



Genetic analysis of multiple system atrophy and related movement disorders

Thesis submitted for the degree of Doctor of Philosophy
(PhD)

by Lucia Valentina Schottlaender

Supervisors Prof Henry Houlden, Prof Andrew Lees and

Dr Conceição Bettencourt

UCL, Institute of Neurology

2018

DECLARATION OF AUTHORSHIP

I, Lucia Valentina Schottlaender, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, or others have contributed, I confirm this has been indicated in the thesis.

ABSTRACT

The understanding of the pathophysiology of most neurodegenerative movement disorders has been elusive. Such is the case of multiple system atrophy (MSA) and primary familial brain calcification (PFBC). In this thesis I used a range of genetic technologies and functional strategies to unravel the genetic basis of MSA and PFBC.

First, I describe the work performed in MSA and related atypical movement disorders initially by investigating candidate genes. My key findings were: i) negative results when attempting to replicate the association between *COQ2* and the risk of MSA, by Sanger sequencing the largest pathologically confirmed MSA cohort the largest pathologically confirmed MSA cohort; ii) reduced levels of Coenzyme Q10 (CoQ10) in the cerebellum of MSA patients with a cerebellar or mixed MSA subtypes when compared to normal controls and other neurodegenerative movement disorders, when I measured the levels in post-mortem brain tissue of MSA and other patients and controls by high performance liquid chromatography (HPLC); iii) identification of three *C9orf72* repeat expansions and one intermediate expansion in patients presenting with a corticobasal and progressive supranuclear palsy syndrome, and confirmation of the absence of the expansion in pathologically proven MSA, corticobasal degeneration (CBD) and progressive supranuclear palsy (PSP); iv) identification of a *LRRK2* protective variant in MSA by case control analysis of genotyping of *LRRK2* candidate variants.

Second, I detail my work applying next generation sequencing technologies (i.e. whole exome sequencing (WES)) to the study of genetic risk factors in MSA: i) Initially I analysed a definite MSA family and ii) later I performed the largest WES study so far in sporadic MSA. This study included 450 cases out of which 298 were

pathologically confirmed. These data were first investigated for candidate genes linked to MSA, other synucleinopathies and related neurodegenerative disorders, and later by performing a case control association study for common and rare variants. The results of this work were not able to replicate previous findings that linked MSA to *COQ2* or *SNCA*, notwithstanding it revealed interesting candidates that require follow up.

Third, I studied genetically patients with PFBC. My key findings were: i) a pathogenic *SLC20A2* mutation segregating with the disease in an interesting family, found by investigating recently discovered candidate genes I identified by Sanger sequencing; ii) I detail how I studied two independent primary brain calcification consanguineous families by means of homozygosity mapping and WES. I was able to identify a homozygous nonsense mutation segregating with the disease in both families in *JAM2*, a gene encoding the Junction adhesion molecule 2, a tight junction protein. This is a novel gene previously unreported as a cause of human disease. Through collaborations with other scientists, I showed the absence of the expression of the *JAM2* protein in a fibroblast cell line of a homozygous patient compared to a heterozygous carrier and 2 independent controls. Additionally we studied a knock out *JAM-b* (ortholog of human *JAM2*) mouse model and showed gait abnormalities and abnormal brain histopathology.

In conclusion, by applying genetic technologies and related methods, I generated important insights into the CoQ10 pathway in MSA, I generated the largest dataset of WES in MSA and I discovered a new gene for PFBC. My findings are discussed in light of the recent literature and future directions of research into each subject.

ACKNOWLEDGEMENTS

I am extremely thankful to my primary supervisor, Prof Henry Houlden, for having given me the opportunity of working in the neurogenetics lab at such an exciting time for this field. Prof Henry Houlden patiently supported me during all these years and has always been an excellent mentor encouraging my work with new ideas. I am grateful to my secondary supervisor Prof Andrew Lees for giving me the opportunity of coming to London in the first place and being a great teacher in clinic. I am also very grateful to my third supervisor and friend, Dr Conceição Bettencourt, for her constant teaching, support and advice in the lab.

I would also like to thank Prof John Hardy and Prof Nicholas Wood, the heads of the department of molecular neuroscience, for accepting me to work in the department and for both always being inspirational scientists giving me important advice and support on my projects.

I want to specially thank Dr Ricardo Reisin for his collaborative efforts and especially for his permanent understanding and support. I would like to thank Dr Marcelo Merello for introducing me to Professor Lees several years ago. I would also like to thank Dr Chris Turner for his advice before I started my PhD.

I would also like to mention my fellow colleagues in the lab: Dr Anna Sailer, Dr Niccolo Mencacci, Dr Josh Hersheson, Dr Sarah Wiethoff, Miss Debbie Hughes, Dr Alan Pittman, Mr Mark Gaskins, Mr Hallgeir Jonvick, Dr Lee Stanyer, Dr Jana Vandrovцова, Dr Sarah Morgan, Dr Viorica Chelban, Miss Stephanie Efthymiou, Dr Iain Hargreaves, Dr David Lynch, Dr Arianna Tucci, Dr Boniface Mok, Dr Rosella Abeti, Dr Paola Giunti, Dr Zane Jaunmuktane.

I would also like to thank my colleagues from the queen square brain bank: Dr Aoife Kiely, Dr Sandrine Wauters and Prof Janice Holton. And also, Prof Lees team of fellows: Dr Laura Silveira Moriyama, Dr Helen Ling, Dr Atbin Djamshidian-Tehrani, Dr Karen Doherty.

I am also grateful to the colleagues in the NIH that received me in their lab and collaborated during these years: Dr Andy Singleton, Dr Monia Ben-Hamad Hammer, Dr Celeste Sassi, Dr Monica Federoff, Dr Dena Hernandez, Dr Raphael Gibbs, Dr Bryan Traynor, Dr Sonja Scholz.

I would also like to thank my supervisors Prof Muntoni and Dr Mariacristina Scoto and the team at the Dubowitz neuromuscular centre for their support while I was writing up my thesis.

This thesis is dedicated to my children Fermin and Paz who were both born during my PhD years. They are a permanent source of happiness and joy. I would like to specially thank my husband Cristian who has walked with me through this journey. Cristian gave me boundless support at all times.

I would like to thank my parents for teaching me that education is and should always be a top priority. No matter what I choose or what I do, I will always work responsibly and gratefully for the opportunity that I have been given.

And last but not least, I would like to acknowledge and thank all patients and their families for participating in research and being so collaborative and generous. I hope that my small contribution to the field will help in the understanding of the diseases studied and this eventually improves the treatments offered to patients.

TABLE OF CONTENTS

<u>DECLARATION OF AUTHORSHIP</u>	<u>2</u>
<u>ABSTRACT</u>	<u>3</u>
<u>ACKNOWLEDGEMENTS</u>	<u>5</u>
<u>TABLE OF CONTENTS.....</u>	<u>7</u>
<u>LIST OF FIGURES.....</u>	<u>15</u>
<u>LIST OF TABLES.....</u>	<u>21</u>
<u>LIST OF ABBREVIATIONS</u>	<u>23</u>
<u>LIST OF PUBLICATIONS</u>	<u>29</u>
<u>1 CHAPTER 1: GENERAL INTRODUCTION</u>	<u>31</u>
1.1 INTRODUCTION TO NEUROLOGICAL DISORDERS	31
1.2 INTRODUCTION TO GENETICS AND GENOMICS	31
1.3 PRECISION MEDICINE	32
1.4 GENETICS OF MENDELIAN AND COMPLEX DISEASE	33
1.5 RECENT ADVANCES IN NEUROGENETICS	35
1.6 THE GENETIC BASIS OF NEUROLOGICAL DISEASE	40
1.7 THESIS AIMS AND OUTLINE	41
<u>2 CHAPTER 2: INTRODUCTION TO MSA AND RELATED MOVEMENT DISORDERS.....</u>	<u>43</u>
2.1 PARKINSON'S DISEASE AND ATYPICAL PARKINSONISM.....	43
2.2 MULTIPLE SYSTEM ATROPHY.....	46

2.2.1	HISTORY, DEFINITION AND EPIDEMIOLOGY OF MSA	46
2.2.2	EPIDEMIOLOGY.....	46
2.2.3	SPECTRUM OF CLINICAL PHENOTYPE.....	47
2.2.4	PATHOLOGY.....	49
2.2.5	PATHOPHYSIOLOGY	50
2.2.6	GENETICS	51
2.2.6.1	SNCA	53
2.2.6.2	COQ2.....	55
2.2.6.2.1	COQ2 and MSA	57
2.2.6.3	MSA GWAS.....	60
2.2.6.4	Other associations	61
2.2.7	INVESTIGATIONS	61
2.2.8	DIAGNOSIS AND PROGNOSIS.....	62
2.2.8.1	Differential diagnosis	62
2.2.8.2	Prognosis.....	63
2.2.9	TREATMENT.....	63
2.2.9.1	Non-Medical treatment	63
2.2.9.2	Medical and surgical treatment.....	64
2.2.9.3	Neuroprotection	65
2.2.10	FUTURE AND EXPERIMENTAL DEVELOPMENTS	66
2.2.10.1	Anti-inflammatory approaches	66
2.2.10.2	Neurotransplantation	67
2.2.10.3	Mesenchymal autologous stem cell therapy	67
2.3	PRIMARY FAMILIAL BRAIN CALCIFICATION	68
2.3.1	HISTORY, CLASSIFICATION AND EPIDEMIOLOGY.....	68
2.3.2	MOLECULAR GENETICS.....	69
2.3.3	CLINICAL ASPECTS	71
2.3.4	TREATMENT.....	72
3	<u>CHAPTER 3: MATERIALS AND METHODS</u>	<u>73</u>

3.1 ETHICS APPROVAL	73
3.2 SAMPLE COLLECTION.....	73
3.3 GENETIC STUDIES	74
3.3.1 GENETIC NOMENCLATURE.....	74
3.3.2 DNA EXTRACTIONS	74
3.3.2.1 Peripheral blood	74
3.3.2.2 Brain.....	74
3.3.3 DNA CONCENTRATION AND PURITY	75
3.3.4 PRIMER DESIGN, PCR AND SANGER SEQUENCING.....	75
3.3.4.1 Primer design	76
3.3.4.2 Polymerase chain reaction (PCR).....	76
3.3.4.3 Agarose gel electrophoresis	78
3.3.4.4 PCR purification	78
3.3.4.5 BigDye sequencing reaction	78
3.3.4.6 Sequencing reaction purification.....	79
3.3.4.7 Sequencing data analysis.....	79
3.3.5 FRAGMENT ANALYSIS.....	80
3.3.6 HOMOZYGOSITY MAPPING.....	82
3.3.7 NEXT GENERATION SEQUENCING (NGS).....	84
3.3.7.1 Whole exome sequencing (WES).....	86
3.3.7.1.1 TruSeq.....	87
3.3.7.1.1.1 Library preparation	89
3.3.7.1.1.2 Sequencing.....	93
3.3.7.1.2 Nextera rapid capture.....	94
3.3.7.2 Bioinformatic processing of NGS data	96
3.3.8 VARIANT PRIORITIZATION, MUTATION CONFIRMATION AND ANALYSES OF COMPLEX DISEASE	98
3.3.8.1 Quality of the variant.....	98
3.3.8.2 Known genes and pathways	99
3.3.8.3 Coding versus non-coding.....	99
3.3.8.4 Frequency in the population.....	99

3.3.8.5	Mode of Inheritance	100
3.3.8.6	In silico predictions and conservation scores	100
3.3.8.7	Target regions	101
3.3.8.8	Complex disease	101
3.3.9	BIOINFORMATIC STEPS FOR ANALYSIS OF NGS FOR ASSOCIATION STUDIES	102
3.3.9.1	Power calculation	102
3.3.9.2	Generating a Joint VCF file.....	102
3.3.9.3	Genotype quality control (QC).....	102
3.3.9.4	Variant and sample quality control	103
3.3.9.4.1	Per-marker QC or Variant QC	103
3.3.9.4.2	Sample QC.....	104
3.3.9.5	Single locus association study.....	106
3.3.9.6	Rare variant testing.....	107
3.4	MOLECULAR BIOLOGY AND FUNCTIONAL STUDIES	108
3.4.1	SOUTHERN BLOT	108
3.4.2	COENZYME Q 10 MEASUREMENT	108
3.4.3	FUNCTIONAL CHARACTERISATION OF A CANDIDATE GENE IDENTIFIED IN FAMILIES WITH PFBC	110
3.4.4	FIBROBLAST CELL LINES.....	110
3.4.5	WESTERN BLOT	110
ANIMAL	STUDIES	
	111
3.4.5.1	Generation of Jam-B deficient mice	111
3.4.5.2	Behavioural studies.....	112
3.4.5.2.1	Beam walking test.....	112
3.4.5.2.2	Footprint analysis	113
3.4.5.3	Mice culling and CNS dissection	113
3.4.5.4	Mice brain pathology.....	113
3.4.5.4.1	H&E.....	114
3.4.5.4.2	GFAP	115
3.4.5.4.3	NeuN.....	116

4	CHAPTER 4. CANDIDATE GENE STUDIES IN MSA.....	117
4.1	<i>EIF4G1</i>.....	117
4.1.1	BACKGROUND	117
4.1.2	SUBJECTS, MATERIALS AND METHODS	117
4.1.3	RESULTS.....	117
4.1.4	DISCUSSION.....	118
4.2	<i>VPS35</i>.....	118
4.2.1	BACKGROUND	119
4.2.2	SUBJECTS, MATERIALS AND METHODS	119
4.2.3	RESULTS.....	119
4.2.4	DISCUSSION	119
4.3	<i>LRRK2</i>.....	119
4.3.1	BACKGROUND	120
4.3.2	PATIENTS, MATERIALS AND METHODS	120
4.3.3	RESULTS.....	120
4.3.4	DISCUSSION	121
4.4	<i>C9ORF72</i>.....	121
4.4.1	BACKGROUND	122
4.4.2	SUBJECTS MATERIAL AND METHODS.....	123
4.4.3	RESULTS.....	123
4.4.4	DISCUSSION	127
4.5	<i>COQ2</i>.....	128
4.5.1	BACKGROUND	128
4.5.2	<i>COQ2</i> PAPER ANALYSIS.....	129
4.5.2.1	<i>COQ2</i> transcripts.....	129
4.5.2.2	Linkage region genomic location	131
4.5.3	SEQUENCING <i>COQ2</i> IN DEFINITE MSA CASES OF CAUCASIAN ANCESTRY.....	132
4.5.3.1	Subjects materials and methods.....	132
4.5.3.2	Results.....	134

4.5.3.3	Discussion	138
4.5.4	COENZYME LEVELS Q10 IN BRAIN TISSUE	140
4.5.4.1	Background	140
4.5.4.2	Subjects, materials and methods.....	141
4.5.4.3	Results.....	142
4.5.4.4	Discussion	147
5	<u>CHAPTER 5: EXOME SEQUENCING IN MSA</u>	149
5.1	BACKGROUND.....	149
5.2	WES IN FAMILIAL MSA	149
5.2.1	SUBJECTS, MATERIALS AND METHODS	150
5.2.2	RESULTS.....	151
5.2.3	DISCUSSION.....	159
5.3	WES IN SPORADIC MSA	160
5.3.1	SUBJECTS, MATERIALS AND METHODS	160
5.3.1.1	Samples.....	160
5.3.1.2	WES chemistries and initial steps for analysis.	161
5.3.2	QUALITY CONTROL	162
5.3.2.1.1	Steps for Variant and Sample QC.....	162
5.3.3	CANDIDATE GENES RESULTS	168
5.3.4	ASSOCIATION STUDIES	174
5.3.4.1	Single locus	174
5.3.4.1.1	Power calculation	174
5.3.4.1.2	Quantile-quantile plot	175
5.3.4.1.3	Single variant results and Manhattan plot	176
5.3.4.2	Rare variant tests.....	181
5.3.5	DISCUSSION.....	188
6	<u>CHAPTER 6. PRIMARY FAMILIAL BRAIN CALCIFICATION</u>	192

6.1	CANDIDATE GENE ANALYSIS.....	192
6.1.1	<i>SLC20A2</i>	192
6.1.1.1	Background	192
6.1.1.2	Subjects, materials and methods.....	192
6.1.1.3	Results.....	192
6.1.1.4	Discussion	194
6.1.2	<i>PDGFB</i>	195
6.1.2.1	Discussion	195
6.2	<i>JAM2</i> VARIANTS CAUSES PFBC IN RECESSIVE FAMILIES	195
6.2.1	STATEMENT OF CONTRIBUTION.....	195
6.2.2	BACKGROUND	196
6.2.3	SUBJECTS MATERIALS, AND METHODS	196
6.2.4	RESULTS.....	196
6.2.4.1	Homozygosity mapping	198
6.2.4.2	Exome sequencing	199
6.2.4.3	Sanger sequencing and segregation study	200
6.2.4.4	Western blot	200
6.2.4.5	<i>JAM2</i> KO mice	200
6.2.4.5.1	Behavioural tests	201
6.2.4.5.2	Histopathological analysis	201
6.2.5	DISCUSSION.....	202
7	<u>CHAPTER 7: CONCLUSIONS</u>	<u>205</u>
7.1	MSA	205
7.2	PFBC.....	206
	<u>REFERENCES.....</u>	<u>207</u>
	<u>APPENDIX</u>	<u>237</u>

7.3	EXOME SEQUENCING SCRIPTS.....	237
7.3.1	SCRIPTS USED TO ANALYSE EXOMES	237
7.3.1.1	Example of script to analyse an individual exome	237
7.3.1.2	Script to align multiple samples	240
7.3.1.3	Script to filter the vcf file	242
7.3.1.4	Script to join multiple samples	243
7.3.1.5	Script used to annotate the Joint VCF file with annovar	244
7.3.2	SCRIPTS USED FOR THE MSA EXOME STUDY QC.....	244
7.3.2.1	Single association study	252
7.3.2.2	Script for rare variant tests	253
7.3.2.3	Script for QQ plot and Manhattan plot in R.....	253
7.3.2.4	Script used to merge MSA cases and controls in a VCF file, annotate with annovar and select candidate genes to be studied.....	254
7.4	PRIMER SEQUENCES.....	255
7.4.1	<i>JAM2</i>	255
7.4.2	<i>SLC20A2</i> (FROM (244)).....	255
7.4.3	<i>PDGFB</i>	256
7.4.4	<i>EIF4G1</i>	256
7.4.5	<i>VPS35</i>	256
7.4.6	<i>COQ2</i>	257
7.4.7	<i>C9ORF72</i> (FROM (348)).....	257
7.4.8	<i>LRRK2</i> EXON 34	258

LIST OF FIGURES

Figure 1-1: Timeline showing the evolution of DNA sequencing during the Sanger and the post-Sanger eras (modified from(1)).....	36
Figure 1-2: Schematic representation of the genetic architecture of disease and the applicability of current methods to finding risk and causative alleles.....	39
Figure 2-1: Immunohistochemistry for α -synuclein presenting many glial cytoplasmic inclusions in the pontine base (a) as well as neuronal threadlike processes (b). Inset in (b) shows a neuronal cytoplasmic inclusion in the inferior olivary nucleus. Modified from (87)	50
Figure 2-2: SNCA associated SNPs from Scholz and colleagues (168)	54
Figure 2-3: Combined data from Scholz and colleagues (176) and Ross and colleagues (118) for pathologically confirmed MSA samples for the SNP rs111931074.....	54
Figure 2-4: SNCA SNP analysis from Al-Chalabi and colleagues showing a SNP in LD with rs111931074 associated with increased risk of MSA (116).....	54
Figure 2-5: No association of rs111931074 in the SNCA gene in the Korean population. From (123).....	55
Figure 2-6: Biosynthetic pathway of CoQ10(103).	56
Figure 2-7: Japanese MSA families(185).	59
Figure 2-8: A: Yeast complementation assay; B: COQ2 activity in lymphoblastoid cell lines; C: Levels of CoQ10 in brain tissue. (185).	60
Figure 3-1: Sanger sequencing result of a section of the gene SLC20A2 exhibiting reference sequence in 2 samples.....	80
Figure 3-2: C9orf72 repeat expansion screening protocol and workflow.	81
Figure 3-3: Example of a patient negative for the C9orf72 repeat expansion. Top graph: repeat primed PCR (RP-PCR). Bottom graph : sizing PCR; both showing the two different alleles both presenting less than 10 repeats.	82
Figure 3-4: Homozygosity mapping example showing significant homozygosity in chromosome 7.....	84

Figure 3-5: Simplified workflow of WES and WGS. Sample preparation is similar. Genomic DNA is fragmented and adaptors are added which allow each fragment to be hybridized to the flow cell for sequencing. WES then hybridizes the fragments to probes that are complementary to the targeted regions (i.e. the exome). These are captured and the remaining DNA is washed away. The following steps of sequencing are the same for WES and WGS. Modified from (9). 86

Figure 3-6: Tru Seq Enrichment workflow composed by two successive rounds of enrichment (Courtesy of Illumina, Inc.reproduced from www.illumina.com). 88

Figure 3-7: Index plates used for NGS indexing that allows sample pooling. A) Index A from 1 to 12. B) Index B from 1 to 12. C) Ninety-six well plate to combine each library with a unique combination of indexes A and B. In this plate one can index 96 samples at a time. The numbers of samples that will be pooled together before sequencing will depend on the desired coverage. Courtesy of Illumina, Inc., reproduced from Illumina.com. 90

Figure 3-8: Example 300 bp Insert Library Distribution (pre-enrichment) on a High Sensitivity DNA chip of the bioanalyzer. 91

Figure 3-9: Automatic cluster generation. Courtesy of Illumina, Inc., reproduced from https://www.illumina.com/documents/products/datasheets/datasheet_cbot.pdf 93

Figure 3-10: Sequencing by synthesis. Courtesy of Illumina, Inc., reproduced from Illumina.com. 94

Figure 3-11: Sequencing workflow and demultiplexing. Courtesy of Illumina, Inc., reproduced from Illumina.com. 94

Figure 3-12: Nextera rapid capture workflow. Courtesy of Illumina, Inc., reproduced from https://www.illumina.com/documents/products/datasheets/datasheet_nextera_rapid_capture_exome.pdf 95

Figure 3-13: Coenzyme Q10 measurement by HPLC. 109

Figure 3-14: A: Walking beam test. Mice are placed on the beam and their ability to traverse it is considered to be an indication of their balance. Paw slips and traverse time are used as the measurements. B: Footprint analysis. The paws are dipped in

ink or paint, so that the mice leave a trail of footprints as they walk or run along a corridor to a goal box. Measurements of stride length, base width, and fore and hind paw overlap give an indication of gait. Modified from (270).	112
Figure 3-15: Normal brain tissue stained with H&E visualised on a medium magnification. The cortex of the cerebrum is shown with triangular neurons and scattered small darker glial cells in a pink neuropil background. Reproduced from (273)	115
Figure 3-16: Visualization of astroglial, oligodendroglial, and microglial cell bodies in the parietal neocortex from an adult donor. Arrows show staining astrocytes immunohistochemically positive with GFAP. Modified from (275).	116
Figure 3-17: Immunohistochemical stainings for NeuN as a neuronal marker.	116
Figure 4-1: Family tree of the patient (II-3, arrow) with a 27-repeat allele. The patient's father (I-1) and sister (II-1) were diagnosed with dementia, and the patient's mother (I-2) was diagnosed with Parkinson's disease. Reproduced from (302).....	125
Figure 4-2: This figure illustrates fragment analysis example of a heterozygous carrier of the C9orf repeat expansion. Top graph: Repeat-primed PCR. Bottom graph: Sizing PCR.....	126
Figure 4-3: Southern blot (with BsU36I restriction digest) showing 2 of the expanded CBS cases. Abbreviations: C, control no expansion; CBS, corticobasal syndrome; E, expanded. Modified from (302).....	126
Figure 4-4: Fig. 2. Top figure: RP-PCR showing the case with heterozygous 27 repeats on 1 allele, and 5 repeats on the other allele. Bottom figure: Southern blot confirmation of different repeat sizes from 20 to 27 showing the Southern blot appearance of different fragments. The number of repeats was also confirmed by fluorescent PCR of the C9orf72 repeat and fragment analysis. Reproduced from (302)	127
Figure 4-5: Expression levels of COQ2 corresponding to the Affymetrix probe located in the 5' end of exon 1. Data obtained from BRAINEAC(310).....	130
Figure 4-6: Mean expression levels of COQ2 in different regions of control brains. Data obtained from BRAINEAC(310).	131

Figure 4-7: Distance between the linkage region in chromosome 4 and the SNCA gene, and also between COQ2 and SNCA. Green lines highlight the beginning of COQ2 and the end of SNCA and the red box shows the limits of the linkage region. 132

Figure 4-8: Collaborating teams and brain banks that contributed samples for this project. 133

Figure 4-9: A: Boxplot presenting CoQ10 levels in the cerebellum of MSA, CBD, DLB, IPD, CRB_ATX and controls. B: Boxplot presenting CoQ10 levels in the cerebellum of MSA cases subdivided by pathological subtypes MSA_SND, MSA_mixed and MSA_OPCA, and also CRB_ATX cases subdivided into SCA, FRDA and other_ATX. C: Boxplot presenting CoQ10 levels in the frontal cortex of MSA, DLB, IPD, CRB_ATX and controls. D: Boxplot presenting CoQ10 levels in the frontal cortex of MSA cases subdivided by pathological subtypes MSA_SND, MSA_mixed and MSA_OPCA, and also CRB_ATX cases subdivided into SCA, FRDA and other_ATX. Each dot represents one individual, and dots beyond the boxplot whiskers represent outliers. Reproduced from (322) 146

Figure 5-1: Pedigree of the German MSA family. Modified from (316) 151

Figure 5-2: Quality control steps performed for this association study in MSA. 162

Figure 5-3: Scatterplot exhibiting the results of gender calculations in the MSA cases and controls. Samples in the intersection between X and Y axis correspond to 13 samples with missing genotypes and therefore no data for sex check. They were removed. All samples between 0.3 and 0.7 (ambiguous sex) and all samples with discordant sex information were removed. 163

Figure 5-4: Scatterplot showing missingness scores before removing SNPs with low call rates. Samples over 0.2 cut off = 98. 164

Figure 5-5: Scatterplot showing the distribution of missingness per variant. Variants with a score under 0.1 were removed. 164

Figure 5-6: Scatterplot presenting sample missingness after removing SNPs with a call rate under 90%. Samples over 0.2 cut-off = 34. 165

Figure 5-7: Scatterplot showing the heterozygosity rate per sample. 166

Figure 5-8: Scatterplot of identity by descent score. A score of 1 or close to 1 is probably due to sample duplicates.....	166
Figure 5-9: Scatterplot of identity by state score.....	167
Figure 5-10: Principal component analysis plot (PCA) presenting population stratification in the dataset.	167
Figure 5-11: Figure presenting estimated study power according to disease allele frequency.	174
Figure 5-12: Figure presenting estimated study power according to genotype relative risk.	175
Figure 5-13: Quantile-quantile plot (QQ plot).....	176
Figure 5-14: Manhattan plot presenting the association p-values from a simple χ^2 allelic test on MSA cases and controls.	177
Figure 5-15: Genomic view of the chromosome 7 at position 35288459 corresponding to the TBX20 gene.	179
Figure 5-16: Genomic view of the chromosome 1 at position 31905889 corresponding to the SERINC2 gene.	180
Figure 5-17: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing all variants and comparing pathologically confirmed MSA cases versus controls.	182
Figure 5-18: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing rare variants (MAF under 0.01) and comparing pathologically confirmed MSA cases versus controls.	182
Figure 5-19: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing all variants and comparing all MSA cases versus controls.	182
Figure 5-20: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing rare variants (MAF under 0.01) and comparing all MSA cases versus controls.	183
Figure 6-1: Sanger sequencing result of the heterozygous c.1723G>A p.Glu575Lys pathogenic missense mutation in SLC20A2.....	193
Figure 6-2: Axial CT scan presenting brain calcification in the basal ganglia, thalamus, frontal and parietal sub cortical areas, and the cerebellum of a patient presenting a p.Glu575Lys mutation in SLC20A2.	194

Figure 6-3: Family tree of a patient presenting with familial brain calcification caused by a p.Glu575Lys mutation in SLC20A2.	194
Figure 6-4: Family tree of the consanguineous family of a patient presenting with Fahr's syndrome.	197
Figure 6-5: Family tree of a consanguineous family with primary familial brain calcification. Symbols ++ show subject homozygous for the p.R229X mutation in JAM2. +/- are heterozygous unaffected carriers.	197
Figure 6-6: CT scan of the affected probands of the two consanguineous families with Fahr's syndrome exhibiting widespread calcification of the basal ganglia, thalamus, cerebral cortex and cerebellum.	198
Figure 6-7: Genome-wide illustration of homozygous regions in the Irish family with brain calcification. The chromosomes numbers are depicted in green and the homozygous regions in red.	199
Figure 6-8: Western blot showing absence of JAM2 protein in homozygous mutation containing fibroblasts compares to a heterozygous carrier and 2 unrelated controls.	200
Figure 6-9: Behavioural study on JAM-B KO mice. Gait analysis results.	201
Figure 6-10: Behavioural study on JAM-B KO mice. Beam walking results.	201
Figure 6-11: Histopathological analysis of JAM-B KO mice compared to wild type mice exhibiting vacuolization of the midbrain and glial activation in the KO mice.	202

LIST OF TABLES

Table 1-1: Glossary of basic concepts in genetics	34
Table 1-2: Advantages and disadvantages of NGS technologies.	40
Table 2-1: Table presenting some clinical clues to differentiate parkinsonian disorders.	44
Table 2-2: Protein accumulation in atypical parkinsonian disorders.....	45
Table 2-3: Table presenting genetic studies in MSA. They are almost all candidate gene/locus approach except for the MSA GWAS and the COQ2 study. Table modified from (106).	52
Table 2-4 Table presenting symptomatic treatment of MSA. Modified from (87)	64
Table 2-5 Clinical trials in MSA enrolling 100 MSA patients or more.	66
Table 2-6: List of genes linked to PFBC.	70
Table 3-1: Example of PCR mix recipe.	76
Table 3-2: An example of a PCR 65 touchdown 55 cycling programme.....	77
Table 3-3: Sequencing reaction recipe.....	79
Table 3-4: Table showing output data from Genome studio software (Illumina) and how this needs to be adapted for homozygosity mapper.	83
Table 4-1: Variants in exon 8 of EIF4G1 in MSA.	118
Table 4-2: LRRK2 coding variants analysed for association in 177 definite MSA cases and 768 controls. Values in bold are significant.....	121
Table 4-3: Results of C9orf72 expansion screening in atypical parkinsonism(301).....	124
Table 4-4: Clinical characteristics of the three cases with the C9orf repeat expansion and of the case with an intermediate repeat.....	124
Table 4-5: COQ2 coding variants compared to the WT controls. Results in bold are considered statistically significant.	136
Table 4-6: COQ2 coding variants compared to the 1000 genomes project data merged and unmerged to the WT controls. No significant associations were detected.....	137
Table 4-7: Testing for Hardy-Weimberg equilibrium.	137
Table 4-8: Characterization of the samples included in the CoQ10 study.	143

Table 4-9: Multinomial logistic regression estimates for the association between different diagnosis groups and CoQ10 levels in human brain tissue. Note: * and bold highlight significant values (p<0.05). All models were adjusted for age, gender, and post-mortem delay.....	144
Table 5-1: Summary metrics of WES in the MSA family.....	151
Table 5-2: Different filtering strategies in the German MSA family.....	152
Table 5-3: Variants present in I1, II1 and III3.....	152
Table 5-4: Variants present in I1 and II1 and absent in II2 and II4.....	154
Table 5-5: Variants present in I1 and II1.....	154
Table 5-6: Variants present in I1, II1 and III1 and absent in II2 and II4.....	158
Table 5-7: Variants present in I1 and II1 and absent in II2, II4 and III1.....	158
Table 5-8: Origin of MSA samples included in this exome sequencing study. Ref: IOP=Institute of Psychiatry; QSB=Queen Square Brain Bank; NHNN=National Hospital for Neurology and Neurosurgery. A list of collaborators from each centre is available upon request.....	161
Table 5-9: Table showing the number of variants excluded by SNP QC steps in the MSA association study.....	165
Table 5-10: Table showing the number of samples excluded in the different sample QC steps of the MSA association study.....	165
Table 5-11: Table showing number of samples excluded by sample QC steps.....	168
Table 5-12: Sex information on the MSA samples included in the case-control association study.....	168
Table 5-13: MSA subtype of the samples included in the MSA case-control association study.....	168
Table 5-14: List of candidate genes studied in the MSA exome sequencing project.	169
Table 5-15: List of filtered rare potentially pathogenic variants in candidate genes studied in the MSA case control exome study.....	171
Table 5-16: Top 10 SNPs of the MSA case control single variant association study.....	177
Table 5-17: Top 10 SNPs of the MSA case control single variant association study after adjusting for multiple testing.....	178

Table 5-18: Significant results of SKAT-O test investigating association of rare variation in MSA cases versus controls..... 183

Table 5-19. Table presenting the protein encoded by significant genes in the MSA case control SKAT-O analysis and the function of these proteins. 187

LIST OF ABBREVIATIONS

AD	Autosomal dominant
ALS	Amyotrophic lateral sclerosis
AR	autosomal recessive
BBB	Blood-brain-barrier
bp	basepair
CBD	Corticobasal degeneration
CBS	Corticobasal syndrome
cDNA	Complementary DNA
CGH	Comparative genomic hybridization
Chr	Chromosome
CNS	Central nervous system
CNV	copy number variant
CSF	Cerebrospinal fluid

CT	Computed tomography
dbSNP	Database of short genetic variation
ddH2O	Double distilled water
ddNTP	Dideoxynucleotides
DLB	Dementia with Lewy bodies
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ESP	Exome sequencing project (NHLBI)
EVS	Exome variant server (NHLBI)
ExAC	Exome Aggregation Consortium
FA	Friedreich's ataxia
FDR	False discovery rate
FTD	Frontotemporal dementia
GATK	Genome Analysis Toolkit
GCTA	Genome-wide Complex Trait Analysis
gDNA	Genomic DNA
GFAP	Glial fibrillary acid protein (expression)
GQ	Genotype quality
gVCF	Genomic VCF

GWAS	Genome wide association study
H&E	Hematoxylin & eosin
HapMap	Haplotype map
HD	Huntington's disease
het	Heterozygous
hom	Homozygous
HWE	Hardy-Weinberg Equilibrium
IBD	Identity by descent
Indels	Insertions/deletions
ION	Institute of Neurology
IPD	idiopathic Parkinson's disease
kb	Kilobase
KO	Knock out
LD	Linkage disequilibrium
LOD	Logarithm of the odds
MAF	Minor allele frequency
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS	Multiple sclerosis
MSA	Multiple system atrophy

N/A NA	Not available/ not applicable
NeuN	Neuronal nuclear protein (monoclonal antibody)
NGS	Next generation sequencing
NHNN	National Hospital for Neurology and Neurosurgery
NIH	National Institutes of Health
NRES	National Research Ethics Service
OMIM	Online Mendelian Inheritance in Man
OPCA	Olivopontocerebellar atrophy
OR	Odds ratio
PAF	Primary autonomic failure
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD	Parkinson's disease
PFBC	Primary familial brain calcification
PSP	Progressive supranuclear palsy
QC	Quality control
QQ	Quantile-quantile
QSBB	Queen Square Brain Bank
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid

rpm	revolutions per minute
RP-PCR	Repeat-primed PCR
RR	Relative risk
SCA	Spinocerebellar ataxia
SD	Standard deviation
SKAT	Sequence-based kernel association test
SKAT-O	Optimal SKAT
SND	Striatonigral degeneration
SNP	single nucleotide polymorphism
SNV	Single nucleotide variation
SPB	Sample purification bead
UCL	University College London
UK	United Kingdom
USA/US	United States (of America)
UTR	Untranslated region
VCF	Variant call format
VQSR	Variant quality score recalibration
WB	Western Blot
WES	Whole exome sequencing
WGS	Whole genome sequencing

WT	wildtype
(W)ES	(Whole) exome sequencing
1000g	1000 Genomes project

LIST OF PUBLICATIONS

First author publications arising from this thesis:

1. Schottlaender LV, Houlden H, Multiple-System Atrophy (MSA) Brain Bank Collaboration. Mutant COQ2 in multiple-system atrophy. *N Engl J Med*. 2014 Jul 3;371(1):81.
2. Schottlaender LV, Holton JL, Houlden H. Multiple system atrophy and repeat expansions in C9orf72. *JAMA Neurol*. 2014 Sep;71(9):1190–1.
3. Heckman MG*, Schottlaender L*, Soto-Ortolaza AI, Diehl NN, Rayaprolu S, Ogaki K, et al. LRRK2 exonic variants and risk of multiple system atrophy. *Neurology*. 2014 Dec 9;83(24):2256–61.
4. Schottlaender LV, Polke JM, Ling H, MacDoanld ND, Tucci A, Nanji T, et al. Analysis of C9orf72 repeat expansions in a large series of clinically and pathologically diagnosed cases with atypical parkinsonism. *Neurobiol Aging*. 2015 Feb;36(2):1221.e1-6.
5. Schottlaender LV, Bettencourt C, Kiely AP, Chalasani A, Neergheen V, Holton JL, et al. Coenzyme Q10 Levels Are Decreased in the Cerebellum of Multiple-System Atrophy Patients. *PLoS One*. 2016;11(2):e0149557.
6. 7. Schottlaender Lucia V., Sailer Anna, Ahmed Zeshan, Dickson Dennis W., Houlden Henry, Ross Owen A. Multiple System Atrophy: Clinical, Genetics, and Neuropathology. *Neurodegeneration* [Internet]. 2017 Feb 18 [cited 2018 Mar 27]; Available from: <https://onlinelibrary.wiley.com/doi/10.1002/9781118661895.ch7>

*shared first authors.

Manuscripts in preparation:

1. Lucia V Schottlaender, Rosella Abeti, Zane Jaumuktane, Conceicao Bettencourt, Marc Soutar, Debbie Hughes, Alan Pittman, Bernardett Kalmar, Orlando Swayne, Patrick Morrison, John McKinley, Lynch T, Raeburn Forbes, Gavin Mc Donnell, Michel Aurrand Lyons, Sebastien Bradner, Paola Giunti, and Henry Houlden. Mutation in *JAM2* cause recessive primary familial basal ganglia calcification.

Other publications (co)authored during this thesis:

1. Mok KY, Koutsis G, Schottlaender LV, Polke J, Panas M, Houlden H. High frequency of the expanded C9ORF72 hexanucleotide repeat in familial and sporadic Greek ALS patients. *Neurobiol Aging*. 2012 Aug;33(8):1851.e1-5.
2. Hammer MB, Eleuch-Fayache G, Schottlaender LV, Nehdi H, Gibbs JR, Arepalli SK, et al. Mutations in *GBA2* cause autosomal-recessive cerebellar ataxia with spasticity. *Am J Hum Genet*. 2013 Feb 7;92(2):245–51.
3. Fawcett K, Mehrabian M, Liu Y-T, Hamed S, Elahi E, Revesz T, et al. The frequency of spinocerebellar ataxia type 23 in a UK population. *J Neurol*. 2013 Mar;260(3):856–9.

4. Guerreiro R, Kara E, Le Ber I, Bras J, Rohrer JD, Taipa R, et al. Genetic analysis of inherited leukodystrophies: genotype-phenotype correlations in the CSF1R gene. *JAMA Neurol.* 2013 Jul;70(7):875–82.
5. Schottlaender LV, Petzold A, Wood N, Houlden H. Diagnostic clues and manifesting carriers in fukutin-related protein (FKRP) limb-girdle muscular dystrophy. *J Neurol Sci.* 2015 Jan 15;348(1–2):266–8.
6. Federoff M, Schottlaender LV, Houlden H, Singleton A. Multiple system atrophy: the application of genetics in understanding etiology. *Clin Auton Res Off J Clin Auton Res Soc.* 2015 Feb;25(1):19–36.
7. Gami P, Murray C, Schottlaender L, Bettencourt C, De Pablo Fernandez E, Mudanohwo E, et al. A 30-unit hexanucleotide repeat expansion in C9orf72 induces pathological lesions with dipeptide-repeat proteins and RNA foci, but not TDP-43 inclusions and clinical disease. *Acta Neuropathol (Berl).* 2015 Oct;130(4):599–601.
8. Kara E, Tucci A, Manzoni C, Lynch DS, Elpidorou M, Bettencourt C, et al. Genetic and phenotypic characterization of complex hereditary spastic paraplegia. *Brain J Neurol.* 2016 Jul;139(Pt 7):1904–18.
9. Sailer A, Scholz SW, Nalls MA, Schulte C, Federoff M, Price TR, et al. A genome-wide association study in multiple system atrophy. *Neurology.* 2016 Oct 11;87(15):1591–8.
10. Chelban V, Manole A, Pihlstrøm L, Schottlaender L, Efthymiou S, OConnor E, et al. Analysis of the prion protein gene in multiple system atrophy. *Neurobiol Aging.* 2017 Jan;49:216.e15-216.e18.
11. Batla A, Tai XY, Schottlaender L, Erro R, Balint B, Bhatia KP. Deconstructing Fahr's disease/syndrome of brain calcification in the era of new genes. *Parkinsonism Relat Disord.* 2017 Apr;37:1–10.
12. Selikhova M, Fedotova E, Wiethoff S, Schottlaender LV, Klyushnikov S, Illarioshkin SN, et al. A 30-year history of MPAN case from Russia. *Clin Neurol Neurosurg.* 2017 Aug;159:111–3.

1 CHAPTER 1: GENERAL INTRODUCTION

1.1 INTRODUCTION TO NEUROLOGICAL DISORDERS

Neurological disorders include all the diseases that affect primarily the nervous system. Neuroscience is the branch of science that studies the nervous system. Physicians treating neurological conditions rely on neuroscience research to provide the best care for their patients. Translational research aims to join together the work of clinicians, researchers and other related fields, to improve healthcare and optimise prevention, diagnosis and treatment of patients.

Neurological disorders can affect one or many portions of the nervous system. In this thesis, I have worked mostly on multiple system atrophy (MSA) that is a complex movement disorder that affects both the central and peripheral nervous system, and presents with atypical parkinsonism, cerebellar ataxia and autonomic dysfunction.

1.2 INTRODUCTION TO GENETICS AND GENOMICS

According to the world health organization genetics is the study of heredity and genomics is defined as the study of genes, their functions, and related techniques. The main difference between genomics and genetics is that genetics scrutinizes the functioning and composition of the single gene while genomics addresses all genes and their inter-relationships in order to identify their combined influence on any organism.

The human genome is composed of around 6 billion base pairs (bp) stored in 23 chromosome pairs. There are approximately 21000 protein-coding genes contained in the human genome. This constitutes around ~1-2% of the genome and the set of human exons is called "the exome". The remaining portions of the genome consist of intronic sequence, RNA genes, regulatory sequences, and repetitive DNA in which the function is less understood(1).

The most common type of variation in the DNA are single-nucleotide polymorphisms (SNPs). Each human present approximately 3 million of these variants when compared to the

reference sequence. These SNPs arise on average every 1000 bps. SNPs where the minor allele frequency (MAF) occurs in more than 1% of the population are classified as common variants, and in less than 1% are rare variants and between 1% and 5% are considered variants of low frequency. Other causes of genomic variation can be the insertion or deletion of multiple bp in the range from 1 to 1000 bp. These can cause further alterations of the coding DNA sequence by producing a frameshift of the sequence that may result in a truncated protein product. Larger (than 1000 bp) are referred to as copy number variants (CNV) and can also be common or rare. Inversion and translocation of genomic regions can result in structural changes affecting many genes(1).

Genetic variation can be present in the germline cells or be acquired somatically. Germline variation can be inherited or can occur *de novo* during meiosis or just after fertilization. Variation occurring in somatic cells is acquired and can arise randomly or secondary to environmental factors.

1.3 PRECISION MEDICINE

Precision medicine aims to identify small groups of patients that share certain genetic and molecular characteristics to tailor the best diagnostic and therapeutic methods. The “omics”, including genomics, transcriptomics, proteomics and metabolomics are increasingly becoming a part of the investigations that can potentially be studied in each patient. They provide enormous information that can be directly applied to the patient’s management. For example, people at risk for surgical procedures (e.g.: *RYR1* gene mutations and risk of malignant hyperthermia(2)) or cancer patients and their response to treatments and monitoring of disease progression(3).

In Neurology, we are still a step back compared to cancer research. The area of greater advances in precision medicine in Neurology are perhaps, the epilepsies, where through the understanding of pharmacogenomics, neurologists can guide the best drugs for their patients. Moreover, the epilepsies have many useful disease models in which drugs can be tested(4). However, it is paramount that large databases are shared and put together to

ensure data analysis provides reliable and reproducible results before applying these to the patients.

1.4 GENETICS OF MENDELIAN AND COMPLEX DISEASE

Neurogenetics is a subspecialty of neurology that aims to understand, diagnose and treat neurological conditions that harbour a genetic basis.

The genetic background of a disorder is classified into “Mendelian” and “non-Mendelian”(5). The classical Mendelian modes of inheritance are monogenic where the presence or absence of a genotype at a single locus are both sufficient and necessary for the character to be expressed. These diseases can present different patterns of inheritance: Autosomal dominant (AD) if manifested in the heterozygous state, autosomal recessive (AR) if manifested in the homozygous or compound heterozygous state, or X-linked if linked to the X chromosome or Y-linked to the Y chromosome. For example, a Mendelian disorder that we are going to study later in this thesis is primary familial brain calcification (PFBC) that can present with both AD or AR inheritance.

Non-mendelian types of inheritance include maternal (or mitochondrial) and multifactorial inheritance. Mitochondrial inheritance is also called maternal because the mitochondria is inherited from the oocyte and therefore from the mother. So, affected fathers do not transmit the condition to their offspring. Complex, or multifactorial inheritance describes the process by which multiple genes in combination with lifestyle and/or environmental factors can cause or increase risk of a trait or disorder. Many neurological disorders such as Parkinson’s disease (PD) can present in some cases Mendelian inheritance (e.g. *LRRK2* families) and other multifactorial or complex (i.e. Idiopathic Parkinson’s disease (IPD))(6).

In this thesis, we will study the genetic basis of monogenic diseases as well as the genetic risk factors of complex disorders.

I have included a table of basic concepts in genetics that will be used throughout this thesis: Table 1-1.

Table 1-1: Glossary of basic concepts in genetics

Penetrance	The frequency with which a genotype manifests itself in a given phenotype.
Linkage disequilibrium (LD)	LD is a statistical association between particular alleles at separate but linked loci. Genetic loci are in LD if, across the population as a whole, they are found together on the same haplotype more often than expected. In general, two loci in linkage disequilibrium will also present genetic linkage, but the reverse is not necessarily true.
Haplotype	A series of alleles found at linked loci on a single chromosome and inherited together from a single parent. Also, the term "haplotype" can refer to the inheritance of a cluster of SNPs that can be analysed to understand the pattern of genetic variation in a group of individuals.
Genotype	The set of genes of an organism and the complete list of alleles present at one or a number of loci.
Phenotype	The observable characteristics of a cell or organism, including the result of any test that is not a direct test of the genotype.
Heritability	The proportion of the causation of a character that is due to genetic causes.

Genetic phase	Refers to the allelic combinations that an individual received from its parents. If two alleles originated from the same parent they are said to be in <i>cis</i> phase. If each allele originated from a different parent they are in <i>trans</i> phase. <i>Cis</i> -acting regulatory factors are on the same DNA molecule or chromosome as opposed to <i>trans</i> -acting regulatory factors that can control their target irrespective of their chromosome location.
Locus heterogeneity	Determination of the same disease or phenotype by mutations at different loci.
Pleiotropy	The common situation where variation in one gene affects several different aspects of the phenotype that can be differently expressed in different organisms or subjects.

1.5 RECENT ADVANCES IN NEUROGENETICS

Since the completion of the human genome project was established in the early 2000's(7), the advances in genetics have increased exponentially.

The current widespread availability of next-generation sequencing (NGS) technologies has significantly improved the precision and speed of genetic research and genetic testing for diagnostic purposes. NGS include the technologies developed in the post-Sanger era. They have mostly been developed with whole-genome sequencing (WGS) in mind, but they have a whole range of applications including, but not limited to, whole exome sequencing (WES), targeted re-sequencing and RNA sequencing(1). Figure 1-1 presents a timeline of the advances in genetic and genomic technologies in the past decades.

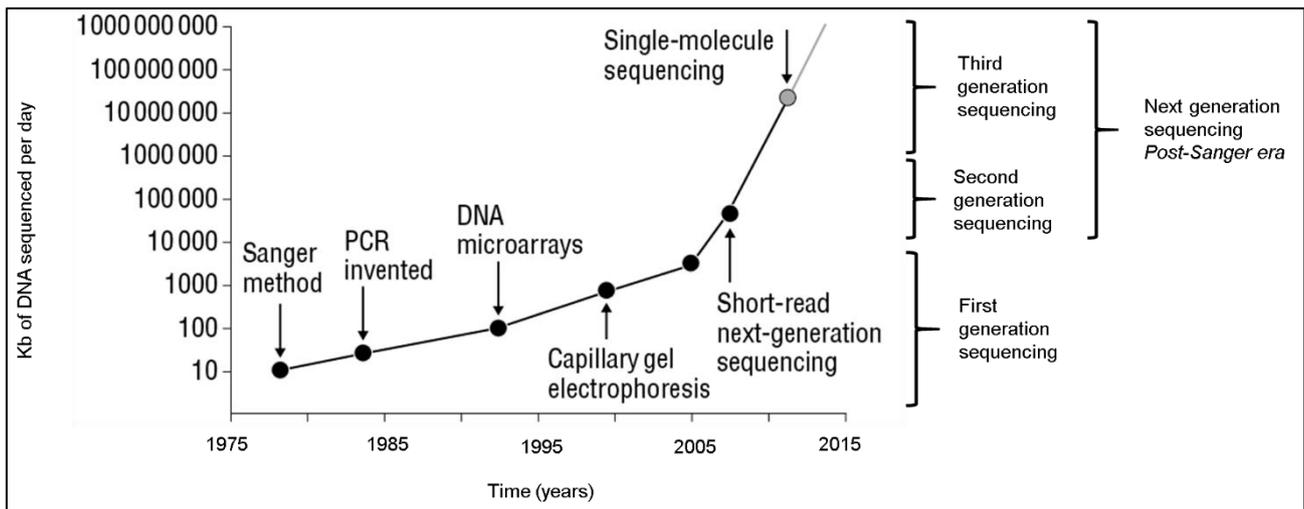


Figure 1-1: Timeline showing the evolution of DNA sequencing during the Sanger and the post-Sanger eras (modified from(1)).

The ever-increasing number of disorders that now have a clear genetic cause was unthinkable a few decades ago.

The approach to investigating different neurological disorders is guided by the estimated genetic risk that at disease may have. Well established genome wide association studies (GWAS) focus on common variants, and are based on the “common disease-common variant” hypothesis where common variants have a small effect size and would modulate risk of developing the disease. Studies usually require large sample sets to achieve powerful results in which large number of samples are genotyped for a large number of markers. It is important to note that GWAS identify loci and not genes, and the direct link to a gene or a pathway is not generally obvious and sometimes resides in intergenic regions. Therefore, risk markers identified by a GWAS always needs further biological studies and support(8). GWAS have been very successful in unravelling genetic risk in many neurological disorders including PD, Alzheimer’s disease and multiple sclerosis (MS)(9).

Traditional linkage studies and positional cloning of targeted regions were the approach when the disease was most likely caused by high risk low frequency alleles(8), such as the *APP* or *PS* mutations that cause Alzheimer’s disease with a Mendelian mode of inheritance.

Linkage analysis is used to map genetic loci by use of observations of related individuals to detect the chromosomal location of presumably disease associated genes. Two genetic loci are linked if they are transmitted together from parent to offspring more often than what

is expected under independent inheritance. Because if they are located close to each other, they tend to be inherited together during meiosis they are genetically “linked”. Sometimes, when recombination occurs during meiosis it is possible that close genes on the same chromosome are not inherited together and this produces genetic variability. However, this occurs through the exchange of genetic material among homologous chromosomes and happens more often in distantly located genes.

Genotyping a number of markers and segments that are identical by descent (IBD) can provide information on the haplotypes of a sample that can then be used to compare the haplotypes of affected patients to unaffected family members. The logarithm of the odds (LOD) score is a function of the recombination fraction or chromosomal position measured in centimorgans (cM). The higher the LOD score, the greater the evidence for linkage. Traditionally, a score of 3 was regarded as significant evidence of linkage associated to the disease or trait studied. However, achieving such a number is not always possible for various reasons, being the most limiting one the size and information available of the family studied. To reach a score of 3 one need at least data from 10 affected family members. Linkage analysis can be used for Mendelian (parametric) and complex traits (non-parametric, or model-free).

Linkage was followed by positional cloning of the genes in the linked region. This was very laborious in the first decades of linkage studies (1980-2000) and was immensely facilitated by the completion of the human genome project. Basically, in the first decades one delimited a linked region and had to then clone the genes in that region and probably all or most of them where previously unknown. After the completion of the human genome project, once the researcher obtains a linked region, they can rely on public databases to study the region of interest and identify candidate genes.

An efficient strategy for mapping human genes that cause recessive traits has been devised and uses mapped linkage analysis data, and the DNA of affected children from consanguineous marriages. This facilitated the concept of homozygosity mapping (or autozygosity mapping) where recessively inherited disease haplotypes are mapped according to the disease status of affected and unaffected family members(10,11).

Obtaining data for linkage and homozygosity studies was in the first decades performed by restriction fragment length polymorphisms (RFLPs)(10) and as methods developed genome wide single nucleotide genotyping assays became more cost effective(11). In parallel, initial gene discovery by cloning and Sanger sequencing have been replaced by next generation sequencing technologies. The later are the technologies used in this thesis(8).

A good example of gene discovery by linkage analysis and positional cloning before the human genome project era are mutations in the *APP* gene causing Alzheimer's disease(12). Similar technologies (but in this case complemented by analysis of chromosomal rearrangements that where affecting the exact gene position) was the case of the discovery of the first (and one of the largest) neurological disease gene, causing Duchenne muscular dystrophy(13). Nowadays this is usually the result of combining genome wide SNP genotyping arrays with exome or genome sequencing(14,15).

Moderate risk low frequency alleles constitute a bigger challenge. Such is the case of an increased risk of PD conferred by *GBA* variants. This link was detected by traditional methods because it was based on the clinical observation of a higher incidence of PD in the parents of children affected by Gaucher's disease(8).

Rarer variants conferring low risk, or rare structural changes such as CNVs or copy neutral variations (loss of heterozygosity) are more difficult to be established with current methods. However, thanks to the advances in whole genome sequencing (WGS) that is closer to be a cost-effective technology, and the developments of new bioinformatics pipelines able to identify these type of changes, it is becoming more and more feasible. See Figure 1-2.

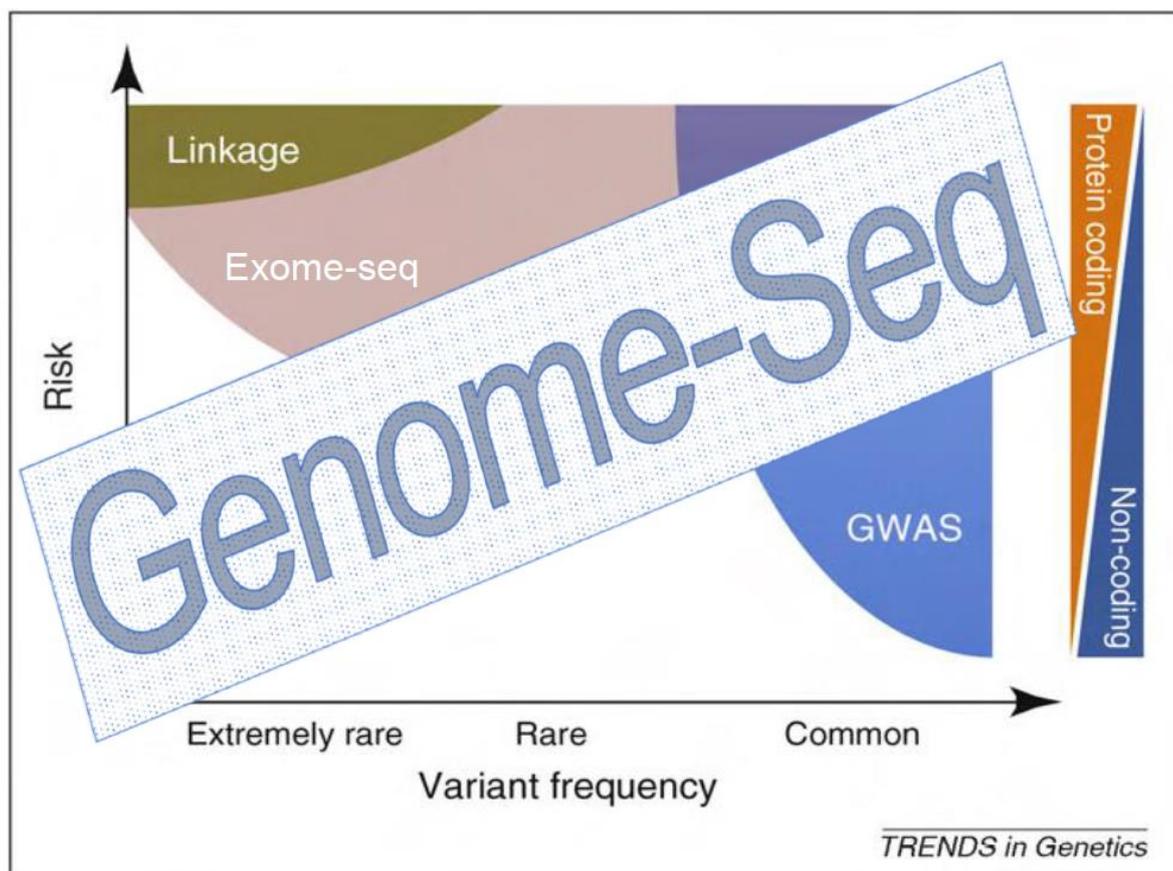


Figure 1-2: Schematic representation of the genetic architecture of disease and the applicability of current methods to finding risk and causative alleles. Ref: Exome-seq: whole-exome sequencing, Genome-Seq: whole-genome sequencing, GWAS: genome-wide association study. Modified from (8).

Recent advances in genetic technology, targeting the whole genome or the whole exome have altered the design of some genetic experiments allowing for hypothesis free projects interrogating the whole genome of ethnically matched subjects with reliable quality control measures. This is tackling the false positive results that were sometimes detected by targeted locus studies that failed to be replicated(9). Table 1-2 gives a list of advantages and limitations of NGS technologies.

Another approach that became widely available with NGS is the study of trios and the possibility to identify *de novo* mutations in the offspring of unaffected parents. This was particularly successful in a study of early onset PD cases(16).

Third generation sequencing refers to a different approach of NGS that avoids the need of DNA amplification and sequences the DNA molecules in real time (sometimes called next

NGS)(1). An example of third generation sequencing is the oxford nanopore. In this thesis, we used second generation sequencing and we are therefore going to describe this method and always refer to NGS (instead of second-generation sequencing).

Table 1-2: Advantages and disadvantages of NGS technologies.

	Advantages	Limitations
Next generation sequencing	Cost effective	Does not detect repeat expansions regions accurately
	Rapid turnaround time (compared to Sanger sequencing)	Cannot detect accurately large CNVs
	Require smaller amounts of DNA (than Sanger and in particular for WGS)	Can present false positives (especially in repetitive regions) and false negatives
	Can study a group of genes simultaneously (e.g.: for linked regions, or specific pathways of interest)	Cannot establish genetic phase
	Can be used for different approaches according to the study hypothesis.	

1.6 THE GENETIC BASIS OF NEUROLOGICAL DISEASE

In the same way that our own genetic architecture determines who we are, our eye colour or height, genetic risk factors affect the risk of neurological disorders as well as the way diseases manifests in each person.

Many neurological disorders have a clear genetic cause. Such is the case of Charcot Marie Tooth type 1A (CMT 1A), spinocerebellar ataxia type 1 (SCA1) or Huntington’s disease (HD) that are all inherited in an autosomal dominant manner. Other neurological disorders like

MS or stroke have known genetic risk factors that are probably acting in combination with the environment. An increased risk in family members of MS patients is well established but in most of the cases it is not inherited in a Mendelian fashion(17). Other disorders, such as amyotrophic lateral sclerosis (ALS), can present monogenic familiar forms and also polygenic complex forms.

Interestingly, although pure Mendelian forms of disease in neurology are rare, the discovery of the faulty genes causing them has constituted the basis of most of our understanding of the pathobiological pathways involved in neuroscience. Further, this information can be then applied to other sporadic diseases(9).

Genetic changes can affect the presence of the disease and/or its phenotype. Such is the case in disorders where anticipation is present in subsequent generations (for example, myotonic dystrophy type 1 through a maternal line of inheritance, HD through a paternal line).

With the advances in genetic technologies it is now possible to study many disorders where this was previously unpractical, unsuccessful or too expensive.

1.7 THESIS AIMS AND OUTLINE

Genes and the way genes interact play a central role in the pathogenesis of neurological disorders. Genetic changes can cause, increase risk, protect, predict course, guide treatment and act as biomarkers.

The scientific knowledge is exponentially boosted after unravelling genetic causes or risk factors for the disease. Understanding the pathophysiology of disease is key for providing medical care. Initially, it will probably be limited to genetic counselling and family planning, but later can lead to potential treatments or preventative measures.

The aim of this thesis is to apply modern genetic technologies in the understanding of two rare neglected neurological disorders, namely MSA and PFBC.

I have provided a statement of contribution at each results chapter and stated when work has been performed in collaboration with other colleagues or researchers.

2 CHAPTER 2: INTRODUCTION TO MSA AND RELATED MOVEMENT DISORDERS

2.1 PARKINSON'S DISEASE AND ATYPICAL PARKINSONISM

Movement disorders are a group of neurological diseases that cause abnormal hyperkinetic and hypokinetic movements. Hyperkinetic movement disorders are tremor, tics, chorea, myoclonus and dystonia. Hypokinetic movement disorders comprise akinesia (slowness and fatigue of movement) and rigidity. The akinetic-rigid syndromes are defined by paucity and slowness of movement accompanied by muscle stiffness and resistance to passive movement. The akinetic-rigid syndrome is typical of idiopathic Parkinson's disease (IPD), so is often described as the syndrome of parkinsonism.

The most common form of parkinsonism is Parkinson's disease (PD) that presents a lifetime risk of developing the disease of 1.5%. The incidence increases with age from 17.4 per 100,000 in the fifth decade to 93.1 per 100,000 in people over 70 years of age. The median age of onset is 60 years although a 10% of patients present younger than 45 years of age. Death occurs usually within 15 years of disease onset.

PD is characterized by bradykinesia associated with at least one other motor sign: rigidity, tremor or gait disturbance. The course is progressive, presents unilaterally with asymmetrical signs, and a number of non-motor symptoms accompanying or preceding the motor onset (e.g. hyposmia, constipation, genito-urinary symptoms, REM-behavioural sleep disorders). These plus a good and sustained response to levodopa support a diagnosis of PD according to the consensus criteria(18). Dementia in PD presents later in the disease and when present before or during the first year of symptoms a differential diagnosis of dementia with Lewis bodies should be considered.

The neuropathological hallmark is a region-specific selective loss of dopaminergic, neuromelanin-containing neurons from the pars compacta of the substantia nigra associated to intraneuronal α -synuclein immunoreactive inclusions: the Lewy bodies(19).

The cause of PD is unknown but up to 10% of patients have a family history and there are significant genetic risk factors that contribute to the disease. Both high risk monogenic forms (*SNCA*, *LRRK2*, *Parkin*, *PINK1*, *VPS35* among others), intermediate risk (for example *GBA*) as well as a long list of low risk variants have been reported as associated with PD(20).

The main differential diagnosis after suspected parkinsonism are the atypical parkinsonian syndromes: MSA, progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and dementia with Lewy bodies (DLB). Secondary causes of parkinsonism (such as drug induced, metabolic, infectious, traumatic, vascular, tumoral or toxic) should always be excluded. And finally, there are other rarer primary degenerative causes of parkinsonism to be considered: Alzheimer’s disease, Huntington’s disease, frontotemporal dementia, basal ganglia calcification (Fahr’s disease), spinocerebellar ataxia (SCAs) among others.

Table 2-1: Table presenting some clinical clues to differentiate parkinsonian disorders.

