

Electronic Thesis and Dissertation Repository

8-5-2016 12:00 AM

Gene Discovery in Mendelian and Complex Diseases

Sali Farhan, *The University of Western Ontario*

Supervisor: Dr. Robert A. Hegele, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry

© Sali Farhan 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Computational Biology Commons](#), [Genetics Commons](#), [Genomics Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Farhan, Sali, "Gene Discovery in Mendelian and Complex Diseases" (2016). *Electronic Thesis and Dissertation Repository*. 4077.

<https://ir.lib.uwo.ca/etd/4077>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Through the Finding of Rare Disease Genes in Canada (FORGE Canada) initiative, individuals affected with rare Mendelian diseases were clinically ascertained with a goal of identifying the genetic origin of their disease. Herein, I describe the methods for identifying the genetic basis of four Mendelian diseases. The application of next generation sequencing led to the discovery of non-synonymous variation in the DNA of individuals affected by rare diseases. The effects of the candidate variants were assessed using a series of functional experiments to complement the human genetics data. The variants observed in patients' cells are extremely rare, were consistently predicted to be pathogenic by multiple *in silico* predictive programs, segregated with disease status in the family, and affected the biological properties of their respective gene products, as measured by functional assays.

Having successfully identified genetic variants underlying the Mendelian diseases, we sought to use the same approach to extract the genetic variation that may predispose individuals to complex diseases, primarily neurodegenerative disorders. We designed a neurodegeneration specific gene panel that utilizes next generation sequencing chemistry. We sequenced patients diagnosed with one of five neurodegenerative diseases: 1) Alzheimer's disease; 2) amyotrophic lateral sclerosis (ALS); 3) frontotemporal dementia (FTD); 4) Parkinson's disease; or 5) vascular cognitive impairment, as part of the Ontario Neurodegenerative Disease Research Initiative (ONDRI). We were successful in detecting rare variants in a large fraction of cases that may be related to the neurodegenerative phenotypes.

We also independently ascertained three large, unique families affected with familial ALS and FTD across multiple generations. The three families are diagnosed with ALS and/or FTD and in which the same hexanucleotide repeat expansion in the *C9orf72* gene has been observed in all three pedigrees, but their phenotypes vary significantly. We sequenced affected individuals and observed several, distinct variants in these families that may explain the additional neurodegenerative phenotypes observed. In summary, the application of next generation sequencing has successfully identified novel genetic loci in both Mendelian and complex diseases.

Keywords

Genetic variation; massively parallel sequencing, next generation sequencing, exome sequencing, targeted sequencing, Mendelian disease, complex disease, rare disease, neurodevelopment, metabolic disease, neurodegeneration, dementia, amyotrophic lateral sclerosis, frontotemporal dementia, autosomal recessive, autosomal dominant, oligogenic inheritance.

Co-Authorship Statement

In all manuscripts listed here, I am the first author and I performed the experiments, unless specified in the Materials and methods section of each Chapter. I also analyzed the data and wrote each manuscript with guidance from Dr. Robert A. Hegele and Dr. Michael J. Strong.

Dr. Robert A. Hegele (primary supervisor) provided funding, supervision, contributed to study design, manuscript preparation, and critical revision for all Chapters. Dr. Michael J. Strong (co-supervisor) provided funding, supervision, contributed to study design, manuscript preparation, and critical revision for Chapters 6 and 7.

Chapter 2 contains material from the manuscript entitled, ‘Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency’ published in *Molecular Genetics & Genomic Medicine* on January 10, 2014 and co-authored by Dr. Jian Wang, John F. Robinson, Dr. Piya Lahiry, Dr. Victoria M. Siu, Dr. Chitra Prasad, Dr. Jonathan B. Kronick, Dr. David A. Ramsay, Dr. C. Anthony Rugar, and Dr. Robert A. Hegele (PMID: 24498631).

Chapter 3 contains material from the manuscript entitled, ‘Old gene, new phenotype: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses’ published in *Journal of Medical Genetics* on August 5, 2015 and co-authored by Dr. Jian Wang, John F. Robinson, Dr. Asuri N. Prasad, Dr. C. Anthony Rugar, Dr. Victoria M. Siu, and Dr. Robert A. Hegele (PMID: 26246518).

Chapter 4 contains material from the manuscript entitled, ‘Linkage analysis and exome sequencing identify a novel mutation in KCTD7 in patients with progressive myoclonus epilepsy with ataxia’ published in *Epilepsia* on July 24, 2014 and co-authored by Lisa M. Murphy, John F. Robinson, Dr. Jian Wang, Dr. Victoria M. Siu, Dr. C. Anthony Rugar, Dr. Asuri N. Prasad, and Dr. Robert A. Hegele (PMID: 25060828).

Chapter 5 contains material from the manuscript entitled, ‘TMTC3 is a synaptic protein involved in seizure susceptibility and intellectual disability’ which has been submitted for publication, and co-authored by Kevin C.J. Nixon, Michelle Everest, Tara Edwards, Shirley Long, Dmitri Segal, Maria J. Knip, Dr. Heleen H. Arts, Dr. Rana Chakrabarti, Dr. Jian Wang, John F. Robinson, Dr. C. Anthony Rupa, Dr. Victoria M. Siu, Dr. Michael O. Poulter, Dr. Robert A. Hegele, and Dr. Jamie M. Kramer.

Chapter 6 contains material from the manuscript entitled, ‘The ONDRISeq panel: custom designed next generation sequencing of genes related to neurodegeneration’ accepted in *npj Genomic Medicine* on August 5, 2016 and co-authored by Allison A. Dilliot, Dr. Mahdi Ghani, Christine Sato, Eric Liang, Dr. Ming Zhang, Adam D. McIntyre, Dr. Henian Cao, Dr. Lemuel Racacho, John F. Robinson, Dr. Michael J. Strong, Dr. Mario Masellis, Dr. Peter St. George-Hyslop, Dr. Dennis E. Bulman, Dr. Ekaterina Rogava, and Dr. Robert A. Hegele.

Chapter 7 contains material from the manuscript entitled, ‘Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia’ which has been submitted for publication, and co-authored by Dr. Tania F. Gendron, Dr. Leonard Petrucelli, Dr. Robert A. Hegele, and Dr. Michael J. Strong.

In loving memory of my big brother, Doureid

Acknowledgements

During graduate school, I have had the privilege to work with and be mentored by excellent scientists and clinicians, extraordinary teachers, inspiring trainees, and I have also developed incredible friendships that I hope to maintain throughout my life.

First, I would like to thank my supervisor, Dr. Rob Hegele, for providing me with an excellent training environment. Thank you for your guidance and support throughout graduate school. I sincerely appreciate the autonomy you have given me during experimental planning and project management. On a personal note, your resiliency and composure during the most difficult situations are the greatest lessons you have unknowingly taught me.

I would also like to acknowledge my co-supervisor, Dr. Michael Strong, for welcoming me into his lab and teaching me about amyotrophic lateral sclerosis and frontotemporal dementia. Thank you for always including me in lab meetings and social gatherings. I sincerely appreciate your willingness to mentor me throughout graduate school and in my career.

Thank you to Dr. Dave Litchfield and Dr. Greg Gloor, for providing me with excellent mentorship and support during committee meetings and elsewhere. I genuinely value the feedback and support you have provided me while I prepared for my qualifying exam and completing my Doctoral dissertation. Sincere thanks to Dr. Murray Huff for also providing me with spontaneous mentorship in the Robarts hallways, and for allowing me to borrow various lab tools that I fully intend on returning one day.

To the members of the Hegele lab, thank you for your collegiality and support throughout my time at Robarts. Thank you to Dr. Jian Wang, John Robinson, Adam McIntyre, and Dr. Henian Cao, for their scientific and technical expertise during experimental planning and data analysis.

To the members of the Strong lab, thank you for welcoming me into your lab. Thank you Dr. Kathryn Volkening, Cheryl Leystra-Lantz, and Wendy Strong, for guiding me during experimental planning and data analysis. Thank you Dr. Danae Campos-Melo and Dr. Cristian Droppelmann for being inspirational mentors during lab meetings, journal clubs, and elsewhere.

I am grateful to the Department of Biochemistry for providing me with research and administrative support. Thank you to Barb Green and Melita Hayes for their guidance throughout graduate school.

I would also like to thank my Robarts friends: Amy, Alex, Kevin, Michael, Arielle, Jacqueline, Zack, and Josh for first being great colleagues and then great friends. Thank you for the laughs, the memories, and the great conversations whether it was at 2 pm or 2 am. To all of my wonderful friends outside of graduate school who are always eager to listen to me talk science, I am forever grateful to you.

To my family, thank you for always supporting me to pursue higher education and for guiding me to a life of happiness. Thank you to my loving parents who have sacrificed their careers, home, and dreams for their children. Your comfort and happiness are always my priorities. Thank you to my sweet and caring big brother, Doureid, for being a wonderful part of my life. To my big sisters, Sara and Jasmine, thank you for your love, support, and friendship.

Finally, to the apple of my eye, Colin, thank you for being a compassionate and reassuring partner, and for walking with me to the mountaintops and through the valleys.

Funding Acknowledgements

The work completed in this PhD thesis was supported by grants from the Canadian Institutes of Health Research (CIHR), Finding of Rare Disease Genes in Canada (FORGE Canada), and Genome Canada, and the Ontario Neurodegenerative Disease Research Initiative.

I am thankful to CIHR for supporting me through the CIHR Strategic Training Program, the CIHR Frederick Banting and Charles Best Doctoral Award, and the CIHR Michael Smith Foreign Study Supplement award. I am also thankful to the Government of Ontario for their support through the Ontario Graduate Scholarship (OGS) program. Finally, I am grateful for the financial support from the Schulich School of Medicine and Dentistry and the School of Graduate and Postdoctoral Studies at Western University.

Table of Contents

Title Page	i
Abstract	ii
Keywords	iii
Co-Authorship Statement	iv
Dedication	vi
Acknowledgements	vii
Table of Contents	x
List of Tables	xix
List of Figures	xx
List of Appendices	xxii
List of Abbreviations	xxiii
Chapter 1 - Introduction	
1 Overview	1
1.1 Variation in the human genome	1
1.1.1 Inherited and <i>de novo</i> variation	2
1.1.2 Germline and somatic variation	2
1.1.3 Structural variation	
1.1.3.1 Chromosomal abnormalities: numerical and structural chromosomal aberrations	3
1.1.3.2 Copy number variation	3
1.1.3.3 Variable number of tandem repeats	4
1.1.3.4 Insertions and deletions	5
1.1.4 Single nucleotide variation	5
1.2 Genetic diseases in humans	
1.2.1 Monogenic (Mendelian) diseases	6
1.2.2 Digenic diseases	7
1.2.3 Polygenic (complex) diseases	8
1.3 Phenotypic variation	9
1.3.1 Heritability	10
1.4 Epigenetics and human diseases	10
1.5 Approaches to studying genetic diseases	11
1.5.1 Family studies	
1.5.1.1 Linkage studies	12
1.5.1.2 Candidate gene approaches	13
1.5.2 Population studies	
1.5.2.1 Association studies	13
1.6 Tools to study genetic variation in disease	14
1.6.1 Sequencing versus genotyping: benefits and pitfalls	14
1.6.2 Massively parallel sequencing	14

1.6.3	Molecular studies and model organisms	15
1.7	Variant causality criteria	16
1.8	Thesis outline	19
1.8.1	Finding of Rare Disease Genes (FORGE) Canada	
1.8.1.1	Project rationale	20
1.8.1.2	Project overview	20
1.8.1.3	Project aims	20
1.8.1.4	Project significance	21
1.8.1.5	My role in the project	21
1.8.2	The Ontario Neurodegenerative Disease Research Initiative (ONDRI)	
1.8.2.1	Project rationale	22
1.8.2.2	Project overview	22
1.8.2.3	Project aims	23
1.8.2.4	Project significance	24
1.8.2.5	My role in the project	25
1.8.3	Familial ALS and FTD study: defining the role of oligogenic inheritance in neurodegenerative disease	
1.8.3.1	Project rationale	25
1.8.3.2	Project overview	26
1.8.3.3	Project aims	26
1.8.3.4	Project significance	27
1.8.3.5	My role in the project	27
1.9	Thesis hypotheses	27
1.10	Thesis aims	28
1.11	Conclusion	28
1.12	References	29

Chapter 2 - Novel phenotype, novel gene: NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency

2	Study rationale	36
2.1	Overview	36
2.2	Introduction	37
2.3	Materials and methods	
2.3.1	Ethics	38
2.3.2	Patients and biological materials	38
2.3.3	DNA isolation	39
2.3.4	Enzymology assays	39
2.3.5	Histology	39
2.3.6	SNP genotyping	40
2.3.7	Autozygosity mapping	40
2.3.8	Exome sequencing	40
2.3.9	Sequence alignment	41

2.3.10 Variant calling and annotation	41
2.3.11 Variant discovery	
2.3.11.1 Prioritization of homozygous, non-synonymous, and rare variants	42
2.3.11.2 <i>in silico</i> analyses	42
2.3.12 Variant validation	43
2.3.12.1 PCR	43
2.3.12.2 Imaging, purifying, and sequencing of PCR products	43
2.3.13 Population screening	44
2.3.13.1 Genotyping of local population	44
2.3.13.2 Genotyping of an ethnically diverse cohort	45
2.3.13.2.1 TaqMan genotyping assay	45
2.3.13.2.2 Pooling samples	45
2.3.13.2.2.1 Restriction enzyme digestion	46
2.3.13.2.2.2 Visualizing and inferring genotypes	46
2.3.14 Cell culture	46
2.3.15 Immunoblotting	46
2.3.16 Quantitative PCR	47
2.3.17 Co-immunoprecipitation	47
2.3.18 Antibodies	48
2.3.19 Statistics	48
2.4 Results	
2.4.1 Clinical characterization of IMC23D	48
2.4.2 Histopathological evaluation of affected individuals	55
2.4.3 Autozygosity mapping generated 20p11.2-q13.1 as a candidate region	57
2.4.4 Exome sequencing and <i>in silico</i> analyses generate candidate variants	59
2.4.5 Population screens demonstrate the rarity of NFS1 p.Arg72Gln	62
2.4.6 Deficiency in NFS1 expression in patients with IMC23D	62
2.4.7 Characterization of IMC23D disease summary: comparing NFS1 and CDH22	65
2.5 Discussion	68
2.6 Conclusion	71
2.7 References	71

Chapter 3 - Novel phenotype, known gene: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses

3 Study rationale	76
3.1 Overview	76
3.2 Introduction	77
3.3 Materials and methods	

3.3.1 Ethics	78
3.3.2 Patients and biological materials	78
3.3.3 DNA isolation	79
3.3.4 Genotyping	79
3.3.5 Autozygosity mapping	79
3.3.6 Exome sequencing	79
3.3.7 Sequence alignment	79
3.3.8 Variant calling and annotation	80
3.3.9 Variant discovery	
3.3.9.1 Prioritization of homozygous, non-synonymous, and rare variants	80
3.3.9.2 <i>in silico</i> analyses	80
3.3.10 Variant validation	
3.3.10.1 PCR	81
3.3.10.2 Imaging, purifying, and sequencing of PCR products	81
3.3.11 Population screening	
3.3.11.1 Genotyping of local population	81
3.3.11.2 Identifying overall MAF of <i>EXT2</i> variants	81
3.3.12 Cell culture	
3.3.12.1 Patient cells	82
3.3.12.2 Mutagenesis	82
3.3.12.3 Transfection	82
3.3.13 Protein isolation	83
3.3.14 Immunoblotting	83
3.3.15 Reverse transcriptase (RT)-PCR	83
3.3.16 Antibodies	83
3.3.17 Statistics	84
3.4 Results	
3.4.1 Clinical description of patients with SSM syndrome	84
3.4.2 Genome-wide autozygosity mapping reveals highly significant homozygous regions	91
3.4.3 Compound homozygous <i>EXT2</i> variants in patients with SSM syndrome	93
3.4.4 <i>EXT2</i> variants segregate in SSM syndrome family	95
3.4.5 Population screening reveals <i>EXT2</i> variants are ultra-rare	97
3.4.6 Decreased <i>EXT2</i> expression and activity in patients with SSM syndrome	97
3.4.7 Both <i>EXT2</i> variants are necessary for the development of SSM syndrome	99
3.5 Discussion	102
3.6 Conclusion	108
3.7 References	109

Chapter 4 - Known phenotype, known gene: linkage analysis and exome sequencing identify a novel mutation in *KCTD7* in patients with progressive myoclonus epilepsy with ataxia

4 Study rationale	112
4.1 Overview	113
4.2 Introduction	113
4.3 Materials and methods	
4.3.1 Ethics	114
4.3.2 Patients and biological materials	115
4.3.3 DNA isolation	115
4.3.4 Genotyping	115
4.3.5 Autozygosity mapping	115
4.3.6 Exome sequencing	115
4.3.7 Sequence alignment	116
4.3.8 Variant calling and annotation	116
4.3.9 Variant discovery	
4.3.9.1 Prioritization of homozygous, non-synonymous, and rare variants	116
4.3.9.2 <i>in silico</i> analyses	116
4.3.10 Variant validation	116
4.3.10.1 PCR	117
4.3.10.2 Imaging, purifying, and sequencing of PCR products	117
4.3.11 Population screening	
4.3.11.1 Identifying overall MAF of <i>KCTD7</i> variant	117
4.4 Results	
4.4.1 Clinical description of patients with progressive myoclonus epilepsy with ataxia	117
4.4.2 Electroencephalogram features	123
4.4.3 Genome-wide autozygosity mapping generated a high priority region on chromosome 7p12.1-7q11.22	127
4.4.4 <i>KCTD7</i> p.Tyr276Cys in patients with progressive myoclonus epilepsy with ataxia	129
4.4.5 <i>KCTD7</i> p.Tyr276Cys segregates with disease phenotype in the family	131
4.4.6 <i>KCTD7</i> is a known disease gene	133
4.5 Discussion	135
4.6 Conclusion	138
4.7 References	139

Chapter 5 - Known phenotype, novel gene: loss of a novel synaptic protein, *TMTC3*, in patients with nocturnal seizures with developmental delay

5 Study rationale	143
5.1 Overview	143
5.2 Introduction	144
5.3 Materials and methods	
5.3.1 Ethics	146
5.3.2 Patients and biological materials	146
5.3.3 DNA isolation	146
5.3.4 Exome sequencing	146
5.3.5 Sequence alignment	146
5.3.6 Variant calling and annotation	146
5.3.7 Variant discovery	
5.3.7.1 Prioritization of autosomal recessive, non-synonymous, and rare variants	147
5.3.7.2 <i>in silico</i> analyses	147
5.3.8 Variant validation	
5.3.8.1 PCR	147
5.3.8.2 Imaging, purifying, and sequencing of PCR products	147
5.3.9 Cell culture	148
5.3.10 Protein isolation	148
5.3.11 Immunoblotting	148
5.3.12 Reverse transcriptase (RT)-PCR	148
5.3.13 Model organisms	148
5.3.13.1 Gateway cloning technology	148
5.3.13.2 <i>Drosophila</i> stocks	149
5.3.13.3 RT-qPCR	149
5.3.13.4 Bang sensitivity	150
5.3.13.5 Immunohistochemistry, image acquisition, and analysis	150
5.3.14 Antibodies	152
5.3.15 Statistics	152
5.4 Results	
5.4.1 Clinical presentation of patients with nocturnal seizures with developmental delay	153
5.4.2 Compound heterozygous <i>TMTC3</i> variants in patients with nocturnal seizures with developmental delay	156
5.4.3 Loss of <i>TMTC3</i> in patients with nocturnal seizures and developmental delay	158
5.4.4 Neuronal knockdown of <i>Drosophila tmtc3</i> causes increased susceptibility to mechanically induced seizures	161
5.4.5 <i>TMTC3</i> is localized at presynaptic terminals in rat brains	167
5.5 Discussion	170
5.6 Conclusion	172
5.7 References	172

Chapter 6 - The ONDRISeq panel: custom designed next generation sequencing of genes related to neurodegeneration as part of the Ontario Neurodegenerative Disease Research Initiative

6 Study rationale	176
6.1 Overview	176
6.2 Introduction	177
6.3 Materials and methods	
6.3.1 Design of ONDRISeq	179
6.3.2 Sample collection and DNA isolation	187
6.3.3 Library preparation	187
6.3.4 Next generation sequencing	188
6.3.5 Sequence alignment	188
6.3.6 Variant calling	189
6.3.7 Variant annotation	189
6.3.8 Variant classification and prioritization	189
6.3.9 <i>APOE</i> genotyping	190
6.3.10 Variant validation	191
6.3.10.1 Variant validation 1: NeuroX	191
6.3.10.2 Variant validation 2: TaqMan allelic discrimination	192
6.3.10.3 Variant validation 3: Sanger sequencing	192
6.3.10.3.1 PCR	193
6.3.10.3.2 Imaging, purifying, and sequencing of PCR products	193
6.3.10.4 Variant validation 4: <i>SOD1</i> testing	193
6.3.11 <i>C9orf72</i> genotyping	193
6.3.12 Statistical analysis	194
6.4 Results	
6.4.1 Study subjects	194
6.4.2 Quality assessment of ONDRISeq data	196
6.4.3 ONDRISeq is concordant with NeuroX, TaqMan allelic discrimination assay, and Sanger sequencing	198
6.4.4 Genetic variation in patients with neurodegenerative disease	198
6.4.4.1 <i>C9orf72</i> hexanucleotide expansion in patients with neurodegenerative disease	199
6.4.4.2 Genetic variation identified in patients with neurodegenerative disease using ONDRISeq	199
6.4.4.3 <i>APOE</i> genotypes in patients with neurodegenerative disease	200
6.4.4.4 Case study: strong evidence of pathogenicity for APP p.Ala713Thr in AD patient	203
6.5 Discussion	206
6.6 Conclusion	207

6.7 References	208
----------------	-----

Chapter 7 - Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia

7 Study rationale	212
7.1 Overview	213
7.2 Introduction	213
7.3 Materials and methods	
7.3.1 Ethics	215
7.3.2 Sample collection clinical assessment	215
7.3.3 DNA isolation	215
7.3.4 <i>C9orf72</i> testing	216
7.3.4.1 Amplicon length analysis	216
7.3.4.2 Repeat-primed PCR	216
7.3.4.3 Immunohistochemistry	217
7.3.4.4 Southern immunoblotting	217
7.3.5 <i>ATXN2</i> expansion testing	218
7.3.6 Next generation sequencing	218
7.3.7 Sequence alignment	219
7.3.8 Variant calling	219
7.3.9 Variant annotation	219
7.3.10 Variant classification and prioritization	219
7.3.11 Variant validation	219
7.3.11.1 PCR	220
7.3.11.2 Imaging, purifying, and sequencing of PCR products	220
7.4 Results	
7.4.1 Clinical description	220
7.4.2 Variants identified in patients with ALS	227
7.4.3 <i>C9orf72</i> dipeptide immunostaining	236
7.4.4 Oligogenic inheritance in patients	238
7.5 Discussion	241
7.6 Conclusion	245
7.7 References	245

Chapter 8 - Discussion

8.0 Summary	248
8.1 Context of study findings	
8.1.1 Chapter 2: <i>NFS1</i> , a novel disease gene underlying a novel metabolic disease	248
8.1.2 Chapter 3: Expanding the biological function of <i>EXT2</i> , a known disease gene in a novel neurodevelopmental disease	249

8.1.3 Chapter 4: Confirming the association of <i>KCTD7</i> in progressive myoclonus epilepsy	250
8.1.4 Chapter 5: The application of a model organism to understand the function of a novel disease gene <i>TMTC3</i> , in neurodevelopment	251
8.1.5 Chapter 6: Developing a custom sequencing based method to study neurodegeneration	252
8.1.6 Chapter 7: The effect of multiple genetic variants on neurodegenerative disease risk	253
8.2 Methodological considerations	254
8.2.1 Study strengths and implications	255
8.2.2 Additional study caveats	256
8.3 Future directions	257
8.4 Final conclusions	260
8.5 References	260
Appendices	266
Curriculum Vitae	277

List of Tables

Table 1.7.1 Population databases of healthy controls	17
Table 1.7.2 Disease databases	17
Table 1.7.3 Examples of <i>in silico</i> predictive tools	18
Table 2.4.1 Clinical and biochemical findings of patients with IMC23D	53
Table 2.4.7 Comparison of <i>NFS1</i> and <i>CDH22</i> as the cause of IMC23D	65
Table 3.4.1 Clinical description of patients with SSM syndrome	87
Table 3.5.1 Hereditary diseases with abnormal heparan sulfate levels	104
Table 4.4.1 Clinical description of patients with epilepsy with ataxia	121
Table 4.4.6 Comparison of all reported patients with <i>KCTD7</i> variants	134
Table 5.4.1 Clinical features of patients with nocturnal seizures with ID	155
Table 6.3.1 Genes associated with amyotrophic lateral sclerosis, frontotemporal dementia, Alzheimer's disease, Parkinson's disease, or vascular cognitive impairment as represented on the ONDRISeq targeted resequencing panel	181
Table 6.4.1 Patient demographics	195
Table 6.4.2 Quality control metrics for sequencing runs on ONDRISeq	197
Table 6.4.4.1 Other risk variants identified in a cohort of 216 disease cases	201
Table 6.4.4.2.1 Diagnostic yield of ONDRISeq in a cohort of 216 disease cases	201
Table 6.4.4.2.2 Variants identified in a cohort of 216 disease cases as detected by ONDRISeq	202
Table 7.4.4 Summary of clinical and genetic information	240
Table 7.5 ALS and/or FTD patients with multiple genetic variants	243

List of Figures

Figure 2.4.1 Pedigree of IMC23D	52
Figure 2.4.2 Pathological findings in patients with IMC23D	56
Figure 2.4.3 Autozygosity mapping of IMC23D family	58
Figure 2.4.4 Mapping and exome sequencing of IMC23D family identifies a highly conserved and destabilizing missense mutation, p.(Arg72Gln) in <i>NFS1</i>	61
Figure 2.4.6 Depletion in NFS1 Protein and Transcript Levels in Patients with IMC23D	64
Figure 3.4.1 Pedigree with four children affected with SSM syndrome	86
Figure 3.4.2 Autozygosity mapping generates highly significant homozygous regions	92
Figure 3.4.3 Schematic of the genetic and bioinformatic studies in the SSM family	94
Figure 3.4.4 <i>EXT2</i> variants segregate with affected individuals in the family	96
Figure 3.4.6 Decreased <i>EXT2</i> expression and activity in patients with SSM syndrome	98
Figure 3.4.7 Both <i>EXT2</i> variants are necessary for the development of SSM syndrome	101
Figure 4.4.1 Pedigree with three daughters affected with epilepsy with ataxia	120
Figure 4.4.2 EEG recording of the affected individuals	126
Figure 4.4.3 Homozygous region on chromosome 7 unique to affected individuals	128
Figure 4.4.4 Schematic of mutation discovery	130
Figure 4.4.5 <i>KCTD7</i> g.661041A>G segregates with affected individuals in the family	132

Figure 5.4.1 Pedigree with four children affected with nocturnal seizures with developmental delay	154
Figure 5.4.2 Compound heterozygous <i>TMTC3</i> variants in patients with nocturnal seizures with developmental delay	157
Figure 5.4.3 Depleted TMTC3 protein and reduced transcript expression in cells of patients affected by nocturnal seizures with developmental delay	160
Figure 5.4.4 <i>Tmtc3</i> deficiency confers susceptibility to seizures in <i>Drosophila melanogaster</i>	165
Figure 6.4.4.4 <i>APP</i> variant in AD case	205
Figure 7.4.1 Pedigrees of three families affected with ALS or ALS-FTD	226
Figure 7.4.2.1 Genetic analyses of <i>C9orf72</i> genotypes of representative individuals from each family	229
Figure 7.4.2.2 Genetic analyses of <i>C9orf72</i> expansion profiles of representative individuals from each family	231
Figure 7.4.2.3 Genetic analysis of <i>ATXN2</i> genotyping profiles of family 2	233
Figure 7.4.2.4 Validation of variants in <i>OPTN</i> and <i>ARHGEF28</i>	235
Figure 7.4.3 Cerebellar dipeptide repeat protein pathology	237

List of Appendices

Appendix A - Ethics approval	266
Appendix B - Chapter 6: Primer list	269
Appendix C - Chapter 7: Expanded genetic and clinical information on all individuals within the study	270
Appendix D - Journal copyright approval	275

List of Abbreviations

ACD, autosomal co-dominant

ACMG, American College of Medical Genetics and Genomics

AD, Alzheimer's disease

ADm, autosomal dominant

ALS, amyotrophic lateral sclerosis

ALSoD, ALS online database

aMCI, amnesic single or multidomain mild cognitive impairment

AMRF, action myoclonus-renal failure

AR, autosomal recessive

AST, aspartate aminotransferase

bvFTD, behavioural variant FTD

BWA, Burrows-Wheeler Aligner

CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy

CARASIL syndrome, cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy

CH, compound heterozygous

CK, creatine kinase

CLIA, clinical laboratory improvement amendments

CMT disease, Charcot-Marie-Tooth disease

CNV, copy number variation

COX, cytochrome C oxidase

dbGaP, the database of Genotypes and Phenotypes

DIF, digoxigenin

DMSO, dimethylsulfoxide

DNA, deoxyribonucleic acid

DOB, date of birth

ENFL, nocturnal frontal lobe epilepsy
ExAC, Exome Aggregation Consortium
FORGE, Finding of Rare Disease Genes
FTD, frontotemporal dementia
GATK, Genome Analysis Toolkit
gDNA, genomic DNA
GMT, Gomori modified trichrome
GTEx, Genotype-Tissue Expression project
GWAS, genome-wide association study
HDLS, leukoencephalopathy, diffuse hereditary, with spheroids
HET4, hereditary essential tremor, 4
HGMD, Human Gene Mutation Database
HME, hereditary multiple exostoses
HMN7B, neuropathy, distal hereditary motor, type VIIB
HPS, hematoxilin, phloxin, and saffron
HSN1E, hereditary sensory neuropathy type 1E
HZ, homozygous
IMC23D, infantile mitochondrial complex II/III deficiency
Indels, insertions and deletions
IR, inverted repeats
LBD, Lewy body dementia
LINEs, long interspersed transposon derived elements
LOD, logarithm of odds
MAF, minor allele frequency
MELAS, mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes
MERF, myoclonic epilepsy with ragged red fibers
MND, motor neuron disease
MoCA, Montreal Cognitive Assessment
NADH, reduced nicotinamide adenine dinucleotide-tetrazolium reductase

NARP, neurogenic muscle weakness, ataxia, and retinitis pigmentosa
NBIA2A, neurodegeneration with brain iron accumulation 2A
NBIA2B, neurodegeneration with brain iron accumulation 2B
NCL, neuronal ceroid lipofuscinoses
NHLBI ESP, National Heart, Lung, and Blood Institute Exome Sequencing Project
OMIM, Online Mendelian Inheritance in Man
ONDRI, Ontario Neurodegenerative Disease Research Initiative
PAS, periodic acid-Schiff (PAS)
PCR, polymerase chain reaction
PD, Parkinson's disease
PKU, phenylketonuria
PMID, PubMed identification
PolyPhen-2, Polymorphism Phenotyping version 2
PSS Potocki-Shaffer syndrome
qRT-PCR, quantitative reverse transcriptase PCR
RNA, ribonucleic acid
RT-PCR, reverse transcriptase PCR
SDH, succinic dehydrogenase
SDS, sodium dodecyl sulfate
SIFT, Sorting Intolerant From Tolerant
SINEs, short interspersed transposon derived elements
SNP, single nucleotide polymorphism
SNV, single nucleotide variant
SSC, standard sodium citrate
SSM, seizures-scoliosis-macrocephaly
STR, short tandem repeats
TCAG, The Centre for Applied Genomics
VCF, variant calling format
VCI, vascular cognitive impairment

VNTR, variable number of tandem repeats

YV syndrome, Yunis-Varon syndrome

Chapter 1 - Introduction

1 Overview

Materials from the following texts with appropriate modifications, were incorporated in Chapter 1:

(1) **Farhan, SMK.**, and Hegele, RA. “Genetics 101 for cardiologists: Rare genetic variants and monogenic cardiovascular disease”. *Canadian Journal of Cardiology*. 2013; 29(1): 18-22 (PMID: 23200093).

(2) **Farhan, SMK.**, and Hegele, RA. “Exome sequencing: new insights into lipoprotein disorders”. *Current Cardiology Reports*. 2014; 16(7): 507-517 (PMID: 24893940).

(3) **Farhan, SMK.**, and Prasad, AN. “Exploring the Epileptome I: Genetics of Age dependent Epileptic Encephalopathies”. *Pediatric Epilepsy*, 4th edition, Chapter 7.

(4) Prasad, AN., and **Farhan, SMK.** “Exploring the Epileptome II: Approaching the Complex Epilepsies”. *Pediatric Epilepsy*, 4th edition, Chapter 8.

(5) **Farhan, SMK.**, ONDRI Investigators, Strong, MJ. “The Ontario Neurodegenerative Disease Research Initiative (ONDRI)”. *Canadian Journal of Neurological Sciences*. 2016; Accepted.

