0.19	0.95	*TNXB* mediates interactions between cells and the extracellular matrix. Mutations in this only present in heterozygous state and mainly expressed in connective tissue
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	ENST00000360004.5: exon3:c.A379C: p.K127Q	Nonsynonymous SNV	het	recorded once in dbSNP	0.11	0.015	*HLA-DRB1* is involved in the presentation of foreign antigens to the immune system unlikely candidate for a complex dystonia syndrome

HLA Human leucocyte antigen system

Two rare heterozygous variants in the *HLA* locus were present in IV:4. The HLA locus is one of the most complex regions in the human genome due to the extreme levels of polymorphism and linkage disequilibrium. The region contains more than 200 protein-coding genes that play key roles in immune function and defense. The genes encode cell-surface antigen-presenting proteins among other functions.

HLA-A

HLA-A belongs to the *HLA* class I heavy chain molecule in the major histocompatibility complex (MHC). This class I molecule is a heterodimer consisting of a heavy chain and a light chain (beta-2 microglobulin). The heavy chain is anchored in the membrane. Class I molecules play a central role in the immune system by presenting peptides derived from the endoplasmic reticulum lumen. They are expressed in nearly all cells.

HLA-DRB1

HLA-DRB1 is a HLA class II molecule in the major histocompatibility complex (MHC). It encodes a cell surface alpha/beta heterodimeric protein whose function is to present processed foreign antigens to T cells. *HLA-DRB1* is highly polymorphic with 58 known alleles in humans. For the two variants in the HLA region review of coverage data did not reveal any uncovered regions for these genes where a second variant might be present. In any event, the HLA genes were felt to be an unlikely cause of autosomal recessive dystonia.

TNXB

This gene encodes a member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin, which is adhesive. This protein is thought to function in matrix maturation during wound healing. *TNXB* is expressed in brain, most highly in the cerebellum. Defects in *TNXB* are the cause of tenascin-X deficiency (MIM 606408). Tenascin-X deficiency leads to an Ehlers-Danlos-like syndrome characterized by hyperextensible skin, hypermobile joints, and tissue fragility. Review of the coverage of *TNXB* revealed it was complete.

No new variants were revealed when synonymous variants were considered in the regions of linkage, or when the linkage regions for the model ignoring the non-paternity.

6.3.3.5 Sanger sequencing of uncovered regions

A custom script was used to identify exons with a coverage of less than 4 reads within the chromosome 6 and 13 linkage regions which take into account the non-paternity. All coding exons and their flanking intronic region identified by this were sanger sequenced in IV:2. However this did not reveal any rare or novel variants.

6.3.3.6 Copy number variant analysis

Inspection of the OmniExpress data for the affected individuals did not reveal any copy number variation within the regions of linkage. ExomeDepth[275] (Chapter 2) was used to detect CNV calls from exome sequence data, however it did not reveal any rare or novel CNVs within regions of linkage.

6.3.3.7 Exploring the possibility of two diseases in the family

There is a wide clinical spectrum between affected family members. IV:2 and IV:4 both present with severe generalised dystonia and spastic paraparesis. However, IV:3 has only a mild spastic paraparesis. Intra-familial phenotypic variability is readily recognized in many Neurogenetic disorders. However, in this family it is likely that there are multiple loops of consanguinity in preceding generations that are not known about. Therefore, it is possible that the spastic paraparesis and generalised dystonia are separate conditions and result from autosomal recessive mutations in different genes. The occurrence of more than one genetic condition in individuals has been reported several times by WES.[33]

To explore whether a autosomal recessive mutations in another gene has resulted in the generalised dystonia phenotype, regions of homozygosity shared only by IV:2 and IV:4 were identified: Hg19 chr1:216,649,667-223,209,450; chr6:36,725,464-41,527,065, chr7:43,140,000-49,850,666; chr12:50:453,641-51,575,615;

chr15:43,589,078-33,912,633. WES filtering was performed as before however this did not reveal any novel or rare homozygous variants.

6.3.4 Discussion

In this consanguineous kindred with generalised dystonia and spastic paraparesis linkage analysis has identified a ~11.5Mb region on chromosome 6 and a ~3.5Mb region on chromosome 13, in which the causal variant is likely to be located. Neither of these regions has been associated with autosomal recessive generalised dystonia before, suggesting this family has a novel cause for this phenotype.

WES and linkage analysis has identified 3 heterozygous variants within these linkage regions. The function of each of these 3 genes has been well studied. Two of the genes lie in the human major histocompatibility complex (MHC). *HLA-A* is a class I HLA gene and interacts with T-cell receptor molecules, as well as killer immunoglobulin-like receptors expressed on natural killer cells and some T-cells.[333] HLA subtypes are associated with susceptibility to HIV-1,[334] birdshot chorioretinopathy,[335] haemochromatosis,[336] early-onset Alzheimer disease,[337] Stevens Johnson Syndrome,[338] infectious mononucleosis,[339] type 1 diabetes[340] and multiple sclerosis.[341] *HLA-DRB1* a cell surface alpha/beta heterodimeric protein that presents processed foreign antigens to T cells. It is associated with susceptibility to severe malaria,[342] bullous pemphigoid,[343] acute graft-versus-host disease,[344], hepatitis C,[345] HIV-1,[334] multiple sclerosis[341, 346] amongst other diseases. Currently, the two main disease associations with the *HLA* genes are response to infectious disease and autoimmune diseases. There are rare reports of autoimmune disease in association with sporadic forms of dystonia for e.g., antinuclear antibodies and orofacial-cervical dystonia[347] and atypical dystonia in association with antibasal ganglia antibodies.[348] TNXB has been established to mediate interactions between cells and the extracellular matrix. Loss of function mutations cause autosomal recessive Ehlers-Danlos like syndrome, where connective tissue problems resulted in hypermobility, hyperelastic skin and easy bruising.

There is a diverse range of pathways implicated in the pathogenesis of primary dystonia: dopamine signaling (*GCH1*, *TH*, *SPR*, *GNAL*, *TOR1A*), regulation of gene

expression and cell cycle dynamics (*THAP1* and *TAF3*), cell structure (*TUBB4A*, *SGCE*), synaptic dysfunction (*PRRT2*) and neuronal excitability (*ANO3*). Pathways involved in heredodegenerative dystonias include neurometabolic pathways (*HPRT* in Lesch-Nyan disease), mitochondrial metabolism (*TIMM8A* in dystonia deafness syndrome), DNA repair (*AOA1*, *AOA2*, *ATM* in autosomal recessive ataxias with dystonia). Although it is possible that a novel gene for heredodegenerative dystonia could implicate a novel pathway, the putative function of *HLA-A*, *HLA-DRB1* and *TNXB* still seem unlikely candidates based on their function. Additionally, in this consanguineous family, the overwhelming likelihood is that the variant is likely to be identical by state and therefore homozygous.

Therefore, WES and linkage analysis has failed to identify the causal variant in this family. Sanger sequencing of uncovered/poorly covered coding exons in the linked regions has not identified the causal variant. The general technical and analytical reasons for failure of WES and mapping data to identify the causal variant in a family with a Mendelian disorder are as follows:

Technical Limitations:

1. Part or all, of the causative gene is not in the target definition (for example, it is not a known gene, or there is a failure in the bait design). The probes in sequence capture methods are designed based on the sequence information from gene annotation databases such as the consensus coding sequence (CCDS) database and Refseq database; therefore, unknown or yet-to-annotate exons cannot be captured. There may also be a failure in bait design so that an exonic region is not captured. Selectively sequencing the exome- which is, to our knowledge, the most likely region of the genome to contain pathogenic mutations – also excludes noncoding regions. The contribution of mutations in non-coding regions to Mendelian disease has yet to be determined. Finally, it is recognized that microRNAs, promotors and ultra-conserved elements may be associated with disease, reflecting this, commercial kits are increasingly targeting these sequences
2. Inadequate coverage of the region that contains a causal variant. A certain minimum depth of coverage is required for sufficient accuracy of variant

detection; that is, positions or regions in the genome of the individual that are different from the reference human genome sequence. Typically, a minimum coverage of 8-10 reads per base is required for high-confidence detection of a heterozygous single nucleotide variant. Regions with repetitive sequences are more poorly characterized, as repetitive sequences may have prevented inclusion of a probe, or the reads originating from these regions cannot easily be mapped to a single position in the reference genome. Additionally probes may be poorly performing in GC-rich regions. There may be inadequate coverage of the region that contains the causal variant, for example, because of poor capture or poor sequencing

3. The causal variant is covered but not accurately called, for example, in the presence of a small but complex indel. Furthermore, exome sequencing is unable to detect structural variants or chromosomal rearrangements, which are believed to be important for Mendelian disorders.

Analytical Limitations:

4. There may be more than one disease in the family e.g. autosomal recessive spastic paraparesis and autosomal recessive generalised dystonia.
5. A variant may have been excluded from a homozygous region as it is present in a publically available database at an appreciable frequency because it is a false positive variant in that database. Currently, more than 17 million SNPs in the human genome have been documented in dbSNP with a false positive rate of ~15-17%.[59]
6. The causal variant could be synonymous and have been excluded in the variant filtering process.

Future work

As the likely position of the causal variant has been narrowed down to two genomic regions, this family have been submitted for whole-genome sequencing. A fibroblast cell line has been established to facilitate future studies if the causal gene is identified.

6.4 FAMILY 5: WHOLE EXOME SEQUENCING IN A FAMILY WITH AUTOSOMAL RECESSIVE GENERALISED DYSTONIA, SPASTIC PARAPARESIS AND CEREBELLAR SIGNS.

6.4.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

I examined all family members and performed homozygosity mapping and analysed WES data. I performed confirmatory Sanger sequencing and segregation analysis. I wrote the manuscript, which arose from this work, which was published in the journal of Neurology.[349]

6.4.2 BACKGROUND

Prof Bhatia identified a family (Family 5, figure 6.2) with an unusual phenotype consisting of generalised dystonia, spastic paraparesis and cerebellar signs. The mode of inheritance in this family was presumed to be autosomal recessive; this was suggested by the presence of parental consanguinity, two affected siblings with infantile age of onset and phenotypically normal parents. Autozygosity mapping and WES lead to the identification of a homozygous nonsense mutation in *ALS2*. Generalised dystonia and cerebellar signs have not been reported in association with *ALS2*-related disease, therefore this widened the recognized phenotype for this disorder.

Figure 6.2: Pedigree of family 5

6.4.3 MATERIALS AND METHODS

6.4.3.1 Samples

I collected DNA samples from this family. All participants gave written informed consent. The study was approved by the local ethics board.

6.4.3.2 Autozygosity Mapping

Autozygosity mapping was performed for regions >0.5Mb in II:1, II:2 and II:3 as outlined in the Materials and Methods chapter (chapter 2) using HumanOmniExpress beadchips.

6.4.3.3 Whole Exome Sequencing

Individual II:2 was selected for WES in family 5 as she was the most severely affected individual. WES was performed at UCL Institute of Neurology as described in the material and Methods chapter (chapter 2).

6.4.3.4 Variant Filtering

Only homozygous variants within regions of homozygosity shared only by the two affected individuals (II:2 and II:3, figure 6.2) were used for filtering. Synonymous variants and any variant present at a frequency >1% in a range of publically

available databases of sequence variation (dbSNP, 1000 Genomes, Complete Genomic 69 Database and NHLBI Exome Sequencing Project database) was filtered out, in addition to variants present >3 times in our own in-house database of exome sequencing in individuals (n=200) with an unrelated non-neurological disease.

6.4.3.5 Sanger Sequencing

Sanger sequencing was performed for exon 10 of the *ALS2* gene (transcript ID: ENST00000264276) in family 5 and described in Materials and Methods (chapter 2).

6.4.4 RESULTS

6.4.4.1 Phenotypic Characterisation

The index case (Figure 6.2, II:2) is of Bangladeshi descent. Her early motor milestones were normal, however she failed to walk independently, prompting assessment aged 2 years. Examination at this time revealed a mild spastic diplegia with global developmental delay and microcephaly (below 2nd centile). She presented to our center aged 13 years, her examination revealed a few beats of nystagmus, a facetious smile and anathria. There was marked spasticity and contractures in the limbs with dystonic posturing of the hands. Global muscle weakness was present with distal lower limb wasting suggested lower motor neuron involvement. She requires a motorized wheelchair. Subsequently, surgical intervention was required for a rapidly progressive scoliosis. A trial of L-dopa was not helpful.

II:3 achieved normal motor milestones up until 12 months of age when he started toe-walking. Examination aged 7 years revealed microcephaly (2nd-9th centile), nystagmoid jerks and intermittent head titubation. In the upper and lower limbs there was spasticity with clonus and dystonic posturing of the arms and trunk. Gait examination showed truncal sway, suggestive of ataxia. Neck flexors were weak as were proximal and distal muscle groups. Reflexes were pathologically brisk with bilaterally extensor plantar responses. There was a mild scoliosis.