Clinical finding	Differential to consider
Cerebellar ataxia	MSA, SCAs
Early or marked dysautonomia	MSA
Supranuclear gaze palsy, slow saccades	PSP, CBD
Early falls	PSP, MSA
Early dementia	Alzheimer's disease, DLB
Asymmetric apraxia	CBD
No sustained levodopa response, despite adequate trial	All of the above

Atypical parkinsonian disorders are less common than PD with a prevalence for MSA and PSP between 2-5 per 100,000. DLB is the second most common type of dementia (after Alzheimer’s disease) and accounts for 5% of cases of dementia in older populations(21), and in a recent study had an incidence of 3.5 per 100,000(22). PSP is characterized by symmetrical parkinsonism, cognitive changes, supranuclear palsy of vertical gaze, early falls, dysarthria, and dysphagia. The corticobasal syndrome (CBS) consists of a constellation of extrapyramidal and frontoparietal cortical features and constitutes the classic clinical

presentation of CBD, but many CBS cases turn out to have alternate neuropathology(23). DLB patients exhibit a similar pattern of cognitive impairment to that seen in PD but with an onset in the early disease stages, either before the motor symptoms or within the first year. There is both pathological and genetic overlap between DLB and Alzheimer’s disease. Pathologically, DLB brains present Lewy bodies as well as neurofibrillary tangles and amyloid plaques characteristic of Alzheimer’s disease. Genetically, mutations in five PD genes have been linked to the DLB phenotype, including genetic variation in *GBA*, *LRRK2*, *MAPT*, *SCARB2* and *SNCA*. Also, the *APOE* ϵ 4 allele is a significant risk factor for DLB, and familial Alzheimer dementia cases due to *APP*, *PSEN1* and *PSEN2* mutations occasionally present with mixed Alzheimer and Lewy body pathology. Remarkably, a recent study of 111 pathologically conformed DLB patients found a ~25% of cases carrying a pathogenic mutation or risk variant in *APP*, *GBA* or *PSEN1*(24). The main clinical clues in the diagnosis are presented in Table 2-1.

Likewise PD, the primary atypical parkinsonian disorders are linked to the accumulation of misfolded proteins in the brain and neuronal degeneration. The proteins are α -synuclein and tau (Table 2-2). The different isoforms of the protein tau are 3 repeat tau and 4 repeat tau. Both PSP and CBD have 4 repeat tau and FTD accumulates 3 and 4 repeat tau(25). In addition to the genes mentioned in the previous paragraph, the known genetic risk factor for PSP is the H1 haplotype in *MAPT*, and there are newly proposed links with *EIF2AK3*, *STX6*, and *MOBP*(26). The H1/H1 *MAPT* haplotype is associated with both PSP and CBD(26–28) and likewise, *MAPT* variants such as p.N410H(29) and p.A152T, have been linked to pathologically confirmed CBD and PSP(27,30).

Table 2-2: Protein accumulation in atypical parkinsonian disorders.

α -synucleinopathies	Tauopathies
PD (Lewy bodies, neuronal)	PSP
DLB (Lewy bodies, neuronal)	CBD
MSA (Glial cytoplasmic inclusions)	FTD parkinsonism linked to chromosome 17

Ref: PD: Parkinson’s disease; DLB: dementia with Lewy bodies; MSA: multiple system atrophy; PSP: progressive supranuclear palsy; CBD: corticobasal degeneration; FTD: fronto temporal dementia.

In this thesis I will focus primarily in MSA and primary basal ganglia calcification and therefore these disorders will be described in detail in the following sections.

2.2 MULTIPLE SYSTEM ATROPHY

2.2.1 History, definition and epidemiology of MSA

The term Multiple System Atrophy (MSA) was first introduced by Graham and Oppenheimer in 1969(31) joining together three diseases that were until then considered distinct entities: olivopontocerebellar atrophy(32), Shy-Drager syndrome(33) and striatonigral degeneration(34). In 1989, the clinical phenotype of MSA was unified and the first clinical diagnostic criteria proposed(35). In the same year, Papp and Lantos permanently consolidated these three disorders after discovering glial argyrophilic cytoplasmic inclusions (GCIs) as the pathological hallmark of the disease(36). In 1999 the main component of GCIs was identified as misfolded, hyperphosphorylated, fibrillar α -synuclein(37). These findings have classified MSA as an α -synucleinopathy, neuropathologically linked to other synucleinopathies: idiopathic Parkinson's disease (IPD), dementia with Lewy Bodies (DLB)(38–40) and primary autonomic failure (PAF)(41).

MSA is a sporadic progressive neurodegenerative disease and a distinct clinicopathological entity. Clinically, MSA is characterized by a variable combination of autonomic dysfunction, parkinsonism, cerebellar ataxia and pyramidal signs, and may change its clinical emphasis as it evolves(35). Patients can be designated MSA-P or MSA-C according to the initial predominant parkinsonian or cerebellar feature(42). Often MSA-mixed is used when cerebellar and parkinsonian signs equally occur at presentation. Pathologically, MSA consists of positive α -synuclein GCIs in the basal ganglia, cerebellum, pons, inferior olivary nuclei and spinal cord accompanied by neurodegeneration and gliosis(38).

2.2.2 Epidemiology

Although MSA is a rare disease it may account for up to 10% of cases with primary parkinsonism(18,43) and up to 34% of MSA cases may remain misdiagnosed until autopsy(43). MSA has an estimated prevalence of 3-4 per 100,000 individuals of 50-99 years of age(44). MSA mostly affects both sexes equally and the two types of MSA, MSA-P and

MSA-C, show a different distribution within populations. European and North American MSA presents in a 60-68% as MSA-P and 13-32% as MSA-C(45,46). The remaining 27% in North America was classified as a mixed-type(46). Studies in Japan have reported frequencies for MSA-P of 33% and 16% and MSA-C 67% and 84% in two retrospective series(47,48). Surprisingly, in a large Korean retrospective series, the predominant variant of MSA was similar to that of western populations(49).

Although some environmental exposures have been reported in connection to MSA, there is no clear environmental cause for MSA, and there is a largely accepted concept that most likely the aetiology of the disorder will lie within a spectrum of environmental and genetic contributing factors (44).

2.2.3 Spectrum of clinical phenotype

The core features of MSA are parkinsonism, cerebellar ataxia, autonomic dysfunction and pyramidal signs.

Parkinsonism is defined by the presence of bradykinesia and rigidity, accompanied by postural instability and tremor(43). MSA patients often present with an akinetic-rigid syndrome. Postural instability occurs earlier and progresses more rapidly than in PD(50).

Although levodopa response is a red flag for the clinical diagnosis of MSA, some degree of efficacy with levodopa treatment has been reported in 28-65% of MSA cases(43,51–56). A 51% symptom improvement due to levodopa was found in clinical MSA cases by the German Competence Network on PD(56) and a trial of this treatment is warranted in MSA-P patients.

Many patients with MSA present dystonia. The most common sign of dystonia is antecollis that is present in up to 12% of definite MSA cases(43,51,57); other features seen in MSA are torticollis, focal limb dystonia, axial dystonia and orofacial dystonia(51,58).

Progressive gait ataxia is the most common cerebellar feature of MSA. It is often accompanied by cerebellar dysarthria and oculomotor dysfunction; limb ataxia is less prominent(50).

Early oculomotor abnormalities in MSA may include square wave jerks, nystagmus with a jerky pursuit and dysmetric saccades. Other eye signs are nystagmus and limitation of up and down-gaze(18,50,59). Excessive square-wave jerks, mild hypometria of saccades, impaired vestibulo-ocular reflex, spontaneous or positional downbeat nystagmus are diagnostic clues of MSA but slow saccades or moderate to severe gaze restriction suggest a diagnosis of PSP(60).

Dysarthria is present in most cases of MSA and can present as an hypophonic, monotonous parkinsonian speech or a cerebellar quivering, high-pitched, strained, slurred and scanning dysarthria or a combination of both(43,59).

Pyramidal signs are a feature in MSA. An extensor plantar response can be seen in up to 41% of MSA cases and hyperreflexia in 46%(45,59).

Dysphagia is frequent in MSA, and accounts for high morbidity(43,61).

Non-motor symptoms (urinary problems, erectile dysfunction (ED), syncope, REM sleep behaviour disorder, stridor, faecal incontinence, constipation and sudomotor disturbances) occurring before any motor disorder, are present in 31% of MSA cases(62). Gastrointestinal symptoms, pain, urinary symptoms, orthostatic hypotension (OH), sleep disturbances, fatigue, attention and memory impairment, and psychiatric disorders were present in more than half of MSA cases in an observational multicentre study in Italy(63).

Symptomatic dysautonomia was present in 99% of patients in the final analysis of the European MSA Registry and autonomic dysfunction was the most frequent feature of MSA(64). Autonomic failure typically comprises urogenital dysfunction and OH(65). MSA usually begins with bladder dysfunction in females, and with ED in males(66). Although early ED is nearly universal in men with MSA, ED is an unspecific sign(67). The most common gastrointestinal symptom seen in MSA is constipation and can be severe. In addition, anhidrosis is a frequent finding(68).

Nearly all patients with MSA have some form of sleep disruption. Sleep apnoea and laryngeal stridor are some of the breathing problems that affect MSA patients. Stridor occurs in a large proportion of MSA patients(69). Obstructive sleep apnoea can also cause

sudden death(59,62). REM sleep behaviour disorder represents the most common clinical and polysomnographic finding in MSA(69). Other sleep disorders in MSA are sleep apnoea, excessive daytime sleepiness, insomnia and restless leg syndrome(45,70).

Cognitive impairment has been largely considered unrelated to MSA and the diagnostic criteria regard dementia as a non-supportive feature(50). However, a prospective cohort of 372 patients with MSA showed a 20% impairment in the Mattis Dementia Rating Scale and 32% on the Frontal Assessment Battery(71). Furthermore, MSA-P patients seem to show more severe and widespread cognitive dysfunctions than MSA-C patients(72). Depression and anxiety are present in MSA in approximately 40% of cases(45,73).

Mild hyposmia was found in clinical MSA cases(74–77), but a retrospective analysis of 4 definite MSA cases showed normal olfaction in all of them(78). This can be helpful information when differentiating MSA from IPD.

2.2.4 Pathology

Pathologically, depending on the anatomic brain regions of predominant neuronal loss, MSA is classified into two variants: olivopontocerebellar atrophy (OPCA) and striatonigral degeneration (SND) that correspond to the clinical subtypes of MSA-C and MSA-P respectively. Macroscopically, the overall weight of an MSA brain is usually not altered. Atrophy of the cerebellum, middle cerebellar peduncle and pontine base are seen in OPCA cases and atrophy and dark discolouration of the putamen, more pronounced posteriorly are seen in SND cases(79). Histopathologically, neuronal loss, gliosis, myelin pallor and axonal degeneration are consistent findings in MSA. Neuronal loss is more severe in the substantia nigra, dorsolateral zone of the caudal putamen, locus coeruleus, vermis, cerebellar hemisphere, inferior olivary nucleus, intermediolateral cell column and Onuf's nucleus(51,80). In addition, neuronal loss is also present in the motor and supplementary motor cortex of MSA brains(81,82). White matter tracts associated with striatonigral and olivopontocerebellar regions, such as the external capsule, striato-pallidal fibres, transverse pontine fibres and cerebellar white matter have a significant reduction in myelin staining(36,79). Moreover, reactive astrocytes(83) and activated microglia are common and important histological findings(79,80).

The neuropathological hallmark of MSA is the presence of α -synuclein immunoreactive GCIs which occur in oligodendrocytes(79). They are a highly complex proteinaceous aggregates(79,84) and their main component consists of misfolded, fibrillar, hyperphosphorylated α -synuclein(39,40). The frequency of α -synuclein GCIs correlates significantly with the severity of neuronal loss and disease duration(80). In contrast to Lewy bodies, GCIs are rarely if ever observed in normal individuals lacking the clinical manifestations of a movement disorder(85). An example of a section of the brain showing GCIs is presented in Figure 2-1.

A pathological variant of MSA, minimal change MSA, has been described in two cases with unusual age of onset (under 40 years) that pathologically presented with widespread GCIs and neuronal loss restricted to the substantia nigra and locus coeruleus(86). This is an important finding that shows that the accumulation of α -synuclein in GCIs are an early event in MSA pathogenesis.

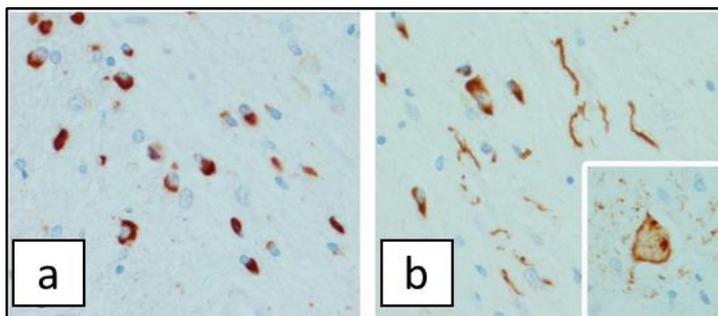


Figure 2-1: Immunohistochemistry for α -synuclein presenting many glial cytoplasmic inclusions in the pontine base (a) as well as neuronal threadlike processes (b). Inset in (b) shows a neuronal cytoplasmic inclusion in the inferior olivary nucleus. Modified from (87). Magnification 400X.

2.2.5 Pathophysiology

α -synuclein aggregates seem to play a fundamental role in MSA pathogenesis. However, the origin of the α -synuclein deposited in GCIs is not yet understood(65). α -synuclein expression is absent in oligodendrocytes of normal brains(39,88), and although an overexpression of α -synuclein in this cells has been proposed as a possible hypothesis(89), α -synuclein mRNA is not expressed in oligodendrocytes of MSA brains or normal controls(90), suggesting an ectopic accumulation. One hypothesis postulates that oligodendrocytes might be unable to degrade α -synuclein or that the accumulation of α -synuclein may overcome their ability to degrade it(65).

Early myelin alterations in MSA brains have been demonstrated by the presence of altered myelin basic protein and p25 α (tubulin polymerization promoting protein TPPP)(65,91). P25 α colocalizes with α -synuclein positive GCIs and is abnormally accumulated in oligodendrocytes(65,92–95). P25 α is redistributed from myelin to the cell soma preceding α -synuclein aggregations, and this is accompanied by an increase in cell body size, suggesting that p25 α might play a role in early events during the formation of GCIs(91,95). Most of these findings have strengthened the hypothesis that MSA is a primary oligodendroglialopathy with GCI accumulation causing oligodendroglia-myelin degeneration(91).

In mouse models, microgliosis is a severe and consistent feature(96) and microglial activation seems to play an important role in MSA pathogenesis. Moreover, microglial numbers and activation were suppressed with minocycline in a mouse model(97).

However, there are also accumulations of fibrillar α -synuclein in neurons and these seem to be relevant to the disease process(98–100). Based on these findings and in similarities with other synucleinopathies, two coexisting degenerative processes in MSA have been proposed: GCI-linked oligodendroglialopathy with secondary neurodegeneration and neuronal α -synucleinopathy associated with development of aggregates(100). Nevertheless, considering minimal change MSA, where GCIs are widespread and neuronal loss is confined to striatonigral or olivopontocerebellar systems, and given that neuronal inclusions show no correlation with GCI distribution, a primary event in glial cells seems to be more likely(65,91).

2.2.6 Genetics

Even though only a handful of patients have ever reported a family history of MSA, there are definite MSA families reported in the literature with different possible modes of inheritance (AD, AR). The family with likely dominant inheritance is a German family described in detail in chapter 5. The recessive family is of Japanese origin and described in chapter 4. And there is a British family with MSA in first cousins that can be interpreted as dominant with reduced penetrance(101–104).

Moreover, familial aggregation with a history of parkinsonism in first degree relatives has been found in MSA in a French study(105). Given the prevalence of MSA, the chance of an MSA patient of having a first degree relative with MSA, is approximately 6×10^{-5} . Therefore, although MSA mainly presents as a sporadic disorder, I hypothesize that there are genetic risk factors that contribute to the disease. Additionally, the other synucleinopathies (namely PD and DLB) that share the accumulation of alpha-synuclein in the brain have both been linked to numerous genetic risk factors.

Most of the genetic studies so far have been candidate gene/locus approaches targeting mostly genes linked to other synucleinopathies or cerebellar ataxias. A comprehensive list presenting these studies is shown in Table 2-3:

Table 2-3: Table presenting genetic studies in MSA. They are almost all candidate gene/locus approach except for the MSA GWAS and the COQ2 study. Table modified from (106).

Positive findings	Negative findings	
Ataxia genes		
SCA3(107)	SCA1–3, 6–8, 12,17 (108–111)	
SCA1 (112)		
SCA8 (113)	FMR1 (114,115)	
Parkinson's disease genes		
SNCA (116–118)	SNCA (119–124)	SNCAIP (120)
MAPT (124–126)	MAPT (117,120)	CYP1A, GMST1, NAT2, DAT1 (127)
GBA (128)	GBA (129–132)	CYP2D6 (133,134)
	LRRK2 (135–138)	PARK2, PINK1(139)
Genes associated with oxidative stress and inflammation		
SLC1A4, SQSTM1, EIF4EBPI (140)	CHOP, ATF3, ATF4, CEBPB, CARS (140)	
IL1A (141)	IL-1RA, IL-6, IL-10, TGF- β 1, TNF (142)	
IL1B, TNF (142)	IGF1, HLA (108)	
IL8/ICAM-1 (143)		
ACT-A/A (144)		
Other genes		
ADH1C (145)	ADH7 (146)	mtDNA (147)
SHC2 (148)	SHC2 (149)	APOE (120)
C9orf72 (150)	DM2 (151)	PGRN (152)
Potentially interesting in GWAS: FBXO47, ELOVL7, EDN1, MAPT (124)	BDNF (153)	
COQ2 (154–157)	COQ2 (124,158–163)	UCHL-1 (164)
	PRNP (165)	CNTF, hiGIRK (108)
	PRNP (166)	DBH (167)

The largest study that was ever performed in MSA, a GWAS that included 918 patients and 3,864 controls of European ancestry, failed to identify any significant associations after Bonferroni correction for multiple testing(168).

In this section I will first present the most important genes that have been linked to MSA and I will then comment on the MSA GWAS.

2.2.6.1 SNCA

The *SNCA* gene has been subject to extensive research due to the deposition of α -synuclein in the brain and peripheral nervous system of several neurodegenerative disorders including MSA, PD, DLB and PAF. The function of α -synuclein is not completely understood, but a role on neurotransmission has been proposed(89). Although the pathophysiology of MSA is still largely under investigation, the exact mechanism by which α -synuclein accumulates in the brain and its links to neuronal degeneration is still subject of extensive research. Notwithstanding, *SNCA* remains an interesting candidate gene in MSA studies.

Genomic duplications of *SNCA* have been observed to cause parkinsonism, dementia and autonomic dysfunction with inclusions in the brain, consistent with the MSA phenotype(169,170). However, no pathogenic mutation has been found in MSA, and direct sequencing of the exons of the *SNCA* gene in eleven pathologically confirmed MSA cases did not reveal coding variants(171). Although this hypothesis is supported by increased levels of the α -synuclein protein in MSA blood plasma, two other studies show no differential *SNCA* mRNA expression pattern in brain samples of patients with MSA(172–175).

A recessive association of a SNP (rs111931074) in the 3' untranslated region (UTR) of the *SNCA* gene which increases the risk for MSA by ~6-fold in a subset of pathologically confirmed cases(Figure 2-2)(117) has been reported. This association was independently confirmed in a second pathological MSA series when these data was merged with the original study by Scholz et al (Figure 2-3)(118). A third study has also suggested that variants across the *SNCA* locus increase the disease risk(Figure 2-4)(116). However, this association could not be replicated in a Korean cohort of MSA cases (Figure 2-5) or in the largest genome-wide study on MSA that was recently published(168).

Table. Nine Most Significantly Associated Single Nucleotide Polymorphisms

SNP ID	Chromosome	Gene	Risk Allele	Screening Stage		Replication Stage		Combined	
				p_{min} (test model)	OR (95% CI) [RR vs (RP + PP)]	$p_{recessive}$	OR (95% CI) [RR vs (RP + PP)]	$p_{recessive}$	OR (95% CI) [RR vs (RP + PP)]
rs11931074	4q22.1	Downstream of SNCA	T	1.7E-07(recessive) ^b	5.4 (2.7-11.1)	1.6E-04 ^a	6.6 (2.15-19.93)	5.5E-12	6.2 (3.4-11.2)
rs3857059	4q22.1	SNCA	G	6.9E-04(recessive)	3.8 (1.7-8.5)	1.3E-06 ^a	9.8 (3.20-29.78)	2.1E-10	5.9 (3.2-10.9)
rs9480154	6q25.1	Downstream of PPP1R14C	A	1.6E-05(recessive) ^b	5.0 (2.2-11.2)	0.99	1.0 (0.12-8.81)	1.3E-04	3.9 (1.8-8.2)
rs2794256	6q22.31	LOC728727	T	1.7E-03(recessive)	1.7 (1.2-2.5)	0.17	1.6 (0.81-3.19)	4.0E-04	1.7 (1.3-2.4)
rs2042079	2p24.2	Intergenic	A	2.7E-03(recessive)	1.7 (1.2-2.5)	0.21	1.6 (0.77-3.18)	8.0E-04	1.7 (1.3-2.4)
rs13139027	4p16.2	Upstream of MSX1	A	2.5E-03(recessive)	3.9 (1.5-10.1)	0.53	1.5 (0.41-5.63)	1.8E-03	3.2 (1.5-6.9)
rs2515501	8p23.2	MCPH1	T	6.5E-04(recessive)	2.4 (1.4-4.1)	0.45	0.6 (0.13-2.52)	7.0E-03	1.9 (1.2-3.2)
rs2896159	7q31.2	Intergenic	T	3.0E-03(recessive)	0.7 (0.5-1.1)	0.38	1.3 (0.73-2.26)	0.43	1.3 (1.1-1.6)
rs2856336	12p13.2	ETV6	C	1.6E-08(recessive) ^b	4.6 (2.6-8.3)	0.12	— ^c	2.4E-05	3.1 (1.8-5.5)

Figure 2-2: SNCA associated SNPs from Scholz and colleagues (modified from (168))

Series	Affection Status	Samples, No.	GG, No. (%)	GT, No. (%)	TT, No. (%)	T allele, No. (%)	Odds Ratio (95% CI)	p Value
Mayo	MSA	58	46 (79)	9 (16)	3 (5)	15 (13)	4.71 (1.03-21.65)	0.06
	Control	350	304 (87)	42 (12)	4 (1)	50 (7)		
Scholz et al ¹	MSA	92	66 (71)	20 (22)	6 (7)	32 (17)	12.26 (4.85-31.01)	<0.00001
	Control	3889	3303 (85)	564 (15)	22 (1)	608 (8)		
Combined	MSA	150	112 (75)	29 (19)	9 (6)	47 (16)	9.32 (4.03-21.55)	<0.00001
	Control	4239	3607 (85)	606 (14)	26 (1)	658 (8)		

Figure 2-3: Combined data from Scholz and colleagues (176) and Ross and colleagues (modified from (118)) for pathologically confirmed MSA samples for the SNP rs111931074.

SNP	Assoc p value	After correction for multiple testing	Replication p value	p value MSAC
rs3822086 (in LD with rs11931074)	0.0044	0.047	0.0347	0.024
rs3775444	0.011	0.18		0.0016

Figure 2-4: SNCA SNP analysis from Al-Chalabi and colleagues showing a SNP in LD with rs11931074 associated with increased risk of MSA (modified from (116))

TABLE: Allele Frequencies of SNCA rs111931074 in MSA and Controls								
Series	Affected Status	Number of Samples	GG [n (%)]	GT [n (%)]	TT [n (%)]	T allele [n (%)]	Odds Ratio (95% CI)	p-Value
Present study	MSA	100	21 (21)	42 (42)	37 (37)	116 (58)	0.92 (0.52–1.63)	0.77
	Control	100	23 (23)	38 (42)	39 (39)	116 (58)		
Ross et al.	MSA	58	46 (79)	9 (16)	3 (5)	15 (13)	4.71 (1.03–21.65)	0.06
	Control	350	304 (87)	42 (12)	4 (1)	50 (7)		
Scholz et al. ^{1,a}	MSA	501	415 (83)	67 (13)	19 (4)	105 (10)	6.2 (3.4–11.2)	5.5E–12
	Control	4,423	3,753 (85)	642 (15)	28 (1)	698 (8)		

Figure 2-5: No association of rs111931074 in the SNCA gene in the Korean population. From (modified from (123))

Furthermore, point mutations in SNCA have been reported in AD families exhibiting some features of MSA and presenting pathological criteria of MSA and PD(177,178).

2.2.6.2 COQ2

Coenzyme Q10 (CoQ10), is a lipophilic molecule located in the inner mitochondrial membrane and functions as an electron transporter from complexes I and II to III in the respiratory chain(179). In its reduced form (ubiquinol) it also has an antioxidant role. Moreover, CoQ10 also transports electrons in extramitochondrial systems and participates in the regulation of mitochondrial permeability pores(179). CoQ10's deficiency is suspected when the activities of complexes I+III and or II+III are deficient in the presence of normal activities of individual complexes(180). CoQ10 is a member of a homologous family of ubiquinones that comprise a benzoquinone ring and a polyprenyl "tail" that acts as an anchor attaching the coenzyme to lipid membranes. The numerical subscript used in coenzyme Q nomenclature refers to the number of isoprenyl units in the tail of the molecule. The predominant coenzyme Q in humans is CoQ10(179). The ring is derived from tyrosine or phenylalanine and the mevalonate pathway generates the decaprenyl tail. The intermediates 4-hydroxibenzoate and decaprenyl diphosphate are then condensed by 4-hydroxibenzoate: polyprenyl transferase, an enzyme encoded by COQ2. After that it is thought that there are eight subsequent steps that modify the 6-carbon ring and its side chains to form the functional CoQ10 molecule as shown in Figure 2-6(179,181).

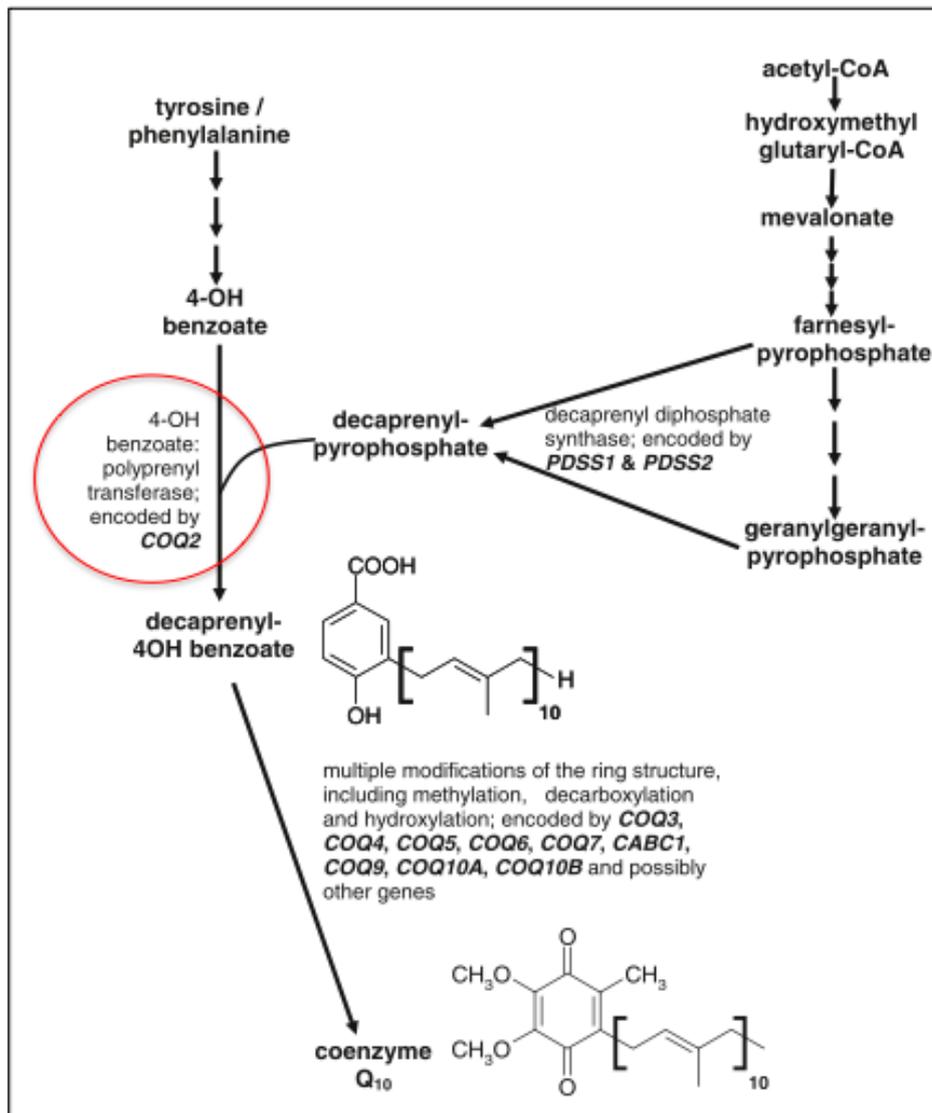


Figure 2-6: Biosynthetic pathway of CoQ10 (modified from (103)).

COQ2 encodes an enzyme called 4-hydroxybenzoate: polyprenyl transferase that condenses 4-hydroxybenzoate and decaprenyl diphosphate. This has been highlighted in red in Figure 2-6 that presents the pathway that synthesises CoQ10. Mutations in *COQ2* have been reported to cause primary coenzyme Q10 deficiency. Mutations described so far include: homozygous Y297C(182), N401fsX4(183), S146N(184), and compound heterozygotes R197H+N228S(184). Phenotypes include cases with isolated nephropathy and infantile multisystem disorder(180).

2.2.6.2.1 COQ2 and MSA

An international collaboration led by Tsuji et al. from Japan recently published a paper that links *COQ2* with MSA(185). The study began studying MSA families of Japanese origin. There were 6 families (Figure 2-7). Family 1 and family 8 have definite MSA. Families 2,3,4 and 12 have probable MSA. Family 1 consists of 2 affected siblings born to a consanguineous marriage strongly suggestive of autosomal recessive inheritance. The affected cases in family 1 also presented retinitis pigmentosa. The rest of the families have 2 siblings affected and they are also interpreted as recessive in the paper. Intriguingly, although it is well established that MSA-C subtype is the most frequent in the Japanese population (in contrast to European populations) these families comprise mostly MSA-P patients.

They first performed whole-genome linkage analysis using Affymetrix SNP 6.0 arrays. They analysed the data doing parametric linkage analysis of the six pedigrees and they could not find any linkage locus with recessive inheritance. They later analysed these data with both parametric analysis allowing for heterogeneity and with non-parametric linkage analysis and they noted a positive LOD score in individual pedigrees in an overlapping region in families 1,2,4, and 12, with family 1 having the highest LOD score of 1.93. The linkage region in family 1 spanning 80 Mb includes the following regions: in chromosome 4 (72.795Mb to 89.616Mb), chromosome 5 (149.50Mb to 168.32Mb), chromosome 6 (85.499Mb to 87.382Mb), chromosome 7 (62.754Mb to 64.907Mb), chromosome 9 (99.781Mb to 115.484Mb), and chromosome 13 (75.849Mb to 98.253Mb).

They performed whole genome sequencing in the proband of family 1 and they filtered for SNPs and indels in the linkage regions, present in exons or splice sites. And finally, they focused on novel variants not present in dbSNP130. They found 4 missense variants: p.K707R in *SHROOM3*, p.M78V and p.V393A in *COQ2*, and p.R231G in *SCEL*.

They screened for the four novel variants in 180 Japanese controls and they did not observe the variant p.M78V in *COQ2* but they did find the other 3 variants. This is the reason why they considered this variant as a candidate of susceptibility to familial MSA and they investigated this gene in all the families.

When investigating *COQ2* in all the Japanese MSA families, they detected homozygous missense p.M78V-p.V343A variants in the 2 affected cases of family 1 (and the unaffected sibling tested was a non-carrier) and compound heterozygous variants in both affected siblings of family 12: p.R337X + p.V393A. The mother was heterozygous for p.V393A, one unaffected sibling did not carry the variant and the other unaffected sibling that was tested (participant II-2) was heterozygous for p.R337X.

They also sequenced other genes in CoQ10 pathway including: *PDSS1*, *PDSS2*, *COQ3*, *COQ4*, *COQ5*, *COQ6*, *COQ7*, *ADCK3*, *COQ9*, *COQ10A*, and *COQ10B*. They did not find any variants that segregate with disease in these genes.

They have later extended their study to sporadic MSA and controls and detected variants listed in sporadic MSA. They performed an association study of the frequency of the variant p.V393A in sporadic MSA versus controls. When comparing 363 MSA patients with 2383 controls they found an odds ratio of 2.23 ($p = 6.0 \times 10^{-5}$).

With these genetic data they decided to do functional work. They performed a yeast complementation assay and detected decreased growth rates in *COQ2* mutants. They later investigated *COQ2* activity in lymphoblastoid cell lines and found a significant decreased activity in *COQ2* homozygotes and compound heterozygotes. Lastly, they looked at the CoQ10 levels in brain tissue of 3 MSA cases and 3 controls and they observed a decreased level in MSA (Figure 2-8).

Finally, they performed an association study with the variants that were functionally impaired in sporadic MSA cases versus controls. They merged their data in Japanese MSA and Caucasian cases and they found a significant association of rare functionally impaired variants with increased risk of MSA. The odds ratio was 11.97 with a p value of 0.004.

The authors concluded that homozygous or compound heterozygous *COQ2* mutations are a cause of familial MSA, and *COQ2* variants (in particular p.V393A) constitute a susceptibility factor of sporadic MSA.

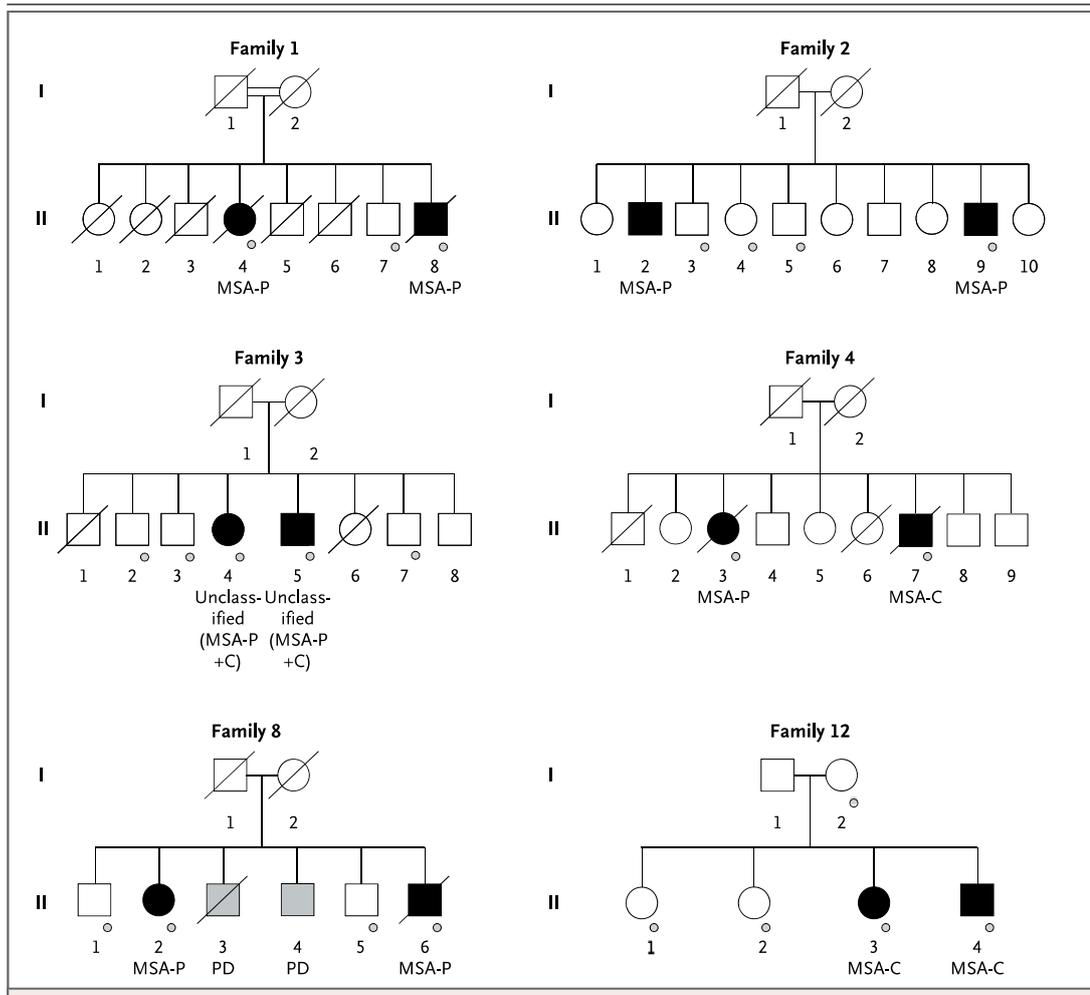


Figure 2-7: Japanese MSA families (Reproduced with permission from (185), Copyright Massachusetts Medical Society.

Ref: MSA-P: multiple system atrophy with predominant parkinsonism; MSA-C: multiple system atrophy with predominant cerebellar signs; PD: Parkinson's disease.

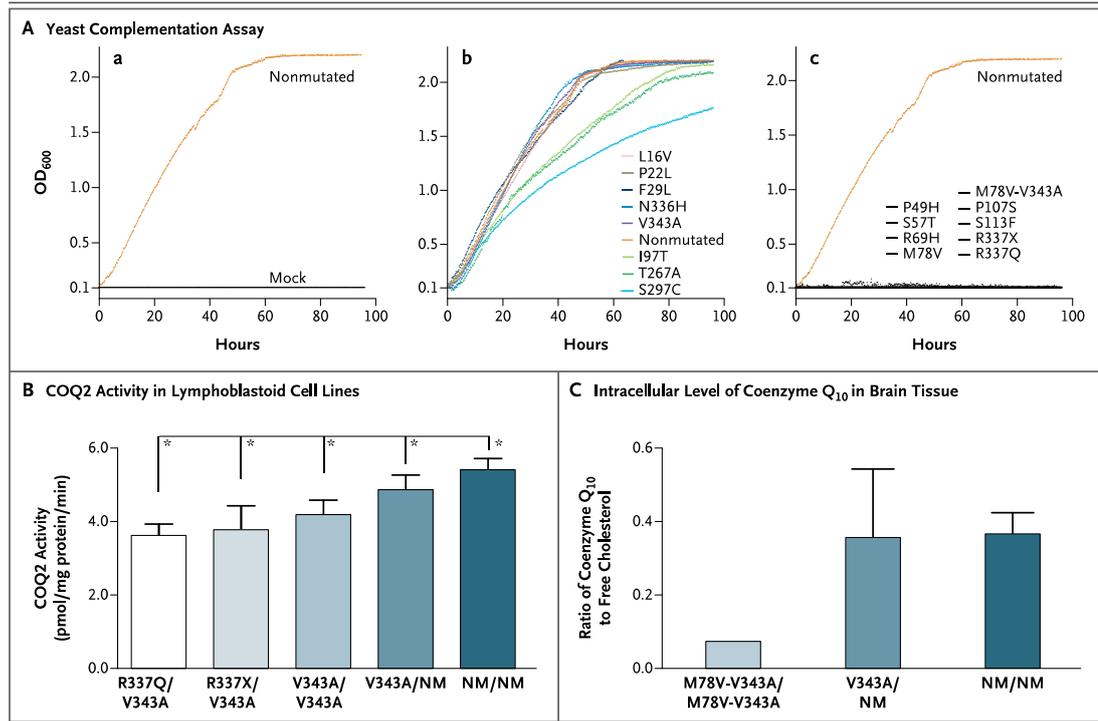


Figure 2-8: A: Yeast complementation assay; B: COQ2 activity in lymphoblastoid cell lines; C: Levels of CoQ₁₀ in brain tissue. Reproduced with permission from (185) Copyright Massachusetts Medical Society.

Ref: OD: optic density.

2.2.6.3 MSA GWAS

After a large collaborative effort, the largest study that was ever performed in MSA was published recently. After quality control, this GWAS included 918 patients and 3,864 controls of European ancestry. Unfortunately, it failed to identify any significant associations after correction for multiple testing(168). However, a number of genes in *loci* with the lowest *p*-values that did not reach the stringent Bonferroni threshold were proposed are candidates: *FBXO47*, *ELOVL7*, *EDN1*, and *MAPT*. *MAPT* remains an interesting candidate because association with this gene has been previously reported(125,126) and also tau has been found in MSA inclusions(186). This finding supports the notion that a genetic overlap with PD/parkinsonism exists and may promote the study of other PD-related genes in MSA.

2.2.6.4 Other associations

A recent multinational study found an increased risk of MSA with *GBA* variants but this result awaits replication(128).

SHC2 CNVs have been proposed in a study that began studying a couple of discordant MSA identical twins and later found striking CNVs in 10 out of 31 clinically diagnosed MSA cases(148). This association however, failed to replicate in a study investigating clinically well characterised MSA cases(149).

2.2.7 Investigations

When a clinical diagnosis of MSA is suspected and other mimics such as a tumour and paraneoplastic phenomenon have been excluded, imaging, autonomic function tests and urology are important to help support a diagnosis of MSA(187). Bladder function assessment often detects early abnormalities consistent with neurogenic dysfunction. Urodynamic tests frequently indicate detrusor hyperreflexia and abnormal urethral sphincter function followed later in disease progression by increased residual urine volume as detected by ultrasound. Cardiovascular autonomic dysfunction in MSA can be investigated by either a standing blood pressure test, or by tilt-table testing(38). Imaging of cardiac innervation with single photon emission computed tomography (SPECT) and [123I]metaiodobenzylguanidine (MIBG) and with positron emission tomography (PET) and [18F]fluorodopa have shown preserved sympathetic postganglionic neurons in MSA, in contrast to PD(50,188). However, some denervation has been reported in MSA(189) and more recently severe cardiac denervation was reported in MSA using PET and [11C]hydroxyepinephrine(190).

Magnetic resonance imaging (MRI) demonstration of putaminal, pontine, and middle cerebellar peduncle atrophy is helpful(191). Posterior putaminal hypointensity, hyperintense lateral putaminal rim, hot cross bun sign, and middle cerebellar peduncle hyperintensity on T2-weighted images can also be useful in MSA(50). Unfortunately, although the hot cross bun sign and the slit-like void signal are features of MSA, they are non-specific findings). Striatal or brainstem hypometabolism demonstrated by functional imaging with PET [18F] fluorodeoxyglucose can also aid diagnosis(50,192). Finally, the

consensus criteria have defined neuroimaging pointers that can aid to MSA diagnosis in possible cases(50).

2.2.8 Diagnosis and prognosis

Definite diagnosis of MSA requires pathology. However, great efforts have been made to approach an accurate diagnosis in living patients. Clinical diagnostic criteria were first proposed by Quinn in 1989(35) and divided MSA into three categories: possible, probable and definite. In 1998 and 2008, the first and second consensus criteria were published and the subdivision of MSA-P and MSA-C was introduced(50,193). Definite cases have pathologically confirmed CNS α -synuclein positive GCIs with neurodegenerative changes in striatonigral or olivopontocerebellar structures. Probable MSA have progressive adult-onset autonomic abnormalities, poorly levodopa-responsive parkinsonism or cerebellar ataxia; whereas possible cases have progressive adult onset disease including parkinsonism or cerebellar ataxia and at least one feature suggesting autonomic dysfunction plus one other feature that may be a clinical or a neuroimaging abnormality. In addition, the consensus conference pointed out supporting and non-supporting features useful to differentiate MSA cases from other diseases(50).

2.2.8.1 Differential diagnosis

The most common diagnostic pitfall for MSA-P patients is to be confused with PD. This can account for up to 55% of misdiagnosed MSA-P. PSP and CBD are the next differentials to be considered(194). Moreover, cases of DLB can present with prominent autonomic features and could also be confused with MSA-P(195). When approaching an MSA-C case one should always consider the dominantly inherited spinocerebellar ataxias 1,2,3,6, and 7, even in the absence of a family history(50). About 24% of cases with late onset cerebellar ataxia will turn out to have confirmed MSA(196).

Likewise, fragile X tremor ataxia syndrome needs to be ruled out(194). Furthermore, in patients with an aggressive clinical course of cerebellar syndrome, even in the absence of malaise, a paraneoplastic disorder should be investigated(50). Finally, cerebrovascular diseases and primary progressive multiple sclerosis may mimic MSA and can be ruled out with imaging(194). Although less likely, cases of ALS can also be misdiagnosed as MSA(71).

Patients that present initially with autonomic symptoms are often considered to have PAF. Important research in these disorders have recently pointed to specific clinical features that can help identify patients that are more likely to phenoconvert from PAF to MSA(197).

2.2.8.2 Prognosis

MSA-P has a much faster disease progression when compared to PD(46), and the unified MSA rating-scale (UMSARS) shows an annual decline in these patients(198). Reported median survival times are from 6.2 to 10 years (range 0.5-24) (47,49,59,199,200) and mean disease duration from 3.2 to 7.9 years (range 1-17)(56,64,201–204). Early autonomic failure, older age at onset(33,43,121), a shorter interval from disease onset to reaching the first clinical milestone, and not being admitted to residential care have been reported as factors predicting shorter disease duration in a study that included 83 definite MSA cases(64). In addition, a study on 230 Japanese patients found that evolution from initial symptoms to MSA within 3 years strongly predicted a shorter survival(47). Early development of autonomic failure in a study that analysed 49 patients with definite MSA showed to be a factor for shorter survival(205). Stridor onset can also play a role in predicting poor prognosis in MSA(206).

The most common causes of death are, sudden death (cardiopulmonary arrest), urinary infections, infectious pneumonia, aspiration pneumonia and wasting syndrome(207–209).

2.2.9 Treatment

In the European MSA study group (EMSA-SG) final analysis of the European MSA registry, which enrolled 437 MSA patients, the management of MSA was inconsistent and different between centres. Only 36% of patients with dysautonomia and 82% with parkinsonism received pharmacological treatment(45).

2.2.9.1 Non-Medical treatment

Physiotherapy is very important in helping patients with balance and maintaining mobility, preventing contractures and improving functional abilities. Speech therapy is essential for communication purposes and improving swallowing. Occupational therapy has shown

amelioration of motor impairment and quality of life(210–212). Patients may also require psychological input.

2.2.9.2 Medical and surgical treatment

Motor impairment, autonomic dysfunction and depression are associated with a poor health-related quality of life in MSA; hence therapeutic management should target these features(210–213). Moreover, breathing problems could be a cause of sudden death in MSA patients and therefore they should be given appropriate consideration(209). Deep brain stimulation (DBS) is currently not recommended for MSA, it can have significant complications and if any benefit is seen, it is usually time-limited(38,210,214).

Table 2-4 presents the usual symptomatic treatment of MSA patients. Ideally all patients should have access to expert advice in a multidisciplinary setting.

Table 2-4 Table presenting symptomatic treatment of MSA. (Modified from (87)).

<p><i>Non-medical treatment:</i></p> <p>Should be offered to all MSA patients.</p>		<ul style="list-style-type: none"> • Clinical nurse specialist advice • Physiotherapy • Occupational therapy • Speech and language therapy • Psychotherapy
<p><i>Medical treatment:</i></p> <p>In the absense of efficacious neuroprotective or preventive treatment, MSA’s management is mainly symptomatic and based on experts experience.</p>		
Movement disorders	– Parkinsonism	<ul style="list-style-type: none"> • First choice: Levodopa (up to 1000 mg/day, if tollerated and neccessary) with domperidone to prevent nausea and vomiting. It should be noted that levodopa can cause worsening of orthostatic hypotension (OH), hypersexuality, delirium and dyskinesias. • Second choice: Dopamine agonists (iPD titration schemes with extreme caution) • Third choice: Amantadine (100 mg tid) previous check of QTc on ECG.
	– Dystonia	<ul style="list-style-type: none"> • Botulinum toxin injection in oro-facial (caution in antecollis because if risk of severe dysphagia) and in limb-dystonia.
Cerebellar ataxia		<ul style="list-style-type: none"> • No drug therapy available.

Autonomic symptoms	– Orthostatic hypotension	<ul style="list-style-type: none"> • 1st choice: nonpharmacological strategies: elastic support stockings or tights, high-salt diet, frequent small meals, head-up tilt of the bed at night, ingestion of water. • If needed: add midodrine (2.5-30 mg tid) or fludrocortisone (0.1-0.3 mg) starting at night • If needed: Droxidopa (100 mg tid)
	– Urinary failure-postvoid residue <100ml	<ul style="list-style-type: none"> • Anticholinergics for detrusor hyperactivity [Trospium chloride (20 mg bid or 15 mg tid) , Oxybutinin (2.5-5 mg bid to tid), Tolterodine (2 mg bid)]. Special attention should be placed on central side effects with anticholinergic drugs. • Alpha-adrenergic antagonists for urethra hypertony (prazosin or tamsulosin). Be careful with exacerbation of OH. • Alternative treatment can be intra-detrusor or urethral sphincter botulinum toxin injection.
	– Urinary failure-postvoid residue >100 ml	<ul style="list-style-type: none"> • All patients should try clean intermittent self catheterization (CISC) • In the advanced stages of MSA urethral or suprapubic permanent catheterization may become necessary • A last option for MSA patients who do not tolerate CISC is urinary surgery
	– Erectile dysfunction	<ul style="list-style-type: none"> • First choice: Sildenafil (50-100 mg) (165). Be careful with worsening of OH. Alprostadil in case of severe OH.
	– Constipation	<ul style="list-style-type: none"> • High fluid and fibre intake, laxatives.
Other treatments	– Breathing problems	<ul style="list-style-type: none"> • CPAP, BiPAP (for prominent stridor and sleep apnoea) • Tracheostomy (in case of life-threatening and/or daytime stridor or abnormal vocal cord mobility on laryngoscopy)
	– Swallowing and nutritional problems	<ul style="list-style-type: none"> • Dietician advice • PEG when necessary
	– Drooling	<ul style="list-style-type: none"> • Botulinum toxin to the salivary glands for sialorrhea.
	– Depression	<ul style="list-style-type: none"> • Serotonine reuptake inhibitors. • Psychotherapy
	– REM sleep behaviour disorder	<ul style="list-style-type: none"> • Clonazepam • Melatonin

2.2.9.3 Neuroprotection

In the last decade, thanks to the creation of International MSA networks (EMSA-SG(215) , the Japanese MSA consortium, the NAMSA-SG(216) and the Chinese MSA study group) a number of clinical trials could be performed(217–221), NCT01146548, NCT00750867, NCT01287221, NCT00977665. Moreover, the implementation UMSARS also facilitates the unification of outcome measures. Unfortunately, there are still no clear neuroprotective

strategies that can be translated into clinical improvement, but novel therapies are under study.

2.2.10 Future and experimental developments

Only a handful of trials have been performed in MSA with ≥ 100 patients and this is in part due to the difficulties in recruiting patients with such a rare disease. By far the largest trial performed in MSA was the NIPPS study including 398 MSA patients. Unfortunately, the results from the drug treatment (Riluzole versus placebo) were negative. However, this study served as a proof of concept of the feasibility of this approach and also, was successful in obtaining natural history data in atypical parkinsonism, the validation of clinical and imaging rating scales and the generation of a biobank (with DNA used to replicate the association of variants in *SNCA* with MSA).

The trials that enrolled 100 or more MSA patients are presented in Table 2-5.

Table 2-5 Clinical trials in MSA enrolling 100 MSA patients or more.

Drug	Study	Trial number	Patients enrolled	Type of study	Outcome measures	Result
Riluzole	NIPPS	NCT00211224	398 (MSA-C and MSA-P)	Multicenter, randomized, double-blind, placebocontrolled	Three-years survival, rate of motor decline	Ineffective(218)
Rasagiline	Clinical Trial to Assess Efficacy, Safety, and Tolerability of Rasagiline in Patients With MSA-P	NCT00977665	174 (MSA-P)	Multicenter, randomized, double-blind, placebocontrolled	Safety and tolerability; 12-months change: UMSARS-total, putaminal diffusivity	Ineffective(222)
Rifampicin	Study of Rifampicin in MSA	NCT01287221	100 (MSA-C and MSA-P)	Multicenter, randomized, double-blind, placebocontrolled	12-months change: UMSARS I and II and COMPASS	Ineffective(223)

2.2.10.1 Anti-inflammatory approaches

As MSA pathophysiology accounts for important inflammatory processes, and a strong dose risk relation between increased aspirin intake and decreased risk of MSA(224) has been reported, a clinical trial with minocycline was performed but unfortunately brought negative results(220). However, preliminary results on myeloperoxidase inhibition in a

transgenic mouse model showed that this inhibition could reduce motor impairment and be protective against neurodegeneration(225) and a recent study showed failure of neuroprotection despite microglial suppression in a model of advanced MSA(226) A phase 2 trial investigating safety and tolerability with a myeloperoxidase inhibitor in MSA patients has been completed although the results have not been published yet (NCT02388295).

2.2.10.2 Neurotransplantation

Evidence in double toxin-double lesion mouse models of MSA has raised the possibility of restoring levodopa responsiveness in MSA-P by striatal allografting(227,228). However, the role of host α -synuclein pathology on grafts and of pro-inflammatory responses on host striatum and its effects on functional outcomes and graft survival is not completely elucidated(229).

2.2.10.3 Mesenchymal autologous stem cell therapy

An open and unblinded study, that consecutively injected intra-arterial and intravenous autologous mesenchymal stem cells in 29 MSA patients, showed feasibility and safety over a 12 month follow up period. The authors report a delayed progression of neurological deficits with achievement of functional improvement measured by UMSARC and PET-scan(172). However, the design of the study, the possibility of confounding effects, the lack of preclinical experimental evidence on underlying mechanisms of action are to take these results with caution(230). A double-blind placebo-controlled, randomized clinical trial with autologous mesenchymal stem cells was later performed with promising results but significant safety concerns, in particular for ischemic lesions in the brain(231). Further studies with mesenchymal cells are ongoing.

Other clinical trials have been completed and reviewed in (89). Novel treatment strategies targeting α -synuclein synthesis, degradation and accumulation are being studied and tested, some in animal models and others in clinical trials. A comprehensive review has been recently published(232).

2.3 PRIMARY FAMILIAL BRAIN CALCIFICATION

2.3.1 History, classification and epidemiology

Primary familial brain calcification (PFBC) is a neurodegenerative disorder with the common characteristic of calcium deposits in the basal ganglia and other brain regions that were originally described by neuropathology. Nowadays, this disorder can be detected in neuroimaging studies. The clinical presentation is heterogeneous, usually occurs in adulthood and mostly during the third to fifth decade of life. The manifestations can include but are not restricted to movement disorders, seizures, migraine and neuropsychiatric symptoms. The condition is inherited and most families exhibit an AD pattern although some unusual recessive families have been reported(233). The clinical picture and severity is often variable between and within families. Penetrance can be incomplete and varies depending on whether the affected status is considered by the presence of isolated brain calcification or in addition to clinical symptoms(234).

Brain calcification was originally reported by Delacour in 1850(235). In 1930, Fahr presented a well described case and hence the naming of the disease “Fahr’s disease”(236). However, it was later challenged by the suspicion that this description corresponded to a secondary cause of brain calcification (actually not affecting the basal ganglia) rather than what Fahr’s disease is commonly attributed to nowadays, which consists of primary calcification of the basal ganglia and other regions of the brain(237). Many names have been proposed for this entity and more recently, other nomenclatures are used. Among these we can find “Idiopathic basal ganglia calcification”, “primary familial brain calcification”, and “familial basal ganglia calcification”. For the purposes of this thesis and to simplify this issue to the reader we will refer this disease as “Primary familial brain calcification (PFBC)”(234). “Primary” because we do not include secondary causes of brain calcification, “familial” because it is presumed genetic and often inherited and, “brain calcification” because the calcification is not usually confined to the basal ganglia and other areas of the CNS are usually affected.

Neuroimaging studies can detect incidental calcifications on the basal ganglia and other brain regions with a frequency of 1% to 20%(238). Most of these are asymptomatic, and maybe related to normal ageing processes. However, in many cases there are affected family members that are not accounted for, who may still be asymptomatic or exhibit unrecognized signs and symptoms. A recent systematic review found that between 24%-54% of cases with reported mutations remained asymptomatic(239).

2.3.2 Molecular genetics

Linkage studies have associated PFBC with three loci: 2q37 (LOD score: 2 and 2.44)(240), 14q (LOD score: 3.37 and 4.95)(241) and 8p21.1-q11.23 (LOD score >3)(242).

More recently four genes have been linked to this disease and have explained the cause of the condition in ~60% of familial patients(243). The genes are listed in Table 2-6.

The first gene to be described was *SLC20A2* (OMIM 158378), located in the previously linked 8p21.1–8q11.23 region and was discovered by next generation sequencing. Mutations in five families of different ethnic origins (3 Chinese, 1 Brazilian and 3 Spanish) were unravelled. *SLC20A2* encodes for a type III sodium-dependent inorganic phosphate transporter (PiT-2) and their work showed that mutations in this gene cause impaired phosphate uptake in *Xenopus oocytes*(244).

The second gene published was *PDGFRB* (OMIM 173410) located in 5q32 and encodes for the platelet derived-growth factor receptor- β . It was first identified in a large French family as well as a sporadic case by exome sequencing(245).

The third gene *PDGFB* was discovered in six human families exhibiting brain calcification and also showed that mice carrying hypomorphic *Pdgfb* alleles develop brain calcifications with increased age(246).

The fourth gene linked to PFBC, is *XPR1*. Mutations were found in an American family of Swedish ancestry and detected 5 other variants, although 2 of these are likely rare polymorphisms(247).

Overall there are 2 apparent pathogenic mechanisms for this disease. On one side, it is related to phosphate homeostasis and links *SLC20A2* and *XPR1*. Inhibition of phosphate uptake by mutations in *SLC20A2* may lead to deposition of calcium in the vascular extracellular matrix, and inhibition of phosphate export associated with *XPR1* mutations is expected to increase intracellular phosphate concentration and provoke calcium phosphate precipitation(247).

On the other side, loss of function of *PDGFRB* and *PDGFB* could lead to the impairment of the pericytes function and blood brain barrier (BBB) integrity, causing vascular and perivascular calcium accumulation by BBB dysregulation(248).

However, this disease presents with locus heterogeneity. The family reported to present significant linkage to a locus in chromosome 14q(241) was later discovered to harbour an *SLC20A2* mutation instead. So finally, the cause of the condition in this family is in chromosome 8. This is possibly due to inaccuracies assessing affected status or calcification on CT scan, presence of phenocopies, incomplete penetrance or sample identification errors. It gives a good example on how these studies can sometimes become very difficult and, if the data is not clearly convincing, a re-assessment from the beginning can help(249).

In summary, the mutations found in these four genes account for more than 60%(243) of cases of PFBC. The most common genetic cause of PFBC is *SLC20A2* which accounts for up to 55% in some series, followed by *PDGFB* 10-31% and *PDGFRB* 5-11%(239,249,250).

Table 2-6: List of genes linked to PFBC.

Gene	Locus	Protein	Protein function	Mutations	Inheritance	Year of discovery
<i>SLC20A2</i>	8p11.21	Type III sodium-dependent inorganic phosphate transporter 2	Phosphate transporter	75	AD	2012

<i>PDGFRB</i>	5q32	Platelet-derived growth factor receptor-beta	Maintaining integrity of the BBB	13	AD	2013
<i>PDGFB</i>	22q13.1	Platelet-derived growth factor subunit B	Pericyte homeostasis, BBB regulation	43	AD	2013
<i>XPR1</i>	1q25.3	Xenotropic and polytropic retrovirus receptor 1	Phosphate homeostasis	6	AD	2015

Ref: BBB = blood-brain barrier; AD: autosomal dominant.

2.3.3 Clinical aspects

The clinical presentation frequently includes psychiatric signs. The disease onset can affect the type of presentation, given that the earlier the disease onset is the more likely psychiatric and cognitive features develop; and if the onset is later in life, patients most likely present with movement disorders.

The *SLC20A2* carriers usually exhibit more severe calcifications and *PDGFRB* the less severe, with carriers of *PDGFB* mutations being in the intermediate region. The best predictor of brain calcification is the combination of age (older being higher), sex (male being higher) and gene. And the calcification score is correlated with the symptomatic status. Importantly a study showed that the range on ages is very widespread and is distributed from early infancy until late adulthood(250).

A recent meta-analysis of genotype-phenotype correlation in PFBC related genetic mutations showed that although there is significant overlap in terms of clinical and radiological features, there are features significantly associated with specific mutations. With regards to significant distinctive neurological features, parkinsonism is more common with *SLC20A2* mutations and headache with *PDGFB*. Depression was more often reported in *PDGFRB*, and cognitive impairment and parkinsonism tended to occur with late onset of disease (>45 years) while younger onset cases more commonly had hyperkinetic movement disorders such as chorea and dystonia(239).

Usually a CT scan is enough to detect brain calcification. Other investigations should be performed to rule out secondary causes of brain calcification before it can be classified as primary brain calcification.

2.3.4 Treatment

There is currently no disease modifying treatment for this disease and medical treatment is mostly symptomatic and in specific cases can target parkinsonism, epilepsy, depression and headaches according to the clinical presentation of each patient.

3 CHAPTER 3: MATERIALS AND METHODS

3.1 ETHICS APPROVAL

Ethics was approved by the University College London Hospitals (UCLH) ethics committee (06/N076 movement disorders, 04/N034 cerebellar ataxias). Written informed consent was obtained from all participants or legal representatives where applicable.

All tissue stored in the Queen Square brain bank (QSBB) and in the Department of Molecular Neuroscience is under a license from the Human Tissue Authority and has been donated for research according to protocols approved by the NRES Committee London-Central.

3.2 SAMPLE COLLECTION

Sample collection varied significantly in different subprojects of this thesis and will be addressed in each chapter in detail.

Overall, samples came from different sources in the UK and overseas, and were of different types: human DNA, peripheral blood, skin biopsies and/or brain tissue as well as tissue from mice. In all occasions, I have followed standard protocols for preserving samples at the best quality possible. All samples have been stored either in the diagnostic laboratory of the National Hospital for Neurology and Neurosurgery (NHNN) or at the research laboratory of the Department of Molecular Neuroscience, UCL Institute of Neurology.

DNA from peripheral blood from patients seen at the NHNN was obtained from the neurogenetics laboratory. Collaborators are all mentioned in the relevant results chapters.

Brain samples were collected from collaborating brain banks (largely from Queen Square Brain Bank, but also from others in the UK and abroad). I produced the material tissue agreements where appropriate. This information will be provided in detail on the chapter on MSA.

Skin biopsies for growing fibroblasts in the PFBC family were obtained by Professor Henry Houlden and myself.

Mice used in chapter 6.2 were obtained from a French collaborator, Prof Michel Aurrand-Lions. Details of this collaboration as well as other members of the lab that had an active participation in this project are detailed in the PFBC chapter.

3.3 GENETIC STUDIES

3.3.1 Genetic nomenclature

Standard nomenclature from the Human Genome Variation Society (<http://varnomen.hgvs.org/>) was used to name all variants described in this thesis. The first time a variant is named, all the information of the base change, the amino acid change and the corresponding transcript are mentioned. After that, variants are named with the amino acid change or the SNP rs number when the variant is present in the dbSNP database (<https://www.ncbi.nlm.nih.gov/projects/SNP/>).

3.3.2 DNA extractions

3.3.2.1 *Peripheral blood*

Genomic DNA (gDNA) extraction from EDTA blood containing tubes was performed using 5-10 ml of fresh or frozen whole blood samples using Flexigene kit (Qiagen) according to manufacturer's instructions. In some cases, I performed this extraction and in others it was done by the diagnostic team of the neurogenetics laboratory at the NHNN. For more cost-effective extractions, DNA from some samples was extracted by a service provided by the LGC Genomics laboratory in Germany, using the S PLUS XL kit (LGC Genomics) according to the manufacturer's protocol.

3.3.2.2 *Brain*

DNA from brain tissue from some samples was extracted by myself using the DNeasy Blood & Tissue Kits (Qiagen), following the manufacturer's instructions. Other samples were extracted by a service provided by the LGC Genomics laboratory in Germany using the sbeadex® tissue kit (LGC Genomics) according to the manufacturers protocol. This was because it was more cost-effective to send them there.

3.3.3 DNA concentration and purity

The concentration and quality of DNA for all samples was measured using a NanoDrop ND-1000 spectrophotometer following the manufacturer's instructions (NanoDrop Technologies). Concentration was assessed at 260 nm. Purity was estimated by the 260/280 and 260/230 absorbance ratios, and the spectrums of the ratios between 1.8-2.0 and 1.8-2.2 respectively were considered of good quality. The concentration of the DNA samples was adjusted to appropriate values, according to the technique to be used afterwards, by diluting the samples with autoclaved distilled H₂O (dH₂O).

For the purposes of next generation sequencing (NGS) library preparation, the measurement of DNA concentration required a higher precision than the one provided by the nanodrop. The Qubit® 2.0 Fluorometer from Thermo Fisher Scientific detects the amount of fluorescent dye bound to DNA, therefore providing a direct measurement of double stranded DNA quantity. The reagents used by the Qubit do not bind to degraded DNA or other molecules such as proteins, and are more accurate than the NanoDrop. DNA concentration of samples for NGS was measured with the Qubit and diluted with dH₂O until reaching the required concentration for the library preparation used in each case. This usually required a few rounds of measurement/dilution.

3.3.4 Primer design, PCR and Sanger sequencing

Sanger sequencing was first developed in 1977 by Frederick Sanger and became the main method for sequencing DNA until the development of next generation sequencing only a few years ago. Sanger sequencing, based on the chain-terminator method, still remains the gold standard method for mutation confirmation. This method has been extensively used in this thesis, and is based on DNA polymerase randomly inhibited by a small amount of modified dideoxynucleotides (ddNTPs) which are mixed with normal deoxyribonucleotides (dNTPs). This produces newly synthesized DNA fragments of different lengths with different terminators, which enables one subsequently determine its sequence by capillary electrophoresis.