1.1 Variation in the human genome

Variation in the human genome, which is composed of approximately three billion base pairs, can lead to benign or pathogenic biological effects. These DNA variations often classified as neutral, adaptive, or deleterious, can be under positive selection (Darwinian selection) or negative selection (purifying selection) based on their effect on an individual's fitness (Fay et al., 2001). DNA variations that increase an individual's fitness are under positive selection; conversely, deleterious variants, which decrease an

individual's fitness, are under negative selection. Importantly, some pathogenic variants can escape natural selection by exerting their effects post sexual reproduction. In this section, I will introduce the various classes of genetic variation observed in the human genome and provide examples of diseases caused by each class of variation.

1.1.1 Inherited and *de novo* variation

Inherited variations are DNA changes transmitted through the germline from parent to offspring. Conversely, *de novo* or sporadic variants spontaneously arise in the offspring's germ cells and are absent from the parents' germ cells. Importantly, if *de novo* variations are not deleterious, they can be transmitted to the progeny and are reclassified as inherited variation. *De novo* variations are the most rare class of genetic variation and have recently been of great interest in neurodevelopmental diseases, primarily autism spectrum disorders (Ku et al., 2012; Neale et al., 2012; Yuen et al., 2015).

1.1.2 Germline and somatic variation

Errors during DNA replication that escape polymerase proofreading can occur within germ or somatic cells. Unlike somatic variation, germline variation can be transmitted to the progeny. Familial diseases, which reappear in multiple generations in pedigrees, are often linked to germline variation. An example of a familial disease is Marfan syndrome (OMIM 154700), an autosomal dominant disorder characterized by defective connective tissue leading to increased height, disproportionate limbs, and multiple cardiovascular anomalies; and often caused by heterozygous variation in the *FBNI* gene (Dietz et al., 1991). Conversely, pathogenic somatic variation can give rise to sporadic phenotypes like non-familial cancer (Greenman et al., 2007).

1.1.3 Structural variation

1.1.3.1 Chromosomal abnormalities: numerical and structural chromosomal aberrations

Chromosomal abnormalities can be either numerical or structural aberrations, which can be easily visualized using cytogenetic techniques. Numerical chromosomal aberrations also termed chromosomal aneuploidy, are any cases when the karyotype is not 46, XY or 46, XX for males and females, respectively. Therefore, they are caused by the absence (monosomy) or addition (trisomy, triploidy) of chromosomes in cells. An example of a numerical chromosomal aberration is Down syndrome (Trisomy 21, OMIM 190685), a neurodevelopmental disorder characterized by distinct facial features and intellectual disability in addition to multiple systemic anomalies, due to three copies of chromosome 21 (or chromosome 21q) (Wiseman et al., 2009). Chromosomal abnormalities are usually the result of nondisjunction at the first or second meiotic division (Ioannou and Tempest, 2015).

Individuals with structural chromosomal aberrations may still have a normal karyotype however, sizable portions of chromosomes have been deleted, duplicated, or inverted. Additionally, portions of DNA may have merged and translocated with other chromosomes (Weckselblatt and Rudd, 2015). An example of a disease caused by structural chromosomal abnormalities is DiGeorge syndrome (22q11.2 deletion syndrome, OMIM 188400), a characteristic dysmorphic facial syndrome with congenital heart defects (Michaelovsky et al., 2012).

1.1.3.2 Copy number variation

Copy number variation (CNV) were originally classified as structural DNA segments that are >10 Kb in the genome and are variations that are more frequently observed than chromosomal aberrations (Redon et al., 2006; Sebat et al., 2004). Importantly, CNV can differ in size: <1 Kb - <1Mb (Park et al., 2010). CNVs can be benign variation among unrelated individuals or can contribute to disease based on their proximity to genetic elements. Their imbalance can alter gene dosage leading to gene amplification or deletion

if encompassing the entire gene or localized within regulatory elements (Stranger et al., 2007). CNV can also influence select regions within genes such as exons, and can sometimes encode an abnormal protein. Among a multitude of examples, CNVs within chromosome 16p13.1 have been associated with schizophrenia (Ingason et al., 2011), autism spectrum disorders, and intellectual disability (Ullmann et al., 2007).

1.1.3.3 Variable number of tandem repeats

One third of the human genome is comprised of repetitive DNA sequences, which are in the form of large segmental duplications (low CNVs) or long and short interspersed transposon derived elements (LINEs and SINEs) (Kozlowski et al., 2010). In addition, variable number of tandem repeats (VNTR) are a class of genetic variation that are due to multiple nucleotides repeating consecutively within genes. There are two main types: minisatellites and microsatellites, which typically occur outside of coding regions, in non-coding sequences that often flank genes. Minisatellites are sometimes referred to as VNTR and range in size from 10-60 nucleotides and have been observed in thousands of locations in the human genome (Kozlowski et al., 2010). Similarly, microsatellites, often referred to as short tandem repeats (STR), are typically composed of 2-10 nucleotides and are hypermutable sites across the genome (Kozlowski et al., 2010). Interestingly, these common repeat regions served as markers for genome-wide linkage studies prior to the emergence of the International HapMap Project (International HapMap, 2005).

Importantly, when these repeats are located in regulatory or coding regions and exceed the normal, stable range, they can result in aberrant gene products, eventually causing disease (Lopez Castel et al., 2010). Examples of diseases due to multinucleotide expansions include Fragile X syndrome (OMIM 300624), an intellectual disability caused by CGG repeats within the 5'-untranslated region of the *FMRI* gene (Bagni et al., 2012). In addition, Huntington's disease (OMIM 143100) is a neurodegenerative disease caused by CAG repeats within the coding region of the *HTT* gene producing a polyglutamine tract (Mahalingam and Levy, 2014).

1.1.3.4 Insertions and deletions

Insertions or deletions (indels) of nucleotides in the DNA sequence may have an impact on gene expression or protein function. Indels can range from 1-50 base pairs with the majority consisting of <10 base pair changes. Depending on their location, indels can cause inframe or frameshift variants, or splicing variants. For example, the most common variant in cystic fibrosis (OMIM 219700), an autosomal recessive respiratory disease characterized by defective mucociliary clearance, is a three-nucleotide deletion in the *CFTR* gene, causing the loss of phenylalanine at amino acid position 508, within the ABC transporter domain (Flume and Van Devanter, 2012). While >90% of intron-containing genes undergo alternative splicing to regulate gene expression and protein synthesis, indels within or near the exon-intron junctions alternating the 5', 3' splice sites or the branch point, can cause disease (Singh and Cooper, 2012). These DNA changes can potentially prevent the spliceosomes, the splicing machinery, from recognizing these elements ultimately inducing inadvertent exon skipping or intron retention (Singh and Cooper, 2012)

1.1.4 Single nucleotide variation

Single nucleotide variation (SNV), which is also referred to as point mutation or substitution, is the most common class of variation. They can be either transitions where the SNV is an interchange between purines (adenine, C <-> guanine, G) or pyrimidines (cytosine, C <-> thymine, T); or transversions, which interchanges purine and pyrimidines (A<->C, A<->T, G<->C, or G<->T). If an SNV is within the codon frame, it can result in synonymous or non-synonymous changes. The degeneracy of the genetic code allows for multiple unique codons to generate the same amino acid. Therefore, SNVs within codons, especially in the third nucleotide position may not alter the amino acid as proposed in 'the wobble hypothesis' (Crick, 1966). These synonymous changes are often referred to as silent or neutral variants as they typically do not modify the phenotype. Conversely, SNVs can exert non-synonymous changes in which the codon codes for an alternate amino acid or results in a stop codon. These classes of variation are

termed missense or nonsense variants, respectively. SNVs can also cause splicing defects if located within or near the exon/intron boundaries and can alter the splice donor or acceptor recognition sequence, rendering them unrecognizable by splicing machinery (Cartegni et al., 2002). An example among many SNV-mediated human diseases is Duchenne muscular dystrophy (OMIM 310200), an X-linked muscle disorder, most often caused by nonsense variation within the *DMD* gene leading to protein truncation (Hoffman et al., 1987).

1.2 Genetic diseases in humans

1.2.1 Monogenic (Mendelian) diseases

Mendelian traits, first characterized by Austrian monk Gregor Mendel in the mid-1800s, explain simple and apparent parent to child transmission patterns (Henig, 2000; van der Waerden, 1968). Mendel first studied these traits in plants by observing height and colour differences in subsequent generations (Henig, 2000; van der Waerden, 1968). In doing so, the laws of Mendel emerged, which attempt to explain these noticeable differences. The first law, known as ‘the law of independent assortment’, states that during gamete formation, different pairs of alleles segregate independently of each other (Castle, 1903; Monaghan and Corcos, 1984). The second law, ‘the law of segregation’, suggests that the two alleles for each trait separate (segregate) during gamete formation, and then unite at random (Castle, 1903; Monaghan and Corcos, 1984). The third law, ‘the law of dominance’, states that recessive alleles are always masked by dominant alleles that determine the phenotype (Castle, 1903; Monaghan and Corcos, 1984).

Although additional complexities and exceptions have been recognized since they were originally proposed, these principles are still in use and are applied today in research and more importantly, in clinical genetic testing (Farhan and Hegele, 2013). Mendelian diseases are considered simple genetic diseases as they are typically caused by variants in a single gene (monogenic). Mendelian diseases are also recognized by their distinct inheritance patterns including: autosomal dominant, co-dominant, or recessive and X-linked (sex-linked) dominant or recessive (Farhan and Hegele, 2013). In dominant

disorders, one variant is sufficient to cause disease; this is in contrast to recessive disorders, where two defective alleles are necessary.

Mendelian diseases are typically not affected by exogenous factors such as the patient's environment, gene by environment interactions, or epigenetic modifications however; these factors can exacerbate or alleviate the phenotype. For example, individuals diagnosed with phenylketonuria (PKU, OMIM 261600), an autosomal recessive, metabolic disease commonly caused by variations in the *PAH* gene, can exacerbate or alleviate the intellectual disability phenotype caused by an accumulation of tyrosine, if an affected individual maintains a high or low protein diet, respectively (Demirkol et al., 2011). Mendelian diseases are mostly rare in the general population as they are the result of a highly penetrant variant in a highly conserved and physiologically essential protein (Farhan and Hegele, 2013). Individuals affected by severe Mendelian diseases such as Tay-Sachs disease (OMIM 272800), a severe early-onset neurological disorder caused by variants in the *HEXA* gene, do not reach the age of sexual maturity; and if they do, are unable to reproduce due to many biological and social factors (Fernandes Filho and Shapiro, 2004).

1.2.2 Digenic diseases

Digenic inheritance, considered the simplest form of oligogenic inheritance, is characterized by the presence of two independent variants that jointly modify the phenotype. Digenic inheritance is a relatively unrecognized type of disease inheritance, as its definition has not been clearly defined and often mistaken with epistasis, the interaction of genes. In 2013, Alejandro Schäffer suggested a universal definition of digenic inheritance in medical genetics: 'inheritance is digenic when the variant genotypes at two loci explain the phenotypes of some patients and their unaffected (or more mildly affected) relatives more clearly than the genotypes at one locus alone' (Schaffer, 2013). In recognition of the importance of digenic inheritance in human diseases, the 'digenic disease database' was created with the intent of aggregating comprehensive information on reports of digenic inheritance in various diseases (Gazzo et al., 2016).

An example of a disease with digenic inheritance is long QT syndrome (OMIM 192500), a congenital heart disease characterized by a prolonged QT interval as detected by electrocardiography (Millat et al., 2006). While variants in either *KCNQ1* or *KCNH2* can cause long QT syndrome, the condition can also result from co-occurrence of variants in both genes (Berthet et al., 1999). Furthermore, the digenic inheritance observed in long QT syndrome, has been replicated in other studies (Millat et al., 2006; Tester et al., 2005). Importantly, digenic traits can also be considered complex diseases as they can be modulated by non-genetic factors, as described in the subsequent section.

1.2.3 Polygenic (complex) diseases

Complex diseases are affected by various factors including a burden of genetic variation, the environment, gene-environment interactions, or epigenetic modifications (Cordell, 2009; Iyengar and Elston, 2007; Petrovski and Kwan, 2013). Complex diseases are more prevalent in the population therefore, they often follow a ‘common variant, common disease’ genetic principle. Traditionally, common variants are primarily studied in complex diseases however, they only marginally contribute to disease susceptibility. Potentially, the presence of multiple genetic variants with modest but cumulatively, significant effects can confer risk to disease (Do et al., 2012; Petrovski and Kwan, 2013). Despite these biological consequences, low effect variants are able to persist in the population, as they are individually not large enough to be targeted by natural selection.

Based on these observations, conventional genome-wide association studies (GWAS) used single nucleotide polymorphism (SNP) arrays, which did not capture any rare variation and were based on *a priori* knowledge of common susceptibility SNPs (Koboldt et al., 2013). However, there has been growing interest in assessing the effects of rare variants in complex diseases to identify the missing heritability (discussed in subsequent sections). The availability of next generation sequencing technologies (discussed in upcoming sections) allows for unbiased detection of both common and novel rare variation. Likewise, exome-wide genotyping arrays are commercially available at significantly lower costs and permit large scale rare and common variant association testing (Koboldt et al., 2013).

As initially observed by Francis Galton and Karl Pearson in the 1900s, quantitative traits typically follow a Gaussian (normal) distribution (Galton and Galton, 1997). Individuals residing at the tail ends of the distribution are often studied to identify any genetic variation contributing to the trait of interest. This approach, referred to as extreme phenotype sampling, is robust in detecting rare variation enriched in cases (extreme phenotypes), than controls (individuals within the normal range) (Auer and Lettre, 2015). Moreover, in the absence of large effect variants, which can underlie Mendelian diseases as previously discussed, individuals with extreme phenotypes may carry a burden of multiple small effect variants that collectively increase risk for disease. The burden of genetic risk is often quantified by summation of risk alleles to produce a ‘polygenic risk score’ (Dudbridge, 2013). Finally, in addition to genetic predisposition, age, sex, puberty, pregnancy, or diet can influence disease manifestation.

1.3 Phenotypic variation

As discussed in the preceding section, complex disorders are influenced by multiple factors and therefore, can have multiple etiologies. The phenotypic variation in an individual with a complex trait can be attributed to genetic variation, environmental conditions, or an interplay of both potential influences (Zuk et al., 2012). Simply, the phenotypic value (P) is equal to the sum of the genotypic value (G) and the environmental deviation (E). Therefore, changes in G and/or E influence the variability in P:

$$P = G + E$$

Importantly, G is comprised of three components: 1) the additive genetic value (A); 2) the dominance component (D); and 3) the epistasis (I, interaction of genes).

$$G = A + D + I$$

These genetic influences make up the heritability of a quantitative trait, which I will discuss next.

1.3.1 Heritability

To determine the effect of genetic variation on phenotypic variation (V_P), we often rely on the heritability of the phenotype of interest. Heritability is the variation in an individual's phenotype that is attributed to the genetic variation transmitted through the parents' germ cells (Zuk et al., 2012). Specifically, heritability is the ratio of genetic variance (V_G) to phenotypic variance (V_P). There are two main types of heritability: broad sense and narrow sense. Broad sense heritability (H^2) is all the genetic variance:

$$H^2 = V_G/V_P = (V_A + V_D + V_I)/V_P.$$

Alternatively, narrow sense heritability (h^2) is only the additive genetic variance:

$$h^2 = V_A/V_P$$

1.4 Epigenetics and human diseases

In addition to genetic variations that contribute to benign human diversity or disease, the expression of genes can be modulated by epigenetic modification machinery. In general, epigenetics is the study of mechanisms that control gene expression without altering the DNA sequence. These potentially heritable processes are stable and are particularly important during development. There are multiple types of epigenetic modifications grouped into three main categories: 1) DNA methylation; 2) histone modification; 3) and nucleosome positioning (Handy et al., 2011; Portela and Esteller, 2010).

DNA methylation is the process of adding a methyl group to a CpG site, yielding a transcriptionally inactive gene. Therefore, methylation and demethylation respectively silence or activate gene expression ultimately affecting protein synthesis (Handy et al., 2011). Similarly, histone modifications can modulate gene expression and protein synthesis. Specifically, the tail ends of histones can undergo post-translational modifications including acetylation, methylation, phosphorylation, ubiquitination, SUMOylation, or ADPribosylation (Portela and Esteller, 2010). Finally, positioning of nucleosomes can regulate gene expression by blocking transcription start sites, ultimately preventing transcription (Portela and Esteller, 2010).

An example of a disease with epigenetic modification is ATR-X syndrome (OMIM 301040), an X-linked dominant disease characterized by intellectual disability and alpha-thalassemia (Badens et al., 2006). ATR-X syndrome is caused by heterozygous (hemizygous) variants in the *ATRX* gene and males are predominantly affected however, carrier females may have skewed X-inactivation (Badens et al., 2006). ATRX binds DAXX to form a chromatin-remodeling complex, and pathogenic variants in either protein can lead to telomere dysfunction, genomic instability, or altered gene expression (Leung et al., 2013; Watson et al., 2013).

1.5 Approaches to studying genetic diseases

The relationship between disease penetrance and allele frequencies can help in prioritizing which approach to pursue when studying genetic diseases. As previously discussed, highly penetrant phenotypes are more likely to be caused by very rare genetic variation. While low-to-intermediate penetrant phenotypes can also be attributed to very rare genetic variation, these phenotypes are typically not severe enough to be clinically ascertained (McCarthy et al., 2008). In contrast, common variants typically result in low penetrant phenotypes, and are often the targets of GWAS (McCarthy et al., 2008).

In this section, I will introduce conventional approaches to study genetic diseases in related individuals (family studies applying forms of linkage analysis) and in unrelated individuals (population studies applying forms of association analysis). While linkage studies, candidate gene approaches, and GWAS have been successful in identifying loci that are causal or confer risk to disease, massively parallel sequencing is able to rapidly and economically identify causal variants, which I will discuss in subsequent sections.

1.5.1 Family studies

1.5.1.1 Linkage studies

Genetic linkage, first postulated by three British geneticists namely, William Bateson, Edith Rebecca Saunders, and Reginald Punnett, is when two loci are transmitted together

from parent to offspring, more often than expected (Dawn Teare and Barrett, 2005; Edwards, 2012). Across a population, when alleles at separate loci are associated with each other at a significantly higher frequency than would be expected by chance, they are in *linkage disequilibrium* and therefore, can be evidence of common ancestry (Dawn Teare and Barrett, 2005). Using linkage maps, which were initially developed by Thomas Hunt Morgan and his student, Alfred Sturtevant, helped us further refine linkage studies. Morgan and Sturtevant hypothesized that recombination of genes was dependent on the physical distance between them, measured in centimorgans (cM; 1 cM is approximately 1 million base pairs) (Sturtevant et al., 1919). This proposition allowed for linkage studies to be applied in humans to identify regions of the genome containing disease-associated genes. Furthermore, a specific subtype of linkage analysis known as the parametric of inheritance model-based analysis, determines co-segregation of genetic markers in families (Dawn Teare and Barrett, 2005). As the proximity of loci on a chromosome reduces the probability that they will separate by recombination, it is likely that a closely related disease cohort would co-inherit an entire section within the chromosome. As aforementioned, this phenomenon known as ‘linkage disequilibrium’ defies Mendel’s law of segregation. Using this approach, a genetic region unique to affected individuals and absent from healthy controls, can be identified. This approach also assigns a logarithm of the odds (LOD) score for each linkage event with significantly large LOD scores indicative of linkage; and low or negative values are evidence of genetic recombination. LOD scores of ≥ 3 are equivalent to P -values of ≤ 0.001 , and are traditionally considered statistically significant.

1.5.1.2 Candidate gene approaches

Prior knowledge of genes implicated in a disease has a central advantage over traditional linkage studies and GWAS. Candidate gene approaches are hypothesis driven and are based on a well-established relationship of a gene with a disease, prior knowledge from genetic analysis (genetic variant present in affected individuals); or a high likelihood region identified via linkage analysis (Jorgensen et al., 2009). Targeting candidate genes by sequencing or genotyping individuals for variation in genes implicated in the disease,

can significantly expedite genetic discoveries. However, candidate gene approaches are generally not suited for novel diseases or diseases exhibiting locus heterogeneity.

1.5.2 Population studies

1.5.2.1 Association studies

Association approaches are more useful for elucidating susceptibility alleles underlying complex diseases in multiple unrelated individuals. This unbiased approach facilitated the era of GWAS, which used genotyping arrays to screen a large number of cases and controls, for thousands of common variants across the genome (Auer and Lettre, 2015). To identify genuine disease-associated variants, the variant frequency in cases must be significantly higher than in controls, which at first glance, likely favours linkage disequilibrium (Auer and Lettre, 2015). The International HapMap Project assessed common variation in a global, ethnically diverse cohort, and determined that when considering linkage disequilibrium, there are approximately one million unique loci in the human genome (International HapMap, 2005). Following a Bonferonni correction for the million unique loci, which are treated as independent tests, a genome-wide significant P -value = $<5 \times 10^{-8}$, became the standard threshold. In addition to observing genome-wide significance of an association between an allele and a phenotype, it is important to replicate the association in an unbiased population sample to eliminate possible confounding variables such as population stratification, which may occur when subpopulations within the study cohort carry the allele due to differences in ancestral origin, rather than disease state (McCarthy et al., 2008). Furthermore, association of the variant and disease should be followed by ensuring the signal is replicable and harboured in or near a gene biologically relevant to the disease mechanism.

1.6 Tools to study genetic variation in disease

In this section, I will introduce tools available to study genetic variation with an emphasis on sequencing technology.

1.6.1 Sequencing versus genotyping: benefits and pitfalls

To identify the genetic basis of human diseases, we often use sequencing or genotyping-based assays. In sequencing approaches, we indiscriminately identify the full nucleotide sequence of genes. In genotyping approaches, we use a known set of genetic markers to determine their presence or absence in individuals. Each approach is selected based on the study hypothesis and objectives. Traditionally, GWAS used arrays to rapidly genotype cases and controls, as previously discussed. Knowledge of the entire genetic sequence is preferred however, genotyping approaches are often selected to reduce costs. In my PhD thesis, I used sequencing approaches to discover novel disease loci and genotyping approaches to screen for the newly discovered loci in cases and controls.

1.6.2 Massively parallel sequencing

Today, we are in the era of massively parallel DNA sequencing often referred to as next generation sequencing. Next generation sequencing technologies is a broad umbrella term for technologies that facilitate rapid, efficient sequencing of multiple genomes by performing millions of reactions simultaneously leading to high throughput of data (Farhan and Hegele, 2014). This is in contrast to Sanger sequencing, now considered first generation sequencing, which has been widely and successfully used in genetic studies most notably in elucidating the code of the first human genome (Lander et al., 2001). However, Sanger sequencing is labour intensive and can be prohibitively expensive for studies with a large sample size. The advent of next generation sequencing technologies has accelerated gene discoveries in virtually all diseases. There are three main types of next generation sequencing applications for sequencing DNA variants: 1) whole genome sequencing; 2) exome sequencing; and 3) targeted gene panels.

Whole genome sequencing is an indiscriminate approach that decodes the genetic information in an individual's genome. In contrast, exome sequencing targets only the protein-coding regions or the "exome" of the genome by designing probes that are unique

to the exons. Targeting only the protein-coding regions of the genome originates from the observation that a large fraction of human genetic diseases are caused by non-synonymous variants in evolutionarily considered protein-coding genes (Chong et al., 2015; Ng et al., 2009). Moreover, the difference in cost between the two methods; and the computational power necessary to reassemble the human exome (1-2% of the human genome), is significantly less resource intensive (Farhan and Hegele, 2014).

Notably, while next generation sequencing has led to the discovery of genes not previously implicated in human diseases, in general, the majority of findings are novel variants in known disease-causing genes. These trends have consequently led to the development of custom designed next generation sequencing gene panels where disease-specific genes are preselected and are screened without sequencing other regions of the genome. This allows for a prioritized, economical, and rapid genetic diagnostic approach without the burden of incidental and secondary findings in known disease-causing genes not relevant to the disease of interest. I describe the approach of developing a targeted neurodegeneration gene panel in Chapter 6.

1.6.3 Molecular studies and model organisms

The use of molecular approaches and model systems allows us to recapitulate genetic variation found in human diseases and provides a clearer understanding of the biological consequence exerted by the variants (Nabbout and Dulac, 2011). There are multiple molecular biology approaches to investigate the effect of variants beginning primarily with the patients' tissue (*ex vivo* experiments). Gene expression and protein interaction assays are examples of experimental approaches to investigate the pathogenicity of candidate genetic variants within patient derived cells. However, *ex vivo* experiments are not always possible especially when studying neurological diseases as many neurological disease-causing genes are predominantly expressed in the central nervous system and require invasive extraction procedures (Ottman et al., 2010; Prasad and Hoffmann, 2010). To overcome this limitation, *in vitro* assays can be performed to assess whether the gene product or its equivalent, is defective and disrupts any downstream pathways. In *in vitro* assays, the variant is genetically engineered into DNA and introduced into an artificial

system using established cell lines. Using this approach, we can determine whether the variant exerts a loss or a gain of protein function and the resulting effect on the connecting pathways.

Finally, *in vivo* models, where the candidate gene is disrupted or the variant is genetically engineered and implanted into embryos that develop into model organisms, have led to the advancement of human genetics research. Specifically, various non-mammalian organisms including *Amoeba proteus* (amoeba), *Caenorhabditis elegans* (worms), *Drosophila melanogaster* (fruit flies), and *Danio rerio* (zebrafish) have been routinely used as model organisms in research given their relatively simple development and short lifespan. The absence of these characteristics in transgenic non-human animal models makes them significantly more challenging to study (Cunliffe et al., 2014). However, despite the long generation times of non-human animal model organisms such as *Mus musculus* (mice), *Sus scrofa* (pigs), or *Pan troglodytes* (chimpanzee), they are more related genetically and physiologically, to humans and therefore, these animals are more likely to generate a phenotype consistent with the human disease (Kullmann et al., 2014). The application of both types of model organisms has greatly advanced our knowledge of human diseases and has ushered the successful development of drugs used to treat the patients.

1.7 Variant causality criteria

Naturally, a single human genome contains numerous rare and common variations. Therefore, prioritization criteria are needed to distinguish a potentially causative variant from merely common benign genetic variation. Being able to ascribe causality to a particular DNA variant with a high degree of likelihood is essential for genetically diagnosing patients with inherited diseases. In this PhD thesis, I have applied the following conservative criteria, which includes multiple lines of evidence, to increase the likelihood that the genetic variants observed in patients in Chapters 2-5, are disease causing. Importantly, given the complexity in the diseases described in Chapters 6 and 7, I did not apply all steps of the variant causality criteria.

(1) The genetic variant is rare in the population

Allele frequency is an important metric of variant deleteriousness. Intuitively, rare diseases are more likely to be caused by rare variants, as they are relatively depleted in the population, as previously discussed. Upon obtaining next generation sequencing data, I prioritized rare variants, which often have a minor allele frequency (MAF) $\ll 1\%$ or may not have a MAF, in ascertained samples in population databases. To determine the allele frequency, I surveyed multiple population databases containing thousands of healthy controls (Table 1.6.1.1) as well as disease specific databases (Table 1.6.1.2). It is important to mention that some of the databases listed here were not initially available during the duration of my PhD studies. Furthermore, in addition to incorporating thousands of individuals, it is important to ensure the reference databases used are ethnically diverse and accurately catalogue human diversity. Population ascertainment bias can occur and can lead to misinterpretation of the allele frequency. For example, European populations are enriched in currently available reference sources. Therefore, the absence of a variant in European populations should be interpreted with caution, as the same variant can be more frequent in African or Middle Eastern populations, which are currently not adequately represented.

Table 1.7.1 Population databases of healthy controls.

Databases	Sample #	Type of sequence	Reference
ExAC	60,706	Exome	Lek et al., 2015
NHLBI ESP	6,500	Exome	Fu et al., 2013
1000 Genomes	2,500 (originally 1092)	Genome	Abecasis et al., 2012
dbSNP		Multiple	Sherry et al., 2001

ExAC, Exome Aggregation Database; NHLBI ESP, National Heart, Lung, and Blood Institute Exome Sequencing Project; dbSNP, database of single nucleotide polymorphisms. Empty cells depict unknown information.

Table 1.7.2 Disease databases.

Databases	Type of disease	Type of sequences	Reference
ClinVar	Mendelian and complex	Multiple	Landrum et al., 2014
OMIM	Mendelian	Multiple	Hamosh et al., 2002
HGMD	Mendelian and complex	Multiple	Cooper et al., 1998
DECIPHER	Mendelian and complex	Multiple	Bragin et al., 2013

OMIM, Online Mendelian Inheritance in Man; HGMD, Human Gene Mutation Database; DECIPHER, Database of genomic variation and phenotype in humans using Ensembl resources.

(2) The genetic variant is predicted to affect the protein function using *in silico* predictive programs

Conservation is another important metric of variant deleteriousness. The conservation of the wild type allele throughout evolution illustrates its importance in gene function. Accordingly, I used multiple *in silico* predictive software (Table 1.6.1.3) to assess the conservation of the wild type residues and also the potential biological effect of the amino acid change. Intuitively, a drastic change in amino acid properties is more likely to affect protein function potentially by disrupting folding, ultimately altering or abolishing ligand-binding sites. Ultimately, the results of these programs are merely prediction and were always interpreted with caution, as their algorithm basis is the same across all genes, independent of gene function or expression.

Table 1.7.3 Examples of *in silico* predictive tools.

Tool	Category	Algorithm basis	Reference
PolyPhen-2	Missense	Protein structure and conservation	Adzhubei et al., 2013
SIFT	Missense	Conservation	Ng et al., 2003
MutationTaster	Missense	Protein structure and conservation	Schwarz et al., 2010
CADD	Missense	Simulated variants and fixed alleles	Kircher et al., 2014
Condel	Missense	Implements PolyPhen-2, SIFT, and MutationAssessor	González-Pérez et al., 2011
MutationAssessor	Missense	Conservation	Reva et al., 2011
PANTHER	Missense	Conservation	Mi et al., 2016
Asseda	Splicing	Combined gene information content	Mucaki et al., 2013
Human Splicing Finder	Splicing	Proximity to exon/intro	Desmet et al., 2009
MaxEntScan	Splicing	Entropy	Yeo et al., 2004
GERP	Conservation	Genomic conservation rates	Cooper et al., 2005
phyloP	Conservation	Alignment and phylogenetic trees	Pollard et al., 2010
ClustalW*	Conservation	Alignment and phylogenetic trees	Larkin et al., 2007

*Not predictive software, through alignment, researchers can observe the conservation of wild type residues.

(3) The genetic variant segregates with disease status

Association of the genetic variant with disease status is important in establishing causality. Herein, I prioritized variants that were significantly overrepresented in cases than controls. I screened all available family members to determine whether the variants observed segregate with the disease status in the pedigrees, in the predetermined mode of inheritance.

(4) Functional validation of the genetic variant

Appropriate assays should be used to experimentally validate the pathogenicity of the candidate variants. These experiments can demonstrate whether the variant exerts a gain or loss of protein function and how this is related to the biology of disease. Experiments using patient-derived tissue (*ex vivo*), genetically engineered cell lines expressing the mutants (*in vitro*), or model organisms (*in vivo*) that clearly recapitulate the human disorder, are useful in assessing variant pathogenicity. This is often the rate-limiting step in implicating DNA variation in human disease. In Chapters 2, 3, and 5, I used multiple experimental approaches to objectively evaluate the effect of the candidate variants.

1.8 Thesis outline

In this section, I will introduce the projects and opportunities to study monogenic diseases and complex diseases (familial and sporadic). More details are provided in Chapters 2-7.

1.8.1 Finding of Rare Disease Genes (FORGE) Canada

1.8.1.1 Project rationale

Rare genetic diseases affect the lives of approximately 500,000 children in Canada (Mackenzie and Boycott, 2012). While individually rare, collectively they account for a significant proportion of hospitalizations (Farhan and Hegele, 2013). Their disproportionate prevalence in certain communities is often the result of years of isolation, which eventually leads to low genetic variation. In these cases, we can infer the

disease-causing variant is harbored within a highly homozygous region in affected individuals and is transmitted through asymptomatic carriers (Broman and Weber, 1999; Lander and Botstein, 1987; Puffenberger et al., 2012). Previously, the use of standard genetic techniques such as homozygosity genetic mapping and candidate gene sequencing have been successful although costly and laborious, in identifying the genetic basis of monogenic disorders (Lahiry et al., 2009; Puffenberger et al., 2004). The application of next generation sequencing has identified hundreds of disease-causing genes and has greatly advanced the study of rare inherited disorders.

1.8.1.2 Project overview

The FORGE Canada Consortium was a nationwide initiative to ascertain and study the genetic basis of multiple inherited pediatric onset rare genetic disorders (Beaulieu et al., 2014). The Children's Hospital of Eastern Ontario Research Institute at the University of Ottawa was established as the lead institution. Gene discovery was an overall objective of the consortium. Accordingly, there were five categories in which to group the multiple disorders submitted by Canadian clinicians. These include: 1) multiple unrelated individuals affected by the same rare disorder; 2a) consanguineous families; 2b) autosomal dominant families; 3) non-consanguineous families with two or more affected individuals; and 4) single affected individuals with no family history. Overall, 264 projects were funded to study their genetic basis. The 264 disorders were distinct clinical phenotypes affecting multiple systems and observed in either single or multiple pedigrees.