MRI brain imaging in II:2 showed mild lack of white matter bulk and some immaturity of the white matter signal. CSF and an extensive metabolic screen were normal. A muscle biopsy showed angular atrophic fibres with grouping of fast and slow fibres in keeping with a neurogenic component.

6.4.4.2 Autozygosity mapping results

5 regions of homozygosity were identified, totaling almost ~54Mb, shared only by the affected individuals II:2 and II:3 (Table 6.4).

Table 6.4: Homozygous regions >0.5Mb concordant in II:2 and II:3

Chromosome	Start	End	Size (bp)
2	171,210,691	189,879,843	18,669,152
2	191,979,359	207,890,135	15,910,776
2	223,594,372	232,343,950	8,749,578
7	0	5,452,756	5,452,756
12	0	5,811,469	5,811,469

6.4.4.3 Whole Exome Sequencing

Alignment metrics for WES in II:2 revealed good coverage (table 6.5)

Table 6.5: WES Metrics for II:2, Family 5

	Family 5 II:2
Total number of reads	47,076,032
Non-duplicated reads	35,872,402
Reads aligned to target	85%
Mean Target Coverage	28
% Target covered by 2 reads	92%
% Target covered by 10 reads	80%
% Target covered by 20 reads	58%
Total number of variants	20,485

6.4.4.4 Variant filtering in regions of shared homozygosity

Following WES, filtering of variants within the regions of homozygosity shared only by II:2 and II:3, revealed a single novel nonsense homozygous variant remained: c.G2002T p.G668X in *ALS2* (Transcript: ENST00000264276).

6.4.4.5 Sanger sequencing confirmation of candidate variant

Sanger sequencing confirmed the variant was present in the homozygous state in both affected siblings (II:2 and II:3). Their parents (I:1 and I:2) are heterozygous carriers and the unaffected sibling (II:1) does not carry the mutation.

6.4.5 DISCUSSION

In this small consanguineous family WES and autozygosity mapping has revealed a novel homozygous nonsense mutation in the gene *ALS2,* as the cause of the generalised dystonia, spastic paraparesis and cerebellar signs in this family. The affected individuals shared >54Mb of homozygosity meaning that Sanger sequencing of candidate genes in the region would not have been feasible in this family. This case demonstrates the power of WES applied with mapping strategies. Subsequently, a collaborator identified a homozygous frameshift mutation in *ALS2* in another family with generalised dystonia and a spastic paraparesis confirming that the phenotype of *ALS2* mutations is wider than first thought and includes generalised dystonia.[349]

Mutations in the *ALS2* gene (OMIM 606352) cause autosomal recessive motor neuron diseases, including juvenile-onset amyotrophic lateral sclerosis (JALS),[350] juvenile-onset primary lateral sclerosis (JPLS) and infantile-onset ascending hereditary spastic paraplegia (IAHSP).[351, 352] In JALS both upper and lower motor neurons are affected, whereas neurodegeneration only involves upper motor neurons in JPLS and IAHSP. Despite these differences in neuropathology, almost all mutations in *ALS2* described to date result in a clinical phenotype of infantile onset limb and facial weakness, accompanied by bulbar symptoms, which generally progress to paraplegia during childhood. Rarely patients with JALS have been reported with lower motor neuron involvement.[353] *ALS2* mutations are distributed widely across the entire coding sequence and mostly result in loss of

protein function.[353-355] No clinical diagnostic criteria exist for *ALS2* related diseases. However, there is a significant overlap in the clinical features of both affected family members in family 5 with cases already described in the literature. Both affected family members have an infantile onset disease with clinical involvement of both upper and lower motor neurons and bulbar involvement, in keeping with JALS.

The manuscript detailing Family 5 is the first report of generalised dystonia in association with *ALS2* related disease.[349] Other novel clinical findings include microcephaly and cerebellar signs, however currently it is not clear if these are features of *ALS2*-related disease. Interestingly, in a wild-type mice model of the disease, ALS2 was highly expressed in the granular and Purkinje layers of the cerebellum and *ALS2*-null mice have been found to develop an age-dependent slowly progressive loss of cerebellar Purkinje cells, suggesting that cerebellar signs may well be part of the extended phenotype.[356] The ALS2 protein binds to a small GTPase RAB5 and functions as a guanine nucleotide exchange factor (GEF) for RAB5 and plays a role in intracellular endosomal trafficking,[356] highlighting that this may be an important biological pathway or mechanism through which dystonia may occur. This report adds to the growing literature of widening of the phenotypic spectrum of genetic disorders using next-generation sequencing. We propose that the *ALS2* gene should be screened for mutations in patients who present with a similar phenotype.

6.5 FAMILY 6: WHOLE EXOME SEQUENCING IN A CONSANGUINEOUS FAMILY WITH GENERALISED DYSTONIA

6.5.1 STATEMENT OF CONTRIBUTION TO THIS RESEARCH

I examined all family members, performed autozygosity mapping and analysed the WES data. I performed confirmatory Sanger sequencing and segregation analysis.

6.5.2 BACKGROUND

I identified a patient from the Neurogenetics database who had a provisional diagnosis of cerebral palsy. The differential diagnosis included an autosomal recessive generalised dystonia, a genetic disorder had been suggested by the presence of parental consanguinity and infantile onset of symptoms. However since she was the only affected family member, there was uncertainty as to the correct diagnosis. A combination of WES and autozygosity mapping was used to explore the possibility of a genetic aetiology. This identified a novel, homozygous, missense mutation in *GCDH*, mutations in which cause Glutaric-aciduria type 1. The patient was subsequently reviewed by the Metabolic disease team at NHNN and underwent metabolic tests which confirmed the diagnosis of Glutaric-aciduria type 1.

Figure 6.3: Pedigree of family 6

6.5.3 METHODS

6.5.3.1 Samples

I collected DNA samples from this family. All participants gave written informed consent. The study was approved by the local ethics board.

6.5.3.2 Phenotypic Characterisation

Family 6 (Figure 6.3) were examined by a movement disorder specialist Dr. Patricia Limousin and myself.

6.5.3.3 Autozygosity Mapping

Autozygosity mapping for regions >0.5Mb was performed on all siblings (IV:1-IV:3, figure 6.3) using a Human Omni express beadchip as discussed in Materials and Methods (chapter 2).

6.5.3.4 Whole exome sequencing

WES was performed in IV:1 at UCL Institute of Neurology using the Illumina Truseq Exome Enrichment kit (62Mb target).

6.5.3.5 Variant Filtering

Only homozygous variants within regions of homozygosity were used for filtering. I filtered out synonymous variants and any variant present in a range of publically available databases of sequence variation above a frequency of 1% (dbSNP, 1000 Genomes, Complete Genomic 69 Database and NHLBI Exome Sequencing Project database), as well as variants that were present more than 3 times in our own in-house database of exome sequencing of individuals (n=200) with an unrelated non-neurological disease.

6.5.3.6 Sanger Sequencing

Sanger sequencing was performed for exon 6 of the *GCDH* gene (transcript ID: ENST00000222214) in family 6 to confirm the putative causal variant and to determine segregation with the disease in the family.

6.5.4 RESULTS

6.5.4.1 Phenotypic Characterisation

The index case (Figure 6.3, IV:1) is of Lebanese descent. The pregnancy with IV:1 was complicated by oligohydramnios and she was born by caesarean section 3 months premature. She achieved normal motor milestones. She was able to stand whilst holding onto objects at the age of 9 months. At 10 months of age, she developed an ear infection followed a day or so later by an acute neurological deterioration. She had several seizures with eye deviation, during which she became rigid and unresponsive. She was admitted to intensive care for 1 month. On discharge from ITU, her family noticed she had lost head control and the ability to sit and stand and required a wheelchair. Subsequently, she developed prominent orolingual dystonia, which improved somewhat with Synacthen. Her neurological condition remained static for the next 7 years. Aged 8 years, she developed slowly progressive involuntary limb movements consisting of dystonia, chorea and ballism. The movements would keep her awake all night and she lost purposeful movements in the upper limbs. She was treated with Madopar 250mg tds, tetrabenazine 25mg bd, baclofen 25mg tds, tizanidine 2mg nocte with mild clinical benefit. Due to the poor response to medical therapy she had bilateral globus pallidus pars interna (GPi) deep brain stimulation (DBS) aged 19 years. This initially helped a little with ballism and painful spasms in the right arm. However this effect soon disappeared and the patient had a capsular response with difficulty swallowing. Despite several trials of adjusting the stimulator, DBS was felt to be exacerbating the involuntary movements and was stopped 2 years after implantation.

Examination of IV:1 aged 18 years revealed severe mobile generalised dystonia with prominent bulbar and lingual involvement. She had a scoliosis and flexion contractures of the limbs. The involuntary movements were particularly severe in the left arm with dystonia, ballism and myoclonus. There was no spasticity at rest but she had numerous painful spasms. Reflexes were normal, the left plantar response was extensor and right plantar response flexor. Currently, she is wheelchair bound and is anarthric. The rest of the neurological examination was normal and there is no macrocephaly. MRI brain showed putaminal hyperintensity in the lateral aspects bilaterally indicating striatal necrosis. Routine bloods, ammonia, acanthocytes, copper studies, urinary and plasma organic acids, long chain fatty acids and white cell enzymes were normal. Genetic testing for *DYT1*

and *PANK2* mutations were negative. EMG and nerve conduction studies were normal.

Neurological examination of IV:2, IV:3, III:1 and III:2 was normal.

6.5.4.2 Autozygosity Mapping Results

33 regions of potential regions of autozygosity, shared by IV:1 alone were identified, totaling > 49Mb (table 6.6).

Table 6.6: Homozygous regions >0.5Mb present only in IV:1

Chromosome	Start	End	Size (bp)
1	190,385,304	191,424,963	1,040
1	187,840,501	188,624,613	784
1	113,338,649	113,914,390	576
2	87,430,727	90,240,473	2,810
3	41,216,488	41,822,054	606
7	22,155,376	28,376,691	6,221
7	139,163,463	140,254,890	1,091
7	84,010,005	84,866,308	856
7	145,928,429	146,516,749	588
7	1,806,723	2,346,115	539
8	47,757,290	49,047,878	1,291
8	33,020,810	33,994,378	974
10	15,629,360	18,302,192	2,673
10	23,661,291	24,693,878	1,033
11	46,393,757	47,303,275	910
11	22,536,864	23,125,291	910
11	116,518,960	117,151,412	632
13	19,058,717	22,844,305	3,786
13	89,510,330	90,196,888	687
13	113,461,004	114,092,549	632
14	88,765,833	89,438,126	672

14	68,208,677	68,843,605	635
14	81,933,447	82,551,186	618
14	75,159,007	75,750,200	591
15	55,226,861	56,362,968	1,136
15	52,100,215	52,712,988	613
18	18,556,505	22,765,009	4,209
18	13,600,497	14,381,966	781
18	25,728,693	26,264,471	536
20	9,077,491	16,248,453	7,171
21	47,557,074	48,100,155	543
22	48,281,630	49,911,182	1,630
22	46,362,396	47,940,908	1,579

6.5.4.3 Whole exome sequencing

Alignment metrics for WES in IV:1 revealed good coverage (table 6.7)

Table 6.7: Summary metrics for WES

	Family 6B IV:1
Total number of reads	131,717,454
Non-duplicated reads	82,829,037
Reads aligned to target	68,562,726
Mean Target Coverage	65
% Target covered by 2 reads	96
% Target covered by 10 reads	91
% Target covered by 20 reads	85
Total number of variants	22,191

After filtering of variants within the regions of homozygosity only present in individual IV:1, 7 homozygous novel variants remained (see table 6.8). The variant in the gene *GCDH,* was of immediate significance as it was associated with a neurometabolic disorder: autosomal recessive Glutaric Aciduria type 1. The variant was predicted to be pathogenic using two in-silico prediction software prediction

programs. The amino acid at which the substitution takes place is highly conserved with a GERP score of 5.14. The variant c.368A>G p.Y123C (transcript ID ENST00000222214) was confirmed by sanger sequencing to be homozygous in IV:1, IV:2 is a heterozygote carrier and IV:3 does not carry the mutation. Their parents (III:1 and III:2) are heterozygous carriers.

In view of the genetic results, further metabolic testing was performed to explore a diagnosis of Glutaric aciduria type I caused by mutations in *GCDH*. This again showed a normal amino acid profile in the plasma. Urine organic acids were normal, glutaric acid and 3-hydroxyglutaric acid were not detected on three separate occasions. Plasma carnitine profile showed a low plasma carnitine at 25umol/L (26-62) with a free carnitine of 22umol/L (22-50) and raised plasma glutarylcarnitine at 0.55umol/L (normal range <0.08), in keeping with a diagnosis of glutaric aciduria type 1. The patient was given carnitine supplementation, advised on a high carbohydrate and low protein diet and given details of an oral emergency regime, in the event of an intercurrent illness.