3.3.4.1 Primer design

Practically, in order to perform Sanger sequencing one first needs to design oligonucleotide primers to target the desired region of the genome. I first download the target region including exons and flanking introns of the longest transcript from <http://www.ensembl.org/index.html>, and then used the online software: <http://primer3.ut.ee/>. This software allows researchers to design specific primers for a region under specific conditions but for 1 region at a time. Optimum primer size is usually around 20 BP, with optimal melting temperatures set for between 55°C and 65°C, the primer GC content is set at around 50 % (30-70), and low self-complementarity is preferable (to avoid a large amount of primer dimers). I then blasted them on <https://genome.ucsc.edu/cgi-bin/hgPcr> to ensure they were specific for the target region, and do not bind unspecific sites, and finally on online databases to ensure there are no common SNPs within the primer sequence. Primer sequences used in this thesis can be found in the Appendix.

3.3.4.2 Polymerase chain reaction (PCR)

The regions targeted are then amplified by polymerase chain reaction (PCR). I used a combination of Roche PCR MasterMix (Roche) (that contains DNA polymerase, deoxynucleotides, magnesium and buffer), forward and reverse primers, dH₂O and DNA of the desired samples or controls as appropriate. An example of volumes used per single reaction is presented in Table 3-1.

Table 3-1: Example of PCR mix recipe.

Reagents	volume (uL) per reaction
Roche Fast start Master mix	10
dH ₂ O	5
Forward_primer (5uM)	2

Reverse_primer(5uM)	2
DNA (~25-50ng/uL)	1
Total	20

PCR reaction mixes were then loaded onto an Eppendorf Mastercycler thermal cycler. Different cycling programs can be used, usually by means of touchdown temperature according to the optimal annealing temperature of the primers. They usually contained a first cycle of denaturation of the double-stranded DNA, then annealing of the primers, and a final elongation step adding each deoxynucleotide (dGTP, dCTP, dATP and dTTP). An example is shown on Table 3-2.

Table 3-2: An example of a PCR 65 touchdown 55 cycling programme.

Step	Temperature	Time	Number of cycles
Denaturation	94°C	10 minutes	X1
Denaturation	94°C	30 seconds	X8
Annealing	65°C	30 seconds	
Elongation	72°C	45 seconds	
Denaturation	94°C	30 seconds	X16
Annealing	65°C (-0.7°C per cycle)	30 seconds	
Elongation	72°C	45 seconds	
Denaturation	94°C	30 seconds	X16
Annealing	55°C	30 seconds	
Elongation	72°C	45 seconds	
Elongation	72°C	5 minutes	X1

Hold	4°C		
------	-----	--	--

In regions of the genome with a high content of GC bonds or “GC rich regions” other reagents were used for the purpose of breaking those three hydrogen bonds: dimethyl sulfoxide (DSMO) and/or Betaine solution (Sigma, UK).

3.3.4.3 Agarose gel electrophoresis

PCR products are then visualized by electrophoresis on 2% agarose gels with added GelRed Nucleic Acid Gel Stain (Biotium, US). A porous gel in which DNA fragments can move is produced combining agarose powder (Sigma, UK) with TBE buffer and GelRed. When an electrical current is applied across the gel, the negatively charged DNA will travel at a rate relative to its size. Contamination screening was also performed by visualization of the same reaction containing dH₂O instead of DNA in the same gel for all reactions. A 1 kb DNA ladder (Qiagen) was used to judge the size of the amplified fragments and ensuring they matched the target region. The PCR products were loaded onto the gel with a loading buffer and the electrophoresis was run at 120V for 30 minutes. Visualization of the bands was done under an Ultraviolet (UV) transilluminator, and digital photographs were taken using the Syngene GeneGenius image acquisition system and GeneSnap software (Synoptics).

3.3.4.4 PCR purification

After confirming that the PCR had amplified the desired region in the agarose gel, PCR products were purified with a user prepared Exo-Sap enzyme mix and protocol. Briefly, a combination of alkaline phosphatase (Fast-AP, Thermofisher scientific) and an Exonuclease-I (Thermofisher scientific) with distilled water plus the PCR products were ran on the thermal cycler for 30 minutes at 37°C followed by 15 minutes at 80°C.

3.3.4.5 BigDye sequencing reaction

Once the PCR is purified, we prepared the sequencing reactions using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with the following protocol: ABI sequencing buffer, dH₂O water, primer (in this case either forward or reverse), ABI Big Dye terminator, and the purified PCR product. This recipe is shown on Table 3-3.

Table 3-3: Sequencing reaction recipe.

Reagents	volume (uL) per reaction
ABI sequencing Buffer	2
dH ₂ O	3.5
Primer (F or R) (5uM)	1
ABI BigDye	0.5
Purified PCR product	3
Total	10

The sequencing reaction is then loaded into the PCR cycler and the following program is run: 1 cycle of denaturation at 94°C for 1 minute, followed by 25 cycles of: denaturation at 94°C for 30 seconds, annealing at 50°C for 15 seconds, and elongation at 60°C for 4 minutes.

3.3.4.6 Sequencing reaction purification

Sequencing products were purified with user-prepared Sephadex® plates. For each plate of 96 reactions, 2.9 g Sephadex® G-100 (Sigma) was dissolved in 40 ml distilled water and allowed to hydrate for 30 minutes at room temperature. The solution was mixed well and 350 uL added to each well of a Corning® glass filter plate (Corning® Filtrex™ 0.66 mm glass fibre filter, Sigma). This plate was centrifuged at 710g for 3 minutes. The entire volume (10 uL) of each sequencing reaction was pipetted onto the centre of a Sephadex® column and the plate placed onto a final 96-well collection plate before being centrifuged a second time at 910g for 5 minutes. A heat seal was placed over the purified sequencing products and the plate was loaded onto the 3730XL DNA Sequencer (Applied Biosystems). The purified sequencing reactions were read by capillary electrophoresis using the protocol 3730BDTv3-KB-DeNovo_v5.2

3.3.4.7 Sequencing data analysis

Sequence data was analysed using Sequencer 5.1 DNA Sequence Assembly Software and compared to the reference genome retrieved from Ensembl (<http://www.ensembl.org>) or

NCBI (<https://www.ncbi.nlm.nih.gov>). The sequences were checked for single nucleotide changes and insertions and deletions in coding and flanking regions of the target gene. Sequence variation was checked to see if it was previously reported and at which frequency utilizing online databases including ensembl, the exome variant server (<http://evs.gs.washington.edu/EVS>), dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), the 1000genomes project (<http://browser.1000genomes.org/index.html>), and the Exome Aggregation Consortium (<http://exac.broadinstitute.org/>). Rare and suspiciously pathogenic variants were then always confirmed on an independent sequencing reaction with fresh DNA obtained from the original tube.

An example of the output of Sanger sequencing is presented in Figure 3-1.

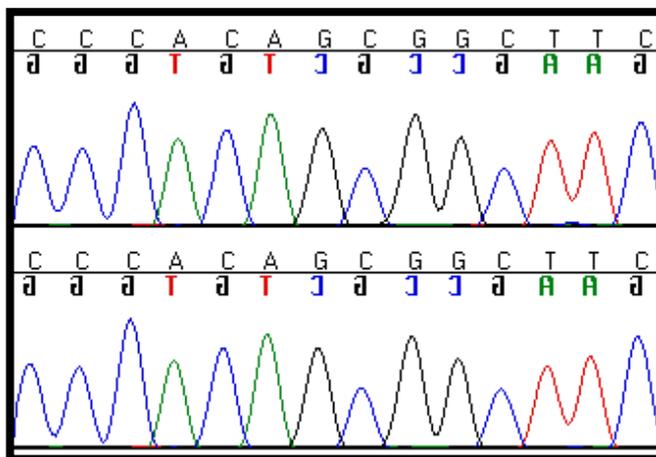


Figure 3-1: Sanger sequencing result of a section of the gene *SLC20A2* exhibiting reference sequence in 2 samples.

3.3.5 Fragment analysis

Fragment analysis was used to determine the presence of the expansion in the hexanucleotide repeat of GGGGCC in *C9orf72* with a qualitative assessment. Figure 3-2 presents the protocol and workflow.

We first performed a repeat-primed-PCR reaction. The repeat was amplified with a PCR reaction performed in the presence of 1M betaine, extensor long PCR master-mix (Thermo Scientific), using a previously optimized cycling program(251) and three primer sequences (forward, reverse and anchor) as previously published(252) (primer sequences also available on the appendix). The forward primer was fluorescently labelled. In cases where

we detected an expanded allele and cases where we couldn't detect both normal alleles we further characterized the expansion with a sizing PCR reaction. The sizing PCR was carried out in a mixture containing extensor long PCR master-mix (Thermo Scientific), 1M betaine solution, 5% dimethylsulfoxide, and 7-deaza-2-deoxy GTP, Q solution and the forward and reverse primers.

PCR products were then mixed with Liz 500 size standard (Applied Biosystems) and HiDi formamide (Applied Biosystems), denatured at 95 °C for 3 minutes, and then put immediately on ice for 5 minutes. This mixture was then analysed by fragment length analysis on an automated ABI3730 DNA-analyser and allele identification and scoring was accomplished using GeneMapper v3.7 software (Applied Biosystems). Figure 3-3Figure 3-3 presents a negative example of this technique.

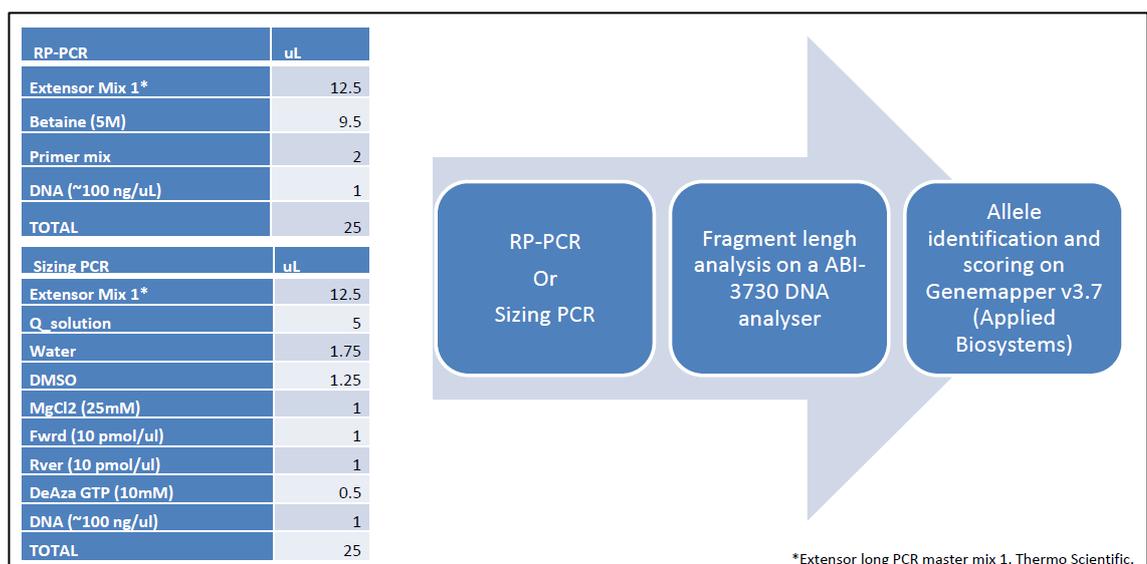


Figure 3-2: C9orf72 repeat expansion screening protocol and workflow.

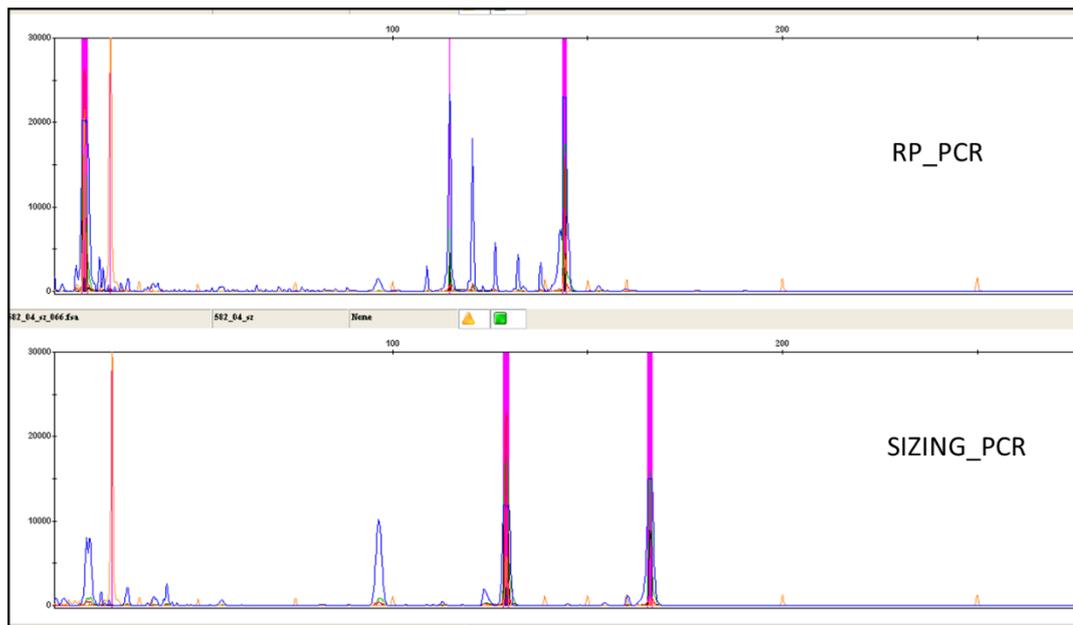


Figure 3-3: Example of a patient negative for the C9orf72 repeat expansion. Top graph: repeat primed PCR (RP-PCR). Bottom graph : sizing PCR; both showing the two different alleles both presenting less than 10 repeats.

3.3.6 Homozygosity mapping

Genome wide genotyping data was generated using the HumanCytoSNP-12v2-1_H (Illumina), which contains probes for over 200,000 markers distributed across the genome.

Ten μ l of DNA at a concentration of ~ 75 ng/ μ l were prepared by myself. After that samples were processed, hybridized and scanned in accordance with the manufacturer's instructions at UCL Genomics. Data analysis was performed by myself.

A project was created in genome studio software 2010.3 (Illumina) for clustering, normalization and obtaining genotype calls using default parameters. Input data were .idat files, and output data from genome studio was a final .txt report file containing data for all samples processed at 293,870 markers. These files were then modified to meet the requirements of the online Homozygosity Mapper software (http://www.homozygositymapper.org/sample_files.html) as shown in Table 3-4, and data were analysed with this online tool specifying affected and unaffected family members (10).

Table 3-4: Table showing output data from Genome studio software (Illumina) and how this needs to be adapted for homozygosity mapper.

Final report from genome studio					
SNP Name	Sample ID	Allele1 - AB	Allele2 - AB		
rs1000002	1	A	B		
rs1000003	1	A	A		
Data formatted for homozygosity mapper					
DBSNP	1	2	3	4	5
rs1000002	AB	BB	AB	BB	BB
rs1000003	AA	AA	AB	AA	AA

Regions of homozygosity are highlighted by this software according to their homozygosity scores. Homozygosity mapper calculates the length of the homozygous block (in SNPs) at each marker for each sample. The values of the cases are then added to get the homozygosity score for a marker. The maximum length for each block is set by the software to default limits according to the SNP array used. This is to reduce the effect of very long blocks in one or few samples. In our case, our sample presented with genetic homogeneity because it was one single family. Regions are excluded when the same homozygous

haplotype is found in any controls(253). An example of the output from the tutorial in homozygosity mapper is shown in Figure 3-4.

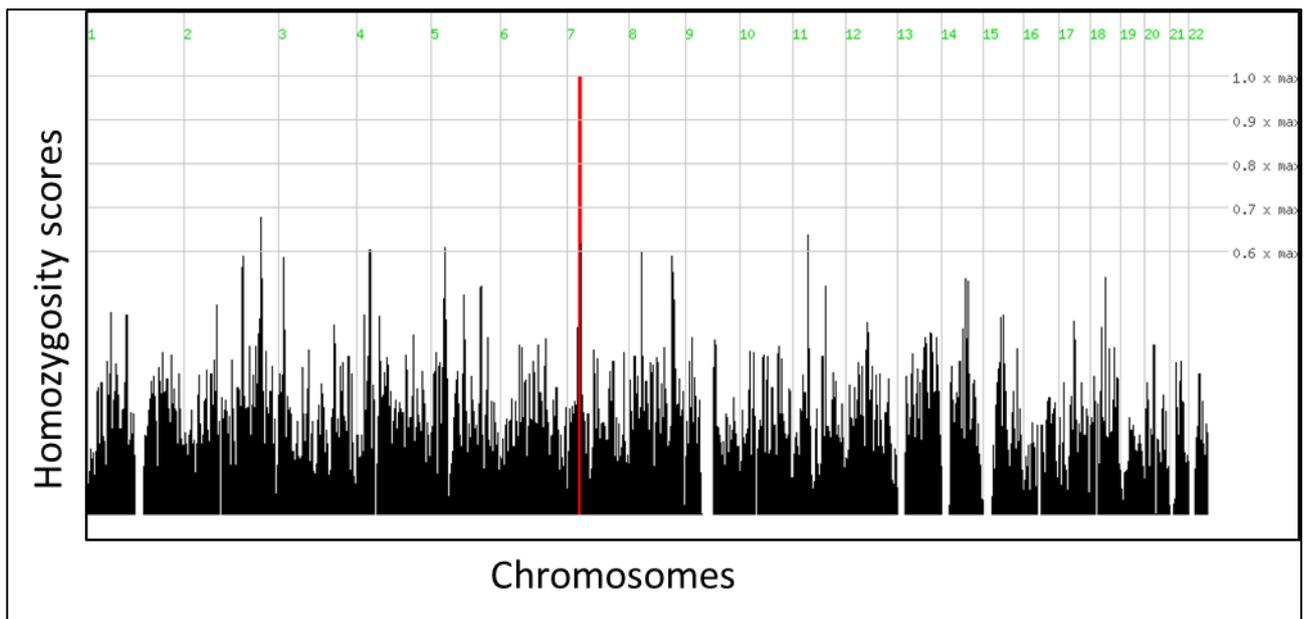


Figure 3-4: Homozygosity mapping example showing significant homozygosity in chromosome 7 (with permission from tutorial in <http://www.homozygositymapper.org/documentation.html>).

3.3.7 Next generation sequencing (NGS)

Next generation sequencing (NGS) technologies have allowed for massive parallelization of sequencing reactions allowing the sequencing of many millions of target molecules in parallel and for a substantial reduction in cost (compared to Sanger sequencing). In NGS, the DNA molecules are immobilized on a solid surface and are sequenced *in situ*. NGS platforms use the clonal amplification of template DNA to generate “clusters” of identical DNA molecules followed by sequencing through a stepwise incorporation of fluorescently labelled nucleotides or oligonucleotides. Through this thesis we used NGS and this is described next.

Through massive parallel-targeted sequencing NGS can effectively generate data of a target region, a human exome or even a genome in a matter of days. In the case of an exome, the coding portion of genomic DNA is selected from a pool of DNA fragments by hybridization with labelled probes that are complementary to the target.

The exome accounts for ~1-2% of the human genome and includes the coding sequence. It consists of >21,000 genes and ~180,000 exons that correspond to ~35-50 million base pairs. After data analysis we usually obtain >20,000 called variants per sample in Caucasian individuals. Since 2010(254), WES is proving to be a cost-effective technology in the successful discovery of causative genes in both recessive(255) and dominant(256) Mendelian diseases.

More recently, it has also been utilized to investigate the missing heritability in complex diseases and in families that are not large enough for traditional linkage studies. WES is also a novel tool useful for the study of rare variants associated with moderate risk of sporadic disease(257,258).

NGS can be performed in different platforms and with different chemistries and can target different parts of the human genome. Different capture technologies will select different targets. These can be all the exons of a genome (i.e. “the whole exome=WES”), the whole genome (WGS) or a targeted group of genes or loci of interest. Targeted NGS capture can be either custom-made or provided as a standard commercial kit. WES capture technologies target the exome and different kits will also include splice sites, some intronic regions, UTRs and non-coding RNAs. A simplified workflow of NGS is presented in Figure 3-5.

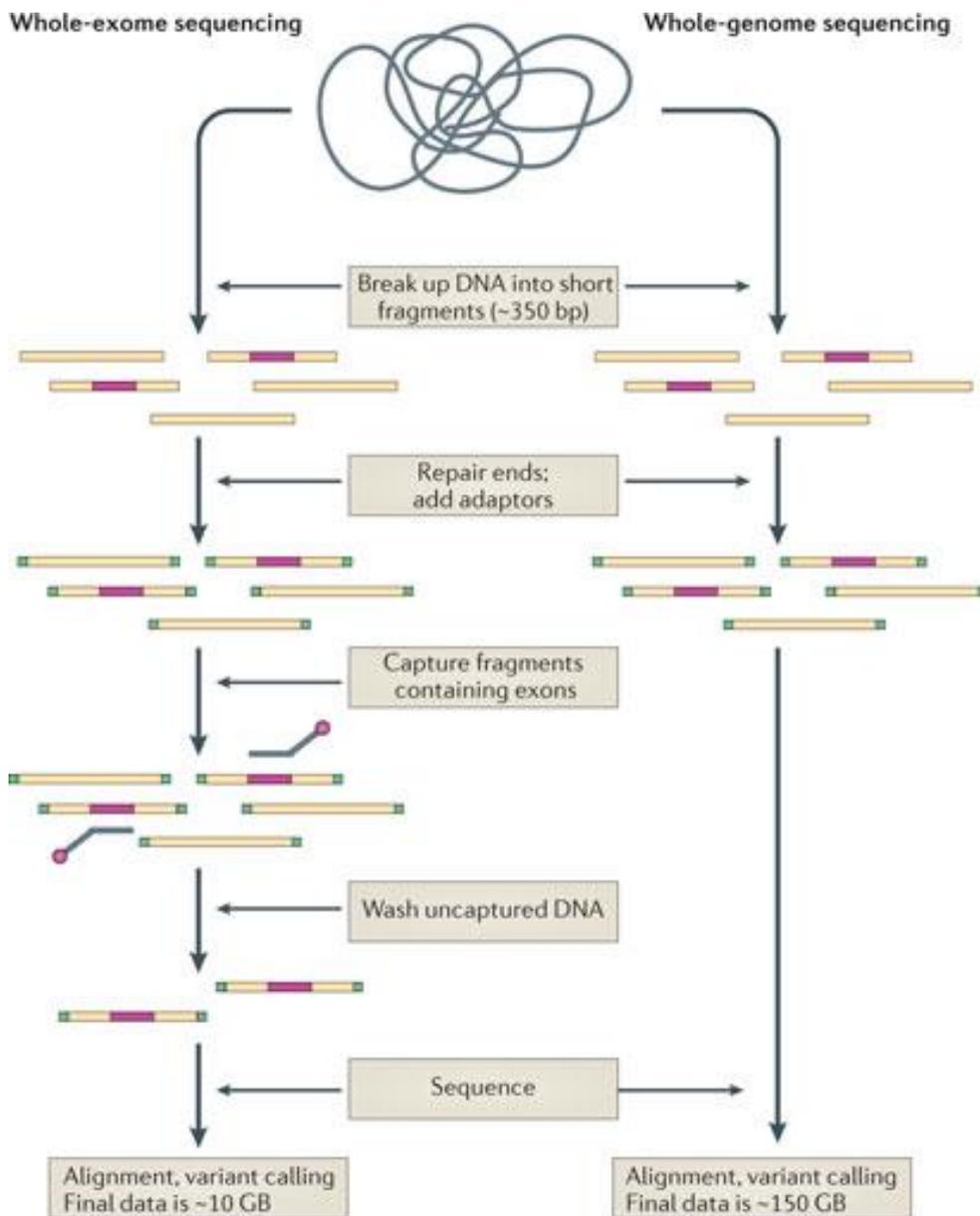


Figure 3-5: Simplified workflow of WES and WGS. Sample preparation is similar. Genomic DNA is fragmented and adaptors are added which allow each fragment to be hybridized to the flow cell for sequencing. WES then hybridizes the fragments to probes that are complementary to the targeted regions (i.e. the exome). These are captured and the remaining DNA is washed away. The following steps of sequencing are the same for WES and WGS. Modified from (9).

3.3.7.1 Whole exome sequencing (WES)

As explained earlier in this thesis, WES consists of the parallel sequencing of the exome or protein coding section of the genome. Through this thesis we have used different capture

methods, and I will describe them in this chapter. All kits used enable the automated preparation of dozens of reactions at the same time and are suitable for large scale projects.

3.3.7.1.1 TruSeq

The TruSeq technology (Illumina) was used for a large portion of our samples that were processed for WES, and all steps will be described below.

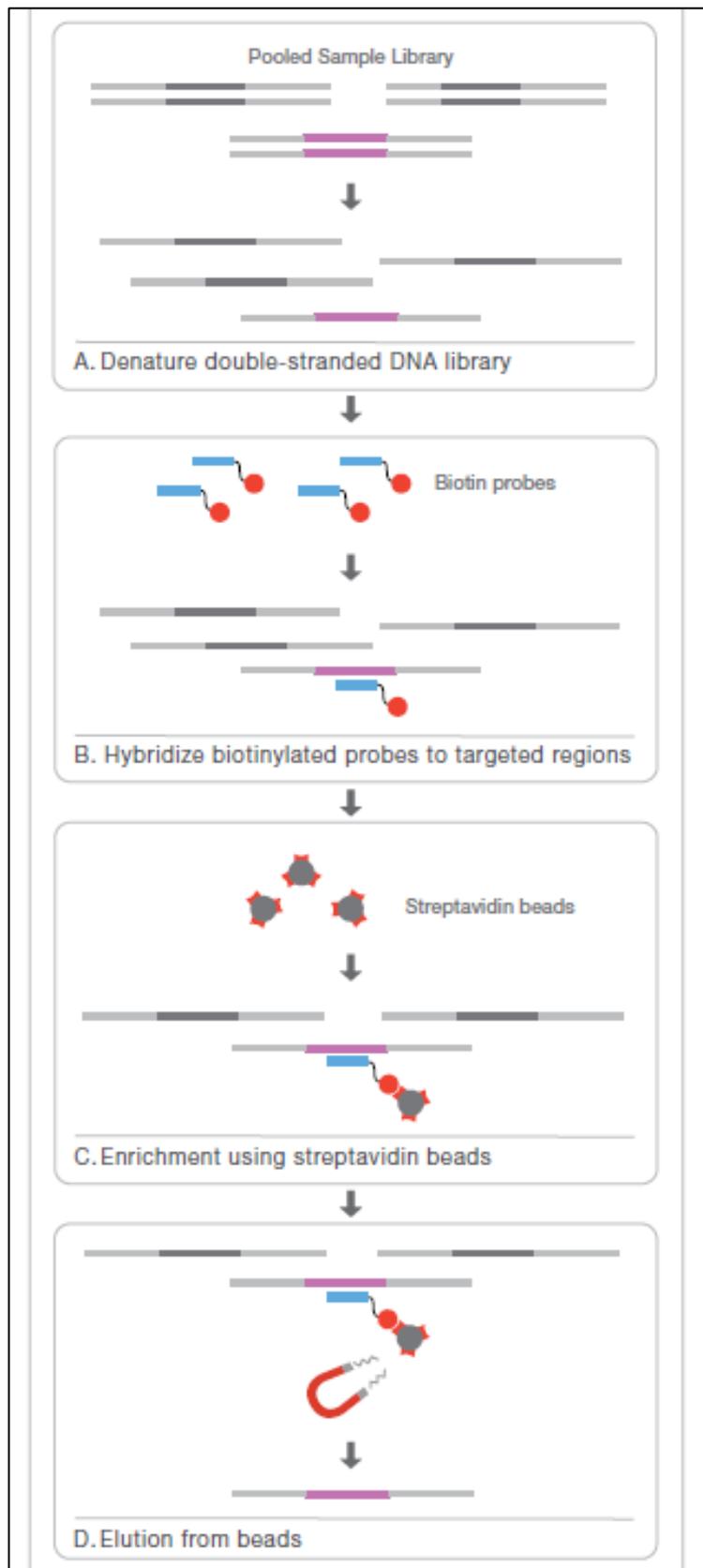


Figure 3-6: TruSeq Enrichment workflow composed by two successive rounds of enrichment (Courtesy of Illumina, Inc. reproduced from www.illumina.com).

3.3.7.1.1.1 Library preparation

Library preparation is the preparation of genomic DNA that is to be sequenced. The TruSeq enrichment kit contains master-mixed reagents, optimized index adapter design, a gel-free protocol, and a flexible workflow for preparing multiplexed samples that are pooled prior to sequencing. The workflow is presented in Figure 3-6.

3.3.7.1.1.1.1 DNA fragmentation and wash

After DNA normalization with the Qubit (as described above in section 3.3.3) each sample was sheared in covaris tubes/wells on a covaris sonicator. This will generate fragmented double stranded DNA. Then the samples are transferred to a midi plate and incubated with sample purification beads (SPB) on a magnetic stand and purified with ethanol.

3.3.7.1.1.1.2 Repair ends and select library size

This process converts the overhangs resulting from fragmentation into blunt ends using ERP3 (End Repair Mix). The 3' to 5' exonuclease activity of this mix removes the 3' overhangs and the 5' to 3' polymerase activity fills in the 5' overhangs. Following end repair, the library size is selected using SPB (Sample Purification Beads).

3.3.7.1.1.1.3 Adenylate 3' Ends

A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation. For this purpose, each sample is combined with A-tailing mix and a buffer.

3.3.7.1.1.1.4 Ligate adapters

This process ligates multiple indexing adapters to the ends of the DNA fragments, which prepares them for hybridization onto a flow cell. Each sample will then be mixed with unique adapter indexes, ligation mix, and buffer so that samples can be then pooled together. After incubation with the adapters, a stop ligation buffer stops the reaction. The

index adapter plate is presented in Figure 3-7. The number of samples that are pooled together before sequencing will be dependent on the desired coverage. The more samples the lower coverage. For most of the libraries prepared for this work the number of pooled libraries was usually 12.

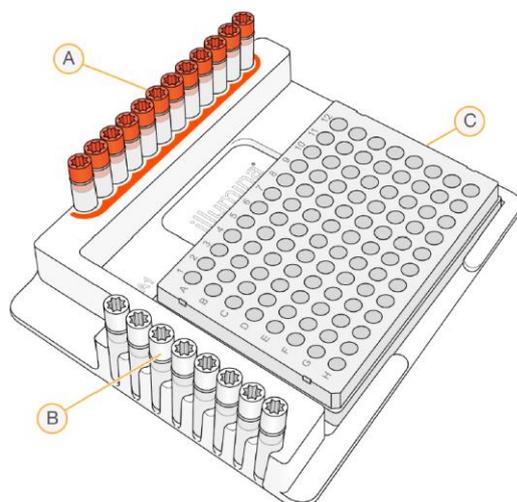


Figure 3-7: Index plates used for NGS indexing that allows sample pooling. A) Index A from 1 to 12. B) Index B from 1 to 12. C) Ninety-six well plate to combine each library with a unique combination of indexes A and B. In this plate one can index 96 samples at a time. The numbers of samples that will be pooled together before sequencing will depend on the desired coverage. (Courtesy of Illumina, Inc., reproduced from Illumina.com).

3.3.7.1.1.1.5 Enrichment

This step uses PCR to amplify DNA fragments that have been ligated with adapters on each end of each molecule. PCR is performed with PPC (PCR Primer Cocktail) that anneals to the ends of the adapters.

It is important to note that both adapters need to be ligated. This is because fragments with no adapters cannot hybridize to surface-bound primers in the flow cell and fragments with an adapter on 1 end can hybridize to surface bound primers, but cannot form clusters.

The PCR reaction is loaded onto the thermal cycler and PCR products are purified with SPB.

3.3.7.1.1.1.6 Validate Libraries

Library quantification can be performed with the Qubit. Library quality control is then performed on the Bioanalyzer 2100 (Agilent technologies). The bioanalyzer will assess the

size and concentration of the libraries. An example is shown on Figure 3-8. The bioanalyzer can also be used in the previous steps of the library preparation if desired.

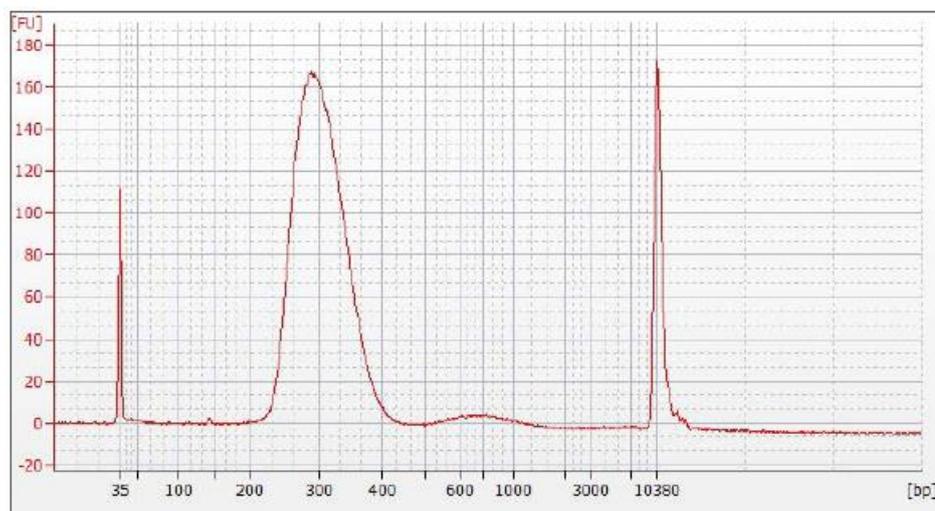


Figure 3-8: Example 300 bp Insert Library Distribution (pre-enrichment) on a High Sensitivity DNA chip of the bioanalyzer.

3.3.7.1.1.1.7 Pooling and first hybridization

This step combines DNA libraries containing unique indexes into a single pool, and then binds targeted regions of the DNA with capture probes (exons and flanking regions in WES). Pooling the libraries requires specific amounts of each library according to the number of samples to make a single pool of 40 uL of 12 libraries. This step might require that some samples are precipitated and concentrated to ensure equal amounts of DNA in each library. Then the library pool is combined with the CEX (Coding Exome Oligos) and a capture buffer, and loaded onto a thermal cycler.

3.3.7.1.1.1.8 Capture Hybridized Probes

This step uses SMB (Streptavidin Magnetic Beads) to capture probes hybridized to the targeted regions of interest. Two heated washes remove nonspecific binding from the beads. The exons enriched library is then eluted from the beads and prepared for a second round of hybridization.

3.3.7.1.1.1.9 Second hybridization and second capture

The second hybridization binds targeted regions of the enriched DNA with capture probes a second time and ensures high specificity of the captured regions. The second capture uses SMB (Streptavidin Magnetic Beads) to capture probes hybridized to the targeted regions of interest. Two heated washes remove nonspecific binding from the beads. The enriched library is then eluted from the beads and prepared for sequencing.

3.3.7.1.1.1.10 Clean Up Captured Library and amplification

This step uses SPB (Sample Purification Beads) to purify the captured library before PCR amplification. The amplification is then performed with an 8-cycle PCR program to amplify the enriched library adding an enrichment amplification mix and a PCR primer cocktail.

The PCR is followed by purification with SPB of the enriched library.

3.3.7.1.1.1.11 Validate enriched libraries

Sample quantification and quality control are performed again on the Qubit and the bioanalyzer. The distribution of DNA fragments expected here is within a size range from ~200 bp to ~400 bp.

3.3.7.1.1.1.12 Clustering

After amplification, the pooled library is ready for clustering. Clustering is performed in the cluster station that is an oligo-derivatized surface of a flow-cell. This step comprises single molecule amplification and starts with an adapter library. The flow-cell is an 8-channel sealed glass microfabricated device that allows bridge amplification of fragments on its surface, and uses DNA polymerase to produce multiple DNA copies of each fragment, also called clusters. Libraries may be run individually or in combination with others. Each cluster contains approximately one million copies of the original fragment which will be sufficient for detection during sequencing. This is shown on Figure 3-9.

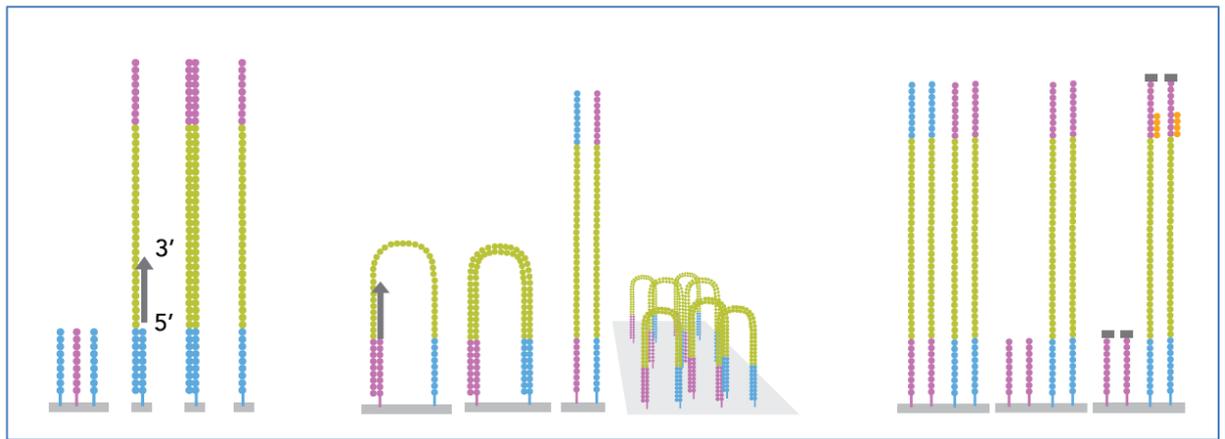


Figure 3-9: Automatic cluster generation. Courtesy of Illumina, Inc., reproduced from https://www.illumina.com/documents/products/datasheets/datasheet_cbot.pdf

3.3.7.1.1.2 Sequencing

Sequencing performed on the Illumina Hi Seq platform is a process of massive parallel sequencing by synthesis where all four nucleotides are added simultaneously to the flow-cell channels, along with DNA polymerase, for incorporation into the oligo-primed cluster fragments. The DNA is linearized by cleaving one adapter and denatured to obtain single strands. Sequencing primers and four reversible terminators are added. The nucleotides carry a fluorescent label, and the 3'-OH group is chemically blocked so that each addition is a unique event. An imaging step follows each base incorporation step, during which each flow cell lane is imaged in three 100-tile segments by the instrument optics at a cluster density per tile of 30,000. After each imaging step, the 3' blocking group is chemically removed to prepare each strand of DNA for the next incorporation of a nucleotide by the DNA polymerase. This series of steps continues for a specific number of cycles, as determined by user-defined instrument settings, which permits discrete read lengths of 25-100 bases (for example, 75 cycles will create seventy-five base pair sequencing reads. Most exons will be covered by 75-100 read lengths). The same sequencing process is performed from both DNA strand ends to create paired-end reads. This is particularly useful for accurate mapping and detection of structural variation. A figure explaining this process is exhibited on Figure 3-10. Data demultiplexing is the next step after sequencing in order to obtain files for alignment and analysis (Figure 3-11).

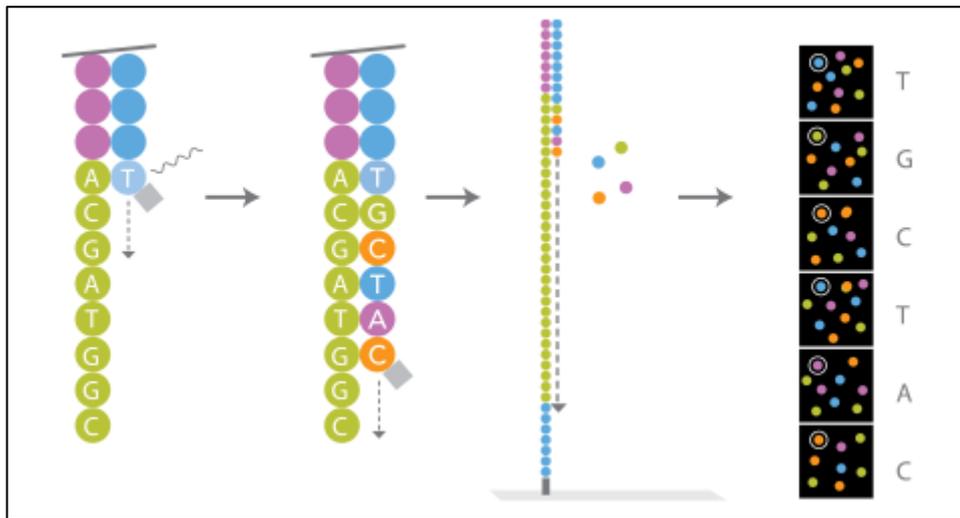


Figure 3-10: Sequencing by synthesis. Courtesy of Illumina, Inc., reproduced from Illumina.com.

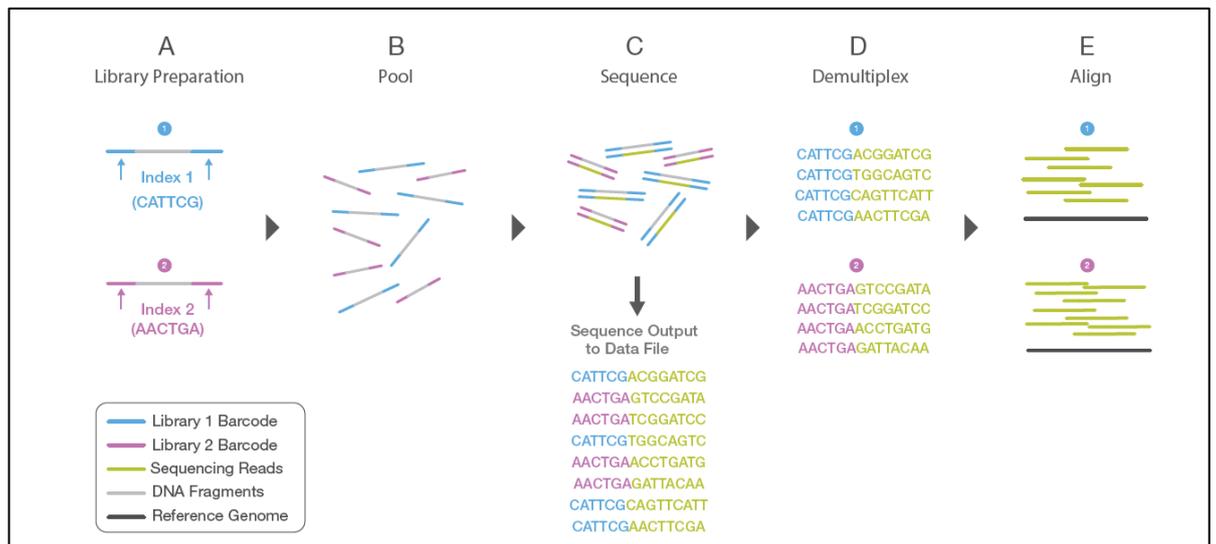


Figure 3-11: Sequencing workflow and demultiplexing. Courtesy of Illumina, Inc., reproduced from Illumina.com.

3.3.7.1.2 Nextera rapid capture

The Illumina Nextera rapid capture kit was introduced after the TruSeq kit, and was used for WES library preparation of more recent samples. In comparison with TruSeq, it was significantly improved in terms of the amount of DNA required and time required to prepare the libraries. They were both massively reduced. The workflow is shown in Figure 3-12 and I will describe these methods shortly because there are similarities to the Tru Seq protocol that do not need to be repeated.

Nextera Rapid Capture Enrichment library preparation uses an enzymatic DNA fragmentation step and thus can be more sensitive to DNA input compared to mechanical fragmentation methods (as in TruSeq). The ultimate success of enrichment strongly depends on using an accurately quantified amount of input DNA. Therefore, accurate quantification of the gDNA is essential and usually requires at least 2 or 3 measurements and dilutions with the Qubit.

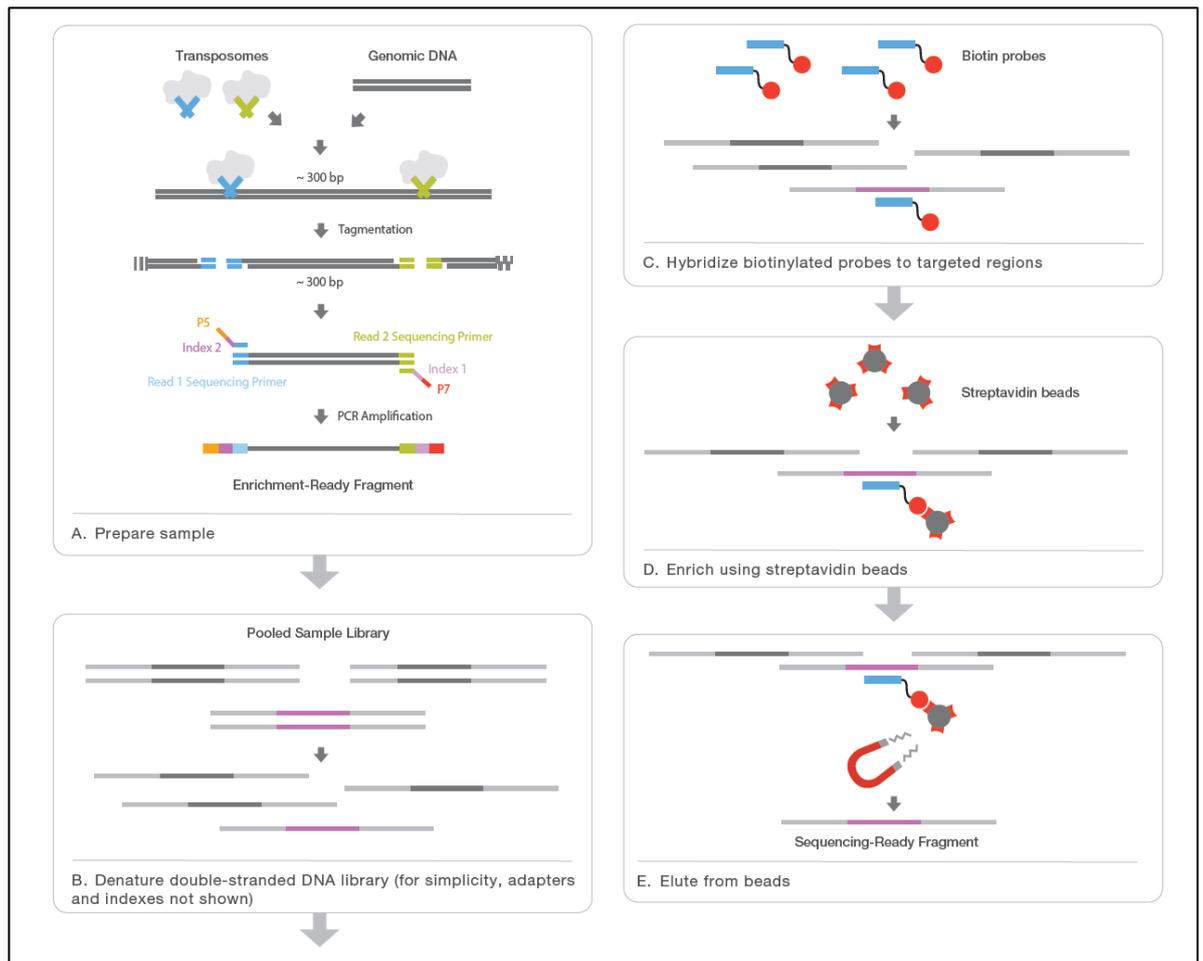


Figure 3-12: Nextera rapid capture workflow. Courtesy of Illumina, Inc., reproduced from https://www.illumina.com/documents/products/datasheets/datasheet_nextera_rapid_capture_exome.pdf

After normalisation of the DNA, the next step is the tagmentation of the gDNA. This step uses the Nextera transposome to tagment gDNA, which is a process that fragments DNA and then tags the DNA with adapter sequences in a single step. This is followed by clean up with sample purification beads on a magnetic stand.

The tagmentation is followed by a first PCR amplification with index primers and a nextera library amplification mix added to each sample. This is loaded onto a thermal cycler and later purified with SPB.

Sample quality is checked on the Bioanalyzer and fragments should be 150-1000 bp in size at this stage.

Library amplification is followed by the first hybridization where the DNA library binds to the biotinylated oligos (baits). Before hybridization samples are pooled into libraries of 12 samples at a volume of 40 uL. Sample amount needs to be precise and even (i.e. 500 ng per sample).

Post hybridization capture of the gDNA-bait hybrids is performed with SPB in a midi plate on a magnetic stand to remove nonspecific bindings. The library is then eluted from the beads.

This process is followed by a second cycle of hybridization, capture and purification.

The second PCR amplification (in this case of the enriched library) is performed afterwards in the presence of the amplification mix and a primer cocktail. The reaction is loaded onto the thermal cycler and then purified with SPB.

Libraries are now ready for clustering and sequencing as described in the previous section.

3.3.7.2 Bioinformatic processing of NGS data

The analyses begin with data demultiplexing with CASAVA tool provided by Illumina. Paired-end sequencing reads in the form of FASTQ files are then aligned to the human genome reference (hg19) using Novoalign (Novocraft technologies). A BAM file is generated using SAMtools (<http://samtools.sourceforge.net/>). The removal of duplicate reads and the generation of statistics are performed with picard tools (<http://picard.sourceforge.net/>). Local realignment if indels and variant calling was performed with the Genome Analysis Toolkit (GATK) <https://software.broadinstitute.org/gatk/>.

Variant annotation was performed with ANNOVAR (<http://www.openbioinformatics.org/annovar/>). ANNOVAR is fully customizable and allows

for annotation of the location of each variant (exonic, splice site, intronic, etc.), determination of its functional effect (non-synonymous change, stopgain, stoploss, frameshift, etc), adding population frequency from different databases (1000 genomes, Exac, EVS, dbSNP, etc), in silico predictions (Polyphen, mutation taster, CADD scores, etc), adding OMIM numbers, among other possibilities.

Other useful online software packages used through this thesis are the GATK (<http://www.broadinstitute.org/gatk/>), VCFtools (<http://vcftools.sourceforge.net/>), PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>), RV test (<http://zhanxw.github.io/rvtests/>) and LASER ancestry server (<https://laser.sph.umich.edu/index.html>).

The types of files we obtain from NGS data are:

1. Fastq: Human readable sequences with associated Phred scores. This is the native sequence or “raw data”. Fastq are usually paired-end. This means we get 2 files per sample (a forward and a reverse). The Phred score gives an estimate of the correctness of the corresponding base call.
2. SAM: Human readable mapped sequences, phred scores and coordinates to the reference sequence.
3. BAM: Binary version of SAM (compressed version of SAM). This file can be visualized with the Integrative Genomics Viewer (<http://software.broadinstitute.org/software/igv/>) or the genome browser (<http://goldenhelix.com/products/GenomeBrowse/>).
4. VCF: Variant calling format. The VCF format is a tab delimited format for storing variant calls and individual genotypes. It can store both SNPs and indels.
5. Text files: After annotation with ANNOVAR the output can be retrieved in an CSV file that can be filtered with a text editor or also loaded onto Excel. Depending of the aim of each chapter and the number of samples to be analysed this was done in different ways along this thesis.

The specific bioinformatics analysis pipeline for each project will be commented upon in each results chapter. Moreover, relevant scripts are attached in the Appendix.

3.3.8 Variant prioritization, mutation confirmation and analyses of complex disease

NGS technologies generates huge amounts of data. It is essential to understand how to analyse these data. Genetic sequencing is a field that is moving faster than functional studies and therefore, knowing with certainty the difference between a polymorphism and a mutation is many times very challenging.

For example, a whole exome today produces over 60,000 variants. Being able to find the biologically relevant ones for a specific patient is like finding the needle in the haystack. All variants suspected to be disease causing have been visualised with the IGV viewer or the genome browser and then Sanger sequenced as this technique is still the gold standard for mutation confirmation.

The approach to variant or gene prioritization is based in some principles:

3.3.8.1 Quality of the variant

Quality scores (based on Phred scores), segmental duplication scores (refers to homology greater or equal than 96% and means that these regions are likely to contain many false positive calls), and depth (number of reads at a specific location) information are provided by the analysis pipeline. They will help in understanding if a variant is true or is a false positive. NGS data is usually more reliable for single nucleotide changes than for indels. However, it is becoming more and more reliable with small indels.

Furthermore, we have an in-house collection of human exomes in the department and we also filter variants against this database. If a variant appears novel or rare in online databases but is common in our data from other disorders or controls it is more likely to be an artefact of our pipeline rather than a disease-related change.

Depending on the strategy of the specific project the filtering according to quality scores was different. For example, looking into families and small genes, I usually keep all the data in there as all variants selected will then be visually inspected and if considered potentially pathogenic Sanger sequenced. A different scenario is when performing association studies. The data needs to be curated for well called variants present in cases as well as controls.

3.3.8.2 Known genes and pathways

Usually, when analysing NGS data, the first step consists in investigating genes that are known to cause the phenotype of interest or similar disorders to those that patients have. Additionally, genes that have a role in the pathway of known genes for the disease are also prioritized. I usually do this before filtering for other principles to ensure I am not missing any known mutations. For example, this is what I did in a family with PFBC in Chapter 6.

3.3.8.3 Coding versus non-coding

The human exome accounts for only around 1-2% of the human genome. It is estimated that 85% of disease-causing mutations lie within the exome. This is the rationale behind performing WES experiments as a cost-effective technology. With 1% of the data we can get 85% of the relevant variation.

We therefore used exome capture methods in NGS and this gave us output of data containing exons, splice-sites, UTRs, non-coding RNA sequences and intergenic regions. From these data, while filtering for novel causes of disease, I focused the analysis on the exons and the splice-sites. The number of variants usually obtained after this step is approximately 21,000-22,000 or above.

3.3.8.4 Frequency in the population

The population frequency of variants is a good way of prioritising variants. The frequency accepted for recessive and dominant disorders is different. For instance, the acceptable frequency of recessive variants will be higher than for dominant disorders. This is because a variant present in the general population at a higher frequency than a rare disorder that is supposed to be inherited in a dominant fashion is unlikely to be playing an important role because if that was the case, the disease should be more frequent. In recessive disorders, one can accept a slightly higher frequency because there will be unaffected carriers in the population.

Filtering by population frequency is performed by obtaining data from large databases. In the beginning of this theses this was done mostly with data from the 1000 genomes project, the exome variant server and the cg69 database. Nowadays, we are also using the Exac

database. The Exac database contains data on 60,706 unrelated individuals. This changed a little our approach to variant frequency, because a few years ago, variants for rare disorders were likely to be novel, whereas today, a variant can be found in this database, but at a very low frequency. An interpretation of the variant frequency in the population in relation to the frequency of the disease and the suspected mode of inheritance is required.

dbSNP is also a useful tool linked to ClinVar that is helpful for variants with clinical information submitted by genetic laboratories and also with data from published papers.

3.3.8.5 Mode of Inheritance

In families where a recessive inheritance is suspected we usually look for homozygous or compound heterozygous variants, whereas in dominant disorders we would target heterozygous variants. In X linked we would be looking for variants in the X chromosome in males with absence of male-to-male transmission in the pedigree. It is important not to confuse dominant families with incomplete penetrance with recessive disorders. And in my experience, to ensure the analysis is thorough, I use different strategies for each family.

3.3.8.6 In silico predictions and conservation scores

In silico predictions are obtained with prediction software. These software use different strategies to predict the pathogenicity of a variant. They use physical and comparative considerations. They base their analysis in information such as if the variant is causing a truncating protein, if the amino acid change is to an amino acid with a different structure (such as proline) or hydrophilic for hydrophobic or vice versa, if the change is located in a relevant domain of the protein, if it changes the phosphorylation status of a protein, and/or the conservation among species. Polyphen (<http://genetics.bwh.harvard.edu/pph2/>) and SIFT (<http://sift.icvi.org/>) can query non-synonymous changes and Mutation taster (<http://www.mutationtaster.org/>) also works for small insertions and deletions. Each one has a different strategy and scoring available on their websites. These software, as I said, have different uses and applications and are all three of them focused on coding variation. Two new algorithms have been developed that integrate previous software with known variation in coding and non-coding regions. These are the Combined annotation-dependent depletion (CADD, <http://cadd.gs.washington.edu/info>) and the deleterious annotation of

genetic variants using neural networks (DANN, <https://omictools.com/deleterious-annotation-of-genetic-variants-using-neural-networks-tool>) scoring.

Mutation prediction software have an estimated accuracy of up to 80% (but usually lower) for missense variants(259) and can sometimes over or underestimate the biological relevance of specific changes. Such is the case of the p.Ala53Thr mutation in the *SNCA* gene that is predicted to be benign but has a known pathogenic role in human disease(260,261). The main reason for this wrong “benign” prediction is due to software reliability on conservation and the fact that other amino acid substitutions can be compensating for these changes in orthologous species(260). Also, in some cases, a gain of function at a poorly conserved residue can cause disease. Given that is not the loss of the wild type that is causing the disease, if this residue is not conserved, the change will be categorised as tolerated when it might not be.

In silico predictions are not always correct. Moreover, they sometimes present conflicting results and some of them will categorise a variant as pathogenic while others will call it benign or tolerated. In general, when all predictions are similar, this can be somewhat reassuring, however, predictions should always be considered with caution and used for what they are: predictions, never confused with functional data.

3.3.8.7 Target regions

Regions of the genome that present significant linkage (as calculated by their LOD scores) in families can be a good target for variant investigation and filtering through WES. Linkage data analysed according to the pattern of inheritance can provide information on shared regions in affected family members compared with the unaffected relatives.

Another method for defining target regions in recessive families is homozygosity mapping, and for example, in the Chapter 6.2 we have used a genome-wide SNP array and obtained homozygosity mapping data in a family with PFBC. Once the homozygous regions were delimited, we filtered the exome data to find the variants in those regions.

3.3.8.8 Complex disease

Complex disease analysis of NGS data is more challenging than studying families.

The principles described previously in this chapter are used but under a different approach. The investigation is performed by analysing a large number of cases sharing a disease phenotype and comparing them with a large number of controls.

Two approaches are currently used for this purpose: by single variant association of common variants and by looking for association of rare variants or burden tests. This strategy was used in Chapter 5.2.3 to understand the genetic risk underlying MSA, and the methods are discussed in the next section.

3.3.9 Bioinformatic steps for analysis of NGS for association studies

Initial steps of data demultiplexing, alignment of sequence reads to generate BAM files from FASTQ files, removal of duplicates and realignment of indels were performed as described in the section 3.3.7.2. Scripts are presented in the Appendix.

3.3.9.1 Power calculation

The power calculation was performed using an online software called Genetic association study Power Calculator available at: http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html

3.3.9.2 Generating a Joint VCF file

In the initial steps of data analysis described above, each BAM file was re-analysed with Haplotype Caller for calling SNPs and indels by local re-assembly of haplotypes. This step would improve the quality of the variant calling. The output is a genomic VCF (gVCF).

The gVCF files are then merged with the CombineGVCFs script (based on GATK) to obtain a joint VCF file containing all the data together.

3.3.9.3 Genotype quality control (QC)

After that, the VCF file is filtered according to quality with the Filter_VCF.sh script. The standards used in our lab filter out genotypes with a depth<8, genotype quality (GQ)<20, GQ_Mean<35, variant quality score recalibration score (VQSR) >99. For association analysis, the quality of the genotypes needs to be very strict to avoid spurious associations and samples processed with different chemistries are analysed in batches(262).

- Depth is defined as the number of reads in a specific position.
- Genotype quality (GQ) is a quality score of each genotype that is generated by GATK and consists of a Phred scaled value representing that a called genotype is true. The GQ represents the Phred-scaled confidence that the genotype assignment is correct and it is derived from the genotype Phred scaled likelihood.
- GQ_Mean is the average GQ at a specific location in a batch of samples.
- Although the name is a little misleading, the variant quality score recalibration (VQSR) is not a recalibration score. Instead it stands for a new well calibrated quality score. The VQSR method, uses machine learning algorithms to learn from each dataset what is the annotation profile of good variants vs. bad variants, and does so in a way that integrates information from multiple dimensions. Each variant will then get a VQSLOD which is the log odds ratio of being a true variant versus being false under the trained Gaussian mixture model of the program. The threshold of 99.0 used here is said to represent a 99% sensitivity for a true variant.

The output of this analysis is a filtered concatenated VCF file of all cases and controls which only contains ID and genotypes.

3.3.9.4 Variant and sample quality control

In order to obtain a reliable dataset and avoid systematic bias, we need to perform further variant and sample quality QC steps(263).

This is performed in different steps to curate the data. The output aims to obtain reliable NGS data for association analysis.

3.3.9.4.1 Per-marker QC or Variant QC

Variant QC consists of two main steps:

- 1) Identification of SNPs with an excessive missing genotype. Variants with a high missing genotype call rate in the whole dataset. The missingness per variant is calculated by PLINK across the dataset provided. Different thresholds can be applied by the researcher and we excluded calls with a call rate of at least 90%.

- 2) Identification of SNPs demonstrating a significant deviation from Hardy-Weinberg equilibrium (HWE) in controls. HWE states that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences. A significant deviation from HWE can indicate a genotyping error or an association with disease. Therefore, HWE is calculated by PLINK only on controls to remove variants where HWE p values are <0.001 .

3.3.9.4.2 Sample QC

Sample QC consists of the following steps:

- Identification of individuals with outlying missing genotypes. Samples that still had a high proportion of missing genotypes after removing bad SNPs during variant QC are also removed. A sample with a high missingness may denote bad quality in the DNA or the library.
- Identification of individuals with discordant sex information: Sex check. In this step, we provide with the sample gender we have recorded and PLINK will compare this to the actual genotype sex and give a score. If there is a mismatch between the ascertained sex and the genotype sex we have to exclude those samples. Checking sex in PLINK is performed by analysing the proportion of homozygosity of the X chromosome. The output gives us F values. An F estimate smaller than 0.2 yield female calls, and values larger than 0.8 yield male calls. We were a little less stringent because PLINK is designed for array data as opposed to sequencing data and we also tried the parameters 0.3 and 0.7 as limits.
- Identification of individuals with outlying heterozygosity rates. The heterozygosity is the proportion of heterozygous genotypes for a given individual. This aims to identify individuals with an excessive or reduced proportion of heterozygote genotypes, which may be indicative of DNA sample contamination or inbreeding, respectively. This rate is calculated in PLINK for the dataset. Standard limits proposed can be -0.2 to 0.2 but this needs to be guided by the actual visualization of the data.

- Identity by state (IBS): To identify duplicate and related individuals, IBS is calculated for each pair of individuals based on the average proportion of alleles shared in common at genotyped SNPs (excluding the sex chromosomes). The method works best when only independent SNPs are included in the analysis. To achieve this, regions of extended linkage disequilibrium (LD) (such as the HLA) are entirely removed from the dataset and remaining regions are typically pruned so that no pair of SNPs within a given window is correlated. Duplicates will have a IBS value=1.
- Identification of related individuals. A basic feature of standard population-based case-control association studies is that all the samples are unrelated (i.e. the maximum relatedness between any pair of individuals is less than a second-degree relative). To calculate the identity by descent (IBD) the degree of shared ancestry between individuals is estimated in PLINK. For the purpose of this analysis pruning of SNPs in LD needs to be performed beforehand (as described in the IBS section above). The IBD results can be interpreted as: IBD =1 for monozygotic twins or sample duplicates, IBD=0.5 for first degree relatives, IBD=0.25 for second-degree relatives and IBD=0.125 for third-degree relatives. Again, visualization of the data is very important.
- Identification of individuals of divergent ancestry. Population stratification can be a confounding factor, in which genotypic differences between cases and controls are generated because of different population origins rather than any effect on disease risk. The most common method for identifying (and subsequently removing) individuals with large-scale differences in ancestry is principal components analysis (PCA). In different Results chapter we investigated samples ancestry with PCA with the software [eigenstrat](#)(264). These data are visually inspected and compared with the results of the other QC steps. Most of the failing samples fail in more than 1 step and in those cases the sequencing quality is not optimal.

3.3.9.5 *Single locus association study*

After exclusion of all samples in the QC steps, duplicates, related samples, bad SNPs and bad sample removal, and having a European cohort we are ready to start testing for association.

For single locus association analysis, only common variation was selected for investigation (i.e. MAF >0.05).

First, we generated a log quantile-quantile (QQ) plot that compares the observed distribution of p -values for all the SNPs on the y axis with their expected values on the x axis under the null hypothesis of no association. By illustrating that the majority of the results follow the null distribution and that only a handful deviate from the null, the QQ plot gives confidence in the quality of the data and the robustness of the analysis. A deviation on the bottom of the $y=x$ line can suggest the presence of excessive false positive associations due to a systematic bias. For example, due to population stratification. A sharp deviation from the $y=x$ line on the smallest p -values (top right corner) shows potentially true associations.

Second, we performed an association analysis for single variants. This was done following PLINK commands. Scripts are available in the Appendix. This test was a standard chi square test including common variants (MAF >0.05). Subsequently, we adjusted for multiple testing with standard Bonferroni correction. The adjusted p -value, after correction for multiple testing, below 0.05 was considered significant.

A Manhattan plot is a scatterplot where genomic coordinates are displayed along the x-axis, with the negative logarithm of the association p -value for each SNP displayed on the y-axis, meaning that each dot on the Manhattan plot indicates a SNP. The strongest associations have the smallest p -values (e.g., 10^{-15}) therefore their negative logarithms will be the greatest (e.g., 15). The name Manhattan plot derives from the similarity of this plot to the Manhattan skyline, with chromosomes (in the x-axis) being like skyscrapers.