1.8.1.3 Project aims

The overall aims of FORGE, which are recapitulated here, are to:

- 1) Assist clinicians and researchers in recruiting patients with rare genetic diseases.
- 2) Use next generation sequencing technologies and strategic functional tests to identify the genetic basis of disease in patients.

- 3) Organize and facilitate a national data coordination centre to accelerate and improve the interpretation of large-scale sequence data.
- 4) Establish ethical guidelines for analyzing and interpreting sequence data for the purposes of genetic counseling.

1.8.1.4 Project significance

Identifying disease genes is important for several reasons. First, it can help us understand the dynamic function and mechanism of the gene product in the normal state. Second, we can implement presymptomatic testing and carrier screening for high-risk populations. Third, depending on the biochemical pathway, the gene may be druggable. While these disorders are each clinically distinct, the process used to identify the causative gene is the same in each case. Publishing causative genes and variations for which the confidence in their causality is high will allow research groups locally and globally to move the discovery process forward using different lines of experimentation in which they are expert. These efforts can eventually lead to targeted therapeutic interventions.

1.8.1.5 My role in the project

Members of the Hegele lab worked with clinicians and scientists who have ascertained multiple, distinct rare disorders that became approved by the FORGE Canada Consortium. Four disorders: 1) infantile mitochondrial complex II/III deficiency (IMC23D); 2) seizures scoliosis macrocephaly (SSM) syndrome; 3) progressive myoclonus epilepsy with ataxia; and 4) nocturnal seizures with developmental delay became independent research chapters within my PhD thesis. All four phenotypes are described in great detail in subsequent chapters.

1.8.2 The Ontario Neurodegenerative Disease Research Initiative (ONDRI)

1.8.2.1 Project rationale

Dementia is the single greatest cause of neurological disability in our senior population with a global incidence of 47.5 million cases (2008; 2015). The annual cost of caring for individuals affected with dementia in Canada is \$15 billion – a figure expected to grow to nearly \$153 billion by 2038 - including a tenfold increase in demand for long-term care (Herrmann et al., 2015). The cumulative global cost of Alzheimer's disease and dementia is currently estimated to be \$605 billion, which is equivalent to 1% of the entire world's gross domestic product. This figure will continue to grow unless disease-modifying interventions are found to alter these alarming trends (2008; Herrmann et al., 2015). Key to any such initiative will be the ability to identify early and accurately those individuals who are at risk for developing a dementia, either as an independent disease process or as a co-morbidity of a related neurodegenerative or neurovascular disorder.

1.8.2.2 Project overview

ONDRI is a prospective cohort study that will use a multimodal approach to predicting the occurrence or progression of cognitive or neuropsychological impairment in a defined patient population. Disease progression will be monitored in 600 patients from across Ontario, Canada, for up to three years using rigorous evaluations across multiple assessment platforms, including neuroimaging, detailed neuropsychological evaluations (including comprehensive speech and language assessments), genomics, evaluations of cognitive control of eye movements and retinal layer thickness and morphology, gait performance, and neuropathology.

Patients will be enrolled in the study with either 1) Alzheimer's disease (AD) or amnesic single or multidomain mild cognitive impairment (aMCI), 2) amyotrophic lateral sclerosis (ALS), 3) frontotemporal dementia (FTD), 4) Parkinson's disease (PD) or 5) vascular cognitive impairment (VCI). These disorders were selected based on their

prevalence within the aging population and the frequent occurrence of neuropsychological dysfunction in each at various stages over their course. Amongst these, AD pathology is the most common, underlying approximately 63% of all dementia cases (Herrmann et al., 2015; Sperling et al., 2011). While objective memory loss characterizes aMCI, unlike AD, the cognitive impairment does not significantly disrupt daily functioning (Gomersall et al., 2015). Although traditionally considered to be a disorder of upper and lower motor neurons, over 50% of ALS patients will exhibit neuropsychological deficits that range from subtle syndromes of cognitive or behavioural dysfunction, including impairments in social cognition, to FTD (Strong et al., 2009). When present, frontotemporal dysfunction in ALS is associated with a reduction in survival by approximately a year (Woolley and Strong, 2015). Parkinson's disease affects approximately 1 in 1,000 in the general population and 1 in 100 individuals over age 65 years (Dorsey et al., 2007). The prevalence of PD is expected to double by the year 2030 (Dorsey et al., 2007). Similar to ALS, a significant proportion of PD patients present with or will develop neuropsychological dysfunction (Aarsland et al., 2005) and, over time, dementia (Parkinson's Disease Dementia; PDD) develops in over 80% (Hely et al., 2008). FTD, including subtypes of primary progressive aphasia, progressive non-fluent aphasia, semantic dementia and behavioural variant FTD (bvFTD), accounts for 20% of early-onset dementia cases with symptom onset commencing at 45-65 years of age (Kertesz, 2004). Finally, VCI is the second most common form of dementia with 30% of stroke patients developing dementia (Pendlebury et al., 2012; Saposnik et al., 2009; Saposnik et al., 2011). The risk of stroke and dementia rises exponentially with each decade over 65, with a third of our population expected to have a stroke, dementia, or both in their lifetime (Hachinski et al., 2006).

1.8.2.3 Project aims

The multimodal approach will rapidly detect meaningful neuropsychological change over a short time interval, allowing for the early and accurate prediction of the presence of, or progression of, dementia. Furthermore, we may identify different forms or profiles of dementia that will map onto specific neural circuits, some shared and some differing

across disorders depending on the hubs primarily affected. ONDRI will also examine the contribution of small vessel pathology to each of these disorders given the increasing recognition of the prevalence and potential synergistic effects of this important comorbidity.

1.8.2.4 Project significance

ONDRI is a unique, prospective multimodality study undertaken to improve understanding of the pathogenesis of neuropsychological deficits across a broad range of neurodegenerative disorders. Beyond this, ONDRI has already facilitated the development of standardized assessment protocols, which will allow for direct comparisons to be made across a range of neurodegenerative and neurovascular disorders. These tools will allow for a systems-based approach to identify not only unique clinical, imaging, ocular, genetic, and other biological markers associated with each of these disorders, but also to define where there are significant biological overlaps. While the numbers of patients being studied is small (600) relative to larger prospective studies, patients enrolled in ONDRI will be a unique resource as they will be deeply endophenotyped and their de-identified data publically available by request through the ONDRI publications and data analysis committee.

Furthermore, current therapeutic approaches in neurodegenerative diseases, where available, tend to be directed towards single biological mechanisms that may be inadequate given the complexity of these multifaceted diseases in addition to the spectrum of mechanisms by which neurodegeneration is expressed (Iadecola and Anrather, 2011). Through this integrated discovery approach, we have the unique opportunity to identify multiple markers of brain health that will contribute to the development of: 1) biomarkers for neurodegenerative diseases that may ultimately be used in the identification of pre-symptomatic individuals; 2) improved identification of overlap syndromes amongst neurodegenerative diseases for both clinical and research purposes; and, 3) personalized treatments, which may require a “cocktail” similar to the approach that has been applied in cancer, and are likely to be more efficacious in halting disease progression than current drug regimens (Lang, 2010). These may vary depending

on the stage of the disorder being treated, the genetic predispositions discovered to be contributing to an individual's disease, and the possible role of mixed pathologies. Hence, this longitudinal exploratory study has the potential to contribute to significant impacts on health care in the rapidly growing area of neurodegenerative and neurovascular diseases.

1.8.2.5 My role in the project

Within ONDRI, ONDRI Genomics focuses on identifying genetic variants that may contribute to disease onset and progression. The ONDRI Genomics team, which is led by Dr. Hegele, developed the ONDRISeq panel, which is a custom designed neurodegeneration gene panel specific to the ONDRI diseases. I played a major role in the development, optimization, and integration of ONDRISeq in the ONDRI study, which I describe in depth in Chapter 6.

1.8.3 Familial ALS and FTD study: defining the role of oligogenic inheritance in neurodegenerative disease

1.8.3.1 Project rationale

While the exact cause of ALS is unknown, there is substantial evidence that a significant proportion has a genetic basis, including ALS cases without a family history of the disease (sporadic ALS). Increasingly, some of these affected individuals are observed to have more than one genetic abnormality – a phenomenon known as oligogenic inheritance as previously discussed. While many of these genes are thought to be causative of ALS, many are considered to be disease modifiers. However, even within individual families carrying at least one genetic variant, the disease manifestations can vary dramatically from a relatively pure ALS phenotype with only motor neuron deficits, to dementia with minimal motor involvement, or phenotypically unaffected carriers. This raises the possibility that while a single gene may be dominant in inducing the disease, its effect will be profoundly modified by the presence of additional genetic variation.

1.8.3.2 Project overview

We have access to three previously unreported ALS families with widely differing disease manifestations, each of which carries a pathological expansion of *C9orf72*. These families provide a unique opportunity to understand the role of multiple gene inheritance in ALS, how the genes interact with each other, and how this interaction may lead to a degenerative disease state.

Although a significant number of genetic variations are known to be associated with ALS, either as causative or disease modifying, it is not clear that the wide phenotypic variability that has become a hallmark of the disease, can be fully explained by individual genes acting independently of one another. The most glaring example of this is the wide phenotypic variability observed with pathological hexanucleotide expansions of *C9orf72*, in which affected individuals may develop ALS, FTD, or some combination thereof. While oligogenic inheritance is held forth as an explanation, the biological mechanism(s) underlying such interactions has yet to be defined. Our hypothesis is that the clinical disease state accompanying oligogenic inheritance is the net effect of the interactions of multiple genes in a hierarchical manner such that while one gene defect will be dominant, its effect will be modified by the presence of additional genetic aberrations.

1.8.3.3 Project aims

Using these pedigrees, we will address the following specific aims:

1. Determine the phenotypic effect(s) of patient specific variations individually and describe the cumulative effects of variations. This can be established by correlating patient clinical data with the presence of variant(s) and post-mortem tissue pathology.
2. Examine the role of each of the proteins encoded by the mutated genes and the changes in binding partners between the mutant protein and the wild type protein in ALS affected spinal cord tissue.

3. As a long-term goal, determine the effect of these variations in an *in vivo* model system using *Drosophila melanogaster*, to assess neuron-specific effects.

1.8.3.4 Project significance

While the concept of oligogenic inheritance in ALS is not novel, these experiments represent a unique opportunity to determine the biological effect of oligogenic inheritance in a clinically relevant scenario and describe how this can lead to disease variability among related and unrelated individuals. Upon completion of all long-term goals, we will have investigated: 1) the effect of each variant in neuronal cells; 2) the hierarchy of neurotoxicity for each combination of genes when they are expressed simultaneously; and 3) how many of these are expressed simultaneously to produce a neurodegenerative phenotype.

1.8.3.5 My role in the project

I performed the experiments and analysis with assistance by other personnel, as outlined in Chapter 7. I also assisted Dr. Strong and Mrs. Anne Row in collecting blood samples from the families, in preparing the ethics documentation for the study, and overall project management.

1.9 Thesis hypotheses

The research aims presented in this PhD thesis seek to assess the following hypotheses:

1) The four Mendelian diseases, IMC23D, SSM syndrome, progressive myoclonus epilepsy with ataxia, and nocturnal seizures with developmental delay (Chapters 2-5) are caused by highly penetrant genetic variants inherited in an autosomal recessive manner.

2) Conversely, the neurodegenerative disorders (Chapters 7) are complex genetic diseases partly caused by single, rare, highly penetrant variants or multiple, low-to-modest effect variants acting synergistically to mediate disease expression.

1.10 Thesis aims

1) I propose to apply next generation sequencing (exome sequencing) to identify genetic variants that may be underlying the Mendelian diseases in Chapters 2-5.

2) I propose to design a neurodegeneration specific gene panel to identify genetic variants underlying familial and sporadic neurodegenerative disorders, ascertained through the ONDRI project and the familial ALS and FTD study.

3) I plan to apply conservative variant causality criteria on any candidate variants observed in the sequencing experiments.

1.11 Conclusion

In this PhD thesis, I studied the genetic basis of Mendelian and complex diseases. In the subsequent chapters, I describe the application of next generation sequencing in identifying highly penetrant variants in patients affected by rare Mendelian diseases ascertained through the FORGE Canada Consortium. In addition, I also describe the development of a custom designed gene panel specific to neurodegenerative diseases, which is currently used to sequence all patients within the ONDRI study as well as patients in the familial ALS and FTD study. The collective research experiments presented here have successfully identified novel genetic loci in multiple human diseases.

1.12 References

- (2015). WHO takes up the baton on dementia. *Lancet Neurol* 14, 455.
- Aarsland, D., Zaccai, J., and Brayne, C. (2005). A systematic review of prevalence studies of dementia in Parkinson's disease. *Mov Disord* 20, 1255-1263.
- Auer, P.L., and Lettre, G. (2015). Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 7, 16.
- Badens, C., Martini, N., Courrier, S., DesPortes, V., Touraine, R., Levy, N., and Edery, P. (2006). ATRX syndrome in a girl with a heterozygous mutation in the ATRX Zn finger domain and a totally skewed X-inactivation pattern. *Am J Med Genet A* 140, 2212-2215.
- Bagni, C., Tassone, F., Neri, G., and Hagerman, R. (2012). Fragile X syndrome: causes, diagnosis, mechanisms, and therapeutics. *J Clin Invest* 122, 4314-4322.
- Beaulieu, C.L., Majewski, J., Schwartzenuber, J., Samuels, M.E., Fernandez, B.A., Bernier, F.P., Brudno, M., Knoppers, B., Marcadier, J., Dymment, D., *et al.* (2014). FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet* 94, 809-817.
- Berthet, M., Denjoy, I., Donger, C., Demay, L., Hammoude, H., Klug, D., Schulze-Bahr, E., Richard, P., Funke, H., Schwartz, K., *et al.* (1999). C-terminal HERG mutations: the role of hypokalemia and a KCNQ1-associated mutation in cardiac event occurrence. *Circulation* 99, 1464-1470.
- Broman, K.W., and Weber, J.L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet* 65, 1493-1500.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3, 285-298.
- Castle, W.E. (1903). Mendel's Law of Heredity. *Science* 18, 396-406.
- Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., *et al.* (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* 97, 199-215.
- Cordell, H.J. (2009). Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 10, 392-404.
- Crick, F.H. (1966). Codon-anticodon pairing: the wobble hypothesis. *J Mol Biol* 19, 548-555.

- Cunliffe, V.T., Baines, R.A., Giachello, C.N., Lin, W.H., Morgan, A., Reuber, M., Russell, C., Walker, M.C., and Williams, R.S. (2014). Epilepsy research methods update: Understanding the causes of epileptic seizures and identifying new treatments using non-mammalian model organisms. *Seizure*.
- Dawn Teare, M., and Barrett, J.H. (2005). Genetic linkage studies. *Lancet* *366*, 1036-1044.
- Demirkol, M., Gizewska, M., Giovannini, M., and Walter, J. (2011). Follow up of phenylketonuria patients. *Mol Genet Metab* *104 Suppl*, S31-39.
- Dietz, H.C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M., *et al.* (1991). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* *352*, 337-339.
- Do, R., Kathiresan, S., and Abecasis, G.R. (2012). Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet* *21*, R1-9.
- Dorsey, E.R., Constantinescu, R., Thompson, J.P., Biglan, K.M., Holloway, R.G., Kieburtz, K., Marshall, F.J., Ravina, B.M., Schifitto, G., Siderowf, A., *et al.* (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology* *68*, 384-386.
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet* *9*, e1003348.
- Edwards, A.W. (2012). Punnett's square. *Stud Hist Philos Biol Biomed Sci* *43*, 219-224.
- Farhan, S.M., and Hegele, R.A. (2013). Genetics 101 for cardiologists: rare genetic variants and monogenic cardiovascular disease. *Can J Cardiol* *29*, 18-22.
- Farhan, S.M., and Hegele, R.A. (2014). Exome sequencing: new insights into lipoprotein disorders. *Curr Cardiol Rep* *16*, 507.
- Fay, J.C., Wyckoff, G.J., and Wu, C.I. (2001). Positive and negative selection on the human genome. *Genetics* *158*, 1227-1234.
- Fernandes Filho, J.A., and Shapiro, B.E. (2004). Tay-Sachs disease. *Arch Neurol* *61*, 1466-1468.
- Flume, P.A., and Van Devanter, D.R. (2012). State of progress in treating cystic fibrosis respiratory disease. *BMC Med* *10*, 88.
- Galton, D.J., and Galton, C.J. (1997). Francis Galton: his approach to polygenic disease. *J R Coll Physicians Lond* *31*, 570-573.

Gazzo, A.M., Daneels, D., Cilia, E., Bonduelle, M., Abramowicz, M., Van Dooren, S., Smits, G., and Lenaerts, T. (2016). DIDA: A curated and annotated digenic diseases database. *Nucleic Acids Res* 44, D900-907.

Gomersall, T., Astell, A., Nygard, L., Sixsmith, A., Mihailidis, A., and Hwang, A. (2015). Living With Ambiguity: A Metasynthesis of Qualitative Research on Mild Cognitive Impairment. *Gerontologist* 55, 892-912.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., *et al.* (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153-158.

Hachinski, V., Iadecola, C., Petersen, R.C., Breteler, M.M., Nyenhuis, D.L., Black, S.E., Powers, W.J., DeCarli, C., Merino, J.G., Kalaria, R.N., *et al.* (2006). National Institute of Neurological Disorders and Stroke-Canadian Stroke Network vascular cognitive impairment harmonization standards. *Stroke* 37, 2220-2241.

Handy, D.E., Castro, R., and Loscalzo, J. (2011). Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation* 123, 2145-2156.

Hely, M.A., Reid, W.G., Adena, M.A., Halliday, G.M., and Morris, J.G. (2008). The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord* 23, 837-844.

Henig, R.M. (2000). *The monk in the garden : the lost and found genius of Gregor Mendel, the father of genetics* (Boston: Houghton Mifflin).

Herrmann, N., Harimoto, T., Balshaw, R., and Lanctot, K.L. (2015). Risk Factors for Progression of Alzheimer Disease in a Canadian Population: The Canadian Outcomes Study in Dementia (COSID). *Can J Psychiatry* 60, 189-199.

Hoffman, E.P., Brown, R.H., Jr., and Kunkel, L.M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 51, 919-928.

Iadecola, C., and Anrather, J. (2011). Stroke research at a crossroad: asking the brain for directions. *Nat Neurosci* 14, 1363-1368.

Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietilainen, O.P., Buizer-Voskamp, J.E., Strengman, E., Francks, C., Muglia, P., *et al.* (2011). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry* 16, 17-25.

International HapMap, C. (2005). A haplotype map of the human genome. *Nature* 437, 1299-1320.

Ioannou, D., and Tempest, H.G. (2015). Meiotic Nondisjunction: Insights into the Origin and Significance of Aneuploidy in Human Spermatozoa. *Adv Exp Med Biol* 868, 1-21.

- Iyengar, S.K., and Elston, R.C. (2007). The genetic basis of complex traits: rare variants or "common gene, common disease"? *Methods Mol Biol* 376, 71-84.
- Jorgensen, T.J., Ruczinski, I., Kessing, B., Smith, M.W., Shugart, Y.Y., and Alberg, A.J. (2009). Hypothesis-driven candidate gene association studies: practical design and analytical considerations. *Am J Epidemiol* 170, 986-993.
- Kertesz, A. (2004). Frontotemporal dementia/Pick's disease. *Arch Neurol* 61, 969-971.
- Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., and Mardis, E.R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27-38.
- Kozlowski, P., Sobczak, K., and Krzyzosiak, W.J. (2010). Trinucleotide repeats: triggers for genomic disorders? *Genome Med* 2, 29.
- Ku, C.S., Vasiliou, V., and Cooper, D.N. (2012). A new era in the discovery of de novo mutations underlying human genetic disease. *Hum Genomics* 6, 27.
- Kullmann, D.M., Schorge, S., Walker, M.C., and Wykes, R.C. (2014). Gene therapy in epilepsy-is it time for clinical trials? *Nat Rev Neurol* 10, 300-304.
- Lahiry, P., Wang, J., Robinson, J.F., Turowec, J.P., Litchfield, D.W., Lanktree, M.B., Gloor, G.B., Puffenberger, E.G., Strauss, K.A., Martens, M.B., *et al.* (2009). A multiplex human syndrome implicates a key role for intestinal cell kinase in development of central nervous, skeletal, and endocrine systems. *Am J Hum Genet* 84, 134-147.
- Lander, E.S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567-1570.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Lang, A.E. (2010). Clinical trials of disease-modifying therapies for neurodegenerative diseases: the challenges and the future. *Nat Med* 16, 1223-1226.
- Leung, J.W., Ghosal, G., Wang, W., Shen, X., Wang, J., Li, L., and Chen, J. (2013). Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *J Biol Chem* 288, 6342-6350.
- Lopez Castel, A., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* 11, 165-170.
- Mackenzie, A., and Boycott, K.M. (2012). The future is now for rare genetic diseases. *CMAJ* 184, 1603.

- Mahalingam, S., and Levy, L.M. (2014). Genetics of Huntington disease. *AJNR Am J Neuroradiol* 35, 1070-1072.
- McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9, 356-369.
- Michaelovsky, E., Frisch, A., Carmel, M., Patya, M., Zarchi, O., Green, T., Basel-Vanagaite, L., Weizman, A., and Gothelf, D. (2012). Genotype-phenotype correlation in 22q11.2 deletion syndrome. *BMC Med Genet* 13, 122.
- Millat, G., Chevalier, P., Restier-Miron, L., Da Costa, A., Bouvagnet, P., Kugener, B., Fayol, L., Gonzalez Armengod, C., Oddou, B., Chanavat, V., *et al.* (2006). Spectrum of pathogenic mutations and associated polymorphisms in a cohort of 44 unrelated patients with long QT syndrome. *Clin Genet* 70, 214-227.
- Monaghan, F., and Corcos, A. (1984). On the origins of the Mendelian laws. *J Hered* 75, 67-69.
- Nabbout, R., and Dulac, O. (2011). Epilepsy. Genetics of early-onset epilepsy with encephalopathy. *Nat Rev Neurol* 8, 129-130.
- Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., *et al.* (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242-245.
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272-276.
- Ottman, R., Hirose, S., Jain, S., Lerche, H., Lopes-Cendes, I., Noebels, J.L., Serratos, J., Zara, F., and Scheffer, I.E. (2010). Genetic testing in the epilepsies--report of the ILAE Genetics Commission. *Epilepsia* 51, 655-670.
- Park, H., Kim, J.I., Ju, Y.S., Gokcumen, O., Mills, R.E., Kim, S., Lee, S., Suh, D., Hong, D., Kang, H.P., *et al.* (2010). Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat Genet* 42, 400-405.
- Pendlebury, S.T., Mariz, J., Bull, L., Mehta, Z., and Rothwell, P.M. (2012). MoCA, ACE-R, and MMSE versus the National Institute of Neurological Disorders and Stroke-Canadian Stroke Network Vascular Cognitive Impairment Harmonization Standards Neuropsychological Battery after TIA and stroke. *Stroke* 43, 464-469.
- Petrovski, S., and Kwan, P. (2013). Unraveling the genetics of common epilepsies: approaches, platforms, and caveats. *Epilepsy Behav* 26, 229-233.

- Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat Biotechnol* 28, 1057-1068.
- Prasad, A.N., and Hoffmann, G.F. (2010). Early onset epilepsy and inherited metabolic disorders: diagnosis and management. *Can J Neurol Sci* 37, 350-358.
- Puffenberger, E.G., Hu-Lince, D., Parod, J.M., Craig, D.W., Dobrin, S.E., Conway, A.R., Donarum, E.A., Strauss, K.A., Dunckley, T., Cardenas, J.F., *et al.* (2004). Mapping of sudden infant death with dysgenesis of the testes syndrome (SIDDT) by a SNP genome scan and identification of TSPYL loss of function. *Proc Natl Acad Sci U S A* 101, 11689-11694.
- Puffenberger, E.G., Jinks, R.N., Sougnez, C., Cibulskis, K., Willert, R.A., Achilly, N.P., Cassidy, R.P., Fiorentini, C.J., Heiken, K.F., Lawrence, J.J., *et al.* (2012). Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One* 7, e28936.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., *et al.* (2006). Global variation in copy number in the human genome. *Nature* 444, 444-454.
- Saposnik, G., Black, S.E., Hakim, A., Fang, J., Tu, J.V., and Kapral, M.K. (2009). Age disparities in stroke quality of care and delivery of health services. *Stroke* 40, 3328-3335.
- Saposnik, G., Cote, R., Rochon, P.A., Mamdani, M., Liu, Y., Raptis, S., Kapral, M.K., and Black, S.E. (2011). Care and outcomes in patients with ischemic stroke with and without preexisting dementia. *Neurology* 77, 1664-1673.
- Schaffer, A.A. (2013). Digenic inheritance in medical genetics. *J Med Genet* 50, 641-652.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., *et al.* (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525-528.
- Singh, R.K., and Cooper, T.A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* 18, 472-482.
- Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Jr., Kaye, J., Montine, T.J., *et al.* (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7, 280-292.
- Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., *et al.* (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848-853.

- Strong, M.J., Grace, G.M., Freedman, M., Lomen-Hoerth, C., Woolley, S., Goldstein, L.H., Murphy, J., Shoesmith, C., Rosenfeld, J., Leigh, P.N., *et al.* (2009). Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* *10*, 131-146.
- Sturtevant, A.H., Bridges, C.B., and Morgan, T.H. (1919). The Spatial Relations of Genes. *Proc Natl Acad Sci U S A* *5*, 168-173.
- Tester, D.J., Will, M.L., Haglund, C.M., and Ackerman, M.J. (2005). Compendium of cardiac channel mutations in 541 consecutive unrelated patients referred for long QT syndrome genetic testing. *Heart Rhythm* *2*, 507-517.
- Ullmann, R., Turner, G., Kirchhoff, M., Chen, W., Tonge, B., Rosenberg, C., Field, M., Vianna-Morgante, A.M., Christie, L., Krepischi-Santos, A.C., *et al.* (2007). Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* *28*, 674-682.
- van der Waerden, B.L. (1968). Mendel's experiments. *Centaurus* *12*, 275-288.
- Watson, L.A., Solomon, L.A., Li, J.R., Jiang, Y., Edwards, M., Shin-ya, K., Beier, F., and Berube, N.G. (2013). Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. *J Clin Invest* *123*, 2049-2063.
- Weckselblatt, B., and Rudd, M.K. (2015). Human Structural Variation: Mechanisms of Chromosome Rearrangements. *Trends Genet* *31*, 587-599.
- Wiseman, F.K., Alford, K.A., Tybulewicz, V.L., and Fisher, E.M. (2009). Down syndrome--recent progress and future prospects. *Hum Mol Genet* *18*, R75-83.
- Woolley, S.C., and Strong, M.J. (2015). Frontotemporal Dysfunction and Dementia in Amyotrophic Lateral Sclerosis. *Neurol Clin* *33*, 787-805.
- World Alzheimer Report 2013. Journey of caring. An analysis of long-term care of dementia: <https://www.alz.uk/research/WorldAlzheimerReport2013.pdf> (accessed August 10, 2016).
- Yuen, R.K., Thiruvahindrapuram, B., Merico, D., Walker, S., Tammimies, K., Hoang, N., Chrysler, C., Nalpathamkalam, T., Pellecchia, G., Liu, Y., *et al.* (2015). Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med* *21*, 185-191.
- Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* *109*, 1193-1198.

Chapter 2 - Novel phenotype, novel gene: NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency.

2 Study rationale

As part of the FORGE Canada Consortium, we have ascertained a presumed novel disorder classified here as ‘infantile mitochondrial complex II/III deficiency’ (IMC23D). Testing of mitochondrial and nuclear genes associated with the biochemical phenotype observed in the affected children did not yield any candidate mutations. Multiple clinical and biochemical geneticists could not diagnose the patients with any known disease. The pedigree as presented in addition to the known history of endogamy observed in the Old Order Mennonite community, led us to postulate that IMC23D is an autosomal recessive disease. Accordingly, we applied autozygosity mapping to generate a highly significant homozygous region. Using whole exome sequencing, we identified two genetic variants and we treated both as candidate variants until functional evidence led us to isolate NFS1 as the probable disease-causing gene.

Chapter 2 was published as an independent study: “Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency” in *Molecular Genetics and Genomic Medicine* in 2014 (Farhan et al., 2014, PMID: 24498631).

2.1 Overview

Iron-sulfur (Fe-S) clusters are a class of highly conserved and ubiquitous prosthetic groups with unique chemical properties that allow the proteins that contain them, Fe-S proteins, to assist in various key biochemical pathways. Mutations in Fe-S proteins often disrupt Fe-S cluster assembly leading to a spectrum of severe disorders such as Friedreich’s ataxia or ISCU myopathy. Herein, we describe infantile mitochondrial

complex II/III deficiency, a novel autosomal recessive mitochondrial disease characterized by lactic acidemia, hypotonia, respiratory chain complex II and III deficiency, multisystem organ failure, and abnormal mitochondria. Through autozygosity mapping, exome sequencing, *in silico* analyses, population studies, and functional tests, we identified c.215G>A, p.Arg72Gln in NFS1 as the likely causative mutation. We describe the first disease in man caused by deficiency in NFS1, a cysteine desulfurase that is implicated in respiratory chain function and iron maintenance by initiating Fe-S cluster biosynthesis. Our results further demonstrate the importance of sufficient NFS1 expression in human physiology.

2.2 Introduction

Iron-sulfur (Fe-S) clusters are a class of ubiquitous prosthetic groups highly conserved in essentially all organisms (Sheftell et al., 2010). Their unique chemical properties allows the proteins that contain them to assist in vital and diverse biochemical pathways such as iron metabolism and homeostasis as well as DNA damage repair (Kispal et al., 1999; Lill et al., 2012; Rudolf et al., 2006). Eukaryotic Fe-S proteins are localized to the mitochondria (Lill et al., 2012), cytosol (Tong and Rouault, 2006), endoplasmic reticulum (Lill et al., 2006), and nucleus (Naamati et al., 2009). Fe-S proteins initiate biogenesis of the Fe-S cluster complex, which can transfer electrons, act as catalysts, stabilize protein structures and execute regulatory functions (Kispal et al., 1999; Rudolf et al., 2006). Given the central role of Fe-S clusters in many fundamental biochemical pathways, deficiency in genes encoding Fe-S proteins are expected to cause human disease. Classic examples of Fe-S cluster deficiency related to human disease include: Friedreich's ataxia (FRDA (OMIM 229300), the most common autosomal recessive ataxia due to a homozygous expansion of a GAA trinucleotide repeat in intron 1 of the *FXN* gene (Koeppen, 2011) and ISCU myopathy (OMIM 255125), due to a splicing defect in the *ISCU* gene that alters the C-terminus ultimately reducing protein expression (Kollberg et al., 2009). Other Fe-S cluster related diseases include those caused by mutations in the *NFUI* gene or the *BOLA3* gene, which ultimately disrupt Fe-S enzyme

lipoate synthase activity (Cameron et al., 2011). This leads to a severe infantile multisystem disorder known as multiple mitochondrial dysfunctions syndrome, which is characterized by hyperglycinemia, lactic acidosis, hypertension, and failure to thrive (Cameron et al., 2011).

In this Chapter, we describe infantile mitochondrial complex II/III deficiency, a novel autosomal recessive mitochondrial disease. We used autozygosity mapping, exome sequencing, *in silico* analyses, population studies, and functional assays to isolate c.215G>A, p.Arg72Gln in NFS1, as the probable genetic cause of disease. NFS1 is a cysteine desulfurase implicated in respiratory chain function and iron homeostasis via initiating Fe-S cluster assembly, an essential and highly conserved biochemical process.

2.3 Materials and methods

2.3.1 Ethics

Ethics approval was obtained for all FORGE projects in accordance with the Research Ethics Board (18152) at Western University, London, Ontario, Canada. Ethics approval was obtained from the parents on behalf of all six family members: two parents, three affected children, and one unaffected child. Of the three affected children, the first two, IV-1 and IV-2, are deceased. We also obtained ethical approval from all DNA samples tested in our study.

2.3.2 Patients and biological materials

Blood and tissue (skin-derived fibroblast cells) samples were collected from all family members. An autopsy was conducted on IV-1 and IV-2, and muscle and liver tissue were also obtained.

2.3.3 DNA isolation

DNA was isolated from 4 ml of blood or cultured cells collected from every participant using the Genra Puregene Blood kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. DNA quality and concentration were measured using NanoDrop-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.3.4 Enzymology assays

The following assays were conducted in Dr. Rupal's laboratory at Children's Health Research Institute in London, Ontario. Mitochondrial electron transport chain enzyme activities were measured using muscle and liver autopsy or biopsy samples taken from the first two affected individuals. Mitochondrial enzyme activities were measured using a previously described protocol with minor modifications (Trounce et al., 1996).

2.3.5 Histology

Pathologist, Dr. David Ramsay at University Hospital, performed the following procedures. First, tissue from the muscle biopsy from affected individual IV-II was frozen and fixed in glutaraldehyde. Second, using standard methods, frozen sections were treated with the following stains and histochemical methods: Gomori modified trichrome (GMT); hematoxylin, phloxin and saffron (HPS); acid phosphatase; ATPase (at pH 4.2 and 9.6), cytochrome C oxidase (COX); reduced nicotinamide adenine dinucleotide-tetrazolium reductase (NADH); periodic acid-Schiff (PAS) and succinic dehydrogenase (SDH). Finally, the tissues fixed in glutaraldehyde were post-fixed in osmium tetroxide and embedded in Epon; thin sections (silver interference) were mounted on standard grids and stained with uranyl acetate.