Table 6.8: showing the remaining variants following exome variant filtering

Gene Symbol	Full Gene Name	Gene Function	Variant Annotation	Exonic Function	Call	Present in Publically Available Databases	SIFT Score	Poly Phen2 Score	GERP Score
CAPN8	calpain 8	Involved in membrane trafficking in the gastric surface mucus cells.	uc009xee.2:c.T1682A:p.I561N	nonsynonymous	hom	Novel	0	0.84	NA
SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	Sulfotransferase enzymes catalyze the sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds.	uc003hen.2:c.C772T:p.R258C	nonsynonymous	hom	Novel	0	1	1.58
AGAP8	ArfGAP with GTPase domain, ankyrin repeat and PH domain 8	Putative GTPase-activating protein	uc009xnv.2:c.C946G:p.R316G	nonsynonymous	hom	Novel	0	NA	NA
C11orf82	Chromosome 11 open reading frame 82	Acts as an anti-apoptotic factor and its absence increases cell death under normal and stress conditions.	uc010rss.1:c.C278T:p.S93L	nonsynonymous	hom	Novel	0	0.983	4.3

ERCC4	Excision repair cross-complementing rodent repair deficiency, complementation group 4	This complex is a structure specific DNA repair endonuclease. Defects in this gene are a cause of Xeroderma Pigmentosum	uc002dce.2:c.G335C:p.R112T	nonsynonymous	hom	Novel	0	0.996	4.98
BFAR	bifunctional apoptosis regulator	Apoptosis regulator. Has anti-apoptotic activity, both for apoptosis triggered via death-receptors and via mitochondrial factors	uc002dcm.2:c.A410G:p.N137S	nonsynonymous	hom	Novel	0	0.309	4.78
GCDH	glutaryl-CoA dehydrogenase	The protein catalyzes the oxidative decarboxylation of glutaryl-CoA to crotonyl-CoA and CO(2). Defects in GCDH are the cause of glutaric aciduria type 1, an autosomal recessive metabolic disorder characterized by progressive dystonia and athetosis due to gliosis and neuronal loss in the basal ganglia	uc010xms.1:c.A269G:p.Y90C	nonsynonymous	hom	Novel	0	1	5.14

6.10 DISCUSSION

Mutations in *GCDH* cause autosomal recessive Glutaric Aciduria Type 1 (GA-1) (MIM 231670) an inborn error of lysine, hydroxylisine, and tryptophan metabolism caused by deficiency of glutaryl-CoA dehydrogenase, an enzyme found in the mitochondrial matrix and converts glutaryl-CoA to crotonyl-CoA. The combined worldwide frequency of GA-1 based on newborn screening is 1:100,000 infants.[357] The disorder is characterized by macrocephaly at birth or shortly after. Patients are vulnerable to striatal damage resulting in severe generalised dystonia between 6 and 18 months of age,[358] usually during an intercurrent febrile illness. It is still unclear what causes this selective sensitivity. Patients remain severely disabled after the initial decompensation and will not be able to walk. MRI imaging typically shows increased signal intensity on T2-weighted imaging. The metabolic hallmark of GA-1 is combined elevation of glutaric acid, 3-hydroxyglutaric acid and glutarylcarnitine (C5-DC) in urine, plasma and CSF, but in a group of patients ("low excreters") metabolic screening may show (intermittently) normal results. This phenomenon is well demonstrated by our patient who had normal glutaric acid and 3-hydroxyglutaric acid, despite repeated determinations of organic acids. The diagnosis in our patient was confirmed by the presence of raised glutarylcarnitine. Identification of patients with *GCDH* mutations is important since dietary intervention (low protein diet, diet restricted in lysine/tryptophan, supplementation with carnitine/riboflavin) prevents neurological damage in the majority of patients.[359]

The novel *GCDH* missense mutation (p.Y123C) in our patient, is highly likely to be pathogenic for a number of reasons. It is absent from publically available databases, it segregates with disease in the family, it is predicted to be damaging by 5 in-silico prediction programs (SIFT, Align GVGD, polyphen2, mutation taster and panther), it is a non-conservative amino acid substitution at a highly conserved residue within a functional domain of GCDH and there are reported pathogenic missense mutations at both adjacent codons.[360, 361]

IV:1's clinical presentation broadly fits with the phenotype described for GA-1, her development was normal until an intercurrent febrile illness aged 9 months after

which she developed acute onset seizures followed by generalised dystonia. She has imaging features that are consistent with a diagnosis of GA-1 and her metabolic investigations are also supportive. However, the progressive movement disorder in IV:1 is atypical in GA-1 as commonly patients with GA-1 have a static neurological deficit. Both IV:1 and a recently reported case demonstrate that a progressive movement disorder may be a manifestation of GA-1, even without recurrent metabolic decompensations.[362] DBS has only been reported in one other patient with GA-1, who experienced improvement in dystonia and functionality following unilateral GPi DBS for dystonia.[363] Our patient had little clinical improvement with bilateral GPi DBS, however further studies with larger numbers of patients will be needed to determine if GPi DBS is helpful in patients with GA-1 related movement disorders. Early identification of GA-1 is important as dietary intervention and advice on preventative measures during intercurrent infections reduces morbidity and mortality. Genetic testing of affected patients siblings may identify at risk individuals in whom neurological damage can be prevented.

This case has demonstrated how a combination of autozygosity mapping and WES has been able to correctly identify the underlying cause of a rare dystonia syndrome in a patient that presented a diagnostic challenge. It has facilitated genetic counseling in the family, revising the initial diagnosis of athetoid cerebral palsy and provided an option for predictive testing in other family members, which is important since dietary intervention can prevent neurological damage. It also highlights the possible potential of exome sequencing in the genetic diagnostics setting, where many genes can be sequenced simultaneously and in genetically heterogeneous disorders such as inherited neuropathies this may be more cost effective than sequential Sanger sequencing of genes. Additionally, this case highlights that findings from exome sequencing may influence therapeutic management decisions. This has been reported for other families in the literature including giving high-dose riboflavin supplementation to a patient with mutations in *ACAD9*,[364] making a definitive decision about allogenic stem-cell transplant in a patient with a mutation in X-linked inhibitor of apoptosis gene,[365] starting patients with alpha-methylacyl-CoA racemase deficiency on phytanic and pristantic acid restricted diets,[366] management of early infections and cancer screening in patients

with *ATM* mutations[367] and initiation of riboflavin therapy in a patient with Brown-Vialetto-Van Laere syndrome.[368]

CHAPTER 7. CONCLUSIONS AND FUTURE DIRECTIONS

My PhD research focused on the use of WES to dissect the aetiologies of Mendelian forms of Parkinson's disease and Dystonia. Through the work on the families presented herein in this thesis I have been able to appreciate the strengths, pitfalls and challenges of WES. I have demonstrated successes in identifying the causal variant in autosomal dominant Parkinson's disease (chapter 4), autosomal recessive generalised parkinsonism syndromes (chapter 5), and autosomal recessive generalised dystonia (chapter 6). I have also presented examples of failures of WES in autosomal recessive early-onset Parkinson's disease (chapter 5), and autosomal recessive generalised dystonia (chapter 6). Additionally I used Sanger sequencing to confirm that mutations in *VPS35* are a cause of autosomal dominant Parkinson's disease (chapter 3). I will briefly summarize the major findings of these projects and highlight out future directions:

VPS35 Screening in a Parkinson's Disease Cohort (Chapter 3)

VPS35 was the first Mendelian gene to be shown to cause PD using next-generation sequencing technologies. As with all genes identified by WES, it is important that replication studies are performed to confirm the finding in ethnically diverse populations, and to determine whether other mutations that those reported in the initial study are pathogenic. We screened a large UK PD series for mutations in *VPS35.* We identified one large kindred with the already reported pathogenic *VPS35* mutation. The kindred had highly penetrant autosomal dominant PD, which was similar to idiopathic PD.

Whole Exome Sequencing in Autosomal Dominant Parkinson's Disease (Chapter 4)

In this family I identified a novel mutation in *DCTN1* as the cause of the autosomal dominant Parkinson's disease. This family serves as a good example of how WES, without the use of linkage analysis, was able to identify the cause of disease in a family with autosomal dominant neurodegenerative disease. Furthermore it widens the recognized phenotype associated with Perry Syndrome.

Whole exome Sequencing and Autozygosity Mapping in Autosomal Recessive Parkinsonism Disorders (Chapter 5)

In family 1, I was unable to identify the causal variant for autosomal recessive Parkinson's Disease in a consanguineous kindred with PD, despite the use of autozygosity mapping which significantly reduced the genomic regions to be considered. Several candidate variants were identified, however, screening of a large series of early-onset PD patients did not reveal another patient with homozygous or compound heterozygous mutations in the exon in which the variant occurred in the candidate genes. This family demonstrate one of the limitations of WES, namely that in small families WES and mapping strategies may still leave several candidate variants and that identifying the pathogenic variant may require large scale screening, which will be best done using targeted NGS panels. It is possible that the causal variant was not identified due to a number of technical or analytical reasons.

In family 2, I was able to identify the cause of a complicated parkinsonian syndrome in a singleton case using a combination of WES and autozygosity mapping. This demonstrated the utility of WES as a timely method to simultaneously screen several genes that can cause juvenile onset complex parkinsonian syndromes.

In family 3, the causal variant in a small non-consanguineous family with early-onset PD was not confidently identified using WES and genome-wide linkage analysis. The finding of a novel, predicted pathogenic variant in *LRRK2,* which segregated with the disease in the family, brought into question the mode of inheritance in the family.

Whole exome sequencing in autosomal recessive generalised dystonia (chapter 6)

In family 4, two genomic regions were shown to be linked with a complex phenotype of autosomal recessive generalised dystonia and spastic paraparesis. WES several heterozygous variants within the linkage regions, however, none were felt to be causal. The analysis in this family was complicated by non-paternity

and the fact that the phenotype could result from two separate genetic aetiologies. The family will be a good candidate for whole genome sequencing.

In family 5, WES and autozygosity mapping revealed a novel nonsense *ALS2* mutation as the cause of generalised dystonia and upper and lower motor neuron signs in two siblings. This work widens the known phenotypic spectrum in association with *ALS2* related disease.

In family 6, autozygosity mapping and WES in a consanguineous family revealed a novel *GCDH* mutation as the cause of progressive generalised dystonia in singleton case. This revised the original diagnosis in the index case and furthermore changed the management of the patient with institution of dietary measures to prevent further neurological relapse.

Future Directions

In two of the families presented in this thesis, WES has widened the known phenotype associated with a Mendelian genetic disorder (families with *DCTN1* and *ALS2* mutations). Other examples of widened phenotype spectrum using WES include the identification of *VCP* mutations and TDP-43 pathology in a family with autosomal dominantly inherited amyotrophic lateral sclerosis.[34] Prior to this *VCP* mutations had only been associated with families with Inclusion body myopathy, Paget's disease and Frontotemporal dementia. Mutational screening of *VCP* in additional familial ALS samples suggested that *VCP* mutations may account for ~1-2% of familial ALS. This work implicated defects in the ubiquitin/protein degradation pathway in motor neuron degeneration for the first time. These examples show that broadening the phenotype associated with mutations using next-generation sequencing has the potential to provide information on the aetiological basis of disorders by uniting what is known about the biological underpinnings of apparently unrelated disorders into a single model.

Most of the successes of WES have been in studies on rare Mendelian disorders, caused by variants in high penetrance segregating in families. In the near future, the power of WES used together with mapping strategies should enable the

identification of genes underlying a large fraction of Mendelian disorders that are currently unsolved.

The work on family 1 and 3 illustrates a problem of narrowing the list of candidate variants from WES data in families from ethnicities that are not well represented in control population databases such as Exome Variant Server. Results from the 1000 Genome project have revealed the differences in the numbers of known and novel variants in Europeans, Asians and Africans.[369] In assessing the effects of rare variants, it is important to consider the allele frequency of the variant in ethnically matched individuals in relation to the disease frequency in that particular ethnic group. For patients from ethnicities underrepresented in control databases, filtering of candidate pathogenic variants for a rare Mendelian disorder on the basis of their absence from public databases is insufficient, sequencing of a sufficient number of unaffected individuals from the same ethnic group is needed to determine if a variant is a rare benign polymorphism or potentially causal. Expansion of these control databases to include data from different ethnic groups will greatly aid gene discovery.

It is not clear yet what proportion of Mendelian disorders are caused by mutations in the non-coding part of the genome, as previous estimates are likely to have been skewed upward by protein-centric studies. The next challenge in Mendelian disease research will be to systematically study the role of variation in the non-coding part of our genome in health and disease. Estimates of the success of WES are in the large part absent from the literature as negative studies tend not to be published, however, at a large groups working on rare Mendelian disorders have estimated their success rate at 60-80%.[370]

It is likely that WGS will be able to reveal the genetic cause of families that have not been solved with WES. This is because WGS does not have the problem of 'missing exons' as a result of incomplete capture, variants in highly conserved non-coding regions can be readily explored for unexplained cases and WGS covers the non-coding genome (intronic and intergenic) and allows better detection of CNVs.