Both the QQ plot and the Manhattan plot were plotted using R commands. Scripts are presented in the appendix.

3.3.9.6 Rare variant testing

The largest MSA GWAS lacked power to achieve significant results(265). As MSA is a rare disorder, it will be very difficult to increase substantially the number of samples (especially pathologically confirmed ones) for this approach. With the advent of NGS, some different strategies have emerged and rare variants have been identified as susceptibility factors in complex diseases (257).

A new approach that has been developed with the advent of NGS is variant aggregation according to different strategies. For example, variants can be aggregated by locus, by gene, by function or *in silico* predictions(266).

Rvtests, which stands for Rare Variant tests, is a flexible software package for genetic association studies. It includes a variety of association tests (e.g. single variant score test, burden test, variable threshold test, sequence kernel association test (SKAT) test, fast linear mixed model score test). It takes VCF format as genotype input file, also takes PLINK format phenotype file, and a covariate file(267). This package has been used to perform rare variant testing in chapter 5. Scripts are available in the Appendix.

A burden test is a powerful test to test for association of rare variants in a genomic or exomic region that are causal of the phenotype, and the effects of all variants are in the same direction (i.e. all lead to increased risk of the disease or all are protective). A burden test collapses rare variants in a genetic region into a single burden variable and then regresses the phenotype on the burden variable to test for the cumulative effects of rare variants in the region.

A SKAT test is a variance-component test, which is more powerful when a large fraction of the variants in a region are non-causal or the effects of the associated variants are in different directions. A SKAT aggregates the associations between variants and the phenotype through a kernel matrix and can allow for SNP-SNP interactions. It is especially powerful when a genetic region has both protective and deleterious variants or even many non-causal variants. More recently, a SKAT test for the combined effect of rare and common variants has been proposed. The proposed test is also computationally efficient and would

evaluate the cumulative effect of rare and common variants over disease susceptibility. This would potentially increase the power of a given sequencing study substantially. This test has been proposed as a powerful method for complex diseases such as MSA. An optimal unified SKAT test (SKAT-O) is a proposed computationally efficient method that maintains power in both scenarios. SKAT-O automatically behaves like the burden test when the burden test is more powerful and behaves like the SKAT when the SKAT is more powerful. This test can be applied to sequencing association studies(268).

3.4 MOLECULAR BIOLOGY AND FUNCTIONAL STUDIES

3.4.1 Southern blot

Southern blot is a method used to detect specific DNA sequences in DNA samples by hybridization that was first described by Edwin Southern(269) and hence the origin of its name. A sample population of purified DNA is digested with one or more restriction endonucleases generating fragments that are several hundred to thousands of BP in length. The restriction fragments are separated according to size by agarose gel electrophoresis, denatured and transferred to a filter nitrocellulose or nylon membrane. Labelled probes are hybridized to the membrane-bound target DNA, and the positions of the labelled heteroduplexes are revealed by autoradiography(6).

This method was used in this thesis for confirmation of the *C9orf72* expansion and sizing. Further details of this experiment are described in chapter 4.4.2.

3.4.2 Coenzyme Q 10 measurement

The levels of coenzyme Q10 were measured in flash frozen brain tissue samples.

Flash-frozen brain tissue sections were prepared including similar amounts of grey and white matter. The tissue was homogenized with a buffer that contained sucrose, EDTA and Tris (tris[hydroxymethyl]aminomethane). The PH was 7.4 and the buffer was isotonic with the brain cells. Forty uL of the homogenates were mixed with an internal standard (IS). The IS used was di-propoxy-CoQ10 which is a synthetic non-physiological ubiquinone. The IS was added to the brain homogenates which were then vortex mixed, frozen in liquid nitrogen,

and thawed twice to ensure maximal release of cellular CoQ10. A solvent of hexane:ethanol (70:30 % (v/v)) was added to the sample, vortex mixed for 1 minute, and then centrifuged (5 min x 14,000 g, 25°C). Following centrifugation, the upper organic layer containing lipophilic compounds including CoQ10 was removed and evaporated to dryness. The sample was reconstituted in ethanol and measured by high-performance liquid chromatography (HPLC) which was linked to a UV detector set at 275 nm (nanometers). Ubiquinones, including CoQ10 have a characteristic absorbance at 275 nm and the HPLC machine is calibrated with a quantified CoQ10 and IS prior to analysing the biological samples. This method was described previously in (187). A picture of a CoQ10 measurement with the IS is shown in Figure 3-13.

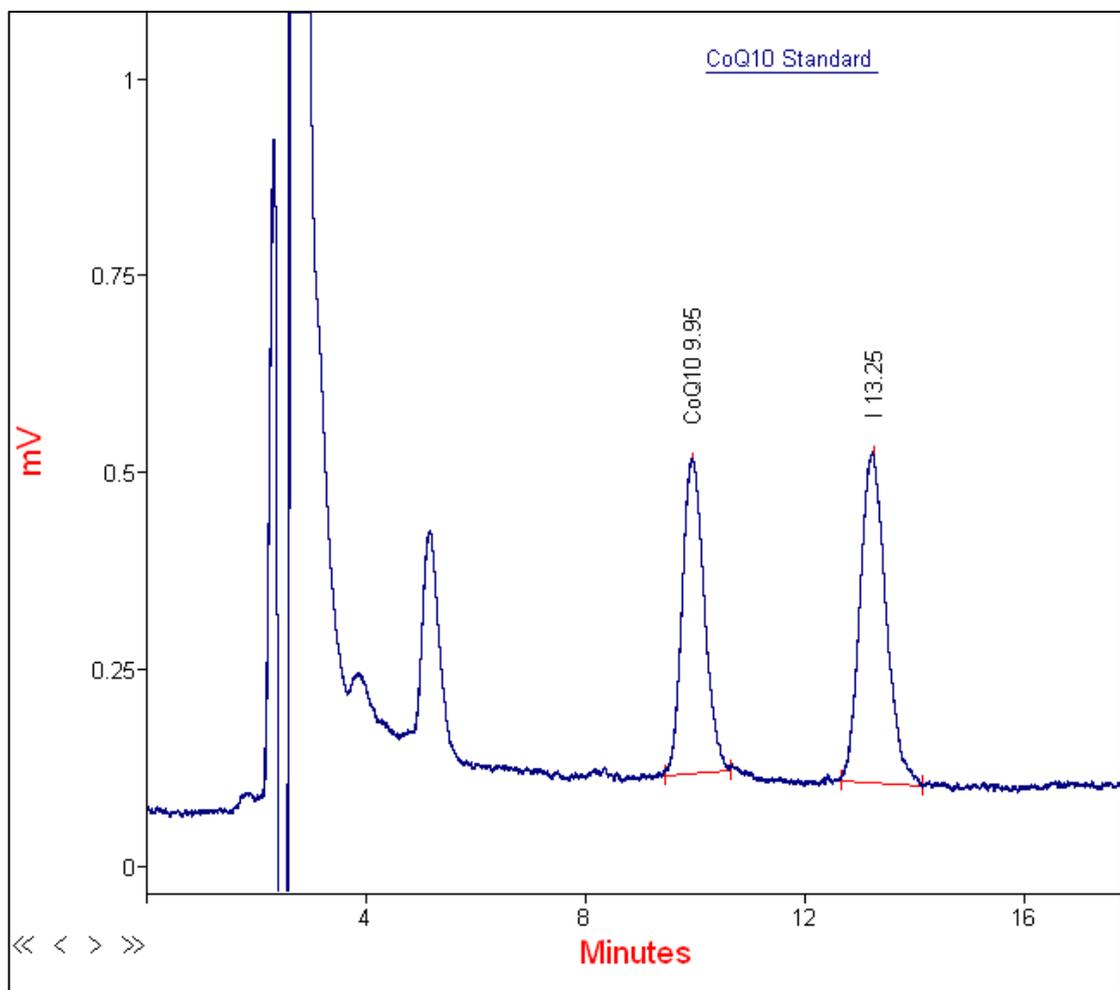


Figure 3-13: Coenzyme Q10 measurement by HPLC (courtesy image provided by Dr Iain Hargreaves).

3.4.3 Functional characterisation of a candidate gene identified in families with PFBC

In this section we investigated the role of a mutation in the *JAM2* gene identified in 2 independent families with PFBC. We first performed a western blot in fibroblast cell lines and we later studied the neurological phenotype of a knock-out mouse model. The relevant collaborators that contributed to this work have been detailed in each result chapter.

3.4.4 Fibroblast cell lines

Fibroblasts were generated from skin biopsies. Three mm punch skin biopsies were obtained from skin overlying the triceps. They were cut into small (1mm) pieces, and cultured in Modified Dulbecco's Eagle Medium (DMEM, Sigma) containing 10 % foetal bovine serum (FBS) and antibiotics in 10 cm² plates at 37°C. After 2-3 weeks in culture, fibroblasts grew from the skin biopsy pieces onto the tissue culture dish. When the dishes were confluent, they were split using 2 mL of 0.05% 1x Trypsin-EDTA (Life Technologies) per plate and maintained in DMEM/FBS without antibiotics at 37°C in an incubator with 5 % of CO₂. The plate was then placed in the incubator at 37 °C for 5 minutes to ensure sufficient detachment. Subsequently, cells were gently scrapped from the plate with a cell scraper and transferred to a new sterile Eppendorf tube. After spinning down the pellet by centrifugation at 4000 rpm for 1 minute, the cell suspension was transferred to another new sterile Eppendorf tube, and used straightaway or stored at -80 °C.

3.4.5 Western blot

Cell lysates were prepared by washing cells with cold phosphate buffered saline (PBS), followed by lysis buffer addition (50 mM Tris, 0.1 mM EGTA, 1mM EDTA, 0.27 M sucrose, 1 % Triton X-100, Protease Inhibitor Cocktail (Roche) Phosphatase Inhibitor Cocktail (Sigma)). Cell lysates were then collected and clarified (removing nucleus and unbroken cells) by centrifugation at 12,500 rpm for 10 minutes at 4°C. Proteins were separated on Novex precasted Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) Bis-Tris 4-12% gels (Invitrogen), using MES running buffer (Invitrogen). Proteins were then transferred to Polyvinylidene difluoride (PVDF) membranes (Millipore), blocked with 5 % milk and blotted with *JAM2* (abcam ab156586) and actin (sigma A2228 loading control) antibodies.

Animal studies The mouse strain used in this thesis is $Jam2^{tm1.2Rha}$. Jam-b (ortholog of human *JAM2*) knockout (KO) and wild type mice were obtained from Professor Michel Aurrand-Lions, a collaborator in the Centre de Recherche en Cancérologie de Marseille, France. Mice were housed in the Institute of Neurology University College London animal facility and were maintained on a 12-hour light dark cycle at a constant temperature and humidity. All animal experiments were carried out according to the UK Animal Act 1986 and approved by the UCL Animal Care Committee. We received 16 mice. Nine Jam-b KO, 2 male Jam-b KO and 9 female wild type.

3.4.5.1 Generation of Jam-B deficient mice

This colony was developed some time ago and was previously described in the literature(270). Briefly, a jam-b gene fragment containing exon 5 was flanked by loxP recognition sites for Cre recombinase. Within the loxP-flanked region, a murine jam-b cDNA fragment (codons 165-298, PEY...SFII*) followed by the bovine growth hormone polyadenylation signal (pA) was fused to a BspE1 site in exon 5. The construct also contained a neomycin resistance cassette surrounded by frt sites enabling Flp recombinase-mediated removal, as well as long (9 kb, 5') and short (1.65 kb, 3') arms for homologous recombination. Electroporated and G418-selected embryonic stem cell clones were analysed by Southern blot hybridization and polymerase chain reaction (PCR). For PCR screening, a primer pair, derived from exon 4 and the intronic jam-b sequence flanking the 5' loxP site, amplified a 345 bp band from WT chromosomes and a 450bp band from transgenic chromosomes (primers were 5'-AGACCGTGCTGAGATGATAGA-3' and 5'-CCGAAGGAAGTGTCTAGTAAT-3'). Three independent lines were generated and maintained in a mixed 129Sv × C57BL/6 background. jam-b deficient mutants were generated by cross-breeding with the PGK-Cre line followed by interbreeding jam-bKO/+ heterozygotes. Mice used in the present study were backcrossed for more than 12 generations on C57Bl/6 background.

3.4.5.2 Behavioural studies

Behavioural studies were performed in the Jam-b mice and compared to controls. All mice were trained in a standard way before the relevant tests. Figure 3-14 presents 2 cartoons of the tests used in this thesis.



Figure 3-14: A: Walking beam test. Mice are placed on the beam and their ability to traverse it is considered to be an indication of their balance. Paw slips and traverse time are used as the measurements. B: Footprint analysis. The paws are dipped in ink or paint, so that the mice leave a trail of footprints as they walk or run along a corridor to a goal box. Measurements of stride length, base width, and fore and hind paw overlap give an indication of gait. Reproduced with permission from (271).

3.4.5.2.1 Beam walking test

The balance beam assesses a mouse's ability to maintain balance while traversing a narrow beam to reach a safe platform. It was originally designed to assess motor deficits in aged rats and has proved useful in assessing motor coordination and balance in mice. Measurements recorded can include the time taken to cross the beam and the number of paw faults or slips. Some versions use a range of cross sections and diameters to vary the task difficulty. Others use a beam that becomes progressively narrower as it approaches the safety platform. In early versions of the test, which use simple square and round cross sections, the mice may occasionally fall, and the frequency of falls becomes an additional dependent variable. Two useful modifications of beam design can promote a mouse's willingness to progress rapidly across to the escape platform rather than simply cling on to prevent falling: an additional ledge can be placed either side of the platform so that even

when the paw slips grip is maintained; and an inclined beam can be used instead of a horizontal beam, as mice seem to have a natural tendency to run upwards to escape(271).

In this thesis, we used the beam walking test to assess the number of missteps of the Jam-2 KO compared to controls when they walked through a round beam.

3.4.5.2.2 Footprint analysis

A detailed analysis of motor coordination and synchrony is provided by examining gait during normal walking. The more commonly used method for assessing gait is the 'footprint' test. The fore and hind paws are painted with dyes of different colours and the mouse is encouraged to walk in a straight line (typically in a narrow corridor) over absorbent paper. The footprint patterns are then analysed for a range of measurements, including stride length, base width, overlap between fore and hind paws, and paw and finger splay.

Gait analysis is not only simple, and the scoring straightforward, but also it is one of the few tests that translates directly from animal to human studies. Although automated video analysis is possible, it is too early to determine whether this matches the sensitivity of the time-consuming manual analysis(271).

In this thesis, we used footprint analysis to assess the stance, stride and sway length comparing Jam-b KO to wild type mice.

3.4.5.3 Mice culling and CNS dissection

Mice were culled by injection of a barbiturate overdose (pentobarbital). The brain and the spinal cord were harvested and fixed in paraffin for histopathological analysis. Sections were performed with a microtome in the following way: the brain was sampled with a mid-sagittal cut and distributed into 2 different blocks, to ensure that sufficient material was available. The spinal cord was prepared according to a standard cutup protocol which separates cervical, thoracic, and lumbar segments and distribute them into 3 separate blocks. That way, all spinal cord material will be sampled. Each mouse generated 5 blocks.

3.4.5.4 Mice brain pathology

Histopathology is the microscopic study of diseased tissue, also called microscopic anatomy. It is performed by examining a thin tissue section under the light microscope. The technique

consists shortly of paraffin embedding, sectioning, and staining. Other techniques are available and could be performed on tissues for specific purposes (special stains, immunohistochemistry, *in situ* hybridization, etc.)(272).

In this thesis, mouse brain and spinal cord samples were sectioned and fixed in paraffin as described in chapter 3.4.5.2, and then stained with Haematoxylin and Eosin (H&E). After morphological evaluation of the tissue, markers of neuronal cells (NeuN) and glial cells (GFAP) were used for further investigations.

Basophilic and acidophilic staining: Acidic dyes react with cationic or basic components in cells. Proteins and other components in the cytoplasm are basic, and will bind to acidic dyes.

Basic dyes react with anionic or acidic components in cells. Nucleic acids are acidic, and therefore bind to basic dyes(273).

3.4.5.4.1 H&E

Haematoxylin is nearly a specific stain for chromatin and it is therefore referred to as a "Basic" stain. Although the stain itself is not basic, the dye called hematein is used in combination with aluminium ions that bind to acidic (or basophilic) structures. It stains the nuclear network and chromosomes in colour blue. It may be used after almost any fixative and is a permanent stain.

DNA (heterochromatin and the nucleolus) in the nucleus, and RNA in ribosomes and in the rough endoplasmic reticulum are both acidic, and so haematoxylin binds to them and stains them purple.

Eosin is a red general cytoplasmic stain. It combines with haemoglobin to give an orange colour. It is an acid dye and the terms acidophilic, oxyphilic and eosinophilic are often used interchangeably. It may be used after any fixative and is used as a counter-stain in many combinations in addition to haematoxylin.

Most proteins in the cytoplasm are basic, and so eosin binds to these proteins and stains them pink. This includes cytoplasmic filaments in muscle cells, intracellular membranes, and extracellular fibres.

H&E staining is used routinely in pathology for recognising tissue types and morphological changes. It achieves this by clearly staining cell structures including the cytoplasm, nucleus, and organelles and extra-cellular components(272,273). An example of brain tissue stained with H&E is shown in Figure 3-15.

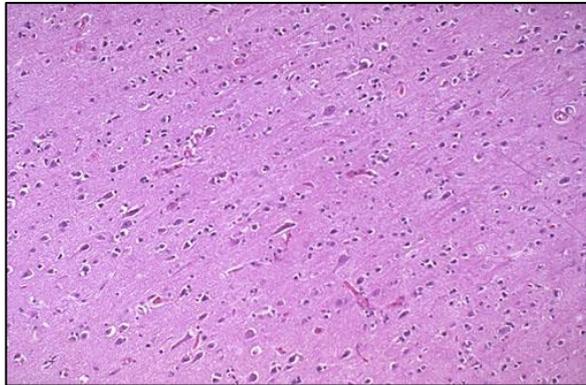


Figure 3-15: Normal brain tissue stained with H&E visualised on a medium magnification (100X). The cortex of the cerebrum is shown with triangular neurons and scattered small darker glial cells in a pink neuropil background. Reproduced from (274) (permission requested).

3.4.5.4.2 GFAP

Expression of glial fibrillary acid protein (GFAP) has become a prototypical marker for immunohistochemical identification of astrocytes. GFAP was first isolated as a protein highly concentrated in old demyelinated plaques from multiple sclerosis patients and was then found to be associated immunohistochemically with reactive astrocytes in such plaques and in other pathological contexts. GFAP expression is a sensitive and reliable marker that labels reactive astrocytes that are responding to CNS injuries(275). An example of brain tissue stained with GFAP is shown on Figure 3-16.

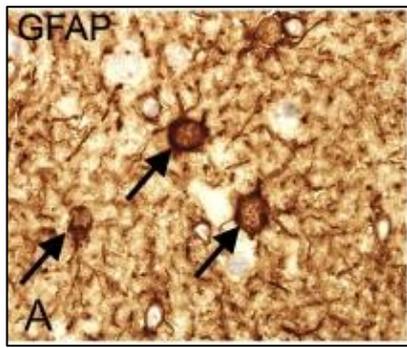


Figure 3-16: Visualization of astroglial, oligodendroglial, and microglial cell bodies in the parietal neocortex from an adult donor. Arrows show staining astrocytes immunohistochemically positive with GFAP. Reproduced from (276) (no permission needed). Magnification 400X.

3.4.5.4.3 NeuN

A monoclonal antibody against the neuronal nuclear protein (NeuN) is used in immunohistochemistry as a specific neuronal marker (Figure 3-17). This marker has not been detected in tissues other than nervous ones. Moreover, the protein has never been detected in glial cells and makes it a reliable marker of neuronal tissue in the brain. NeuN has been successfully used for more than 20 years as a reliable marker of postmitotic neurons in studies of neuronal differentiation and in the assessment of neuronal status both in norm and pathology(277).

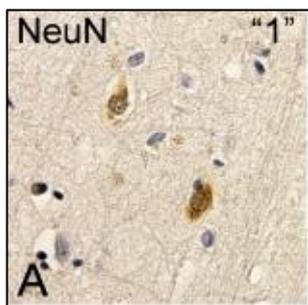


Figure 3-17: Immunohistochemical stainings for NeuN as a neuronal marker. Reproduced under a Creative Commons Attribution License from (277) Magnification 400X.

4 CHAPTER 4. CANDIDATE GENE STUDIES IN MSA

This chapter consists of the analysis of different candidate genes that have been linked to MSA, PD or other relevant neurodegenerative disorders. Depending of the type of mutation this has been achieved by different methods that will be explained in each section.

4.1 *EIF4G1*

Statement of contribution: Samples were collected by Prof Henry Houlden and Dr Anna Sailer. Experiments and data analysis were performed by myself.

4.1.1 Background

The *EIF4G1* gene has been recently linked to PD with an autosomal dominant inheritance(278). Mutations were first found in a French family and later detected in other 4 families out of 96, and in 9 out of 4708 cases with sporadic PD. Chartier and colleagues showed *in vitro* that mutations in *EIF4G1* increased the vulnerability of cells to oxidative stress and suggested translation initiation as an affected pathway in PD(278). We therefore decided to investigate if mutations in *EIF4G1* are also a cause of MSA.

4.1.2 Subjects, materials and methods

Two-hundred and fifty-one pathologically confirmed MSA cases were included. Most of the mutations lie within *EIF4G1* exons 8, 10 and 22 so we decided to focus on those exons for our analysis. Primers covering exons and splice sites are described in the Appendix. The reference sequence used was NM_182917.3. Exon 8 is very large and therefore we divided it into 2 parts, A and B.

Sanger sequencing experiments and analysis was performed as described in chapter 3.3.4.

4.1.3 Results

We successfully sequenced exons 8, 10 and 22 of *EIF4G1* in 251 definite MSA cases. We failed to detect any known PD mutations. We couldn't find any variant in exon 10 and 22 and changes in exon 8 are listed in Table 4-1.

Table 4-1: Variants in exon 8 of EIF4G1 in MSA.

Variant	Exon	AA	call	Number of cases	MAF MSA	MAF dbsnp (EUR 1000genomes)	p value (χ^2)*
rs2178403	8	p.M432V	het hom	96 16	0.254	0.243	0.6223
novel in-frame deletion	8	p.E453del	het	1			
rs111659103	8	p.G466_A468del	het hom	7 2	0.021	0.033	0.2490
novel	8	p.P486S	het	1			

* p value result from a χ^2 test comparing the minor allele frequency of the SNPs detected in our study with the data from the 1000 genomes project.

4.1.4 Discussion

Two variants (rs2178403 and rs111659103) are common changes previously reported in the literature and the MAF in our cohort are similar to those reported in the European cohort of the 1000genomes project.

Two coding variants are novel. One is an in-frame deletion close to another one that is common in the population (rs111659103) and possibly well tolerated too, and the last one is a missense mutation predicted by online software Polyphen2 (<http://genetics.bwh.harvard.edu/pph2/index.shtml>) to and SIFT (<http://sift.jcvi.org/>) to be benign.

Importantly I would like to comment that there is recent controversy of the role of *EIF4G1* variants in PD and there is a recent paper highlighting the presence of these variants in normal controls at a higher frequency than in PD(278). In conclusion, I could not find a link between *EIF4G1* and MSA.

4.2 VPS35

Statement of contribution: Samples were collected by Prof Henry Houlden and Dr Anna Sailer. Experiments and data analysis were performed by myself.

4.2.1 Background

The gene *VPS35* encoding for vacuolar protein sorting 35 that is a component of the retromer complex and mediates retrograde transport between endosomes and the trans-Golgi network. This gene has been recently reported as a rare cause of autosomal dominant PD by two groups originally finding mutations in a Swedish and an Austrian family(279,280). As MSA belongs to the group of the α -synucleinopathies together with PD and DLB we decided to investigate this gene in our pathologically confirmed cases with a diagnosis of MSA and DLB. We selected exon 15 for our analysis because it is the one containing the originally described PD causing mutation (p.D620N).

4.2.2 Subjects, materials and methods

Primer sequences to cover the exon 15 and intronic flanking regions of *VPS35* are presented in the Appendix. Sanger sequencing was performed as described in chapter 3.3.4. Two-hundred forty-nine pathologically confirmed MSA cases and fifty-one pathologically confirmed DLB cases were included.

4.2.3 Results

Screening of exon 15 of the *VPS35* gene in 249 MSA cases and 51 DLB cases failed to detect pathogenic mutations. We found one synonymous change (rs138794859, p.A627=) in a case with a diagnosis of DLB.

4.2.4 Discussion

VPS35 is a rare cause of PD and is probably not linked to other synucleinopathies like MSA or DLB.

4.3 LRRK2

Statement of contribution. Data from the UK cohort was obtained by myself either by analysing genotyping data or by Sanger sequencing of the missing data. Data analysis of the merged cohort was done by our collaborators from the Mayo clinic Dr Mike Heckman and Dr Owen Ross.

4.3.1 Background

LRRK2, encoding for the leucine-rich repeat kinase 2 is the most common genetic cause of PD and also presents variants that have risk-modifying variants to this disorder(281). Previous studies have looked into the p.G2019S mutation in MSA and a Taiwanese study investigated MSA risk with the p.G2385R variant, both with negative findings(282,283). In this section of this thesis, in collaboration with Dr Mike Heckman, we have investigated common *LRRK2* coding variants for association with MSA.

4.3.2 Patients, materials and methods

We included data from 85 UK patients with pathologically confirmed MSA from the QSBB collected between 1987 and 2009 and 352 population matched controls from the British 1958 Birth Cohort. All our MSA patients were of European descent.

Genotyping data was obtained by extracting the data from the custom-built Immunochip. Data was extracted using the software Genome studio and missing data was completed by Sanger sequencing.

The data obtained in this UK cohort was merged with data from our collaborators in the Mayo clinic in Jacksonville (US series) and they performed the final analysis of the merged cohort of MSA patients and compared for association with controls.

Logistic regression under a dominant model was used to test for association using the software R (version 2.14.0). To account for multiple testing a single step minP permutation adjustment was used after which p values ≤ 0.0044 (for the US series), ≤ 0.0054 (UK series), and ≤ 0.0047 (combined series) were considered statistically significant.

4.3.3 Results

Seventeen coding *LRRK2* common variants with $MAF \geq 5\%$ were analysed in the combined series of UK and US patients (177 patients and 768 controls). The results are presented in Table 4-2. The variant p.M2397T presented a significant association (after adjustment for multiple testing) with a protective effect towards risk of MSA (OR 0.6 CI 0.43-0.84; $p = 0.002$)

Table 4-2: *LRRK2* coding variants analysed for association in 177 definite MSA cases and 768 controls. Values in **bold** are significant. (From (284) (no permission needed).

Variant	Amino acid	Position	MA	OR	95% CI	p value
rs10878245	L153L	38918058	T	0.99	(0.71, 1.39)	0.95
rs7308720	N551K	38943967	G	0.88	(0.54, 1.42)	0.59
rs10878307	I723V	38958256	G	1.05	(0.65, 1.69)	0.85
rs7966550	L953L	38974962	C	0.68	(0.45, 1.02)	0.064
rs7133914	R1398H	38989178	A	0.83	(0.51, 1.35)	0.45
rs11175964	K1423K	38989254	A	0.89	(0.55, 1.45)	0.64
rs35507033	R1514Q	38994045	A	NA		1
rs33958906	P1542S	38994128	T	0.88	(0.42, 1.83)	0.72
rs1427263	G1624G	39000101	C	0.63	(0.45, 0.88)	0.006
rs11176013	K1637K	39000140	A	0.8	(0.57, 1.14)	0.22
rs35303786	M1646T	39000166	C	1.64	(0.63, 4.24)	0.31
rs11564148	S1647T	39000168	A	0.83	(0.59, 1.15)	0.25
rs10878371	G1819G	39002527	T	0.84	(0.60, 1.20)	0.34
rs33995883	N2081D	39026953	G	0.15	(0.02, 1.08)	0.01
rs10878405	E2108E	39028521	A	0.97	(0.70, 1.35)	0.87
rs33962975	G2385G	39043597	G	1.36	(0.95, 1.94)	0.096
rs3761863	M2397T	39044919	T	0.6	(0.43, 0.84)	0.002

Ref: MA: Minor allele; OR: odds ratio; CI: confidence interval.

4.3.4 Discussion

In this investigation we found an association for MSA with the variant p.M2397T (rs3761863). This variant has not been previously associated with PD but in this large study we found it was associated with a protective effect on MSA risk. This is the first study to report a significant association of MSA with variants in *LRRK2*. More studies will be needed to replicate these findings in a larger cohort and across populations.

The role of *LRRK2* in MSA has been recently highlighted in view of the publication of a neuropathologically confirmed MSA case carrying p. Ile1371Val mutation in *LRRK2*. This variant is not significantly associated with PD in association studies but it was previously reported in 5 PD families(285).

4.4 C9orf72

Statement of contribution: DNA of PSP samples were provided by Dr Rohan de Silva and the CBD cohort was provided by Dr Rohan de Silva and Dr Helen Ling. I collected MSA samples, I performed DNA extractions, RP-PCR, sizing-PCR experiments and data analysis. Dr James Polke performed the Southern blot.

4.4.1 Background

Recently, a breakthrough in genetic research of neurodegenerative diseases has occurred. An intronic expansion in *C9orf72* was discovered as a major cause of familial and sporadic ALS and frontotemporal dementia (FTD)(252,286). These cases exhibit TAR DNA-binding protein-43 (TDP-43) pathology.

There is extensive evidence that supports the concept of pathological overlap between neurodegenerative disorders and up to 35% of these cases may exhibit signs consistent with PSP and CBS. In particular, they may present slowness of vertical saccades, parkinsonism, frontal dementia and abnormal DATSCANS(287–293).

Screening for the *C9orf72* mutation in other neurodegenerative diseases has been performed, and expansions detected in patients with FTD-ALS and parkinsonism, clinical PD, CBS, clinical-PSP, clinical-OPCA and clinical DLB (289,294,295). The expansion has been also detected in few families with Alzheimer's disease (AD) which in some cases exhibit AD pathology but in most of them could represent a clinical misdiagnosis of FTD(296–298). A large UK study showed an association of the *C9orf72* expansion with AD, Huntington's disease (HD)-like patients, and other neurodegenerative diseases despite a large frequency in controls(299). One expansion in PD was found in a pathological study in the UK but this case also exhibited TDP-43 pathology(300).

MSA, PSP and CBD conform a group of diseases known as atypical parkinsonian disorders that present overlapping features, and therefore we thought it would be interesting to include samples with a diagnosis of PSP and CBD in this section of our project.

Clinically, PSP patients present with postural instability with early falls, cognitive dysfunction and abnormalities of vertical gaze(293). Pathologically, PSP cases exhibit neuronal loss, astrocytosis and neurofibrillary tangles in basal ganglia and brainstem nuclei. The tau deposition is predominantly 4-repeat tau into both neuronal and glial inclusions.

The CBS phenotype consists of asymmetric parkinsonism, cortical signs (apraxia, cortical sensory loss, alien limb), and dystonia and myoclonus(293). CBD cases are characterized by cortical and striatal 4 repeat tau-positive neuronal and glial lesions, especially astrocytic

plaques and thread-like lesions in both white and grey matter, with neuronal loss in cortical regions and in the substantia nigra(301).

4.4.2 Subjects material and methods

We screened for the *C9orf72* hexanucleotide repeat expansion in 96 pathologically confirmed MSA patients. We included 177 PSP, 18 CBD, 37 CBS and 22 clinical PSP samples as comparison groups(302).

To provide a qualitative assessment of the presence of an expanded hexanucleotide repeat of GGGGCC in *C9orf72*, we performed a repeat-primed-PCR reaction and fragment analysis as described in the methods chapter. In cases where we detected an expanded allele and cases where we couldn't detect both normal alleles we further characterized the expansion with a sizing PCR reaction. Genotyping primers using one fluorescently labelled primer were the forward from the repeat-primed PCR reaction and reverse genotyping primer that was previously published and are presented In the Appendix(252).

PCR products were analysed by fragment length analysis on an automated ABI3730 DNA-analyser and allele identification and scoring was accomplished using GeneMapper v3.7 software (Applied Biosystems) as described in the methods.

Haplotype analysis to determine the presence of the expansion-associated risk haplotype was performed by Sanger sequencing of the rs3849942 marker(303).

Statistical analysis was performed with Open Source Epidemiologic Statistics for Public Health software (http://www.openepi.com/Menu/OE_Menu.htm).

4.4.3 Results

We detected an expansion in 3 cases fulfilling clinical criteria of CBS. There are a significant difference when compared to published data on controls(304) ($p < 0,001$). These results can be found in Table 4-3. Interestingly we also found an expansion of 27 repeats in another case presenting with atypical parkinsonism. An example of a RP-PCR and a sizing PCR of a heterozygous expanded case is provided in Figure 4-2.

Haplotype analysis performed in the three expanded patients as well as the 27 repeat patient detected one A allele of the SNP rs3849942 in all of them and confirmed they are all carriers of the associated risk haplotype(303).

The clinical characteristics of the expansion carriers are described in

Table 4-4 and the family tree of the patient with an intermediate repeat of 27 is shown in Figure 4-1.

Table 4-3: Results of C9orf72 expansion screening in atypical parkinsonism. Reproduced from (305) under a Creative Commons Attribution License.

Diagnosis	Number of samples	Expanded
MSA (pathologically confirmed)	96	0
CBD (pathologically confirmed)	18	0
PSP (pathologically confirmed)	177	0
CBS (clinical)	37	3 (p < 0.001*)
PSP (clinical)	22	1 (27 repeats)
British controls (clinical)	7579	11

*Ref: CBD, corticobasal degeneration; CBS, corticobasal syndrome; MSA, multiple system atrophy; PSP, progressive supranuclear palsy. * Fisher exact test comparing our CBS cohort with previously published British controls*

Table 4-4: Clinical characteristics of the three cases with the C9orf repeat expansion and of the case with an intermediate repeat. Reproduced from (305) under a Creative Commons Attribution License.

Age of onset, gender	Repeat	Initial presentation	Previous psychiatric features	Other clinical features	Working diagnosis	CBD variant (consensus criteria)	Family history	Imaging
51, F	Expanded	Falls and personality change	Depression	Akinetic-rigid syndrome with cognitive decline and asymmetrical upper-motor neuron signs. Progressed with minimal L-dopa response and atrophy on MRI brain scan.	CBS	Probable-CBS	Father with dementia in his early fifties	MRI: generalized volume loss. DATSCAN: bilateral severe deficit of presynaptic dopamine transporter in the basal ganglia.
44, F	Expanded	Parkinsonism after head trauma	No	Atypical PD with cognitive decline after 9 years, asymmetrical limb manifestations, falls, hypometric saccades and frontal liberation signs, dysphagia and dysarthria	CBS	Probable-CBS	Mother diagnosed with Pick's disease	MRI: atrophy.

60, M	Expanded	Parkinsonism (hypomimia, shuffling gait and reduced arm swing)	Possibly	Tremor, dysphagia and dysarthria, impaired memory and reduced concentration, staring expression with frontalis over-activity. Asymmetrical parkinsonism poorly responsive to L-dopa.	CBS	Probable-CBS (Frontal-behavioural-spatial syndrome)	No	MRI: generalized volume loss with marked involvement of the frontal and temporal lobes more severe on the left.
74, F	Intermediate allele of 27 repeats	Writing difficulties and falls	No	Parkinsonism, restriction of vertical gaze, cognitive decline, echolalia, brisk reflexes, dysphagia	Clinical-PSP	Probable PSP	See pedigree Figure 4-1.	MRI: the superior cerebellar peduncles were found slender, general atrophy.

Ref: CBS, corticobasal syndrome; PSP, progressive supranuclear palsy.

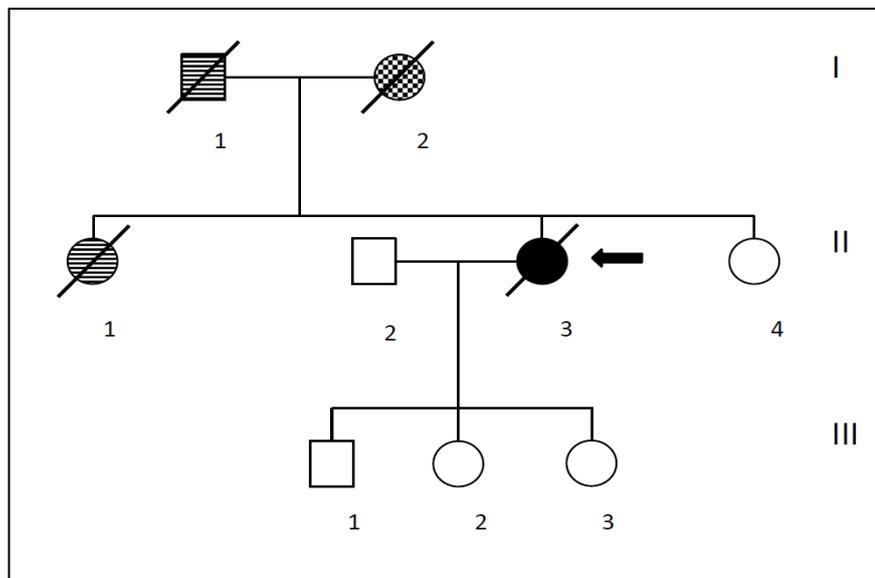


Figure 4-1: Family tree of the patient (II-3, arrow) with a 27-repeat allele. The patient's father (I-1) and sister (II-1) were diagnosed with dementia, and the patient's mother (I-2) was diagnosed with Parkinson's disease. Reproduced from (305) under a Creative Commons Attribution License.

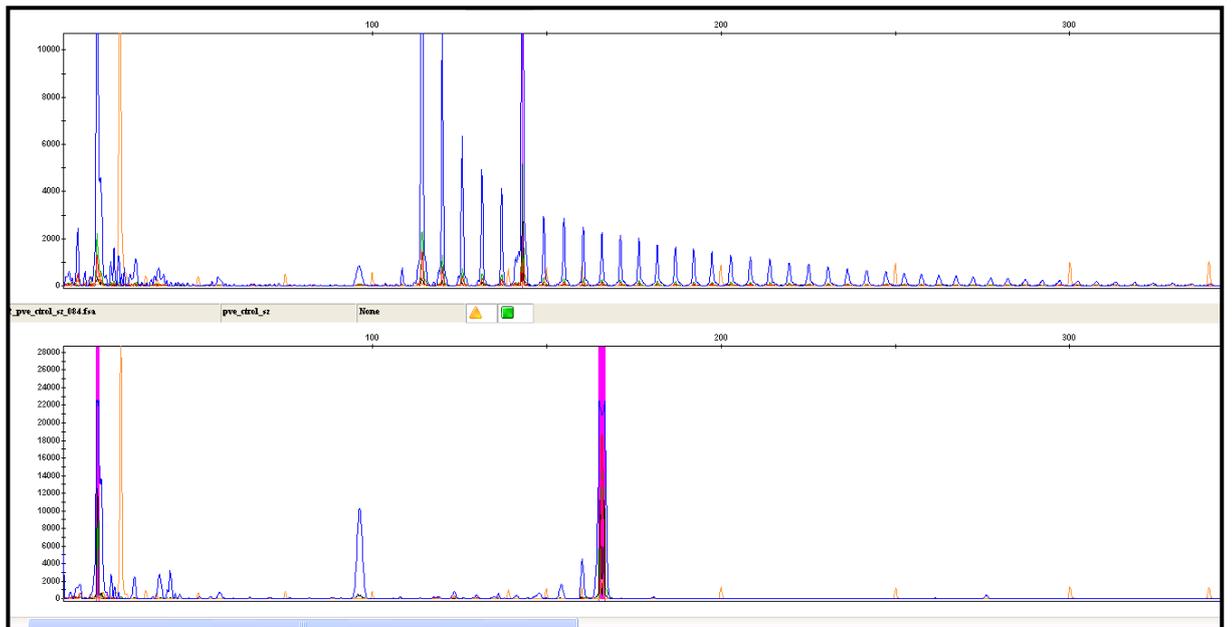


Figure 4-2: This figure illustrates fragment analysis example of a heterozygous carrier of the C9orf repeat expansion. Top graph: Repeat-primed PCR. Bottom graph: Sizing PCR. Reproduced from (305) under a Creative Commons Attribution License.

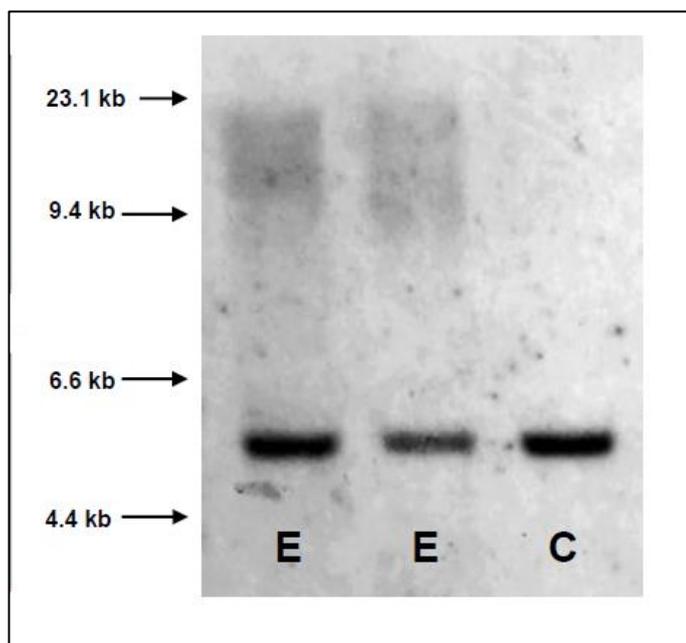


Figure 4-3: Southern blot (with *BsU36I* restriction digest) showing 2 of the expanded CBS cases. Abbreviations: C, control no expansion; CBS, corticobasal syndrome; E, expanded. Reproduced from (305) under a Creative Commons Attribution License.

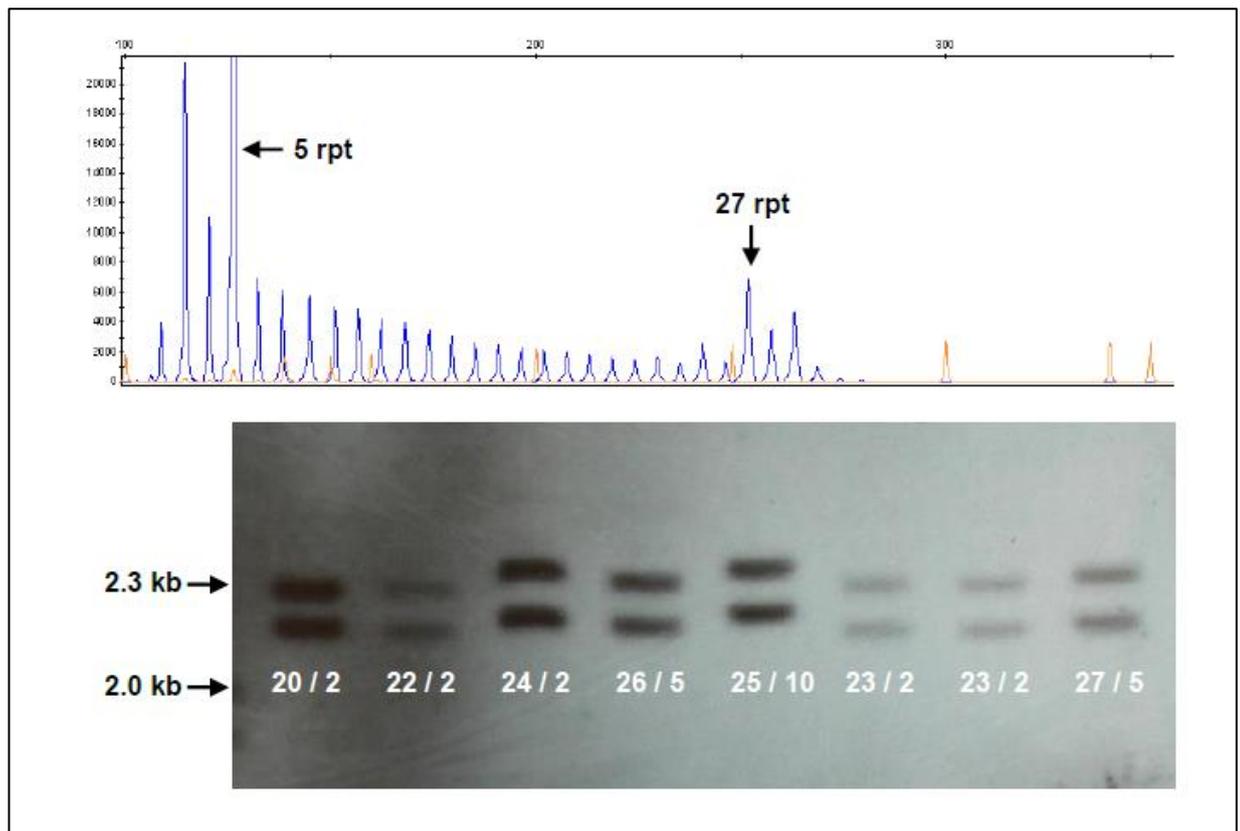


Figure 4-4: Fig. 2. Top figure: RP-PCR showing the case with heterozygous 27 repeats on 1 allele, and 5 repeats on the other allele. Bottom figure: Southern blot confirmation of different repeat sizes from 20 to 27 showing the Southern blot appearance of different fragments. The number of repeats was also confirmed by fluorescent PCR of the C9orf72 repeat and fragment analysis. Reproduced from (305) under a Creative Commons Attribution License.

Ref: PCR, polymerase chain reaction; RP-PCR, repeat-primed polymerase chain reaction

4.4.4 Discussion

Cases with 20-22 repeats were reported in FTD without clinical differences with expanded cases and proved segregation(306). Repeats between 7-24 have been shown to strongly correlate with the C9ORF72 expansion risk haplotype. Gene expression studies suggested that the intermediate repeats result in a significantly reduced promoter activity(307). The C9ORF72 intermediate repeat copies were found to be a significant risk factor for clinical-PD in a Spanish study(308). Four cases with intermediate repeats (21,23,24,38) were found in a USA study of 781 PD cases without segregation in the latter(309); and 4 out of 289 in a Canadian study(298). No segregation of 2 intermediate repeats in 2 AD families were described in the same study. No segregation in 1 homozygous PD case and no repeat instability over 2 generations in an Australian study(310), and also no repeat instability in a

Belgian study(307). A possible risk factor over many generations has been proposed(307,310). In contrast, intermediate repeat instability in controls over 3 generations was found in a UK study(299).

A recent interesting report described the C9orf72 repeat expansion in a British-Canadian family in whom the father who is 89 years old and carries a 70-repeat allele is not affected by the disease and four children with a larger expansion ~1750 are affected by ALS. Epigenetic studies showed that the offspring expansions were methylated and the fathers were not. RNA foci were detected in the offspring but again not in the father. This interesting study denotes the complexity of the different mechanisms for disease manifestation that may involve multiple mechanisms related and not related to the number of repeats(311).

The significance of short/intermediate repeats is still unknown. Unfortunately, we are unable to investigate segregation in this family because of lack of DNA in deceased patients, but we could confirm the presence of the risk haplotype in this patient.

To sum up, the *C9ORF72* expansion can be detected in clinical cases presenting with parkinsonian syndromes that may have overlapping features with the FTD-ALS spectrum. These data show a significant association when compared to controls. We failed to detect an expansion in pathologically proven cases with MSA, PSP or CBD, and this confirms that *C9ORF72* is not related to α -synuclein or tau pathology and the significance of intermediate repeats is under study. Although the common European risk haplotype(303) is not always present in expansion carriers(312), we think it is supporting the role that an intermediate repeat presents in our clinical PSP case.

4.5 COQ2

4.5.1 Background

In this chapter I investigated the role of *COQ2* variants in MSA samples of European origin. I first began by analysing the paper describing the association between *COQ2* variants and the risk of MSA(185), then went on to sequence the gene in all our pathologically confirmed

MSA cases and compared them to controls and analysed GWAS data for association. Finally, I measured Coenzyme Q10 levels in the brain as the final product of the pathway.

4.5.2 *COQ2* paper analysis

The original paper has been described in chapter 2.2.6.2. Here, I will comment on some considerations I think are important when reading this interesting paper.

First of all, the data shown by the Japanese paper presents an association of *COQ2* variants and MSA risk, and functional work done in rare variants in patients with familial MSA of which only 1 family was pathologically confirmed. Secondly, the association of the p.V393AV393A variant with increased risk of MSA was significant in Japanese patients but was not replicated in other populations. They only show rare variants detected in Europeans but not significant results when compared to controls(185). Also, they filtered against dbSNP130 which contains benign polymorphism data as well as disease-causing mutations. In this way, it is possible that a relevant genomic change could have been filtered out.

4.5.2.1 *COQ2* transcripts

The reference genome used was NCBI36 and for *COQ2* analysis they focused in transcript ENST00000311461 that is not the longest coding transcript. This particular one considers the first 136 base pairs of exon 1 as non-coding and therefore their analysis skips a large and potentially important section of the gene. The protein encoded by this transcript is 371 amino acid long, whereas the longest one contains 421 amino acids (ENST00000311469).

Forsgren et al(313) compared the functional activity of different human CoQ2 isoforms by using a yeast model with disrupted endogenous *COQ2* gene. Their results suggest that the first 50 amino acids are required for a functional active human CoQ2 protein.

In order to understand if the first 50 amino acids play a role in the human brain, I analysed brain expression data from normal subjects obtained from a publicly available database, BRAINEAC(314). These data are presented in Figure 4-5 and Figure 4-6. Figure 4-5 shows the expression of *COQ2* from a probe located within the first portion of exon 1, a part that is not included in the transcript ENST00000311461 used by the Japanese group. Figure 4-6 presents the mean expression levels of *COQ2* in different brain regions. These data confirm

that all 7 exons of *COQ2*, including the 5' ending of exon 1 are expressed in human brain tissue.

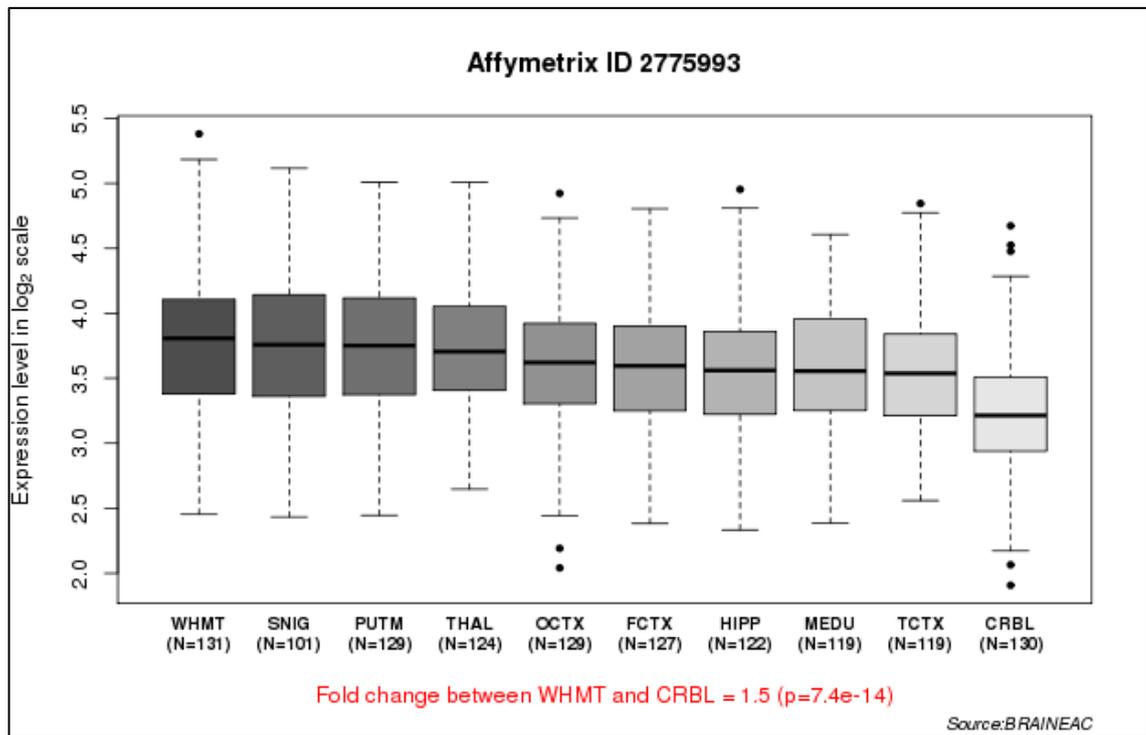


Figure 4-5: Expression levels of *COQ2* corresponding to the Affymetrix probe located in the 5' end of exon 1. Data publicly available obtained from BRAINEAC(314).

Ref: WHMT: white mater; SNIG: substatia nigra; PUTM: putamen; THAL: thalamus; OCTX: occipital cortex; FCTX: frontal cortex; HIPP: hippocampus; MEDU: medulla; TC TX: temporal cortex; CRBL: cerebellum.

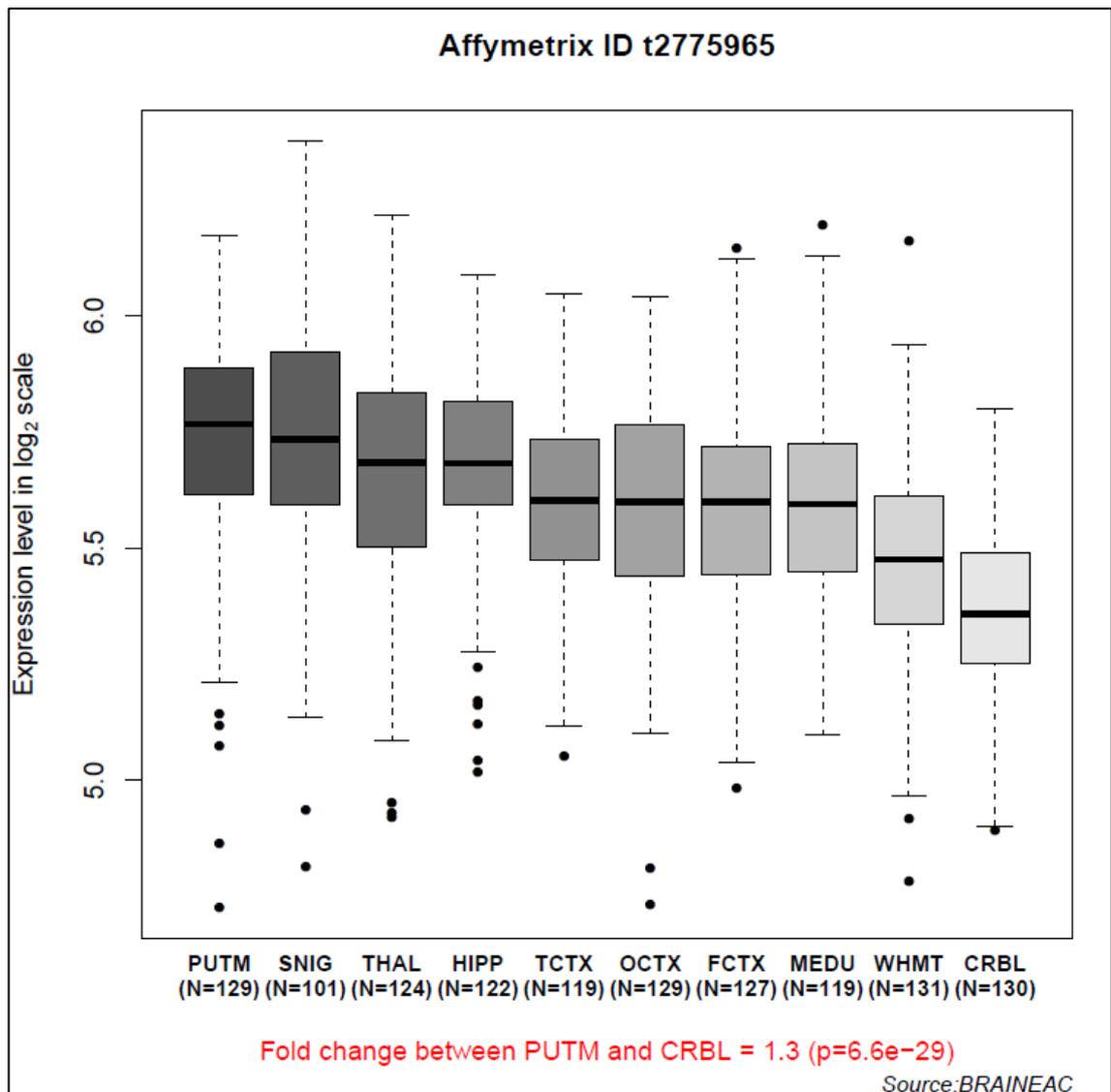


Figure 4-6: Mean expression levels of *COQ2* in different regions of control brains. Data publicly available obtained from BRAINEAC(314).

Ref: WHMT: white mater; SNIG: substatia nigra; PUTM: putamen; THAL: thalamus; OCTX: occipital cortex; FCTX: frontal cortex; HIPP: hippocampus; MEDU: medulla; TC TX: temporal cortex; CRBL: cerebellum.

4.5.2.2 Linkage region genomic location

The proposed region of genetic linkage, is very close to the *SNCA* locus. Although we know that at least in Family 1 this gene was sequenced before(315), we do not know if they excluded *SNCA* CNVs in these families though. Using the online software genome browser, I highlighted in Figure 4-7 the *SNCA* gene location lying only 1.2 Mbase away from the linkage region and this is very close to *COQ2*. We therefore think it would be important to exclude CNVs in *SNCA* in this family.



Figure 4-7: Distance between the linkage region in chromosome 4 and the SNCA gene, and also between COQ2 and SNCA. Green lines highlight the beginning of COQ2 and the end of SNCA and the red box shows the limits of the linkage region.

4.5.3 Sequencing COQ2 in definite MSA cases of Caucasian ancestry

Statement of contribution: Samples were collected by Dr Henry Houlden, Dr Anna Sailer and myself. All experiments and data analysis were performed by me with supervision from Dr Conceição Bettencourt. The map of collaborating brain banks was created by myself and edited by the graphic designer Sol Gonzalez Clement.

4.5.3.1 Subjects materials and methods

The first approach of our replication was to Sanger sequence the entire coding region (including splice sites) of COQ2 in our definite MSA cases. All samples were of European ancestry confirmed by genome-wide genotyping(168) or exome sequencing data.

We included a total of 305 pathologically confirmed MSA samples from different brain banks around the world (see map on Figure 4-8) and 264 controls from the Wellcome Trust 1958 British birth cohort as a comparison group.



Figure 4-8: Collaborating teams and brain banks that contributed samples for this project. Reproduced with permission from (163). Copyright Massachusetts Medical Society.

COQ2 consists of 7 coding exons. We utilized the reference sequence from NCBI37 and Ensemble transcript ENST00000311469 was selected because it provides the largest coverage of the coding region of this gene which is shown to be expressed in human brain

tissue and encodes for the longest functional protein isoform (see previous section). Sanger sequencing was performed as described in the methods chapter and primers are provided in the Appendix.

The sequences were analysed for base pair changes and insertions and deletions in the coding regions and flanking intronic regions of *COQ2*. Variant frequency was compared for association to our sequenced controls as well as data from online databases. Of note, we always selected data from European populations during the analysis.

For statistical analysis, we utilized an online open source software (<http://www.openepi.com/>) and we used the two by two table to compare allele frequencies. We used mid P exact test because in many cases genotype count was under 5. We also run the analysis in SPSS but this is not presented here because there was no difference in the results with the 2 software.

4.5.3.2 Results

Five samples had to be excluded because of non-European ancestry and 1 sample because it had a different diagnosis (PSP). One sample failed for exons 1,2,3,4 and we believe this is secondary to bad DNA quality. And another sample was excluded because of failure sequencing exon 1. Exon 1 was a particularly difficult GC rich exon that required various different sets of primers and PCR conditions. The PCR of the sample excluded was repeated more than 10 times and with different protocols. Although our first thought is that this failure is probably due to these difficulties with the amplification of exon 1, one cannot exclude a CNV with this method. This sample was sent to us by Dr Owen Ross from the Mayo clinic in Jacksonville.

The eight coding variants detected are listed in Table 4-5.

We failed to detect any variants that were reported in the Japanese population of MSA cases and controls. We found one variant, p.S57T, that in the paper(185) was reported in a European MSA case. This variant was present in our study in one MSA case and one control.

We identified a *COQ2* nonsense variant that was present at a higher frequency in controls than in MSA (p.R22X, 24 control vs 9 MSA alleles, $p < 0.0024$) and two other variants which

were also found at a significantly higher frequency in controls (rs6818847 and rs6535454). No association between the synonymous *COQ2* SNPs rs183012002 and rs1129617 was observed. Four heterozygous rare coding variants were detected: the p.S57T mutation reported by Tsuji and colleagues(185) was present in an MSA case and a control, the p.P68S variant was present in one MSA case as was rs121918231, and the rare SNP rs183012002 was identified in 2 MSA cases and 6 controls.

To increase the power of the study we compared MSA versus data from the European section of the 1000 genomes project (EUR) and we also merged our WT control data with the European data of the 1000 genomes project. There was no significant result under this analysis and values are shown in Table 4-6.

We tested control data sets for Hardy–Weinberg equilibrium and these results are shown in Table 4-7. It is of note that when any individual genotype is <5 this method is not accurate. rs6818847 and rs1129617 are not in equilibrium in the WT controls. Also, rs6818847 is not in equilibrium in the EVS data. All SNPs are in equilibrium in the 1000genomes project which makes those data the most reliable. rs183012002, rs112033303, rs121918231 and variants p.S57T and p.P68S have a genotype count under 5.

Table 4-5: COQ2 coding variants compared to the WT controls. Results in bold are considered statistically significant. Reproduced with permission from (163). Copyright Massachusetts Medical Society.

Variant	Function	Status	n	Alleles	Allele frequencies	OR (maf MSA vs maf WT)	P value*
rs183012002, p.R10R	synonymous	MSA_cases	299	G	0.997	0.29	0.13
				A	0.003		
		WT_controls	261	G	0.989		
				A	0.011		
rs112033303, p.R22X	nonsense	MSA_cases	299	A	0.985	0.32	0.002
				T	0.015		
		WT_controls	261	A	0.954		
				T	0.046		
rs6818847, p.V66L	missense	MSA_cases	299	T	0.696	0.75	0.02
				G	0.304		
		WT_controls	261	T	0.630		
				G	0.370		
p.S107T	missense	MSA_cases	299	G	0.998	0.87	0.93
				C	0.002		
		WT_controls	261	G	0.998		
				C	0.002		
p.P118S	missense	MSA_cases	299	C	0.998	N/A	N/A
				T	0.002		
		WT_controls	261	C	1		
				T	0		
rs121918231±, p.R197H	missense	MSA_cases	299	G	0.998	N/A	N/A
				A	0.002		
		WT_controls	262	G	1		
				A	0		
rs6535454, p.D298D	synonymous	MSA_cases	299	C	0.846	0.73	0.02
				T	0.309		
		WT_controls	262	C	0.671		
				T	0.337		
rs1129617, p.S330S	synonymous	MSA_cases	299	C	0.741	0.8	0.09
				T	0.259		
		WT_controls	262	C	0.695		
				T	0.305		

Ref: Minor allele frequency (MAF), 1000 genomes project (1000g), Wellcome trust controls from the British 1958 cohort (WT), odds ratio (OR), not available (NA). *p values are calculated with 2 tailed mi d P exact test.

Table 4-6: COQ2 coding variants compared to the 1000 genomes project data merged and unmerged to the WT controls. No significant associations were detected.

Variant	Function	MAF_MSA	MAF_WT	MAF 1000g	MAF_1000g plus WT	OR_MSA_vs_ 1000g	p value*	OR_MSA_vs_ 1000g plus WT	p value*
rs183012002	synonymous	0.0033	0.0115	0.0040	0.0070	0.8445	0.8792	0.4739	0.3567
rs112033303	nonsense	0.0151	0.0460	0.0185	0.0297	0.8120	0.6410	0.4994	0.0539
rs6818847	missense	0.3043	0.3697	0.3113	0.3352	0.9677	0.7828	0.8679	0.1843
c.G170C, p.S57T	missense	0.0017	0.0019	unreported	unreported	N/A	N/A	N/A	N/A
c.C202T, p.P68S	missense	0.0017	0.0000	unreported	unreported	N/A	N/A	N/A	N/A
rs121918231	missense	0.0017	0.0000	unreported	unreported	N/A	N/A	N/A	N/A
rs6535454	synonymous	0.3089	0.3373	0.2652	0.2942	1.0120	0.9206	0.8804	0.2516
rs1129617	synonymous	0.2592	0.3053	0.2533	0.2746	1.0310	0.8043	0.9244	0.4859

Ref: Minor allele frequency (MAF), 1000 genomes project (1000g), Wellcome trust controls from the British 1958 cohort (WT), odds ratio (OR), not available (NA). *p values are calculated with 2 tailed mid P exact test.