Routine pediatric autopsies were performed on individuals IV-I and IV-II. Representative areas from all organs were examined microscopically. The brains were retained and fixed in 20% formalin. Next, a standard neuropathological examination was conducted, including widespread sampling of various regions for microscopic

examination. All sections were stained with hematoxylin and eosin; in addition, selected sections from the liver were stained with oil red O.

2.3.6 SNP genotyping

Genomic DNA (gDNA) extracted from all family members was genotyped for single nucleotide polymorphisms (SNPs) using Affymetrix® (Santa Clara, CA, USA) protocols as instructed by the manufacturer. 500 ng of genomic DNA was digested with two restriction enzymes, *NspI* and *StyI* separately. All DNA fragments were ligated to sequence specific adaptors complementary to a PCR primer. Each digest was PCR amplified then combined and purified. Finally, PCR products were fragmented, labeled, and hybridized to the Affymetrix® Genome-Wide Human SNP Array 6.0. SNP genotypes were determined using the Birdseed v2 algorithm employed in the Affymetrix® Genotyping Console™ Software 4.1.1 (Korn et al., 2008; Rabbee and Speed, 2006).

2.3.7 Autozygosity mapping

GeneSpring GT v2.0 (Agilent Technologies, Santa Clara, CA, USA) software was used to identify regions of homozygosity that are identical by descent in family members. SNP allele frequencies from healthy Caucasian controls were used to estimate logarithm of the odds (LOD) scores for each SNP. Location scores were calculated by summation of LOD scores of accumulated homozygous regions in the genome (Broman and Weber, 1999; Lander and Botstein, 1987). Homozygous regions unique to the affected individuals with significantly high location scores were treated as candidate loci.

2.3.8 Exome sequencing

Exome sequencing was performed using gDNA from two affected individuals (IV-2 and IV-3) on the Illumina HiSeq 2000 with 2x100 paired end chemistry in accordance with library protocols used at The Centre for Applied Genomics (TCAG) at the Hospital for

Sick Children, Toronto, Ontario. The SureSelect Target Enrichment kit v1 (Agilent Technologies, Santa Clara, CA, USA) was used to enrich 40 Mb of exome region.

Briefly, DNA is enzymatically fragmented and adapters are added on to the sequence, and placed on a flow cell. The Illumina platform uses sequencing-by-synthesis employing reversible dye terminators, which is a DNA polymerase-dependent method that allows for real-time sequencing (Metzker, 2010; Voelkerding et al., 2009). Specifically, the template encounters bridge amplification, as the adapters are complementary to the sequence attached on the flow cell (Metzker, 2010). Extending the vacant 3' OH with a fluorescently labeled primer that is complementary to the adapter hybridizes, and light is emitted and detected, revealing the identity of the nucleotide base (Metzker, 2010).

2.3.9 Sequence alignment

Raw sequencing reads in the form of FASTQ files were aligned to the Hg19 reference genome using the Burrows-Wheeler Aligner (BWA) algorithm to generate SAM/BAM format files, which are sorted by genomic coordinates (Li and Durbin, 2009; Li et al., 2009; Quinlan and Hall, 2010). A BAM file is a binary form of a SAM file, which holds information on all sequencing reads and their respective genomic coordinates. Post-alignment processing steps included flagging and removing duplicate reads generated by PCR amplification bias, using Picard (<http://picard.sourceforge.net/>). Furthermore, local realignment was performed around insertions and deletions as these types of genetic variation can produce sequencing artifacts due to misalignment in previous steps. Finally, recalibration of base quality score was performed to ensure each base call is adequate as downstream variant calling algorithms are score dependent.

2.3.10 Variant calling and annotation

Genome Analysis Toolkit (GATK) was used to generate the coverage of the whole exome in the form of variant calling format (VCF) files (DePristo et al., 2011; McKenna

et al., 2010) in accordance with GATK Best Practices (<https://www.broadinstitute.org/gatk/guide/best-practices>). In conjunction, CLC Bio Workbench (Aarhus, Denmark) was used to generate VCF files for comparison. Variants within VCF files were annotated using ANNOVAR (Wang et al., 2010) and GATK Haplotype Caller (https://www.broadinstitute.org/gatk/guide/tooldocs/org_broadinstitute_gatk_tools_walkers_haplotypecaller_HaplotypeCaller.php), which included context and functional annotations. Context annotations provided information regarding the quality of the sequence reads. Functional annotations provided information on the biological effect of the variants, including: 1) type of variant (synonymous, non-synonymous, splicing, insertions, or deletions); 2) coordinates (genomic, transcript, and protein); 3) zygosity in patients; 4) and the associated allele frequency based on 1000 Genomes (Genomes Project et al., 2012) and dbSNP (Sherry et al., 2001).

2.3.11 Variant discovery

2.3.11.1 Prioritization of homozygous, non-synonymous, and rare variants

Given the rarity and severity of disease and the pedigree structure, we used a homozygous, non-synonymous, and rare variant (minor allele frequency [MAF] <1%) according to NCBI dbSNP and 1000 Genomes, approach to identify any variants. Because we identified a highly significant homozygous region on chromosome 20, we prioritized variants within 20p11.2-q13.1.

2.3.11.2 *In silico* analyses

To determine the predicted biological effect of non-synonymous variants on protein function, all non-synonymous variants identified as candidate variants (NFS1 p.Arg72Gln and CDH22 p.Arg167Cys) were assessed using the following tools: Polymorphism Phenotyping version 2 (PolyPhen-2) (Adzhubei et al., 2010), Sorting Intolerant From Tolerant (SIFT) (Kumar et al., 2009), MutationTaster2 (Schwarz et al.,

2014), and PMut (Ferrer-Costa et al., 2005). To determine amino acid conservation, human protein sequences were aligned to other protein sequences from a diverged set of species using ClustalW2 (Thompson et al., 1994).

2.3.12 Variant validation

To determine whether p.Arg72Gln in *NFS1* and p.Arg167Cys in *CDH22* co-segregated with disease status in the pedigree, we genotyped family members using standard Sanger sequencing methods.

2.3.12.1 PCR

Forward and reverse primers specific to amplify g.1573G>A in *NFS1* were: 5'-TTGGTGGGTAGTTTTGTGGG-3' and 5'-CCATTCTACCTCCATGCAC-3', respectively. Forward and reverse primers specific to amplify g.10682C>T in *CDH22* were: 5'-CAAGTACACCATCTCAGGCGAGG-3' and 5'-TCTTTTCTGTCCCTCCCCAGAGG-3'. PCR conditions for *NFS1* consisted of: initial denaturation at 95°C for 5 minutes; 30 cycles of denaturation at 95°C for 30 seconds; annealing at 58°C for 30 seconds; extension 72°C for 30 seconds, and a final extension at 72°C for 7 minutes. PCR conditions for *CDH22* consisted of: initial denaturation at 95°C for 7 minutes; 30 cycles of denaturation at 95°C for 20 seconds; annealing at 60°C for 20 seconds; extension 72°C for 20 seconds and a final extension at 72°C for 7 minutes.

2.3.12.2 Imaging, purifying, and sequencing of PCR products

PCR products were loaded on a 1.5% agarose gel with SybrSafe DNA gel stain (Life Technologies, Carlsbad, CA, USA) and processed against a 100 bp DNA Ladder (Thermo Fisher Scientific, Cambridge, MA, USA). The gel was maintained at 100 V for 45 minutes using GIBCO BRL Horizontal Gel Electrophoresis Apparatus (Life technologies, Carlsbad, CA, USA) and imaged using Variable Intensity UV

transilluminator VWR (BioDoc-It Imaging System, Upland, CA, USA). Once PCR products aligned to the correct size fragment as determined by the pre-calculated PCR product size and the corresponding ladder fragment size, PCR products were purified using calf intestinal phosphatase/*ExoI* (New England BioLabs, Ipswich, MA, USA) according to manufacturer's instruction.

Purified PCR products were sequenced on an ABI 3730 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). Electropherograms produced were analyzed using Applied Biosystems SeqScape[®] Software version 2.6 with the reference sequence of each gene obtained from NCBI GenBank database.

2.3.13 Population screening

Importantly, while we were conducting this study, there were many resources not available to us that later became available as next generation sequencing and the corresponding annotation pipelines became commonplace. For example, the availability of public databases such as the National Heart, Lung, and Blood Institute Exome Sequencing Project (NHLBI ESP) (<http://evs.gs.washington.edu/EVS/>) and Exome Aggregation Consortium (ExAC) (<http://exac.broadinstitute.org/>) were not yet available or coalesced. As a result, we conducted our own genotyping to determine the MAF of the variants. When we finished our study, the NHLBI ESP database became available and we compared the allele frequencies reported to the allele frequencies we obtained. Furthermore, to the best of our knowledge, these individuals were free from any mitochondrial-like diseases.

2.3.13.1 Genotyping of local population

To determine the allele frequency in the local Old Order Mennonite community, *NFS1* g.1573G>A and *CDH22* g.10682C>T were both screened in 40 healthy Old Order Mennonite controls using standard Sanger sequencing as previously described.

2.3.13.2 Genotyping of an ethnically diverse cohort

To determine the allele frequency in a genetically heterogeneous population, we screened for *NFS1* g.1573G>A and *CDH22* g.10682C>T in a large ethnically diverse cohort. DNA from individuals with ancestral origins from North America, Europe, Asia, and Africa collected by Dr. Hegele and collaborators were used in our study.

2.3.13.2.1 TaqMan genotyping assay

NFS1 g.1573G>A was genotyped in 3,033 individuals from an ethnically diverse cohort using a custom-designed TaqMan genotyping assay (Life Technologies, Carlsbad, CA, USA). Primers used during the custom probe assay were 5'-CGCTCAAGGAAAATGATCTGTTCTG-3' and 5'-GGAGTGTGGGTTCCCATAGTAGTT-3'. SNP genotyping was performed using an allelic discrimination assay on the 7900HT Fast Real-Time PCR System (Life Technologies, Carlsbad, CA, USA) and genotypes were identified using automated software (SDS 2.3; Life Technologies). Reactions were processed in 5 µl volumes with an amplification protocol of: 95°C for 10 minutes; 50 cycles for 95°C for 15 seconds, and 60°C for 1.5 minutes.

2.3.13.2.2 Pooling samples

For technical reasons, we were unable to genotype individuals for *CDH22* g.10682C>T using the same TaqMan genotyping assay previously described. We were also unable to genotype individuals using Sanger sequencing as the cumulative cost of genotyping a well powered cohort would have been ~\$10,000. Therefore, we pooled 781 DNA samples in batches of 8-10 depending on sample availability, quality, or quantity and performed restriction enzyme based genotyping.

2.3.13.2.2.1 Restriction enzyme digestion

2-5 ul of ~100 ng/ul DNA was added to an eppendorf tube in batches of 8-10 samples for technical reasons. Samples were enzymatically digested using endonuclease *AciI* to detect an abolished cut site, in accordance with the manufacturer's instructions (New England BioLabs®, Ipswich, MA, USA.). Post enzyme digestion, products were loaded on a 1.5% agarose gel and maintained at 100 V for 45 mins as previously described.

2.3.13.2.2.2 Visualizing and inferring genotypes

PCR product size was compared to a neighboring 1 Kb DNA ladder (Thermo Fisher Scientific, Cambridge, MA, USA). Genotype of samples was inferred from observing a single band (homozygous) or two bands (heterozygous). Any sample that appeared heterozygous was independently reprocessed by enzymatically digesting the DNA to determine patient zygosity. Three controls were included in every gel: 1) DNA from a CDH22 p.Arg167Cys homozygote; 2) a heterozygote, and 3) a homozygote for the wild type allele.

2.3.14 Cell culture

Skin fibroblasts from patients and healthy controls were cultured in Petri plates with Dulbecco's modified Eagle's medium (GIBCO, Carlsbad, CA, USA) containing 50 units/ml penicillin and 50 µg/ml streptomycin, supplemented with 10% fetal bovine serum and maintained at 37°C/5% CO₂ until 60-70% confluency.

2.3.15 Immunoblotting

Protein levels were measured using immunoblotting and quantified by densitometry. Proteins isolated from cultured fibroblasts were supplemented with 50 mM Tris-Cl pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue and 10% glycerol, and heated to 95°C for 5 min before resolving on 4-20% SDS-PAGE and transferred onto

polyvinylidene difluoride membranes (Invitrogen, Carlsbad, CA, USA). The membranes were blocked in TBS containing 0.1% Tween-20 and 5% fat-free dry milk for 1 hr and then incubated with their respective primary antibody overnight at 4°C. The membranes were then incubated with peroxidase-conjugates anti-species specific secondary for 1 hr, followed by primary antibody visualization with enhanced chemiluminescent detection with horseradish peroxidase and luminol reagent according to the manufacturer's instructions (Santa Cruz Biotechnology, CA, USA).

2.3.16 Quantitative PCR

RNA was isolated from fibroblast cells using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen, Valencia, CA, USA). RNA quality was measured using the Agilent 2100 Bioanalyzer (Agilent Genomics, Santa Clara, CA, USA). RNA was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit (Life Technologies, Carlsbad, CA, USA) to subsequently perform quantitative reverse-transcription PCR (qRT-PCR) using custom designed probes specific to the beginning and end of the *NFS1* transcript 5'-AGCGCACTCTTCTATCAGGTTTGGGA-3' and 5'-AGCTACAACCTCTCTGGACCCCGG-3' and also for the *CDH22* transcript 5'-TGGACCCCAAGACCGGCGTAATCCG-3' and 5'-TTGACATCCAAGACAACACCGCTG C-3' (Life Technologies). Gene expression assays were performed on the 7900HT Fast Real-Time PCR System (Life Technologies), and gene expression profiles were determined using automated software (SDS 2.3; Life Technologies).

2.3.17 Co-immunoprecipitation

Co-immunoprecipitations were performed to isolate wild type and p.Arg72Gln NFS1 proteins along with protein interactors from the lysate samples of cells of controls and patients, respectively. Experimental procedures were in accordance with the manufacturer's instructions with slight modifications using AminoLink Plus Resin (The

Pierce Co-Immunoprecipitation Kit, Thermo Fisher Scientific, Waltham, MA, USA). Co-immunoprecipitations were performed with anti-NFS1 (Santa Cruz Biotechnology, Santa Cruz, CA, USA) using 10 μ g of cell lysate followed by immunoblotting with anti-ISD11 (Aviva Systems Biology, San Diego, CA, USA).

2.3.18 Antibodies

Anti-NFS1 (1:200, Santa Cruz Biotechnology); anti-CDH22 (1:200, LifeSpan Biosciences, Inc, Seattle, WA, USA); and anti-ISD11 (1:100, Aviva Systems Biology) were used. For a positive control, α -tubulin (1:10,000, Sigma-Aldrich, Saint Louis, MO, USA) was used. Anti-mouse (1:10,000, Jackson ImmunoResearch Labs, West Grove, PA, USA) was used as a secondary antibody.

2.3.19 Statistics

The Student's t-test was used to determine the significance of the differences between case and control. All error bars represent SEM. The data are representative of four separate, independent experiments. A *P* value of less than 0.05 was considered significant.

2.4 Results

2.4.1 Clinical characterization of IMC23D

The content in this section was compiled and transcribed by me and extensively reviewed by Dr. Prasad, Dr. Rugar, Dr. Kronick, Dr. Siu, and Dr. Hegele.

Three children born to healthy Canadian Old Order Mennonite third cousins have been extensively studied for a suspected autosomal recessive inherited mitochondrial disease (Figure 2.4.1). All three children were born after an uncomplicated pregnancy at full term by spontaneous vaginal delivery. Birth height, weight, and head circumference

were in the normal range for the third affected child. There were no dysmorphic features in any of the children. Cytogenetic tests did not reveal any structural or numerical chromosomal abnormalities (Table 2.4.1).

The first child was admitted to London Health Sciences Centre (London, Ontario) at seven months of age. She was dehydrated and hypoglycemic after suffering a brief episode of pharyngitis. One day prior to admission, she presented with lethargy, anorexia and oliguria. She had a generalized seizure followed by metabolic acidosis, respiratory decompensation and severe myocardial failure. Her blood lactate, liver enzymes, and amylase levels were elevated (Table 2.4.1). She developed persistent hypotension, worsening oliguria, and peripheral edema. She died of cardiac failure three days post-admission.

The second child was admitted with respiratory failure and cyanosis at six weeks of age. He was normoglycemic but had elevated levels of lactate, aspartate aminotransferase (AST) and creatine kinase (CK) (Table 2.4.1). At 15 weeks, he was readmitted with truncal hypotonia, feeding problems, vomiting and acidosis. He developed multisystem organ failure including cardiorespiratory failure, which required intubation, ventilation and inotropic support. He had focal seizures with cerebral infarction documented by computed tomography, renal failure and disseminated intravascular coagulation. He was treated with a mitochondrial cofactor therapy consisting of riboflavin (50 mg twice daily [BID]), coenzyme Q10 (30 mg BID), thiamine (50 mg BID), vitamin C (100 mg BID), vitamin E (400 IU once daily [OD]), and vitamin K (2 mg OD), which led to a rapid and substantial clinical and biochemical improvement allowing extubation and discharge from hospital. He did continue to have mild hypotonia. At seven months, symptoms recurred, he worsened rapidly and died of cardiac failure three days post admission, despite intensive treatment.

A third child was born to the couple and was monitored for the same suspected mitochondrial disease. Neonatal neurological examination revealed symmetrical mild hypotonia of the upper limbs and to a lesser extent, of the lower limbs. He had symptoms in infancy consistent with those observed in his siblings, which included hypotonia, lethargy, weight loss and abnormal lactate levels (Table 2.4.1). Similarly, he was treated with a mitochondrial cofactor therapy from six months until about 11 years of age. He is

generally healthy today. However, his blood lactate, AST and CK levels remain elevated (Table 2.4.1). While he has no history of seizures or problems related to his heart, he still has some mild developmental and gross motor delay, consisting primarily of a minor tremor and overall low energy. The lactic acidemia has persisted albeit fluctuated, for approximately eight years but has recently abated. The results of metabolite analyses in all three children support the diagnosis of a mitochondrial disease. All three children were negative for causative mutations of four known mitochondrial disorders, mainly (1) mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS [OMIM 540000]); (2) cardiomyopathy with or without skeletal myopathy (OMIM 590050); (3) myoclonic epilepsy with ragged red fibers (MERF [MIM 545000]) and (4) neurogenic muscle weakness, ataxia, and retinitis pigmentosa (NARP/Leigh syndrome [OMIM 551500]) (Table 1). The couple has since conceived a fourth child who does not have the mitochondrial disease but has cystinuria (OMIM 220100), for which she is currently receiving treatment.

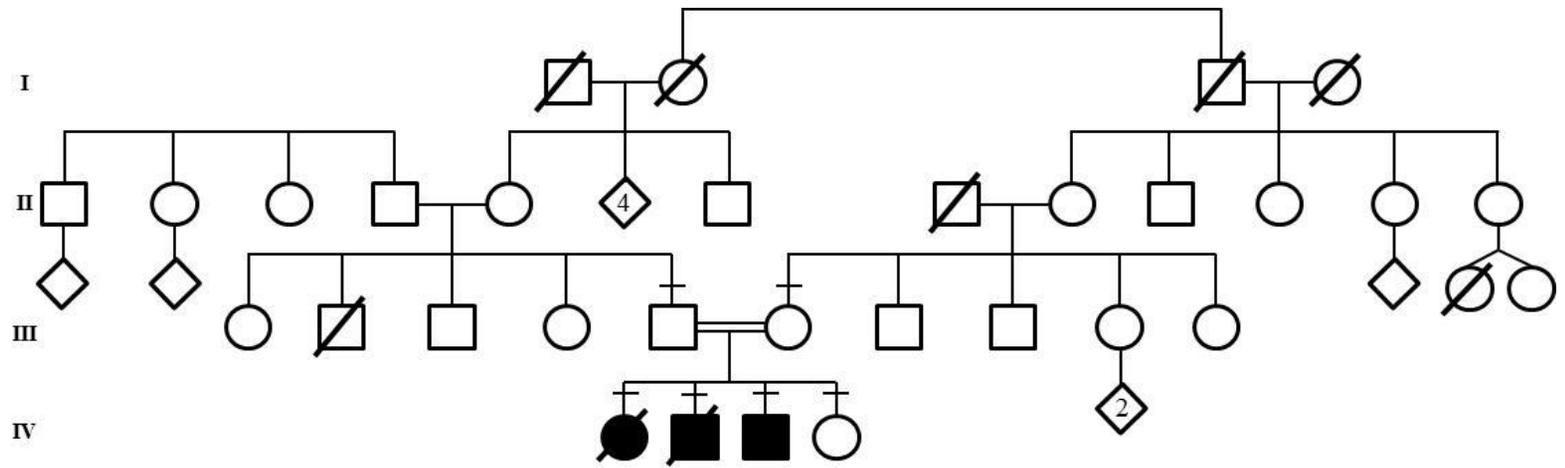


Figure 2.4.1 Pedigree of IMC23D family.

An Old Order Mennonite pedigree showing a union between two healthy third cousins. Three out of four children are affected with autosomal recessive infantile mitochondrial complex II/III deficiency. Affected individuals are shown in shaded squares (male) and circles (female). Diagonal lines across symbols indicate deceased individuals. A double line between two individuals shows a consanguineous union. Horizontal dashes above symbols indicate individuals who underwent DNA analysis. Diamonds indicate unspecified genders.

Table 2.4.1 Clinical and biochemical findings of patients with IMC23D.

Clinical Features	Affected Individuals (year of birth)		
	IV-I (1992)	IV-II (1994)	IV-III (1999)
Gestation	Full term	Full term	Full term
Sex	F	M	M
Karyotype	46,XX	46,XY	46,XY
Facial/limb dysmorphism	-	-	-
Lethargy/anorexia/hypotonia	+	+	+
Autopsy	+	+	+
Age at last assessment	7 months, deceased	7 months, deceased	12 years
Height (cm)			154.9 (50 th percentile)
Weight (kg)			50 (75 th -90 th percentile)
Associated diagnoses			
Respiratory failure	+	+	-
Cardiac failure	+	+	-
Hemorrhagic pancreatitis	+	-	-
Cerebral infarction	-	+*	-
Renal failure	+	+	-
Disseminated intravascular coagulation	+	+	-
Seizures	+	+*	-
Differential diagnoses			
Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), c.3243A>G	-	-	-
Myopathy, c.3260A>G	-	-	-
Cardiomyopathy, c.3303C>T	-	-	-
<i>Myoclonic epilepsy with ragged red fibers (MERF), c.8344A>G</i>	-	-	-
Neurogenic muscle weakness, ataxia, and retinitis pigmentosa (NARP)/Leigh syndrome, c.9883T>G/C	-	-	-
Biochemical findings			
Hypoglycemia	+	+	-

Increased lactate (normal: 0.5-2.2 mmol/L)	+	+	+ (ranged between 4-6 mmol/L)
Increased aspartate aminotransferase	+	+	+
Increased amylase	+	-	
Increased creatine kinase	+	+	+
Increased plasma amino acid concentration	+ (most amino acids)	+	+ (small increase in Alanine)
Urine organic acids	+	N	+
Amino aciduria	-	+	+
Metabolic acidosis	+	+	+
Respiratory chain enzymes			
Muscle mitochondria (nmol/min/mg)			
Complex I + III (range: 37-99, mean: 71)	23 ^a	33 ^b , 29 ^a	
Complex II + III (range: 85-214, mean: 152)	5 ^a	27 ^a , 23 ^b	
Complex IV (range: 193-354, mean: 264)	156 ^a	181 ^b , 111 ^a	
Citrate synthase (range: 170-481, mean: 339)	229 ^a	846 ^b , 540 ^a	
Liver mitochondria (nmol/min/mg)			
Complex I + III (range: 2-14, mean: 7)	7 ^a	19 ^b	
Complex II + III (range: 18-70, mean: 45)	133 ^a	13 ^b	
Complex IV (range: 15-100, mean: 41)	39 ^a	35 ^b	
Citrate synthase (range: 15-53, mean: 33)	37 ^a	14 ^b	

Abbreviations are as follows: N, normal; +, present; -, absent; a, autopsy; b, biopsy; *, at 15 weeks of age. Blank cells indicate unavailable information.

2.4.2 Histopathological evaluation of affected individuals

No specific findings relevant to the molecular abnormalities were noted from the general autopsy materials. However, both cases exhibited hepatic steatosis, a common nonspecific finding in various, including metabolic, disorders. Hemorrhagic pancreatitis with secondary mesenteric fat necrosis, a nonspecific complication of 'critical illness', was present in individual IV-I and there was evidence of acute myocardial ischemic injury in individual IV-II, a common agonal finding. No other definite histological abnormalities were noted in the myocardium although this does not exclude the presence of a significant mitochondrial disorder in the heart. With respect to the brain, no significant or specific abnormalities were seen in either case. There was, however, an old medial left occipital infarct in individual IV-II whose significance is uncertain. This is an uncommon lesion in infants and its cause in this case is unknown; however, the lesion did not have the typical superficial location at the crest of the gyri of the infarcts associated with MELAS nor did it show any similarities in appearance or location to the areas of hemorrhagic vasoproliferation that characterize the central nervous system lesions in Leigh's disease. Its localized nature suggests it is not directly related to the primary metabolic disorder.

Under the light microscope, the majority of the muscle fibers in the biopsy from individual IV-II were unusually small and many of them contained small, faintly basophilic granules in the section stained with HPS; these granules were magenta in the section stained with GMT and exhibited increased enzyme activity in the sections treated with NADH and COX histochemistry. The SDH activity was extremely weak in comparison to normal controls. Focally prominent subsarcolemmal collections of PAS-positive glycogen were observed. The ultrastructural abnormalities are illustrated in Figure 2.4.2 and described in the corresponding legend.

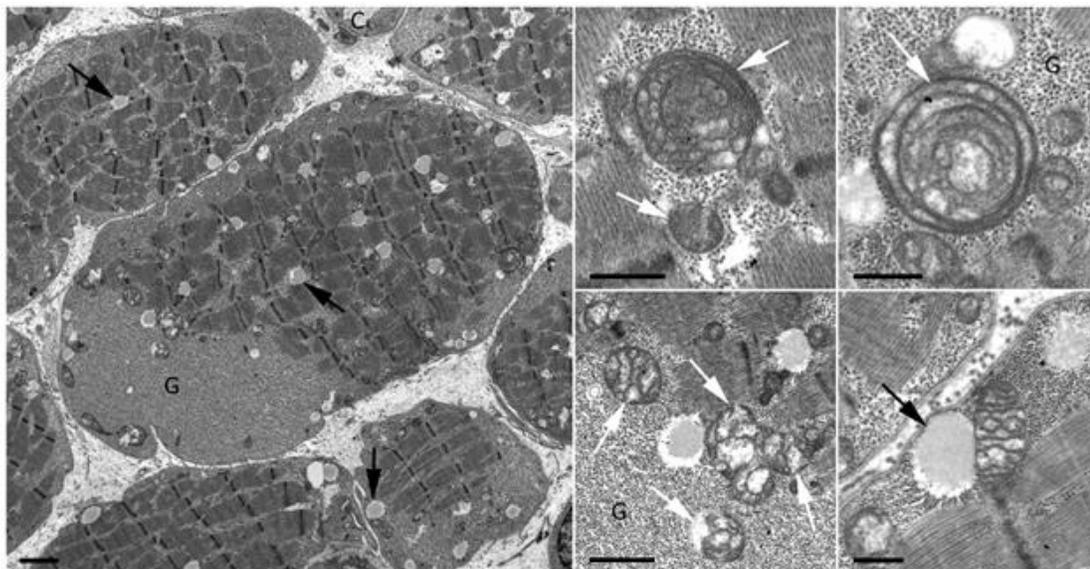


Figure 2.4.2 Pathological findings in patients with IMC23D.

These electron micrographs show scattered lipid droplets (black arrows) and abundant glycogen (G), which are common nonspecific findings. Note also the lack of capillaries (left) - there should be at least three but only one can be seen (C). The lipid droplets are often close to directly or attached to the mitochondria (bottom right). There are a variety of mitochondrial abnormalities (white arrows), including concentric cristae (top middle and right), a honeycomb arrangement of the cristae (top middle and right, bottom middle), vacuolated or 'blown' mitochondria reminiscent of artifacts but possibly also evidence of 'metabolic fragility' (bottom centre), scant cristae (bottom right). Scale bars 2 μm (left), 1 μm (bottom centre), 500 nm (top centre and right, bottom right).

2.4.3 Autozygosity mapping generated 20p11.2-q13.1 as a candidate region

Genome-wide autozygosity mapping generated a locus on chromosome 20p11.2-q13.1, approximately 27.7 Mb with a highly significant location score of 1754 (Figure 2.4.3). The long segment of homozygosity was unique to the affected individuals, which was consistent with an overwhelming likelihood that the causative mutation was harboured within (Figure 2.4.3). The autozygous region contained 4,239 SNPs belonging to 453 genes bordered by SNPs rs4305333:A>C and rs6125184:A>G, which correspond to 20:18891229 and 20:46576714, respectively (Figure 2.4.3).

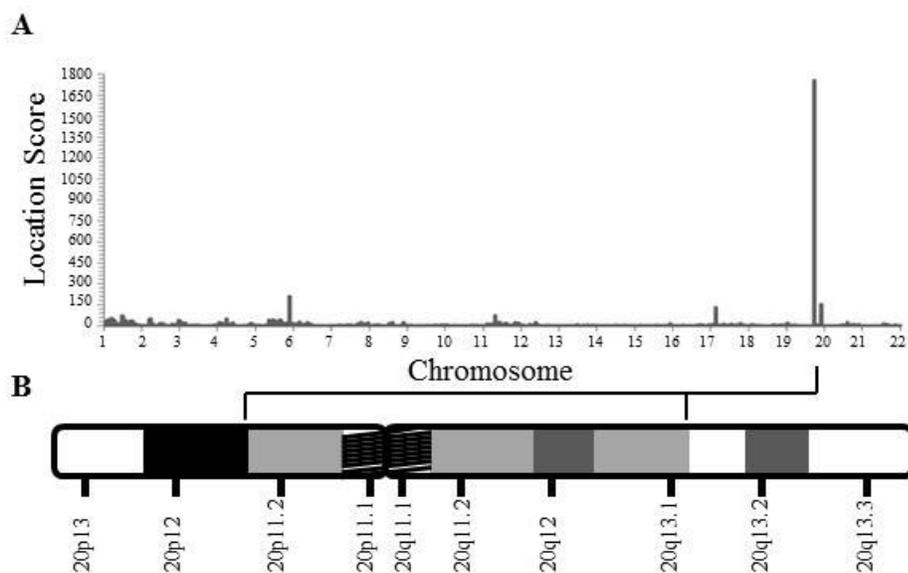


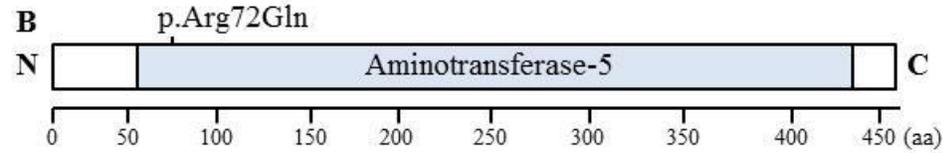
Figure 2.4.3 Autozygosity mapping of IMC23D family

A: Genome-wide autozygosity mapping generated a long homozygous segment unique to the affected individuals on chromosome 20 with a highly significant location score of 1754. B: Ideogram depicting the homozygous segment unique to the three affected individuals, chromosome 20p11.2-q13.1, spanning 27.7 Mb.

2.4.4 Exome sequencing and *in silico* analyses generate candidate variants

Exome sequencing together with bioinformatic and *in silico* analyses were applied to determine suspected disease-causing variants within the autozygous region. A novel homozygous variant was identified within the 27.7 Mb homozygous segment: c.215G>A, p.Arg72Gln in NFS1 (Figure 2.4.4). It was not reported in NCBI dbSNP, 1000 Genomes, or HGMD. However, we surveyed the NHLBI ESP when it became available and the variant was reported to have a MAF of 0.015%. NFS1 p.Arg72Gln is located within the aminotransferase class V domain and PMUT, PolyPhen-2 and SIFT predicted the variant to be ‘pathogenic’. Interestingly, the wild type residue in NFS1 is conserved across all phylogeny including prokaryotes (Figure 2.4.4). It also co-segregated with disease status in the family (Figure 2.4.4). All three patients were homozygous for p.Arg72Gln in NFS1, whereas the father and the mother were heterozygous (Figure 2.4.4). The unaffected sister was homozygous for the wild type allele (Figure 2.4.4).

During our analysis, we identified a second variant within the autozygous region, c.499C>T, p.Arg167Cys in *CDH22*, a calcium-dependent cell adhesion protein predominately expressed in the brain (Saarimaki-Vire et al., 2011; Sugimoto et al., 1996). *CDH22* p.Arg167Cys was predicted to be pathogenic by *in silico* analyses, co-segregated with disease status in the family, and has a MAF of 0.062% according to NHLBI ESP. However, our subsequent functional studies suggest p.Arg167Cys in *CDH22* may not have a role in disease pathology given its specific tissue expression.