As the cost of next-generation sequencing continues to fall, the field will move from whole-exome to whole-genome sequencing. However, taking advantage of these more comprehensive data for disease gene discovery and molecular diagnostics in patients crucially depends on the development of analytical strategies for making sense of non-coding variation. Renewed effort in several areas is likely to help additional gene identification by exome and whole-genome sequencing. Firstly, there needs to be proper curation of phenotypes, particularly in the context of Mendelian disorders. Secondly, there is a need for improved technical, statistical and bioinformatics methods for reducing the rate of false-positive and false negative variant calls; calling indels; prioritizing candidate causal variants; and predicting and annotating the potential functional impact for disease gene discovery or molecular diagnostics. Identifying novel genes or candidate genes for autosomal recessive and autosomal disorders is a realistic goal presently, however narrowing down candidate gene lists is likely to require unprecedented cooperation and coordination in Neurogenetics, with the development of large consortia groups for data sharing and with sophisticated approaches to conduct candidate prioritization and screening in replication cohorts.

REFERENCES

1. Mayeux R, Marder K, Cote LJ, et al. The frequency of idiopathic Parkinson's disease by age, ethnic group, and sex in northern Manhattan, 1988-1993. Am J Epidemiol. 1995 Oct 15;142(8):820-7.
2. Vossius C, Nilsen OB, Larsen JP. Parkinson's disease and nursing home placement: the economic impact of the need for care. Eur J Neurol. 2009 Feb;16(2):194-200.
3. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7.
4. Moore GE. Cramming more components onto integrated circuits (Reprinted from Electronics, pg 114-117, April 19, 1965). P Ieee. 1998 Jan;86(1):82-5.
5. Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. J Mol Diagn. 2010 Sep;12(5):539-51.
6. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov 6;456(7218):53-9.
7. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387-402.
8. Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989 Sep 8;245(4922):1066-73.
9. Rommens JM, Iannuzzi MC, Kerem B, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science. 1989 Sep 8;245(4922):1059-65.
10. Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989 Sep 8;245(4922):1073-80.
11. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Available from: http://omim.org/.
12. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009 Sep 10;461(7261):272-6.
13. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010 Jan;42(1):30-5.
14. Pruitt KD, Harrow J, Harte RA, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res. 2009 Jul;19(7):1316-23.
15. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010 Sep;42(9):790-3.

16. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011 Nov;12(11):745-55.
17. Meyer E, Ricketts C, Morgan NV, et al. Mutations in FLVCR2 are associated with proliferative vasculopathy and hydranencephaly-hydrocephaly syndrome (Fowler syndrome). Am J Hum Genet. 2010 Mar 12;86(3):471-8.
18. Bilguvar K, Ozturk AK, Louvi A, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature. 2010 Sep 9;467(7312):207-10.
19. http://www.ncbi.nlm.nih.gov/projects/SNP/. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD) [cited 2012].
20. 1000 Genomes Project. [cited 2012]; Available from: http://www.1000genomes.org/.
21. NHLBI Exome Sequencing Project (ESP) [cited 2012]; Available from: http://evs.gs.washington.edu/EVS/.
22. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003 Mar;33 Suppl:228-37.
23. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum Mol Genet. 2010 Oct 15;19(R2):R125-30.
24. Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform. 2011 Jan;12(1):41-51.
25. Cooper GM, Goode DL, Ng SB, et al. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat Methods. 2010 Apr;7(4):250-1.
26. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.
27. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9.
28. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010 Aug;7(8):575-6.
29. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res. 2005 Jul;15(7):978-86.
30. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet. 2011 Apr 8;88(4):440-9.
31. Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. Clin Chem. 2004 Nov;50(11):1974-8.
32. Xu B, Roos JL, Dexheimer P, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet. 2011 Sep;43(9):864-8.
33. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013 Oct 17;369(16):1502-11.
34. Johnson JO, Mandrioli J, Benatar M, et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron. 2010 Dec 9;68(5):857-64.

35. Bras J, Verloes A, Schneider SA, Mole SE, Guerreiro RJ. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. Hum Mol Genet. 2012 Jun 15;21(12):2646-50.
36. Zimprich A, Benet-Pages A, Struhal W, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet. 2011 Jul 15;89(1):168-75.
37. Vilarino-Guell C, Wider C, Ross OA, et al. VPS35 mutations in Parkinson disease. Am J Hum Genet. 2011 Jul 15;89(1):162-7.
38. Charlesworth G, Plagnol V, Holmstrom KM, et al. Mutations in ANO3 cause dominant craniocervical dystonia: ion channel implicated in pathogenesis. Am J Hum Genet. 2012 Dec 7;91(6):1041-50.
39. Fuchs T, Saunders-Pullman R, Masuho I, et al. Mutations in GNAL cause primary torsion dystonia. Nat Genet. 2013 Jan;45(1):88-92.
40. Chen WJ, Lin Y, Xiong ZQ, et al. Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. Nat Genet. 2011 Dec;43(12):1252-5.
41. Crow JF. The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet. 2000 Oct;1(1):40-7.
42. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007 Aug;8(8):610-8.
43. Hoischen A, van Bon BW, Gilissen C, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet. 2010 Jun;42(6):483-5.
44. Hoischen A, van Bon BW, Rodriguez-Santiago B, et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. Nat Genet. 2011 Aug;43(8):729-31.
45. Lindhurst MJ, Sapp JC, Teer JK, et al. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med. 2011 Aug 18;365(7):611-9.
46. Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. Nat Genet. 2010 Dec;42(12):1109-12.
47. O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011 Jun;43(6):585-9.
48. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012 May 10;485(7397):237-41.
49. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012 May 10;485(7397):242-5.
50. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012 May 10;485(7397):246-50.
51. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012 Apr 26;74(2):285-99.
52. Girard SL, Gauthier J, Noreau A, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet. 2011 Sep;43(9):860-3.
53. McClellan J, King MC. Genomic analysis of mental illness: a changing landscape. JAMA. 2010 Jun 23;303(24):2523-4.

54. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013 Jan 10;368(2):117-27.
55. Guerreiro RJ, Lohmann E, Bras JM, et al. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. JAMA Neurol. 2013 Jan;70(1):78-84.
56. Jonsson T, Stefansson K. TREM2 and neurodegenerative disease. N Engl J Med. 2013 Oct 17;369(16):1568-9.
57. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257-68.
58. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20;72(2):245-56.
59. Day IN. dbSNP in the detail and copy number complexities. Hum Mutat. 2010 Jan;31(1):2-4.
60. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science. 1987 Jun 19;236(4808):1567-70.
61. Gibbs JR, Singleton A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. PLoS Genet. 2006 Oct 6;2(10):e150.
62. Benayoun L, Spiegel R, Auslender N, et al. Genetic heterogeneity in two consanguineous families segregating early onset retinal degeneration: the pitfalls of homozygosity mapping. Am J Med Genet A. 2009 Feb 15;149A(4):650-6.
63. Lezirovitz K, Pardono E, de Mello Auricchio MT, et al. Unexpected genetic heterogeneity in a large consanguineous Brazilian pedigree presenting deafness. Eur J Hum Genet. 2008 Jan;16(1):89-96.
64. Frishberg Y, Ben-Neriah Z, Suvanto M, et al. Misleading findings of homozygosity mapping resulting from three novel mutations in NPHS1 encoding nephrin in a highly inbred community. Genet Med. 2007 Mar;9(3):180-4.
65. Ducroq D, Shalev S, Habib A, Munnich A, Kaplan J, Rozet JM. Three different ABCA4 mutations in the same large family with several consanguineous loops affected with autosomal recessive cone-rod dystrophy. Eur J Hum Genet. 2006 Dec;14(12):1269-73.
66. Laurier V, Stoetzel C, Muller J, et al. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. Eur J Hum Genet. 2006 Nov;14(11):1195-203.
67. Miano MG, Jacobson SG, Carothers A, et al. Pitfalls in homozygosity mapping. Am J Hum Genet. 2000 Nov;67(5):1348-51.
68. Pannain S, Weiss RE, Jackson CE, et al. Two different mutations in the thyroid peroxidase gene of a large inbred Amish kindred: power and limits of homozygosity mapping. J Clin Endocrinol Metab. 1999 Mar;84(3):1061-71.
69. Genin E, Todorov AA, Clerget-Darpoux F. Optimization of genome search strategies for homozygosity mapping: influence of marker spacing on power and threshold criteria for identification of candidate regions. Ann Hum Genet. 1998 Sep;62(Pt 5):419-29.

70. Paisan-Ruiz C, Bhatia KP, Li A, et al. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. Ann Neurol. 2009 Jan;65(1):19-23.
71. Krebs CE, Karkheiran S, Powell JC, et al. The Sac1 Domain of SYNJ1 Identified Mutated in a Family with Early-Onset Progressive Parkinsonism with Generalized Seizures. Hum Mutat. 2013 Sep;34(9):1200-7.
72. Quadri M, Fang M, Picillo M, et al. Mutation in the SYNJ1 Gene Associated with Autosomal Recessive, Early-Onset Parkinsonism. Hum Mutat. 2013 Sep;34(9):1208-15.
73. Edvardson S, Cinnamon Y, Ta-Shma A, et al. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. PLoS One. 2012;7(5):e36458.
74. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955 Sep;7(3):277-318.
75. Fahn S. Description of Parkinson's disease as a clinical syndrome. Ann N Y Acad Sci. 2003 Jun;991:1-14.
76. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry. 1992 Mar;55(3):181-4.
77. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. Arch Neurol. 1999 Jan;56(1):33-9.
78. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry. 1988 Jun;51(6):745-52.
79. Braak H, Del Tredici K, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging. 2003 Jan 1;24(2):197-211.
80. Allam MF, Campbell MJ, Hofman A, Del Castillo AS, Fernandez-Crehuet Navajas R. Smoking and Parkinson's disease: systematic review of prospective studies. Mov Disord. 2004 Jun;19(6):614-21.
81. Elbaz A, Tranchant C. Epidemiologic studies of environmental exposures in Parkinson's disease. J Neurol Sci. 2007 Nov 15;262(1-2):37-44.
82. Thacker EL, Chen H, Patel AV, et al. Recreational physical activity and risk of Parkinson's disease. Mov Disord. 2008 Jan;23(1):69-74.
83. Tanner CM, Aston DA. Epidemiology of Parkinson's disease and akinetic syndromes. Curr Opin Neurol. 2000 Aug;13(4):427-30.
84. Duvoisin RC, Eldridge R, Williams A, Nutt J, Calne D. Twin study of Parkinson disease. Neurology. 1981 Jan;31(1):77-80.
85. Marttila RJ, Kaprio J, Koskenvuo M, Rinne UK. Parkinson's disease in a nationwide twin cohort. Neurology. 1988 Aug;38(8):1217-9.
86. Ward CD, Duvoisin RC, Ince SE, Nutt JD, Eldridge R, Calne DB. Parkinson's disease in 65 pairs of twins and in a set of quadruplets. Neurology. 1983 Jul;33(7):815-24.
87. Laihinen A, Ruottinen H, Rinne JO, et al. Risk for Parkinson's disease: twin studies for the detection of asymptomatic subjects using [18F]6-fluorodopa PET. J Neurol. 2000 Apr;247 Suppl 2:II110-3.
88. Piccini P, Burn DJ, Ceravolo R, Maraganore D, Brooks DJ. The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. Ann Neurol. 1999 May;45(5):577-82.

89. Polymeropoulos MH, Higgins JJ, Golbe LI, et al. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. Science. 1996 Nov 15;274(5290):1197-9.
90. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science. 1997 Jun 27;276(5321):2045-7.
91. Kruger R, Kuhn W, Muller T, et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet. 1998 Feb;18(2):106-8.
92. Zarranz JJ, Alegre J, Gomez-Esteban JC, et al. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol. 2004 Feb;55(2):164-73.
93. Golbe LI, Di Iorio G, Bonavita V, Miller DC, Duvoisin RC. A large kindred with autosomal dominant Parkinson's disease. Ann Neurol. 1990 Mar;27(3):276-82.
94. Proukakis C, Dudzik CG, Brier T, et al. A novel alpha-synuclein missense mutation in Parkinson disease. Neurology. 2013 Mar 12;80(11):1062-4.
95. Appel-Cresswell S, Vilarino-Guell C, Encarnacion M, et al. Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease. Mov Disord. 2013 Jun;28(6):811-3.
96. Lesage S, Anheim M, Letournel F, et al. G51D alpha-synuclein mutation causes a novel parkinsonian-pyramidal syndrome. Ann Neurol. 2013 Mar 22.
97. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. Science. 2003 Oct 31;302(5646):841.
98. Ibanez P, Lesage S, Janin S, et al. Alpha-synuclein gene rearrangements in dominantly inherited parkinsonism: frequency, phenotype, and mechanisms. Arch Neurol. 2009 Jan;66(1):102-8.
99. Fuchs J, Nilsson C, Kachergus J, et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology. 2007 Mar 20;68(12):916-22.
100. Devine MJ, Gwinn K, Singleton A, Hardy J. Parkinson's disease and alpha-synuclein expression. Mov Disord. 2011 Oct;26(12):2160-8.
101. Nishioka K, Ross OA, Ishii K, et al. Expanding the clinical phenotype of SNCA duplication carriers. Mov Disord. 2009 Sep 15;24(12):1811-9.
102. Shin CW, Kim HJ, Park SS, Kim SY, Kim JY, Jeon BS. Two Parkinson's disease patients with alpha-synuclein gene duplication and rapid cognitive decline. Mov Disord. 2010 May 15;25(7):957-9.
103. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. Nature. 1997 Aug 28;388(6645):839-40.
104. Bertoncini CW, Fernandez CO, Griesinger C, Jovin TM, Zweckstetter M. Familial mutants of alpha-synuclein with increased neurotoxicity have a destabilized conformation. J Biol Chem. 2005 Sep 2;280(35):30649-52.
105. Chen L, Feany MB. Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a Drosophila model of Parkinson disease. Nat Neurosci. 2005 May;8(5):657-63.
106. Cuervo AM, Stefanis L, Fredenburg R, Lansbury PT, Sulzer D. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. Science. 2004 Aug 27;305(5688):1292-5.