Table 4-7: Testing for Hardy-Weimberg equilibrium.

	n < 5	WT		1000g	
		x ²	p	x ²	p
rs183012002	yes	0.0353	0.8510	0.0060	0.9383
rs112033303	yes	0.6062	0.4362	0.1342	0.7141
rs6818847	no	14.4702	0.0001	0.8023	0.3704
c.G170C, p.S57T	yes	0.0010	0.9753	N/A	N/A
c.C202T, p.P68S	yes	N/A	N/A	N/A	N/A
rs121918231	yes	N/A	N/A	N/A	N/A
rs6535454	no	3.5990	0.0578	0.1891	0.6636
rs1129617	no	4.8648	0.0274	0.3958	0.5292

4.5.3.3 Discussion

None of the detected variants in our replication study present with an increased risk of MSA.

Overall, we think the results are variable when compared to different controls (WT and 1000 genomes project) because of the fact that our controls are not completely ethnically matched. Also, there is possible relatedness within the 1958 Wellcome Trust cohort because they are all subjects born in the Cambridgeshire region in the UK. This however, doesn't change our thoughts or interpretation of the data as explained in the following paragraph.

We believe it is unlikely that a gene that is suggested to increase risk of MSA by loss of function(185) will present a frequent nonsense variant (rs112033303) in the beginning of the first exon that would be "protective" for MSA and so common among the WT cohort with a MAF of 0.046. More recently, this variant was also found in other populations in the Exac database at a total MAF of 0.0299 (European maf 0.05521 including 8 homozygotes). Furthermore, this variant is in position +64 in exon 1 with transcript ENST00000311469, but would be in the 5'UTR according to the reference the Japanese team used in their paper (ENST00000311461), so that section of the gene was not analysed in that study.

In Table 4-8 I present all the replication studies performed in MSA to investigate the role of CoQ2 variants. In short, no Caucasian association has been detected (including the largest pathologically confirmed cohort which is the work presented in this chapter) and the associations found in Asians are all from candidate gene studies and in MSA-C patients. It could be possible that there are population specific differences that are contributing to these results so a well powered population corrected study is warranted to validate these data.

Table 4-8 Replication studies performed on the relation of MSA and COQ2 variants.

Ethnicity	Sample	Increased risk of MSA	Source
Japanese Sequencing exons 1,2,6,7	133 clinical MSA, 200 controls	p.L25V associated with MSA ($p=0.04$, OR 3.54) (MSAC driven)	<i>Sun et al, Neurol Genet, 2016(155)</i>
East Asians (Han Chinese) Selected variants And meta-analysis	82 probable MSA, 484 controls	p.V393A associated with MSA (OR 4.17, $p=0.004$) (MSAC driven) Meta-analysis: p.V393A associated with MSA (OR 2.05, $p=0.002$)	<i>Zhao et al, Neurol Sci, 2016(158)</i>
Chinese Sequencing all exons	153 MSAC, 798 controls	p.V393A associated with MSA (OR 3.1, $p<0.001$) (MSAC driven)	<i>Lin et al, Mov disord, 2015(316)</i>
Ethnic Chinese Sequencing all exons	312 clinical MSA, 100 controls	Negative	<i>Chen et al, NBA, 2015(158)</i>
North American (European) Sequencing all exons and CNV analysis	97 definite MSA 58 clinical MSA, 300 controls	Negative	<i>Ogaki et al. Molecular Neurodegeneration 2014(159)</i>
Chinese Sequencing all exons	116 MSAC, 192 controls	Negative	<i>Wen et al, CNS Neuroscience & Therapeutics, 2016(160)</i>
Korean Sequencing all exons	299 MSA, 365 controls	Negative	<i>Jeon et al, nejm 2014(161)</i>

European Genotyped variants	selected 788 clinical controls	MSA, 600	Negative	<i>Sharma et al, nejm 2014(162)</i>
UK (European) Sequencing all exons	300 definite controls	MSA, 262	p.R22X (OR 0.32, p = 0.002); p.V66L (OR 0.75, p = 0.02); p.D298D (OR 0.73, p = 0.02) all more frequent in controls than in MSA.	<i>Schottlaender et al., nejm 2014(163)</i>

4.5.4 Coenzyme levels Q10 in brain tissue

Statement of contribution to this project: Brain samples were collected by myself. Brain experiments were performed by Dr Iain Hargreaves and me. Data analysis was performed by myself and supervised by Dr Conceição Bettencourt.

4.5.4.1 Background

The mitochondrial complex I is dysfunctional in PD and PSP and CoQ10 acts as a cofactor in this complex(317,318). There is increasing evidence that impairment of mitochondrial function and oxidative damage are contributing factors to the pathophysiology of PD(319,320) and PSP(321). A reduced level of CoQ10 in cerebral cortex of PD brains has been reported(319). A study of antioxidant markers in PD cases found decreased CoQ10 levels and CoQ10 was proposed as a biomarker of antioxidant status in PD(320). A small study revealed reduced levels of CoQ10 in serum of DLB patients(322).

Many attempts have been made and are underway to assess CoQ10 as a therapeutic approach in these groups of patients.

A clinical trial in 2002 showed promising results with CoQ10 in patients with PD. There was reduced disability especially at higher doses (1200mg)(323), and higher doses of 2400mg

proved to be safe(324). A later trial of symptomatic improvement with CoQ10 in PD in 132 patients was completed in 2007 and showed negative results(325). Finally, a larger trial that recruited 600 patients and tested 1200 mg or 2400 mg of CoQ10 vs placebo, was terminated after 16 months and did not achieve significant benefits(326).

A short trial of CoQ10 in PSP (5mg/kg) vs placebo recruited 20 patients and showed an improved cerebral energy metabolism and a slight but significant improvement in the PSP rating scale and the Frontal Assessment Battery(317). A phase 3 trial of CoQ10 during 12 months in PSP is ongoing (NCT00382824) had negative results.

Finally, by analysing a small sample group, the authors of the paper proposing *COQ2* as a cause of familial MSA and a risk factor of sporadic MSA suggested decreased levels of CoQ10 in brain tissue of MSA cases (n = 3) in comparison with controls (n = 3)(185). In view of this result, we decided to determine the level of CoQ10 in brain tissue from a large cohort of pathologically confirmed MSA cases and compare these levels to those of normal controls as well as to patients with other neurodegenerative movement disorders as comparison groups.

4.5.4.2 Subjects, materials and methods

We analysed pathologically confirmed MSA cases (n=20) and pathologically normal elderly controls (n=37). With the purpose of comparing to other α -synucleinopathies we also analysed DLB cases (n=20) and IPD cases (n=7) and to compare to a tauopathy we used CBD cases (n=15). Finally, to compare to other degenerative diseases that affect the cerebellum we decided to include cerebellar ataxia (CRB_ATX) cases (n=18). A description of the cases is given in Table 4-9. We used flash frozen brain tissue from these cases and we measured CoQ10 in the cerebellum (cerebellar cortex) and frontal cortex (Brodmann areas 8 or 9). Cerebellum was selected as a severely affected region in MSA, particularly in OPCA and mixed pathological subtypes, and frontal cortex as an overall less affected brain region in this disease. Basal ganglia regions, which would be more severely affected in MSA-SND pathological subtype, were not available in most cases and were therefore excluded. All MSA cases included in this study have been sequenced for the entire coding region of the

COQ2 gene, with no *COQ2* variants associated with increased risk of MSA being found in the entire cohort (see chapter 4.5.3).

The method for CoQ10 measurement has been performed as described in the methods section.

Data were analysed with SPSS (v.22). For both cerebellar and frontal cortex tissue, one-way analysis of variance was performed to check for omnibus significant differences in the CoQ10 levels between the main diagnosis groups (MSA, CRB_ATX, CBD, DLB, IPD, and controls). A post hoc Tukey HSD test was then used for pairwise comparisons. To reach a normal distribution of the CoQ10 levels, this variable was log-transformed prior to the abovementioned analyses. These analyses were also performed considering subgroups for MSA (MSA-OPCA, MSA-Mix and MSA-SND) and CRB_ATX (SCA, FRDA and other ataxias). Additionally, multinomial logistic regression was used to infer the magnitude of the association between the outcome (diagnosis) and levels of CoQ10, corrected for potential confounding factors. Controls were used as the reference group, except for comparisons between disease groups only; in this case MSA was taken as the reference. All models were adjusted for age (continuous), gender (binomial), and post-mortem delay (continuous). For all the analyses performed, a p-value <0.05 was considered significant.

4.5.4.3 Results

We measured the CoQ10 levels in brain tissue from the cerebellum and the frontal cortex of MSA, CBD, DLB, IPD, CRB_ATX, and elderly normal controls. Table 4-10 and Figure 4-9 show a summary of the results. The analysis of variance revealed an omnibus significant difference between groups for the CoQ10 levels in the cerebellum [$F(5,111) = 9.434$, $P < 0.001$], but not in the frontal cortex [$F(4,88) = 1.976$, NS]. Post hoc pairwise comparisons revealed that this difference is due to lower mean cerebellar CoQ10 levels in MSA when compared to all other diagnosis groups (Tukey test, $P \leq 0.002$). The levels of CoQ10 were also mildly reduced in CRB_ATX cases compared to controls (Table 4-10 and Figure 4-9), but this did not reach statistical significance.

Table 4-9: Characterization of the samples included in the CoQ10 study. Reproduced from (327) under the Creative Commons Attribution License.

Diagnosis groups	MSA all	MSA-OPCA	MSA-MIX		MSA-SND	CBD	DLB	IPD	CRB_ATX ^a	Contr
Demographic features										
Number of cases	20	9	5		6	15	20	7	18	
Cases with CRBL available	20	9	5		6	15	20	7	18	
Cases with FCTX available	20	9	5		6	0	20	7	18	
Gender (% of female)	70%	56%	80%		83%	40%	25%	14%	50%	
Mean age at death, years (range)	64.55 (51; 74)	64.1 (57; 72)	65.4 (57; 73)		64.5 (51; 74)	69.60 (48; 90)	77.72 (66; 92)	77.86 (65; 84)	59.89 (36; 88)	81.32
Mean post-mortem delay, hours	53.44	44.96	67.86		54.14	53.43	21.23	54.89	33.25	
Unadjusted levels of CoQ10 (pmol/mg)										
Mean CRBL (±SD)	169.30* (±49.71)	150.52** (±29.12)	163.44** (±73.40)		202.33 (±41.79)	271.18 (±76.21)	288.37 (±133.72)	262.47 (±28.84)	233.08 (±46.97)	241.8
Mean FCTX (±SD)	260.44 (±70.22)	264.87 (±75.19)	283.98 (±88.91)		234.17 (±44.12)	-	256.94 (±75.20)	276.02 (±71.37)	330.12 (±96.14)	259.3 (±107)

Ref: MSA = multiple system atrophy; MSA-OPCA = MSA olivopontocerebellar atrophy; MSA-MIX= MSA mixed type; MSA-SND = MSA striatonigral degeneration; CBD = corticobasal degeneration; DLB = dementia with Lewy bodies; IPD = idiopathic Parkinson's disease; CRB_ATX = cerebellar ataxia [includes: SCA = spinocerebellar ataxia (n = 9); FRDA = Friedreich's ataxia (n = 5); other ataxias (miscellaneous) (n = 4)]; CRBL = cerebellar cortex; FCTX = frontal cortex; SD = standard deviation.

Note: *MSA presents significantly lower cerebellar CoQ10 levels when compared to all other diagnosis groups (Tukey test, $p \leq 0.002$); **when dividing by disease subgroups, only MSA-OPCA and MSA-Mix show significantly lower levels than controls, CBD, DLB, FRDA^a, and IPD (Tukey test, $p \leq 0.02$); no other significant differences were observed between groups for CoQ10 levels.

Table 4-10: Multinomial logistic regression estimates for the association between different diagnosis groups and CoQ10 levels in human brain tissue. Note: * and bold highlight significant values ($p < 0.05$). All models were adjusted for age, gender, and post-mortem delay. Reproduced from (327) under the Creative Commons Attribution License.

A) All disease groups versus controls in the Cerebellum tissue		
Diagnosis	OR (95% CI)	p value
MSA	0.97 (0.95-0.99)	0.001*
CBD	1.01 (1.00-1.02)	0.223
DLB	1.01 (1.00-1.02)	0.058
IPD	1.01 (1.00-1.02)	0.405
CRB_ATX	1.00 (0.99-1.01)	0.935
B) MSA and CRB_ATX subdivided into their respective subtypes versus controls in the Cerebellum tissue		
Diagnosis	OR (95% CI)	p value
MSA_SND	0.98 (0.96-1.01)	0.128
MSA_Mixed	0.96 (0.93-0.99)	0.005*
MSA_OPCA	0.95 (0.92-0.98)	0.001*
SCA	0.99 (0.97-1.01)	0.147
FRDA	1.01 (1.00-1.03)	0.062
other_ataxias	1.00 (0.97-1.02)	0.885
C) All other degenerative diseases versus MSA in the Cerebellum tissue		
Diagnosis	OR (95% CI)	p value
CBD	1.04 (1.02-1.06)	<0.001*
DLB	1.04 (1.02-1.06)	<0.001*
IPD	1.04 (1.02-1.06)	<0.001*
CRB_ATX	1.03 (1.01-1.05)	0.001*

D) All disease groups versus controls in the Frontal cortex tissue		
Diagnosis	OR (95% CI)	p value
MSA	1.00 (1.00-1.01)	0.846
DLB	1.00 (1.00-1.01)	0.792
IPD	1.00 (1.00-1.01)	0.543
CRB_ATX	1.01 (1.00-1.01)	0.123

Ref: MSA = multiple system atrophy; MSA_SND = MSA striatonigral degeneration; MSA_mixed = MSA mixed; MSA_OPCA = MSA olivopontocerebellar atrophy; CBD = corticobasal degeneration; DLB = dementia with Lewy bodies; IPD = idiopathic Parkinson's disease; CRB_ATX = cerebellar ataxia; SCA = spinocerebellar ataxia; FRDA = Friedrich's ataxia; other_ataxias = other ataxias of miscellaneous origin; OR = odds ratio; 95% CI = 95% confidence interval.

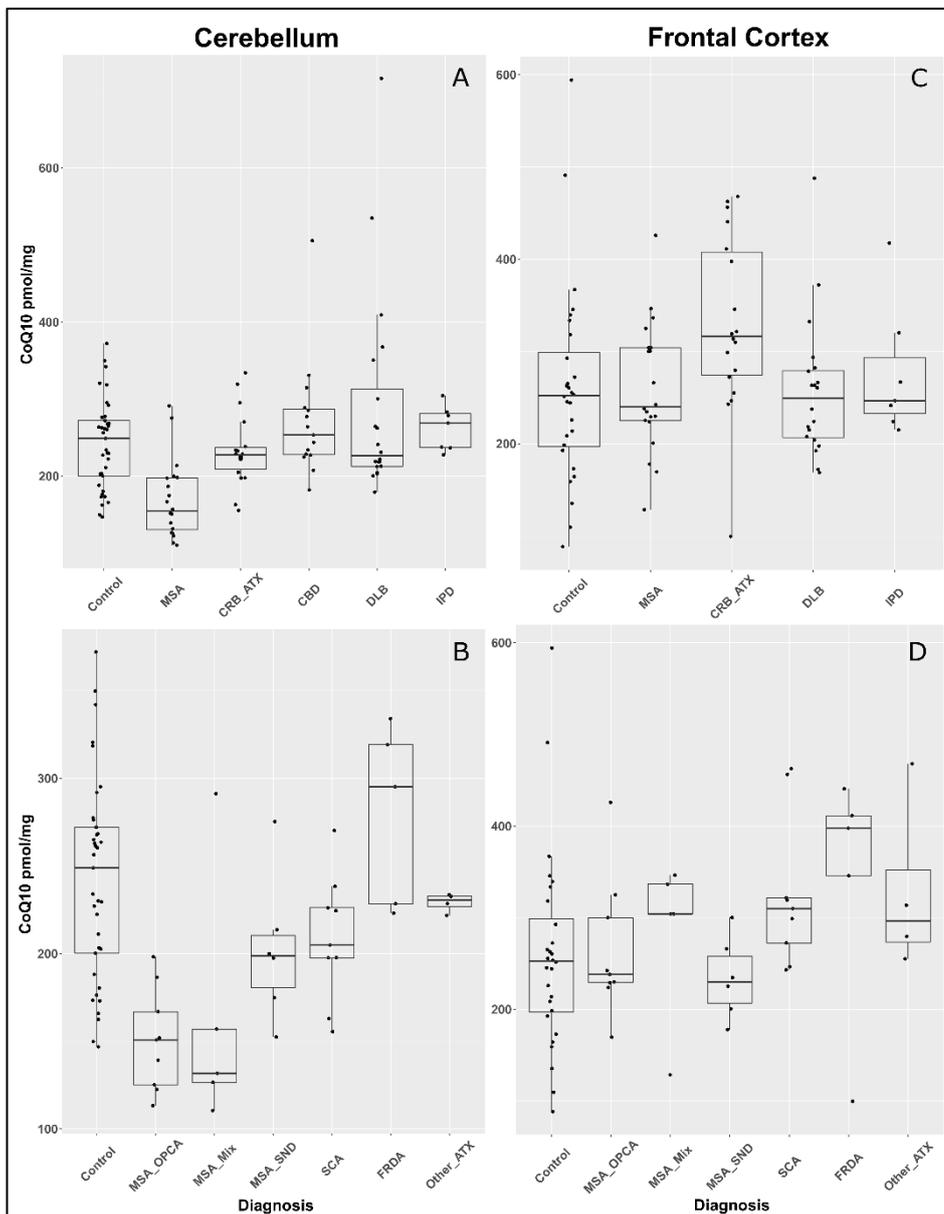


Figure 4-9: A: Boxplot presenting CoQ10 levels in the cerebellum of MSA, CBD, DLB, IPD, CRB_ATX and controls. B: Boxplot presenting CoQ10 levels in the cerebellum of MSA cases subdivided by pathological subtypes MSA_SND, MSA_mixed and MSA_OPCA, and also CRB_ATX cases subdivided into SCA, FRDA and other_ATX. C: Boxplot presenting CoQ10 levels in the frontal cortex of MSA, DLB, IPD, CRB_ATX and controls. D: Boxplot presenting CoQ10 levels in the frontal cortex of MSA cases subdivided by pathological subtypes MSA_SND, MSA_mixed and MSA_OPCA, and also CRB_ATX cases subdivided into SCA, FRDA and other_ATX. Each dot represents one individual, and dots beyond the boxplot whiskers represent outliers. Reproduced from (327) under the Creative Commons Attribution License.

Ref: MSA = multiple system atrophy; MSA_SND = MSA striatonigral degeneration; MSA_mixed = MSA mixed; MSA_OPCA = MSA olivopontocerebellar atrophy; CBD = corticobasal degeneration; DLB = dementia with Lewy bodies; IPD = idiopathic Parkinson's disease; CRB_ATX = cerebellar ataxia; SCA = spinocerebellar ataxia; FRDA = Friedreich's ataxia; other_ATX = other ataxias of miscellaneous origin.

Following adjustment for potential confounders, we found a significant association between the diagnosis and cerebellar CoQ10 levels (Table 4-10), with MSA presenting 3% less CoQ10 than controls (Table 4-10 A; OR=0.97, $p = 0.001$) and less 3-4% than other disease groups (DLB, IPD, CBD, and CRB_ATX; Table 4-10C; $p \leq 0.001$). Given the findings in the cerebellum, we also subdivided MSA samples into the 3 pathological subtypes, and the CRB_ATX cases into spinocerebellar ataxia (SCA), Friedreich's ataxia (FRDA) and miscellaneous, and compared these subgroups to controls (Table 4-9 and Table 4-10 B). Significantly lower levels of CoQ10 in the cerebellar tissue were detected only in OPCA (OR=0.95, $p=0.001$) and the mixed type (OR=0.96, $p=0.005$) MSA cases but not in SND cases when compared to controls (Table 4-10 and Figure 4-9 B). Within the cerebellar diseases (CRB_ATX), even though SCAs presented with the lowest levels of CoQ10, this reduction was still not statistically significant when compared to the control group (Table 4-10 B). In the frontal cortex samples, we found no association between the diagnosis and CoQ10 levels (Table 4-10 D, and Figure 4-9 C).

4.5.4.4 Discussion

In this section of my thesis, we assessed the CoQ10 levels in the cerebellar and frontal cortices from MSA patients and compared these results to elderly controls and samples from CBD, IPD, DLB and CRB_ATX.

The unavailability of basal ganglia tissue as well as the relatively small sample size of some of the disease groups (e.g. IPD) was a limitation of this study for the interpretation of the results. We found, however, significantly decreased levels of CoQ10 in the cerebellar cortex of MSA patients, particularly of the OPCA and mixed pathological subtypes, when compared to controls and all the other disease groups. No differences were detected in the frontal cortex. Although the role of the *COQ2* variants as a cause of MSA is yet to be replicated, the specific and significant decrease of CoQ10 in the cerebellar cortex of the MSA cohort, but not of cerebellar ataxias (CRB_ATX), suggests that a perturbation in the CoQ10 biosynthetic pathway might be involved in the pathogenesis of MSA. In both cerebellar ataxias and MSA-OPCA and mixed Purkinje cell loss will be present. Our results suggest that although CoQ10

reduction may reflect Purkinje cell loss it is likely that other factors are contributing to the observed effect in MSA. More detailed studies would be required to correlate the degree of Purkinje cell depletion with CoQ10 levels and elucidate the cause and specificity of the CoQ10 biosynthesis impairment in MSA.

CoQ10 related pathways have been previously related to neurodegenerative diseases. The mitochondrial respiratory chain complex I, of which CoQ10 is a cofactor(328,329), has been found to be dysfunctional in several neurodegenerative diseases, including IPD and progressive supranuclear palsy (PSP). There is increasing evidence that impairment of mitochondrial function and oxidative damage are contributing factors to the pathophysiology of those diseases (330–332). Furthermore, reduced levels of CoQ10 in cerebral cortex and in lymphocytes of IPD brains have been previously reported(330), and CoQ10 has been also proposed as a biomarker of the antioxidant status in PD(331). A recent study(333), which measured brain energy metabolism in the basal ganglia of clinically diagnosed MSA-P cases, is not in support of mitochondrial dysfunction playing a primary role in the pathophysiology of MSA. Unfortunately, that study did not include comparisons with the MSA-C clinical subtype to help understanding whether this could be relevant to MSA when the cerebellum is the main affected brain region.

The treatment of MSA is limited and purely symptomatic. Whether the reduction in CoQ10 is linked specifically to the aetiology of MSA or is related to the degree of neurodegeneration in the cerebellum of MSA patients is uncertain. However, our results suggest that it may be worth undertaking further studies to evaluate the efficacy of CoQ10 and/or Idebenone in the treatment of MSA given that these quinones are reported to be safe and well tolerated in patients(334,335).

The role of *COQ2* variants in the aetiology of MSA remains debatable. However, our data suggests that a deficiency in cerebellar CoQ10 status may be involved in the pathophysiology of MSA. More work is required before we can elucidate whether this consists of a primary involvement as a cause of MSA or is a secondary finding due to neurodegeneration.

Additionally an independent study replicated my findings by studying 12 MSA, 9 Parkinson disease (PD), 9 essential tremor (ET) patients, and 12 controls. This study revealed CoQ10 deficiency in MSA cerebellum, which was associated with impaired CoQ biosynthesis and increased oxidative stress in the absence of *COQ2* mutations and with a normal mitochondrial mass(336).

5 CHAPTER 5: EXOME SEQUENCING IN MSA

5.1 BACKGROUND

Whole exome sequencing (WES) has proven to be an effective technology when applied to familial studies and also association studies of rare variants in other neurodegenerative disorders(256,257).

More recently, it is also being utilized to investigate the missing heritability in complex diseases and in families that are not large enough for traditional linkage studies. WES is a novel tool useful for the study of rare variants associated with moderate risk of sporadic disease(257,258).

The largest genomic study in MSA has failed to identify significant risk loci for this disease through a GWAS(168). Although still awaiting replication, NGS were used to identify *COQ2* variants associated to Japanese MSA(185). Given that the estimated heritability from common variation in MSA is 2.09-6.65%, and lower to that of ALS, IPD and Alzheimer's disease, we hypothesise that the genetic component might be unravelled through a rare variants approach study. Therefore, we performed the largest exome sequencing study in MSA by first studying an MSA family and later investigating risk factors in sporadic cases.

5.2 WES IN FAMILIAL MSA

Statement of contribution to this project: This project was started by Dr Anna Sailer and I followed up after her. DNA was provided by Professor Wüellner (University of Bonn,

Germany). Experiments were performed by Dr Anna Sailer and Dr Sonja Scholz in the NIH, and I worked on data analysis. Raw data analysis was performed by Dr Vincent Plagnol and downstream analysis by myself.

5.2.1 Subjects, materials and methods

The German MSA family is extremely interesting because it is pathologically confirmed in one individual, and consists of three generations with two affected individuals in two successive generations(337). The family tree is provided in Figure 5-1. Case I-1 presented in her late sixties and had an MSA-P phenotype. Case II-1, presented in her forties with MSA-C. Both cases showed typical MRI findings of MSA. The disease duration was of 16 years in case I-1 and post-mortem brain examination revealed neuropathological changes consistent with a diagnosis of definite MSA. By personal communication, case III-1 presented mild ataxia on examination in her late twenties. She was not further investigated and has not been examined since then. Thus, the significance of this finding remains currently unclear. She is under the typical age of disease onset for MSA.

Of note, genetic testing for SCA 1,2,3,6,7, and 17 was performed with negative results. The *SNCA* gene was investigated for non-synonymous changes as well as CNV and no mutation was detected.

Genomic DNA from 5 members (marked with an * on the pedigree shown on Figure 5-1 of the family) was included in the study. Sequences corresponding to all annotated human exons were enriched by hybridization using the NimbleGen kit (Roche NimbleGen, Madison, WI, USA) and sequenced on the Illumina Genome Analyser II and HiSeq 2000. The DNA samples were sequenced on a paired-end 50 base-pair run.

Raw sequencing reads were aligned to the UCSC hg19 build of the reference genome using the software novoalign. Calling was performed using samtools 0.18 and the resulting calls were annotated using ANNOVAR.

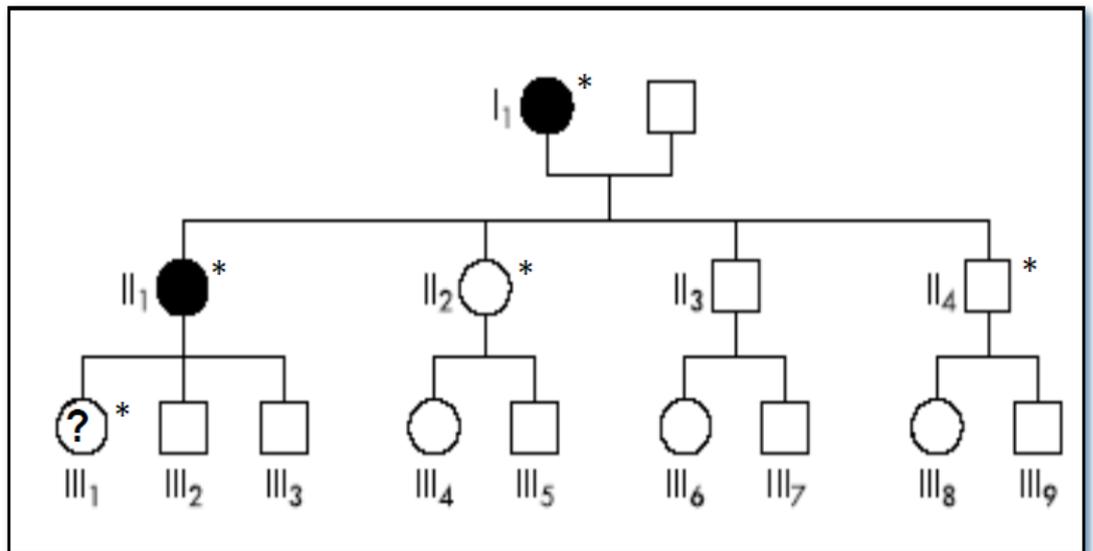


Figure 5-1: Pedigree of the German MSA family. Modified from (316)

5.2.2 Results

Summary metrics on the German MSA kindred are shown in Table 5-1. Unfiltered analysis of known genes causative of PD and cerebellar ataxias were first investigated and no pathogenic mutations were found.

We have provided different strategies of filtering within this family. The numbers of variants are presented in Table 5-2Table 5-1 and the variants are listed in Table 5-3, Table 5-4, Table 5-5, Table 5-6 and Table 5-7. We focused on rare variants (MAF <5%) and variants that had good quality and depth with a heterozygous call because this was based on interpreting the family as presenting autosomal dominant inheritance.

Of note we first looked at shared variants between affected cases because there could be reduced penetrance. However, as the number of variants was still too large, we later decided to filter excluding variants carried by unaffected family members.

Table 5-1: Summary metrics of WES in the MSA family.

Patient	Affected	Target size (bp)	Total number of unique reads	Mean target coverage	Total number of variants	Heterozygous, unreported, depth >20, quality >30. Excluding

						synonymous changes
I1	Yes	33989968	156383446	88.461946	15400	131
II1	Yes	33989968	102246602	67.868065	15095	121
II2	No	33989968	111339445	75.489936	15212	145
II4	No	33989968	124010439	63.061651	15147	124
III1	Unknown	33989968	153925228	89.268034	15724	141

Ref: Target size = unique number of target bases in the experiment; total number of unique reads = number of reads that are not marked as duplicates after alignment; mean target coverage = mean coverage of targets that received at least coverage depth = 2 at one base.

Table 5-2: Different filtering strategies in the German MSA family.

Present in	I1 and II1	I1 and II1	I1, II1 and III1	I1, II1 and III1	I1 and II1
Absent in		II2 and II4		II2 and II4	II2, II4 and III1
Number of variants	68	10	32	2	8

Table 5-3: Variants present in I1, II1 and III3.

	Gene	Full gene name	Variant
1	<i>NADK</i>	NAD kinase	uc010nyv.1:c.1238_1240del:p.413_414del
2	<i>PEG3</i>	paternally expressed 3	uc002qnt.2:c.C402G:p.N134K
3	<i>MUC6</i>	mucin 6, oligomeric mucus/gel-forming	uc001lsw.2:c.G6136A:p.V2046I
4	<i>RXFP2</i>	relaxin/insulin-like family peptide receptor 2	uc010aba.2:c.C1363T:p.R455C
5	<i>C14orf159</i>	chromosome 14 open reading frame 159	uc001xyw.2:c.C1015G:p.L339V
6	<i>ATP11B</i>	ATPase, class VI, type 11B	uc003fla.2:c.C40T:p.P14S
7	<i>LEPREL1</i>	leprecan-like 1	uc003fsg.2:c.C1363T:p.P455S
8	<i>OR2B3</i>	olfactory receptor, family 2, subfamily B, member 3	uc003nlx.2:c.A97G:p.I33V
9	<i>EGFL8</i>	EGF-like-domain, multiple 8	uc003oab.1:c.G34A:p.G12S

10	<i>NOL9</i>	nucleolar protein 9	uc001ans.2:c.G594C:p.L198F
11	<i>THAP11</i>	THAP domain containing 11	uc002euo.2:c.372_373insCAA:p.Q124del sQQ
12	<i>KLK3</i>	kallikrein-related peptidase 3	uc002pts.1:c.G260A:p.G87D
13	<i>TTC30B</i>	tetratricopeptide repeat domain 30B	uc002uln.2:c.G710A:p.R237H
14	<i>CRYBB3</i>	crystallin, beta B3	uc003abo.1:c.G455A:p.R152H
15	<i>PRCC</i>	papillary renal cell carcinoma (translocation-associated)	uc001fqa.2:c.C1135G:p.P379A
16	<i>CD1A</i>	CD1a molecule	uc001frt.2:c.C553T:p.R185C
17	<i>OR4C3</i>	olfactory receptor, family 4, subfamily C, member 3	uc010rhv.1:c.169_177del:p.57_59del
18	<i>HNF1A</i>	HNF1 homeobox A	NA
19	<i>PABPC3</i>	poly(A) binding protein, cytoplasmic 3	uc001upy.2:c.878_879insACTAGGG:p.V29 3fs
20	<i>TSC22D1</i>	TSC22 domain family, member 1	uc001uzn.3:c.C182G:p.P61R
21	<i>OR4N2</i>	olfactory receptor, family 4, subfamily N, member 2	uc001yuf.2:c.208_209insT:p.L70fs
22	<i>OR4N4</i>	olfactory receptor, family 4, subfamily N, member 4	uc010tzv.1:c.448_449insT:p.L150fs
23	<i>MYH11</i>	myosin, heavy chain 11, smooth muscle	uc002ddw.2:c.G3806A:p.S1269N
24	<i>JMJD5</i>	jumonji domain containing 5	uc010bxw.2:c.G517T:p.D173Y
25	<i>HIRIP3</i>	HIRA interacting protein 3	uc002dve.2:c.G756C:p.E252D
26	<i>ZNF423</i>	zinc finger protein 423	uc010vgn.1:c.G2308A:p.G770S
27	<i>HIF3A</i>	hypoxia inducible factor 3, α subunit	NA
28	<i>SEC14L3</i>	SEC14-like 3 (<i>S. cerevisiae</i>)	uc003ahy.2:c.A1118G:p.H373R

29	<i>KLHDC8B</i>	kelch domain containing 8B	uc003cwh.2:c.G1052A:p.R351H
30	<i>FLJ00114</i>	NA	uc010vfk.1:c.T626A:p.V209D
31	<i>ESPNL</i>	NA	uc002vxq.3:exon6:c.1102+1G>T
32	<i>DKFZ</i>	NA	p586D0922

Table 5-4: Variants present in I1 and II1 and absent in II2 and II4.

	Gene	Full gene name	Variant
1	<i>URB2</i>	URB2 ribosome biogenesis 2 homolog (<i>S. cerevisiae</i>)	uc001hts.1:c.A266G:p.N89S
2	<i>MBTPS1</i>	membrane-bound transcription factor peptidase, site 1	uc002fhh.2:c.C1274T:p.P425L
3	<i>NOTCH3</i>	notch 3	uc002nao.1:c.A3196T:p.N1066Y
4	<i>C19orf44</i>	chromosome 19 open reading frame 44	uc002nef.1:c.T1405A:p.S469T
5	<i>PEG3</i>	paternally expressed 3	uc002qnt.2:c.C402G:p.N134K
6	<i>STK32B</i>	serine/threonine kinase 32B	uc003gih.1:c.A151G:p.M51V
7	<i>EVC2</i>	Ellis van Creveld syndrome 2	uc003gij.2:c.G2248A:p.E750K
8	<i>REV3L</i>	REV3-like, catalytic subunit of DNA polymerase zeta (yeast)	uc003puy.3:c.C3092T:p.P1031L
9	<i>APAF1</i>	apoptotic peptidase activating factor 1	uc009zto.2:c.C308G:p.T103S
10	<i>NADK</i>	NAD kinase	uc010nyv.1:c.1238_1240del:p.413_414del

Table 5-5: Variants present in I1 and II1.

	Gene	Full gene name	Variant
1	<i>NADK</i>	NAD kinase	uc010nyv.1:c.1238_1240del:p.413_414del

2	<i>URB2</i>	URB2 ribosome biogenesis 2 homolog (<i>S. cerevisiae</i>)	uc001hts.1:c.A266G:p.N89S
3	<i>APAF1</i>	apoptotic peptidase activating factor 1	uc009zto.2:c.C308G:p.T103S
4	<i>MBTPS1</i>	membrane-bound transcription factor peptidase, site 1	uc002fhh.2:c.C1274T:p.P425L
5	<i>NOTCH3</i>	notch 3	uc002nao.1:c.A3196T:p.N1066Y
6	<i>C19orf44</i>	chromosome 19 open reading frame 44	uc002nef.1:c.T1405A:p.S469T
7	<i>PEG3</i>	paternally expressed 3	uc002qnt.2:c.C402G:p.N134K
8	<i>STK32B</i>	serine/threonine kinase 32B	uc003gih.1:c.A151G:p.M51V
9	<i>EVC2</i>	Ellis van Creveld syndrome 2	uc003gij.2:c.G2248A:p.E750K
10	<i>REV3L</i>	REV3-like, catalytic subunit of DNA polymerase zeta (yeast)	uc003puy.3:c.C3092T:p.P1031L
11	<i>NRD1</i>	nardilysin (N-arginine dibasic convertase)	uc001ctd.3:c.A2869G:p.T957A
12	<i>C8B</i>	complement component 8, beta polypeptide	uc001cyp.2:c.G164A:p.S55N
13	<i>MUC6</i>	mucin 6, oligomeric mucus/gel-forming	uc001lsw.2:c.G6136A:p.V2046I
14	<i>RXFP2</i>	relaxin/insulin-like family peptide receptor 2	uc010aba.2:c.C1363T:p.R455C
15	<i>MMP14</i>	matrix metalloproteinase 14 (membrane-inserted)	uc001whc.2:c.G446A:p.R149H
16	<i>CDH24</i>	cadherin 24, type 2	uc001wil.2:c.G1052A:p.R351Q
17	<i>C14orf21</i>	chromosome 14 open reading frame 21	uc001wom.1:c.A128T:p.H43L
18	<i>AKAP6</i>	A kinase (PRKA) anchor protein 6	uc001wrq.2:c.C827T:p.T276M
19	<i>C14orf159</i>	chromosome 14 open reading frame 159	uc001xyw.2:c.C1015G:p.L339V
20	<i>BLMH</i>	bleomycin hydrolase	uc010wbn.1:c.C445T:p.R149X
21	<i>CNTN6</i>	contactin 6	uc011asj.1:c.A1369C:p.I457L
22	<i>ATP11B</i>	ATPase, class VI, type 11B	uc003fla.2:c.C40T:p.P14S

23	<i>LEPREL1</i>	leprecan-like 1	uc003fsg.2:c.C1363T:p.P455S
24	<i>RGS12</i>	regulator of G-protein signaling 12	uc010icv.2:c.C233T:p.S78F
25	<i>OR2B3</i>	olfactory receptor, family 2, subfamily B, member 3	uc003nlx.2:c.A97G:p.I33V
26	<i>EGFL8</i>	EGF-like-domain, multiple 8	uc003oab.1:c.G34A:p.G12S
27	<i>DOCK8</i>	dedicator of cytokinesis 8	uc003zgz.2:c.A391G:p.I131V
28	<i>NOL9</i>	nucleolar protein 9	uc001ans.2:c.G594C:p.L198F
29	<i>TDRD5</i>	tudor domain containing 5	uc001gnf.1:c.A727G:p.T243A
30	<i>DUSP10</i>	dual specificity phosphatase 10	uc001hmy.1:c.A628G:p.I210V
31	<i>PSEN2</i>	presenilin 2 (Alzheimer disease 4)	uc009xeo.1:c.C205G:p.P69A
32	<i>ASCC1</i>	activating signal cointegrator 1 complex subunit 1	uc001jsr.1:c.G65A:p.G22D
33	<i>THAP11</i>	THAP domain containing 11	uc002euo.2:c.372_373insCAA:p.Q124delinsQQ
34	<i>ADAMTS18</i>	ADAM metalloproteinase with thrombospondin type 1 motif, 18	uc002ffe.1:c.C32T:p.T11I
35	<i>KLK3</i>	kallikrein-related peptidase 3	uc002pts.1:c.G260A:p.G87D
36	<i>TTC30B</i>	tetratricopeptide repeat domain 30B	uc002uln.2:c.G710A:p.R237H
37	<i>XKR7</i>	XK, Kell blood group complex subunit-related family, member 7	uc002wxe.2:c.C233G:p.A78G
38	<i>CRYBB3</i>	crystallin, beta B3	uc003abo.1:c.G455A:p.R152H
39	<i>UNC13B</i>	unc-13 homolog B (C. elegans)	uc003zww.2:c.C115T:p.R39C
40	<i>LANCL3</i>	LanC lantibiotic synthetase component C-like 3 (bacterial)	uc004ddp.1:c.A680G:p.Y227C
41	<i>AWAT1</i>	acyl-CoA wax alcohol acyltransferase 1	uc004dxy.2:c.C647T:p.P216L

42	<i>PHKA1</i>	phosphorylase kinase, α 1 (muscle)	uc010nll.2:c.G607A:p.A203T
43	<i>DNAJC6</i>	DnaJ (Hsp40) homolog, subfamily C, member 6	uc001dcd.1:c.A2486G:p.K829R
44	<i>OLFM3</i>	olfactomedin 3	uc001duf.2:c.C689T:p.T230I
45	<i>PRCC</i>	papillary renal cell carcinoma (translocation-associated)	uc001fqa.2:c.C1135G:p.P379A
46	<i>CD1A</i>	CD1a molecule	uc001frt.2:c.C553T:p.R185C
47	<i>OR4C3</i>	olfactory receptor, family 4, subfamily C, member 3	uc010rhv.1:c.169_177del:p.57_59del
48	<i>HNF1A</i>	HNF1 homeobox A	NA
49	<i>PABPC3</i>	poly(A) binding protein, cytoplasmic 3	uc001upy.2:c.878_879insACTAGG G:p.V293fs
50	<i>TSC22D1</i>	TSC22 domain family, member 1	uc001uzn.3:c.C182G:p.P61R
51	<i>ANKRD10</i>	ankyrin repeat domain 10	uc001vrn.2:c.C592T:p.H198Y
52	<i>OR4N2</i>	olfactory receptor, family 4, subfamily N, member 2	uc001yuf.2:c.208_209insT:p.L70fs
53	<i>OR4N4</i>	olfactory receptor, family 4, subfamily N, member 4	uc010tzv.1:c.448_449insT:p.L150fs
54	<i>TP53BP1</i>	tumor protein p53 binding protein 1	uc010udp.1:c.A3002G:p.E1001G
55	<i>MYH11</i>	myosin, heavy chain 11, smooth muscle	uc002ddw.2:c.G3806A:p.S1269N
56	<i>JMJD5</i>	jumonji domain containing 5	uc010bxw.2:c.G517T:p.D173Y
57	<i>HIRIP3</i>	HIRA interacting protein 3	uc002dve.2:c.G756C:p.E252D
58	<i>ZNF423</i>	zinc finger protein 423	uc010vgn.1:c.G2308A:p.G770S
59	<i>HIF3A</i>	hypoxia inducible factor 3, α subunit	NA
60	<i>SEC14L3</i>	SEC14-like 3 (<i>S. cerevisiae</i>)	uc003ahy.2:c.A1118G:p.H373R

61	<i>KLHDC8B</i>	kelch domain containing 8B	uc003cwh.2:c.G1052A:p.R351H
62	<i>ACSL1</i>	acyl-CoA synthetase long-chain family member 1	uc011ckn.1:c.G608A:p.R203Q
63	<i>VPS13A</i>	vacuolar protein sorting 13 homolog A (<i>S. cerevisiae</i>)	uc004akq.3:c.9237_9251del:p.3079_3084del
64	<i>IARS</i>	isoleucyl-tRNA synthetase	uc010mqt.2:c.C1247T:p.T416I
65	<i>LOC401308</i>	NA	uc009zbd.1:c.1174_1175insACTCG:p.L392fs
66	<i>FLJ00114</i>	NA	uc010vfk.1:c.T626A:p.V209D
67	<i>ESPNL</i>	NA	uc002vxq.3:exon6:c.1102+1G>T
68	<i>DKFZ</i>	NA	p586D0922

Table 5-6: Variants present in I1, II1 and III1 and absent in II2 and II4.

	Gene	Full gene name	Variant
1	<i>PEG3</i>	paternally expressed 3	uc002qnt.2:c.C402G:p.N134K
2	<i>NADK</i>	NAD kinase	uc010nyv.1:c.1238_1240del:p.413_414del

Table 5-7: Variants present in I1 and II1 and absent in II2, II4 and III1.

	Gene	Full gene name	Variant
1	<i>URB2</i>	URB2 ribosome biogenesis 2 homolog (<i>S. cerevisiae</i>)	uc001hts.1:c.A266G:p.N89S
2	<i>MBTPS1</i>	membrane-bound transcription factor peptidase, site 1	uc002fhh.2:c.C1274T:p.P425L
3	<i>NOTCH3</i>	notch 3	uc002nao.1:c.A3196T:p.N1066Y
4	<i>C19orf44</i>	chromosome 19 open reading frame 44	uc002nef.1:c.T1405A:p.S469T

5	<i>STK32B</i>	serine/threonine kinase 32B	uc003gih.1:c.A151G:p.M51V
6	<i>EVC2</i>	Ellis van Creveld syndrome 2	uc003gij.2:c.G2248A:p.E750K
7	<i>REV3L</i>	REV3-like, catalytic subunit of DNA polymerase zeta (yeast)	uc003puy.3:c.C3092T:p.P1031L
8	<i>APAF1</i>	apoptotic peptidase activating factor 1	uc009zto.2:c.C308G:p.T103S

5.2.3 Discussion

At a first glance, there are variants in 2 known genes that cause other neurological diseases and these are *PSEN2* and *NOTCH3* and are linked to AD and CADASIL (Cerebral-Autosomal-Dominant-Arteriopathy-with-Subcortical-Infarcts-and-Leukoencephalopathy) respectively. The variant in *PSEN2* is present in dbSNP as rs202133351 and has an allele frequency of 9.078×10^{-05} in Exac (11 het counts of 121,176 alleles) and it is reported as of uncertain significance in ClinVar. This variant needs to be followed up and it is possible that it is rare polymorphism. The variant in *NOTCH3* is in exon 20 (rs376950447) has a frequency in Exac of 5.793×10^{-05} (7 of 120,840 alleles) and also most of CADASIL causing mutations are in exons 2 to 6. These variants are not present in case III-1 so, if this case was affected it would be unlikely to be linked to MSA in this family. Although a mutation in exon 22 of *NOTCH3* has been reported in a case with dementia(338) and it is possible that variants that code for other regions of this protein can be causing different phenotypes, the frequency of this variant in Exac is probably too common to be pathogenic.

One stop-gain variant shared by 2 affected individuals in the *BLMH* gene was detected. This gene encodes for bleomycin hydrolase that is a cytoplasmic cysteine peptidase that is highly conserved through evolution; however, the only known activity of the enzyme is metabolic inactivation of the glycopeptide bleomycin, an essential component of combination chemotherapy regimens for cancer. An association of a polymorphism in *BLMH* and AD has been found in 1996(339) but could not be replicated. Finally, the stop-gain variant in *BLMH* we detected is present also in an unaffected case from our MSA family and not present in our possibly-affected case; so, we conclude that although it would be a functionally interesting variant it is probably not linked to MSA in this family.

Unfortunately, this family is small and we continue to wait for clinical information on case III-1 and therefore this is where the analysis stands so far.

5.3 WES IN SPORADIC MSA

Statement of contribution: Pathologically confirmed samples were collected, sectioned and organized by myself. DNA extractions were performed in part by me and also on a commercial basis at LGC Genomics. The lab work on the pathologically confirmed samples was performed in part in the ION, UCL and in part at the NIH. The experiments were performed by Debbie Hughes, colleagues from the NIH lab and myself.

The clinical samples were collected and prepared by myself. Half of the cohort was from Queen Square and half were shared with us by our French collaborators (Dr Wassillios Meissner and colleagues). Sequencing experiments were performed by Dr Monica Federoff in the NIH.

Data analysis presented in this thesis was performed by Dr Alan Pittmann, Dr Jana Vandrovcova and me.

5.3.1 Subjects, materials and methods

5.3.1.1 Samples

The total number of samples where we were able to obtain exome sequencing data was 1109 (Table 5-8). Six-hundred fifty-nine were controls and 450 were MSA cases. Obtaining this large cohort of pathologically confirmed MSA cases and well-characterised clinically diagnosed cases was possible thanks to an enormous effort from our team and from collaborators in other centres. A complete list of collaborators is available upon request.

Table 5-8: Origin of MSA samples included in this exome sequencing study. Ref: IOP=Institute of Psychiatry; QSBB=Queen Square Brain Bank; NHNN=National Hospital for Neurology and Neurosurgery. A list of collaborators from each centre is available upon request.

Brain bank of origin	Number of samples
Barcelona	13
Bordeaux	3
Emory	2
Harvard	2
Imperial College	3
IOP Kings College	8
Manchester	2
Mayo clinic Jacksonville	41
Miami	10
Munich	17
Netherlands	11
Newcastle	6
Paris	9
QSBB	138
Toronto	3
Upenn	28
Wurzburg	2
Total definite MSA samples	298

Clinical centre of origin	Number of samples
French reference centre for MSA	121
NHNN	31
Total clinical MSA samples	152

Most of the control subjects were sequenced and data kindly shared with us by Dr Andy Singleton and his team from the NIH. The remaining controls were central UCL control data provided by Dr Alan Pitman.

5.3.1.2 WES chemistries and initial steps for analysis.

This project took many years to complete. Therefore, the chemistries used for exome sequencing varied in time. Sequencing methods are described in chapter 3.3.7.

For the first batch of samples we used Tru Seq custom amplicon from Illumina, as well as for most of the controls. The latest batch of samples was sequenced with Nextera focussed

capture. During downstream analysis, we performed different quality control steps to ensure maximal reliability of the final data. In particular, the Nextera focussed exome does not contain as much out of target data as Tru seq (for example 5'UTRs) and we therefore selected the output from nextera genomic coordinates from the bed files to select the final variants for analysis.

The methods for the initial steps for raw data analysis are described in chapter 3.3.9.

5.3.2 Quality control

5.3.2.1.1 Steps for Variant and Sample QC

The first steps of data alignment, merging files and genotype QC have been described in Chapter 3.3.9 and listed in Figure 5-2

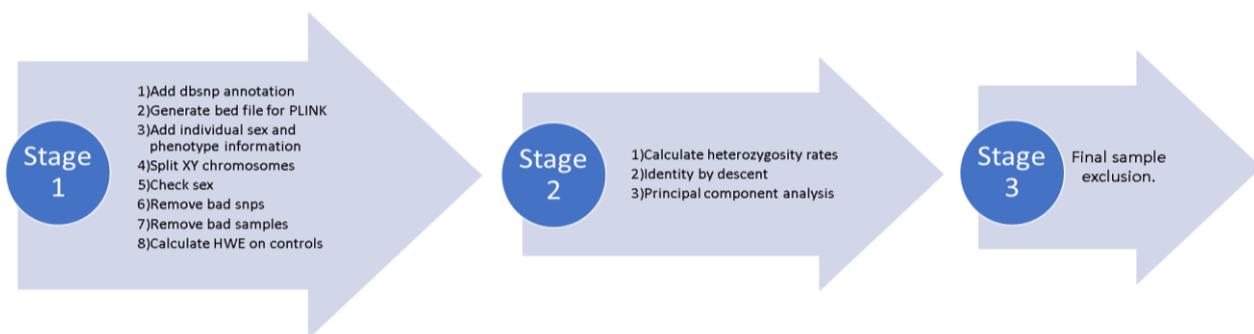


Figure 5-2: Quality control steps performed for this association study in MSA.

In order to perform sample and variant QC we started from a concatenated VCF file of all cases and controls which only contains ID and genotypes. We then annotated the file adding dbSNP annotations onto the VCF file. This was done on GATK. This file now contained genomic positions of variants. Then, we generate a binary (.bed) file for PLINK input. After that we add individual sex and phenotype information with a .ped file. We provide PLINK with 2 files: "MSA_SexInfo.txt" and "MSA_pheno.txt". This will allow sex check and HWE testing the subsequent steps.

Splitting XY chromosomes in PLINK is performed to avoid haploid calls in the X chromosome pseudo-autosomal region of male samples that will otherwise create errors.

Sex check is an important step where we compare the ascertained sex against the genotype sex. If there is a mismatch we have to exclude those samples. Checking sex in PLINK is performed by analysing the proportion of homozygosity of the X chromosome. The output gives us F values. An F estimate smaller than 0.2 yields female calls, and values larger than 0.8 yields male calls. We were a little less stringent because PLINK is designed for array data as opposed to sequencing data and used the parameters 0.3 and 0.7 as limits. The results are shown on Figure 5-3.

SNPs that have a call rate less than 90% are interpreted as “bad SNPs” according to the generated F_MISS score in PLINK. The information from this step is visualised by the researcher. If there are still samples that present high missingness rates even after removal of the bad SNPs (F_MISS) they will have to be subsequently removed. The steps of missingness check by SNP as well as by sample are illustrated in Figure 5-4, Figure 5-5 and Figure 5-6.

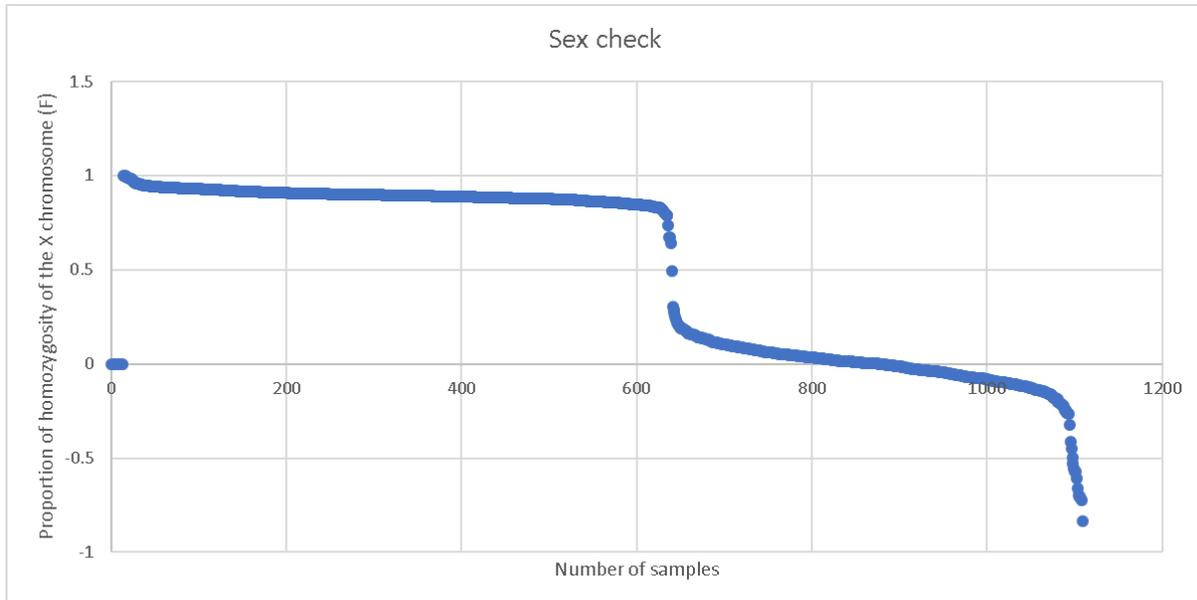


Figure 5-3: Scatterplot exhibiting the results of gender calculations in the MSA cases and controls. Samples in the intersection between X and Y axis correspond to 13 samples with missing genotypes and therefore no data for sex check. They were removed. All samples between 0.3 and 0.7 (ambiguous sex) and all samples with discordant sex information were removed.



Figure 5-4: Scatterplot showing missingness scores before removing SNPs with low call rates. Samples over 0.2 cut off = 98.

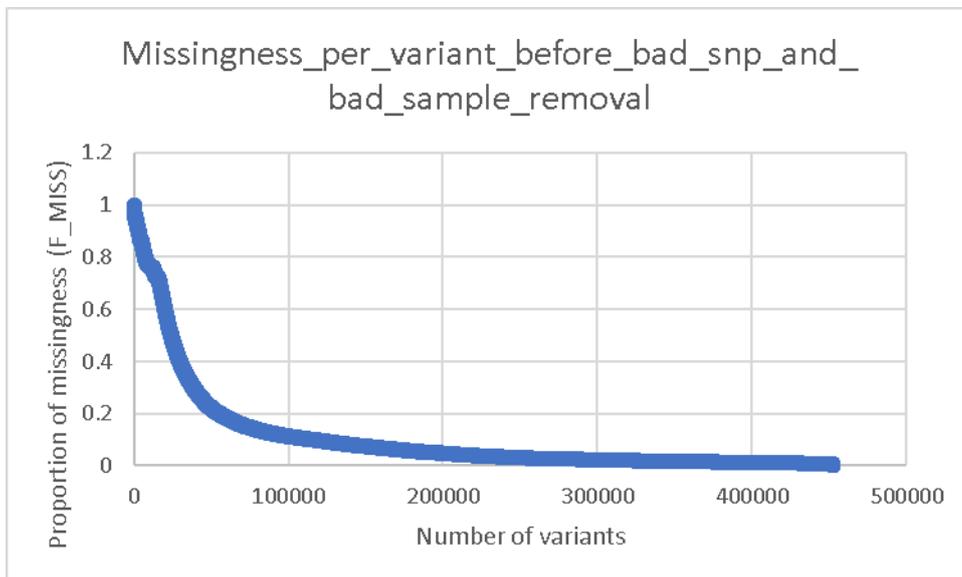


Figure 5-5: Scatterplot showing the distribution of missingness per variant. Variants with a score under 0.1 were removed.

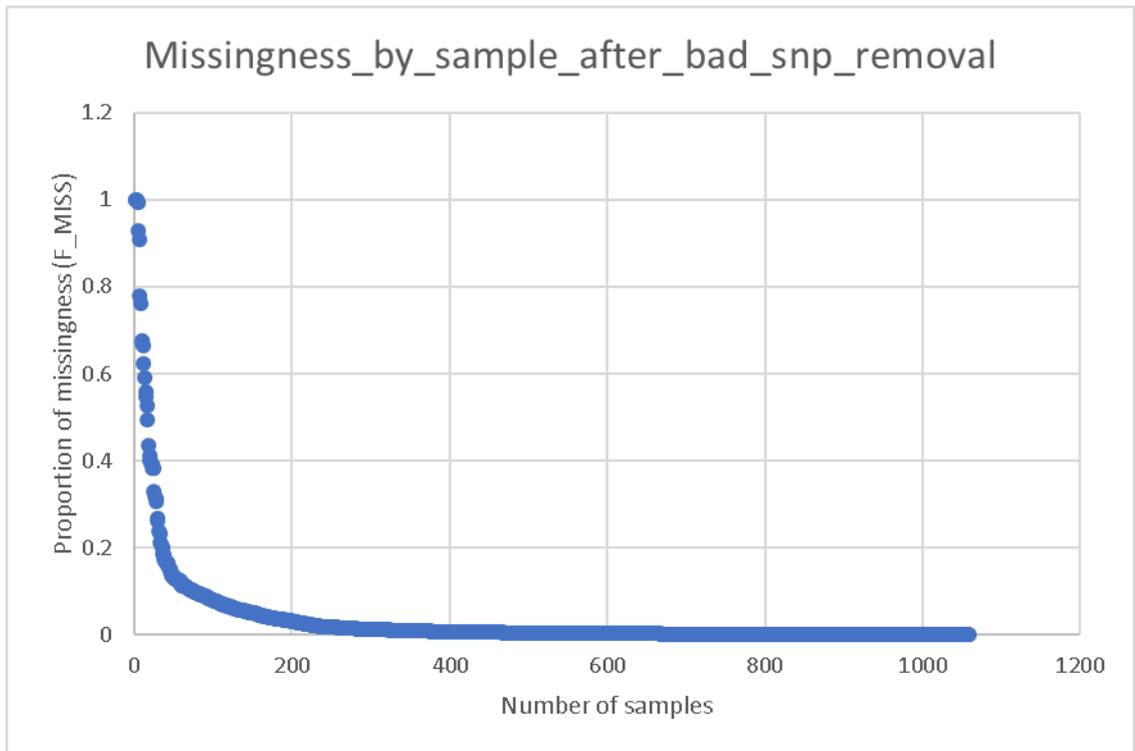


Figure 5-6: Scatterplot presenting sample missingness after removing SNPs with a call rate under 90%. Samples over 0.2 cut-off = 34.

The number of variants excluded in the different steps of variant QC and the number of samples excluded in the steps of sample QC are presented in Table 5-9 and Table 5-10.

Table 5-9: Table showing the number of variants excluded by SNP QC steps in the MSA association study.

QC step	Number of variants
	452168
Exclude SNPs with <90% call rate	116527
HWE	940
Total variants	334701

Table 5-10: Table showing the number of samples excluded in the different sample QC steps of the MSA association study.

QC step	Total samples	MSA cases	Controls	Female	Male	Ambiguous	Number of samples to exclude
Sex check	1109	450	659	460	648	1	51
Exclude samples with >20% missingness	1059	422	637	443	615	1	34
Total samples	1025	402	623	429	596		

The following steps of quality control where the calculations of heterozygosity rates, Identity by descent (IBD), Identity by state (IBS), and principal component analysis (PCA) as described in chapter 3.3.9. The results of these steps are illustrated in Figure 5-7, Figure 5-8, Figure 5-9 and Figure 5-10. The thresholds used were the following: heterozygosity rate (>0.05), IBD (<0.24), IBS (<1), PCA ($x < 0$; $y > 0$).

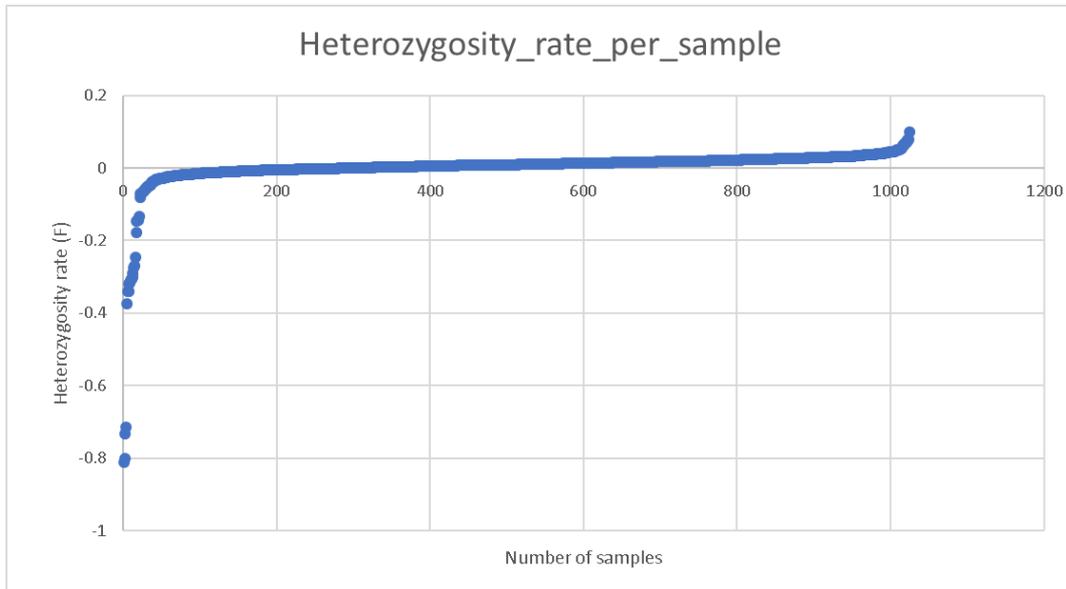


Figure 5-7: Scatterplot showing the heterozygosity rate per sample.

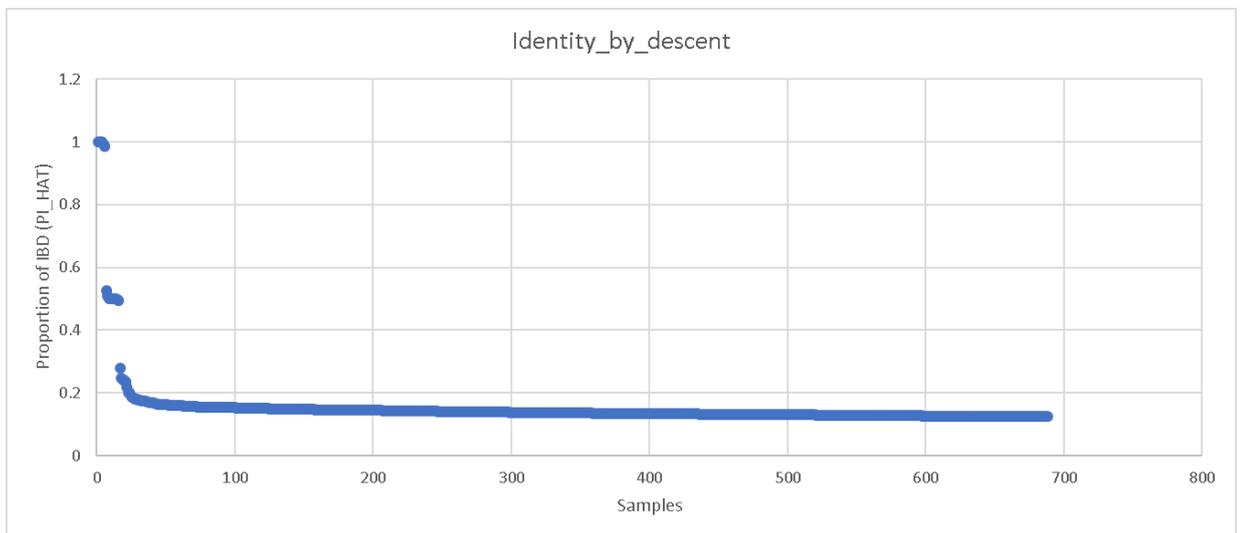


Figure 5-8: Scatterplot of identity by descent score. A score of 1 or close to 1 is probably due to sample duplicates.