C

Human	59	LYMDVQATTPLD	FRVLDAMLPYLINYYGNPHSRTHAYGWESEAA	ERARQQVASLI	114
Chimpanzee	59	LYMDVQATTPLD	FRVLDAMLPYLINYYGNPHSRTHAYGWESEAA	ERARQQVASLI	114
Macaque	59	LYMDVQATTPLD	FRVLDAMLPYLINYYGNPHSRTHAYGWESEAA	ERARQQVASLI	114
Cow	59	LYMDVQATTPLD	FRVLDAMLPYLVNYYGNPHSRTHAYGWESEAA	MECARQQVASLI	114
Horse	59	LYMDVQATTPLD	FRVLDAMLPYLINYYGNPHSRTHAYGWESEAA	ERARQQVASLI	114
Dog	59	LYMDVQATTPLD	FRVLDAMLPYLVNYYGNPHSRTHAYGWESEAA	MEHARQQVASLI	114
Rabbit	53	LYMDVQATTPLD	FRVLDAMLPYLVNYYGNPHSRTHAYGWESEAA	AVERARQQVASLI	109
Rat	60	LYMDVQATTPLD	FRVLDAMLPYLVNYYGNPHSRTHAYGWESEAA	ERARQQVASLI	116
Mouse	60	LYMDVQATTPLD	FRVLDAMLPYLVNYYGNPHSRTHAYGWESEAA	ERARQQVASLI	116
Chicken	47	LYLDVQATTPLD	FRVLDRLPGLTGCYGNPHSRTHAYGWESEAA	TERARQQVADLI	102
Zebrafish	53	LYMDFQATT	PMDFRVLDAMLPYQVNYGNPHSRTHAYGWESE	SAMEKARKQVAGLI	108
Mustard Plant	54	LYLDMQATT	PIDFRVFDAMNASQIHEYGNPHSRTHLYGWEA	ENAVENARNQVAKLI	210
Yeast	89	IYLDVQATT	PVDFRVLDKMLEFYTLGYGNPHSSTHAYGWET	DKEVERARGYIADVI	144
Bacteria	18	IYLDYHATT	PCDFRVEVMLPYLTGDFGNPASGFHVYGRRA	ARAVEMAREQVAALI	74

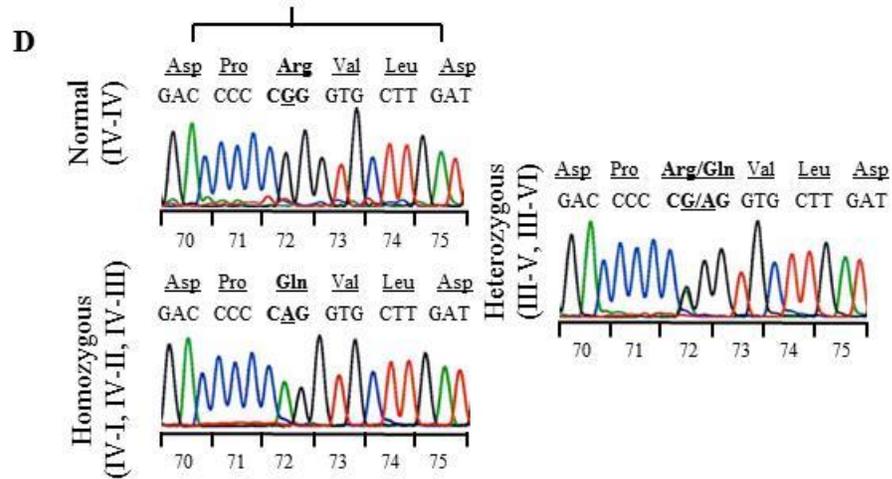


Figure 2.4.4 Mapping and exome sequencing of IMC23D family identifies a highly conserved and destabilizing missense mutation, p.Arg72Gln in NFS1.

A: The *NFS1* gene consists of 13 coding exons with a non-synonymous nucleotide change, c.215T>G in exon one. B: The structure of the NFS1 protein contains one domain, aminotransferase class V domain, shown from the N-terminal to C-terminal end. The amino acid p.Arg72 is harboured within the aminotransferase class V domain. C: Multiple alignments demonstrate high conservation of the amino acid residue p.Arg72 across a set of species-specific NFS1 homologs. A ClustalW analysis of the NFS1 region encompassing the mutation site at residue p.Arg72 (highlighted in red) in aligned homologs with multiple divergent species is shown. The residues shaded in blue indicate fully conserved residues. D: DNA sequence analysis of p.Arg72Gln from DNA of a normal individual (IV-IV, top left electropherogram), a homozygous individual (IV-I, IV-II, IV-III, bottom left electropherogram), and a p.Arg72Gln heterozygous individual (III-V, III-VI, right electropherogram). For each electropherogram, amino acid codes are shown in the top line with nucleotide sequence and codon numbers below. The position of the mutated nucleotide is underlined.

2.4.5 Population screens demonstrate rarity of NFS1 p.Arg72Gln

To determine the MAF of each variant, we began by genotyping 40 otherwise healthy controls from the Old Order Mennonite community. We did not identify any carriers for p.Arg72Gln in NFS1. However, we further genotyped 3,033 individuals from an ethnically diverse cohort and identified one heterozygote (0.016%), which was consistent with the reported MAF in the NHLBI ESP. We were unable to identify carriers for CDH22 p.Arg167Cys in both 40 Old Order Mennonite controls and 781 individuals from an ethnically diverse cohort. However, the NHLBI ESP reported eight heterozygotes in 6,495 individuals, which suggests the likelihood of a homozygous genotype for p.Arg72Gln in NFS1 is 16 times rarer than for p.Arg167Cys in CDH22. Based on the few reported cases of IMC23D, suggesting that p.Arg72Gln in NFS1 is an extremely rare and private mutation within this founder population.

2.4.6 Deficiency in NFS1 expression in patients with IMC23D

We performed immunoblotting to determine whether the variants in *NFS1* and *CDH22* disrupt protein expression. We observed a 73% reduction in NFS1 protein levels (Figure 2.4.6). We could not detect CDH22 protein expression by immunoblotting in wild type or patient cells. Next, we performed qRT-PCR using cDNA synthesized from patient fibroblast RNA, to test gene expression. We observed a 65% reduction in *NFS1* transcript levels relative to a healthy control (Figure 2.4.6). As expected, *CDH22* transcript levels were not detected in both patient and control samples, which is likely due to tissue specificity of *CDH22*, in that it is predominantly expressed in the brain (Saarimaki-Vire et al., 2011; Sugimoto et al., 1996).

Using enzymology assays on muscle and liver tissue, we observed deficiency in mitochondrial respiratory chain complex II and III, which is consistent with decreased NFS1 protein levels and likely, a disruption in Fe-S cluster assembly (Table 2.4.1). Lastly, when wild type and patient fibroblast cells were subjected to co-immunoprecipitation with an antibody against NFS1, ISD11 was co-precipitated with NFS1 in wild type but not in patient cells (Figure 2.4.6). This suggests that the NFS1-

ISD11 complex is disrupted in patient cells. Interestingly, mutations in *LYRM4*, which encodes ISD11, also result in a compromised NFS1-ISD11 complex and leads to a clinical presentation similar to IMC23D (Lim et al., 2013).

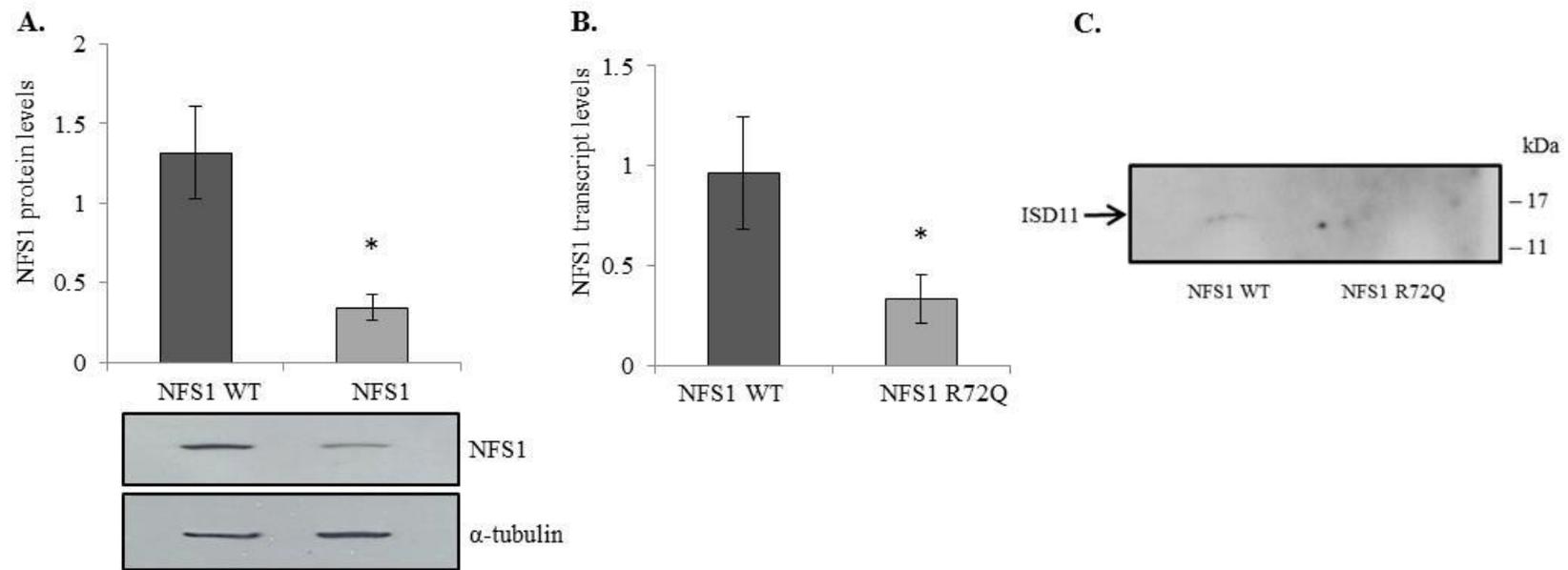


Figure 2.4.6 Decreased NFS1 Protein and Transcript Levels in Patients with IMC23D.

Immunoblotting (A) and qPCR (B) showing reduced NFS1 cellular expression in fibroblast cells of patients affected with IMC23D. Bar graphs indicate means \pm standard deviations from two sets of experiments, showing the relative NFS1 expression in affected individuals normalized to a healthy control quantified by densitometry. The autoradiographs provide a visual representation of NFS1 protein level. The upper blot shows decreased protein expression in NFS1 R72Q relative to NFS1 wild type (WT). The lower blot shows the constitutive expression of α -tubulin in both experiments. C: NFS1-ISD11 interaction is disrupted in patient cells. ISD11 was co-precipitated with NFS1 in wild type (NFS1 WT) but not patient cells (NFS1 R72Q). *represents P-value <0.05.

2.4.7 Characterization of IMC23D summary: *NFS1* and *CDH22*

Table 2.4.7 summarizes the tests performed on variants in *NFS1* and *CDH22* to determine which gene is causal of IMC23D.

Table 2.4.7 Comparison of *NFS1* and *CDH22* as the cause of IMC23D

Test	<i>NFS1</i>	<i>CDH22</i>	Interpretation
Identifying candidate genes in autozygous region			
Chromosome 20 autozygous region (unique to affected individuals only)	Present	Present	Favours both
Type of mutation	Missense	Missense	Favours both, coding region
cDNA location	c.215G>A	c.499C>T	
Amino acid change	p.Arg72Gln (basic → polar)	p.Arg167Cys (basic → polar)	Favours both; amino acid properties changed
CLC BIO verification	Present	Present	Favours both; consistent with AR model
GATK verification	Present	Present	Favours both; consistent with AR model
Bioinformatics approach to determine pathogenicity of mutations			
Location of mutation in protein structure	Aminotransferase class 5 domain	Low complexity region	Favours <i>NFS1</i> as it is harboured in a critical region
Polyphen-2	Probably damaging	Probably damaging	Favours both
SIFT	Damaging	Damaging	Favours both
PMUT	Pathogenic	Pathogenic	Favours both
Reported in HGMD	Not reported	Not reported	Favours both; rare, novel disorder likely the result of a new, private mutation. Consistent with our disease model
Species conservation	Conserved in eukaryotic and prokaryotic domains	Conserved primarily in vertebrate phylum	Favours both, however strengthens the pathogenicity state of <i>NFS1</i> as it is conserved throughout more lineages
Population screening to determine minor allelic frequency (MAF)			
Co-segregated within family	Affected are homozygous; unaffected are heterozygotes or wild type homozygous	Affected are homozygous; unaffected are heterozygotes or wild type homozygous	Favours both; mode of inheritance and AZ mapping is consistent
Allele frequency	1/3,033 (0.016%)	8/6,495 (0.062%)	Favours both. Strengthens support for <i>NFS1</i> as it is a novel,

rare and previously uncharacterized disorder that is likely to be the result of a rare mutation

Measuring functional dysfunction of candidate genes

Tissue expression	Ubiquitous	Pituitary gland, brain, colon, lymphatic tissue. Controversial; likely expressed on in brain	Favours NFS1 as clinical and pathological reports are consistent with many tissues being affected. Does not support CDH22 as clinical presentation and physical examination do not suggest brain, colorectal or lymphatic problems.
Protein localization	Mitochondria and some in the nucleus and cytoplasm	Cell membrane	Favours NFS1 as pathological and clinical tests are highly suggestive of mitochondrial damage. Downstream damage is thought to be the result of mitochondrial deficiency and by-product accumulation
Protein quantification	Decreased expression in affected	No change was detected	Favours NFS1 as clinical data are indicative of mitochondrial deficiency therefore, we would expect candidate protein levels to be decreased and this decrease is likely responsible for the mitochondrial dysfunction. It is also important to mention that skin fibroblasts were used for this experiment, which do not express CDH22.
Transcript quantification	Decreased <i>NFS1</i> transcript levels in affected individuals versus controls	No expression. Skin fibroblast cells were used, this gene is predominantly expressed in brain tissue	Favours <i>NFS1</i> as we see decreased transcript levels in patients, which is consistent with western analysis. Also, it is likely that <i>CDH22</i> is expressed predominantly in brain tissue. Clinical data are in support of ubiquitous <i>NFS1</i> deficiency as there are dysfunctional mitochondria in muscle and liver. It is also important to mention that skin fibroblasts were used for this experiment, which may not express CDH22.
Enzymology assays	- Dysfunction in electron transport chain where NFS1 is located - huNFS1 RNAi experiment enzymology assays are in parallel with ours (Complex II shows	No link to brain, colon, or lymphatic tissue	Favours NFS1; enzyme deficiency is primarily where NFS1 is located/expressed

	decreased function; citrate synthase shows minimal change)		
Modelling dysfunction	- Morphological dysfunction in mitochondria of HeLa cells, recapitulating phenotype - Protein expression observed similar to protein expression we observed	Morphological dysfunction predominantly in adult brain tissue (<i>in vivo</i> assay in rats)	Favours NFS1; NFS1 was specifically knocked down and mitochondrial phenotype is similar to the mitochondrial phenotype seen in patients. These data are collected from other experiments published elsewhere.
Patient clinical presentation			
Clinical data	Directly correlates with mitochondrial dysfunction	Does not directly correlate with brain, colon, or lymphatic tissue	Favours NFS1 as it is targeted to mitochondria
Cause of death: cardiac failure in two of three patients	Highest proportion of mitochondria (NFS1 localized) in cardiomyocytes		Indirectly favours NFS1

2.5 Discussion

Mutations disrupting genes encoding Fe-S proteins are likely to be physiologically harmful due to their ubiquitous expression, and unique and diverse chemical properties, which have been conserved in nearly all organisms, including prokaryotes (Huber et al., 2003; Sheftell et al., 2010). In this present study, we describe a novel autosomal recessive inherited disease observed within the Old Order Mennonites of Ontario likely caused by p.Arg72Gln in NFS1, a cysteine desulfurase that plays a pivotal role in Fe-S cluster biogenesis and assembly (Bridwell-Rabb et al., 2011; Schmucker et al., 2011).

Autozygosity mapping identified a long stretch of identity by descent haplotypes, which mapped to 20p11.2-q13.1, spanning 27.7 Mb. Whole exome sequencing, which generates a list of variants harbored within coding regions of the genome, together with bioinformatic and *in silico* analyses identified two novel homozygous variants within the candidate locus: p.Arg72Gln in NFS1 and p.Arg167Cys in CDH22. We applied multiple *in silico* analyses to predict the pathogenicity of the candidate variants. Based on the markedly higher MAF reported from NHLBI ESP and the specific brain tissue expression of *CDH22*, it is unlikely that CDH22 p.Arg167Cys has a role in the development of infantile mitochondrial complex II/III deficiency (IMC23D). In contrast, NFS1 p.Arg72 is highly conserved, even in prokaryotes, which is consistent with reports describing cysteine desulfurase as a primitive protein involved in evolutionarily ancient pathways (Huber et al., 2003). Population screens demonstrate p.Arg72Gln in NFS1 to likely be a private mutation given its exceptionally low MAF of 0.016% and its disproportionate prevalence in the founder population. Protein and transcript expression assays show a reducing in NFS1 levels, which is likely causative of respiratory chain complex II and III deficiency, in the affected individuals.

The role of NFS1 in Fe-S cluster assembly and iron homeostasis is beginning to be understood. Initially, NFS1 was reported to be predominantly expressed in the mitochondria while its less abundant isoform, which has reduced activity, was localized to the cytosol (Muhlenhoff et al., 2004). However, overexpression of the cytosolic isoform (c-Nfs1) in yeast demonstrated c-Nfs1 is an active cysteine desulfurase that undergoes efficient dimerization equivalent to its mitochondrial counterpart (Tong and

Rouault, 2006). Another study in yeast showed NFS1 is also localized to the nucleus by detecting NFS1 subpopulations first targeted to the mitochondria, then transferred to the cytosol and finally targeted to the nucleus (Naamati et al., 2009). The interaction network of proteins involved in Fe-S cluster assembly is becoming clearer in both yeast and mammalian cells. In particular, NFS1, ISCU, ISD11, and FXN form the Fe-S cluster core complex (Bridwell-Rabb et al., 2011; Prischi et al., 2010; Rouault, 2012; Schmucker et al., 2011; Tsai and Barondeau, 2010). Fe-S cluster formation begins with NFS1 forming a homodimer to which monomers of a scaffold protein ISCU, bind near the top and the bottom (Shi et al., 2010). A co-factor, pyridoxal 5' phosphate, helps NFS1 provide inorganic sulfur from cysteine residues, which then bind to cysteine ligands supplied by ISCU, that further covalently bind to iron (Bandyopadhyay et al., 2008; Raulfs et al., 2008). The core complex then recruits NFS1 binding protein, ISD11 (Adam et al., 2006; Wiedemann et al., 2006) and finally FXN (Prischi et al., 2010; Tsai and Barondeau, 2010). Next, the Fe-S cluster is transferred to recipient apo-proteins via binding of ISCU to chaperone proteins (Craig and Marszalek, 2002; Rouault, 2012). In short, it is clear that Fe-S cluster assembly is a highly conserved multistep process requiring cysteine desulfurases, scaffold proteins, chaperones and iron donors to ultimately maintain iron homeostasis, execute catalysis and gene regulation (Lill et al., 2012; Rouault, 2012).

Cooperatively, Fe-S clusters perform numerous essential cellular functions including electron transfer in oxidative phosphorylation of the respiratory chain, enzymatic catalysis of dehydratases, stabilization of protein structure, and regulation of gene expression, which are essential processes in human biology (Beinert, 2000; Bouton and Drapier, 2003). As a result, mutations in genes encoding Fe-S proteins are expected to cause human disease. For example, in Friedreich's ataxia (FRDA) mutations within FXN lead to abnormal central and peripheral nervous system activity and may lead to heart failure (Campuzano et al., 1996; Koeppen, 2011). In contrast, mutations within ISCU cause ISCU myopathy, which is characterized by lactic acidosis and muscle fatigue (Kollberg et al., 2009). Although IMC23D shares some of its clinical and pathological features with FRDA and ISCU myopathy, it is clinically and genetically distinct. Until now three Fe-S proteins had not been linked to a human disease, namely NFS1, NBP35, and IOP1 (Sheftell et al., 2010). There are arguably two reasons for the lack of

observations: 1) such disorders have not yet been ascertained; or 2) a deficiency in these vital and well-conserved gene products is incompatible with life.

Several studies have examined the cellular localization of NFS1 (Naamati et al., 2009), its role in Fe-S cluster biogenesis and assembly (Tong and Rouault, 2006), as well as its role in other pathways such as donating sulfur to MOCS3 in the molybdenum cofactor biosynthesis pathway (Marelja et al., 2008). However, two studies examined the physiological consequence of cellular NFS1 depletion via small interfering RNA (siRNA)-mediated gene silencing approaches and are of particular interest to us (Biederbick et al., 2006; Fosset et al., 2006). In one study, depletion of mouse NFS1 (m-Nfs1) in murine fibroblasts led to reduced activity of mitochondrial respiratory chain complex I, II, and aconitase, a citric acid cycle protein (Fosset et al., 2006). Interestingly, there was no change in mitochondrial malate dehydrogenase activity or cytosolic lactate dehydrogenase, neither of which contains Fe-S clusters (Fosset et al., 2006). The second study showed severe growth retardation and morphological changes in mitochondria following *NFS1* gene silencing in HeLa cells (Biederbick et al., 2006). Similarly, the authors showed decreased activity in both mitochondrial and cytosolic Fe-S proteins (Biederbick et al., 2006). Interestingly, introduction of m-Nfs1 repaired growth and restored Fe-S protein activity (Biederbick et al., 2006). These findings further demonstrate the importance of proper Fe-S cluster assembly in mammalian cells and are consistent with an NFS1 dependent human deficiency state. These studies also shed light on novel gene therapy approaches using exogenous forms of Fe-S proteins to reverse the physiological consequence of NFS1 depletion.

It is important to mention that although p.Arg167Cys in CDH22 co-segregated with disease status in the family and has a low MAF, there are currently no established mechanisms linking CDH22 to mitochondrial dysfunction. Although we have observed cellular depletion of NFS1 consistent with impaired respiratory chain enzyme activity, we ultimately cannot exclude the possible role of CDH22 in IMC23D. However, it is more likely that p.Arg167Cys in CDH22 is a rare variant with clinically insignificant effects.

2.6 Conclusion

We report a novel variant within *NFS1* as causative of a rare mitochondrial disease characterized by chronic lactic acidemia, hypotonia, respiratory chain complex II and III deficiency, multisystem failure, and abnormal mitochondria. Our detailed clinical description, bioinformatic and *in silico* analyses, population, and functional validation studies have described in part, the etiology of a novel disease. *NFS1* encodes a cysteine desulfurase, which is an Fe-S protein that is highly involved in Fe-S cluster biogenesis, iron maintenance, and DNA repair. Notably, p.Arg72Gln in human mitochondrial NFS1 is likely to disrupt a biochemical pathway owing to its high conservation within essentially all living organisms as well as its role in forming the Fe-S cluster core complex. Our findings highlight the physiological significance of proper Fe-S cluster biogenesis and assembly and their role in human health particularly, Fe-S cluster-related diseases.

2.7 References

- Adam, A.C., Bornhovd, C., Prokisch, H., Neupert, W., and Hell, K. (2006). The Nfs1 interacting protein Isd11 has an essential role in Fe/S cluster biogenesis in mitochondria. *EMBO J* 25, 174-183.
- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nature methods* 7, 248-249.
- Bandyopadhyay, S., Chandramouli, K., and Johnson, M.K. (2008). Iron-sulfur cluster biosynthesis. *Biochem Soc Trans* 36, 1112-1119.
- Beinert, H. (2000). Iron-sulfur proteins: ancient structures, still full of surprises. *J Biol Inorg Chem* 5, 2-15.
- Biederbick, A., Stehling, O., Rosser, R., Niggemeyer, B., Nakai, Y., Elsasser, H.P., and Lill, R. (2006). Role of human mitochondrial Nfs1 in cytosolic iron-sulfur protein biogenesis and iron regulation. *Mol Cell Biol* 26, 5675-5687.
- Bouton, C., and Drapier, J.C. (2003). Iron regulatory proteins as NO signal transducers. *Sci STKE* 2003, pe17.

Bridwell-Rabb, J., Winn, A.M., and Barondeau, D.P. (2011). Structure-function analysis of Friedreich's ataxia mutants reveals determinants of frataxin binding and activation of the Fe-S assembly complex. *Biochemistry* 50, 7265-7274.

Broman, K.W., and Weber, J.L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet* 65, 1493-1500.

Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., *et al.* (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423-1427.

Craig, E.A., and Marszalek, J. (2002). A specialized mitochondrial molecular chaperone system: a role in formation of Fe/S centers. *Cell Mol Life Sci* 59, 1658-1665.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491-498.

Farhan, S.M., Wang, J., Robinson, J.F., Lahiry, P., Siu, V.M., Prasad, C., Kronick, J.B., Ramsay, D.A., Rupar, C.A., and Hegele, R.A. (2014). Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency. *Molecular genetics & genomic medicine* 2, 73-80.

Ferrer-Costa, C., Gelpi, J.L., Zamakola, L., Parraga, I., de la Cruz, X., and Orozco, M. (2005). PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* 21, 3176-3178.

Fosset, C., Chauveau, M.J., Guillon, B., Canal, F., Drapier, J.C., and Bouton, C. (2006). RNA silencing of mitochondrial m-Nfs1 reduces Fe-S enzyme activity both in mitochondria and cytosol of mammalian cells. *J Biol Chem* 281, 25398-25406.

Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.

Huber, C., Eisenreich, W., Hecht, S., and Wachtershauser, G. (2003). A possible primordial peptide cycle. *Science* 301, 938-940.

Koeppen, A.H. (2011). Friedreich's ataxia: pathology, pathogenesis, and molecular genetics. *J Neurol Sci* 303, 1-12.

Kollberg, G., Tulinius, M., Melberg, A., Darin, N., Andersen, O., Holmgren, D., Oldfors, A., and Holme, E. (2009). Clinical manifestation and a new ISCU mutation in iron-sulphur cluster deficiency myopathy. *Brain* 132, 2170-2179.

Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., *et al.* (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 40, 1253-1260.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 4, 1073-1081.

Lander, E.S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567-1570.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Lill, R., Hoffmann, B., Molik, S., Pierik, A.J., Rietzschel, N., Stehling, O., Uzarska, M.A., Webert, H., Wilbrecht, C., and Muhlenhoff, U. (2012). The role of mitochondria in cellular iron-sulfur protein biogenesis and iron metabolism. *Biochim Biophys Acta* 1823, 1491-1508.

Lim, S.C., Friemel, M., Marum, J.E., Tucker, E.J., Bruno, D.L., Riley, L.G., Christodoulou, J., Kirk, E.P., Boneh, A., Degennaro, C., *et al.* (2013). Mutations in LYRM4, encoding iron-sulfur cluster biogenesis factor ISD11, cause deficiency of multiple respiratory chain complexes. *Hum Mol Genet.*

Marelja, Z., Stocklein, W., Nimtz, M., and Leimkuhler, S. (2008). A novel role for human Nfs1 in the cytoplasm: Nfs1 acts as a sulfur donor for MOCS3, a protein involved in molybdenum cofactor biosynthesis. *J Biol Chem* 283, 25178-25185.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20, 1297-1303.

Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nature reviews Genetics* 11, 31-46.

Muhlenhoff, U., Balk, J., Richhardt, N., Kaiser, J.T., Sipos, K., Kispal, G., and Lill, R. (2004). Functional characterization of the eukaryotic cysteine desulfurase Nfs1p from *Saccharomyces cerevisiae*. *J Biol Chem* 279, 36906-36915.

Naamati, A., Regev-Rudzki, N., Galperin, S., Lill, R., and Pines, O. (2009). Dual targeting of Nfs1 and discovery of its novel processing enzyme, Icp55. *J Biol Chem* 284, 30200-30208.

- Prischi, F., Konarev, P.V., Iannuzzi, C., Pastore, C., Adinolfi, S., Martin, S.R., Svergun, D.I., and Pastore, A. (2010). Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nat Commun* 1, 95.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Rabbee, N., and Speed, T.P. (2006). A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22, 7-12.
- Raulfs, E.C., O'Carroll, I.P., Dos Santos, P.C., Unciuleac, M.C., and Dean, D.R. (2008). In vivo iron-sulfur cluster formation. *Proc Natl Acad Sci U S A* 105, 8591-8596.
- Rouault, T.A. (2012). Biogenesis of iron-sulfur clusters in mammalian cells: new insights and relevance to human disease. *Dis Model Mech* 5, 155-164.
- Saarimaki-Vire, J., Alitalo, A., and Partanen, J. (2011). Analysis of Cdh22 expression and function in the developing mouse brain. *Dev Dyn* 240, 1989-2001.
- Schmucker, S., Martelli, A., Colin, F., Page, A., Wattenhofer-Donze, M., Reutenauer, L., and Puccio, H. (2011). Mammalian frataxin: an essential function for cellular viability through an interaction with a preformed ISCU/NFS1/ISD11 iron-sulfur assembly complex. *PLoS One* 6, e16199.
- Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods* 11, 361-362.
- Sheftell, F., Almas, M., Weeks, R., Mathew, N.T., Pitman, V., and Lipton, R.B. (2010). Quantifying the return of headache in triptan-treated migraineurs: an observational study. *Cephalalgia* 30, 838-846.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic acids research* 29, 308-311.
- Shi, R., Proteau, A., Villarroya, M., Moukadiri, I., Zhang, L., Trempe, J.F., Matte, A., Armengod, M.E., and Cygler, M. (2010). Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-protein interactions. *PLoS Biol* 8, e1000354.
- Sugimoto, K., Honda, S., Yamamoto, T., Ueki, T., Monden, M., Kaji, A., Matsumoto, K., and Nakamura, T. (1996). Molecular cloning and characterization of a newly identified member of the cadherin family, PB-cadherin. *J Biol Chem* 271, 11548-11556.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22, 4673-4680.

Tong, W.H., and Rouault, T.A. (2006). Functions of mitochondrial ISCU and cytosolic ISCU in mammalian iron-sulfur cluster biogenesis and iron homeostasis. *Cell Metab* 3, 199-210.

Tsai, C.L., and Barondeau, D.P. (2010). Human frataxin is an allosteric switch that activates the Fe-S cluster biosynthetic complex. *Biochemistry* 49, 9132-9139.

Voelkerding, K.V., Dames, S.A., and Durtschi, J.D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry* 55, 641-658.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 38, e164.

Wiedemann, N., Urzica, E., Guiard, B., Muller, H., Lohaus, C., Meyer, H.E., Ryan, M.T., Meisinger, C., Muhlenhoff, U., Lill, R., *et al.* (2006). Essential role of Isd11 in mitochondrial iron-sulfur cluster synthesis on Isu scaffold proteins. *EMBO J* 25, 184-195.

Chapter 3 - Novel phenotype, known gene: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses.

3 Study rationale

Similar to the disorder described in Chapter 2, seizures-scoliosis-macrocephaly (SSM) syndrome affects an Old Order Mennonite family. The family agreed to participate in genetic testing, and SSM syndrome became another project funded by the FORGE Canada Consortium.

During clinical visits, affected individuals were extensively examined for any known biochemical and genetic abnormalities with no promising diagnosis obtained. Nerve, muscle, and skin biopsies have been performed on two of the affected children with findings weakly suggestive of a mitochondrial disorder. Initially, some of the features observed in the affected children were consistent with FG syndrome (Opitz-Kaveggia syndrome, OMIM 305450), which is an X-linked recessive disorder and is inconsistent with the most probable mode of inheritance in the pedigree. Lastly, the family visited Dr. Holmes Morton at the Clinic for Special Children in Pennsylvania, as he is especially knowledgeable in the genetic disorders of Amish and Mennonite communities. However, he was unable to suggest a diagnosis.

Chapter 3 was published as an independent study: “Old gene, new phenotype: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses” in *Journal of Medical Genetics* in 2015 (Farhan et al., 2015)

3.1 Overview

Heparan sulfate proteoglycans are vital components of the extracellular matrix and are essential for cellular homeostasis. Many genes are involved in modulating heparan sulfate synthesis and when these genes are mutated, they can give rise to early-onset developmental disorders affecting multiple body systems. Herein, we describe a

consanguineous family of 4 sibs with a novel disorder, which we designate as seizures-scoliosis-macrocephaly (SSM) syndrome, characterized by seizures, intellectual disability, hypotonia, scoliosis, macrocephaly, hypertelorism, and renal dysfunction. Our application of autozygosity mapping and exome sequencing allowed us to identify variants in the patients. To confirm the autosomal recessive mode of inheritance, all available family members were genotyped. We also studied the effect of these variants on protein expression and function in patient cells and also using an *in vitro* system. We identified two homozygous variants, p.Met87Arg and p.Arg95Cys in exostosin 2, EXT2, a ubiquitously expressed gene that encodes a glycosyltransferase, which is required for heparan sulfate biosynthesis. In patient cells, we observed diminished EXT2 expression and function. We also performed an *in vitro* assay to determine which variant has a larger effect on protein expression and observed reduced EXT2 expression in constructs expressing either one of the variants but a greater reduction when both residues were mutated. In short, we have identified the genetic basis of a novel recessive disorder, seizures-scoliosis-macrocephaly (SSM) syndrome. Our results have implicated a well-characterized gene in a new developmental disorder and have further illustrated the spectrum of phenotypes that can arise due to errors in glycosylation.