107. Manning-Bog AB, Schule B, Langston JW. Alpha-synuclein-glucocerebrosidase interactions in pharmacological Gaucher models: a biological link between Gaucher disease and parkinsonism. Neurotoxicology. 2009 Nov;30(6):1127-32.
108. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet. 2009 Dec;41(12):1303-7.
109. Simon-Sanchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet. 2009 Dec;41(12):1308-12.
110. Spencer CC, Plagnol V, Strange A, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. Hum Mol Genet. 2011 Jan 15;20(2):345-53.
111. Brice A. Genetics of Parkinson's disease: LRRK2 on the rise. Brain. 2005 Dec;128(Pt 12):2760-2.
112. Lesage S, Durr A, Tazir M, et al. LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. N Engl J Med. 2006 Jan 26;354(4):422-3.
113. Ozelius LJ, Senthil G, Saunders-Pullman R, et al. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. N Engl J Med. 2006 Jan 26;354(4):424-5.
114. Di Fonzo A, Tassorelli C, De Mari M, et al. Comprehensive analysis of the LRRK2 gene in sixty families with Parkinson's disease. Eur J Hum Genet. 2006 Mar;14(3):322-31.
115. Healy DG, Falchi M, O'Sullivan SS, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol. 2008 Jul;7(7):583-90.
116. Goldwurm S, Di Fonzo A, Simons EJ, et al. The G6055A (G2019S) mutation in LRRK2 is frequent in both early and late onset Parkinson's disease and originates from a common ancestor. J Med Genet. 2005 Nov;42(11):e65.
117. Ross OA, Toft M, Whittle AJ, et al. Lrrk2 and Lewy body disease. Ann Neurol. 2006 Feb;59(2):388-93.
118. Zimprich A, Biskup S, Leitner P, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. 2004 Nov 18;44(4):601-7.
119. Borroni B, Cotelli MS, Marchina E, Filosto M, Premi E, Padovani A. Choreo-athetosis in LRRK2 R1441C mutation: Expanding the clinical phenotype. Clin Neurol Neurosurg. 2013 Jul 26.
120. Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. Hum Mutat. 2010 Jul;31(7):763-80.
121. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat. 2012 Sep;33(9):1340-4.
122. Paisan-Ruiz C, Jain S, Evans EW, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron. 2004 Nov 18;44(4):595-600.

123. Zabetian CP, Samii A, Mosley AD, et al. A clinic-based study of the LRRK2 gene in Parkinson disease yields new mutations. Neurology. 2005 Sep 13;65(5):741-4.
124. Di Fonzo A, Rohe CF, Ferreira J, et al. A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. Lancet. 2005 Jan 29-Feb 4;365(9457):412-5.
125. Gilks WP, Abou-Sleiman PM, Gandhi S, et al. A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet. 2005 Jan 29-Feb 4;365(9457):415-6.
126. Aasly JO, Vilarino-Guell C, Dachsel JC, et al. Novel pathogenic LRRK2 p.Asn1437His substitution in familial Parkinson's disease. Mov Disord. 2010 Oct 15;25(13):2156-63.
127. Greggio E, Cookson MR. Leucine-rich repeat kinase 2 mutations and Parkinson's disease: three questions. ASN Neuro. 2009;1(1).
128. Clark LN, Wang Y, Karlins E, et al. Frequency of LRRK2 mutations in early- and late-onset Parkinson disease. Neurology. 2006 Nov 28;67(10):1786-91.
129. Zabetian CP, Hutter CM, Yearout D, et al. LRRK2 G2019S in families with Parkinson disease who originated from Europe and the Middle East: evidence of two distinct founding events beginning two millennia ago. Am J Hum Genet. 2006 Oct;79(4):752-8.
130. Simon-Sanchez J, Marti-Masso JF, Sanchez-Mut JV, et al. Parkinson's disease due to the R1441G mutation in Dardarin: a founder effect in the Basques. Mov Disord. 2006 Nov;21(11):1954-9.
131. Tomiyama H, Li Y, Funayama M, et al. Clinicogenetic study of mutations in LRRK2 exon 41 in Parkinson's disease patients from 18 countries. Mov Disord. 2006 Aug;21(8):1102-8.
132. Goldwurm S, Zini M, Mariani L, et al. Evaluation of LRRK2 G2019S penetrance: relevance for genetic counseling in Parkinson disease. Neurology. 2007 Apr 3;68(14):1141-3.
133. Latourelle JC, Sun M, Lew MF, et al. The Gly2019Ser mutation in LRRK2 is not fully penetrant in familial Parkinson's disease: the GenePD study. BMC Med. 2008;6:32.
134. Gaig C, Ezquerra M, Marti MJ, Munoz E, Valldeoriola F, Tolosa E. LRRK2 mutations in Spanish patients with Parkinson disease: frequency, clinical features, and incomplete penetrance. Arch Neurol. 2006 Mar;63(3):377-82.
135. Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann Neurol. 2002 Mar;51(3):296-301.
136. Giasson BI, Covy JP, Bonini NM, et al. Biochemical and pathological characterization of Lrrk2. Ann Neurol. 2006 Feb;59(2):315-22.
137. Gaig C, Marti MJ, Ezquerra M, Rey MJ, Cardozo A, Tolosa E. G2019S LRRK2 mutation causing Parkinson's disease without Lewy bodies. J Neurol Neurosurg Psychiatry. 2007 Jun;78(6):626-8.
138. Marti-Masso JF, Ruiz-Martinez J, Bolano MJ, et al. Neuropathology of Parkinson's disease with the R1441G mutation in LRRK2. Mov Disord. 2009 Oct 15;24(13):1998-2001.
139. Ujiie S, Hatano T, Kubo S, et al. LRRK2 I2020T mutation is associated with tau pathology. Parkinsonism Relat Disord. 2012 Aug;18(7):819-23.

140. Hasegawa K, Stoessl AJ, Yokoyama T, Kowa H, Wszolek ZK, Yagishita S. Familial parkinsonism: study of original Sagamihara PARK8 (I2020T) kindred with variable clinicopathologic outcomes. Parkinsonism Relat Disord. 2009 May;15(4):300-6.
141. MacLeod D, Dowman J, Hammond R, Leete T, Inoue K, Abeliovich A. The familial Parkinsonism gene LRRK2 regulates neurite process morphology. Neuron. 2006 Nov 22;52(4):587-93.
142. Gehrke S, Imai Y, Sokol N, Lu B. Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. Nature. 2010 Jul 29;466(7306):637-41.
143. Piccoli G, Condliffe SB, Bauer M, et al. LRRK2 controls synaptic vesicle storage and mobilization within the recycling pool. J Neurosci. 2011 Feb 9;31(6):2225-37.
144. Sheerin UM, Charlesworth G, Bras J, et al. Screening for VPS35 mutations in Parkinson's disease. Neurobiol Aging. 2012 Apr;33(4):838 e1-5.
145. Lesage S, Condroyer C, Klebe S, et al. Identification of VPS35 mutations replicated in French families with Parkinson disease. Neurology. 2012 May 1;78(18):1449-50.
146. Sharma M, Ioannidis JP, Aasly JO, et al. A multi-centre clinico-genetic analysis of the VPS35 gene in Parkinson disease indicates reduced penetrance for disease-associated variants. J Med Genet. 2012 Nov;49(11):721-6.
147. Ando M, Funayama M, Li Y, et al. VPS35 mutation in Japanese patients with typical Parkinson's disease. Mov Disord. 2012 Sep 15;27(11):1413-7.
148. Bonifacino JS, Rojas R. Retrograde transport from endosomes to the trans-Golgi network. Nat Rev Mol Cell Biol. 2006 Aug;7(8):568-79.
149. Seaman MN. Recycle your receptors with retromer. Trends Cell Biol. 2005 Feb;15(2):68-75.
150. Healy DG, Abou-Sleiman PM, Casas JP, et al. UCHL-1 is not a Parkinson's disease susceptibility gene. Ann Neurol. 2006 Apr;59(4):627-33.
151. Lincoln S, Vaughan J, Wood N, et al. Low frequency of pathogenic mutations in the ubiquitin carboxy-terminal hydrolase gene in familial Parkinson's disease. Neuroreport. 1999 Feb 5;10(2):427-9.
152. Chartier-Harlin MC, Dachsel JC, Vilarino-Guell C, et al. Translation initiator EIF4G1 mutations in familial Parkinson disease. Am J Hum Genet. 2011 Sep 9;89(3):398-406.
153. Tucci A, Charlesworth G, Sheerin UM, Plagnol V, Wood NW, Hardy J. Study of the genetic variability in a Parkinson's Disease gene: EIF4G1. Neurosci Lett. 2012 Jun 14;518(1):19-22.
154. Nuytemans K, Bademci G, Inchausti V, et al. Whole exome sequencing of rare variants in EIF4G1 and VPS35 in Parkinson disease. Neurology. 2013 Mar 12;80(11):982-9.
155. Lesage S, Condroyer C, Klebe S, et al. EIF4G1 in familial Parkinson's disease: pathogenic mutations or rare benign variants? Neurobiol Aging. 2012 Sep;33(9):2233 e1- e5.
156. Schulte EC, Mollenhauer B, Zimprich A, et al. Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics. 2012 Aug;13(3):281-5.

157. Siitonen A, Majounie E, Federoff M, Ding J, Majamaa K, Singleton AB. Mutations in EIF4G1 are not a common cause of Parkinson's disease. Eur J Neurol. 2013 Apr;20(4):e59.
158. Ramirez-Valle F, Braunstein S, Zavadil J, Formenti SC, Schneider RJ. eIF4GI links nutrient sensing by mTOR to cell proliferation and inhibition of autophagy. J Cell Biol. 2008 Apr 21;181(2):293-307.
159. Silvera D, Arju R, Darvishian F, et al. Essential role for eIF4GI overexpression in the pathogenesis of inflammatory breast cancer. Nat Cell Biol. 2009 Jul;11(7):903-8.
160. Farrer MJ, Hulihan MM, Kachergus JM, et al. DCTN1 mutations in Perry syndrome. Nat Genet. 2009 Feb;41(2):163-5.
161. Wider C, Dachsel JC, Farrer MJ, Dickson DW, Tsuboi Y, Wszolek ZK. Elucidating the genetics and pathology of Perry syndrome. J Neurol Sci. 2010 Feb 15;289(1-2):149-54.
162. Newsway V, Fish M, Rohrer JD, et al. Perry syndrome due to the DCTN1 G71R mutation: a distinctive levodopa responsive disorder with behavioral syndrome, vertical gaze palsy, and respiratory failure. Mov Disord. 2010 Apr 30;25(6):767-70.
163. Ohshima S, Tsuboi Y, Yamamoto A, et al. Autonomic failures in Perry syndrome with DCTN1 mutation. Parkinsonism Relat Disord. 2010 Nov;16(9):612-4.
164. Aji BM, Medley G, O'Driscoll K, Larner AJ, Alusi SH. Perry syndrome: a disorder to consider in the differential diagnosis of Parkinsonism. J Neurol Sci. 2013 Jul 15;330(1-2):117-8.
165. Perry TL, Bratty PJ, Hansen S, Kennedy J, Urquhart N, Dolman CL. Hereditary mental depression and Parkinsonism with taurine deficiency. Arch Neurol. 1975 Feb;32(2):108-13.
166. Perry TL, Wright JM, Berry K, Hansen S, Perry TL, Jr. Dominantly inherited apathy, central hypoventilation, and Parkinson's syndrome: clinical, biochemical, and neuropathologic studies of 2 new cases. Neurology. 1990 Dec;40(12):1882-7.
167. Roy EP, 3rd, Riggs JE, Martin JD, Ringel RA, Gutmann L. Familial parkinsonism, apathy, weight loss, and central hypoventilation: successful long-term management. Neurology. 1988 Apr;38(4):637-9.
168. Lechevalier B, Chapon F, Defer G, et al. [Perry and Purdy's syndrome (familial and fatal parkinsonism with hypoventilation and athymhormia)]. Bull Acad Natl Med. 2005 Mar;189(3):481-90; discussion 90-2.
169. Bhatia KP, Daniel SE, Marsden CD. Familial parkinsonism with depression: a clinicopathological study. Ann Neurol. 1993 Dec;34(6):842-7.
170. Elibol B KT, Atac FB, Hattori N, Sahin G, Gurer G, et al. Familial Parkinsonism with apathy, depression and central hypovenilation (Perry Syndrome). In: Mapping the progress of Alzheimer's and Parkinson's disease. Y M, editor. Boston, MA: Kluwer Academic/Plenum publishers; 2002.
171. Wszolek ZK TY, farrer M, Utti RJ, Hutton ML. Hereditary Tauopathies and parkinsonism. Advances in Neurology, Parkinson disease. Philadelphia, PA: Lippincott Williams and Wilkins 2003. p. 153-63.
172. Wider C, Dickson DW, Stoessl AJ, et al. Pallidonigral TDP-43 pathology in Perry syndrome. Parkinsonism Relat Disord. 2009 May;15(4):281-6.

173. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature. 1998 Apr 9;392(6676):605-8.
174. Matsumine H, Saito M, Shimoda-Matsubayashi S, et al. Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. Am J Hum Genet. 1997 Mar;60(3):588-96.
175. Takahashi H, Ohama E, Suzuki S, et al. Familial juvenile parkinsonism: clinical and pathologic study in a family. Neurology. 1994 Mar;44(3 Pt 1):437-41.
176. Farrer M, Chan P, Chen R, et al. Lewy bodies and parkinsonism in families with parkin mutations. Ann Neurol. 2001 Sep;50(3):293-300.
177. Lucking CB, Durr A, Bonifati V, et al. Association between early-onset Parkinson's disease and mutations in the parkin gene. N Engl J Med. 2000 May 25;342(21):1560-7.
178. Clark IE, Dodson MW, Jiang C, et al. Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin. Nature. 2006 Jun 29;441(7097):1162-6.
179. Park J, Lee SB, Lee S, et al. Mitochondrial dysfunction in Drosophila PINK1 mutants is complemented by parkin. Nature. 2006 Jun 29;441(7097):1157-61.
180. Narendra D, Tanaka A, Suen DF, Youle RJ. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. J Cell Biol. 2008 Dec 1;183(5):795-803.
181. Healy DG, Abou-Sleiman PM, Gibson JM, et al. PINK1 (PARK6) associated Parkinson disease in Ireland. Neurology. 2004 Oct 26;63(8):1486-8.
182. Rogaeva E, Johnson J, Lang AE, et al. Analysis of the PINK1 gene in a large cohort of cases with Parkinson disease. Arch Neurol. 2004 Dec;61(12):1898-904.
183. Valente EM, Salvi S, Ialongo T, et al. PINK1 mutations are associated with sporadic early-onset parkinsonism. Ann Neurol. 2004 Sep;56(3):336-41.
184. Bonifati V, Rohe CF, Breedveld GJ, et al. Early-onset parkinsonism associated with PINK1 mutations: frequency, genotypes, and phenotypes. Neurology. 2005 Jul 12;65(1):87-95.
185. Cazeneuve C, San C, Ibrahim SA, et al. A new complex homozygous large rearrangement of the PINK1 gene in a Sudanese family with early onset Parkinson's disease. Neurogenetics. 2009 Jul;10(3):265-70.
186. Li Y, Tomiyama H, Sato K, et al. Clinicogenetic study of PINK1 mutations in autosomal recessive early-onset parkinsonism. Neurology. 2005 Jun 14;64(11):1955-7.
187. Camargos ST, Dornas LO, Momeni P, et al. Familial Parkinsonism and early onset Parkinson's disease in a Brazilian movement disorders clinic: phenotypic characterization and frequency of SNCA, PRKN, PINK1, and LRRK2 mutations. Mov Disord. 2009 Apr 15;24(5):662-6.
188. Samaranch L, Lorenzo-Betancor O, Arbelo JM, et al. PINK1-linked parkinsonism is associated with Lewy body pathology. Brain. 2010 Apr;133(Pt 4):1128-42.
189. Youle RJ, Narendra DP. Mechanisms of mitophagy. Nat Rev Mol Cell Biol. 2011 Jan;12(1):9-14.

190. Pankratz N, Pauciulo MW, Elsaesser VE, et al. Mutations in DJ-1 are rare in familial Parkinson disease. Neurosci Lett. 2006 Nov 20;408(3):209-13.
191. Canet-Aviles RM, Wilson MA, Miller DW, et al. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. Proc Natl Acad Sci U S A. 2004 Jun 15;101(24):9103-8.
192. Junn E, Taniguchi H, Jeong BS, Zhao X, Ichijo H, Mouradian MM. Interaction of DJ-1 with Daxx inhibits apoptosis signal-regulating kinase 1 activity and cell death. Proc Natl Acad Sci U S A. 2005 Jul 5;102(27):9691-6.
193. Macedo MG, Anar B, Bronner IF, et al. The DJ-1L166P mutant protein associated with early onset Parkinson's disease is unstable and forms higher-order protein complexes. Hum Mol Genet. 2003 Nov 1;12(21):2807-16.
194. Miller DW, Ahmad R, Hague S, et al. L166P mutant DJ-1, causative for recessive Parkinson's disease, is degraded through the ubiquitin-proteasome system. J Biol Chem. 2003 Sep 19;278(38):36588-95.
195. Annesi G, Savettieri G, Pugliese P, et al. DJ-1 mutations and parkinsonism-dementia-amyotrophic lateral sclerosis complex. Ann Neurol. 2005 Nov;58(5):803-7.
196. Ramirez A, Heimbach A, Grundemann J, et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nat Genet. 2006 Oct;38(10):1184-91.
197. Bruggemann N, Hagenah J, Reetz K, et al. Recessively inherited parkinsonism: effect of ATP13A2 mutations on the clinical and neuroimaging phenotype. Arch Neurol. 2010 Nov;67(11):1357-63.
198. Farias FH, Zeng R, Johnson GS, et al. A truncating mutation in ATP13A2 is responsible for adult-onset neuronal ceroid lipofuscinosis in Tibetan terriers. Neurobiol Dis. 2011 Jun;42(3):468-74.
199. Wohlke A, Philipp U, Bock P, et al. A one base pair deletion in the canine ATP13A2 gene causes exon skipping and late-onset neuronal ceroid lipofuscinosis in the Tibetan terrier. PLoS Genet. 2011 Oct;7(10):e1002304.
200. Di Fonzo A, Dekker MC, Montagna P, et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology. 2009 Jan 20;72(3):240-5.
201. Burchell VS, Nelson DE, Sanchez-Martinez A, et al. The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. Nat Neurosci. 2013 Sep;16(9):1257-65.
202. Paisan-Ruiz C, Li A, Schneider SA, et al. Widespread Lewy body and tau accumulation in childhood and adult onset dystonia-parkinsonism cases with PLA2G6 mutations. Neurobiol Aging. 2012 Apr;33(4):814-23.
203. McPherson PS, Garcia EP, Slepnev VI, et al. A presynaptic inositol-5-phosphatase. Nature. 1996 Jan 25;379(6563):353-7.
204. Koroglu C, Baysal L, Cetinkaya M, Karasoy H, Tolun A. DNAJC6 is responsible for juvenile parkinsonism with phenotypic variability. Parkinsonism Relat Disord. 2013 Mar;19(3):320-4.
205. Ahle S, Ungewickell E. Auxilin, a newly identified clathrin-associated protein in coated vesicles from bovine brain. J Cell Biol. 1990 Jul;111(1):19-29.
206. Goker-Alpan O, Lopez G, Vithayathil J, Davis J, Hallett M, Sidransky E. The spectrum of parkinsonian manifestations associated with glucocerebrosidase mutations. Arch Neurol. 2008 Oct;65(10):1353-7.

207. Neudorfer O, Giladi N, Elstein D, et al. Occurrence of Parkinson's syndrome in type I Gaucher disease. QJM. 1996 Sep;89(9):691-4.
208. Machaczka M, Rucinska M, Skotnicki AB, Jurczak W. Parkinson's syndrome preceding clinical manifestation of Gaucher's disease. Am J Hematol. 1999 Jul;61(3):216-7.
209. Tayebi N, Callahan M, Madike V, et al. Gaucher disease and parkinsonism: a phenotypic and genotypic characterization. Mol Genet Metab. 2001 Aug;73(4):313-21.
210. Tayebi N, Walker J, Stubblefield B, et al. Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism? Mol Genet Metab. 2003 Jun;79(2):104-9.
211. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med. 2009 Oct 22;361(17):1651-61.
212. Nalls MA, Duran R, Lopez G, et al. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. JAMA Neurol. 2013 Jun;70(6):727-35.
213. Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. Lancet Neurol. 2012 Nov;11(11):986-98.
214. Mazzulli JR, Xu YH, Sun Y, et al. Gaucher disease glucocerebrosidase and alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies. Cell. 2011 Jul 8;146(1):37-52.
215. Kruger R, Vieira-Saecker AM, Kuhn W, et al. Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype. Ann Neurol. 1999 May;45(5):611-7.
216. Golbe LI, Lazzarini AM, Spychala JR, et al. The tau A0 allele in Parkinson's disease. Mov Disord. 2001 May;16(3):442-7.
217. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med. 2009 Apr 23;360(17):1759-68.
218. Saad M, Lesage S, Saint-Pierre A, et al. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. Hum Mol Genet. 2011 Feb 1;20(3):615-27.
219. Edwards TL, Scott WK, Almonte C, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. Ann Hum Genet. 2010 Mar;74(2):97-109.
220. Pankratz N, Wilk JB, Latourelle JC, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet. 2009 Jan;124(6):593-605.
221. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet. 2010 Sep;42(9):781-5.
222. Nalls MA, Plagnol V, Hernandez DG, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet. 2011 Feb 19;377(9766):641-9.
223. A two-stage meta-analysis identifies several new loci for Parkinson's disease. PLoS Genet. 2011 Jun;7(6):e1002142.
224. Goedert M. Tau protein and neurodegeneration. Semin Cell Dev Biol. 2004 Feb;15(1):45-9.

225. Martin ER, Scott WK, Nance MA, et al. Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. JAMA. 2001 Nov 14;286(18):2245-50.
226. Pastor P, Ezquerra M, Munoz E, et al. Significant association between the tau gene A0/A0 genotype and Parkinson's disease. Ann Neurol. 2000 Feb;47(2):242-5.
227. Baker M, Litvan I, Houlden H, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet. 1999 Apr;8(4):711-5.
228. Stefansson H, Helgason A, Thorleifsson G, et al. A common inversion under selection in Europeans. Nat Genet. 2005 Feb;37(2):129-37.
229. Charlesworth G, Gandhi S, Bras JM, et al. Tau acts as an independent genetic risk factor in pathologically proven PD. Neurobiol Aging. 2012 Apr;33(4):838 e7-11.
230. Leung JC, Klein C, Friedman J, et al. Novel mutation in the TOR1A (DYT1) gene in atypical early onset dystonia and polymorphisms in dystonia and early onset parkinsonism. Neurogenetics. 2001 Jul;3(3):133-43.
231. Kabakci K, Hedrich K, Leung JC, et al. Mutations in DYT1: extension of the phenotypic and mutational spectrum. Neurology. 2004 Feb 10;62(3):395-400.
232. Risch NJ, Bressman SB, deLeon D, et al. Segregation analysis of idiopathic torsion dystonia in Ashkenazi Jews suggests autosomal dominant inheritance. Am J Hum Genet. 1990 Mar;46(3):533-8.
233. Kamm C, Fischer H, Garavaglia B, et al. Susceptibility to DYT1 dystonia in European patients is modified by the D216H polymorphism. Neurology. 2008 Jun 3;70(23):2261-2.
234. Goodchild RE, Dauer WT. Mislocalization to the nuclear envelope: an effect of the dystonia-causing torsinA mutation. Proc Natl Acad Sci U S A. 2004 Jan 20;101(3):847-52.
235. Torres GE, Sweeney AL, Beaulieu JM, Shashidharan P, Caron MG. Effect of torsinA on membrane proteins reveals a loss of function and a dominant-negative phenotype of the dystonia-associated DeltaE-torsinA mutant. Proc Natl Acad Sci U S A. 2004 Nov 2;101(44):15650-5.
236. Misbahuddin A, Placzek MR, Taanman JW, et al. Mutant torsinA, which causes early-onset primary torsion dystonia, is redistributed to membranous structures enriched in vesicular monoamine transporter in cultured human SH-SY5Y cells. Mov Disord. 2005 Apr;20(4):432-40.
237. Hewett JW, Zeng J, Niland BP, Bragg DC, Breakefield XO. Dystonia-causing mutant torsinA inhibits cell adhesion and neurite extension through interference with cytoskeletal dynamics. Neurobiol Dis. 2006 Apr;22(1):98-111.
238. Almasy L, Bressman SB, Raymond D, et al. Idiopathic torsion dystonia linked to chromosome 8 in two Mennonite families. Ann Neurol. 1997 Oct;42(4):670-3.
239. Roussigne M, Cayrol C, Clouaire T, Amalric F, Girard JP. THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies. Oncogene. 2003 Apr 24;22(16):2432-42.