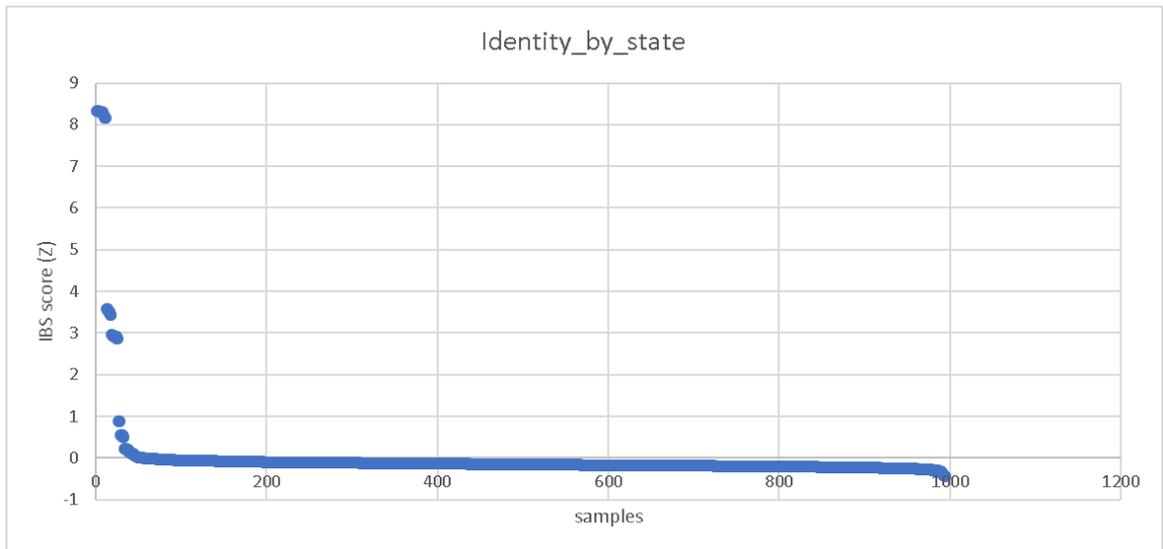


Figure 5-9: Scatterplot of identity by state score.

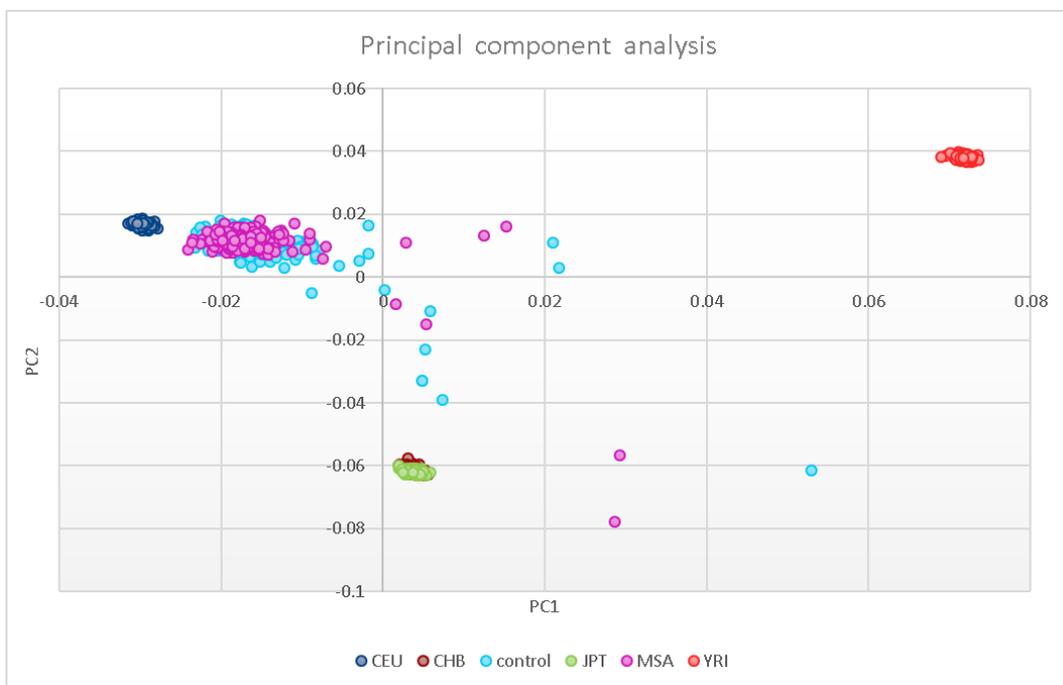


Figure 5-10: Principal component analysis plot (PCA) presenting population stratification in the dataset.

Ref: The HapMap genotype data contains individual genotypes from Europe (CEU=Utah residents with ancestry from northern and western Europe), Asia (CHB=Chinese in Beijing; JPT=Japanese in Tokyo), and Africa (YRI=Yoruba in Ibadan, Nigeria). Ref: PC1 = principal component 1, PC2 = principal component 2, MSA = Multiple system atrophy.

Table 5-11: Table showing number of samples excluded by sample QC steps.

Sample QC step	samples	Cases	controls	outliers
Heterozygosity rate	1025	402	623	30
IBD	995	390	605	32
IBS	995	390	605	52
PCA	995	390	605	131

Ref: IBD=identity by descent; IBS=identity by state; PCA=principal component analysis.

All samples with sexcheck problems presented also PCA problems. One sample had very bad quality and failed for sex, missingness and PCA. All samples with high missingness are PCA flagged. And all samples exhibiting high heterozygosity are flagged on PCA. As expected, all outliers on IBS and outliers on IBD but not the other way around. Two samples flagged on IBS and IBD are also PCA flagged.

The final file after QC contains 953 people (563 males, 390 females) of which 375 were MSA cases (113 clinical and 168 path) and 578 were controls. The clinical information regarding sex and subtype of MSA (where available) is presented in Table 5-12 and Table 5-13.

Table 5-12: Sex information on the MSA samples included in the case-control association study.

	Definite MSA	Clinical MSA	Total
Male	93	60	153
Female	75	53	128
Total	168	113	

Table 5-13: MSA subtype of the samples included in the MSA case-control association study.

	Number of samples
MSA-C	73
MSA-P	118

Ref: MSA-C: multiple system atrophy cerebellar subtype; MSA-P: multiple system atrophy parkinsonian subtype.

5.3.3 Candidate genes results

Candidate genes studied were performed looking for possible known pathogenic alterations linked to MSA or related disorders as well as comparing findings to controls for single association and sequencing artefacts detection.

The list of genes investigated was based on previous associations with MSA(44) and interesting genes reported as pathogenic and/or risk factors of PD and DLB(24,340). A list of the genes analysed can be found in Table 5-14. Scripts used for this analysis can be found in the Appendix.

Table 5-14: List of candidate genes studied in the MSA exome sequencing project.

<i>ACMSD</i>	<i>HLA-DRB5</i>
<i>APOE</i>	<i>LRKK2</i>
<i>APP</i>	<i>MAPT</i>
<i>ATP13A2</i>	<i>MCCC1/LAMP3</i>
<i>BST1</i>	<i>PARKIN/PARK2</i>
<i>CCDC62/HIP1R</i>	<i>PINK1</i>
<i>CHCHD2</i>	<i>PLA2G6</i>
<i>COQ2</i>	<i>PSEN1</i>
<i>DJ1</i>	<i>RAB39B</i>
<i>DNAJC6</i>	<i>SCARB2</i>
<i>EIF4G1</i>	<i>SNCA</i>
<i>FBXO7</i>	<i>STK39</i>
<i>GAK</i>	<i>SYT11</i>
<i>GBA</i>	<i>VPS35</i>
<i>GCH1</i>	

For MSA related genes (*COQ2* and *SNCA*), we analysed all variants present in our cohort of patients including synonymous changes and putatively pathogenic variants. They were all searched for in databases for potentially known pathogenic variants and also tested for single association.

There were 20 variants in *COQ2*. From those, 17 were exonic. One was a stop-gain, 11 non-synonymous changes and 5 were synonymous changes. I will mention here the relevant variants:

1. The stop-gain variant p.R22X (rs112033303) that was previously reported by us with a protective effect on MSA(341) was detected in both MSA and controls with no significant difference in this study. The MAF in MSA was 0.021 and in controls 0.019. The allele count in MSA was 16 and in controls 22 and there were 2 MSA cases homozygous for this change. Very common in exac (with 10 homozygotes out of 23378)

2. The variant p.V393A in *COQ2* (rs148156462) that has been proposed to increase risk of MSA in Japanese(342) was found in one healthy control only.

In *SNCA* we found 2 variants. One synonymous change and one non-synonymous change. Both (rs138969470, p.L100L; rs200056149, p.E123K) were present in one control each. No MSA cases had variants in *SNCA* after QC.

All the rest of the variants manually analysed were first filtered by frequency (MAF<0.01), functional consequence (excluding synonymous changes and changes outside exons and splice sites), and then evaluated on the basis on the clinical information stored in the ClinVar database for known pathogenic alterations. All variants known to be pathogenic or presented as of uncertain significance were analysed. This list together with my interpretation is presented in Table 5-15.

Table 5-15: List of filtered rare potentially pathogenic variants in candidate genes studied in the MSA case control exome study.

Chr:position, transcript	Gene	Function	exon	change	1000g2015_all	ExAC_all	SIFT	Polypeptide	MutationTaster	CADD_raw	CADD_phred	CLINSIG	CLNDBN	msa count (all het)	control count (all het)	comment	relevant role in our study
1:17314656, NM_001141974	ATP13A2	nonsyn	24	c.A2704T, p.I902F	0.000798722	0.001	T	B	D	3.512	23.1	Uncertain significance	not provided	0	2	present only in controls	unlikely
1:17322750, NM_001141973	ATP13A2	nonsyn	14	c.G1337A, p.R446Q	NA	4.96E-05	T	B	N	2.063	16.62	Uncertain significance	not specified	0	1	present only in controls	unlikely
1:155206167, NM_001171811	GBA	nonsyn	7	c.G832A, p.E278K	0.00499201	0.0098	T	B	A	2.173	17.33	Pathogenic	Gaucher's disease	11	9	present in 11 msa cases in the het state and 9 controls in the het state. No comp het. Also no sign difference with chi sq	unlikely
1:155207965, NM_001171811	GBA	nonsyn	5	c.G460A, p.G154R	NA	2.47E-05	T	P	D	4.556	24.4	Pathogenic	Gaucher's disease	1	0	present in 1 msa case in the het state. No compound het	unlikely
3:182810333, NM_020166	MCCC1	nonsyn	3	c.G137A, p.G46E	0.000199681	4.98E-05	T	B	D	2.771	21.2	Pathogenic	Not provided	0	1	present in 1 control only	unlikely
4:84194658, NM_015697	COQ2	nonsyn	3	c.A683G, p.N228S	NA	0.0002	D	D	A	4.178	23.8	Pathogenic	Coenzyme Q10 deficiency	1	1	present in 1 msa case and 1 control. All in the heterozygous state. No compound het	unlikely
4:84194751, NM_015697	COQ2	nonsyn	3	c.G590A, p.R197H	NA	8.29E-06	D	D	A	8.07	35	Pathogenic	Coenzyme Q10 deficiency	1	0	present in 1 msa case in the heterozygous state. No compound het	possible test for association
4:84200234, NM_015697	COQ2	nonsyn	2	c.G437A, p.S146N	NA	5.84E-05	D	D	A	5.994	27.8	Pathogenic	Coenzyme Q10 deficiency	1	0	present in 1 msa case in the heterozygous state. No compound het	possible test for association
4:77084391, NM_001204255	SCARB2	nonsyn	8	c.G956A, p.G319E	0.000798722	0.0003	T	P	N	2.731	21	Uncertain significance	not specified	0	1	present in 1 control. Not the pd related snp (rs6812193)	unlikely
4:77091123, NM_001204255	SCARB2	nonsyn	5	c.T581C, p.M194T	0.000399361	8.25E-05	D	P	D	3.657	23.2	Uncertain significance	not specified	0	1	present in 1 control. Not the pd related snp (rs6812193)	unlikely
4:77100846, NM_005506	SCARB2	framesh ins	4	c.435_436ins AG, p.W146fs	NA	6.60E-05	NA	NA	NA	NA	NA	Pathogenic	Epilepsy progressive myoclonic 4, renal failure	1	0	present in 1 msa case. However this causes myoclonic epilepsy and glomerulosclerosis and is AR. https://www.ncbi.nlm.nih.gov/pubmed/18308289 . Our case is het and no comp het. And this is not the pd related snp rs727502773	unlikely
6:162206852, NM_013988	PARK2	nonsyn	4	c.C376T, p.R126W	0.000399361	0.0021	D	D	A	7.462	34	Pathogenic	Parkinson disease 2	3	4	present in het state in 3 msa and 4 controls. No compound het	unlikely

6:162394435, NM_013988	PARK2	nonsyn	3	c.A186T, p.K62N	NA	1.67E-05	D	D	A	5.959	27.7	Pathogenic	Parkinson disease 2	0	1	present in 1 control . No comp het	unlikely
6:162683724, NM_004562	PARK2	nonsyn	3	c.C245A, p.A82E	0.00119808	0.0047	T	B	A	-0.878	0.029	Pathogenic	Parkinson disease 2	1	1	present in 1 msa and 1 control in het state. Both have another snp (rs1801582) in hom state (msa case) and het state (control) but that one is a known benign anc common snp with a maf of 16%	unlikely
12:40629436, NM_198578	LRRK2	nonsyn	4	c.T356C, p.L119P	0.000798722	0.0013	D	D	D	5.916	27.5	Uncertain significance	Parkinson disease autosomal dominant	8, 6	5	uncertain role in PD. Maf 1%. Present in 6 msa cases and 5 controls	unlikely
12:40645075, NM_198578	LRRK2	nonsyn	9	c.G1000A, p.E334K	0.00139776	0.003	T	B	N	2.897	21.8	Uncertain significance	Parkinson disease autosomal dominant	8, 0	1	likely benign. Maf 9%. Present in 1 control	unlikely
12:40677813, NM_198578	LRRK2	nonsyn	19	c.G2378T, p.R793M	0.000199681	0.0008	D	D	N	4.94	25	Uncertain significance	Parkinson disease autosomal dominant	8, 0	3	probably benign snp. Present in 3 controls	unlikely
12:40687426, NM_198578	LRRK2	nonsyn	21	c.G2769C, p.Q923H	0.000199681	0.0002	D	D	N	1.265	12.09	Uncertain significance	Parkinson disease autosomal dominant	8, 1	0	probably a rare snp present in 1 msa case	Possible test for association
12:40689368, NM_198578	LRRK2	nonsyn	23	c.A3018G, p.I1006M	NA	2.48E-05	D	P	N	3.068	22.4	Uncertain significance	Parkinson disease autosomal dominant	8, 0	1	probably a rare snp present in 1 control	unlikely
12:40697842, NM_198578	LRRK2	nonsyn	27	c.G3683C, p.S1228T	NA	0.0001	D	P	N	2.047	16.51	Uncertain significance	Parkinson disease autosomal dominant	8, 1	1	probably a rare snp present in 1 control and 1 msa case	unlikely
12:40713845, NM_198578	LRRK2	nonsyn	34	c.G4883C, p.R1628P	0.00638978	0.0017	D	D	D	5.999	27.8	Pathogenic	Parkinson disease autosomal dominant	8, 1	2	this is a known risk factor for PD in asians. In our study it was present in 2 controls and 1 msa case. The maf of this variant is 0.2%	unlikely
12:40713899, NM_198578	LRRK2	nonsyn	34	c.T4937C, p.M1646T	0.00479233	0.0092	T	B	D	2.261	17.91	Uncertain significance	Parkinson disease autosomal dominant	8, 10	14	likely benign snp maf 0.5% present in more controls than msa in our series and not significant difference	unlikely
12:40715849, NM_198578	LRRK2	nonsyn	36	c.G5183T, p.R1728L	NA	1.66E-05	D	D	D	7.359	34	Uncertain significance	Parkinson disease autosomal dominant	8, 1	0	probably a rare snp. Present in 1 msa case	possible test for association
12:40734202, NM_198578	LRRK2	nonsyn	41	c.G6055A, p.G2019S	0.000199681	0.0004	D	D	A	7.654	35	Pathogenic	Parkinson disease autosomal dominant	8, 0	1	this known pathogenic alteration was present in 1 control only	unlikely

12:40757242, NM_198578	LRRK2	nonsyn	48	c.C7067T, p.T2356I	NA	0.0002	T	B	N	2.017	16.32	Uncertain significance	Parkinson disease autosomal dominant	8,	0	1	probably a rare snp. Present in 1 control	unlikely
12:40758647, NM_198578	LRRK2	framesh ins	49	c.7185_7186i nsGT, p.E2395fs	NA	9.18E-06	NA	NA	NA	NA	NA	Uncertain significance	Parkinson disease autosomal dominant	8,	0	1	probably a rare snp. Present in 1 control	unlikely
14:55312502, NM_000161	GCH1	nonsyn	5	c.G610A, p.V204I	0.000199681	0.0003	D	D	D	5.863	27.3	Uncertain significance	Dystonia, dopa- responsive		0	1	present in 1 control only	unlikely
19:45411110, NM_000041	APOE	nonsyn	3	c.T137C, p.L46P	0.000798722	0.0024	T	P	N	1.14	11.43	Pathogenic	APOE4(-)- FREIBURG		1	3	citation is not pathogenic. Also present in 3 controls in contrast to 1 msa case	unlikely
22:38509628, NM_001004426	PLA2G6	nonsyn	14	c.G1906A, p.V636I	0.000199681	0.0005	D	D	D	6.859	33	Uncertain significance	Iron accumulation in brain		0	2	present in 2 controls. No comp het	unlikely
22:38516893, NM_001004426	PLA2G6	nonsyn	11	c.G1453A, p.G485S	0.000599042	0.0006	T	B	N	1.413	12.86	Uncertain significance	Iron accumulation in brain		1	1	present in 1 msa case and 1 control in het state. No comp het	unlikely

5.3.4 Association studies

5.3.4.1 Single locus

5.3.4.1.1 Power calculation

The power calculation for a single variant association study was performed retrospectively on http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html. The results are as follows:

Considering a prevalence of MSA of 3.4 per 100,000 and in order to reach a genome-wide significance level of 1.494×10^{-07} (0.05 threshold on 334,701 tests), the study has 15% power at the current sample size (MSA = 375; control = 578) with a relative risk (RR) = 3.0 for a MAF=0.02.

This is illustrated in Figure 5-11 and Figure 5-12 according to allele frequency and relative risk.

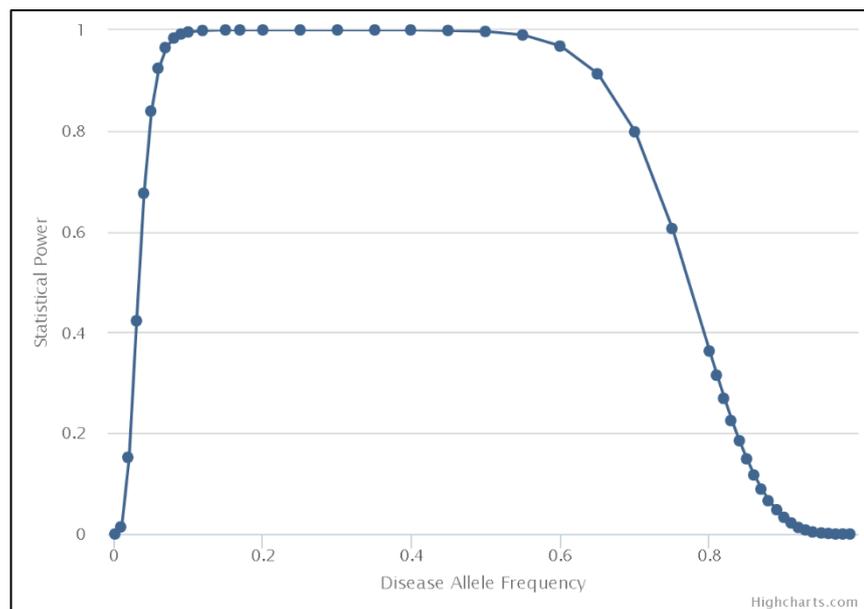


Figure 5-11: Figure presenting estimated study power according to disease allele frequency.

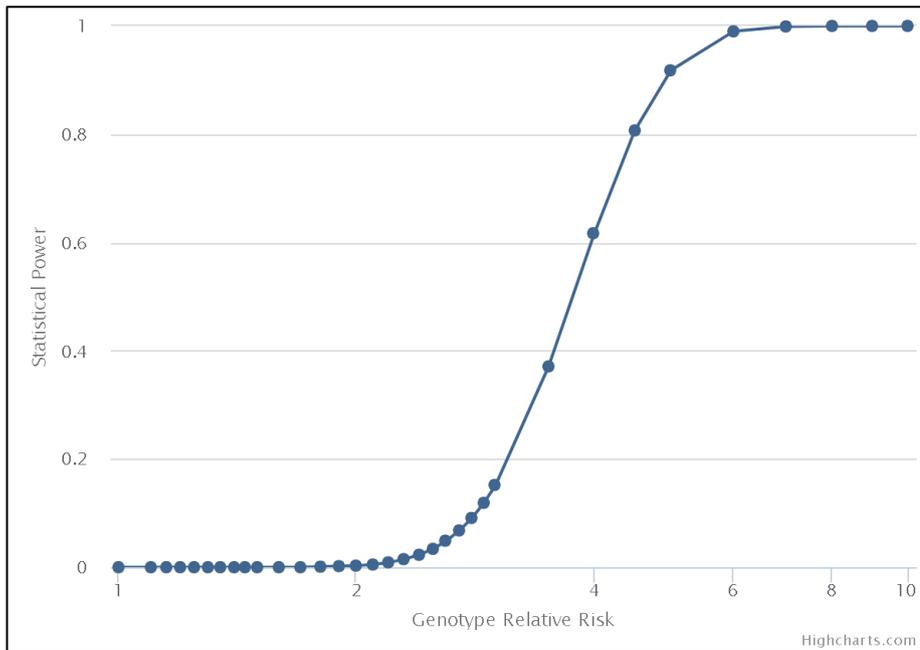


Figure 5-12: Figure presenting estimated study power according to genotype relative risk.

5.3.4.1.2 Quantile-quantile plot

A QQ plot was performed using R commands. The scripts are presented in the appendix. The QQ plot is shown in Figure 5-13. The points deviating from the $y=x$ line may indicate associations.

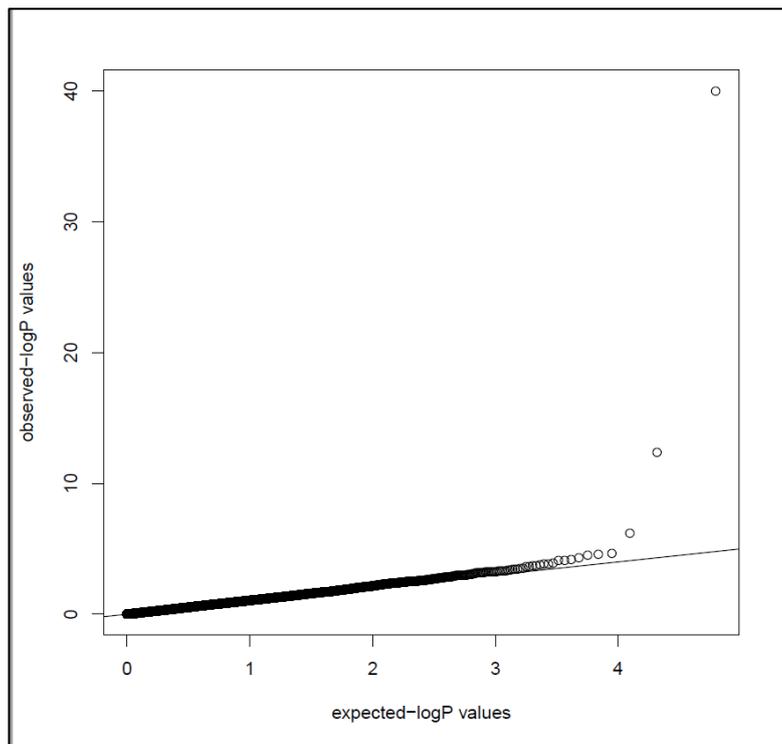


Figure 5-13: Quantile-quantile plot (QQ plot)..

5.3.4.1.3 Single variant results and Manhattan plot

A single association test for variants with a MAF >0.05 was performed and adjusted for multiple testing with Bonferroni correction. The Manhattan plot showing this is presented in Figure 5-14. The top 10 SNPs are presented in Table 5-16 and the results after adjustment for multiple testing in Table 5-17.

There were 2 variants that were significant after correction for multiple testing. These were in chromosomes 1 and 7. Both were called as small indels in two CAG repetitive regions.

The variant in chromosome 7 is present in the Exac database at a MAF of 0.002648 across all populations and of 0.00269 among Europeans. This variant was detected 92 times across our cases and once across our controls. Also, there were 74 samples with missing genotypes. This variant is located in the intronic region of the gene *TBX20* and is multiallelic. I therefore visually inspected the data with the IGV viewer for samples in all categories (alt allele, ref allele, missing genotype). This is presented in Figure 5-15. All samples appear very similar when visually inspecting the, they presented between 60 and 100 reads at this

position and therefore the data is of good quality. The variant is detected by reads in both forward and reverse directions.

The variant in chromosome 1 is called as an inframe insertion in the gene *SERINC2* and presents a MAF across all populations of 0.7309 in the Exac database. In Europeans, the MAF is of 0.7877. This variant is also multiallelic and is present in a repetitive region of a CAG repeat. This variant was found 196 times in cases and 158 times among controls in our cohort. Six MSA cases and 9 controls were homozygous and the rest heterozygous calls. Again, the visual inspection results are presented in Figure 5-16.

Of note, there are no other variants in either region in LD with these variants. This can be visualised in the Manhattan plot.

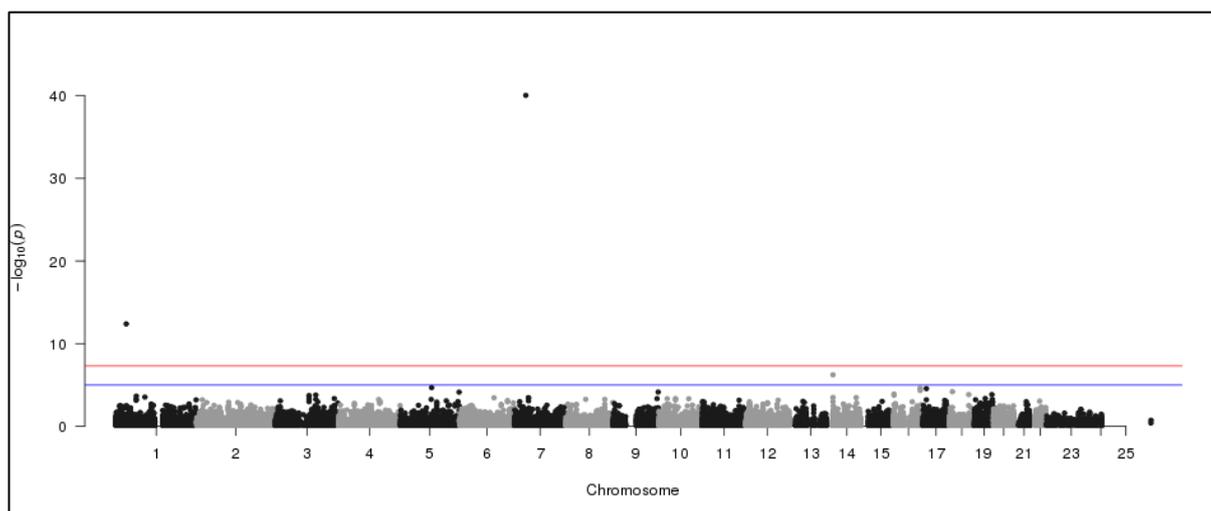


Figure 5-14: Manhattan plot presenting the association p-values from a simple χ^2 allelic test on MSA cases and controls.

Table 5-16: Top 10 SNPs of the MSA case control single variant association study.

Variant	CHR	SNP	BP	A1	C_A	C_U	A2	CHISQ	P	OR
1	7	.	35288458	C	92	1	CAG	178.8	9.03E-41	203.7
2	1	.	31905889	A	194	158	ACAG	52.6	4.09E-13	2.359
3	14	rs112192573	20528448	T	212	490	TCATAGATT TGCTCACTGAC	24.94	5.92E-07	0.6028
4	5	rs6233	95733112	G	336	404	A	18.1	2.10E-05	1.503
5	16	rs12931227	81199538	C	154	348	T	17.75	2.52E-05	0.6273
6	17	rs17811250	10404046	A	189	201	G	17.57	2.77E-05	1.612
7	16	rs12597040	81199555	G	157	348	A	16.6	4.62E-05	0.6384
8	18	rs2240906	9887546	A	321	368	G	15.99	6.36E-05	1.478

9	9	rs2297537	139564474	G	290	548	C	15.77	7.14E-05	0.6782
10	5	rs10073017	180551937	T	72	57	C	15.72	7.36E-05	2.048

Ref: CHR=Chromosome; BP=basepair; A1=allele 1; C_A=Allele 1 count among cases; C_U=Allele 1 count among controls; A2=allele 2; CHISQ= Allelic test chi-square statistic; P=p-value; OR=odds ratio.

Table 5-17: Top 10 SNPs of the MSA case control single variant association study after adjusting for multiple testing.

Variant	CHR	SNP	UNADJ	BONF
1	7	.	9.03E-41	2.81E-36
2	1	.	4.09E-13	1.27E-08
3	14	rs112192573	5.92E-07	0.01842
4	5	rs6233	2.10E-05	0.6545
5	16	rs12931227	2.52E-05	0.7858
6	17	rs17811250	2.77E-05	0.8627
7	16	rs12597040	4.62E-05	1
8	18	rs2240906	6.36E-05	1
9	9	rs2297537	7.14E-05	1
10	5	rs10073017	7.36E-05	1

Ref: CHR=Chromosome; UNADJ=unadjusted p-values; BONF= Bonferroni single-step adjusted

p-values.

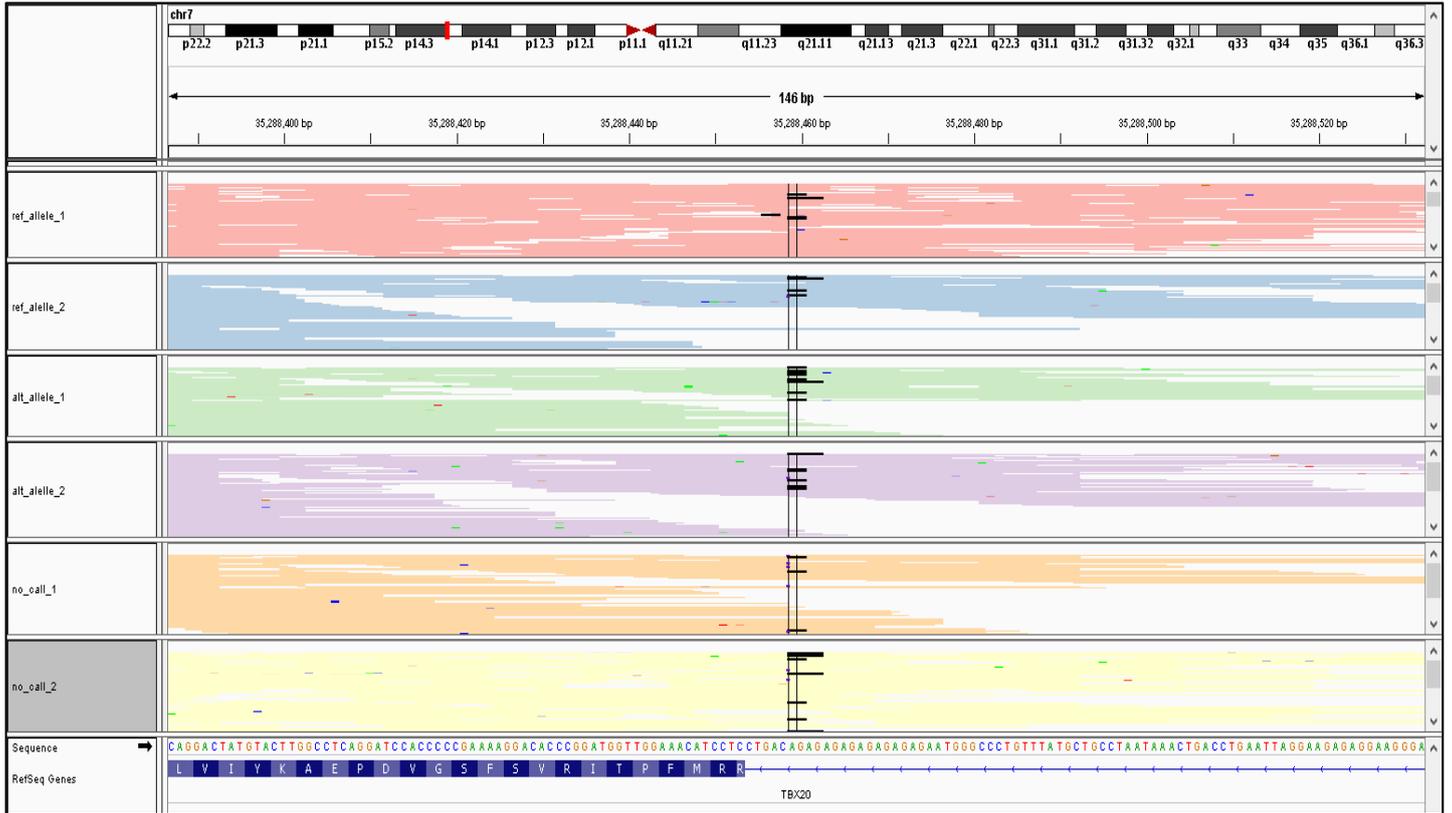


Figure 5-15: Genomic view of the chromosome 7 at position 3528459 corresponding to the TBX20 gene.

Ref: Samples named ref allele were called as reference, alt allele was called as the first alternative allele and no call were missing genotypes; however, when visually inspecting the data they all look very similar and present with a small deletion (black horizontal lines).

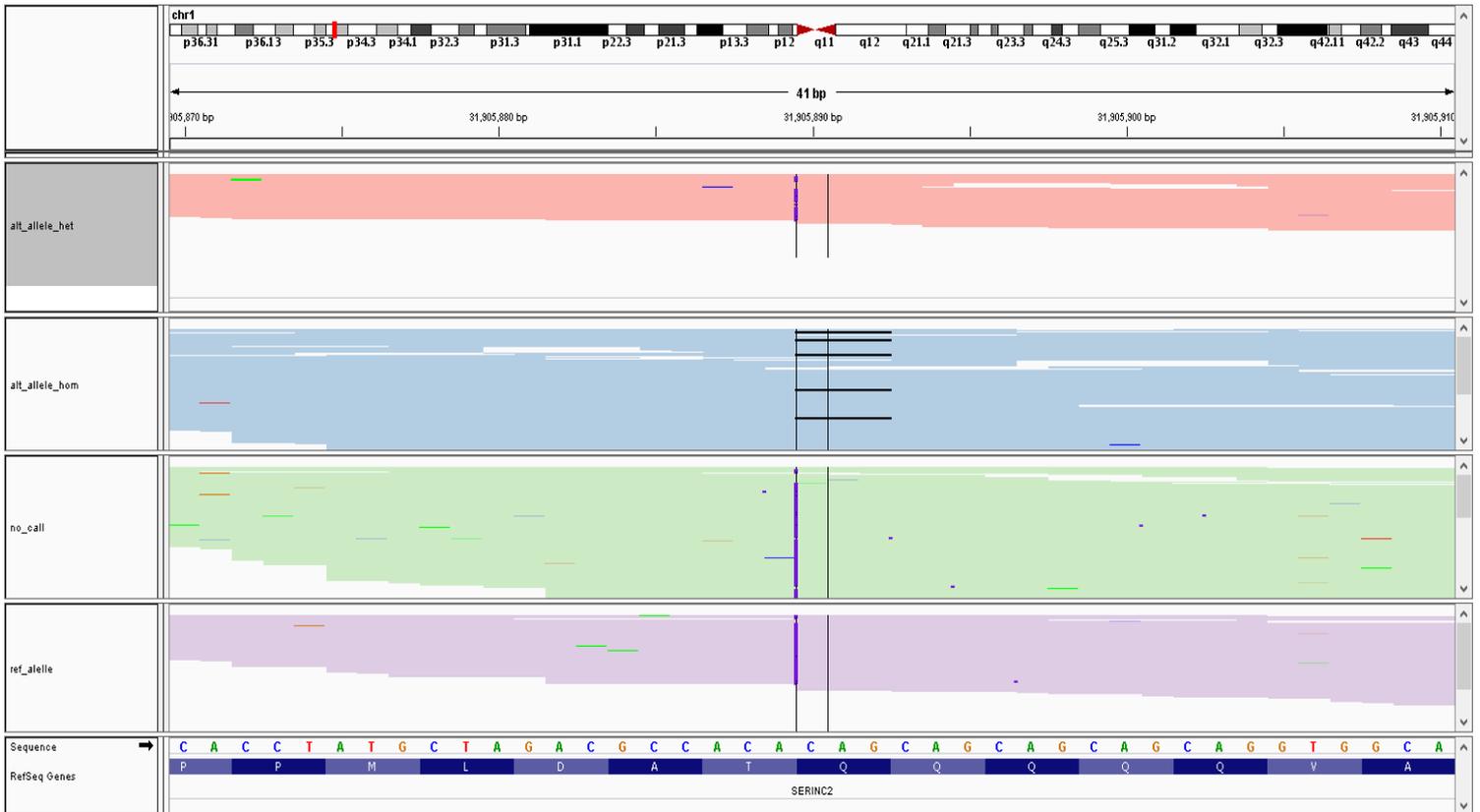


Figure 5-16: Genomic view of the chromosome 1 at position 31905889 corresponding to the SERINC2 gene.

Ref: Samples named *ref allele* were called as reference, *alt allele* was called as the first alternate allele and *no call* were missing genotypes. The sample in the bottom was called as a reference. When visually inspected samples appear to have different genotypes. For instance, the sample in blue resembles a deletion and is was called as a homozygous insertion. Note: black horizontal lines correspond to small deletions and vertical purple lines are insertions.

5.3.4.2 Rare variant tests

A rare variant association test was performed using the software package RVTESTS(267) as described on the methods section 3.3.9.6. RVTEST software was used to annotate the joint VCF file containing all MSA cases and controls that passed QC steps. This annotated file was then tested for a burden test and 2 kernel tests:

- Burden test: CMC test(268).
- Kernel methods: SKAT(343) and SKAT-O tests(344).

The number of variants included in the analysis is 198,567. Variants were grouped into gene units and the number of genes/regions analysed was 55,765. The significance level considered correcting for multiple testing was $p < 0.05$ divided by the number of units tested that was 55,765: 0.0000008966 (8.966E-07).

The analysis was first performed on pathologically confirmed cases only. Later we included the clinical cases in a replication phase and re-analysed the data as a whole dataset.

The strategies used to analyse these data were:

1. Analysis of all variants
2. Analysis of rare (< 0.01 MAF) variants
3. Analysis of all variants excluding synonymous changes

After reviewing the results, I have decided to present the SKAT-O results because these are the ones that less likely show false positive results as the QQ appear to be more reliable. CMC and SKAT results are available upon request.

The QQ plots and Manhattan plots are presented in Figure 5-17, Figure 5-18, Figure 5-19 and Figure 5-20. The significant results by gene are presented in Table 5-18 and the protein product and function of this list is in Table 5-19.

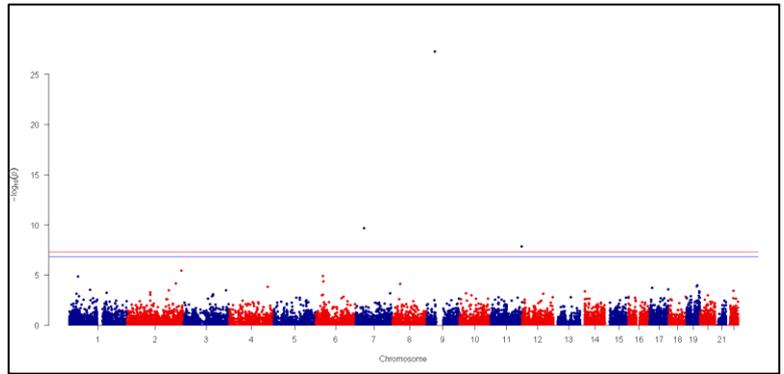
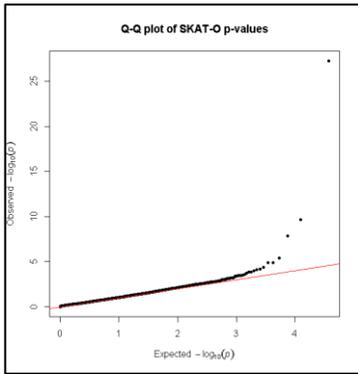


Figure 5-17: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing all variants and comparing pathologically confirmed MSA cases versus controls.

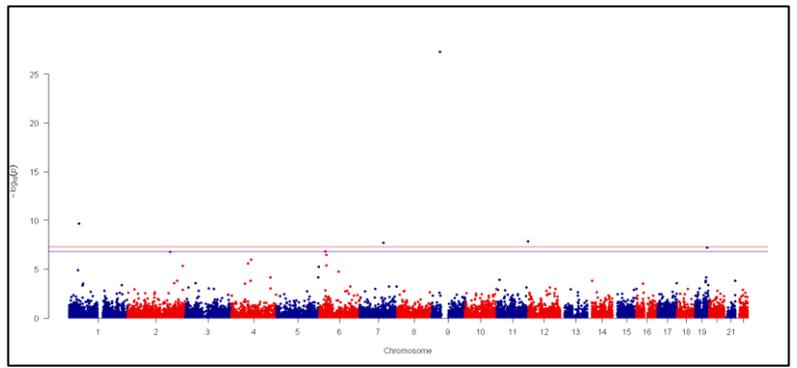
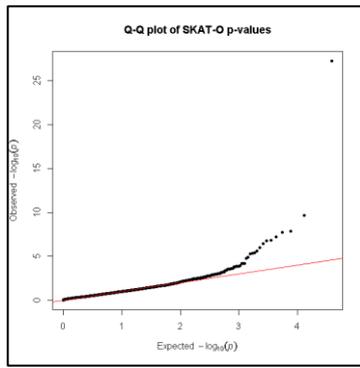


Figure 5-18: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing rare variants (MAF under 0.01) and comparing pathologically confirmed MSA cases versus controls.

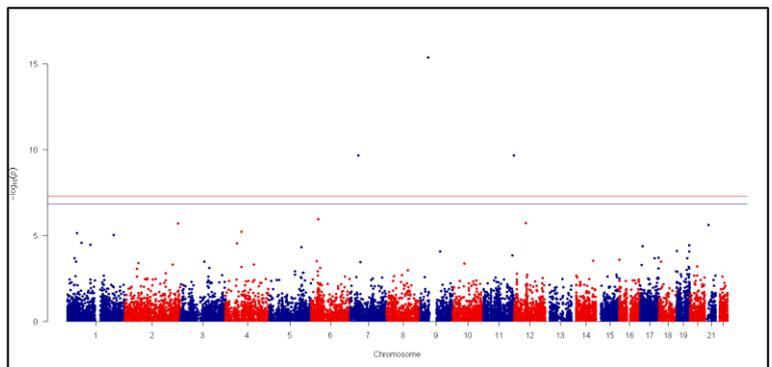
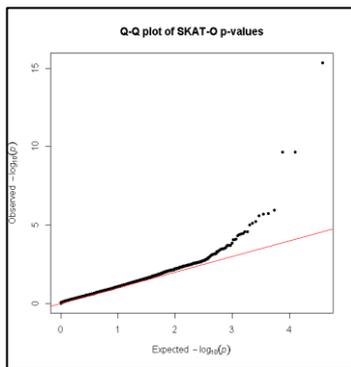


Figure 5-19: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing all variants and comparing all MSA cases versus controls.

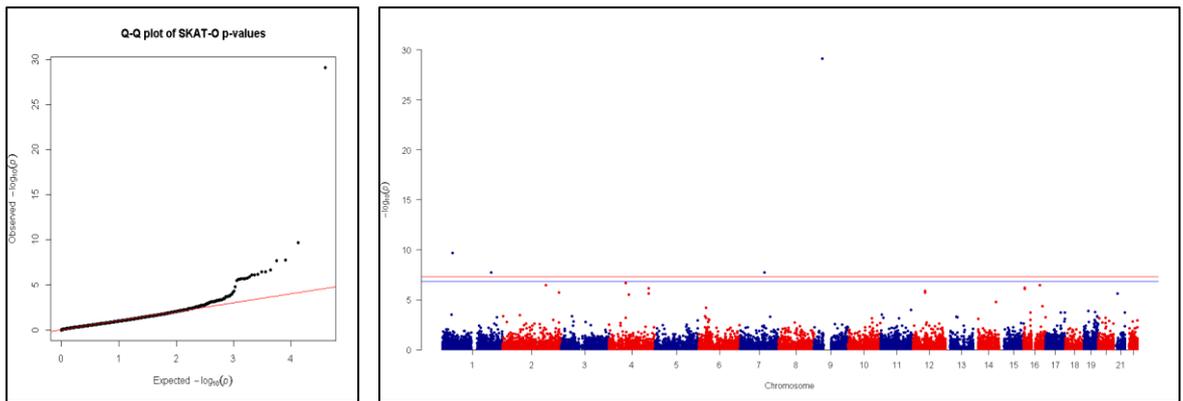


Figure 5-20: QQ plot (left) and Manhattan plot (right) of SKAT-O results analysing rare variants (MAF under 0.01) and comparing all MSA cases versus controls.

Table 5-18: Significant results of SKAT-O test investigating association of rare variation in MSA cases versus controls.

Only definite MSA--All variants						
Gene	RANGE	N_INFORMATIVE	NumVar	NumPolyVar	Q	Pvalue
<i>HRCT1</i>	9:35906188-35907138	841	1	1	121711	5.61E-28
<i>TBX20</i>	7:35242041-35293758,7:35271111-35293242	841	7	7	91764.4	2.15E-10
<i>OPCML</i>	11:132284870-132813566,11:132284874-133402414,11:132289753-133402297,11:132289913-132813663,11:132290086-132812987,11:132526981-132812905,11:132687020-132730672	841	6	6	19566.9	1.50E-08
Only definite MSA--Rare variants (MAF<0.01)						
Gene	RANGE	N_INFORMATIVE	NumVar	NumPolyVar	Q	Pvalue
<i>HRCT1</i>	9:35906188-35907138	841	1	1	336775	5.61E-28

<i>CCDC30</i>	1:42929000-42948667,1:42929024-43119857,1:42929024-43119852,1:42929096-42939088,1:42929592-42939231,1:42933414-42939056,1:42948364-43003656,1:42955599-43119857,1:42962781-43047087,1:42999549-43119861,1:43000559-43120335,1:43002155-43120011,1:43002243-43061244,1:43008459-43119852,1:43047104-43056132	841	20	14	307978	2.15E-10
<i>OPCML</i>	11:132284870-132813566,11:132284874-133402414,11:132289753-133402297,11:132289913-132813663,11:132290086-132812987,11:132526981-132812905,11:132687020-132730672	841	6	6	19699.2	1.43E-08
<i>SRRT</i>	7:100472732-100473557,7:100472747-100486285,7:100472747-100486285,7:100472784-100479326,7:100472806-100486285,7:100472809-100486285,7:100473010-100479343,7:100479384-100481860,7:100480340-100482173,7:100482374-100483208,7:100482612-100486277,7:100482915-100484445,7:100483328-100483806,7:100484180-100484821,7:100484622-100485117,7:100484968-100485994,7:100485323-100486281,7:100485655-100486285	841	31	27	34431.9	1.94E-08
<i>HSD17B14</i>	19:49316273-49339935,19:49316274-49318329,19:49316280-49339767,19:49316365-49339080	841	9	7	22865.1	6.32E-08
<i>BTN1A1</i>	6:26501448-26510650	841	12	10	21373.6	1.61E-07
<i>PRKRA</i>	2:179296140-179315958,2:179296141-179308734,2:179296380-179315878,2:179296380-179315878,2:179296692-179315491,2:179296702-179315484,2:179301000-179310610,2:179308015-179315211,2:179308069-179315786,2:179310152-179316239,2:179310221-179315456,2:179312046-179315840,2:179314805-179315900	841	14	10	26262.4	1.73E-07
<i>HLA-J</i>	6:29974359-29977733,6:29974363-29977733,6:29974372-29977733,6:29974382-29977316	841	4	4	13212.5	3.47E-07
<i>All samples--All variants</i>						

Gene	RANGE	N_INFORMATIVE	NumVar	NumPolyVar	Q	Pvalue
<i>HRCT1</i>	9:35906188-35907138	953	1	1	206090	4.44E-16
<i>OPCML</i>	11:132284870-132813566,11:132284874-133402414,11:132289753-133402297,11:132289913-132813663,11:132290086-132812987,11:132526981-132812905,11:132687020-132730672	953	6	6	42977.4	2.15E-10
<i>TBX20</i>	7:35242041-35293758,7:35271111-35293242	953	7	7	100004	2.15E-10
All samples--Rare variants (MAF<0.01)						
Gene	RANGE	N_INFORMATIVE	NumVar	NumPolyVar	Q	Pvalue
<i>HRCT1</i>	9:35906188-35907138	953	1	1	551117	7.84E-30
<i>CCDC30</i>	1:42929000-42948667,1:42929024-43119857,1:42929024-43119852,1:42929096-42939088,1:42929592-42939231,1:42933414-42939056,1:42948364-43003656,1:42955599-43119857,1:42962781-43047087,1:42999549-43119861,1:43000559-43120335,1:43002155-43120011,1:43002243-43061244,1:43008459-43119852,1:43047104-43056132	953	20	15	716149	2.15E-10
<i>SRRT</i>	7:100472732-100473557,7:100472747-100486285,7:100472747-100486285,7:100472784-100479326,7:100472806-100486285,7:100472809-100486285,7:100473010-100479343,7:100479384-100481860,7:100480340-100482173,7:100482374-100483208,7:100482612-100486277,7:100482915-100484445,7:100483328-100483806,7:100484180-100484821,7:100484622-100485117,7:100484968-	953	31	29	45790.2	1.83E-08

	100485994,7:100485323- 100486281,7:100485655-100486285					
<i>GPR37L1</i>	1:202091985-202102720	953	13	11	42848	1.95E-08
<i>YTHDC1</i>	4:69176104-69215807,4:69179038- 69215807,4:69179041-69215471,4:69181572- 69184439,4:69188209-69195791,4:69203483- 69204130,4:69203557-69215763	953	14	14	25768.7	2.09E-07
<i>PRKRA</i>	2:179296140-179315958,2:179296141- 179308734,2:179296380- 179315878,2:179296380- 179315878,2:179296692- 179315491,2:179296702- 179315484,2:179301000- 179310610,2:179308015- 179315211,2:179308069- 179315786,2:179310152- 179316239,2:179310221- 179315456,2:179312046- 179315840,2:179314805-179315900	953	14	13	32181.4	3.29E-07
<i>SLC9A5</i>	16:67271585-67306093,16:67280817- 67293845,16:67282774-67305961,16:67282852- 67306093,16:67282879-67306093,16:67282881- 67290381,16:67282980-67299051,16:67286638- 67299009,16:67286711-67304883	953	24	22	31746.3	3.30E-07
<i>RP11- 127120.7</i>	16:4780115-4802566	953	28	22	26273.3	6.37E-07
<i>ANP32C</i>	4:165118158-165118863	953	15	7	18058.4	7.61E-07
<i>C16orf71</i>	16:4784272-4799397,16:4784344- 4790346,16:4784480-4799396,16:4784502- 4799397	953	27	21	25994	8.23E-07

I have also analysed with the rvtests the candidate genes that presented a p value below 10^{-6} in the MSA GWAS(168) but were no statistically significant results after correction for multiple testing. None of the genes presented variants for analysis after the QC steps so those genes were skipped during the analysis process.

Finally, I also run the analysis weighting variants by function, including Start_Gain, Stop_Loss, Start_Loss, Essential_Splice_Site, Stop_Gain, Normal_Splice_Site, and Nonsynonymous changes and excluding synonymous changes. This did not show any significant result.

Table 5-19. Table presenting the protein encoded by significant genes in the MSA case control SKAT-O analysis and the function of these proteins.

	Gene	Protein	Function
1	<i>HRCT1</i>	Histidine-rich carboxyl terminus protein 1	Unknown
2	<i>TBX20</i>	T-box transcription factor TBX20	Transcriptional activator and repressor required for cardiac development and may have key roles in the maintenance of functional and structural phenotypes in adult heart
3	<i>OPCML</i>	Opioid-binding protein/cell adhesion molecule	Binds opioids in the presence of acidic lipids; probably involved in cell contact.
4	<i>CCDC30</i>	Coiled-coil domain-containing protein 30	Unknown
5	<i>SRRT</i>	Serrate RNA effector molecule homolog	Mediator between the cap-binding complex and the primary microRNAs processing machinery during cell proliferation
6	<i>HSD17B14</i>	17-beta-hydroxysteroid dehydrogenase 14	NAD-dependent 17-beta-hydroxysteroid dehydrogenase activity. Converts oestradiol to oestrone
7	<i>BTN1A1</i>	Butyrophilin subfamily 1 member A1	May function in the secretion of milk-fat droplets. May act as a membrane-associated receptor for the association of cytoplasmic droplets with the apical plasma membrane. Inhibits the proliferation of CD4 and CD8 T-cells activated by anti-CD3 antibodies, T-cell metabolism and IL2 and IFNG secretion.
8	<i>PRKRA</i>	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	Inhibition of translation and induction of apoptosis. Required for siRNA production by DICER1 and for subsequent siRNA-mediated post-transcriptional gene silencing.
9	<i>HLA-J</i>	Major Histocompatibility Complex, Class I, J (Pseudogene)	Pseudogene
10	<i>GPR37L1</i>	Prosaposin receptor GPR37L1	Receptor for the neuroprotective and glioprotective factor prosaposin. Ligand binding induces endocytosis, followed by an ERK phosphorylation cascade
11	<i>YTHDC1</i>	YTH domain-containing protein 1	Regulator of alternative splicing that specifically recognizes and binds N6-methyladenosine (m6A)-containing RNAs
12	<i>SLC9A5</i>	Sodium/hydrogen exchanger 5	Involved in pH regulation to eliminate acids generated by active metabolism. Major proton extruding system driven by the inward sodium ion chemical gradient. Role in signal transduction
13	<i>RP11-127I20.7</i>	RNA gene	Unknown
14	<i>ANP32C</i>	Acidic leucine-rich nuclear phosphoprotein 32 family member C	Tumor suppressor that can inhibit several types of cancers, including prostate and breast.
15	<i>C16orf71</i>	Uncharacterized protein Chromosome 16 Open Reading Frame 71	Unknown

5.3.5 Discussion

MSA is a rare disease with a low heritability(345). A recent GWAS was unable to detect common variants associated with risk of the disease(265). This is possibly due to lack of power of the study because the number of definite cases was ~300 and that of clinical cases ~700. Unfortunately, given the rarity of the disease it is unlikely that we will be able to increase our cohort significantly for a long period of time, and WES has been proposed as an alternative approach to unravel the genetic contribution of rare variation in rare neurodegenerative disorders such as AD(346) and PD(347).

In the study presented here, we performed WES in the largest MSA cohort available and represents a unique opportunity to improve our understanding of this devastating disorder.

After a stringent QC we studied candidate genes that have been previously linked to PD, DLB and MSA. We did not find any relevant pathogenic variants. In the gene *COQ2*, the variant that has been proposed to increase risk of MSA in Japanese (p.V393A) was found in this study in 1 control sample only. Also, the variant p.R22X that was previously identified with a significant protective effect, in this study did not show significant differences between MSA and controls. There were other variants detected that are known to cause coenzyme Q10 deficiency with an AR mode of inheritance, but were here found in the heterozygous state. These variants warrant further follow up and study.

There were 2 variants in *SNCA* that were interpreted to be unlikely pathogenic or affect risk of MSA (see Table 5-15).

There was one p.G2019S known pathogenic alteration in *LRK2* but this was present in one control sample. We interpreted this mutation as probably of reduced penetrance. All the samples included as controls were reported as healthy at the time of assessment.

There were other variants in the genes studied but they were all interpreted as unlikely related to MSA. The details of each variant and the reasons for this interpretation are presented in Table 5-15.

We also performed a single variant chi square allelic association test of MSA cases and controls to investigate for common variation ($MAF \geq 0.05$). There were 2 signals that appeared significant after Bonferroni correction for multiple testing. They were located in chromosome 7 and chromosome 1. Both variants are called as small indels and do not appear to have other variants in LD close to them. I therefore decided to visually inspect the data and look at the sequence. When analysing the data on the IGV viewer I realised that the actual sequence appears in a repetitive region of CAG repeats. The genotypes that were in my final files were not consistent with what I was seeing in the bam files. I think the most likely explanation is that these 2 variants are artefacts due to the limitation of the WES technique in repetitive regions. Moreover, the 2 variants are common in the population and are multi allelic. For multiallelic sites, Plink 1.9 defaults to keeping only the reference allele and the most common alternative allele; any call involving a lower-frequency alternative allele is treated as missing data. This could have also contributed to these highly significant results because we used plink for the association study and the QC steps.

The next step of analysis performed was investigating the role of rare variants in MSA. This was established by using a software called RVTESTS(267) where multiple tests can be performed. We did a burden test and two sequence kernel association tests (SKAT and SKAT-O). This test was recently successfully implemented in discovering PD associated gene(348) variants. The results of the SKAT-O test were the most reliable ones so this analysis was presented in the results section above.

There were several genes that appear to have significant results with a p value below the level of Bonferroni correction for multiple testing. These are presented in Table 5-18.

The only gene that appears to have a significant p value under all strategies of analysis is *HRCT1*. The signal in the gene *HRCT1* is given by only one change and is present only in MSA cases and absent controls. This variant (rs370606246) has an 8% MAF in Exac. It is a small

deletion present in a repetitive region (CTG repeat) and therefore I think this is likely an artefact due to the limitation of this technique in such areas; a similar scenario to the 2 variants found significant by single variant analysis.

The signal in the gene *TBX20* is mostly given by the common variants in this gene and the it appears to be given by 7 variants. This gene encodes a protein that has been found to present functions in cardiac development. The signal did not replicate when excluding common variation and it is also present in a repetitive region

The genes *CCD30*, *SRRT* and *PRKRA* are significant only when including the variants with a $MAF \leq 0.01$. The protein product of *CCD30*, Coiled-coil domain-containing protein 30, has an unknown function, whereas the one encoded by *SRRT*, Serrate RNA effector molecule homolog, is important during cell proliferation. Interferon-inducible double-stranded RNA-dependent protein kinase activator A, encoded by *PRKRA*, is known to participate in the inhibition of translation and induction of apoptosis. Mutations in *PRKRA* have been shown to cause Dystonia 16 characterized by early-onset progressive limb dystonia, laryngeal and oromandibular dystonia, and non-levodopa responding parkinsonism that is inherited in an AR manner. These genes are interesting and will be followed up in future studies.

The genes *HSD17B14*, *BTN1A1* and *HLA-J* are significant when analysing the pathologically confirmed cases and looking into rare variation. These did not show significant results in the replication phase of this study and therefore we cannot draw many conclusions on that standalone data. Moreover *HLA-J* is a pseudogene and unlikely to be of relevance.

Finally, the genes *GPR37L1*, *YTHDC1*, *SLC9A5*, *RP11-127I20.7*, *ANP32C*, and *C16orf71* appeared only to be significant when analysing the whole cohort including the clinical cases. The gene *GPR37L1* encodes for prosaposin receptor that functions as a neuroprotective and glioprotective factor and may have relevant functions in MSA. However, all the abovementioned genes including *GPR37L1* only showed a significant result with the large cohort including clinically diagnosed patients and not in the initial discovery phase of analysis of the pathologically confirmed samples. Thus, further investigations should be done to understand if they play a role in MSA or not.

A limitation of our WES study is that we were not able to account for age as a covariate. Unfortunately, the data we have for age of these samples was incomplete and in most of the cases it was not clear if it was age of onset, age at DNA extraction or age at death.

Another limitation of this study was the fact that samples were sequenced with different chemistries in different labs. We tried to minimise this bias by applying a stringent QC and focusing on the shared genotypes covered as much as possible and we visually inspected a large number of samples at different positions for coverage intensity.

In summary, these data don't support previous associations of MSA with the genes *COQ2* or *SNCA* and presents interesting genes that require further study to be validated.

6 CHAPTER 6. PRIMARY FAMILIAL BRAIN CALCIFICATION

6.1 CANDIDATE GENE ANALYSIS

6.1.1 *SLC20A2*

Statement of contribution. I performed the sample collection, lab work and data analysis and Dr Niccolo Mencacci supervised the sequences.

6.1.1.1 *Background*

Mutations in *SLC20A2*, encoding for Type III sodium-dependent inorganic phosphate transporter 2, are the most common cause of AD primary familial brain calcification among multiple populations(239). We decided to sequence this gene in our patients with Fahr's disease.

6.1.1.2 *Subjects, materials and methods*

We sequenced by Sanger sequencing the coding sequence and flanking regions of the *SLC20A2* gene in 28 sporadic and 3 familial cases with evidence of CNS calcification. The human gene consists of 11 exons, of which 10 are coding. The transcript used was: ENST00000342228.7.

6.1.1.3 *Results*

We found 2 synonymous changes (rs34124953 and rs111553899), 2 missense variants (rs73675069 and rs79577461) both presenting a MAF over 1% so interpreted an unlikely to be pathogenic and finally one missense change previously reported in 2 individuals of a PFBC family of Spanish origin with functional data supporting its pathogenicity(244). The mutation was heterozygous and in exon 10: c.1723G>A; p.Glu575Lys. The Sanger sequencing result is shown in Figure 6-1.

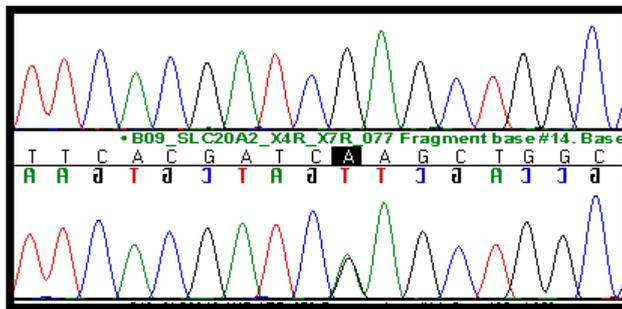


Figure 6-1: Sanger sequencing result of the heterozygous c.1723G>A p.Glu575Lys pathogenic missense mutation in SLC20A2.

Ref: Top: reference sequence. Bottom: mutated sequence highlighted in bold.

The patient exhibiting the heterozygous c.1723G>A p.Glu575Lys missense mutation was male. His symptoms started in his childhood with myoclonic epilepsy and also non-epileptic attacks, psychiatric symptoms, diabetes type I, polyneuropathy, bilateral ptosis, proximal upper and lower-limb weakness. At age 35 he was diagnosed with hypertrophic cardiomyopathy and required an implanted defibrillator. When we examined the patient we also detected a parkinsonian syndrome in this patient.

On neuroimaging, our patient had brain calcification in the basal ganglia, thalamus, frontal and parietal sub cortical areas, and the cerebellum Figure 6-2.

He was largely investigated for mitochondrial dysfunction (including a muscle biopsy and whole mitochondrial genome sequencing, respiratory chain enzyme assay, *POLG* and *PEO1* screening) with negative results.



Figure 6-2: Axial CT scan presenting brain calcification in the basal ganglia, thalamus, frontal and parietal sub cortical areas, and the cerebellum of a patient presenting a p.Glu575Lys mutation in SLC20A2.

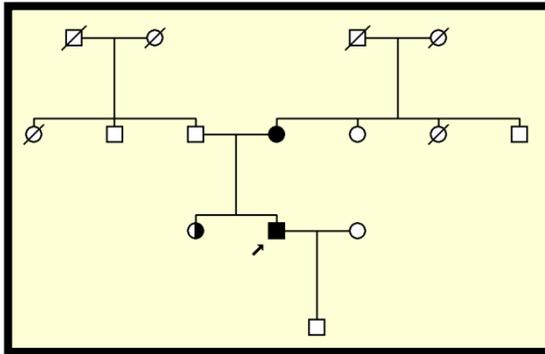


Figure 6-3: Family tree of a patient presenting with familial brain calcification caused by a p.Glu575Lys mutation in SLC20A2.

Other family members at the time of diagnosis were thought to be unaffected. However, when we examined them, the mother who also carries the mutation, had mild neuropsychiatric symptoms and had brain calcification on CT scan. Also, the sibling, a carrier as well, is currently asymptomatic and presents calcification in a brain scan (Figure 6-3).

6.1.1.4 Discussion

In this project we discovered the cause of brain calcification in a patient that had a complex disorder and had been investigated for many years. The patient had a known pathogenic alteration in SLC20A2 (p.Glu575Lys). It is important to mention that the mother of the patient who was previously thought to be unaffected, when undergoing a detailed

assessment also presented neuropsychiatric symptoms and is a carrier of the same mutation.

This patient also presented type 1 diabetes, hypertrophic cardiomyopathy and muscle weakness and this has not been reported in relation with *SLC20A2*. The cause of these features in this patient are still unknown.

6.1.2 PDGFB

Statement of contribution: I performed the sample collection, lab work and data analysis.