3.2 Introduction

Heparan sulfate proteoglycans are ubiquitously expressed at nearly all cell surfaces as well as extracellular matrices, and modulate the activity of many growth and developmental factors such as FGF, BMP, TGF- β , Shh, and Wnt (Astudillo and Larrain, 2014; Kreuger et al., 2006; Perrimon and Bernfield, 2000; Shimokawa et al., 2011; Venero Galanternik et al., 2015). Heparan sulfate proteoglycans are involved in cell migration, neuron elongation, synapse formation, and regulation of intracellular signaling pathways (Bandtlow and Zimmermann, 2000; Bishop et al., 2007). The heparan sulfate backbone is formed by the exostosins, EXT1 and EXT2, which are ubiquitous glycosyltransferases involved in a multistep process in the Golgi apparatus (Ahn et al., 1995; McCormick et al., 2000; Stickens et al., 1996). Previous studies have demonstrated

that heparan sulfate synthesis enzymes EXT1 and EXT2, form a stable hetero-oligomeric complex *in vivo* and together have significantly higher glycosyltransferase activity than either enzyme alone (McCormick et al., 2000). These findings suggest that mutations in either exostosin can disrupt heparan sulfate biosynthesis in addition to halting growth factor signaling cascades (McCormick et al., 2000; Venero Galanternik et al., 2015).

We have recently identified a novel, rare disease designated as seizures-scoliosis-macrocephaly (SSM) syndrome, which was first observed in four Old Order Mennonite sibs. SSM syndrome is characterized by seizures, intellectual disability, hypotonia, scoliosis, macrocephaly, hypertelorism, and renal dysfunction. To the best of our knowledge, this is the first report describing SSM syndrome, which we believe is a heparan sulfate-related disease. Based on the small founder population and endogamy observed in the Mennonite community, leading to low genetic variability, we used homozygosity genetic mapping to identify homozygous regions within the genomes of SSM syndrome patients. We then used exome sequencing and identified two different homozygous, rare missense variants in *EXT2*. We confirmed disease mode of inheritance through genetic studies and performed several functional assays to demonstrate loss of *EXT2* as the cause of SSM syndrome.

3.3 Materials and methods

3.3.1 Ethics

As described in Chapter 2, subsection 2.3.1, ethics approval was obtained from the parents on behalf of all seven members of the family.

3.3.2 Patients and biological materials

Four children of a healthy Old Order Mennonite couple were clinically ascertained because they were experiencing seizures and also had macrocephaly. Blood and tissue

(skin-derived fibroblast cells) samples were collected from all four affected children, an unaffected sibling, and the two parents.

3.3.3 DNA isolation

DNA was isolated from all seven family members as described in Chapter 2, subsection 2.3.3.

3.3.4 Genotyping

Genomic DNA (gDNA) extracted from all family members was genotyped for single nucleotide polymorphisms (SNPs) using Affymetrix[®] protocols as described in Chapter 2, subsection 2.3.6.

3.3.5 Autozygosity mapping

GeneSpring GT v2.0 software was used to identify regions of homozygosity that are identical by descent in family members as described in Chapter 2, subsection 2.3.7.

3.3.6 Exome sequencing

Exome sequencing was performed using gDNA from two affected individuals (II-1 and II-4) using the protocols described in Chapter 2, subsection 2.3.8.

3.3.7 Sequence alignment

Sequence alignment was performed using the same procedures described in Chapter 2, subsection 2.3.9.

3.3.8 Variant calling and annotation

Variant calling and annotation were performed using the same procedures described in Chapter 2, subsection 2.3.10.

3.3.9 Variant discovery

3.3.9.1 Prioritization of homozygous, non-synonymous, and rare variants

We investigated whether there were any non-synonymous, rare variants within the homozygous regions on chromosomes 1, 7, 11, and 13. We also applied an unbiased, exome-wide approach to identify any homozygous, non-synonymous, and rare variants as described in Chapter 2, subsection 2.3.11.1. We scanned for variants in genes known to cause seizures and intellectual disability based on literature and OMIM, as well as all genes.

3.3.9.2 *In silico* analyses

To determine the predicted biological effect of non-synonymous variants on protein function, all non-synonymous variants identified as candidate variants (EXT2 p.Met87Arg and p.Arg95Cys) were assessed using the same tools described in Chapter 2, subsection 2.3.11.2.

3.3.10 Variant validation

To determine whether p.Met87Arg and p.Arg95Cys in EXT2 segregated with disease status in the pedigree, we genotyped family members using standard Sanger sequencing methods as described in Chapter 2, subsection 2.3.12.1 and 2.3.12.2.

3.3.10.1 PCR

Forward and reverse primers specific to amplify g.12424T>G and g.12447C>T in *EXT2* were: 5'-CTCTTCTCCATTGTCCTCC-3' and 5'-TCTGAGATGGCCATGAGC-3', respectively. The PCR conditions were as follows: initial denaturation at 95°C for 5 minutes; 30 cycles of denaturation at 95°C for 30 seconds; annealing at 58°C for 30 seconds; extension 72°C for 30 seconds and a final extension at 72°C for 7 minutes.

3.3.10.2 Imaging, purifying, and sequencing of PCR products

Imaging, purifying, and sequencing of PCR products were performed using the same procedures described in Chapter 2, subsection 2.3.12.2.

3.3.11 Population screening

3.3.11.1 Genotyping of local population

39 healthy Old Order Mennonite controls and 311 Caucasian controls were also genotyped by standard Sanger sequencing for *EXT2* g.12424T>G and g.12447C>T, as aforementioned.

3.3.11.2 Identifying overall MAF of *EXT2* variants

We also used 1000 Genomes (Genomes Project et al., 2012), Exome Variant Server available through the Exome Sequencing Project (Fu et al., 2013), and Exome Aggregation Consortium (ExAC) genome browser (Lek et al., 2015) to identify the carrier frequency of the variants, as these resources became available.

3.3.12 Cell culture

3.3.12.1 Patient cells

Skin fibroblasts from patients and healthy controls were cultured as described in Chapter 2, subsection 2.3.14.

3.3.12.2 Mutagenesis

EXT2 variants: (1) p.Arg95Cys; (2); p.Met87Arg; and (3) p.Met87Arg and p.Arg95Cys, were introduced via site-directed mutagenesis using the Stratagene QuikChange II, QuikChange IIXLor Quik-Change Lightning kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. Wild type and variant constructs were transformed into *E.coli* DH5 α cells (Subcloning efficiency, DH5 α , Invitrogen by Life Technologies, Carlsbad, CA, USA) according to manufacturer's instructions. Miniprep DNA isolation was completed to isolate EXT2 wild type and variant clones (1) p.Arg95Cys; (2); p.Met87Arg; and (3) p.Met87Arg and p.Arg95Cys, from bacterial cultures using QIAprep Spin Miniprep Kit (QIAGEN) in accordance with manufacturer's instructions. All plasmid sequences were verified by standard Sanger sequencing as previously described. All primers were purchased from Sigma-Aldrich (Saint Louis, MO, USA).

3.3.12.3 Transfection

COS-7 and HEK293 cells were also cultured and cells were seeded at a density of 7×10^5 per 60 mm dish 24 hours before transfection. One microgram of (1) wild type; or variant EXT2 plasmid DNA containing: (2) p.Arg95Cys; (3); p.Met87Arg; or (4) p.Met87Arg and p.Arg95Cys, was transfected using Effectene Transfection Reagent and Enhancer (Qiagen, Venlo, Limburg, Netherlands) in accordance with the manufacturer's instructions.

3.3.13 Protein isolation

Transfections: 48 hours post-transfections, cells were washed twice with 1xPBS and lysed in sodium dodecyl sulfate (SDS) buffer (50 mM Tris-Cl, pH 6.8, 2% SDS, 10% glycerol) that had been preheated at 95°C for 5 mins. Samples were then incubated at 95°C for 5 mins. Protein concentrations of whole cell lysates were determined using the DC Protein Assay kit (Bio-Rad Laboratories Inc, Hercules, CA, United States of America). The same protocol was used for isolating protein from patient fibroblast cells.

3.3.14 Immunoblotting

Protein expression was measured using immunoblotting as described in Chapter 2, subsection 2.3.15.

3.3.15 Reverse transcriptase (RT)-PCR

RNA was isolated from patient fibroblast cells using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen, Valencia, CA, USA). RNA quality was measured using the Agilent 2100 Bioanalyzer RNA 6000 Nano. RNA was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit (Life Technologies, Carlsbad, CA, USA) to subsequently perform RT-PCR.

3.3.16 Antibodies

Anti-EXT2 (1:500, Abnova, Taipei City, Taiwan); anti-DDK (tagged to exogenous EXT2) (1:1000, Rockville, MD, USA); and anti-NDST1 (1:200, Santa Cruz Biotechnology, CA, USA) were used. For a positive control, α -tubulin (1:10,000, Sigma-Aldrich, Saint Louis, Missouri) was used. Anti-mouse (1:10,000, Jackson ImmunoResearch Laboratories, Inc) was used as a secondary antibody.

3.3.17 Statistics

Statistics were performed as described in Chapter 2, subsection 2.3.19.

3.4 Results

3.4.1 Clinical description of patients with SSM syndrome

The content in this section was compiled and transcribed by me and extensively reviewed by Dr. Siu, Dr. Rugar, and Dr. Hegele.

To the best of our knowledge, SSM syndrome is a novel developmental disorder. Four children (three males, one female) born to consanguineous (third cousin) Old Order Mennonite parents are affected (Figure 3.4.1). Features seen in all affected individuals include: moderate intellectual disability, seizure disorder with onset between two and five years of age, hypotonia, scoliosis, macrocephaly, coarse facies, bilateral cryptorchidism in males, long hypoplastic philtrum, and hypertelorism. For a full clinical description of patients with SSM syndrome, please refer to Table 3.4.1. Extensive metabolic workup did not yield any diagnostic possibilities (Table 3.4.1).

Case II-1, now age 30, was identified as a floppy and weak child with a macrocephaly and delays in motor milestones. He did not sit until he was one year old, and began to walk at the age of three years. His motor skills showed regression as he entered his teenage years. Cardiac assessment disclosed a ventricular septal defect with a possible subaortic stenosis. While chronic constipation was a feature in early life, diarrhea has been a feature lately. Significant gastroesophageal reflux and associated ulceration of the esophagus were documented. At age three years, he developed seizures that were generalized tonic clonic in nature, lasting two to three minutes. In addition, he has significant scoliosis and kyphosis for which he wears a brace and his gait is stiff with out-toeing on a broad base.

Similarly, case II-3, currently 25 years old, presented with general hypotonia at birth, exhibited dysmorphic features, delayed motor milestones, and developed seizures at the age of three years. He too suffers from dysfunctional gastrointestinal motility, as he

tends to vomit easily and has documented gastroesophageal reflux. At age 18 months, he developed hemolytic uremic syndrome and required peritoneal dialysis. Case II-4 had a very similar phenotype to her brothers' however, she had recurrent episodes of prolonged status epilepticus. During one episode at the age of five, she had a four-hour long seizure and sustained significant brain injury. She was rendered hemiplegic on the right side, and has been wheelchair bound since that event. She experiences breakthrough seizures frequently, which appear to be mostly complex partial and are self-limited events of short duration.

The pregnancy with case II-5 was complicated by polyhydramnios. At four days of age, he was diagnosed with intestinal volvulus, which required surgical intervention. His early motor milestones were apparently delayed and at three years of age, he began having generalized seizures lasting up to two hours and occurring every six weeks. At 17 years of age, after a mild respiratory illness, he developed status epilepticus and subsequently died.

In terms of development, all four affected sibs were capable of independent ambulation although case II-4 lost her ability to walk after an episode of prolonged status epilepticus. They have minimal expressive speech, but are able to comprehend simple verbal commands. From a behavioral perspective, they appear shy, and do not sustain eye contact. The epileptic seizures though lacking a distinctive semiology and appear to be generalized except in subject II-4 where seizures are associated with non-specific slowing of background rhythms and focal spikes in the temporal region on EEG. Apart from the (kypho)scoliosis in all 4 affected sibs and possible segmentation defects in the upper spine in one individual, there are no obvious skeletal abnormalities, and specifically no evidence of exostoses.

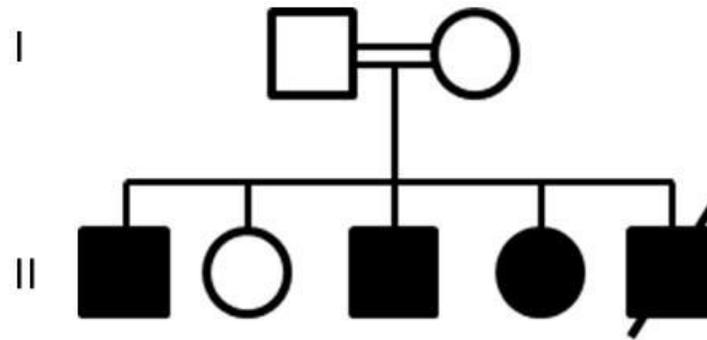


Figure 3.4.1 Pedigree with four children affected with SSM syndrome.

Squares depict males; circles depict females; affected individuals are shaded; diagonal lines across symbols indicate deceased individuals; a double line between two individuals represents consanguineous union.

Table 3.4.1 Clinical description of patients with SSM syndrome.

Clinical Features	Affected individuals (year of birth)			
	II-1 (1985)	II-3 (1990)	II-4 (1992)	II-5 (1994, deceased)
Gestation	Full term	Full term	39 weeks	Full term
Polyhydramnios	+			+
Delivery complications	Breech; caesarean section	SVD	SVD; vertex	Forceps
Gestational hypertension	+		+	+
Sex	M	M	F	M
Karyotype	46,XY	46,XY	46,XX	46,XY
Birth length (cm)			66	
Birth weight (g)	3,221 (70th percentile)	3,583 (78th percentile)	3,583 (78th percentile)	4,037 (88th percentile)
Age at last assessment (years)	19	14	11	10
Head circumference (cm)	59 (>98th percentile)	57 (90th percentile)		58 (98th percentile)
Height (cm)	179 (75th percentile)	157 (10th percentile)		138 (50th percentile)
Weight (kg)	65 (50th percentile)	42 (10-25th percentile)		32 (50-75th percentile)
Hand length (cm)	19.7 (>98th percentile)	17 (25th percentile)	15.8 (25th percentile)	
Middle finger length (cm)	8.5 (>97th percentile)	7.2 (25th percentile)	6.8 (25th percentile)	
Ear length (cm)	6.5 (75th percentile)	6.5 (75th percentile)	5.6 (25th percentile)	
Inner canthal distance (cm)	3.8 (>98th percentile)	3.2 (50-75th percentile)		
Outer canthal distance (cm)	10.7 (>97th percentile)	8 (3rd percentile)		
Associated diagnoses				
Developmental delay	+	+	+	+
Sleep apnea				+
Tremor				Coarse
Seizures, onset (years)	+, 2.5	+, 4-5	+, 2-3	+, 5
Brain hemorrhage		+	+	
Circulation		Cold extremities	Cold extremities	
Strabismus			+, left eye	+, exotropia
Hypotonia	+	+	+	+

Ventricular septal defect	+		+	-
Macrocephaly	+	+	+	+
Cryptorchidism	+	+		+
Precocious puberty				+
Scoliosis	+	+	+	
Hiatus hernia	+			
Bowel dysfunction	+		+	+
Hypertension	+			
Hematuria	+	+, Kidney failure		
Proteinuria	+	+		
Wide-based gait	+		Loss of ambulation	
Overlapping toes	+		+	
Hepatic dysfunction				+
Bowel malrotation		+		+
Volvulus				+
Encopresis	+			
Morning nausea	+	+	+	+
Regression	+			+
Dysphagia	+		Tube-fed	
Low bone density	+		+	+
Sensitive skin	+	+	+	+
Hearing problems	-	-	-	-
Metabolic evaluations				
Mitochondrial respiratory chain	N	N	A	N
NADH: cytochrome C reductase (complexes I +III) (normal: 22-139 nmol/min/mg)	93		26	77
Succinate: cytochrome C reductase (complexes II + III) (normal: 70-268 mmol/min/mg)	210		38	105
Cytochrome oxidase (complex IV) (normal: 173-443 nmol/min/mg)	457		115	193

Mitochondrial DNA mutations				
Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), c.3243A>G	N		N	
Cardiomyopathy, c.3303C>T			N	
Neurogenic muscle weakness, ataxia, and retinitis pigmentosa (NARP)/Leigh syndrome, c.9883T>G/C			N	
mtDNA deletion	N			
Urine oligosaccharides			N	N
Urine organic acids		A	N	N
Urine amino acids	N	N	N	N
Urine mucopolysaccharides screening	N	N	N	N
Plasma amino acids		N	N	N
Plasma carnitine free		N	N	N
Plasma carnitine acyl	N	N	N	N
Lactate	N	N	N	N
Pyruvate	N	N	N	N
Lactate/Pyruvate	N	N	N	N
Medium chain acyl-CoA dehydrogenase deficiency screening	N			
Carbohydrate-deficient glycoprotein syndrome screening	N			
Niemann-Pick disease, Type A and Type B				
Sphingomyelinase (normal: 16-54 nmol/hr/ mg)	66			
Diagnostic imaging				
AP Pelvis/bilateral femurs,	N			

tibia, and fibula	
AP and lateral thoracic and lumbosacral spine	N
Lateral cervical spine	Failure of segmentation of C2-3 and possibly C1-2
AP and lateral skull	N
AP bilateral humerus/radius, ulna, and hands	N
Bilateral AP feet	N

Abbreviations are as follows: A, abnormal; N, normal; +, present; -, absent; AP, anterior posterior. Empty cells indicate unavailable information. Mitochondrial DNA mutation testing was examined on muscle tissue.

3.4.2 Genome-wide autozygosity mapping reveals highly significant homozygous regions

Considering the consanguineous union between the parents, we modelled an autosomal recessive mode of inheritance and first performed homozygosity genetic mapping, which isolated four significant homozygous segments on chromosomes 1, 7, 11, and 13 (Figure 3.4.2).

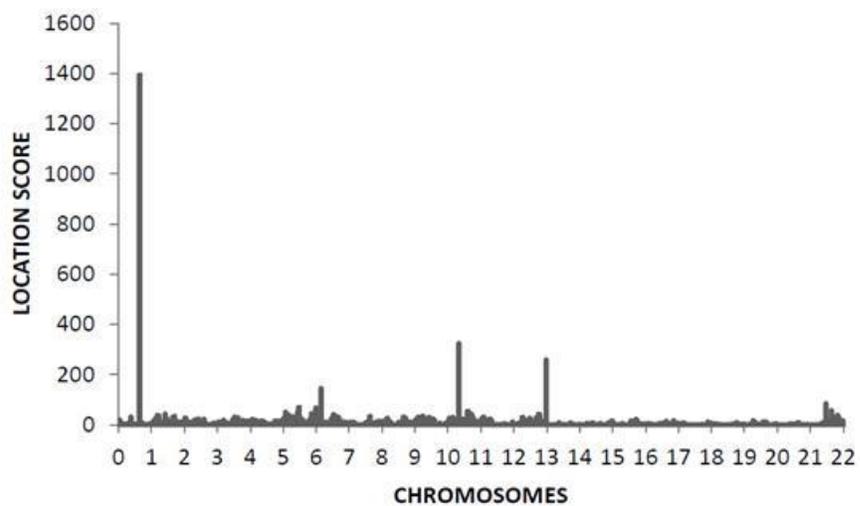


Figure 3.4.2 Autozygosity mapping generates highly significant homozygous regions.

The x- and y-axes show autosomal chromosomes and location scores, respectively.

3.4.3 Compound homozygous *EXT2* variants in patients with SSM syndrome

A non-synonymous, rare variant analysis was applied on exome sequencing data to identify candidate genetic variants within these isolated segments as well as an indiscriminate exome-wide approach. Both analyses yielded similar results and identified two homozygous variants in exostosin 2, *EXT2*: c.260T>G, p.Met87Arg; and c.283C>T, p.Arg95Cys (Figure 3.4.3A). Both variants were consistently predicted to be ‘disease-causing’ by multiple *in silico* software programs and both are near but upstream of the highly conserved exostosin domain (Figure 3.4.3B). Both wild type residues are also well-conserved when aligned against a set of diverged species from the animal kingdom (Figure 3.4.3C).

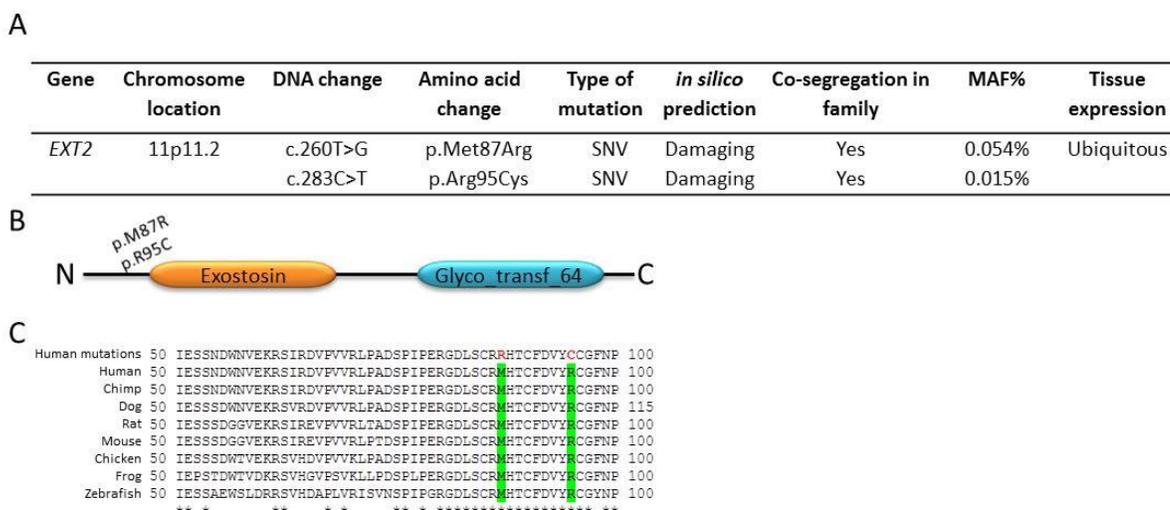


Figure 3.4.3 Schematic of the genetic and bioinformatic studies in the SSM family.

(A) Annotated variants in *EXT2* identified via exome sequencing. *In silico* predictions were determined by gathering outputs of multiple programs. Co-segregation in family was determined via individual allele genotyping. Minor allele frequencies were determined using multiple publically available databases. Tissue expression was determined using The Human Protein Atlas and GeneCards. (B) The structure of the *EXT2* protein is comprised of the exostosin domain and the glucose transferase 64 domain, shown from N- to C-terminal. The variants identified are shown. (C) Multiple alignments demonstrate strong conservation of both wild type amino acid residues, p.Met87 and p.Arg95 across a set of species-specific *EXT2* homologs. A ClustalW analysis of the *EXT2* region encompassing the variant sites (highlighted in red) in aligned homologs with multiple divergent species is shown (in green). The asterisks below indicate fully conserved residues.

3.4.4 *EXT2* variants segregate in SSM syndrome family

Interestingly, the variants co-segregated within the family in an autosomal recessive, compound homozygous inheritance pattern with both variants transmitted to each affected individual from each parent (Figure 3.4.4A). Specifically, both parents were double heterozygotes for both variants, which were on the same chromosome (i.e. in cis) and all affected individuals were homozygous for both variants (Figure 3.4.4A and Figure 3.4.4B). The unaffected sister was homozygous wild type for both variants.

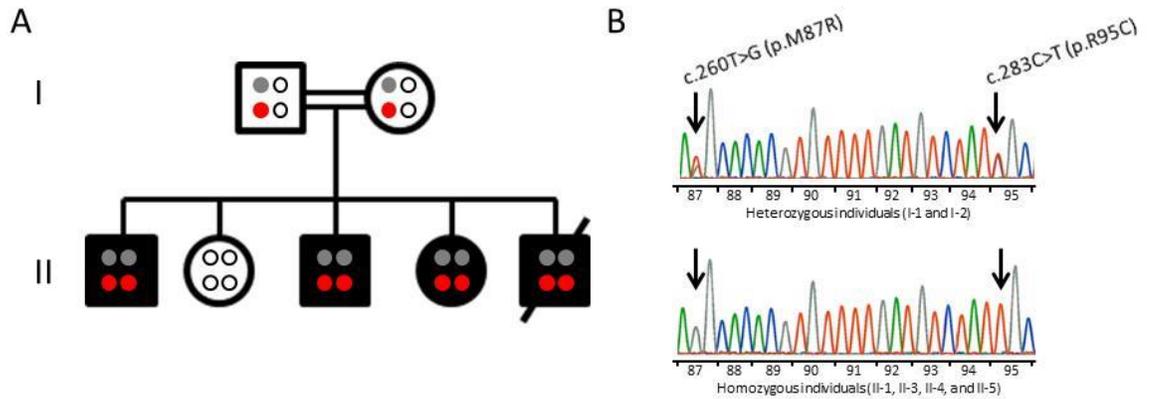


Figure 3.4.4 *EXT2* variants segregate with affected individuals in the family.

(A) Within the pedigree, a grey circle represents one p.Met87Arg allele; a red circle represents one p.Arg95Cys allele; empty circles represent wild type alleles. (B) Electropherograms showing DNA sequence analysis of *EXT2* from genomic DNA of heterozygous and homozygous individuals. The positions of the genetic variants are shown.

3.4.5 Population screening reveals *EXT2* variants are ultra-rare

To identify the local carrier frequency in the Old Order Mennonite community, 39 individuals were subjected to genotyping for both *EXT2* variants with three p.Met87Arg heterozygotes identified, yielding a local carrier frequency of 3.85%. We did not observe any homozygous or heterozygous p.Arg95Cys carriers. No double heterozygotes for p.Met87Arg and p.Arg95Cys were identified. We screened an additional 311 healthy Caucasian controls with no carriers identified for either *EXT2* variant. According to the NHLBI ESP Exome Variant Server, the MAF of p.Met87Arg and p.Arg95Cys in 6500 individuals is 0.054% and 0.015%, respectively (Figure 3.4.3A).

3.4.6 Decreased *EXT2* expression and activity in patients with SSM syndrome

To determine whether the newly identified variants affect protein and transcript expression, we measured *EXT2* expression levels via immunoblotting and RT-PCR, respectively in patient fibroblast-derived cells. We observed diminished *EXT2* protein expression relative to a healthy *EXT2*-WT control ($P < 0.001$) (Figure 3.4.6A). Similarly, we observed a modest but significantly lower difference in *EXT2* transcript levels in patient cells compared to healthy controls ($P < 0.05$) (Figure 3.4.6B). Previous studies have shown that *EXT2* interacting protein, NDST1 is affected in *EXT2* knockdown cell lines (Presto et al., 2008). Consequently, we explored whether the function of *EXT2* is compromised in patient cells by measuring the expression of binding partner NDST1. Our findings indicate that cellular NDST1 protein levels are abolished in patient whole cell lysates (Figure 3.4.6C). Conversely and not surprisingly, *NDST1* transcript expression is not affected in patient cells (Figure 3.4.6D) in agreement with previous studies, suggesting the absence of NDST1 protein expression is a product of an error in post-translational modification (Presto et al., 2008).

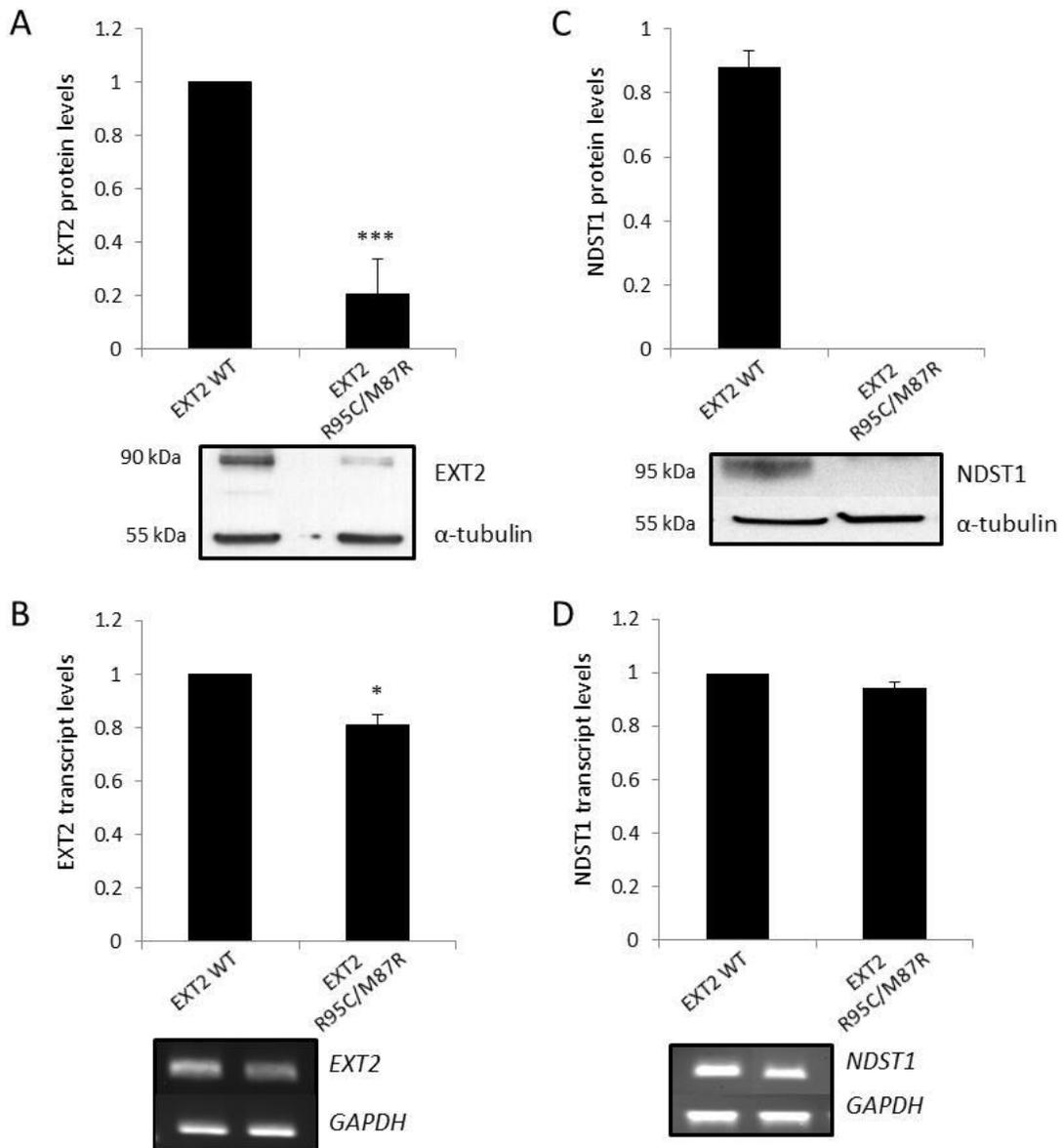


Figure 3.4.6 Decreased EXT2 expression and activity in patients with SSM syndrome.

(A) EXT2 protein expression. (B) *EXT2* transcript expression. (C) NDST1 protein expression. (D) *NDST1* transcript expression. Protein and RNA were isolated from patient cells and a healthy, wild type control. Bars represent mean \pm standard error. All expression targets were normalized with their respective constitutively active gene or gene product (α -tubulin and *GAPDH* for protein and transcript expression, respectively). *represents P -value <0.05 ; **represents P -value <0.01 ; ***represents P -value <0.001 .

3.4.7 Both EXT2 variants are involved in the development of SSM syndrome

Since two homozygous variants (both in the same gene) were identified, we sought to identify which of the two (or both) variants is (are) responsible for the decreased protein expression in the patients. We created three EXT2 constructs: 1) EXT2 p.Arg95Cys; 2) EXT2 p.Met87Arg; and 3) EXT2 p.Arg95Cys and p.Met87Arg. Using anti-EXT2, we measured EXT2 protein expression between endogenous and exogenous wild type EXT2 in two cell lines as an experimental control and observed no significant difference providing confidence in our use of exogenous wild type EXT2 as a reference point (Figures 3.4.7B and Figure 3.4.7D). We observed statistically significant decreased protein levels in all three constructs relative to wild type exogenous and endogenous EXT2 but more noticeably when both residues were mutated (Figures 3.4.7A-D). Interestingly however, the EXT2 p.Arg95Cys variant seems to have a significantly larger effect on protein expression than EXT2 p.Met87Arg ($P < 0.05$) (Figures 3.4.7A-D). The experiments were replicated multiple times (total $n=20$) with two different cell lines and two different antibodies and all produced similar results. These data suggest that the EXT2 variants synergistically but unequally, affect EXT2 protein levels.