240. Kaiser FJ, Osmanoric A, Rakovic A, et al. The dystonia gene DYT1 is repressed by the transcription factor THAP1 (DYT6). Ann Neurol. 2010 Oct;68(4):554-9.
241. Gavarini S, Cayrol C, Fuchs T, et al. Direct interaction between causative genes of DYT1 and DYT6 primary dystonia. Ann Neurol. 2010 Oct;68(4):549-53.
242. Hartzell C, Putzier I, Arreola J. Calcium-activated chloride channels. Annu Rev Physiol. 2005;67:719-58.
243. Huang WC, Xiao S, Huang F, Harfe BD, Jan YN, Jan LY. Calcium-activated chloride channels (CaCCs) regulate action potential and synaptic response in hippocampal neurons. Neuron. 2012 Apr 12;74(1):179-92.
244. Herve D, Levi-Strauss M, Marey-Semper I, et al. G(olf) and Gs in rat basal ganglia: possible involvement of G(olf) in the coupling of dopamine D1 receptor with adenylyl cyclase. J Neurosci. 1993 May;13(5):2237-48.
245. Zhuang X, Belluscio L, Hen R. G(olf)alpha mediates dopamine D1 receptor signaling. J Neurosci. 2000 Aug 15;20(16):RC91.
246. Corvol JC, Studler JM, Schonn JS, Girault JA, Herve D. Galpha(olf) is necessary for coupling D1 and A2a receptors to adenylyl cyclase in the striatum. J Neurochem. 2001 Mar;76(5):1585-8.
247. Hersheson J, Mencacci NE, Davis M, et al. Mutations in the autoregulatory domain of beta-tubulin 4a cause hereditary dystonia. Ann Neurol. 2012 Dec 13.
248. Lohmann K, Wilcox RA, Winkler S, et al. Whispering dysphonia (DYT4 dystonia) is caused by a mutation in the TUBB4 gene. Ann Neurol. 2012 Dec 13.
249. Yen TJ, Machlin PS, Cleveland DW. Autoregulated instability of beta-tubulin mRNAs by recognition of the nascent amino terminus of beta-tubulin. Nature. 1988 Aug 18;334(6183):580-5.
250. Makino S, Kaji R, Ando S, et al. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. Am J Hum Genet. 2007 Mar;80(3):393-406.
251. Muller U. The monogenic primary dystonias. Brain. 2009 Aug;132(Pt 8):2005-25.
252. Waters CH, Faust PL, Powers J, et al. Neuropathology of lubag (x-linked dystonia parkinsonism). Mov Disord. 1993 Jul;8(3):387-90.
253. Evidente VG, Advincula J, Esteban R, et al. Phenomenology of "Lubag" or X-linked dystonia-parkinsonism. Mov Disord. 2002 Nov;17(6):1271-7.
254. Ichinose H, Ohye T, Takahashi E, et al. Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase I gene. Nat Genet. 1994 Nov;8(3):236-42.
255. Brashear A, Dobyns WB, de Carvalho Aguiar P, et al. The phenotypic spectrum of rapid-onset dystonia-parkinsonism (RDP) and mutations in the ATP1A3 gene. Brain. 2007 Mar;130(Pt 3):828-35.
256. Asmus F, Gasser T. Inherited myoclonus-dystonia. Adv Neurol. 2004;94:113-9.
257. Zimprich A, Grabowski M, Asmus F, et al. Mutations in the gene encoding epsilon-sarcoglycan cause myoclonus-dystonia syndrome. Nat Genet. 2001 Sep;29(1):66-9.
258. Esapa CT, Waite A, Locke M, et al. SGCE missense mutations that cause myoclonus-dystonia syndrome impair epsilon-sarcoglycan trafficking to the

plasma membrane: modulation by ubiquitination and torsinA. Hum Mol Genet. 2007 Feb 1;16(3):327-42.
259. Bruno MK, Lee HY, Auburger GW, et al. Genotype-phenotype correlation of paroxysmal nonkinesigenic dyskinesia. Neurology. 2007 May 22;68(21):1782-9.
260. Lee HY, Xu Y, Huang Y, et al. The gene for paroxysmal non-kinesigenic dyskinesia encodes an enzyme in a stress response pathway. Hum Mol Genet. 2004 Dec 15;13(24):3161-70.
261. Bruno MK, Hallett M, Gwinn-Hardy K, et al. Clinical evaluation of idiopathic paroxysmal kinesigenic dyskinesia: new diagnostic criteria. Neurology. 2004 Dec 28;63(12):2280-7.
262. Suls A, Dedeken P, Goffin K, et al. Paroxysmal exercise-induced dyskinesia and epilepsy is due to mutations in SLC2A1, encoding the glucose transporter GLUT1. Brain. 2008 Jul;131(Pt 7):1831-44.
263. Weber YG, Storch A, Wuttke TV, et al. GLUT1 mutations are a cause of paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak. J Clin Invest. 2008 Jun;118(6):2157-68.
264. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002 Jan;30(1):97-101.
265. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011 Jun;12(6):443-51.
266. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010 Jan;11(1):31-46.
267. Lee H, Schatz MC. Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. Bioinformatics. 2012 Aug 15;28(16):2097-105.
268. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.
269. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156-8.
270. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. Genome Res. 2011 Jun;21(6):961-73.
271. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.
272. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol. 2010;6(12):e1001025.
273. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009 Sep;19(9):1553-61.
274. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010 Jan;20(1):110-21.
275. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012 Nov 1;28(21):2747-54.

276. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012 Aug;22(8):1525-32.
277. Sathirapongsasuti JF, Lee H, Horst BA, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics. 2011 Oct 1;27(19):2648-54.
278. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics. 2009;10:80.
279. Ince PG, Clark, B., Holton, J.L., Revesz, T., Wharton, S.,. Disorders of movement and system degenerations. eighth edition ed: Arnold, London; 2008. p. 889-981.
280. Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. Physiol Behav. 1984 Mar;32(3):489-502.
281. Silveira-Moriyama L, Petrie A, Williams DR, et al. The use of a color coded probability scale to interpret smell tests in suspected parkinsonism. Mov Disord. 2009 Jun 15;24(8):1144-53.
282. Alcalay RN, Siderowf A, Ottman R, et al. Olfaction in Parkin heterozygotes and compound heterozygotes: the CORE-PD study. Neurology. 2011 Jan 25;76(4):319-26.
283. Khan NL, Katzenschlager R, Watt H, et al. Olfaction differentiates parkin disease from early-onset parkinsonism and Parkinson disease. Neurology. 2004 Apr 13;62(7):1224-6.
284. Saunders-Pullman R, Stanley K, Wang C, et al. Olfactory dysfunction in LRRK2 G2019S mutation carriers. Neurology. 2011 Jul 26;77(4):319-24.
285. Silveira-Moriyama L, Guedes LC, Kingsbury A, et al. Hyposmia in G2019S LRRK2-related parkinsonism: clinical and pathologic data. Neurology. 2008 Sep 23;71(13):1021-6.
286. Silveira-Moriyama L, Munhoz RP, de JCM, et al. Olfactory heterogeneity in LRRK2 related Parkinsonism. Mov Disord. 2010 Dec 15;25(16):2879-83.
287. Puls I, Jonnakuty C, LaMonte BH, et al. Mutant dynactin in motor neuron disease. Nat Genet. 2003 Apr;33(4):455-6.
288. Levy JR, Sumner CJ, Caviston JP, et al. A motor neuron disease-associated mutation in p150Glued perturbs dynactin function and induces protein aggregation. J Cell Biol. 2006 Feb 27;172(5):733-45.
289. ; Available from: http://genetics.bwh.harvard.edu/pph2/.
290. Wider C, Wszolek ZK. Rapidly progressive familial parkinsonism with central hypoventilation, depression and weight loss (Perry syndrome)--a literature review. Parkinsonism Relat Disord. 2008;14(1):1-7.
291. Puls I, Oh SJ, Sumner CJ, et al. Distal spinal and bulbar muscular atrophy caused by dynactin mutation. Ann Neurol. 2005 May;57(5):687-94.
292. Purdy A, Hahn A, Barnett HJ, et al. Familial fatal Parkinsonism with alveolar hypoventilation and mental depression. Ann Neurol. 1979 Dec;6(6):523-31.
293. Kumazawa R, Tomiyama H, Li Y, et al. Mutation analysis of the PINK1 gene in 391 patients with Parkinson disease. Arch Neurol. 2008 Jun;65(6):802-8.

294. Abou-Sleiman PM, Healy DG, Quinn N, Lees AJ, Wood NW. The role of pathogenic DJ-1 mutations in Parkinson's disease. Ann Neurol. 2003 Sep;54(3):283-6.
295. Chen H, Herndon ME, Lawler J. The cell biology of thrombospondin-1. Matrix Biol. 2000 Dec;19(7):597-614.
296. Neugebauer KM, Emmett CJ, Venstrom KA, Reichardt LF. Vitronectin and thrombospondin promote retinal neurite outgrowth: developmental regulation and role of integrins. Neuron. 1991 Mar;6(3):345-58.
297. DeFreitas MF, Yoshida CK, Frazier WA, Mendrick DL, Kypta RM, Reichardt LF. Identification of integrin alpha 3 beta 1 as a neuronal thrombospondin receptor mediating neurite outgrowth. Neuron. 1995 Aug;15(2):333-43.
298. Zhou X, Huang J, Chen J, et al. Genetic association analysis of myocardial infarction with thrombospondin-1 N700S variant in a Chinese population. Thromb Res. 2004;113(3-4):181-6.
299. Liu XN, Song L, Wang DW, et al. [Correlation of thrombospondin-1 G1678A polymorphism to stroke: a study in Chinese population]. Zhonghua Yi Xue Za Zhi. 2004 Dec 2;84(23):1959-62.
300. Zhou ZQ, Cao WH, Xie JJ, et al. Expression and prognostic significance of THBS1, Cyr61 and CTGF in esophageal squamous cell carcinoma. BMC Cancer. 2009;9:291.
301. Rice AJ, Steward MA, Quinn CM. Thrombospondin 1 protein expression relates to good prognostic indices in ductal carcinoma in situ of the breast. J Clin Pathol. 2002 Dec;55(12):921-5.
302. Bhargava MM, Feigelson M. Studies on the mechanisms of histidase development in rat skin and liver. I. Basis for tissue specific developmental changes in catalytic activity. Dev Biol. 1976 Feb;48(2):212-25.
303. Ghadimi H, Partington MW, Hunter A. A familial disturbance of histidine metabolism. N Engl J Med. 1961 Aug 3;265:221-4.
304. Selden C, Calnan D, Morgan N, Wilcox H, Carr E, Hodgson HJ. Histidinemia in mice: a metabolic defect treated using a novel approach to hepatocellular transplantation. Hepatology. 1995 May;21(5):1405-12.
305. Levy HL, Shih VE, Madigan PM. Routine newborn screening for histidinemia. Clinical and biochemical results. N Engl J Med. 1974 Dec 5;291(23):1214-9.
306. Lam WK, Cleary MA, Wraith JE, Walter JH. Histidinaemia: a benign metabolic disorder. Arch Dis Child. 1996 Apr;74(4):343-6.
307. Dierks T, Schmidt B, Borissenko LV, et al. Multiple sulfatase deficiency is caused by mutations in the gene encoding the human C(alpha)-formylglycine generating enzyme. Cell. 2003 May 16;113(4):435-44.
308. Cosma MP, Pepe S, Annunziata I, et al. The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. Cell. 2003 May 16;113(4):445-56.
309. Dierks T, Schlotawa L, Frese MA, Radhakrishnan K, von Figura K, Schmidt B. Molecular basis of multiple sulfatase deficiency, mucolipidosis II/III and Niemann-Pick C1 disease - Lysosomal storage disorders caused by defects of non-lysosomal proteins. Biochim Biophys Acta. 2009 Apr;1793(4):710-25.