I sequenced by the Sanger method the coding sequence and flanking intronic regions of the *PDGFB* gene in 28 sporadic and 3 familial cases with evidence of CNS calcification. The human gene consists of 7 exons, of which 6 are coding. The transcript used for this experiment was: ENST00000331163.10.

No relevant variants were detected.

6.1.2.1 Discussion

In this project we did not detect any relevant variants causing brain calcification in these patients.

6.2 JAM2 VARIANTS CAUSES PFBC IN RECESSIVE FAMILIES

6.2.1 Statement of contribution

Samples were collected by Prof Henry Houlden and myself in collaboration with Dr Orlando Swayne, Dr Patrick Morrison, Dr Gavin McDonnell, Dr Raeburn Forbes, Dr John Mckinley and colleagues.

Skin biopsies were performed by Prof Henry Houlden and myself.

WES library preparation was performed by Debbie Hughes, Raw data analysis was performed by Dr Alan Pitmann and downstream analysis by myself. Genotyping data was performed in the ICH genomics (UCL), and data analysis was performed by myself. The fibroblasts were grown by Mr Chris Lovejoy and the western blot performed by Dr Marc Soutar while I was on maternity leave.

The Jam-b mice were provided by Prof Michel Aurrand-Lyons and the mice care and tests were performed by Dr Rosella Abeti. Histopathological analysis was done by Dr Zane Jaunmuktane. I led this collaboration and coordinated the data interpretation together with Dr Abeti, Dr Jaunmuktane and Prof Houlden.

6.2.2 Background

Homozygosity mapping in conjunction with WES has proven to be a successful combination in the discovery of disease genes in recessive disorders(349). In this section we used these techniques to investigate the genetic cause of Fahr's disease in two families.

6.2.3 Subjects materials, and methods

Two consanguineous recessive families of the travellers' communities from England and Northern Ireland were studied in this section of the thesis. The patient was from London and Professor Houlden and I assessed him and his family. The second family included is from Northern Ireland and was assessed by our collaborators after we contacted them regarding a previous abstract describing this family presented in the ABN meeting in 2007.

Samples were genotyped with Illumina HumanCytoSNP-12v2-1_H (described in the methods 3.3.6). Samples selected were 3 affected and 2 unaffected from the Irish family (Figure 6-5).

The exome was captured using nextera focussed library as described in the methods chapter 3.3.7.1.2.

Sanger sequencing for segregation was performed as described in the methods chapter 3.3.4.

The fibroblasts were cultured and the western blots were performed as described in the methods chapter 3.4.3 and 3.4.5.

The Jam-b deficient mice were generated and studied as described in the methods 0.

6.2.4 Results

The family from London consisted of one affected patient that presented marked dystonia, ophthalmoplegia, cerebellar ataxia and cognitive impairment. This male patient was

reported to have been born normal and started to show neurological symptoms during his childhood. The family tree is presented in Figure 6-4.

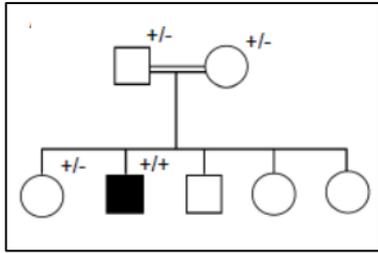


Figure 6-4: Family tree of the consanguineous family of a patient presenting with Fahr's syndrome.

Note: The proband is homozygous (+/+) for the p.R229X mutation in JAM2 and the subjects with a +/- sign are heterozygous unaffected carriers.

The Irish proband presented at age 41 with a 4-year history of akinetic rigid parkinsonism, memory problems and pyramidal signs. There were other 3 members of the family affected by familial brain calcification and they were highly consanguineous although the exact degree of consanguinity in each generation was difficult to assess. The family tree is presented in Figure 6-5.

The CT scans presenting widespread calcification in these patients are shown in Figure 6-6.

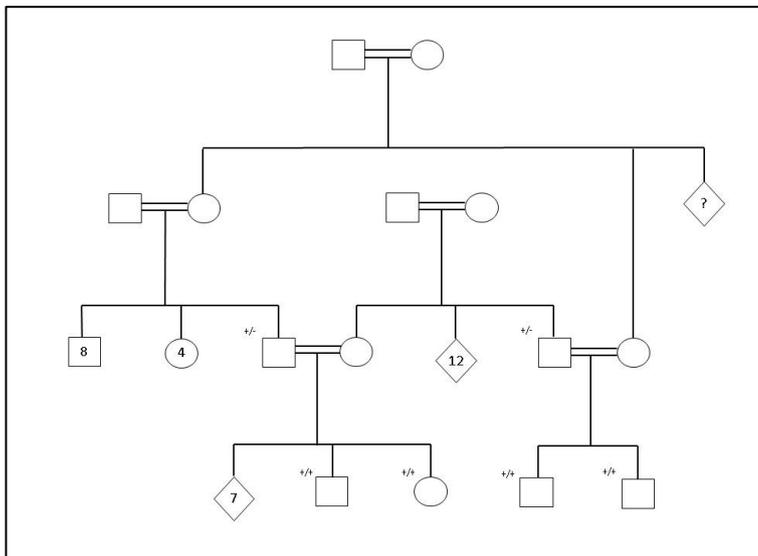


Figure 6-5: Family tree of a consanguineous family with primary familial brain calcification. Symbols +/- show subject homozygous for the p.R229X mutation in JAM2. +/- are heterozygous unaffected carriers.

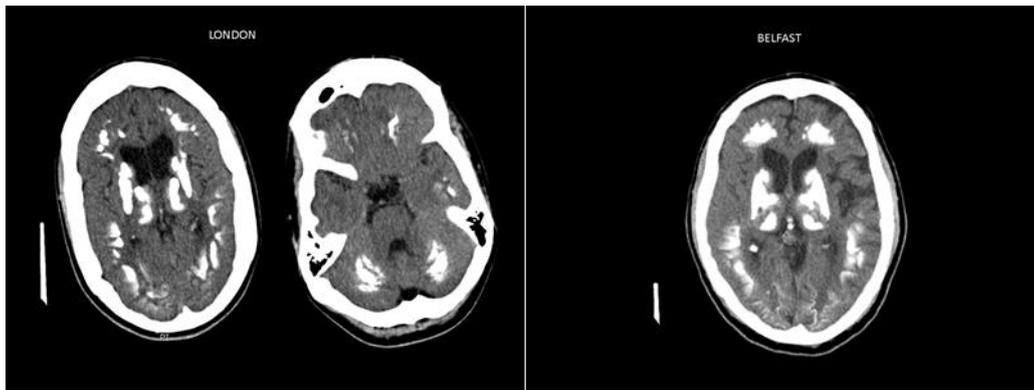


Figure 6-6: CT scan of the affected probands of the two consanguineous families with Fahr's syndrome exhibiting widespread calcification of the basal ganglia, thalamus, cerebral cortex and cerebellum.

6.2.4.1 Homozygosity mapping

Homozygous regions shared among the 3 affected members of the Irish family and absent in the 2 unaffected subjects were investigated using the online tool at <http://www.homozygositymapper.org> (350).

There were 3 homozygous regions in the chromosomes 10, 13 and 21 spanning:

- Chr 10: 37414883-43132376
- Chr 13: 88327643-93518692
- Chr 21: 22370881-28338710

These regions are represented in Figure 6-7.

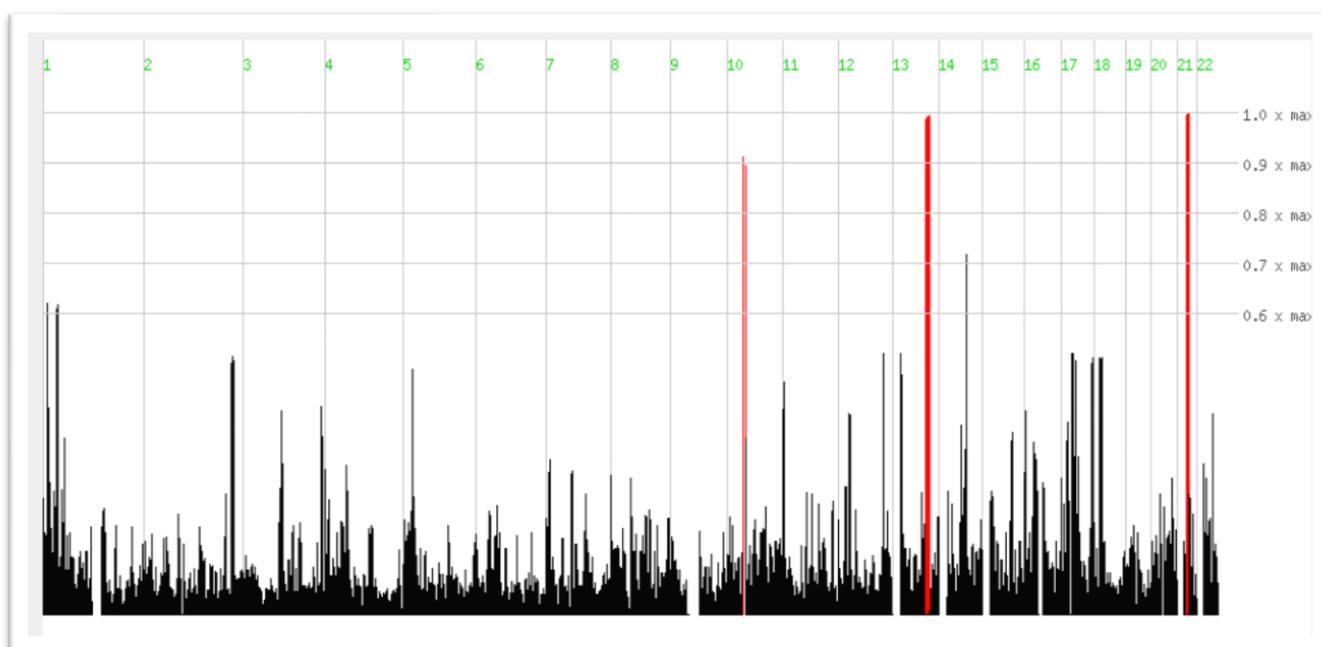


Figure 6-7: Genome-wide illustration of homozygous regions in the Irish family with brain calcification. The chromosomes numbers are depicted in green and the homozygous regions in red.

6.2.4.2 Exome sequencing

After delimiting the homozygous regions in the consanguineous family of interest, we decided to investigate for causal mutations in these regions.

We therefore decided to sequence the coding region of our proband.

The exome was enriched with Illumina's Nextera focused library and sequenced on a HiSeq 2000 to an average sequence depth of 43%; and on average, 93% of targets were covered at least 10X.

Under the assumption that a mutation would be present in the homozygous state in one of the homozygous regions, we determined the putatively damaging mutations (defined as missense, nonsense, frameshift, canonical splice-site) with a minor allele frequency < 1%.

We only found 1 variant under this strategy of filtering.

The variant is a nonsense variant: c.C685T:p.R229X in the gene *JAM2*.

6.2.4.3 Sanger sequencing and segregation study

We thus sequenced this gene in the other recessive family with Fahr's syndrome and detected the same mutation found in the proband of the family from London.

We investigated for segregation in both families and we could confirm this was present in all affected individuals and was either absent or heterozygous in the unaffected (Figure 6-4 and Figure 6-5). Segregation was confirmed in the family from London and was not complete in the family from Ireland due to lack of samples although all unaffected samples available were either not carriers or heterozygous carriers. Some subjects are not shown in the family tree due to the size of the family and the difficulties in ascertaining relationships.

6.2.4.4 Western blot

After confirming the mutation in both families, in collaboration with Dr Marc Soutar we obtained a Western blot showing lack of expression of the JAM2 protein in the affected proband of the family from London in contrast to the expression of this protein in a heterozygous carrier (mother) as well as 2 unrelated controls (Figure 6-8).

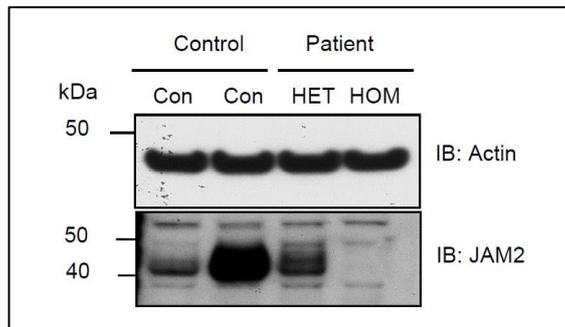


Figure 6-8: Western blot showing absence of JAM2 protein in homozygous mutation containing fibroblasts compares to a heterozygous carrier and 2 unrelated controls.

6.2.4.5 JAM2 KO mice

In order to characterise the link between mutations in *JAM2* and this neurological phenotype, we decided to investigate further using an animal model. There was no invertebrate model covering the region of interest so we assessed the possibility of a mouse model because this would potentially provide us of additional data. Through a collaborative work, we studied a *JAM-B* (ortholog of human *JAM2*) knock out (KO) mouse model.

6.2.4.5.1 Behavioural tests

We investigated the neurological manifestations of these mice by means of behavioural tests on locomotion abnormalities: walking beam performance and gait analysis. This work was done by Dr Rosella Abeti. The *JAM-B* KO animals exhibited significant difficulties when compared to wild type in beam walking test ($p=0.017$) and gait abnormalities (stride length $p<0.001$; sway length $p=0.002$) (Figure 6-9 and Figure 6-10).

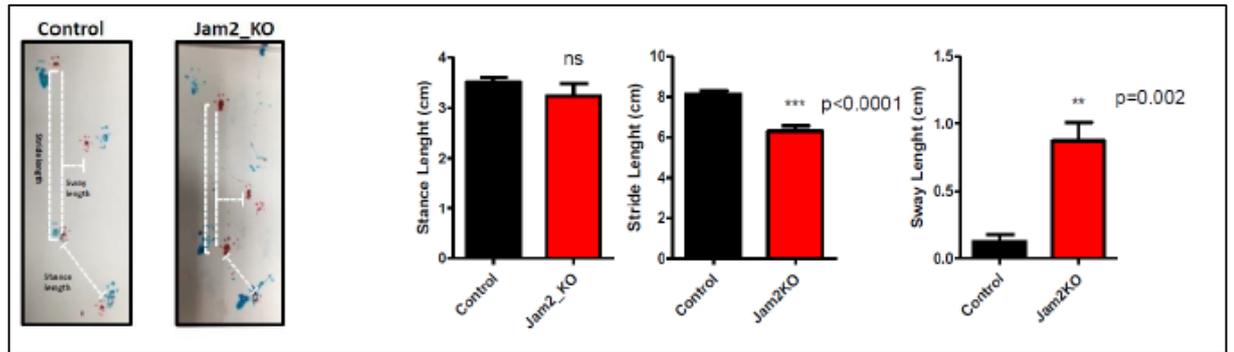


Figure 6-9: Behavioural study on *JAM-B* KO mice. Gait analysis results.



Figure 6-10: Behavioural study on *JAM-B* KO mice. Beam walking results.

A video of the beam walking test of a KO and a wild type have been presented to the examiners and discussed during the viva.

6.2.4.5.2 Histopathological analysis

The histopathological evaluation of *JAM-B* KO mice compared to controls is shown in Figure 6-11. *JAM-B* KO presented marked vacuolization of the brain and cerebellum and in particular this was more evident in the midbrain. Interestingly, this vacuolization was associated to glial activation as seen by immunohistochemistry on GFAP staining. Vacuoles

appear to be localized in the neurons.

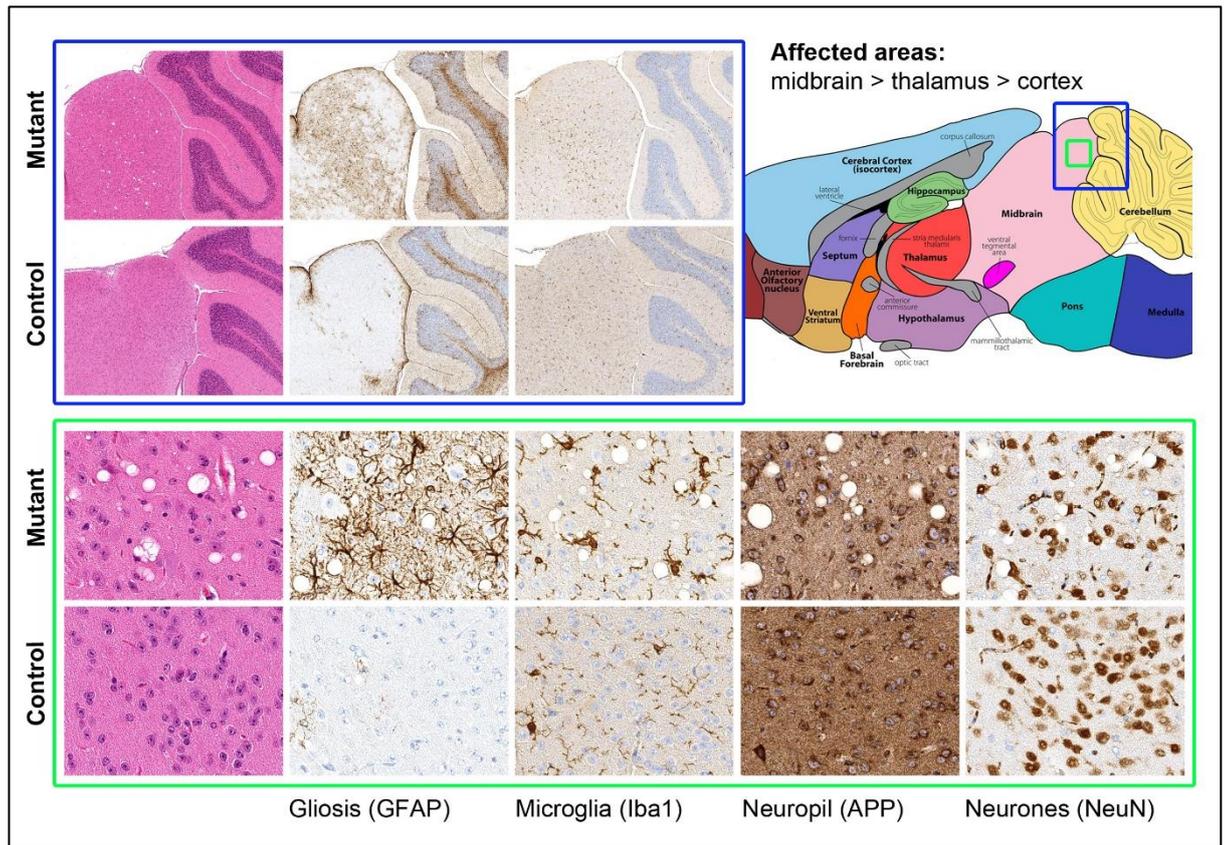


Figure 6-11: Histopathological analysis of JAM-B KO mice compared to wild type mice exhibiting vacuolization of the midbrain and glial activation in the KO mice. Low magnification (above), higher magnification (below).

6.2.5 Discussion

The data presented in this chapter is showing that a nonsense mutation in the gene *JAM2* causes a recessive disorder presenting with severe brain calcification in the absence of secondary causes of brain calcification. Through homozygosity mapping analysis we identified a region of homozygosity in one family and through exome sequencing we discovered the causing mutation. The same mutation was also found in a second family. By investigating a mouse model, we support these data showing that the KO mice present gait and histopathological abnormalities in the brain when compared to controls.

Identifying the genetic cause of neurological disorders where there is no current disease-modifying therapies available is of great importance. On one hand, offering families a precise diagnosis can help them understand the condition and allows the possibility of

genetic counselling and family planning. On the other hand, with the advances in clinical trials and the possibility of targeting different mutations by their mechanism (i.e. Ataluren, currently approved for Duchenne muscular dystrophy could be considered in the future for other nonsense mediated genetic disorders such as the one presented here) may bring light into diseases where the feasibility of a clinical trial would be impossible because they are very rare.

Junctional adhesion molecules are a family of proteins localized at the tight junction of polarized cells and on the cell surface of leukocytes. They play an important role in the regulation of cell polarity, endothelium permeability and leukocyte migration. In particular, an important role of tight junction proteins in the blood brain barrier regulation has been shown when recessive mutations in *JAM3* and *OCN* were linked to neurological disorders presenting with calcification in the brain in addition to other manifestations(351,352).

Additionally, recently *JAM2* was identified as an inhibitor of somatodendritic myelination in spinal cord neurons and this further confirms the importance of this molecule in the neuronal tissue(353).

Mechanistically, the proteins encoded by *SLC20A2* and *XPR1* have been linked to impairment of phosphate homeostasis in genetically determined brain calcification; and *PDGFB* and *PDGFR β* to blood brain barrier control regulation and integrity(239).

JAM2 presents a role in endothelium permeability and binds to platelet-derived growth factor receptor α positive (encoded by *PDGFR α*)(353) and we thus believe the mechanism by which it is causing brain calcification in our families is probably linked to blood brain barrier dis-regulation.

In summary, in collaboration with members of our lab and other teams I discovered that a nonsense mutation in the gene *JAM2* causes recessive brain calcification in 2 consanguineous families with a severe neurological disorder.

The most likely mechanism for the lack of *JAM2* protein in our patient is by nonsense-mediated decay caused by the presence of a premature stop codon. The gait disturbances

in JAM2 KO mice are in line with the phenotype we see in humans where the presenting features are movement disorders.

7 CHAPTER 7: CONCLUSIONS

The aim of this thesis was to investigate the genetic basis of two main diseases: Multiple system atrophy and primary familial brain calcification. This was achieved by both traditional Sanger sequencing and next generation sequencing.

7.1 MSA

In this thesis I investigated the genetic risk factors of MSA. MSA has proven to be a difficult disease to understand and unfortunately all previous genetic links to MSA have not been replicated in larger studies. By Sanger sequencing I found that a nonsense variant in the first exon of *COQ2* (not covered by the original publication linking *COQ2* and MSA) had a significantly increased frequency in controls compared to MSA cases. There was also another protective variant. Overall, this study, that is the largest of its kind including only pathologically confirmed MSA cases, did not support the association of increased risk of MSA with *COQ2* variants.

In contrast, I found a decreased level of coenzyme q10 in the cerebellum of MSA cases with some degree of cerebellar involvement (MSA-C and MSA-mixed cases) when compared to controls as well as cerebellar ataxia cases and other parkinsonian disorders. This interesting link between cerebellar MSA and coenzyme q10 has been replicated by others. However, although some researchers have proposed that this finding is supportive of the role of *COQ2* variants in MSA, in my opinion this still awaits confirmation and the mechanism may not involve *COQ2* variants because at least in our cases this gene was sequenced and not found significant in the same MSA cases.

In a collaborative study, by candidate variant approach we found that a variant in *LRRK2* was significantly protective of MSA. This link is promising but requires further replication.

I also studied the role of the *C9ORF72* repeat expansion in MSA as well as other related disorders. I could not detect any expansion in pathologically confirmed cases with MSA, PSP or CBD, but we found 3 large pathogenic expansions in CBS cases and one intermediate expansion or 27 repeats in a case with a clinical syndrome of PSP. This further supports the

overlapping clinical phenotypes encountered in the FTD spectrum and the importance of using pathologically confirmed cases where possible.

Finally, my main PhD project consisted in the generation and analysis of exome sequencing data in MSA. I first analysed data from an MSA family, and unfortunately, given that the family is small and some members are still young, the filtering strategies are limited and the list of genes is too long to have any conclusive finding. Later, by sequencing a large cohort of MSA cases (including 298 pathologically confirmed cases and 152 clinically diagnosed cases) I studied this population by a candidate gene approach and an association study. First, by analysis of common variation and then by looking into rare variants with a SKAT-O test. Although we were not able to replicate previous findings linking MSA to *COQ2* or *SNCA*, there are several genes of interest to follow up in future studies or with different techniques, however this will require larger collaborative efforts. A strength of this study was the large number of definite MSA cases and the stringent QC steps applied in exome wide data that allowed for a reliable dataset to be analysed.

Future directions in the research of genetic causes and risk factors of MSA require international multicentre collaborations in order to enable adequately powered studies. Genome sequencing is becoming a cost-effective technology that is currently been implemented to study the genetic background of PD, and is an excellent strategy that hopefully will be applied in MSA in the near future.

7.2 PFBC

By studying candidate genes in patients exhibiting primary brain calcification I identified a causative mutation in *SLC20A2* by Sanger sequencing. This is interesting because not only the patient was finally diagnosed after many years of follow up, but also because we detected two additional affected family members with neuropsychiatric symptoms that were never investigated before.

After investigating candidate genes in further samples, I selected a recessive family with no known mutation for further analysis. I performed homozygosity mapping and identified 3 homozygous regions in which I found a nonsense mutation in *JAM2* by WES that segregated

with the disease. I then studied another family and they carry the same mutation. A western blot showed lack of expression of the JAM2 protein in the homozygous state supporting the role of this mutation in these families. In collaboration with others I studied a JAMB-KO (ortholog of JAM2) mouse model and found gait and histopathological abnormalities in the KO mice when compared to wild type. In conclusion, I discovered the cause of brain calcification in these 2 families and supported this link with functional and animal studies.

REFERENCES

1. Pittman A, Hardy J. Genetic analysis in neurology: the next 10 years. *JAMA Neurol*. 2013 Jun;70(6):696–702.
2. Rosenberg H, Pollock N, Schiemann A, Bulger T, Stowell K. Malignant hyperthermia: a review. *Orphanet J Rare Dis*. 2015 Aug 4;10:93.
3. Doherty M, Metcalfe T, Guardino E, Peters E, Ramage L. Precision medicine and oncology: an overview of the opportunities presented by next-generation sequencing and big data and the challenges posed to conventional drug development and regulatory approval pathways. *Ann Oncol Off J Eur Soc Med Oncol*. 2016 Aug;27(8):1644–6.
4. EpiPM Consortium. A roadmap for precision medicine in the epilepsies. *Lancet Neurol*. 2015 Dec;14(12):1219–28.
5. Aiello LB, Chiatti BD. Primer in Genetics and Genomics, Article 4-Inheritance Patterns. *Biol Res Nurs*. 2017 Jan 1;1099800417708616.
6. Human molecular genetics - NLM Catalog - NCBI [Internet]. [cited 2017 Jun 14]. Available from: <https://www.ncbi.nlm.nih.gov/nlmcatalog/101523906>
7. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004 Oct 21;431(7011):931–45.
8. Singleton AB, Hardy J, Traynor BJ, Houlden H. Towards a complete resolution of the genetic architecture of disease. *Trends Genet TIG*. 2010 Oct;26(10):438–42.
9. Bras J, Guerreiro R, Hardy J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci*. 2012 Jun 20;13(7):453–64.

10. 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.
11. Gibbs JR, Singleton A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet*. 2006 Oct 6;2(10):e150.
12. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*. 1991 Feb 21;349(6311):704–6.
13. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*. 1987 Jul 31;50(3):509–17.
14. Hershenson J, Mencacci NE, Davis M, MacDonald N, Trabzuni D, Ryten M, et al. Mutations in the autoregulatory domain of β -tubulin 4a cause hereditary dystonia. *Ann Neurol*. 2013 Apr;73(4):546–53.
15. Hammer MB, Eleuch-Fayache G, Schottlaender LV, Nehdi H, Gibbs JR, Arepalli SK, et al. Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. *Am J Hum Genet*. 2013 Feb 7;92(2):245–51.
16. Kun-Rodrigues C, Ganos C, Guerreiro R, Schneider SA, Schulte C, Lesage S, et al. A systematic screening to identify de novo mutations causing sporadic early-onset Parkinson's disease. *Hum Mol Genet*. 2015 Dec 1;24(23):6711–20.
17. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. *J Autoimmun*. 2015 Nov;64:13–25.
18. 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.
19. Lees AJ, Hardy J, Revesz T. Parkinson's disease. *Lancet Lond Engl*. 2009 Jun 13;373(9680):2055–66.
20. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014 Sep;46(9):989–93.
21. Hogan DB, Fiest KM, Roberts JI, Maxwell CJ, Dykeman J, Pringsheim T, et al. The Prevalence and Incidence of Dementia with Lewy Bodies: a Systematic Review. *Can J Neurol Sci J Can Sci Neurol*. 2016 Apr;43 Suppl 1:S83-95.

22. Savica R, Grossardt BR, Bower JH, Boeve BF, Ahlskog JE, Rocca WA. Incidence of dementia with Lewy bodies and Parkinson disease dementia. *JAMA Neurol.* 2013 Nov;70(11):1396–402.
23. Ling H, O’Sullivan SS, Holton JL, Revesz T, Massey LA, Williams DR, et al. Does corticobasal degeneration exist? A clinicopathological re-evaluation. *Brain J Neurol.* 2010 Jul;133(Pt 7):2045–57.
24. Geiger JT, Ding J, Crain B, Pletnikova O, Letson C, Dawson TM, et al. Next-generation sequencing reveals substantial genetic contribution to dementia with Lewy bodies. *Neurobiol Dis.* 2016 Oct;94:55–62.
25. Atypical parkinsonism - Oxford Medicine [Internet]. [cited 2018 Jan 18]. Available from: <http://oxfordmedicine.com/view/10.1093/med/9780198705062.001.0001/med-9780198705062-chapter-4>
26. Höglinger GU, Melhem NM, Dickson DW, Sleiman PMA, Wang L-S, Klei L, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet.* 2011 Jun 19;43(7):699–705.
27. Houlden H, Baker M, Morris HR, MacDonald N, Pickering-Brown S, Adamson J, et al. Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. *Neurology.* 2001 Jun 26;56(12):1702–6.
28. Kouri N, Ross OA, Dombroski B, Younkin CS, Serie DJ, Soto-Ortolaza A, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat Commun.* 2015 Jun 16;6:7247.
29. Kouri N, Carlomagno Y, Baker M, Liesinger AM, Caselli RJ, Wszolek ZK, et al. Novel mutation in MAPT exon 13 (p.N410H) causes corticobasal degeneration. *Acta Neuropathol (Berl).* 2014 Feb;127(2):271–82.
30. Kara E, Ling H, Pittman AM, Shaw K, de Silva R, Simone R, et al. The MAPT p.A152T variant is a risk factor associated with tauopathies with atypical clinical and neuropathological features. *Neurobiol Aging.* 2012 Sep;33(9):2231.e7-2231.e14.
31. Graham JG, Oppenheimer DR. Orthostatic hypotension and nicotine sensitivity in a case of multiple system atrophy. *J Neurol Neurosurg Psychiatry.* 1969 Feb;32(1):28–34.
32. Dejerine J, Thomas AA. L’atrophie olivo-ponto-cerebelleuse. *Nouv Iconog Salpetriere.* 1900;(13):330–70.
33. SHY GM, DRAGER GA. A neurological syndrome associated with orthostatic hypotension: a clinical-pathologic study. *Arch Neurol.* 1960 May;2:511–27.

34. ADAMS R, VAN BOGAERT L, VAN DER EECKEN H. [Nigro-striate and cerebello-nigro-striate degeneration. (Clinical uniqueness and pathological variability of presenile degeneration of the extrapyramidal rigidity type)]. *Psychiatr Neurol (Basel)*. 1961;142:219–59.
35. Quinn N. Multiple system atrophy--the nature of the beast. *J Neurol Neurosurg Psychiatry*. 1989 Jun;Suppl:78–89.
36. Papp MI, Kahn JE, Lantos PL. Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy-Drager syndrome). *J Neurol Sci*. 1989 Dec;94(1–3):79–100.
37. Dickson DW, Liu W, Hardy J, Farrer M, Mehta N, Uitti R, et al. Widespread alterations of alpha-synuclein in multiple system atrophy. *Am J Pathol*. 1999 Oct;155(4):1241–51.
38. Stefanova N, Bücke P, Duerr S, Wenning GK. Multiple system atrophy: an update. *Lancet Neurol*. 2009 Dec;8(12):1172–8.
39. Spillantini MG, Crowther RA, Jakes R, Cairns NJ, Lantos PL, Goedert M. Filamentous alpha-synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. *Neurosci Lett*. 1998 Jul 31;251(3):205–8.
40. Wakabayashi K, Yoshimoto M, Tsuji S, Takahashi H. Alpha-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. *Neurosci Lett*. 1998 Jun 19;249(2–3):180–2.
41. Hague K, Lento P, Morgello S, Caro S, Kaufmann H. The distribution of Lewy bodies in pure autonomic failure: autopsy findings and review of the literature. *Acta Neuropathol (Berl)*. 1997 Aug;94(2):192–6.
42. Gilman S, Low PA, Quinn N, Albanese A, Ben-Shlomo Y, Fowler CJ, et al. Consensus statement on the diagnosis of multiple system atrophy. *J Auton Nerv Syst*. 1998 Dec 11;74(2–3):189–92.
43. Wenning GK, Ben-Shlomo Y, Magalhães M, Daniel SE, Quinn NP. Clinicopathological study of 35 cases of multiple system atrophy. *J Neurol Neurosurg Psychiatry*. 1995 Feb;58(2):160–6.
44. Federoff M, Schottlaender LV, Houlden H, Singleton A. Multiple system atrophy: the application of genetics in understanding etiology. *Clin Auton Res Off J Clin Auton Res Soc*. 2015 Feb;25(1):19–36.
45. Köllensperger M, Geser F, Ndayisaba J-P, Boesch S, Seppi K, Ostergaard K, et al. Presentation, diagnosis, and management of multiple system atrophy in Europe: final analysis of the European multiple system atrophy registry. *Mov Disord Off J Mov Disord Soc*. 2010 Nov 15;25(15):2604–12.
46. May S, Gilman S, Sowell BB, Thomas RG, Stern MB, Colcher A, et al. Potential outcome measures and trial design issues for multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2007 Dec;22(16):2371–7.

47. Watanabe H, Saito Y, Terao S, Ando T, Kachi T, Mukai E, et al. Progression and prognosis in multiple system atrophy: an analysis of 230 Japanese patients. *Brain J Neurol*. 2002 May;125(Pt 5):1070–83.
48. Yabe I, Soma H, Takei A, Fujiki N, Yanagihara T, Sasaki H. MSA-C is the predominant clinical phenotype of MSA in Japan: analysis of 142 patients with probable MSA. *J Neurol Sci*. 2006 Nov 15;249(2):115–21.
49. Kim H-J, Jeon BS, Lee J-Y, Yun JY. Survival of Korean patients with multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2011 Apr;26(5):909–12.
50. Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ, Trojanowski JQ, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*. 2008 Aug 26;71(9):670–6.
51. Wenning GK, Tison F, Ben Shlomo Y, Daniel SE, Quinn NP. Multiple system atrophy: a review of 203 pathologically proven cases. *Mov Disord Off J Mov Disord Soc*. 1997 Mar;12(2):133–47.
52. Hughes AJ, Colosimo C, Kleedorfer B, Daniel SE, Lees AJ. The dopaminergic response in multiple system atrophy. *J Neurol Neurosurg Psychiatry*. 1992 Nov;55(11):1009–13.
53. Colosimo C, Albanese A, Hughes AJ, de Bruin VM, Lees AJ. Some specific clinical features differentiate multiple system atrophy (striatonigral variety) from Parkinson's disease. *Arch Neurol*. 1995 Mar;52(3):294–8.
54. Fearnley JM, Lees AJ. Striatonigral degeneration. A clinicopathological study. *Brain J Neurol*. 1990 Dec;113 (Pt 6):1823–42.
55. Rajput AH, Rozdilsky B, Rajput A, Ang L. Levodopa efficacy and pathological basis of Parkinson syndrome. *Clin Neuropharmacol*. 1990 Dec;13(6):553–8.
56. Wüllner U, Schmitz-Hübsch T, Abele M, Antony G, Bauer P, Eggert K. Features of probable multiple system atrophy patients identified among 4770 patients with parkinsonism enrolled in the multicentre registry of the German Competence Network on Parkinson's disease. *J Neural Transm Vienna Austria 1996*. 2007 Sep;114(9):1161–5.
57. Quinn N. Disproportionate antecollis in multiple system atrophy. *Lancet*. 1989 Apr 15;1(8642):844.
58. Boesch SM, Wenning GK, Ransmayr G, Poewe W. Dystonia in multiple system atrophy. *J Neurol Neurosurg Psychiatry*. 2002 Mar;72(3):300–3.
59. Wenning GK, Ben Shlomo Y, Magalhães M, Daniel SE, Quinn NP. Clinical features and natural history of multiple system atrophy. An analysis of 100 cases. *Brain J Neurol*. 1994 Aug;117 (Pt 4):835–45.
60. Anderson T, Luxon L, Quinn N, Daniel S, Marsden CD, Bronstein A. Oculomotor function in multiple system atrophy: clinical and laboratory

features in 30 patients. *Mov Disord Off J Mov Disord Soc.* 2008 May 15;23(7):977–84.

61. Fernagut P-O, Vital A, Cannon M-H, Tison F, Meissner WG. Ambiguous mechanisms of dysphagia in multiple system atrophy. *Brain J Neurol* [Internet]. 2011 Aug 8 [cited 2011 Dec 31]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21824966>
62. Colosimo C. Nonmotor presentations of multiple system atrophy. *Nat Rev Neurol.* 2011 May;7(5):295–8.
63. Colosimo C, Morgante L, Antonini A, Barone P, Avarello TP, Bottacchi E, et al. Non-motor symptoms in atypical and secondary parkinsonism: the PRIAMO study. *J Neurol.* 2010 Jan;257(1):5–14.
64. O’Sullivan SS, Massey LA, Williams DR, Silveira-Moriyama L, Kempster PA, Holton JL, et al. Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. *Brain J Neurol.* 2008 May;131(Pt 5):1362–72.
65. Wenning GK, Stefanova N. Recent developments in multiple system atrophy. *J Neurol.* 2009 Nov;256(11):1791–808.
66. Kirchof K, Apostolidis AN, Mathias CJ, Fowler CJ. Erectile and urinary dysfunction may be the presenting features in patients with multiple system atrophy: a retrospective study. *Int J Impot Res.* 2003 Aug;15(4):293–8.
67. Wenning GK, Krismer F, Poewe W. New insights into atypical parkinsonism. *Curr Opin Neurol.* 2011 Aug;24(4):331–8.
68. Donadio V, Nolano M, Elam M, Montagna P, Provitera V, Bugiardini E, et al. Anhidrosis in multiple system atrophy: a preganglionic sudomotor dysfunction? *Mov Disord Off J Mov Disord Soc.* 2008 Apr 30;23(6):885–8.
69. Plazzi G, Corsini R, Provini F, Pierangeli G, Martinelli P, Montagna P, et al. REM sleep behavior disorders in multiple system atrophy. *Neurology.* 1997 Apr;48(4):1094–7.
70. Moreno-López C, Santamaría J, Salamero M, Del Sorbo F, Albanese A, Pellecchia MT, et al. Excessive daytime sleepiness in multiple system atrophy (SLEEMSA study). *Arch Neurol.* 2011 Feb;68(2):223–30.
71. Brown RG, Lacomblez L, Landwehrmeyer BG, Bak T, Uttner I, Dubois B, et al. Cognitive impairment in patients with multiple system atrophy and progressive supranuclear palsy. *Brain J Neurol.* 2010 Aug;133(Pt 8):2382–93.
72. Kawai Y, Suenaga M, Takeda A, Ito M, Watanabe H, Tanaka F, et al. Cognitive impairments in multiple system atrophy: MSA-C vs MSA-P. *Neurology.* 2008 Apr 15;70(16 Pt 2):1390–6.

73. Schrag A, Sheikh S, Quinn NP, Lees AJ, Selai C, Mathias C, et al. A comparison of depression, anxiety, and health status in patients with progressive supranuclear palsy and multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2010 Jun 15;25(8):1077–81.
74. Abele M, Riet A, Hummel T, Klockgether T, Wüllner U. Olfactory dysfunction in cerebellar ataxia and multiple system atrophy. *J Neurol*. 2003 Dec;250(12):1453–5.
75. Nee LE, Scott J, Polinsky RJ. Olfactory dysfunction in the Shy-Drager syndrome. *Clin Auton Res Off J Clin Auton Res Soc*. 1993 Aug;3(4):281–2.
76. Silveira-Moriyama L, Mathias C, Mason L, Best C, Quinn NP, Lees AJ. Hyposmia in pure autonomic failure. *Neurology*. 2009 May 12;72(19):1677–81.
77. Wenning GK, Shephard B, Hawkes C, Petrukevitch A, Lees A, Quinn N. Olfactory function in atypical parkinsonian syndromes. *Acta Neurol Scand*. 1995 Apr;91(4):247–50.
78. Glass PG, Lees AJ, Mathias C, Mason L, Best C, Williams DR, et al. Olfaction in pathologically proven patients with multiple system atrophy. *Mov Disord Off J Mov Disord Soc* [Internet]. 2011 Sep 27 [cited 2012 Jan 2]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21953531>
79. Ahmed Z, Asi YT, Sailer A, Lees AJ, Houlden H, Revesz T, et al. Review: The neuropathology, pathophysiology and genetics of multiple system atrophy. *Neuropathol Appl Neurobiol* [Internet]. 2011 Nov 10 [cited 2012 Jan 3]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22074330>
80. Ozawa T, Paviour D, Quinn NP, Josephs KA, Sangha H, Kilford L, et al. The spectrum of pathological involvement of the striatonigral and olivopontocerebellar systems in multiple system atrophy: clinicopathological correlations. *Brain J Neurol*. 2004 Dec;127(Pt 12):2657–71.
81. Spargo E, Papp MI, Lantos PL. Decrease in neuronal density in the cerebral cortex in multiple system atrophy. *Eur J Neurol*. 1996 Sep 1;3(5):450–6.
82. Tsuchiya K, Ozawa E, Haga C, Watabiki S, Ikeda M, Sano M, et al. Constant involvement of the Betz cells and pyramidal tract in multiple system atrophy: a clinicopathological study of seven autopsy cases. *Acta Neuropathol (Berl)*. 2000 Jun;99(6):628–36.
83. Song YJC, Halliday GM, Holton JL, Lashley T, O’Sullivan SS, McCann H, et al. Degeneration in different parkinsonian syndromes relates to astrocyte type and astrocyte protein expression. *J Neuropathol Exp Neurol*. 2009 Oct;68(10):1073–83.

84. Jellinger KA, Lantos PL. Papp-Lantos inclusions and the pathogenesis of multiple system atrophy: an update. *Acta Neuropathol (Berl)*. 2010 Jun;119(6):657–67.
85. Trojanowski JQ, Revesz T. Proposed neuropathological criteria for the post mortem diagnosis of multiple system atrophy. *Neuropathol Appl Neurobiol*. 2007 Dec;33(6):615–20.
86. Wenning GK, Quinn N, Magalhães M, Mathias C, Daniel SE. “Minimal change” multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 1994 Mar;9(2):161–6.
87. Schottlaender LV, Sailer A, Ahmed Z, Dickson DW, Houlden H, Ross OA. Multiple System Atrophy: Clinical, Genetics, and Neuropathology. In: FMedSci ASM DSc, FRCP, MD ZW, MD TMD, FMedSci NWF, editors. *Neurodegeneration* [Internet]. John Wiley & Sons, Ltd; 2017 [cited 2017 Nov 10]. p. 58–71. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9781118661895.ch7/summary>
88. Solano SM, Miller DW, Augood SJ, Young AB, Penney JB Jr. Expression of alpha-synuclein, parkin, and ubiquitin carboxy-terminal hydrolase L1 mRNA in human brain: genes associated with familial Parkinson’s disease. *Ann Neurol*. 2000 Feb;47(2):201–10.
89. Palma J-A, Kaufmann H. Novel therapeutic approaches in multiple system atrophy. *Clin Auton Res Off J Clin Auton Res Soc*. 2015 Feb;25(1):37–45.
90. Miller DW, Johnson JM, Solano SM, Hollingsworth ZR, Standaert DG, Young AB. Absence of alpha-synuclein mRNA expression in normal and multiple system atrophy oligodendroglia. *J Neural Transm Vienna Austria* 1996. 2005 Dec;112(12):1613–24.
91. Wenning GK, Stefanova N, Jellinger KA, Poewe W, Schlossmacher MG. Multiple system atrophy: a primary oligodendroglialopathy. *Ann Neurol*. 2008 Sep;64(3):239–46.
92. Jellinger KA. P25alpha immunoreactivity in multiple system atrophy and Parkinson disease. *Acta Neuropathol (Berl)*. 2006 Jul;112(1):112.
93. Kovács GG, Gelpi E, Lehotzky A, Höftberger R, Erdei A, Budka H, et al. The brain-specific protein TPPP/p25 in pathological protein deposits of neurodegenerative diseases. *Acta Neuropathol (Berl)*. 2007 Feb;113(2):153–61.
94. Lindersson E, Lundvig D, Petersen C, Madsen P, Nyengaard JR, Højrup P, et al. p25alpha Stimulates alpha-synuclein aggregation and is colocalized with aggregated alpha-synuclein in alpha-synucleinopathies. *J Biol Chem*. 2005 Feb 18;280(7):5703–15.
95. Song YJC, Lundvig DMS, Huang Y, Gai WP, Blumbergs PC, Højrup P, et al. p25alpha relocates in oligodendroglia from myelin to

- cytoplasmic inclusions in multiple system atrophy. *Am J Pathol.* 2007 Oct;171(4):1291–303.
96. Stefanova N, Tison F, Reindl M, Poewe W, Wenning GK. Animal models of multiple system atrophy. *Trends Neurosci.* 2005 Sep;28(9):501–6.
 97. Stefanova N, Reindl M, Neumann M, Kahle PJ, Poewe W, Wenning GK. Microglial activation mediates neurodegeneration related to oligodendroglial alpha-synucleinopathy: implications for multiple system atrophy. *Mov Disord Off J Mov Disord Soc.* 2007 Nov 15;22(15):2196–203.
 98. Wakabayashi K, Mori F, Nishie M, Oyama Y, Kurihara A, Yoshimoto M, et al. An autopsy case of early (“minimal change”) olivopontocerebellar atrophy (multiple system atrophy-cerebellar). *Acta Neuropathol (Berl).* 2005 Aug;110(2):185–90.
 99. Yoshida M. Multiple system atrophy: alpha-synuclein and neuronal degeneration. *Neuropathol Off J Jpn Soc Neuropathol.* 2007 Oct;27(5):484–93.
 100. Nishie M, Mori F, Yoshimoto M, Takahashi H, Wakabayashi K. A quantitative investigation of neuronal cytoplasmic and intranuclear inclusions in the pontine and inferior olivary nuclei in multiple system atrophy. *Neuropathol Appl Neurobiol.* 2004 Oct;30(5):546–54.
 101. Hara K, Momose Y, Tokiguchi S, Shimohata M, Terajima K, Onodera O, et al. Multiplex families with multiple system atrophy. *Arch Neurol.* 2007 Apr;64(4):545–51.
 102. Hohler AD, Singh VJ. Probable hereditary multiple system atrophy-autonomic (MSA-A) in a family in the United States. *J Clin Neurosci Off J Neurosurg Soc Australas.* 2012 Mar;19(3):479–80.
 103. Wüllner U, Schmitt I, Kammal M, Kretschmar HA, Neumann M. Definite multiple system atrophy in a German family. *J Neurol Neurosurg Psychiatry.* 2009 Apr;80(4):449–50.
 104. Itoh K, Kasai T, Tsuji Y, Saito K, Mizuta I, Harada Y, et al. Definite familial multiple system atrophy with unknown genetics. *Neuropathol Off J Jpn Soc Neuropathol.* 2014 Jun;34(3):309–13.
 105. Vidal J-S, Vidailhet M, Derkinderen P, Tzourio C, Alperovitch A. Familial aggregation in atypical Parkinson’s disease: a case control study in multiple system atrophy and progressive supranuclear palsy. *J Neurol.* 2010 Aug;257(8):1388–93.
 106. Stemberger S, Scholz SW, Singleton AB, Wenning GK. Genetic players in multiple system atrophy: unfolding the nature of the beast. *Neurobiol Aging.* 2011 Oct;32(10):1924.e5-14.

107. Nirenberg MJ, Libien J, Vonsattel J-P, Fahn S. Multiple system atrophy in a patient with the spinocerebellar ataxia 3 gene mutation. *Mov Disord Off J Mov Disord Soc.* 2007 Jan 15;22(2):251–4.
108. Bandmann O, Sweeney MG, Daniel SE, Wenning GK, Quinn N, Marsden CD, et al. Multiple-system atrophy is genetically distinct from identified inherited causes of spinocerebellar degeneration. *Neurology.* 1997 Dec;49(6):1598–604.
109. Berciano J, Ferrer I. Glial and neuronal cytoplasmic inclusions in familial olivopontocerebellar atrophy. *Ann Neurol.* 1996 Nov;40(5):819–20.
110. Berciano J, Ferrer I. Glial cell cytoplasmic inclusions in SCA2 do not express alpha-synuclein. *J Neurol.* 2005 Jun;252(6):742–4.
111. Cho JW, Kim SY, Park SS, Jeon BS. Spinocerebellar ataxia type 12 was not found in Korean Parkinsonian patients. *Can J Neurol Sci J Can Sci Neurol.* 2008 Sep;35(4):488–90.
112. Gilman S, Sima AA, Junck L, Kluin KJ, Koeppe RA, Lohman ME, et al. Spinocerebellar ataxia type 1 with multiple system degeneration and glial cytoplasmic inclusions. *Ann Neurol.* 1996 Feb;39(2):241–55.
113. Factor SA, Qian J, Lava NS, Hubbard JD, Payami H. False-positive SCA8 gene test in a patient with pathologically proven multiple system atrophy. *Ann Neurol.* 2005 Mar;57(3):462–3.
114. Kamm C, Healy DG, Quinn NP, Wüllner U, Moller JC, Schols L, et al. The fragile X tremor ataxia syndrome in the differential diagnosis of multiple system atrophy: data from the EMSA Study Group. *Brain J Neurol.* 2005 Aug;128(Pt 8):1855–60.
115. Yabe I, Soma H, Takei A, Fujik N, Sasaki H. No association between FMR1 premutations and multiple system atrophy. *J Neurol.* 2004 Nov;251(11):1411–2.
116. Al-Chalabi A, Dürr A, Wood NW, Parkinson MH, Camuzat A, Hulot J-S, et al. Genetic variants of the alpha-synuclein gene SNCA are associated with multiple system atrophy. *PloS One.* 2009 Sep 22;4(9):e7114.
117. Scholz SW, Houlden H, Schulte C, Sharma M, Li A, Berg D, et al. SNCA variants are associated with increased risk for multiple system atrophy. *Ann Neurol.* 2009 May;65(5):610–4.
118. Ross OA, Vilariño-Güell C, Wszolek ZK, Farrer MJ, Dickson DW. Reply to: SNCA variants are associated with increased risk of multiple system atrophy. *Ann Neurol.* 2010 Mar;67(3):414–5.
119. Lincoln SJ, Ross OA, Milkovic NM, Dickson DW, Rajput A, Robinson CA, et al. Quantitative PCR-based screening of alpha-synuclein multiplication in multiple system atrophy. *Parkinsonism Relat Disord.* 2007 Aug;13(6):340–2.

120. Morris HR, Vaughan JR, Datta SR, Bandopadhyay R, Rohan De Silva HA, Schrag A, et al. Multiple system atrophy/progressive supranuclear palsy: alpha-Synuclein, synphilin, tau, and APOE. *Neurology*. 2000 Dec 26;55(12):1918–20.
121. Ozawa T, Healy DG, Abou-Sleiman PM, Ahmadi KR, Quinn N, Lees AJ, et al. The alpha-synuclein gene in multiple system atrophy. *J Neurol Neurosurg Psychiatry*. 2006 Apr;77(4):464–7.
122. Ozawa T, Takano H, Onodera O, Kobayashi H, Ikeuchi T, Koide R, et al. No mutation in the entire coding region of the alpha-synuclein gene in pathologically confirmed cases of multiple system atrophy. *Neurosci Lett*. 1999 Jul 30;270(2):110–2.
123. Yun JY, Lee W-W, Lee J-Y, Kim HJ, Park SS, Jeon BS. SNCA variants and multiple system atrophy. *Ann Neurol*. 2010 Apr;67(4):554–5.
124. Sailer A, Scholz SW, Nalls MA, Schulte C, Federoff M, Price TR, et al. A genome-wide association study in multiple system atrophy. *Neurology*. 2016 Oct 11;87(15):1591–8.
125. Vilariño-Güell C, Soto-Ortolaza AI, Rajput A, Mash DC, Papapetropoulos S, Pahwa R, et al. MAPT H1 haplotype is a risk factor for essential tremor and multiple system atrophy. *Neurology*. 2011 Feb 15;76(7):670–2.
126. Labbé C, Heckman MG, Lorenzo-Betancor O, Murray ME, Ogaki K, Soto-Ortolaza AI, et al. MAPT haplotype diversity in multiple system atrophy. *Parkinsonism Relat Disord*. 2016 Sep;30:40–5.
127. Nicholl DJ, Bennett P, Hiller L, Bonifati V, Vanacore N, Fabbrini G, et al. A study of five candidate genes in Parkinson's disease and related neurodegenerative disorders. European Study Group on Atypical Parkinsonism. *Neurology*. 1999 Oct 22;53(7):1415–21.
128. Mitsui J, Matsukawa T, Sasaki H, Yabe I, Matsushima M, Dürr A, et al. Variants associated with Gaucher disease in multiple system atrophy. *Ann Clin Transl Neurol*. 2015 Apr;2(4):417–26.
129. Goker-Alpan O, Giasson BI, Eblan MJ, Nguyen J, Hurtig HI, Lee VM-Y, et al. Glucocerebrosidase mutations are an important risk factor for Lewy body disorders. *Neurology*. 2006 Sep 12;67(5):908–10.
130. Segarane B, Li A, Paudel R, Scholz S, Neumann J, Lees A, et al. Glucocerebrosidase mutations in 108 neuropathologically confirmed cases of multiple system atrophy. *Neurology*. 2009 Mar 31;72(13):1185–6.
131. Sun Q, Guo J, Han W, Zuo X, Wang L, Yao L, et al. Genetic association study of glucocerebrosidase gene L444P mutation in essential tremor and multiple system atrophy in mainland China. *J Clin Neurosci Off J Neurosurg Soc Australas*. 2013 Feb;20(2):217–9.

132. Asselta R, Rimoldi V, Siri C, Cilia R, Guella I, Tesei S, et al. Glucocerebrosidase mutations in primary parkinsonism. *Parkinsonism Relat Disord.* 2014 Nov;20(11):1215–20.
133. Bandmann O, Wenning GK, Quinn NP, Harding AE. Arg296 to Cys296 polymorphism in exon 6 of cytochrome P-450-2D6 (CYP2D6) is not associated with multiple system atrophy. *J Neurol Neurosurg Psychiatry.* 1995 Nov;59(5):557.
134. Iwahashi K, Miyatake R, Tsuneoka Y, Matsuo Y, Ichikawa Y, Hosokawa K, et al. A novel cytochrome P-450IID6 (CYPIID6) mutant gene associated with multiple system atrophy. *J Neurol Neurosurg Psychiatry.* 1995 Feb;58(2):263–4.
135. Cho J-W, Kim S-Y, Park S-S, Jeon BS. The G2019S LRRK2 Mutation is Rare in Korean Patients with Parkinson's Disease and Multiple System Atrophy. *J Clin Neurol Seoul Korea.* 2009 Mar;5(1):29–32.
136. Ozelius LJ, Foroud T, May S, Senthil G, Sandroni P, Low PA, et al. G2019S mutation in the leucine-rich repeat kinase 2 gene is not associated with multiple system atrophy. *Mov Disord Off J Mov Disord Soc.* 2007 Mar 15;22(4):546–9.
137. Ross OA, Toft M, Whittle AJ, Johnson JL, Papapetropoulos S, Mash DC, et al. Lrrk2 and Lewy body disease. *Ann Neurol.* 2006 Feb;59(2):388–93.
138. Tan EK, Skipper L, Chua E, Wong M-C, Pavanni R, Bonnard C, et al. Analysis of 14 LRRK2 mutations in Parkinson's plus syndromes and late-onset Parkinson's disease. *Mov Disord Off J Mov Disord Soc.* 2006 Jul;21(7):997–1001.
139. Brooks JA, Houlden H, Melchers A, Islam AJ, Ding J, Li A, et al. Mutational analysis of parkin and PINK1 in multiple system atrophy. *Neurobiol Aging.* 2011 Mar;32(3):548.e5-7.
140. Soma H, Yabe I, Takei A, Fujiki N, Yanagihara T, Sasaki H. Associations between multiple system atrophy and polymorphisms of SLC1A4, SQSTM1, and EIF4EBP1 genes. *Mov Disord Off J Mov Disord Soc.* 2008 Jun 15;23(8):1161–7.
141. Combarros O, Infante J, Llorca J, Berciano J. Interleukin-1A (-889) genetic polymorphism increases the risk of multiple system atrophy. *Mov Disord Off J Mov Disord Soc.* 2003 Nov;18(11):1385–6.
142. Nishimura M, Kawakami H, Komure O, Maruyama H, Morino H, Izumi Y, et al. Contribution of the interleukin-1beta gene polymorphism in multiple system atrophy. *Mov Disord Off J Mov Disord Soc.* 2002 Jul;17(4):808–11.
143. Infante J, Llorca J, Berciano J, Combarros O. Interleukin-8, intercellular adhesion molecule-1 and tumour necrosis factor-alpha gene

- polymorphisms and the risk for multiple system atrophy. *J Neurol Sci*. 2005 Jan 15;228(1):11–3.
144. Furiya Y, Hirano M, Kurumatani N, Nakamuro T, Matsumura R, Futamura N, et al. Alpha-1-antichymotrypsin gene polymorphism and susceptibility to multiple system atrophy (MSA). *Brain Res Mol Brain Res*. 2005 Aug 18;138(2):178–81.
 145. Schmitt I, Wüllner U, Healy DG, Wood NW, Kölsch H, Heun R. The ADH1C stop mutation in multiple system atrophy patients and healthy probands in the United Kingdom and Germany. *Mov Disord Off J Mov Disord Soc*. 2006 Nov;21(11):2034.
 146. Buervenich S, Sydow O, Carmine A, Zhang Z, Anvret M, Olson L. Alcohol dehydrogenase alleles in Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2000 Sep;15(5):813–8.
 147. Zhang J, Montine TJ, Smith MA, Siedlak SL, Gu G, Robertson D, et al. The mitochondrial common deletion in Parkinson's disease and related movement disorders. *Parkinsonism Relat Disord*. 2002 Jan;8(3):165–70.
 148. Sasaki H, Emi M, Iijima H, Ito N, Sato H, Yabe I, et al. Copy number loss of (src homology 2 domain containing)-transforming protein 2 (SHC2) gene: discordant loss in monozygotic twins and frequent loss in patients with multiple system atrophy. *Mol Brain*. 2011 Jun 10;4:24.
 149. Ferguson MC, Garland EM, Hedges L, Womack-Nunley B, Hamid R, Phillips JA, et al. SHC2 gene copy number in multiple system atrophy (MSA). *Clin Auton Res Off J Clin Auton Res Soc*. 2014 Feb;24(1):25–30.
 150. Goldman JS, Quinzii C, Dunning-Broadbent J, Waters C, Mitsumoto H, Brannagan TH 3rd, et al. Multiple System Atrophy and Amyotrophic Lateral Sclerosis in a Family With Hexanucleotide Repeat Expansions in C9orf72. *JAMA Neurol*. 2014 Apr 14;
 151. Lim S-Y, Wadia P, Wenning GK, Lang AE. Clinically probable multiple system atrophy with predominant parkinsonism associated with myotonic dystrophy type 2. *Mov Disord Off J Mov Disord Soc*. 2009 Jul 15;24(9):1407–9.
 152. Yu C-E, Bird TD, Bekris LM, Montine TJ, Leverenz JB, Steinbart E, et al. The spectrum of mutations in progranulin: a collaborative study screening 545 cases of neurodegeneration. *Arch Neurol*. 2010 Feb;67(2):161–70.
 153. Nishimura M, Kuno S, Kaji R, Kawakami H. Brain-derived neurotrophic factor gene polymorphisms in Japanese patients with sporadic Alzheimer's disease, Parkinson's disease, and multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2005 Aug;20(8):1031–3.

154. Multiple-System Atrophy Research Collaboration. Mutations in COQ2 in familial and sporadic multiple-system atrophy. *N Engl J Med*. 2013 Jul 18;369(3):233–44.
155. Sun Z, Ohta Y, Yamashita T, Sato K, Takemoto M, Hishikawa N, et al. New susceptible variant of COQ2 gene in Japanese patients with sporadic multiple system atrophy. *Neurol Genet*. 2016 Apr;2(2):e54.
156. Zhao Q, Yang X, Tian S, An R, Zheng J, Xu Y. Association of the COQ2 V393A variant with risk of multiple system atrophy in East Asians: a case-control study and meta-analysis of the literature. *Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol*. 2016 Mar;37(3):423–30.
157. Lin C-H, Lin H-I, Chen M-L, Wu R-M. COQ2 p.V393A variant, rs148156462, is not associated with Parkinson's disease in a Taiwanese population. *Neurobiol Aging*. 2015 Jan;36(1):546.e17-18.
158. Chen YP, Zhao B, Cao B, Song W, Guo X, Wei Q-Q, et al. Mutation scanning of the COQ2 gene in ethnic Chinese patients with multiple-system atrophy. *Neurobiol Aging*. 2015 Feb;36(2):1222.e7-11.
159. Ogaki K, Fujioka S, Heckman MG, Rayaprolu S, Soto-Ortolaza AI, Labbé C, et al. Analysis of COQ2 gene in multiple system atrophy. *Mol Neurodegener*. 2014 Nov 5;9:44.
160. Wen X-D, Li H-F, Wang H-X, Ni W, Dong Y, Wu Z-Y. Mutation Analysis of COQ2 in Chinese Patients with Cerebellar Subtype of Multiple System Atrophy. *CNS Neurosci Ther*. 2015 Aug;21(8):626–30.
161. Jeon BS, Farrer MJ, Bortnick SF, Korean Canadian Alliance on Parkinson's Disease and Related Disorders. Mutant COQ2 in multiple-system atrophy. *N Engl J Med*. 2014 Jul 3;371(1):80.
162. Sharma M, Wenning G, Krüger R, European Multiple-System Atrophy Study Group (EMSA-SG). Mutant COQ2 in multiple-system atrophy. *N Engl J Med*. 2014 Jul 3;371(1):80–1.
163. Schottlaender LV, Houlden H, Multiple-System Atrophy (MSA) Brain Bank Collaboration. Mutant COQ2 in multiple-system atrophy. *N Engl J Med*. 2014 Jul 3;371(1):81.
164. Healy DG, Abou-Sleiman PM, Quinn N, Ahmadi KR, Ozawa T, Kamm C, et al. UCHL-1 gene in multiple system atrophy: a haplotype tagging approach. *Mov Disord Off J Mov Disord Soc*. 2005 Oct;20(10):1338–43.
165. Shibao C, Garland EM, Gamboa A, Vnencak-Jones CL, Van Woeltz M, Haines JL, et al. PRNP M129V homozygosity in multiple system atrophy vs. Parkinson's disease. *Clin Auton Res Off J Clin Auton Res Soc*. 2008 Feb;18(1):13–9.

166. Chelban V, Manole A, Pihlstrøm L, Schottlaender L, Efthymiou S, OConnor E, et al. Analysis of the prion protein gene in multiple system atrophy. *Neurobiol Aging*. 2016 Oct 3;
167. Cho S, Kim C-H, Cubells JF, Zabetian CP, Hwang D-Y, Kim J-W, et al. Variations in the dopamine beta-hydroxylase gene are not associated with the autonomic disorders, pure autonomic failure, or multiple system atrophy. *Am J Med Genet A*. 2003 Jul 15;120A(2):234–6.
168. Sailer A, Scholz SW, Nalls MA, Schulte C, Federoff M, Price TR, et al. A genome-wide association study in multiple system atrophy. *Neurology*. 2016 Oct 11;87(15):1591–8.
169. Ross OA, Braithwaite AT, Skipper LM, Kachergus J, Hulihan MM, Middleton FA, et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Ann Neurol*. 2008 Jun;63(6):743–50.
170. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science*. 2003 Oct 31;302(5646):841.
171. Ozawa T, Takano H, Onodera O, Kobayashi H, Ikeuchi T, Koide R, et al. No mutation in the entire coding region of the alpha-synuclein gene in pathologically confirmed cases of multiple system atrophy. *Neurosci Lett*. 1999 Jul 30;270(2):110–2.
172. Lee PH, Kim JW, Bang OY, Ahn YH, Joo IS, Huh K. Autologous mesenchymal stem cell therapy delays the progression of neurological deficits in patients with multiple system atrophy. *Clin Pharmacol Ther*. 2008 May;83(5):723–30.
173. Lee PH, Lee G, Park HJ, Bang OY, Joo IS, Huh K. The plasma alpha-synuclein levels in patients with Parkinson's disease and multiple system atrophy. *J Neural Transm Vienna Austria* 1996. 2006 Oct;113(10):1435–9.
174. Ozawa T, Okuizumi K, Ikeuchi T, Wakabayashi K, Takahashi H, Tsuji S. Analysis of the expression level of alpha-synuclein mRNA using postmortem brain samples from pathologically confirmed cases of multiple system atrophy. *Acta Neuropathol (Berl)*. 2001 Aug;102(2):188–90.
175. Vogt IR, Lees AJ, Evert BO, Klockgether T, Bonin M, Wüllner U. Transcriptional changes in multiple system atrophy and Parkinson's disease putamen. *Exp Neurol*. 2006 Jun;199(2):465–78.
176. Scholz SW, Houlden H, Schulte C, Sharma M, Li A, Berg D, et al. SNCA variants are associated with increased risk for multiple system atrophy. *Ann Neurol*. 2009 May;65(5):610–4.
177. Kiely AP, Asi YT, Kara E, Limousin P, Ling H, Lewis P, et al. α -Synucleinopathy associated with G51D SNCA mutation: a link between

Parkinson's disease and multiple system atrophy? *Acta Neuropathol (Berl)*. 2013 May;125(5):753–69.