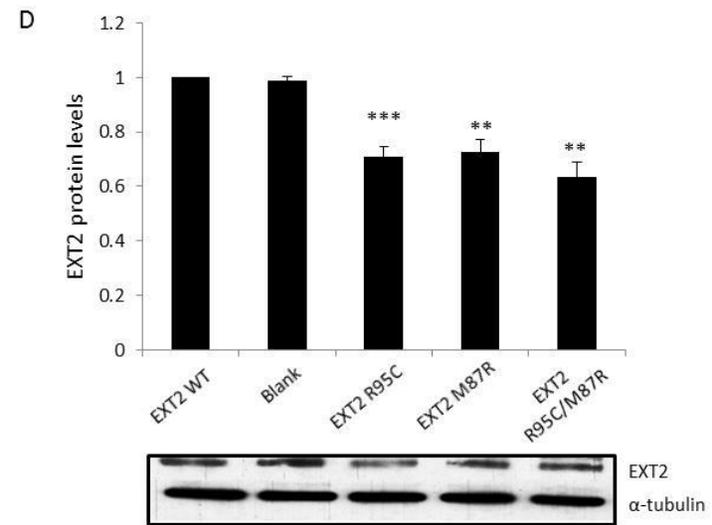
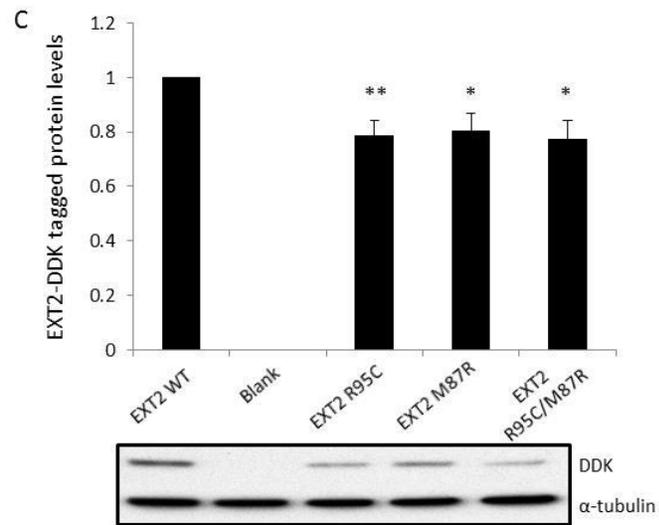
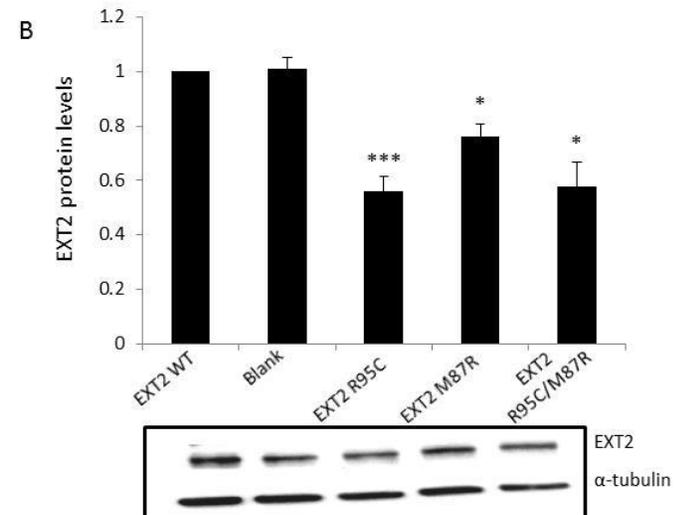
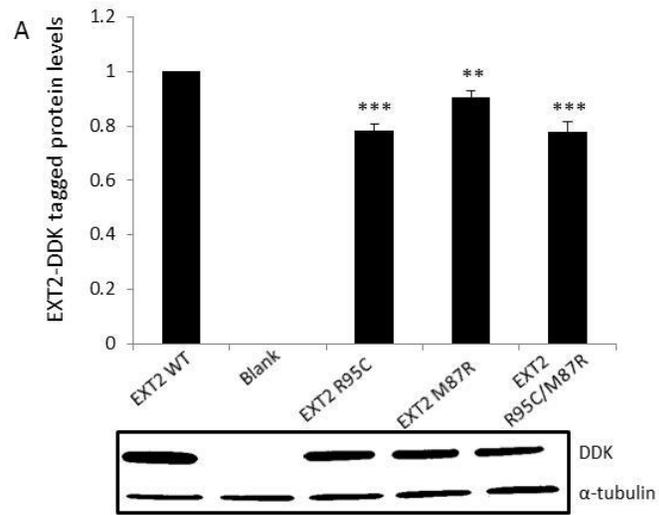


Figure 3.4.7 Both EXT2 variants are necessary for the development of SSM syndrome.

(A) EXT2 protein expression in COS-7 cells using anti-DDK-EXT2. (B) EXT2 protein expression in COS-7 cells using anti-EXT2. (C) EXT2 protein expression in HEK293 cells using anti-DDK-EXT2. (D) EXT2 protein expression in HEK293 cells using anti-EXT2. Protein expression was isolated from four different EXT2 constructs and a negative control (blank). Bars represent mean \pm standard error. All expression assays were normalized with α -tubulin. *represents P -value <0.05 ; **represents P -value <0.01 ; ***represents P -value <0.001 .

3.5 Discussion

Glycosylation is an essential, ubiquitous, and evolutionarily conserved process required for cellular function and development (Nadanaka and Kitagawa, 2008; Shevelev et al., 2012). Not surprisingly, defects in genes encoding heparan sulfate biosynthesis enzymes can lead to a wide spectrum of human disorders including bone abnormalities, cancer, metabolic disorders, and even late-onset neurodegeneration (Nadanaka and Kitagawa, 2008; Maeda et al., 2011; Stickens et al., 2000; Ropero et al., 2004; van Horssen et al., 2003). In this Chapter, we describe a family with children affected with a novel heparan sulfate related disorder designated as SSM syndrome. By coupling autozygosity mapping and exome sequencing, we identified two homozygous variants in *EXT2* that segregated in a distinctive, compound homozygous inheritance pattern in the family. *EXT2* p.Met87Arg and p.Arg95Cys are extremely rare variants, harbored in a highly conserved region within the protein, and as expected, given the drastic change in the amino acid properties, were consistently predicted to be ‘disease-causing’ by multiple *in silico* programs. We were able to assess the expression and function of *EXT2* *ex vivo* in patient fibroblasts and in doing so, we observed decreased *EXT2* expression and activity. In addition, we have shown using an *in vitro* system, that both variants contribute to decreased *EXT2* expression potentially destabilizing the protein, disrupting heparan sulfate biosynthesis, and probably other developmental cascades requiring *EXT2* for propagation, such as FGF and Wnt signalling, as was previously demonstrated in *Ext2*-knockdown studies (Fischer et al., 2011; Norton et al., 2005; Nurcombe et al., 1993). Collectively, our findings have implicated a well-characterized gene in a new developmental disorder and have further exemplified the spectrum of phenotypes that can arise due to errors in glycosylation.

Herein, we have also catalogued all reported heparan sulfate-related diseases due to defects in heparan sulfate genes (Table 3.5.1). In doing so, we have identified 26 genes that when mutated, can cause developmental disorders consisting of intellectual disability, distinct coarse facial features, and seizures—features linked to errors in glycosylation pathways. This exercise helped us distinguish the clinical features observed in all heparan sulfate-related diseases from those unique to SSM syndrome patients. Of

particular interest to us was a recently reported NDST1-associated intellectual disability in a consanguineous family, as NDST1 directly interacts with EXT2, and our findings show abolished NDST1 protein expression in affected individuals (Presto et al., 2008; Reuter et al., 2014).

Table 3.5.1 Hereditary diseases with abnormal heparan sulfate levels.

Gene	Chromosomal location	Protein	Disease	Predominant clinical features	OMIM (locus, phenotype)
<i>B3GALT6</i>	1p36.33	UDP-Gal: beta-gal beta-1,3-galactosyltransferase polypeptide 6	(1) Ehlers-Danlos syndrome, progeroid type 2 (2) Spondyloepimetaphyseal dysplasia with joint laxity, type 1	- Aged appearance, developmental delay, short stature, craniofacial disproportion, generalized osteopenia, hypotonic muscles, and loose skin. - Vertebral abnormalities leading to severe kyphoscoliosis, thoracic asymmetry, and respiratory failure.	615291, 615349, 271640
<i>B3GAT3</i>	11q12.3	Beta-1,3-glucuronyltransferase 3	Multiple joint dislocations, short stature, craniofacial dysmorphism, and congenital heart defects	- Dysmorphic facies, bilateral dislocations of the elbows, hips, and knees, short stature, and cardiovascular defects.	606374, 245600
<i>B4GALT7</i>	5q35.3	Xylosylprotein 4-beta-galactosyltransferase, polypeptide 7	Ehlers-Danlos syndrome, progeroid type 1	- Aged appearance, developmental delay, short stature, craniofacial disproportion, osteopenia, hypotonia, and loose skin.	604327, 130070
<i>COL4A2</i>	13q34	Collagen, type IV, alpha-2	Porencephaly 2	- Delayed early development, poor motor skills, and seizures.	120090, 614483
<i>COL4A3</i>	2q36.3	Collagen, type IV, alpha-3	Alport syndrome, autosomal dominant, autosomal recessive	- End-stage renal failure and hearing loss.	120070, 104200, 203780
<i>DAG1</i>	3p21.31	Dystrophin-associated glycoprotein 1	Muscular dystrophy-dystroglycanopathy (limb-girdle), type C, 9	- Muscular dystrophy, mental retardation, and delayed motor development.	128239, 613818
<i>EXT1</i>	8q24.11	Exostosin glycosyltransferase 1	(1) Hereditary multiple exostoses, type 1 (2) Chondrosarcoma	- Exostoses arising from long bones.	608177, 133700, 215300
<i>EXT2</i>	11q11.2	Exostin glycosyltransferase 2	Hereditary multiple exostoses, type 2	- Exostoses arising from long bones.	608210, 133701
<i>FGF16</i>	Xq21.1	Fibroblast growth factor 16	Metacarpal 4-5 fusion	- Hand and foot malformation.	300827, 309630
<i>GNS</i>	12q14.3	N-acetylglucosamine-6-sulfatase	Mucopolysaccharidosis type IIID	- Mental retardation, coarse facies and hirsutism, developmental delay, and excessive HS in urine.	607664, 252940
<i>GPC3</i>	Xq26.2	Glypican 3	(1) Simpson-Golabi-Behmel syndrome, type 1 (2) Wilms tumor 1	- Pre- and post-natal overgrowth, coarse facies, and congenital heart defect. Stocky appearance with normal intelligence. - Childhood tumors from malignant transformation of renal stem cells.	300037, 312870, 194070

<i>GPC6</i>	13q31.3	Glypican 6	Omodysplasia 1	- Severe congenital micromelia with prominent forehead, frontal bossing. Variable findings include cryptorchidism, hernias, congenital heart defects, and cognitive delay.	604404, 258315
<i>GUSB</i>	7q11.21	Beta-glycuronidase	Mucopolysaccharidosis VII	- Hepatomegaly, skeletal anomalies, coarse facies, and variable degrees of mental impairment.	611499, 253220
<i>HGSNAT</i>	8p11.21	Heparan-alpha-glucosaminide N-acetyltransferase	Mucopolysaccharidosis type IIIC	- Dysmorphic features including coarse facies, hypertelorism, low-set ears, depressed nasal bridge, and coarse hair. Skeletal examination showed high lumbar vertebral bodies, delayed motor development, iliac flaring. Urinary HS was increased.	610453, 252930
<i>HS6ST1</i>	2q14.3	Heparan sulfate 6-O-sulfotransferase 1	Hypogonadotropic hypogonadism 15 with/without anosmia	- Absent or incomplete sexual maturation, other features include cleft palate, sensorineural hearing loss.	604846, 614880
<i>HSPG2</i>	1p36.12	Heparan sulfate proteoglycan of basement membrane	(1) Dyssegmental dysplasia, Silverman-Handmaker type (2) Schwartz-Jampel syndrome, type 1	- Lethal forms of neonatal short limbed dwarfism, anisospondyly. -Myotonia, skeletal dysplasia, reduced stature, kyphoscoliosis, irregular epiphyses.	142461, 224410, 255800
<i>IDS</i>	Xq28	Iduronate 2-sulfatase	Mucopolysaccharidosis II	- Multisystem disorder: airway obstruction, skeletal deformities, cardiomyopathy, and neurological decline.	300823, 309900
<i>IDUA</i>	4p16.3	Alpha-Iduronidase	(1) Mucopolysaccharidosis 1h (2) Mucopolysaccharidosis 1h/s (3) Mucopolysaccharidosis 1s	- Variable phenotypes but generally all 3 disorders encompass coarse facies, corneal clouding, mental retardation, hernias, dysostosis multiplex, and hepatosplenomegaly.	252800, 607014, 607015, 607016
<i>IHH</i>	2q35	Indian hedgehog	Acrocapitofemoral dysplasia	- Short stature with postnatal onset, brachydactyly	600726, 607778
<i>KALI</i>	Xp22.31	KAL1 gene	Kallmann syndrome 1 (Hypogonadotropic hypogonadism 1 with/without anosmia)	- Absent or incomplete sexual maturation. Patients also show anosmia.	300836, 308700
<i>MANBA</i>	4q24	Mannosidase, beta A, lysosomal	Mannosidosis, beta	- Developmental delay and mental retardation.	609489, 248510
<i>NAGLU</i>	17q21.2	N-Acetylglucosaminidase	Mucopolysaccharidosis type IIIB	- Progressive neurodegeneration, behavioral problems, mild skeletal changes, and shorted life span due to accumulation of HS.	609701, 252920
<i>NDST1</i>	5q33.1	N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 1	Autosomal recessive intellectual disability	- Intellectual disability, muscular hypotonia, epilepsy, and postnatal growth deficiency.	600853, PMID: 25125150
<i>SERPINC1</i>	1q25.1	Serpin peptidase inhibitor, clade C, member 1	Thrombophilia due to antithrombin III deficiency	- Venous thrombosis.	107300, 613118
<i>SGSH</i>	17q25.3	N-sulfoglucosamine sulfohydrolase	Mucopolysaccharidosis type IIIA	- CNS degeneration, severe neurological degeneration, due to impaired degradation of HS.	605270, 252900

<i>SMPDI</i>	11p15.4	Sphingomyelin phosphodiesterase 1, acid lysosomal	(1) Niemann-Pick disease, type A (2) Niemann-Pick disease, type B	- Severe infantile form with neurologic degeneration resulting in death by 3 years of age (type A) to later-onset nonneurological form (type B).	607608, 257200, 607616
--------------	---------	---	--	--	---------------------------

Notably however, the clinical features of patients with SSM syndrome are distinct from patients with hereditary multiple exostoses (HME, OMIM 133701), an autosomal dominant disorder characterized by benign, cartilage-capped tumours or “exostoses” localized to the growth plate of long bones (Busse-Wicher et al., 2014; Ryckx et al., 2013). Typically, HME has an onset of 12 years of age and is caused by mutations in *EXT1* or *EXT2*, with the majority of cases linked to *EXT1* (Busse-Wicher et al., 2014; Huegel et al., 2013). None of the affected individuals in our family has evidence of exostoses, nor do either of the parents who are heterozygous for both p.Met87Arg and p.Arg95Cys in *EXT2*. To the best of our knowledge, there have not been any reports of autosomal recessive HME. Interestingly, there has only been one report of HME and seizures in a patient with a mutation in *EXT2* (Stancheva-Ivanova et al., 2011). There have been multiple reports of patients with multiple exostoses, intellectual disability, and seizures, diagnosed with Potocki-Shaffer Syndrome (PSS, OMIM 601224), a rare contiguous gene deletion syndrome due to haploinsufficiency of the 11p11.2 region, encompassing *EXT2* (Kim et al., 2012; Palka et al., 2012). In addition, partial loss of *EXT2* sister protein, *EXT1* and *TRPS1*, its neighboring gene, can give rise to another developmental disorder known as Langer-Giedion syndrome (OMIM 150230), which is characterized by multiple cartilaginous exostoses, dysmorphic facial appearance, and mild intellectual disability (Kozlowski et al., 1977; Ludecke et al., 1995). Together, our results supplemented with previous findings, suggest the phenotype is dependent on the location of the *EXT2* (or *EXT1*) mutation, the consequential enzyme dosage, and whether adjacent genes are affected.

The EXT protein family has also been well characterized in many animal systems including zebrafish, flies, nematodes, and mice. Previous studies have shown Ext1- and Ext2-deficient mutant mouse embryos do not form proper heparan sulfate chains, fail to gastrulate, and ultimately, die at embryonic day 8.5 (Lin et al., 2000; Shimokawa et al., 2011; Stickens et al., 2005). In contrast, some embryos with low expression of the enzymes survive longer as the enzymes can produce shorter but, normal heparan sulfate chains (Kozziel et al., 2004; Yamada et al., 2004). This is due to normal expression of the other EXT protein, which still has glycosyltransferase activity although significantly less than the amount generated by the EXT1 and EXT2 complex as the enzymes cannot

substitute for one another during co-polymerization (Busse et al., 2007; Busse and Kusche-Gullberg, 2003; Lind et al., 1998; McCormick et al., 2000). Therefore, while residual levels of heparan sulfate facilitate some cellular functions, they are not sufficient for proper neural development including brain morphology and midline axon guidance—processes defective in *Ext*-mutants (Maeda et al., 2011; Shevelev et al., 2012). Collectively, these studies have demonstrated the importance of sufficient and spatiotemporal expression of heparan sulfate biosynthesis enzymes in the development of brain size and structure. Importantly, in the context of SSM syndrome patients, these studies help explain why SSM patients are alive but have macrocephaly and are severely delayed in their developmental course.

3.6 Conclusion

In conclusion, we have presented herein the first account of an autosomal recessive EXT2-mediated human disease designated as seizures-scoliosis-macrocephaly (SSM) syndrome. By coupling autozygosity mapping and exome sequencing, we identified two homozygous variants in heparan sulfate biosynthesis enzyme, EXT2, which is only partially expressed and active in patient cells. Our results have illustrated and further confirmed the importance of sufficient expression of heparan sulfate biosynthesis enzymes in human development. These results have also demonstrated the need for heparan sulfate biosynthesis enzyme replacement therapy. As the majority of heparan sulfate disorders are primarily due to a deficiency in heparan sulfate, a simple therapeutic intervention would be to elevate heparan sulfate levels. However, it would be naïve to administer heparan sulfate without knowledge of the patient's genotype as some features of heparan sulfate-related diseases overlap with many other metabolic disorders and it is not clear if excess heparan sulfate is toxic to cells. Alternatively, early genetic diagnosis of a heparan sulfate deficiency potentially *in utero*, coupled with a heparan sulfate production stimulant via enzyme or gene replacement therapy at the proper developmental period, could be a promising therapeutic intervention.

3.7 References

- Ahn, J., Ludecke, H.J., Lindow, S., Horton, W.A., Lee, B., Wagner, M.J., Horsthemke, B., and Wells, D.E. (1995). Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nat Genet* *11*, 137-143.
- Astudillo, P., and Larrain, J. (2014). Wnt signaling and cell-matrix adhesion. *Curr Mol Med* *14*, 209-220.
- Bandtlow, C.E., and Zimmermann, D.R. (2000). Proteoglycans in the developing brain: new conceptual insights for old proteins. *Physiol Rev* *80*, 1267-1290.
- Bishop, J.R., Schuksz, M., and Esko, J.D. (2007). Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* *446*, 1030-1037.
- Busse, M., Feta, A., Presto, J., Wilen, M., Gronning, M., Kjellen, L., and Kusche-Gullberg, M. (2007). Contribution of EXT1, EXT2, and EXTL3 to heparan sulfate chain elongation. *J Biol Chem* *282*, 32802-32810.
- Busse, M., and Kusche-Gullberg, M. (2003). In vitro polymerization of heparan sulfate backbone by the EXT proteins. *J Biol Chem* *278*, 41333-41337.
- Busse-Wicher, M., Wicher, K.B., and Kusche-Gullberg, M. (2014). The exostosin family: proteins with many functions. *Matrix Biol* *35*, 25-33.
- Farhan, S.M., Wang, J., Robinson, J.F., Prasad, A.N., Rupar, C.A., Siu, V.M., Consortium, F.C., and Hegele, R.A. (2015). Old gene, new phenotype: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses. *J Med Genet* *52*, 666-675.
- Fischer, S., Filipek-Gorniok, B., and Ledin, J. (2011). Zebrafish Ext2 is necessary for Fgf and Wnt signaling, but not for Hh signaling. *BMC Dev Biol* *11*, 53.
- Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., *et al.* (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* *493*, 216-220.
- Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* *491*, 56-65.
- Huegel, J., Sgariglia, F., Enomoto-Iwamoto, M., Koyama, E., Dormans, J.P., and Pacifici, M. (2013). Heparan sulfate in skeletal development, growth, and pathology: the case of hereditary multiple exostoses. *Dev Dyn* *242*, 1021-1032.
- Kim, H.G., Kim, H.T., Leach, N.T., Lan, F., Ullmann, R., Silahtaroglu, A., Kurth, I., Nowka, A., Seong, I.S., Shen, Y., *et al.* (2012). Translocations disrupting PHF21A in the

- Potocki-Shaffer-syndrome region are associated with intellectual disability and craniofacial anomalies. *Am J Hum Genet* 91, 56-72.
- Kozziel, L., Kunath, M., Kelly, O.G., and Vortkamp, A. (2004). Ext1-dependent heparan sulfate regulates the range of Ihh signaling during endochondral ossification. *Dev Cell* 6, 801-813.
- Kozlowski, K., Harrington, G., Barylak, A., and Bartoszewica, B. (1977). Multiple exostoses-mental retardation syndrome (Ale-Calo or M.E.M.R. syndrome). *Clin Pediatr (Phila)* 16, 219-224.
- Kreuger, J., Spillmann, D., Li, J.P., and Lindahl, U. (2006). Interactions between heparan sulfate and proteins: the concept of specificity. *J Cell Biol* 174, 323-327.
- Lin, X., Wei, G., Shi, Z., Dryer, L., Esko, J.D., Wells, D.E., and Matzuk, M.M. (2000). Disruption of gastrulation and heparan sulfate biosynthesis in EXT1-deficient mice. *Dev Biol* 224, 299-311.
- Lind, T., Tufaro, F., McCormick, C., Lindahl, U., and Lidholt, K. (1998). The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J Biol Chem* 273, 26265-26268.
- Ludecke, H.J., Wagner, M.J., Nardmann, J., La Pillo, B., Parrish, J.E., Willems, P.J., Haan, E.A., Frydman, M., Hamers, G.J., Wells, D.E., *et al.* (1995). Molecular dissection of a contiguous gene syndrome: localization of the genes involved in the Langer-Giedion syndrome. *Hum Mol Genet* 4, 31-36.
- Maeda, N., Ishii, M., Nishimura, K., and Kamimura, K. (2011). Functions of chondroitin sulfate and heparan sulfate in the developing brain. *Neurochem Res* 36, 1228-1240.
- McCormick, C., Duncan, G., Goutsos, K.T., and Tufaro, F. (2000). The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proc Natl Acad Sci U S A* 97, 668-673.
- Nadanaka, S., and Kitagawa, H. (2008). Heparan sulphate biosynthesis and disease. *J Biochem* 144, 7-14.
- Norton, W.H., Ledin, J., Grandel, H., and Neumann, C.J. (2005). HSPG synthesis by zebrafish Ext2 and Extl3 is required for Fgf10 signalling during limb development. *Development* 132, 4963-4973.
- Nurcombe, V., Ford, M.D., Wildschut, J.A., and Bartlett, P.F. (1993). Developmental regulation of neural response to FGF-1 and FGF-2 by heparan sulfate proteoglycan. *Science* 260, 103-106.

- Palka, C., Alfonsi, M., Mohn, A., Guanciali Franchi, P., Chiarelli, F., and Calabrese, G. (2012). Delayed Diagnosis of Potocki-Shaffer Syndrome in a Woman with Multiple Exostoses and Mental Retardation. *Mol Syndromol* 2, 259-261.
- Perrimon, N., and Bernfield, M. (2000). Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* 404, 725-728.
- Presto, J., Thuveson, M., Carlsson, P., Busse, M., Wilen, M., Eriksson, I., Kusche-Gullberg, M., and Kjellen, L. (2008). Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation. *Proc Natl Acad Sci U S A* 105, 4751-4756.
- Reuter, M.S., Musante, L., Hu, H., Diederich, S., Sticht, H., Ekici, A.B., Uebe, S., Wienker, T.F., Bartsch, O., Zechner, U., *et al.* (2014). NDST1 missense mutations in autosomal recessive intellectual disability. *Am J Med Genet A* 164A, 2753-2763.
- Ryckx, A., Somers, J.F., and Allaert, L. (2013). Hereditary multiple exostosis. *Acta Orthop Belg* 79, 597-607.
- Shevelev, O.B., Rykova, V.I., Fedoseeva, L.A., Leberfarb, E.Y., Dymshits, G.M., and Kolosova, N.G. (2012). Expression of Ext1, Ext2, and heparanase genes in brain of senescence-accelerated OXYS rats in early ontogenesis and during development of neurodegenerative changes. *Biochemistry (Mosc)* 77, 56-61.
- Shimokawa, K., Kimura-Yoshida, C., Nagai, N., Mukai, K., Matsubara, K., Watanabe, H., Matsuda, Y., Mochida, K., and Matsuo, I. (2011). Cell surface heparan sulfate chains regulate local reception of FGF signaling in the mouse embryo. *Dev Cell* 21, 257-272.
- Stancheva-Ivanova, M.K., Wuyts, W., van Hul, E., Radeva, B.I., Vazharova, R.V., Sokolov, T.P., Vladimirov, B.Y., Apostolova, M.D., and Kremensky, I.M. (2011). Clinical and molecular studies of EXT1/EXT2 in Bulgaria. *J Inherit Metab Dis* 34, 917-921.
- Stickens, D., Clines, G., Burbee, D., Ramos, P., Thomas, S., Hogue, D., Hecht, J.T., Lovett, M., and Evans, G.A. (1996). The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. *Nat Genet* 14, 25-32.
- Stickens, D., Zak, B.M., Rougier, N., Esko, J.D., and Werb, Z. (2005). Mice deficient in Ext2 lack heparan sulfate and develop exostoses. *Development* 132, 5055-5068.
- Venero Galanternik, M., Kramer, K.L., and Piotrowski, T. (2015). Heparan Sulfate Proteoglycans Regulate Fgf Signaling and Cell Polarity during Collective Cell Migration. *Cell Rep*.
- Yamada, S., Busse, M., Ueno, M., Kelly, O.G., Skarnes, W.C., Sugahara, K., and Kusche-Gullberg, M. (2004). Embryonic fibroblasts with a gene trap mutation in Ext1 produce short heparan sulfate chains. *J Biol Chem* 279, 32134-32141.

Chapter 4 - Known phenotype, known gene: linkage analysis and exome sequencing identify a novel mutation in *KCTD7* in patients with progressive myoclonus epilepsy with ataxia.

4 Study rationale

Three females initially diagnosed with idiopathic epilepsy with ataxia, were born to a Low-German-Speaking Mexican Mennonite family originally from Durango, Mexico but now all reside in southwestern Ontario. All three females have a history of early onset seizures. They later became ataxic with intention tremor. There are no growth or metabolic abnormalities. The family agreed to participate in genetic testing, and the disorder, which was originally referred to as ‘epilepsy with ataxia’, became another project funded by the FORGE Canada Consortium.

This project provides an excellent example of the genetics results successfully guiding clinicians to correctly diagnose the patients. Once we identified g.661041A>G, c.827A>G, p.Tyr276Cys in *KCTD7*, a known progressive myoclonus epilepsy gene, in the affected individuals, the clinicians reviewed our results, which complemented previous genetic reports of *KCTD7*-mediated progressive myoclonus epilepsy and together, agreed to diagnose the three females with ‘progressive myoclonus epilepsy with ataxia’.

Chapter 4 was published as an independent study: “Linkage analysis and exome sequencing identify a novel mutation in *KCTD7* in patients with progressive myoclonus epilepsy with ataxia” in *Epilepsia* in 2014 (Farhan et al., 2014; PMID: 25060828). Furthermore, the results of Chapter 4 were described and compared to other epilepsy subtypes in “Exploring the Epileptome I: Genetics of Age dependent Epileptic Encephalopathies” in *Pediatric Epilepsy*, 4th edition, Chapter 7 (Farhan and Prasad, 2016, textbook in production).

4.1 Overview

Epilepsy affects approximately 1% of the world's population. Genetic factors and acquired etiologies, as well as a range of environmental triggers, together contribute to epileptogenesis. We have ascertained a family with three daughters affected with progressive myoclonus epilepsy with ataxia. Clinical details of the onset and progression of the neurological presentation, epileptic seizures, and the natural history of progression over a 10-year period are described. Using autozygosity genetic mapping, we identified a highly significant homozygous region on chromosome 7p12.1-7q11.22. We subsequently applied exome sequencing and used a rare, non-synonymous variant prioritization analysis within the homozygous region. We identified g.661041A>G, c.827A>G, p.Tyr276Cys in the potassium channel tetramerization domain-containing 7 gene, *KCTD7*, which is predominantly expressed in the brain. Mutations in *KCTD7* have been previously implicated in epileptic phenotypes due to disturbances in potassium channel conductance. Pathogenicity of the mutation was supported by *in silico* analyses and *KCTD7* p.Tyr276Cys homozygosity segregating with affected individuals only.

4.2 Introduction

Approximately 70 million people worldwide are living with some form of epilepsy, a multifactorial and chronic neurological condition characterized by recurrent seizures due to abnormal excessive or synchronous neuronal activity in the brain (Fisher et al., 2005). Epilepsy is multifactorial in its causation and its chronic nature results in a considerable economic burden estimated at \$100 million per year in North America (Tator et al., 2007). The major expense however, approximately \$700 million, is not directly a result of epilepsy care but from lost productivity (Katchanov and Birbeck, 2012; Newton and Garcia, 2012; Tator et al., 2007). Common forms of epilepsy tend to be complex disorders in which a number of implicated genes and environmental factors together contribute to epileptogenesis (Buono, 2013). Based on this model, the inheritance of multiple susceptibility alleles coupled with high risk environmental conditions can result

in the eventual manifestation of epileptic seizures (Buono, 2013; de Araujo Furtado et al., 2012). Conversely, some relatively rare forms of epilepsy are monogenic, where a mutation in a single gene encodes a dysfunctional protein that compromises a key pathway and produces the disease phenotype (Deng et al., 2013). While such molecular defects are individually rare, collectively they account for a large proportion of familial epilepsies (Deng et al., 2013).

Recently, next generation sequencing technologies have successfully identified the molecular genetic basis of nearly 200 disorders since 2009 (Boycott et al., 2013). Briefly, next generation sequencing is a rapid and cost-effective sequencing approach that can be applied to study any genetic disease and is able to detect most types of genomic variation and in many cases, the casual variant in the affected patient or family (Boycott et al., 2013). One subtype of next generation sequencing, exome sequencing, sequences only protein-coding variants and has been a particularly robust diagnostic tool for monogenic disorders (Boycott et al., 2013).

Herein, we describe a rare neurological disease designated as epilepsy with ataxia, with pediatric onset and an autosomal recessive mode of inheritance. Epilepsy with ataxia was first observed in three of eight children of a non-consanguineous Low-German-Speaking Mexican Mennonite couple and is characterized primarily by tonic-clonic and myoclonic seizures, ataxia, and cognitive and language delay. We used genomic tools including homozygosity genetic mapping and exome sequencing to identify a novel, rare variant, p.Tyr276Cys in KCTD7, an epilepsy-associated gene, as the probable genetic cause of epilepsy with ataxia.

4.3 Materials and methods

4.3.1 Ethics

As described in Chapter 2, subsection 2.3.1, ethics approval was obtained from the parents, the three affected females, and four unaffected siblings.

4.3.2 Patients and biological materials

Three children of a healthy Low-German-Speaking Mexican Mennonite couple were clinically ascertained because they were experiencing epileptic seizures and ataxia. Blood samples were collected from all three affected daughters, four unaffected siblings, and the two parents.

4.3.3 DNA isolation

DNA was isolated from all nine family members as described in Chapter 2, subsection 2.3.3.

4.3.4 Genotyping

Genomic DNA (gDNA) extracted from all family members was genotyped for single nucleotide polymorphisms (SNPs) using Affymetrix[®] protocols as described in Chapter 2, subsection 2.3.6.

4.3.5 Autozygosity mapping

GeneSpring GT v2.0 software was used to identify regions of homozygosity that are identical by descent in family members as described in Chapter 2, subsection 2.3.7.

4.3.6 Exome sequencing

Exome sequencing was performed using gDNA from two affected individuals (II-6 and II-7) using the protocols described in Chapter 2, subsection 2.3.8.

4.3.7 Sequence alignment

Sequence alignment was performed using the same procedures described in Chapter 2, subsection 2.3.9.

4.3.8 Variant calling and annotation

Variant calling and annotation were performed using the same procedures described in Chapter 2, subsection 2.3.10.

4.3.9 Variant discovery

4.3.9.1 Prioritization of homozygous, non-synonymous, and rare variants

We investigated whether there were any non-synonymous, rare variants within the homozygous regions on chromosomes 7q11.21. We also applied an unbiased, exome-wide approach to identify any homozygous, non-synonymous, and rare variants as described in Chapter 2, subsection 2.3.11.1. We scanned for variants in genes known to cause epileptic seizures or ataxia based on literature and OMIM, as well as all genes.