310. Gubser C, Bergamaschi D, Hollinshead M, Lu X, van Kuppeveld FJ, Smith GL. A new inhibitor of apoptosis from vaccinia virus and eukaryotes. PLoS Pathog. 2007 Feb;3(2):e17.
311. Orrenius S, Zhivotovsky B, Nicotera P. Regulation of cell death: the calcium-apoptosis link. Nat Rev Mol Cell Biol. 2003 Jul;4(7):552-65.
312. Pinton P, Rizzuto R. Bcl-2 and Ca2+ homeostasis in the endoplasmic reticulum. Cell Death Differ. 2006 Aug;13(8):1409-18.
313. Najim al-Din AS, Wriekat A, Mubaidin A, Dasouki M, Hiari M. Pallido-pyramidal degeneration, supranuclear upgaze paresis and dementia: Kufor-Rakeb syndrome. Acta Neurol Scand. 1994 May;89(5):347-52.
314. Di Fonzo A, Chien HF, Socal M, et al. ATP13A2 missense mutations in juvenile parkinsonism and young onset Parkinson disease. Neurology. 2007 May 8;68(19):1557-62.
315. Williams DR, Hadeed A, al-Din AS, Wreikat AL, Lees AJ. Kufor Rakeb disease: autosomal recessive, levodopa-responsive parkinsonism with pyramidal degeneration, supranuclear gaze palsy, and dementia. Mov Disord. 2005 Oct;20(10):1264-71.
316. Ning YP, Kanai K, Tomiyama H, et al. PARK9-linked parkinsonism in eastern Asia: mutation detection in ATP13A2 and clinical phenotype. Neurology. 2008 Apr 15;70(16 Pt 2):1491-3.
317. Schneider SA, Paisan-Ruiz C, Quinn NP, et al. ATP13A2 mutations (PARK9) cause neurodegeneration with brain iron accumulation. Mov Disord. 2010 Jun 15;25(8):979-84.
318. Paisan-Ruiz C GR, Federoff M, Hanagasi H, Sina F, Elahi E, Schneider SA, , Schwingenschuh P BN, Emre M, Singleton AB, Hardy J, Bhatia KP, Brandner S, , Lees AJ HH. Early-onset L-dopa-responsive parkinsonism with pyramidal signs due to ATP13A2,
PLA2G6, FBXO7 and spatacsin mutations. Mov Disord. 2010;25(12):1791-800.
319. Eiberg H, Hansen L, Korbo L, et al. Novel mutation in ATP13A2 widens the spectrum of Kufor-Rakeb syndrome (PARK9). Clin Genet. 2012 Sep;82(3):256-63.
320. Crosiers D, Ceulemans B, Meeus B, et al. Juvenile dystonia-parkinsonism and dementia caused by a novel ATP13A2 frameshift mutation. Parkinsonism Relat Disord. 2011 Feb;17(2):135-8.
321. Santoro L, Breedveld GJ, Manganelli F, et al. Novel ATP13A2 (PARK9) homozygous mutation in a family with marked phenotype variability. Neurogenetics. 2011 Feb;12(1):33-9.
322. Park JS, Mehta P, Cooper AA, et al. Pathogenic effects of novel mutations in the P-type ATPase ATP13A2 (PARK9) causing Kufor-Rakeb syndrome, a form of early-onset parkinsonism. Hum Mutat. 2011 Aug;32(8):956-64.
323. Lin CH, Tan EK, Chen ML, et al. Novel ATP13A2 variant associated with Parkinson disease in Taiwan and Singapore. Neurology. 2008 Nov 18;71(21):1727-32.
324. Fei QZ, Cao L, Xiao Q, et al. Lack of association between ATP13A2 Ala746Thr variant and Parkinson's disease in Han population of mainland China. Neurosci Lett. 2010 May 14;475(2):61-3.
325. Mao XY, Burgunder JM, Zhang ZJ, et al. ATP13A2 G2236A variant is rare in patients with early-onset Parkinson's disease and familial Parkinson's

disease from Mainland China. Parkinsonism Relat Disord. 2010 Mar;16(3):235-6.
326. Funayama M, Tomiyama H, Wu RM, et al. Rapid screening of ATP13A2 variant with high-resolution melting analysis. Mov Disord. 2010 Oct 30;25(14):2434-7.
327. Vilarino-Guell C, Soto AI, Lincoln SJ, et al. ATP13A2 variability in Parkinson disease. Hum Mutat. 2009 Mar;30(3):406-10.
328. Podhajska A, Musso A, Trancikova A, et al. Common pathogenic effects of missense mutations in the P-type ATPase ATP13A2 (PARK9) associated with early-onset parkinsonism. PLoS One. 2012;7(6):e39942.
329. Khan NL, Jain S, Lynch JM, et al. Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: clinical, pathological, olfactory and functional imaging and genetic data. Brain. 2005 Dec;128(Pt 12):2786-96.
330. Macedo MG, Verbaan D, Fang Y, et al. Genotypic and phenotypic characteristics of Dutch patients with early onset Parkinson's disease. Mov Disord. 2009 Jan 30;24(2):196-203.
331. Mata IF, Kachergus JM, Taylor JP, et al. Lrrk2 pathogenic substitutions in Parkinson's disease. Neurogenetics. 2005 Dec;6(4):171-7.
332. Alcalay RN, Mejia-Santana H, Tang MX, et al. Motor phenotype of LRRK2 G2019S carriers in early-onset Parkinson disease. Arch Neurol. 2009 Dec;66(12):1517-22.
333. Trowsdale J. Genetic and functional relationships between MHC and NK receptor genes. Immunity. 2001 Sep;15(3):363-74.
334. MacDonald KS, Fowke KR, Kimani J, et al. Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. J Infect Dis. 2000 May;181(5):1581-9.
335. Szpak Y, Vieville JC, Tabary T, et al. Spontaneous retinopathy in HLA-A29 transgenic mice. Proc Natl Acad Sci U S A. 2001 Feb 27;98(5):2572-6.
336. Cardoso CS, Alves H, Mascarenhas M, et al. Co-selection of the H63D mutation and the HLA-A29 allele: a new paradigm of linkage disequilibrium? Immunogenetics. 2002 Mar;53(12):1002-8.
337. Zareparsi S, James DM, Kaye JA, Bird TD, Schellenberg GD, Payami H. HLA-A2 homozygosity but not heterozygosity is associated with Alzheimer disease. Neurology. 2002 Mar 26;58(6):973-5.
338. Ueta M, Sotozono C, Tokunaga K, Yabe T, Kinoshita S. Strong association between HLA-A*0206 and Stevens-Johnson syndrome in the Japanese. Am J Ophthalmol. 2007 Feb;143(2):367-8.
339. McAulay KA, Higgins CD, Macsween KF, et al. HLA class I polymorphisms are associated with development of infectious mononucleosis upon primary EBV infection. J Clin Invest. 2007 Oct;117(10):3042-8.
340. Nejentsev S, Howson JM, Walker NM, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature. 2007 Dec 6;450(7171):887-92.
341. Sawcer S, Hellenthal G, Pirinen M, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011 Aug 11;476(7359):214-9.

342. Hill AV, Allsopp CE, Kwiatkowski D, et al. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991 Aug 15;352(6336):595-600.
343. Delgado JC, Turbay D, Yunis EJ, et al. A common major histocompatibility complex class II allele HLA-DQB1* 0301 is present in clinical variants of pemphigoid. Proc Natl Acad Sci U S A. 1996 Aug 6;93(16):8569-71.
344. Petersdorf EW, Longton GM, Anasetti C, et al. The significance of HLA-DRB1 matching on clinical outcome after HLA-A, B, DR identical unrelated donor marrow transplantation. Blood. 1995 Aug 15;86(4):1606-13.
345. Thursz M, Yallop R, Goldin R, Trepo C, Thomas HC. Influence of MHC class II genotype on outcome of infection with hepatitis C virus. The HENCORE group. Hepatitis C European Network for Cooperative Research. Lancet. 1999 Dec 18-25;354(9196):2119-24.
346. Hafler DA, Compston A, Sawcer S, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007 Aug 30;357(9):851-62.
347. Jankovic J, Patten BM. Blepharospasm and autoimmune diseases. Mov Disord. 1987;2(3):159-63.
348. Edwards MJ, Trikouli E, Martino D, et al. Anti-basal ganglia antibodies in patients with atypical dystonia and tics: a prospective study. Neurology. 2004 Jul 13;63(1):156-8.
349. Sheerin UM, Schneider SA, Carr L, et al. ALS2 mutations: Juvenile amyotrophic lateral sclerosis and generalized dystonia. Neurology. 2014 Feb 21.
350. Yang Y, Hentati A, Deng HX, et al. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. Nat Genet. 2001 Oct;29(2):160-5.
351. Eymard-Pierre E, Lesca G, Dollet S, et al. Infantile-onset ascending hereditary spastic paralysis is associated with mutations in the alsin gene. Am J Hum Genet. 2002 Sep;71(3):518-27.
352. Gros-Louis F, Meijer IA, Hand CK, et al. An ALS2 gene mutation causes hereditary spastic paraplegia in a Pakistani kindred. Ann Neurol. 2003 Jan;53(1):144-5.
353. Shirakawa K, Suzuki H, Ito M, et al. Novel compound heterozygous ALS2 mutations cause juvenile amyotrophic lateral sclerosis in Japan. Neurology. 2009 Dec 15;73(24):2124-6.
354. Lesca G, Eymard-Pierre E, Santorelli FM, et al. Infantile ascending hereditary spastic paralysis (IAHSP): clinical features in 11 families. Neurology. 2003 Feb 25;60(4):674-82.
355. Mintchev N, Zamba-Papanicolaou E, Kleopa KA, Christodoulou K. A novel ALS2 splice-site mutation in a Cypriot juvenile-onset primary lateral sclerosis family. Neurology. 2009 Jan 6;72(1):28-32.
356. Hadano S, Benn SC, Kakuta S, et al. Mice deficient in the Rab5 guanine nucleotide exchange factor ALS2/alsin exhibit age-dependent neurological deficits and altered endosome trafficking. Hum Mol Genet. 2006 Jan 15;15(2):233-50.
357. Lindner M, Kolker S, Schulze A, Christensen E, Greenberg CR, Hoffmann GF. Neonatal screening for glutaryl-CoA dehydrogenase deficiency. J Inherit Metab Dis. 2004;27(6):851-9.

358. Strauss KA, Puffenberger EG, Robinson DL, Morton DH. Type I glutaric aciduria, part 1: natural history of 77 patients. Am J Med Genet C Semin Med Genet. 2003 Aug 15;121C(1):38-52.
359. Kolker S, Greenberg CR, Lindner M, Muller E, Naughten ER, Hoffmann GF. Emergency treatment in glutaryl-CoA dehydrogenase deficiency. J Inherit Metab Dis. 2004;27(6):893-902.
360. Goodman SI, Stein DE, Schlesinger S, et al. Glutaryl-CoA dehydrogenase mutations in glutaric acidemia (type I): review and report of thirty novel mutations. Hum Mutat. 1998;12(3):141-4.
361. Chen J, Wang ZX, Zhang JL, Yang YL, Huang YN. [Mutation analysis of GCDH gene in eight patients with glutaric aciduria type I]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi. 2011 Aug;28(4):374-8.
362. Marti-Masso JF, Ruiz-Martinez J, Makarov V, et al. Exome sequencing identifies GCDH (glutaryl-CoA dehydrogenase) mutations as a cause of a progressive form of early-onset generalized dystonia. Hum Genet. 2012 Mar;131(3):435-42.
363. Air EL, Ostrem JL, Sanger TD, Starr PA. Deep brain stimulation in children: experience and technical pearls. J Neurosurg Pediatr. 2011 Dec;8(6):566-74.
364. Haack TB, Haberberger B, Frisch EM, et al. Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. J Med Genet. 2012 Apr;49(4):277-83.
365. Worthey EA, Mayer AN, Syverson GD, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011 Mar;13(3):255-62.
366. Haugarvoll K, Johansson S, Tzoulis C, et al. MRI characterisation of adult onset alpha-methylacyl-coA racemase deficiency diagnosed by exome sequencing. Orphanet J Rare Dis. 2013;8:1.
367. Mallott J, Kwan A, Church J, et al. Newborn screening for SCID identifies patients with ataxia telangiectasia. J Clin Immunol. 2013 Apr;33(3):540-9.
368. Johnson JO, Gibbs JR, Megarbane A, et al. Exome sequencing reveals riboflavin transporter mutations as a cause of motor neuron disease. Brain. 2012 Sep;135(Pt 9):2875-82.
369. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061-73.
370. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012 May;20(5):490-7.

Web Resources

Exon Primer http://ihg.gsf.de/ihg/ExonPrimer.html

Primer 3 http://bioinfo.ut.ee/primer3-0.4.0/

UCSC http://genome.ucsc.edu/

Ensembl http://www.ensembl.org/index.html

Plink http://pngu.mgh.harvard.edu/~purcell/plink/

Merlin Linkage Analysis

http://www.sph.umich.edu/csg/abecasis/merlin/tour/linkage.html

Pedstats: http://www.sph.umich.edu/csg/abecasis/PedStats/download/

Novoalign www.novocraft.com

FastQC http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

SIFT http://sift.jcvi.org/

Polyphen 2 http://genetics.bwh.harvard.edu/pph2/

Samtools http://samtools.sourceforge.net/

Picard http://sourceforge.net/projects/picard/