178. Pasanen P, Myllykangas L, Siitonen M, Raunio A, Kaakkola S, Lyytinen J, et al. Novel α -synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol Aging*. 2014 Sep;35(9):2180.e1-5.
179. Duncan AJ, Bitner-Glindzicz M, Meunier B, Costello H, Hargreaves IP, López LC, et al. A nonsense mutation in COQ9 causes autosomal-recessive neonatal-onset primary coenzyme Q10 deficiency: a potentially treatable form of mitochondrial disease. *Am J Hum Genet*. 2009 May;84(5):558–66.
180. Emmanuele V, López LC, López L, Berardo A, Naini A, Tadesse S, et al. Heterogeneity of coenzyme Q10 deficiency: patient study and literature review. *Arch Neurol*. 2012 Aug;69(8):978–83.
181. Turunen M, Olsson J, Dallner G. Metabolism and function of coenzyme Q. *Biochim Biophys Acta*. 2004 Jan 28;1660(1–2):171–99.
182. Quinzii C, Naini A, Salviati L, Trevisson E, Navas P, Dimauro S, et al. A mutation in para-hydroxybenzoate-polyprenyl transferase (COQ2) causes primary coenzyme Q10 deficiency. *Am J Hum Genet*. 2006 Feb;78(2):345–9.
183. Mollet J, Giurgea I, Schlemmer D, Dallner G, Chretien D, Delahodde A, et al. Prenyldiphosphate synthase, subunit 1 (PDSS1) and OH-benzoate polyprenyltransferase (COQ2) mutations in ubiquinone deficiency and oxidative phosphorylation disorders. *J Clin Invest*. 2007 Mar;117(3):765–72.
184. Diomedi-Camassei F, Di Giandomenico S, Santorelli FM, Caridi G, Piemonte F, Montini G, et al. COQ2 nephropathy: a newly described inherited mitochondriopathy with primary renal involvement. *J Am Soc Nephrol JASN*. 2007 Oct;18(10):2773–80.
185. The Multiple-System Atrophy Research Collaboration. Mutations in COQ2 in Familial and Sporadic Multiple-System Atrophy. *N Engl J Med*. 2013 Jun 12;
186. Nagaishi M, Yokoo H, Nakazato Y. Tau-positive glial cytoplasmic granules in multiple system atrophy. *Neuropathol Off J Jpn Soc Neuropathol*. 2011 Jun;31(3):299–305.
187. Seppi K, Poewe W. Brain magnetic resonance imaging techniques in the diagnosis of parkinsonian syndromes. *Neuroimaging Clin N Am*. 2010 Feb;20(1):29–55.
188. Courbon F, Brefel-Courbon C, Thalamas C, Alibelli M-J, Berry I, Montastruc J-L, et al. Cardiac MIBG scintigraphy is a sensitive tool for detecting cardiac sympathetic denervation in Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2003 Aug;18(8):890–7.

189. Nagayama H, Hamamoto M, Ueda M, Nagashima J, Katayama Y. Reliability of MIBG myocardial scintigraphy in the diagnosis of Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2005 Feb;76(2):249–51.
190. Raffel DM, Koeppe RA, Little R, Wang C-N, Liu S, Junck L, et al. PET measurement of cardiac and nigrostriatal denervation in Parkinsonian syndromes. *J Nucl Med Off Publ Soc Nucl Med*. 2006 Nov;47(11):1769–77.
191. Seppi K, Schocke MFH, Wenning GK, Poewe W. How to diagnose MSA early: the role of magnetic resonance imaging. *J Neural Transm Vienna Austria 1996*. 2005 Dec;112(12):1625–34.
192. Gilman S. Functional imaging with positron emission tomography in multiple system atrophy. *J Neural Transm Vienna Austria 1996*. 2005 Dec;112(12):1647–55.
193. Gilman S, Low P, Quinn N, Albanese A, Ben-Shlomo Y, Fowler C, et al. Consensus statement on the diagnosis of multiple system atrophy. American Autonomic Society and American Academy of Neurology. *Clin Auton Res Off J Clin Auton Res Soc*. 1998 Dec;8(6):359–62.
194. Quinn NP. How to diagnose multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2005 Aug;20 Suppl 12:S5–10.
195. Horimoto Y, Matsumoto M, Akatsu H, Ikari H, Kojima K, Yamamoto T, et al. Autonomic dysfunctions in dementia with Lewy bodies. *J Neurol*. 2003 May;250(5):530–3.
196. Gilman S, Little R, Johanns J, Heumann M, Kluin KJ, Junck L, et al. Evolution of sporadic olivopontocerebellar atrophy into multiple system atrophy. *Neurology*. 2000 Aug 22;55(4):527–32.
197. Kaufmann H, Norcliffe-Kaufmann L, Palma J-A, Biaggioni I, Low PA, Singer W, et al. Natural history of pure autonomic failure: A United States prospective cohort. *Ann Neurol*. 2017 Feb;81(2):287–97.
198. Geser F, Wenning GK, Seppi K, Stampfer-Kountchev M, Scherfler C, Sawires M, et al. Progression of multiple system atrophy (MSA): a prospective natural history study by the European MSA Study Group (EMSA SG). *Mov Disord Off J Mov Disord Soc*. 2006 Feb;21(2):179–86.
199. Schrag A, Wenning GK, Quinn N, Ben-Shlomo Y. Survival in multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2008 Jan 30;23(2):294–6.
200. Ben-Shlomo Y, Wenning GK, Tison F, Quinn NP. Survival of patients with pathologically proven multiple system atrophy: a meta-analysis. *Neurology*. 1997 Feb;48(2):384–93.
201. Chrysostome V, Tison F, Yekhle F, Sourgen C, Baldi I, Dartigues JF. Epidemiology of multiple system atrophy: a prevalence and pilot risk

- factor study in Aquitaine, France. *Neuroepidemiology*. 2004 Aug;23(4):201–8.
202. Tison F, Yekhlief F, Chrysostome V, Sourgen C. Prevalence of multiple system atrophy. *Lancet*. 2000 Feb 5;355(9202):495–6.
203. Schrag A, Ben-Shlomo Y, Quinn NP. Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. *Lancet*. 1999 Nov 20;354(9192):1771–5.
204. Seo J-H, Yong SW, Song SK, Lee JE, Sohn YH, Lee PH. A case-control study of multiple system atrophy in Korean patients. *Mov Disord Off J Mov Disord Soc*. 2010 Sep 15;25(12):1953–9.
205. Tada M, Onodera O, Tada M, Ozawa T, Piao Y-S, Kakita A, et al. Early development of autonomic dysfunction may predict poor prognosis in patients with multiple system atrophy. *Arch Neurol*. 2007 Feb;64(2):256–60.
206. Silber MH, Levine S. Stridor and death in multiple system atrophy. *Mov Disord Off J Mov Disord Soc*. 2000 Jul;15(4):699–704.
207. Papapetropoulos S, Tuchman A, Laufer D, Papatsoris AG, Papapetropoulos N, Mash DC. Causes of death in multiple system atrophy. *J Neurol Neurosurg Psychiatry*. 2007 Mar;78(3):327–9.
208. Shimohata T, Ozawa T, Nakayama H, Tomita M, Shinoda H, Nishizawa M. Frequency of nocturnal sudden death in patients with multiple system atrophy. *J Neurol*. 2008 Oct;255(10):1483–5.
209. Tada M, Kakita A, Toyoshima Y, Onodera O, Ozawa T, Morita T, et al. Depletion of medullary serotonergic neurons in patients with multiple system atrophy who succumbed to sudden death. *Brain J Neurol*. 2009 Jul;132(Pt 7):1810–9.
210. Flabeau O, Meissner WG, Tison F. Multiple system atrophy: current and future approaches to management. *Ther Adv Neurol Disord*. 2010 Jul;3(4):249–63.
211. Jain S, Dawson J, Quinn NP, Playford ED. Occupational therapy in multiple system atrophy: a pilot randomized controlled trial. *Mov Disord Off J Mov Disord Soc*. 2004 Nov;19(11):1360–4.
212. Wedge F. The impact of resistance training on balance and functional ability of a patient with multiple system atrophy. *J Geriatr Phys Ther* 2001. 2008;31(2):79–83.
213. Winter Y, Spottke AE, Stamelou M, Cabanel N, Eggert K, Höglinger GU, et al. Health-related quality of life in multiple system atrophy and progressive supranuclear palsy. *Neurodegener Dis*. 2011;8(6):438–46.
214. Shih LC, Tarsy D. Deep brain stimulation for the treatment of atypical parkinsonism. *Mov Disord Off J Mov Disord Soc*. 2007 Nov 15;22(15):2149–55.

215. Geser F, Seppi K, Stampfer-Kountchev M, Köllensperger M, Diem A, Ndayisaba JP, et al. The European Multiple System Atrophy-Study Group (EMSA-SG). *J Neural Transm Vienna Austria* 1996. 2005 Dec;112(12):1677–86.
216. Gilman S, May SJ, Shults CW, Tanner CM, Kukull W, Lee VM-Y, et al. The North American Multiple System Atrophy Study Group. *J Neural Transm Vienna Austria* 1996. 2005 Dec;112(12):1687–94.
217. Holmberg B, Johansson J-O, Poewe W, Wenning G, Quinn NP, Mathias C, et al. Safety and tolerability of growth hormone therapy in multiple system atrophy: a double-blind, placebo-controlled study. *Mov Disord Off J Mov Disord Soc.* 2007 Jun 15;22(8):1138–44.
218. Bensimon G, Ludolph A, Agid Y, Vidailhet M, Payan C, Leigh PN. Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: the NNIPPS study. *Brain J Neurol.* 2009 Jan;132(Pt 1):156–71.
219. Seppi K, Peralta C, Diem-Zangerl A, Puschban Z, Mueller J, Poewe W, et al. Placebo-controlled trial of riluzole in multiple system atrophy. *Eur J Neurol Off J Eur Fed Neurol Soc.* 2006 Oct;13(10):1146–8.
220. Dodel R, Spottke A, Gerhard A, Reuss A, Reinecker S, Schimke N, et al. Minocycline 1-year therapy in multiple-system-atrophy: effect on clinical symptoms and [(11)C] (R)-PK11195 PET (MEMSA-trial). *Mov Disord Off J Mov Disord Soc.* 2010 Jan 15;25(1):97–107.
221. Heo J-H, Lee S-T, Chu K, Kim M. The efficacy of combined estrogen and buspirone treatment in olivopontocerebellar atrophy. *J Neurol Sci.* 2008 Aug 15;271(1–2):87–90.
222. Poewe W, Seppi K, Fitzner-Attas CJ, Wenning GK, Gilman S, Low PA, et al. Efficacy of rasagiline in patients with the parkinsonian variant of multiple system atrophy: a randomised, placebo-controlled trial. *Lancet Neurol.* 2015 Feb;14(2):145–52.
223. Low PA, Robertson D, Gilman S, Kaufmann H, Singer W, Biaggioni I, et al. Efficacy and safety of rifampicin for multiple system atrophy: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* 2014 Mar;13(3):268–75.
224. Vidal J-S, Vidailhet M, Elbaz A, Derkinderen P, Tzourio C, Alperovitch A. Risk factors of multiple system atrophy: a case-control study in French patients. *Mov Disord Off J Mov Disord Soc.* 2008 Apr 30;23(6):797–803.
225. Stefanova N, Georgievska B, Eriksson H, Poewe W, Wenning GK. Myeloperoxidase Inhibition Ameliorates Multiple System Atrophy-Like Degeneration in a Transgenic Mouse Model. *Neurotox Res [Internet].* 2011 Dec 8 [cited 2012 Jan 5]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22161470>

226. Kaindlstorfer C, Sommer P, Georgievska B, Mather RJ, Kugler AR, Poewe W, et al. Failure of Neuroprotection Despite Microglial Suppression by Delayed-Start Myeloperoxidase Inhibition in a Model of Advanced Multiple System Atrophy: Clinical Implications. *Neurotox Res*. 2015 Oct;28(3):185–94.
227. Köllensperger M, Stefanova N, Pallua A, Puschban Z, Dechant G, Hainzer M, et al. Striatal transplantation in a rodent model of multiple system atrophy: effects on L-Dopa response. *J Neurosci Res*. 2009 May 15;87(7):1679–85.
228. Puschban Z, Stefanova N, Petersén A, Winkler C, Brundin P, Poewe W, et al. Evidence for dopaminergic re-innervation by embryonic allografts in an optimized rat model of the Parkinsonian variant of multiple system atrophy. *Brain Res Bull*. 2005 Dec 15;68(1–2):54–8.
229. Stefanova N, Hainzer M, Stemberger S, Couillard-Després S, Aigner L, Poewe W, et al. Striatal transplantation for multiple system atrophy—are grafts affected by alpha-synucleinopathy? *Exp Neurol*. 2009 Sep;219(1):368–71.
230. Quinn N, Barker RA, Wenning GK. Are trials of intravascular infusions of autologous mesenchymal stem cells in patients with multiple system atrophy currently justified, and are they effective? *Clin Pharmacol Ther*. 2008 May;83(5):663–5.
231. Lee PH, Lee JE, Kim H-S, Song SK, Lee HS, Nam HS, et al. A randomized trial of mesenchymal stem cells in multiple system atrophy. *Ann Neurol*. 2012 Jul;72(1):32–40.
232. Valera E, Monzio Compagnoni G, Masliah E. Review: Novel treatment strategies targeting alpha-synuclein in multiple system atrophy as a model of synucleinopathy. *Neuropathol Appl Neurobiol*. 2016 Feb;42(1):95–106.
233. Smits MG, Gabreëls FJ, Thijssen HO, 't Lam RL, Notermans SL, ter Haar BG, et al. Progressive idiopathic strio-pallido-dentate calcinosis (Fahr's disease) with autosomal recessive inheritance. Report of three siblings. *Eur Neurol*. 1983;22(1):58–64.
234. Sobrido MJ, Coppola G, Oliveira J, Hopfer S, Geschwind DH. Primary Familial Brain Calcification. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJ, et al., editors. *GeneReviews*(®) [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cited 2016 Jul 21]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1421/>
235. Delacour A. Ossification des capillaires du cerveau. *Ann Med-Psychol*. 1850;2:458–61.
236. Fahr T. Idiopathische Verkalkung der Hirngefäße. *Zentralbl Allg Pathol*. 1930;50:129–30.

237. Klein C, Vieregge P. Fahr's disease--far from a disease. *Mov Disord Off J Mov Disord Soc.* 1998 May;13(3):620–1.
238. Tadic V, Westenberger A, Domingo A, Alvarez-Fischer D, Klein C, Kasten M. Primary familial brain calcification with known gene mutations: a systematic review and challenges of phenotypic characterization. *JAMA Neurol.* 2015 Apr;72(4):460–7.
239. Batla A, Tai XY, Schottlaender L, Erro R, Balint B, Bhatia KP. Deconstructing Fahr's disease/syndrome of brain calcification in the era of new genes. *Parkinsonism Relat Disord.* 2016 Dec 27;
240. Volpato CB, De Grandi A, Buffone E, Facheris M, Gebert U, Schifferle G, et al. 2q37 as a susceptibility locus for idiopathic basal ganglia calcification (IBGC) in a large South Tyrolean family. *J Mol Neurosci MN.* 2009 Nov;39(3):346–53.
241. Geschwind DH, Loginov M, Stern JM. Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). *Am J Hum Genet.* 1999 Sep;65(3):764–72.
242. Dai X, Gao Y, Xu Z, Cui X, Liu J, Li Y, et al. Identification of a novel genetic locus on chromosome 8p21.1-q11.23 for idiopathic basal ganglia calcification. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet.* 2010 Oct 5;153B(7):1305–10.
243. Oliveira JRM, Oliveira MF. Primary brain calcification in patients undergoing treatment with the biphosphonate alendronate. *Sci Rep.* 2016;6:22961.
244. Wang C, Li Y, Shi L, Ren J, Patti M, Wang T, et al. Mutations in SLC20A2 link familial idiopathic basal ganglia calcification with phosphate homeostasis. *Nat Genet.* 2012 Mar;44(3):254–6.
245. Nicolas G, Pottier C, Maltête D, Coutant S, Rovelet-Lecrux A, Legallic S, et al. Mutation of the PDGFRB gene as a cause of idiopathic basal ganglia calcification. *Neurology.* 2013 Jan 8;80(2):181–7.
246. Keller A, Westenberger A, Sobrido MJ, García-Murias M, Domingo A, Sears RL, et al. Mutations in the gene encoding PDGF-B cause brain calcifications in humans and mice. *Nat Genet.* 2013 Sep;45(9):1077–82.
247. Legati A, Giovannini D, Nicolas G, López-Sánchez U, Quintáns B, Oliveira JRM, et al. Mutations in XPR1 cause primary familial brain calcification associated with altered phosphate export. *Nat Genet.* 2015 Jun;47(6):579–81.
248. Taglia I, Bonifati V, Mignarri A, Dotti MT, Federico A. Primary familial brain calcification: update on molecular genetics. *Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol.* 2015 May;36(5):787–94.
249. Hsu SC, Sears RL, Lemos RR, Quintáns B, Huang A, Spiteri E, et al. Mutations in SLC20A2 are a major cause of familial idiopathic basal ganglia calcification. *Neurogenetics.* 2013 Feb;14(1):11–22.

250. Nicolas G, Charbonnier C, de Lemos RR, Richard A-C, Guillin O, Wallon D, et al. Brain calcification process and phenotypes according to age and sex: Lessons from SLC20A2, PDGFB, and PDGFRB mutation carriers. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet*. 2015 Oct;168(7):586–94.
251. Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, et al. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol*. 2012 Jan;11(1):54–65.
252. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011 Oct 20;72(2):245–56.
253. Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Res*. 2009 Jul;37(Web Server issue):W593-599.
254. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010 Jan;42(1):30–5.
255. Hammer MB, Eleuch-Fayache G, Schottlaender LV, Nehdi H, Gibbs JR, Arepalli SK, et al. Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. *Am J Hum Genet*. 2013 Feb 7;92(2):245–51.
256. Wang C, Li Y, Shi L, Ren J, Patti M, Wang T, et al. Mutations in SLC20A2 link familial idiopathic basal ganglia calcification with phosphate homeostasis. *Nat Genet* [Internet]. 2012 Feb 12 [cited 2012 Feb 20]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22327515>
257. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med*. 2013 Jan 10;368(2):117–27.
258. Singleton AB, Hardy J, Traynor BJ, Houlden H. Towards a complete resolution of the genetic architecture of disease. *Trends Genet TIG*. 2010 Oct;26(10):438–42.
259. Walters-Sen LC, Hashimoto S, Thrush DL, Reshmi S, Gastier-Foster JM, Astbury C, et al. Variability in pathogenicity prediction programs: impact on clinical diagnostics. *Mol Genet Genomic Med*. 2015 Mar;3(2):99–110.
260. Azevedo L, Mort M, Costa AC, Silva RM, Quelhas D, Amorim A, et al. Improving the in silico assessment of pathogenicity for compensated variants. *Eur J Hum Genet EJHG*. 2016 Jan;25(1):2–7.

261. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*. 1997 Jun 27;276(5321):2045–7.
262. Carson AR, Smith EN, Matsui H, Brækkan SK, Jepsen K, Hansen J-B, et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics*. 2014 May 2;15:125.
263. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010 Sep;5(9):1564–73.
264. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006 Aug;38(8):904–9.
265. Sailer A, Scholz SW, Nalls MA, Schulte C, Federoff M, Price TR, et al. A genome-wide association study in multiple system atrophy. *Neurology*. 2016 Oct 11;87(15):1591–8.
266. Nicolae DL. Association Tests for Rare Variants. *Annu Rev Genomics Hum Genet*. 2016 Aug 31;17:117–30.
267. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinforma Oxf Engl*. 2016 May 1;32(9):1423–6.
268. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*. 2008 Sep;83(3):311–21.
269. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*. 1975 Nov 5;98(3):503–17.
270. Arcangeli M-L, Frontera V, Bardin F, Obrados E, Adams S, Chabannon C, et al. JAM-B regulates maintenance of hematopoietic stem cells in the bone marrow. *Blood*. 2011 Oct 27;118(17):4609–19.
271. Brooks SP, Dunnett SB. Tests to assess motor phenotype in mice: a user's guide. *Nat Rev Neurosci*. 2009 Jul;10(7):519–29.
272. Slaoui M, Fiette L. Histopathology procedures: from tissue sampling to histopathological evaluation. *Methods Mol Biol Clifton NJ*. 2011;691:69–82.
273. Paxton S, Peckham M, Adele K, Paxton S, Adele K, Peckham M. The Leeds Histology Guide. 2003 [cited 2017 Jun 28]; Available from: <http://histology.leeds.ac.uk/what-is-histology/index.php>
274. Normal Histology [Internet]. [cited 2017 Jun 29]. Available from: <http://library.med.utah.edu/WebPath/HISTHTML/NORMAL/NORM055.html>

275. Sofroniew MV, Vinters HV. Astrocytes: biology and pathology. *Acta Neuropathol (Berl)*. 2010 Jan;119(1):7–35.
276. Lyck L, Dalmau I, Chemnitz J, Finsen B, Schrøder HD. Immunohistochemical markers for quantitative studies of neurons and glia in human neocortex. *J Histochem Cytochem Off J Histochem Soc*. 2008 Mar;56(3):201–21.
277. Gusel'nikova VV, Korzhevskiy DE. NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker. *Acta Naturae*. 2015 Jun;7(2):42–7.
278. Chartier-Harlin M-C, Dachsel JC, Vilariño-Güell C, Lincoln SJ, Leprêtre F, Hulihan MM, et al. Translation initiator EIF4G1 mutations in familial Parkinson disease. *Am J Hum Genet*. 2011 Sep 9;89(3):398–406.
279. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 mutations in Parkinson disease. *Am J Hum Genet*. 2011 Jul 15;89(1):162–7.
280. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, Offman MN, 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.
281. Lesage S, Brice A. Role of mendelian genes in “sporadic” Parkinson’s disease. *Parkinsonism Relat Disord*. 2012 Jan;18 Suppl 1:S66-70.
282. Fujioka S, Ogaki K, Tacik PM, Uitti RJ, Ross OA, Wszolek ZK. Update on novel familial forms of Parkinson’s disease and multiple system atrophy. *Parkinsonism Relat Disord*. 2014 Jan;20 Suppl 1:S29-34.
283. Lu C-S, Chang H-C, Weng Y-H, Chen R-S, Bonifati V, Wu-Chou Y-H. Analysis of the LRRK2 Gly2385Arg variant in primary dystonia and multiple system atrophy in Taiwan. *Parkinsonism Relat Disord*. 2008;14(5):393–6.
284. Heckman MG, Schottlaender L, Soto-Ortolaza AI, Diehl NN, Rayaprolu S, Ogaki K, et al. LRRK2 exonic variants and risk of multiple system atrophy. *Neurology*. 2014 Nov 5;
285. Lee K, Nguyen K-D, Sun C, Liu M, Zafar F, Saetern J, et al. LRRK2 p.Ile1371Val Mutation in a Case with Neuropathologically Confirmed Multi-System Atrophy. *J Park Dis*. 2018;8(1):93–100.
286. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, 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.
287. Boeve BF, Boylan KB, Graff-Radford NR, DeJesus-Hernandez M, Knopman DS, Pedraza O, et al. Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the

- GGGGCC repeat expansion in C9ORF72. *Brain J Neurol.* 2012 Mar;135(Pt 3):765–83.
288. Hodges J. Familial frontotemporal dementia and amyotrophic lateral sclerosis associated with the C9ORF72 hexanucleotide repeat. *Brain J Neurol.* 2012 Mar;135(Pt 3):652–5.
289. Lindquist SG, Duno M, Batbayli M, Puschmann A, Braendgaard H, Mardosiene S, et al. Corticobasal and ataxia syndromes widen the spectrum of C9ORF72 hexanucleotide expansion disease. *Clin Genet.* 2013 Mar;83(3):279–83.
290. Mahoney CJ, Beck J, Rohrer JD, Lashley T, Mok K, Shakespeare T, et al. Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain J Neurol.* 2012 Mar;135(Pt 3):736–50.
291. Majounie E, Renton AE, Mok K, Dopper EGP, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* 2012 Apr;11(4):323–30.
292. O’Dowd S, Curtin D, Waite AJ, Roberts K, Pender N, Reid V, et al. C9ORF72 expansion in amyotrophic lateral sclerosis/frontotemporal dementia also causes parkinsonism. *Mov Disord Off J Mov Disord Soc.* 2012 Jul;27(8):1072–4.
293. Stamelou M, Hoeglenger GU. Atypical parkinsonism: an update. *Curr Opin Neurol.* 2013 Aug;26(4):401–5.
294. Snowden JS, Rollinson S, Lafon C, Harris J, Thompson J, Richardson AM, et al. Psychosis, C9ORF72 and dementia with Lewy bodies. *J Neurol Neurosurg Psychiatry.* 2012 Oct;83(10):1031–2.
295. Lesage S, Le Ber I, Condroyer C, Broussolle E, Gabelle A, Thobois S, et al. C9orf72 repeat expansions are a rare genetic cause of parkinsonism. *Brain J Neurol.* 2013 Feb;136(Pt 2):385–91.
296. Harms M, Benitez BA, Cairns N, Cooper B, Cooper P, Mayo K, et al. C9orf72 Hexanucleotide Repeat Expansions in Clinical Alzheimer Disease. *JAMA Neurol.* 2013 Apr 15;1–6.
297. Majounie E, Abramzon Y, Renton AE, Perry R, Bassett SS, Pletnikova O, et al. Repeat expansion in C9ORF72 in Alzheimer’s disease. *N Engl J Med.* 2012 Jan 19;366(3):283–4.
298. Xi Z, Zinman L, Grinberg Y, Moreno D, Sato C, Bilbao JM, et al. Investigation of c9orf72 in 4 neurodegenerative disorders. *Arch Neurol.* 2012 Dec;69(12):1583–90.
299. Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G, et al. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than

- expected in the UK population. *Am J Hum Genet.* 2013 Mar 7;92(3):345–53.
300. Cooper-Knock J, Frolov A, Highley JR, Charlesworth G, Kirby J, Milano A, et al. C9ORF72 expansions, parkinsonism, and Parkinson disease: A clinicopathologic study. *Neurology.* 2013 Aug 27;81(9):808–11.
301. Dickson DW, Bergeron C, Chin SS, Duyckaerts C, Horoupian D, Ikeda K, et al. Office of Rare Diseases neuropathologic criteria for corticobasal degeneration. *J Neuropathol Exp Neurol.* 2002 Nov;61(11):935–46.
302. Armstrong MJ, Litvan I, Lang AE, Bak TH, Bhatia KP, Borroni B, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology.* 2013 Jan 29;80(5):496–503.
303. Mok K, Traynor BJ, Schymick J, Tienari PJ, Laaksovirta H, Peuralinna T, et al. Chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol Aging.* 2012 Jan;33(1):209.e3-8.
304. Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G, et al. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet.* 2013 Mar 7;92(3):345–53.
305. Schottlaender L, Polke JM, Ling H, MacDoanld ND, .Tucci A, Nanji T, et al. The analysis of C9orf72 repeat expansions in a large series of clinically and pathologically diagnosed cases with atypical parkinsonism. *Neurobiol Aging* [Internet]. [cited 2014 Sep 15]; Available from: <http://www.sciencedirect.com/science/article/pii/S0197458014005375>
306. Gómez-Tortosa E, Gallego J, Guerrero-López R, Marcos A, Gil-Neciga E, Sainz MJ, et al. C9ORF72 hexanucleotide expansions of 20-22 repeats are associated with frontotemporal deterioration. *Neurology.* 2013 Jan 22;80(4):366–70.
307. van der Zee J, Gijssels I, Dillen L, Van Langenhove T, Theuns J, Engelborghs S, et al. A pan-European study of the C9orf72 repeat associated with FTLT: geographic prevalence, genomic instability, and intermediate repeats. *Hum Mutat.* 2013 Feb;34(2):363–73.
308. Nuytemans K, Bademci G, Kohli MM, Beecham GW, Wang L, Young JJ, et al. C9ORF72 Intermediate Repeat Copies Are a Significant Risk Factor for Parkinson Disease. *Ann Hum Genet.* 2013 Jul 12;
309. Majounie E, Abramzon Y, Renton AE, Keller MF, Traynor BJ, Singleton AB. Large C9orf72 repeat expansions are not a common cause of Parkinson's disease. *Neurobiol Aging.* 2012 Oct;33(10):2527.e1-2.

310. Pamphlett R, Cheong PL, Trent RJ, Yu B. Transmission of C9orf72 hexanucleotide repeat expansions in sporadic amyotrophic lateral sclerosis: an Australian trio study. *Neuroreport*. 2012 Jun 20;23(9):556–9.
311. Xi Z, van Blitterswijk M, Zhang M, McGoldrick P, McLean JR, Yunusova Y, et al. Jump from pre-mutation to pathologic expansion in C9orf72. *Am J Hum Genet*. 2015 Jun 4;96(6):962–70.
312. Dobson-Stone C, Hallupp M, Bartley L, Shepherd CE, Halliday GM, Schofield PR, et al. C9ORF72 repeat expansion in clinical and neuropathologic frontotemporal dementia cohorts. *Neurology*. 2012 Sep 4;79(10):995–1001.
313. Forsgren M, Attersand A, Lake S, Grünler J, Swiezewska E, Dallner G, et al. Isolation and functional expression of human COQ2, a gene encoding a polyprenyl transferase involved in the synthesis of CoQ. *Biochem J*. 2004 Sep 1;382(Pt 2):519–26.
314. Trabzuni D, Ryten M, Walker R, Smith C, Imran S, Ramasamy A, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem*. 2011 Oct;119(2):275–82.
315. Hara K, Momose Y, Tokiguchi S, Shimohata M, Terajima K, Onodera O, et al. Multiplex families with multiple system atrophy. *Arch Neurol*. 2007 Apr;64(4):545–51.
316. Lin C-H, Tan E-K, Yang C-C, Yi Z, Wu R-M. COQ2 gene variants associate with cerebellar subtype of multiple system atrophy in Chinese. *Mov Disord Off J Mov Disord Soc*. 2015 Mar;30(3):436–7.
317. Stamelou M, Reuss A, Pilatus U, Magerkurth J, Niklowitz P, Eggert KM, et al. Short-term effects of coenzyme Q10 in progressive supranuclear palsy: a randomized, placebo-controlled trial. *Mov Disord Off J Mov Disord Soc*. 2008 May 15;23(7):942–9.
318. Thomas B, Beal MF. Mitochondrial therapies for Parkinson’s disease. *Mov Disord Off J Mov Disord Soc*. 2010;25 Suppl 1:S155-160.
319. Hargreaves IP, Lane A, Sleiman PMA. The coenzyme Q10 status of the brain regions of Parkinson’s disease patients. *Neurosci Lett*. 2008 Dec 5;447(1):17–9.
320. Mischley LK, Allen J, Bradley R. Coenzyme Q10 deficiency in patients with Parkinson’s disease. *J Neurol Sci*. 2012 Jul 15;318(1–2):72–5.
321. Ries V, Oertel WH, Höglinger GU. Mitochondrial dysfunction as a therapeutic target in progressive supranuclear palsy. *J Mol Neurosci MN*. 2011 Nov;45(3):684–9.
322. Molina JA, de Bustos F, Ortiz S, Del Ser T, Seijo M, Benito-Léon J, et al. Serum levels of coenzyme Q in patients with Lewy body disease. *J Neural Transm Vienna Austria 1996*. 2002 Sep;109(9):1195–201.

323. Shults CW, Oakes D, Kieburtz K, Beal MF, Haas R, Plumb S, et al. Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. *Arch Neurol*. 2002 Oct;59(10):1541–50.
324. Shults CW, Flint Beal M, Song D, Fontaine D. Pilot trial of high dosages of coenzyme Q10 in patients with Parkinson's disease. *Exp Neurol*. 2004 Aug;188(2):491–4.
325. Storch A, Jost WH, Vieregge P, Spiegel J, Greulich W, Durner J, et al. Randomized, double-blind, placebo-controlled trial on symptomatic effects of coenzyme Q(10) in Parkinson disease. *Arch Neurol*. 2007 Jul;64(7):938–44.
326. Jankovic J, Poewe W. Therapies in Parkinson's disease. *Curr Opin Neurol*. 2012 Aug;25(4):433–47.
327. Schottlaender LV, Bettencourt C, Kiely AP, Chalasani A, Neergheen V, Holton JL, et al. Coenzyme Q10 Levels Are Decreased in the Cerebellum of Multiple-System Atrophy Patients. *PloS One*. 2016;11(2):e0149557.
328. Stamelou M, Reuss A, Pilatus U, Magerkurth J, Niklowitz P, Eggert KM, et al. Short-term effects of coenzyme Q10 in progressive supranuclear palsy: a randomized, placebo-controlled trial. *Mov Disord Off J Mov Disord Soc*. 2008 May 15;23(7):942–9.
329. Thomas B, Beal MF. Mitochondrial therapies for Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2010;25 Suppl 1:S155-160.
330. Hargreaves IP, Lane A, Sleiman PMA. The coenzyme Q10 status of the brain regions of Parkinson's disease patients. *Neurosci Lett*. 2008 Dec 5;447(1):17–9.
331. Mischley LK, Allen J, Bradley R. Coenzyme Q10 deficiency in patients with Parkinson's disease. *J Neurol Sci*. 2012 Jul 15;318(1–2):72–5.
332. Ries V, Oertel WH, Höglinger GU. Mitochondrial dysfunction as a therapeutic target in progressive supranuclear palsy. *J Mol Neurosci MN*. 2011 Nov;45(3):684–9.
333. Stamelou M, Pilatus U, Reuss A, Respondek G, Knake S, Oertel WH, et al. Brain energy metabolism in early MSA-P: A phosphorus and proton magnetic resonance spectroscopy study. *Parkinsonism Relat Disord*. 2015 May;21(5):533–5.
334. Parkinson Study Group QE3 Investigators, Beal MF, Oakes D, Shoulson I, Henchcliffe C, Galpern WR, et al. A randomized clinical trial of high-dosage coenzyme Q10 in early Parkinson disease: no evidence of benefit. *JAMA Neurol*. 2014 May;71(5):543–52.
335. Parkinson MH, Schulz JB, Giunti P. Co-enzyme Q10 and idebenone use in Friedreich's ataxia. *J Neurochem*. 2013 Aug;126 Suppl 1:125–41.

336. Barca E, Kleiner G, Tang G, Ziosi M, Tadesse S, Masliah E, et al. Decreased Coenzyme Q10 Levels in Multiple System Atrophy Cerebellum. *J Neuropathol Exp Neurol*. 2016;75(7):663–72.
337. Wüllner U, Schmitt I, Kammal M, Kretschmar HA, Neumann M. Definite multiple system atrophy in a German family. *J Neurol Neurosurg Psychiatry*. 2009 Apr;80(4):449–50.
338. Guerreiro RJ, Lohmann E, Kinsella E, Brás JM, Luu N, Gurunlian N, et al. Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer’s disease. *Neurobiol Aging*. 2012 May;33(5):1008.e17-23.
339. Brömme D, Rossi AB, Smeekens SP, Anderson DC, Payan DG. Human bleomycin hydrolase: molecular cloning, sequencing, functional expression, and enzymatic characterization. *Biochemistry (Mosc)*. 1996 May 28;35(21):6706–14.
340. Delamarre A, Meissner WG. Epidemiology, environmental risk factors and genetics of Parkinson’s disease. *Presse Medicale Paris Fr* 1983. 2017 Mar;46(2 Pt 1):175–81.
341. Schottlaender LV, Houlden H, Multiple-System Atrophy (MSA) Brain Bank Collaboration. Mutant COQ2 in multiple-system atrophy. *N Engl J Med*. 2014 Jul 3;371(1):81.
342. Multiple-System Atrophy Research Collaboration. Mutations in COQ2 in familial and sporadic multiple-system atrophy. *N Engl J Med*. 2013 Jul 18;369(3):233–44.
343. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011 Jul 15;89(1):82–93.
344. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostat Oxf Engl*. 2012 Sep;13(4):762–75.
345. Federoff M, Price TR, Sailer A, Scholz S, Hernandez D, Nicolas A, et al. Genome-wide estimate of the heritability of Multiple System Atrophy. *Parkinsonism Relat Disord*. 2016 Jan;22:35–41.
346. Bellenguez C, Charbonnier C, Grenier-Boley B, Quenez O, Le Guennec K, Nicolas G, et al. Contribution to Alzheimer’s disease risk of rare variants in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls. *Neurobiol Aging*. 2017 Nov;59:220.e1-220.e9.
347. Jansen IE, Gibbs JR, Nalls MA, Price TR, Lubbe S, van Rooij J, et al. Establishing the role of rare coding variants in known Parkinson’s disease risk loci. *Neurobiol Aging*. 2017 Nov;59:220.e11-220.e18.
348. Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson’s disease. *Brain [Internet]*. [cited 2017 Nov 14]; Available

from:

<https://academic.oup.com/brain/article/doi/10.1093/brain/awx285/4617852>

349. Hammer MB, Eleuch-Fayache G, Schottlaender LV, Nehdi H, Gibbs JR, Arepalli SK, et al. Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. *Am J Hum Genet.* 2013 Feb 7;92(2):245–51.
350. Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Res.* 2009 Jul;37(Web Server issue):W593-599.
351. Mochida GH, Ganesh VS, Felie JM, Gleason D, Hill RS, Clapham KR, et al. A homozygous mutation in the tight-junction protein JAM3 causes hemorrhagic destruction of the brain, subependymal calcification, and congenital cataracts. *Am J Hum Genet.* 2010 Dec 10;87(6):882–9.
352. O’Driscoll MC, Daly SB, Urquhart JE, Black GCM, Pilz DT, Brockmann K, et al. Recessive mutations in the gene encoding the tight junction protein occludin cause band-like calcification with simplified gyration and polymicrogyria. *Am J Hum Genet.* 2010 Sep 10;87(3):354–64.
353. Redmond SA, Mei F, Eshed-Eisenbach Y, Osso LA, Leshkowitz D, Shen Y-AA, et al. Somatodendritic Expression of JAM2 Inhibits Oligodendrocyte Myelination. *Neuron.* 2016 Aug 17;91(4):824–36.
354. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011 Oct 20;72(2):245–56.
355. Schottlaender LV, Polke JM, Ling H, MacDoanld ND, Tucci A, Nanji T, et al. Analysis of C9orf72 repeat expansions in a large series of clinically and pathologically diagnosed cases with atypical parkinsonism. *Neurobiol Aging.* 2015 Feb;36(2):1221.e1-6.

APPENDIX

7.3 EXOME SEQUENCING SCRIPTS

7.3.1 Scripts used to analyse exomes

7.3.1.1 Example of script to analyse an individual exome

```
#!/bin/bash
```

```
#$ -S /bin/bash
```

```
#$ -o cluster/out
```

```
#$ -e cluster/error
```

```
#$ -cwd
```

```
date ##to measure the duration
```

```
export
```

```
PATH=${PATH}:/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013:/illumina/pipeline/vincent/Software/tabix-0.2.3
```

```
export PERL5LIB=/illumina/pipeline/vincent/Software/vcftools_0.1.8/lib:
```

```
/illumina/pipeline/vincent/Software/novocraft/novoalign -c 12 -o SAM
```

```
.$@RG\tID:MSA_1632\tSM:MSA_1632\tLB:MSA_1632\tPL:ILLUMINA' --rOQ --hdrhd 3 -H -
```

```
k -a -o Soft -t 320 -F ILM1.8 -f
```

```
/illumina/runs/MSA_New/Unaligned/MSA_1632/1632_MSA_R1.fastq.bz2
```

```
/illumina/runs/MSA_New/Unaligned/MSA_1632/1632_MSA_R2.fastq.bz2 -d
```

```
/illumina/pipeline/vincent/reference_genome/novoalign/human_g1k_v37.fasta.k15.s2.no
```

```
voindex > /illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.sam
```

```
/illumina/pipeline/vincent/Software/samtools-0.1.18/samtools view -bS -t
```

```
/illumina/pipeline/vincent/reference_genome/fasta/human_g1k_v37.fasta -o
```

```

/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.bam
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.sam

## make BAM file

/illumina/pipeline/vincent/Software/samtools-0.1.18/samtools sort -m 500000000
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.bam
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted

## sort

/illumina/pipeline/vincent/Software/samtools-0.1.18/samtools index
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted.bam ##build index

## Now remove duplicates using PICARD

java -Xmx10g -jar /illumina/pipeline/vincent/Software/picard-tools-
1.75/MarkDuplicates.jar TMP_DIR=/illumina/runs/temp/java ASSUME_SORTED=true
REMOVE_DUPLICATES=FALSE

INPUT=/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted.bam
OUTPUT=/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted_unique.bam
METRICS_FILE=/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_picard_metric
s.out

/illumina/pipeline/vincent/Software/samtools-0.1.18/samtools index
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted_unique.bam

##build index

rm /illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted.bam
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted.bam.bai

rm /illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.bam
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.sam

java -Xmx10g -jar /illumina/pipeline/vincent/Software/picard-tools-
1.75/CalculateHsMetrics.jar

```

```

BAIT_INTERVALS=/illumina/pipeline/vincent/reference_genome/query_novopile/TruSeq_
ExomeTarget_hg19_0bp.tab.intList

TARGET_INTERVALS=/illumina/pipeline/vincent/reference_genome/query_novopile/TruS
eq_ExomeTarget_hg19_0bp.tab.intList

INPUT=/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted_unique.bam

OUTPUT=/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632.hybridMetrics

/illumina/pipeline/vincent/Software/samtools-0.1.18/samtools mpileup -q 20 -L 400 -d 400
-ugf /illumina/pipeline/vincent/reference_genome/fasta/human_g1k_v37.fasta
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_sorted_unique.bam |
/illumina/pipeline/vincent/Software/samtools-0.1.18/bcftools/bcftools view -bvcg - >
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_rawVar.bcf

/illumina/pipeline/vincent/Software/samtools-0.1.18/bcftools/bcftools view
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_rawVar.bcf >
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_Var.vcf

rm /illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_rawVar.bcf

/illumina/pipeline/vincent/Software/vcftools_0.1.8/cpp/vcftools --vcf
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_Var.vcf --bed
/illumina/pipeline/vincent/reference_genome/query_novopile/TruSeq_ExomeTarget_hg1
9_0bp.tab.bed --recode --out
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_Var_target

/illumina/pipeline/vincent/Software/samtools-0.1.18/bcftools/vcfutils.pl varFilter
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_Var.vcf | awk '{if ( ($6 >= 18)
|| ( $1 ~ /^#/ ) ) print}' >
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_Var.vcf_filtered

/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013/convert2annovar.pl -
format vcf4 /illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_Var.vcf_filtered -
outfile

```

```
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_annovar/annovar_MSA_1632  
_temp
```

```
cut -f1-8
```

```
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_annovar/annovar_MSA_1632  
_temp >
```

```
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_annovar/annovar_MSA_1632
```

```
/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013/summarize_annovar_VP.pl
```

```
-ver1000g 1000g2012apr -verdb SNP 137 -veresp 6500si -alltranscript -buildver hg19 --
```

```
genetype gencodegene --remove -buildver hg19
```

```
/illumina/runs/MSA_New/v_aligned/MSA_1632/MSA_1632_annovar/annovar_MSA_1632
```

```
/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013/humandb_hg19
```

Date

7.3.1.2 Script to align multiple samples

```
pipeline=/illumina/pipeline/vincent/Software/pipeline/align_pipeline_DNA_v4_1_ATLAS.s
```

```
h
```

```
build=hg19
```

```
###Z:\illumina\runs\MSA_New
```

```
### align all the data
```

```
iFolder=/illumina/runs/MSA_New/Unaligned
```

```
oFolder=/illumina/runs/MSA_New/v_aligned
```

```
myIDs="MSA_1132
```

```
MSA_1386
```

```
MSA_1409
```

```
MSA_15569
```

```
MSA_1561
```

```
240
```

```

if [ ! -e $oFolder ]; then mkdir $oFolder; fi

for nID in $myIDs; do  folder=${iFolder}/${nID}

    inputFiles=""

    nfiles=0

    for file1 in `find $folder -name *R1*bz2`; do

        ((nfiles=nfiles+2))

        file2=`echo $file1 | sed -e 's/R1/R2/g'`

        inputFiles="$inputFiles $file1 $file2"

    done

    inputFiles="$nfiles $inputFiles"

    fasta=/illumina/pipeline/vincent/reference_genome/fasta/human_g1k_v37.fasta

    reference=/illumina/pipeline/vincent/reference_genome/novoalign/human_g1k_v37.fast
a.k15.s2.novoindex

    output=${oFolder}/${nID}/${nID}

    script=/illumina/runs/MSA_New/${nID}_H.sh

    query=/illumina/pipeline/vincent/reference_genome/query_novopile/TruSeq_ExomeTarg
et_hg19_0bp.tab

    baitFile=/illumina/pipeline/vincent/reference_genome/query_novopile/TruSeq_ExomeTar
get_hg19_0bp.tab.intList

    if [ ! -e ${oFolder}/${nID} ]; then mkdir ${oFolder}/${nID}; fi

    align=yes

```

```

summaryStats=yes

pileup=yes

annotate=yes

local=interactive

    echo $nID

    sh ${pipeline} --script ${script} --inputFiles ${inputFiles} --fasta ${fasta} --reference
    ${reference} --output ${output} --annotate ${annotate} --align ${align} --pileup ${pileup}
    ${query} --local ${local} --build ${build} --summaryStats ${summaryStats} --tparam 320 --
    baitFile ${baitFile} --fullPileup --inputFormat ILM1.8 --javaTemp /illumina/runs/temp/java

done

```

7.3.1.3 *Script to filter the vcf file*

```

#!/bin/bash

#$ -S /bin/bash

#$ -o cluster/out

#$ -e cluster/error

#$ -cwd

date ##to measure the duration

export

PATH=${PATH}:/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013:/illumina/pi
pipeline/vincent/Software/tabix-0.2.3

export PERL5LIB=/illumina/pipeline/vincent/Software/vcftools_0.1.8/lib:

/illumina/pipeline/vincent/Software/samtools-0.1.18/bcftools/vcfutils.pl          varFilter
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA

```

```

TK_MSA_Samples_1_Var.vcf | awk '{if ( ($6 >= 18) || ( $1 ~ /^#/ ) ) print}' >
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA
TK_MSA_Samples_1_Var.vcf_filtered

/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013/convert2annovar.pl -
format vcf4
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA
TK_MSA_Samples_1_Var.vcf_filtered -outfile
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA
TK_MSA_Samples_1_annovar/annovar_combinedGATK_MSA_Samples_1_temp
cut -f1-8
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA
TK_MSA_Samples_1_annovar/annovar_combinedGATK_MSA_Samples_1_temp >
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA
TK_MSA_Samples_1_annovar/annovar_combinedGATK_MSA_Samples_1

/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013/summarize_annovar_VP.pl
-ver1000g 1000g2012apr -verdb SNP 137 -veresp 6500si -alltranscript -buildver hg19 --
genetype gencodegene --remove -buildver hg19
/illumina/runs/MSA_Joint_VCF/v_aligned/combinedGATK_MSA_Samples_1/combinedGA
TK_MSA_Samples_1_annovar/annovar_combinedGATK_MSA_Samples_1

/illumina/pipeline/vincent/Software_heavy/annovar_Feb2013/humandb_hg19

date

```

7.3.1.4 Script to join multiple samples

```

java -jar /illumina/pipeline/vincent/Software/GenomeAnalysisTK-1.4-16-
g3ba918a/GenomeAnalysisTK.jar -R
/illumina/pipeline/vincent/reference_genome/fasta/human_g1k_v37.fasta -T
CombineVariants --variant 002-08_Var.vcf_filtered --variant 003-
08_Var.vcf_filtered --variant 004-08_Var.vcf_filtered --variant 004-
10_Var.vcf_filtered --variant 006_08_Var.vcf_filtered --variant 006-

```



```

#####STAGE1 : VARIANT AND SAMPLE QC
#####

#STEP1

#ADD dbSNP annotation to vcf file in GATK

$GATK_VariantAnnotator -V $WorkingDirectory/$inputVCF --dbnp
/data/kronos/NGS_Reference/GATK_refFiles/common_all.vcf -o
$WorkingDirectory/db_$inputVCF

mv db_$inputVCF $studyNAME.vcf

#STEP2

#Make a binary (bed) file for PLINK input:

plink --vcf $studyNAME.vcf --double-id --make-bed --out $studyNAME

#STEP3

#Update individual sex information and phenotype:

plink --bfile $WorkingDirectory/$studyNAME --update-sex $WorkingDirectory/$sexINFO --
make-bed --out $WorkingDirectory/${studyNAME}_updated_sex

plink --bfile $WorkingDirectory/${studyNAME}_updated_sex --pheno
$WorkingDirectory/$phenoINFO --make-bed --out
$WorkingDirectory/${studyNAME}_updated_sex_and_pheno --allow-no-sex

#STEP4

#Split XY chromosomes

plink --bfile $WorkingDirectory/${studyNAME}_updated_sex_and_pheno --double-id --
split-x hg19 --make-bed --out $WorkingDirectory/${studyNAME}_Xsplit --allow-no-sex

#STEP5

```

#Sex Check and removal of ambiguous I added the allow no sex command to keep the samples where there was unknown sex information in the input file

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit --check-sex 0.3 0.6 --out $WorkingDirectory/${studyNAME} --allow-no-sex
```

```
cat $WorkingDirectory/${studyNAME}.sexcheck | awk '($3=="0" || $5=="OK"){print $1"\t"$2}' > Sex_samples_to_keep
```

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit --keep Sex_samples_to_keep --make-bed --out $WorkingDirectory/${studyNAME}_Xsplit_SexPruned --allow-no-sex
```

#STEP6

#remove SNPs that have a call rate less than 90% (probably bad SNPs) (one can also visually inspect distribution)

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit_SexPruned --missing --out $WorkingDirectory/${studyNAME}_missingness_before_bad_SNP_removal
```

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit_SexPruned --geno 0.1 --make-bed --out $WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate
```

#STEP7

#now lets remove bad samples that still have a lot of missingness after removal of the bad SNPs (one can also visually inspect distribution)

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate --missing --out $WorkingDirectory/${studyNAME}_missingness_after_bad_SNP_removal
```

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate --mind 0.2 --make-bed --out $WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate_80pc_sample_call_rate
```

```
plink --bfile $WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate_80pc_sample_call_r  
246
```

```

ate --missing --out
$WorkingDirectory/${studyNAME}_missingness_after_bad_SNP_removal_and_bad_sampl
e_removal

#STEP8

#Hardy-Weinberg Equilibrium

plink --bfile
$WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate_80pc_sample_call_r
ate --hardy --out
$WorkingDirectory/${studyNAME}_missingness_after_bad_SNP_removal_and_bad_sampl
e_removal_HWE

#filter:

plink --bfile
$WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate_80pc_sample_call_r
ate --hwe 0.001 --make-bed --out
$WorkingDirectory/${studyNAME}_Xsplit_SexPruned_90pc_call_rate_80pc_sample_call_r
ate_HWE_filtered

#####STAGE2 : POPULATION
STRATIFICATION#####

#STEP1 - HET

#het outliers (inc.removal)

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC/${studyNAME}_
post_variantANDsampleGenotypeQC --het --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_het

```

```

awk ' $6 >= 0.05 || $6 <= -0.1 '
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_het.het >
$WorkingDirectory/het_samples_to_remove.txt

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC/${studyNAME}_
post_variantANDsampleGenotypeQC --remove
$WorkingDirectory/het_samples_to_remove.txt --make-bed --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails

#STEP2 - IBD Calculation

#delete low MAF SNPs:

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails --
maf 0.01 --make-bed --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails1

##prune_LD_SNPS

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails1 -
-indep-pairwise 50 5 0.5 --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned

###use only RS numbers here!

awk '$1 ~ /^rs/'
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned.prune.in >
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned.prune2.in

```

```

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails --
extract
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned.prune2.in --make-bed --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned

```

#IBD calculation:

```

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned --genome --min 0.1 --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned

```

#STEP3 - IBS Calculation

```

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned --cluster --neighbour 1 5 --out
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned_IBS

```

#----> Neighborhood Z scores (visually Inspect to remove outliers that slip off at the end(low)
)

#STEP4 - PCA visualisation

```

plink --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L
D_Pruned --extract
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.no-at-cg-
snps.txt --make-bed --out

```

```
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L  
D_Pruned.hapmap-snps
```

```
#REMOVE MULTI-ALLELE SNPS (by finding out which ones are multi-allelic first):
```

```
plink --bfile  
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L  
D_Pruned.hapmap-snps --bmerge  
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.founders.  
no-at-cg-snps.bed  
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.founders.  
no-at-cg-snps.bim  
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.founders.  
no-at-cg-snps.fam
```

```
plink --bfile  
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L  
D_Pruned.hapmap-snps --exclude $WorkingDirectory/plink.missnp --make-bed --out  
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L  
D_Pruned.hapmap-snps2
```

```
plink --bfile  
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_less_het_fails_L  
D_Pruned.hapmap-snps2 --bmerge  
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.founders.  
no-at-cg-snps.bed  
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.founders.  
no-at-cg-snps.bim  
/data/kronos/NGS_Reference/HapMap_Refernce/hapmap3r2_CEU.CHB.JPT.YRI.founders.  
no-at-cg-snps.fam --out
```

```
$WorkingDirectory/${studyNAME}_merged_with_HapMap_for_PCA
```

```
#now need to remove warning SNPs: use a bit of awk magic on the log file:
```

```

grep                                'Warning:                                Multiple'
$WorkingDirectory/${studyNAME}_merged_with_HapMap_for_PCA.log                >
$WorkingDirectory/removal1.txt

awk '{print $7}' $WorkingDirectory/removal1.txt > $WorkingDirectory/removal2.txt

sed -i 's/^./' $WorkingDirectory/removal2.txt

sed -i 's/./' $WorkingDirectory/removal2.txt

#great now remove these additional warning SNPs!

plink --bfile $WorkingDirectory/${studyNAME}_merged_with_HapMap_for_PCA --exclude
$WorkingDirectory/removal2.txt --make-bed --out
$WorkingDirectory/${studyNAME}_merged_with_HapMap_for_PCA_clean

#now ready for PCA:

#GCTA

gcta --bfile $WorkingDirectory/${studyNAME}_merged_with_HapMap_for_PCA --make-
grm --autosome --thread-num 10 --out $WorkingDirectory/${studyNAME}_matrix

gcta --grm $WorkingDirectory/${studyNAME}_matrix --pca 4

#####STAGE3 : DECIDE ON FINAL SAMPLE INCLUSION
FOR ASSOCIATION ANALYSIS before re-running #####

mkdir $WorkingDirectory/${studyNAME}_Analysis_Ready_Variants

plink                                --bfile
$WorkingDirectory/${studyNAME}_post_variantANDsampleGenotypeQC_PopulationStrati
fication/${studyNAME}_post_variantANDsampleGenotypeQC_PopulationStratification --
make-bed --remove $SAMPLEEXCLUSION --out
$WorkingDirectory/${studyNAME}_Analysis_Ready_Variants/${studyNAME}_Analysis_Rea
dy_Variants

```

7.3.2.1 *Single association study*

```
#####STAGE1:SINGLE LOCUS - COMMON VARIATION
#####
```

```
mkdir $WorkingDirectory/${studyNAME}_Association_analysis_results
```

```
plink --bfile
$WorkingDirectory/${studyNAME}_Analysis_Ready_Variants/${studyNAME}_Analysis_Ready_Variants
--maf 0.05 --make-bed --out
$WorkingDirectory/${studyNAME}_Analysis_Ready_Variants/${studyNAME}_Analysis_Ready_Variants_common
```

```
plink --bfile
$WorkingDirectory/${studyNAME}_Analysis_Ready_Variants/${studyNAME}_Analysis_Ready_Variants_common
--allow-no-sex --assoc counts --pheno $pheno --out
$WorkingDirectory/${studyNAME}_Association_analysis/singlelocus
```

```
#####STAGE2:QQ PLOT - COMMON VARIATION
#####
```

```
#QQ PLOT...
```

```
plink --bfile
$WorkingDirectory/${studyNAME}_Analysis_Ready_Variants/${studyNAME}_Analysis_Ready_Variants_common
--assoc --allow-no-sex --pheno $pheno --adjust --qq-plot --out
$WorkingDirectory/${studyNAME}_Association_analysis/singlelocus_adjust_forQQ
```

```
echo "#!/usr/bin/env Rscript" >
$WorkingDirectory/${studyNAME}_Association_analysis/${studyNAME}_QQ.R
```

```
echo "QQdata <- read.table(file =
\"$WorkingDirectory/${studyNAME}_Association_analysis/singlelocus_adjust_forQQ.assoc.
adjusted\", header = T)" >>
$WorkingDirectory/${studyNAME}_Association_analysis/${studyNAME}_QQ.R
```

```

echo "plot(-log(QQdata\${QQ}, 10), -log(QQdata\${UNADJ}, 10), xlab = \"expected-logP
values\", ylab = \"observed-logP values\")" >>
\${WorkingDirectory}/\${studyNAME}_Asociation_analysis/\${studyNAME}_QQ.R

echo "abline(a = 0, b = 1)" >>
\${WorkingDirectory}/\${studyNAME}_Asociation_analysis/\${studyNAME}_QQ.R

echo "dev.off()" >>
\${WorkingDirectory}/\${studyNAME}_Asociation_analysis/\${studyNAME}_QQ.R

Rscript \${WorkingDirectory}/\${studyNAME}_Asociation_analysis/\${studyNAME}_QQ.R

```

7.3.2.2 *Script for rare variant tests*

```

#!/bin/sh

RVTESTS="/array/jvandrovcova/software/rvtests/executable"

INPUT="MSA_all_for_RVTESTS.vcf.gz"

\${RVTESTS}/rvtest --inVcf \${INPUT}.gz \

--pheno MSA_all_Analysis_Ready_Variants.fam \

--covar covar_file_for_rvtest.txt --covar-name age,pc1,pc2

--out \${INPUT}.rvtest \

--geneFile refFlat.gencode.v19.gz \

--burden cmc \

--vt price \

--kernel skat,skato,kbac

```

7.3.2.3 *Script for QQ plot and Manhattan plot in R*

```

data <- read.table("all_samples_all_variants_SKAT-O_plot.txt",header=TRUE)

manhattan(data, cex.axis = 1,chr = "CHR", bp = "BP", p = "P", snp = "SNP", col = c("blue4",
"red2"), chrlabs = NULL,suggestiveline = -log10(8.966E-07), genomewideline = -log10(5e-

```


grep

```
"Chr\|ACMSD\|APOE\|APP\|ATP13A2\|BST1\|CCDC62\|CHCHD2\|COQ2\|DNAJC6\|EIF4  
G1\|FBXO7\|GAK\|GBA\|GCH1\|HIP1R\|HLA-  
DRB5\|LAMP3\|LRKK2\|MAPT\|MCCC1\|PARK2\|PARK7\|PINK1\|PLA2G6\|PSEN1\|RAB  
39B\|SCARB2\|SNCA\|STK39\|SYT11\|VPS35"
```

```
merge.annotated.annovar.0902.hg19_multianno.txt > all_candidate_genes_together
```

7.4 PRIMER SEQUENCES

7.4.1 *JAM2*

JAM2_X7_A_F	ATGACTGCATCTGTCCGTGT
JAM2_X7_A_R	AGATGAGCTGGGTGTGTTGG
JAM2_X3_A_F	CCAATTCATGGGACCTGTTGA
JAM2_X3_A_R	TGCCTGAGTTGAGAGAGAAGAG
JAM2_X2_B_F	GGAAAATGCTTAAAGGCCAAAGC
JAM2_X2_B_R	TCTTTGCACATCCGGTCTTT
JAM2_X8_B_F	GCCCCAAAGCCTAAAATGGT
JAM2_X8_B_R	AGTGTTGGTGCCAGGATTGT
JAM2_X9_A_F	GTGCACAAGGCTTCACACTG
JAM2_X9_A_R	TGCCATCCTAGACAGGGTACA
JAM2_X1_B_F	CTTCCCGCCCCAGAAGTTC
JAM2_X1_B_R	TGCATTGGAATCACGTCCCT
JAM2_X5_A_F	GCAGTCTCTATCACAGGGTCC
JAM2_X5_A_R	AGTGCTGCCATCTTTCCTGG
JAM2_X4_A_F	CCATCATGCCTGGCCTTTTG
JAM2_X4_A_R	TGGTCTCTGTAACCTGACTGGT
JAM2_X6_2_F	TGCTAAAATTTGAGGGACAGGG
JAM2_X6_2_R	TGGCTTATTTGGGAGATGGC

7.4.2 *SLC20A2* (from (244))

X2	F	CATGCCAAAGTTAGATCCCA
X2	R	AGAAAATAAATGGTTGCCTGA
X3	F	CGCTTTGTAAAGAAACAATCACA
X3	R	GCTCACGCCTATAATCCTG
X4	F	GTCAGCTCTGCCAAGTCA
X4	R	ACAATTATTCCTCTAACCCCTC
X5	F	CAACAGTGGGCTCTTTGACA
X5	R	TTACTATCAGCCAACAACCTCC
X6	F	TTAAGCACATATTCGCCAGA

X6	R	CTTCCAGTTACTCATGGCAAC
X7	F	CCTGGCCTCAACTTCATTTTCTC
X7	R	CCCCAGTGCCTCCGGTTAG
X8	F	GGCATGGTGTGCGCCTTGAG
X8	R	CCGGCGACCTCCTAGCTTGT
X9	F	CCGCGGCTGTAGTCTCAATTA
X9	R	GGGGCCTGTTAAGTCTGTGC
X10	F	GCGGCCTCTTGTCTGTAAAAT
X10	R	CCCGGAGACCTGGAGAACCT
X11	F	GCTGAAGAGAAGAATCCCCAAAC
X10	F	GCGGCCTCTTGTCTGTAAAAT
X10	R	CCCGGAGACCTGGAGAACCT
X11	F	GCTGAAGAGAAGAATCCCCAAAC

7.4.3 PDGFB

PDGFB-Ex1_F	aggcctgagcgcctgatcgc
PDGFB-Ex1_R	cgctggtgccttcccttaga
PDGFB-Ex2_F	gaggcctttgtgctcctgat
PDGFB-Ex2_R	caagtcccaggtaccaacc
PDGFB-Ex3_F	ctggaaggaggactgttct
PDGFB-Ex3_R	agttcgctcagtcctgaatgt
PDGFB-Ex4_F	taatgacagccaggacttgaaac
PDGFB-Ex4_R	tgcccagtcgaaggaagcctggtca
PDGFB-Ex5_F	ccgggctttcgaggaaagat
PDGFB-Ex5_R	cttgtgtctcagcaagatg
PDGFB-Ex6_F	agaagggtccatggcaggccttggt
PDGFB-Ex6_R	cacaggattctgggcctcagttg

7.4.4 EIF4G1

Exon	basepairs covered	Forward	Reverse
Exon_8A	605	TGGAGTGACTIONTGA GGGTAC	GTCATGAATTTCCACT GTGTG
Exon_8B	662	TCTCGCCGAACCCATAC TG	CAGGGACCCAGAAACA TGTC
Exon_10	292	GAGCAGTGGTCATTCTG CAA	GCCTCCTCTGGCCCTAA TAA
Exon_22	403	TGCTAAGAACAAGGCC CAACAG	CTAGTCCCAAGGCAGCC AATG

7.4.5 VPS35

Exon	Forward	Reverse

Exon_15	TGCTCAACTAGAGGATGGTTG	ACACAAGGCCATGACAACCTG
---------	-----------------------	-----------------------

7.4.6 COQ2

Exon	Forward	Reverse	Alternative forward
1 option a	gtgttgcccgataatggaac	gactcggaggctgctacttg	aaggatgaggaaaggttctg
1 option b	cctagagtaagcgaccacgatg	tgaaggaggccacagaaa	
2	agtaaggggtcctttgtgat	tcactgaatgatcttgttgc	
3	gggccagtctcttcattaa	cttaactccttggctgaaa	
4	aagtcgaaggctaggaagat	aaaatagctaactgctctcc	
5	cactgaacacactccgatg	ccatggaactgtgaatgac	
6	tagtggttaattggtgcaca	gtaaacacagaggcactactg	
7	ctgtttctcctccgtgta	gctctaaatcttcatcttcagg	

7.4.7 C9orf72 (from (354))

C9ORF72_revers e	CAGGAAACAGCTATGACCGGGCCCGCCCCGACCACGCCCCGGCCCCGGC CCCGG
C9ORF72_forwa rd	FAM-CAAGGAGGGAAACAACCGCAGCC
C9ORF72_anch r	CAGGAAACAGCTATGACC

7.4.8 *LRRK2* exon 34

F: GGTACTGTGTTGCACTTGAAAA

R: CAGTAGGAGGTTTACTACTAGAAGC