4.3.9.2 *In silico* analyses

To determine the predicted biological effect of non-synonymous variants on protein function, all non-synonymous variants identified as candidate variants (KCTD7 p.Tyr276Cys) were assessed using the same tools described in Chapter 2, subsection 2.3.11.2.

4.3.10 Variant validation

To determine whether p.Tyr276Cys in KCTD7 segregated with disease status in the pedigree, we genotyped family members using standard Sanger sequencing methods as described in Chapter 2, subsection 2.3.12.1 and 2.3.12.2.

4.3.10.1 PCR

Forward and reverse primers specific to amplify g.661041A>G in *KCTD7* were: 5'-TATGACCTGCTGCACTGCCT-3' and 5'-CTTCAGGGATCTTACTCCC-3', respectively. The PCR conditions were as follows: initial denaturation at 95°C for 5 minutes; 30 cycles of denaturation at 95°C for 30 seconds; annealing at 57°C for 30 seconds; extension 72°C for 30 seconds and a final extension at 72°C for 7 minutes.

4.3.10.2 Imaging, purifying, and sequencing of PCR products

Imaging, purifying, and sequencing of PCR products were performed using the same procedures described in Chapter 2, subsection 2.3.12.2.

4.3.11 Population screening

4.3.11.1 Identifying overall MAF of *KCTD7* variant

We also used 1000 Genomes (Genomes Project et al., 2012) and Exome Variant Server available through the NHLBI Exome Sequencing Project (Fu et al., 2013) to identify the carrier frequency of the variants.

4.4 Results

4.4.1 Clinical description of patients with progressive myoclonus epilepsy with ataxia

The content in this section was compiled and transcribed by Lisa Murphy, a fourth year Biochemistry thesis student and myself, and extensively reviewed by Dr. Siu, Dr. Prasad, Dr. Rupar, and Dr. Hegele.

The three sisters are the youngest of eight children born to a non-consanguineous Low-German-Speaking Mennonite couple originally from Durango, Mexico (Figure 4.4.1). II-6 became symptomatic at age two years when she first presented with seizures. Prior to the onset of seizures, signs of developmental delay, such as walking at 19 months and speech limited to single words, were evident. The initial seizures were associated with febrile events, and later afebrile generalized clonic seizures were reported without any obvious provocative triggers. The generalized seizures were associated with up rolling of the eyeballs, clonic activity in all four limbs and varied in duration but were self-limited. Postictal hemibody weakness was reported for variable periods. Additional seizure types including sudden head and shoulder drops associated with loss of posture and falls as well as episodes of staring and unresponsiveness were also reported after the onset afebrile seizures.

Following the onset of seizures, developmental regression was noted, with loss of expressive language skills and eventual progression to a nonverbal state with limited comprehension of spoken language. The first neurological assessment disclosed the presence of significant hypotonia, motor incoordination, and gait ataxia. There were no signs of ocular movements, nystagmus, or cranial nerve deficits. By age seven, seizure control was vastly improved on a combination of lamotrigine (6 mg/kg/day) and clonazepam (0.02-0.05 mg/kg/day). However, neurological examination confirmed the presence of multifocal myoclonus and development of pyramidal signs (hyperreflexia, extensor plantar responses). A mild right hemiparesis was also noted and was explained by identification of a small epidural hematoma sustained during a fall associated with myoclonic astatic seizure. Addition of levetiracetam (35 mg/kg/day) significantly improved control of myoclonus and drop attacks. Between age 9 and 14 years, additional signs of progression were noted, with loss of independent ambulation, incontinence, ability to feed herself and manage activities of independent living.

During her first evaluation at age three, II-7 was reported to have had normal development of milestones until age two. She then showed generalized clonic seizure activity, head and shoulder drops, and staring spells associated with eyelid fluttering. She began to show developmental regression after the onset of seizures. She lost expressive speech, and had significant sialorrhea (drooling). Her neurological examination disclosed

findings similar to her elder sibling: moderately increased tone in lower extremities, hyperreflexia, and gait ataxia. Her seizures were well controlled on a combination of valproic acid (15-20 mg/kg) and clonazepam (0.03-0.05 mg/kg). Since the age of five years, her seizures have remained well controlled, but her gait and myoclonus have continued to worsen. However, she retains the ability to walk and feed herself.

II-8 presented with a generalized seizure at 15 months. Prior to this, she showed signs of developmental delay in walking, but had a vocabulary of several words. She has continued to experience both febrile and afebrile events, occasional head drops, but no observed staring spells. Her seizures have remained relatively easy to control with monotherapy (valproic acid) initially, and with the appearance of myoclonic jerks clonazepam was added. She is currently 12 years of age, with mild neurological symptoms compared to her sisters with respect to gait, and coordination. She is capable of independent ambulation, and can feed and dress herself. She is cognitively delayed, and has limited expressive speech, but unlike her older sibling she has yet to show further decline in her functioning.

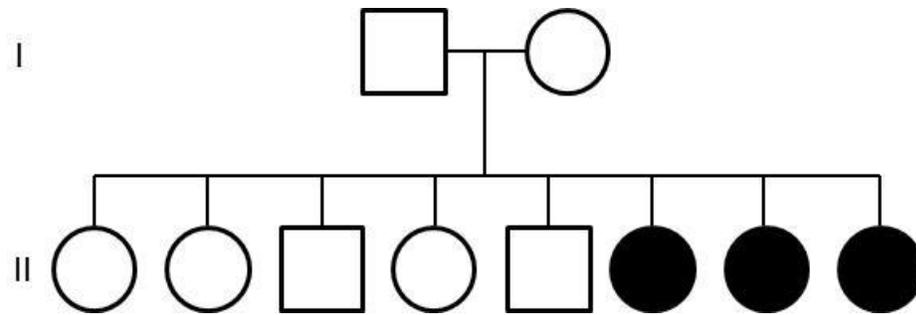


Figure 4.4.1 Pedigree with three daughters affected with epilepsy with ataxia.
Squares depict males and circles depict females. Affected individuals are shaded.

Table 4.4.1 Clinical description of patients with epilepsy with ataxia.

Clinical features	<u>Affected individuals (year of birth)</u>		
	II-6 (2000)	II-7 (2002)	II-8 (2006)
Gestation	Full term	Full term	8.5 months
Sex	F	F	F
Karyotype	46,XX	46,XX	46,XX
Birth weight (kg)	2.7	2.5	2.9
Physical examination			
Facial/limb dysmorphology	-	-	-
Teeth inturned	+	+	+
High-arched palate	+	+	+
Wide palatine ridges	+	+	+
Seizures			
Seizure onset (months)	24	24	17
Seizure type	Febrile and generalized clonic seizures, atypical absences, and head drops, drop attacks	Febrile and generalized clonic seizures, atypical absences, and head drops, drop attacks	Febrile and generalized clonic seizures, atypical absences, and head drops, drop attacks
Intention tremors	+	+	+
Growth and development			
Time of initial sitting (months)	8-9	8-9	8-9
Time of initial walking (months)	19	16	Not walking at 22 months
Time of initial speech (months)	24	16	18
Neurological examination			
	Abnormal	Abnormal	Abnormal
Gait and motor function	Broad based ataxic gait, motor incoordination, distal myoclonus	Ataxic gait, incoordination and distal myoclonus	Ataxia on stressed gait manoeuvres
Toilet training	Later years incontinent	In diapers	In diapers
Functional inquiry			
Vision	Normal	Normal	Normal
Hearing	Normal	Normal	Normal
Investigations			
Metabolic screening tests (Blood and Urine Amino acids, Urine Organic acids, Urine mucopolysaccharides, oligosaccharides)	Negative	Negative	Negative

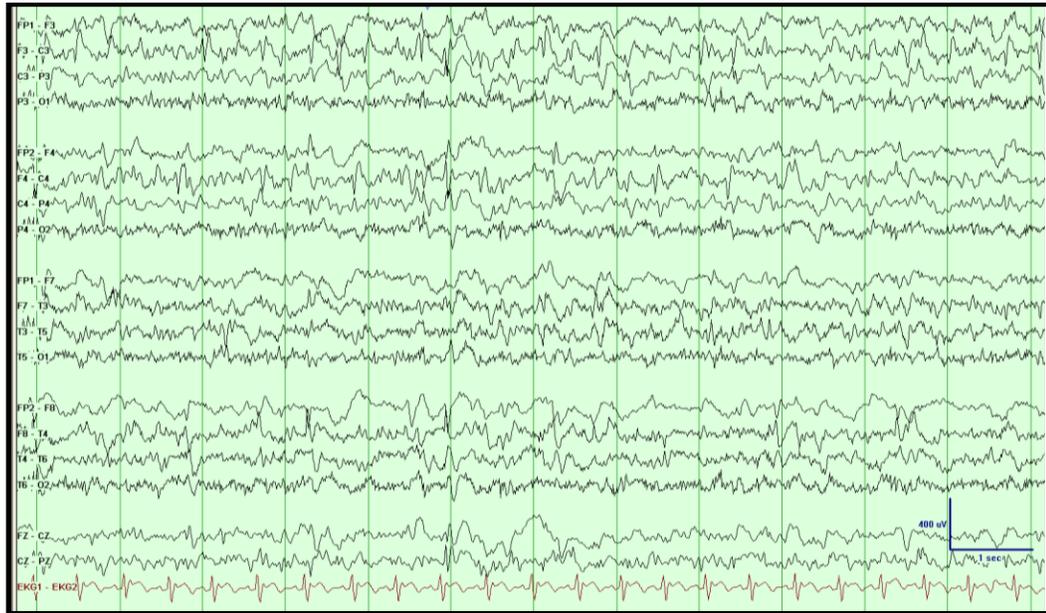
CT scan/MRI	Normal	Normal	Normal
EEG	Generalized epileptiform bursts and multiple independent spike foci, photoparoxysmal response	Multiple independent spike foci, photoparoxysmal response	Multiple independent spike foci, photoparoxysmal response
Skin biopsy EM		Showed no accumulation of any storage material in all cytological elements examined	

Abbreviations are as follows: N, normal; +, present; -, absent. Blank cells indicate unavailable information. Other metabolic testing included ammonia, lactate, complete blood count differential, liver function tests, blood urea nitrogen, creatinine, creatine kinase, and calcium produced normal results.

4.4.2 Electroencephalogram features

Dr. Prasad prepared the results from the electroencephalogram. All three siblings share similarities in their seizure phenotype and electroencephalographic findings (Figure 4.4.2). Background rhythms show poor organization and diffuse non-specific slowing. In the early stages, abundant paroxysmal bursts of generalized spike wave discharges were prominent, however with improving seizure control and increasing serial recordings showed multiple independent spike foci and regional expression of epileptiform activity distributed over bilateral central, temporal and parietal electrode chains, often with shifting asymmetry. In the instance of a single staring spell associated with behavioural arrest captured during a recording, the EEG showed focal spikes, with a posterior distribution and ictal rhythms that progressively spread from the occipital regions to become generalized with an abrupt onset and offset. All three siblings showed a photoparoxysmal response on photic stimulation, which was self-limited in nature (Figure 4.4.2).

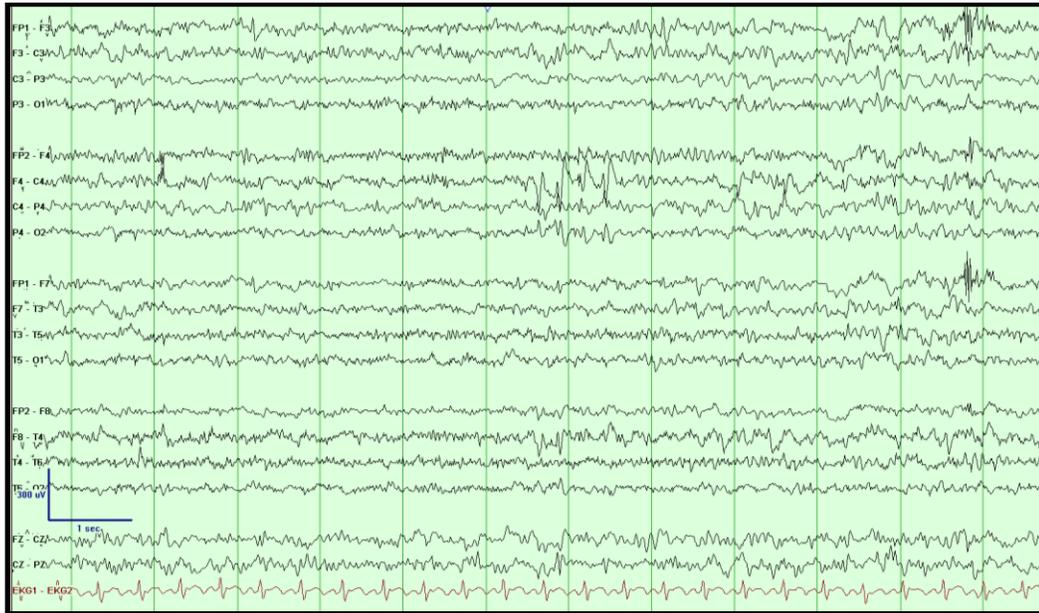
A.



B.



C.



D.

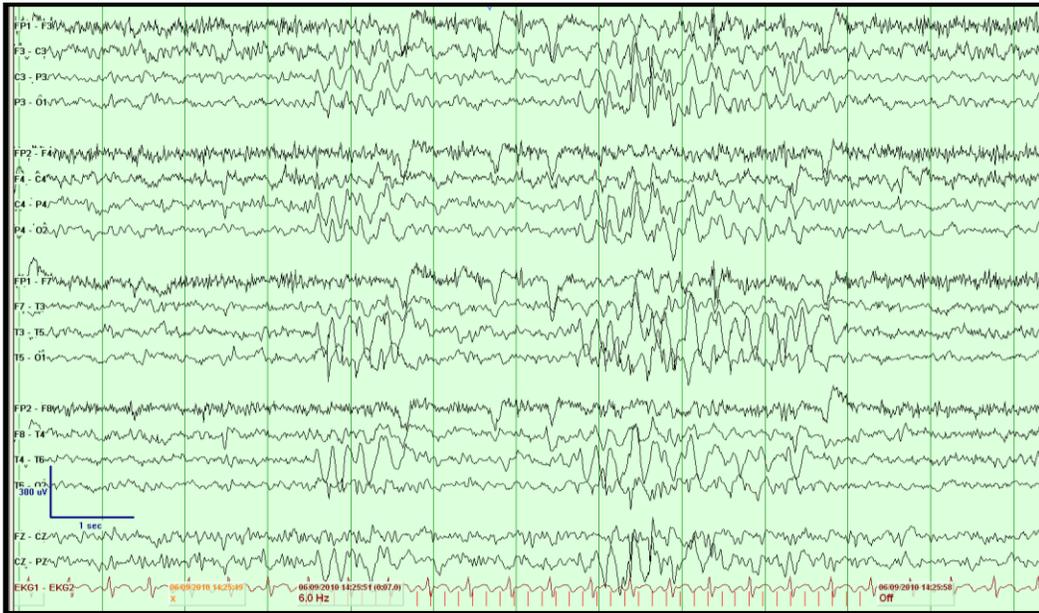


Figure 4.4.2 EEG recording of the affected individuals.

- (A) Patient 1. A snap shot of the EEG recording at age 11 years of age, abundant multifocal independent spikes, as well as synchronized, bihemispheric generalized spike waves superimposed on slow background rhythms with superimposed fast rhythms (beta frequencies) attributable to a benzodiazepine effect. (B) Patient 1. At 14 years of age, on a coronal montage, generalized 2-3 Hz multiphasic spike and wave discharges with a predominant distribution over midline parasagittal and bitemporal regions.
- (C) Patient 3. Excess fast background rhythms are seen, with focal repetitive spike and waves in the right hemisphere, and wave maxima over C4, P4 and T4-T6 electrodes. (D) Patient 3. At 6 Hz flash frequency a generalized and self-limited photoparoxysmal response is seen.

4.4.3 Genome-wide autozygosity mapping generated a high priority region on chromosome 7p12.1-7q11.22

Although the family is unaware of any consanguinity, we modelled an autosomal recessive disease inheritance given the known founder effect within the Mennonite community. Genome-wide autozygosity mapping generated a high priority locus on chromosome 7p12.1-7q11.22, ~14.3 Mb in length with a significant location score of 492 (Figure 4.4.3).

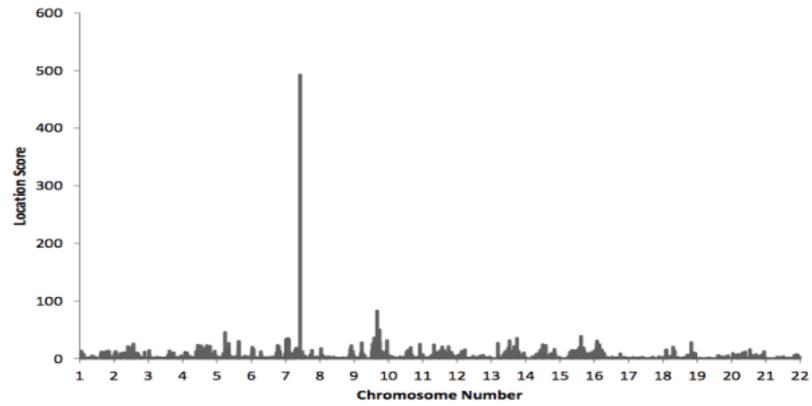


Figure 4.4.3 Homozygous region on chromosome 7 unique to affected individuals.

The x and y axes show autosomal chromosomes and location scores, respectively.

4.4.4 KCTD7 p.Tyr276Cys in patients with progressive myoclonus epilepsy with ataxia

Next, we performed exome sequencing and applied a non-synonymous, rare variant filtering analysis (Figure 4.4.4.A). We identified a novel variant within the autozygous region: g.66104176A>G, c.827A>G, p.Tyr276Cys in *KCTD7*, the potassium channel tetramerization domain-containing 7 gene. *KCTD7* is predominantly expressed in the brain specifically in cortical neurons, in granular and pyramidal cell layers of the hippocampus, and in cerebellar Purkinje cells and is involved in hyperpolarization of the cell membrane via interaction with a component of the ubiquitin ligase complex (Azizieh et al., 2011; Kousi et al., 2012; Staropoli et al., 2012) (Figure 4.4.4.A, 4.4.4.B, and 4.4.4.C). Furthermore, *KCTD7* p.Tyr276Cys was consistently predicted to be ‘disease-causing’ by multiple *in silico* analyses, suggesting deviation from the wild type amino acid residue (tyrosine) may disrupt protein function (Figure 4.4.4.D). According to the Exome Variant Server, the MAF of p.Tyr276Cys in *KCTD7* is 0.008% in the general population.

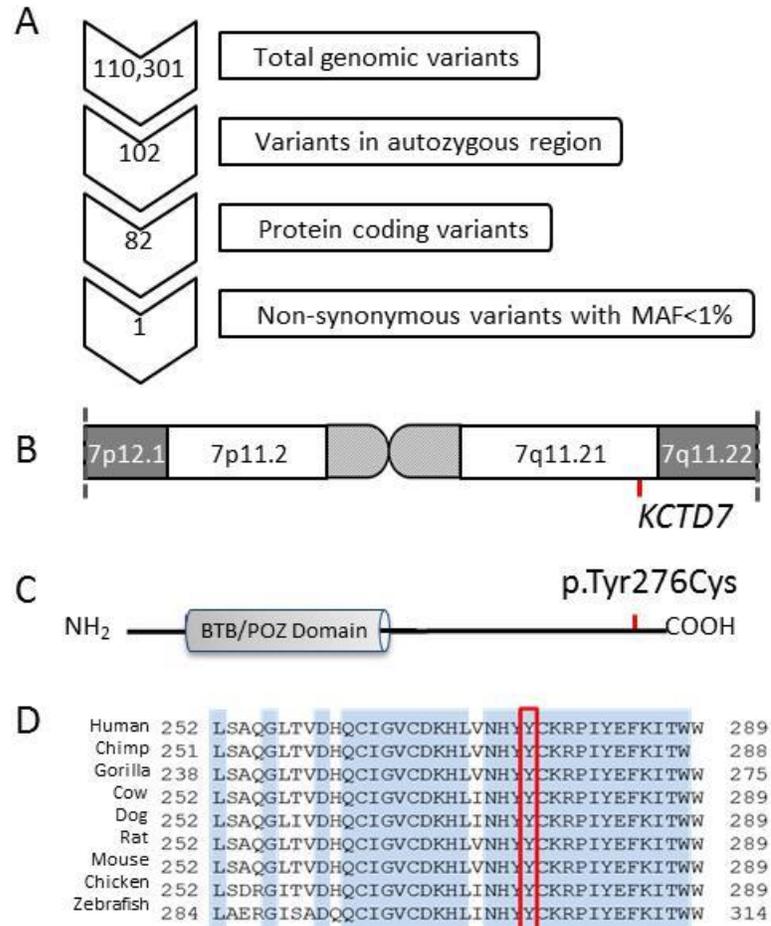


Figure 4.4.4 Schematic of mutation discovery.

(A) Using exome sequencing, we identified a rare, non-synonymous variant. (B) KCTD7 is located on chromosome 7q11.21, within the autozygous region. (C) BTB/POZ domain is the only domain in the KCTD7 protein. (D) Multiple alignments of KCTD7 demonstrate high homology of amino acid residue p.Tyr276 across a set of divergent species. The residues shaded in blue indicate fully conserved residues.

4.4.5 KCTD7 p.Tyr276Cys segregates with disease phenotype in the family

We performed subsequent variant validation analysis and observed *KCTD7* p.Tyr276Cys to segregate with disease status in the family (Figure 4.4.5). All three affected individuals are homozygous for the variant and unaffected siblings are either heterozygous or homozygous for the wild type allele. As expected, both parents are heterozygotes.

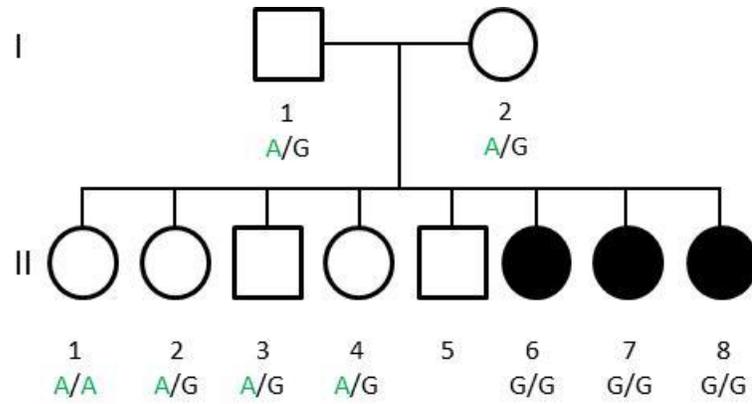


Figure 4.4.5 *KCTD7* g.661041A>G segregates with affected individuals in the family. *KCTD7* g.661041A>G segregated with the three affected females as shown by a 'G/G' genotype, heterozygotes have a 'A/G' genotype, and homozygotes for the wild type allele have a 'A/A' genotype.

4.4.6 *KCTD7* is a known disease gene

When reviewing the *KCTD7* literature, we identified additional *KCTD7* variants observed in individuals affected with progressive myoclonus epilepsy. In total, we identified 11 disease-causing *KCTD7* variants in 19 patients (Blumkin et al., 2012; Kousi et al., 2012; Krabichler et al., 2012; Staropoli et al., 2012; Van Bogaert et al., 2007). These variants and their corresponding phenotypes are summarized in Table 4.4.6.

4.5 Discussion

Progressive myoclonus epilepsy is an umbrella term encompassing a group of central nervous system disorders leading to involuntary muscle contractions and seizures, which worsen with time (Girard et al., 2013). Onset is usually around two years of age and is often accompanied with developmental regression and in some cases, prominent ataxia, speech and language impediments, and intellectual disability (Girard et al., 2013). Notably, progressive myoclonus epilepsy is a genetically and clinically heterogeneous group of disorders however, the clinical presentation and progression often overlap with other progressive myoclonus epilepsy syndromes, which are classified based on their underlying genetic etiology.

Typically, autosomal recessive (homozygous and compound heterozygous) mutations in the potassium channel tetramerization domain containing protein 7 (*KCTD7*) gene, a member of the KCTD gene family, cause progressive myoclonus epilepsy (Deng et al., 2014). In this Chapter, we coupled autozygosity mapping with exome sequencing to identify a novel variant in *KCTD7* in three patients affected with epilepsy with ataxia. Currently, the biological function of *KCTD7* in epilepsy is not fully known; however, given that mutations within *KCTD7* have been previously observed in epileptic patients, and the protein is predominantly expressed in the brain, it is likely that *KCTD7* is involved in a highly regulated electric circuit. Therefore, abnormal *KCTD7* function leads to epileptogenesis.

KCTD7 is a member of the KCTD gene family (Deng et al., 2013). The family of proteins shares an N-terminal BTB/POZ domain that demonstrates sequence homology to the TI domain in voltage gated potassium channels (Staropoli et al., 2012). To date, there have been 11 *KCTD7* mutations observed in 19 different patients presenting with some form of progressive myoclonus epilepsy (Table 4.4.6). These patients share many clinical symptoms including myoclonic seizures and cognitive impairment however, only a subset of these patients have ataxia and/or dystonia, which may suggest that that disorder is more severe depending on the type of mutation and its location (Table 4.4.6). However, all 11 mutations are within the coding region of *KCTD7* with five in its only known functional domain, BTB/POZ. Interestingly, mutations within the BTB/POZ domain do not lead to a more severe clinical phenotype.

The epilepsy phenotype in the three siblings followed over 10 years shows a distinctive natural history. Features of the first phase include motor and speech delay prior to onset of

seizures. In the second phase, initial complex febrile seizures evolve into multiple seizure types including generalized clonic seizures, atypical absences and myoclonic head and shoulder drops. Developmental regression involving loss of language is accompanied by prominent ataxia, incoordination, and non-epileptic myoclonus. There appears to be a plateau in the third phase lasting several years, wherein seizure control improves. Late deterioration was seen in the most severely affected child, involving loss of ambulation and incontinence. There is considerable variability in the severity of symptoms across the 3 siblings, with the youngest being most mildly affected and not yet having experienced significant deterioration in function.

The EEG features again suggest an evolution that mirrors the clinical neurological progression, with abundant multifocal spikes as well as generalized epileptiform activity in the initial stages followed in later years by a more regional expression of spike wave activity. Photic stimulation is associated with the development of photoparoxysmal responses that are self-limited and do not outlast the stimulus train. The seizures appear to respond particularly well to treatment with several agents alone or in combination, including valproic acid, lamotrigine, clonazepam and levetiracetam. The latter has been particularly helpful in reducing myoclonic seizures and falls. The clinical presentation and progression share some features with other progressive myoclonus epilepsy syndromes such as neuronal ceroid lipofuscinoses (Staropoli et al., 2012).

Within the progressive myoclonic epilepsies, there is also Baltic myoclonus epilepsy (Unverricht Lundborg); action myoclonus-renal failure syndrome; Lafora body disease, myoclonic epilepsy associated with ragged-red fibers (MERRF) syndrome, and the neuronal ceroid lipofuscinosis disorders.

Baltic myoclonus epilepsy (also named myoclonic epilepsy of Unverricht and Lundborg for the clinicians who first described the disease in Finland in the late 1800s) is an autosomal recessive disorder characterized by onset of neurodegeneration between 6 and 16 years of age (Kalviainen et al., 2008). It differs from other progressive myoclonus epilepsy as it is progressive only in adolescence. Typically, there is dramatic worsening of myoclonus and ataxia in the first 6 years after onset (Ramachandran et al., 2009). The disease stabilizes in early adulthood, and in some patients, myoclonus and ataxia may even improve while cognitive function is relatively preserved (Ramachandran et al., 2009). Mutations in cystatin B (*CSTB*) have been identified in patients with Baltic myoclonus epilepsy (Joensuu et al., 2008). Cystatin B is a cysteine protein

inhibitor and is thought to be involved in preventing proteinases leaking from lysosomes (Joensuu et al., 2008). In an animal study performed by Pennacchio et al., *Cstb* mice knockouts succumbed to myoclonic seizures and ataxia thus, recapitulating features seen in patients with Baltic myoclonus epilepsy (Pennacchio et al., 1998).

In action myoclonus-renal failure (AMRF) syndrome, patients present with progressive action myoclonus, dysarthria, later development of ataxia, intermittent generalized seizures, and in some patients, renal failure (Badhwar et al., 2004). AMRF syndrome is most often caused by mutations in the scavenger receptor class B, member 2 gene (*SCARB2*) inherited in autosomal recessive pattern (homozygous and compound heterozygous mutations have been observed) (Dibbens et al., 2011; Dibbens et al., 2009). *SCARB2* is involved in lysosomal mannose-6-phosphate-independent trafficking of β -glucocerebrosidase, the enzyme deficient in patients with Gaucher disease (Zeigler et al., 2014).

The Lafora type of progressive myoclonic epilepsy, which was first described by Lafora and Glueck in 1911, is an autosomal recessive disorder characterized by insidious onset of progressive neurodegeneration between 8 and 18 years of age (Ramachandran et al., 2009). Clinical symptoms can include headaches, academic difficulties, myoclonic jerks, generalized seizures, and frequent visual hallucination, which continue to worsen with time (Ramachandran et al., 2009). This is accompanied by progressive cognitive decline, resulting in dementia. Typically, 10 years following disease onset, patients are in near-continuous myoclonus with absence seizures, experience generalized seizures, and become severely demented (Ramachandran et al., 2009). Histologic examination of brain, muscle, liver, and heart biopsies show intracellular Lafora bodies, which are characterized by an accumulation of abnormal glycogen molecules (polyglucosans) (Andrade et al., 2003). Two genes have been linked to Lafora body disease namely, the NHL repeat-containing 1 gene (*NHLRC1*) and the epilepsy, progressive myoclonic 2A gene (*EPM2A*) (Chan et al., 2003; Serratosa et al., 1999). *NHLRC1* encodes malin, which is a single subunit E3 ubiquitin ligase; and *EPM2A* encodes laforin, a protein phosphatase that hydrolyzes phosphotyrosine and phosphoserine/threonine substrates (Gentry et al., 2005; Worby et al., 2006). Gentry et al. determined that laforin is a substrate for malin by showing that malin interacts with laforin leading to its polyubiquitination and eventual degradation (Gentry et al., 2005). As malin regulates cellular protein concentration of laforin, not

surprisingly, mutations in either *NHLRC1* or *EPM2A* disrupt this interaction and lead to the formation of aggregates.

The neuronal ceroid lipofuscinoses (NCL) are a clinically and genetically heterogeneous group of neurodegenerative disorders characterized by an accumulation of lipofuscin in various cell types, which resemble aspects of classical lysosomal storage diseases (Boustany, 2013). The clinical course of patients with NCL includes cognitive and motor impairments, seizures, progressive blindness, and early deaths (Boustany, 2013). There is significant genetic heterogeneity in the NCL disorders as 14 loci have been implicated in different NCL subtypes, each of which varies in age of onset (infantile, late infantile, juvenile, and adult onset) and severity of the disease (Radke et al., 2015). Each subtype is linked to a disruption in a specific gene defect but some pathways have been observed in multiple subtypes including apoptosis (Kim et al., 2006); clearance of degraded proteins (Walus et al., 2010); and most often, disordered intracellular trafficking to lysosomes (Schmiedt et al., 2010). Furthermore, each subtype can be visualized based on the type of lipopigment storage material pattern they produce (Mole et al., 2005). For example, granular osmiophilic deposits (GROD) are seen in the first subtype (CLN1) while CLN2 and CLN3 display 'curvilinear' and 'fingerprint' profiles, respectively and CLN4, CLN5, CLN6, CLN7, and CLN8 show a mixture of granular, curvilinear, fingerprint, and rectilinear patterns (Mole et al., 2005).

Therefore, knowledge of the clinical profile, seizure types, and EEG features of the disease phenotype associated with the specific mutation as described above aids in improving diagnostic precision. Identification of the causative mutation is also important for prognostication and genetic counseling, and provides a sound basis of treatment decisions and symptom management.

4.6 Conclusion

In summary, using autozygosity mapping and exome sequencing, we identified a novel homozygous variant, p.Tyr276Cys in *KCTD7*, in three sisters affected with progressive myoclonus epilepsy with ataxia. *KCTD7* is a known progressive myoclonus epilepsy gene and is predominantly expressed in the brain. Variants in *KCTD7* have been previously implicated in epileptic phenotypes due to disturbances in potassium channel conductance. Our results further

support the role of KCTD7 in epileptic seizures. During our analysis, we also identified three siblings (II-2, II-3, and II-4) to be carriers of KCTD7 p.Tyr276Cys, which can have significant clinical utility. The newly discovered mutation is now a genetic marker and can be used in genetic counseling for the siblings and their prospective mates and eventually, for other relatives from the same Low-German-Speaking Mexican Mennonite community.

4.7 References

- Andrade, D.M., Ackerley, C.A., Minett, T.S., Teive, H.A., Bohlega, S., Scherer, S.W., and Minassian, B.A. (2003). Skin biopsy in Lafora disease: genotype-phenotype correlations and diagnostic pitfalls. *Neurology* *61*, 1611-1614.
- Azizieh, R., Orduz, D., Van Bogaert, P., Bouschet, T., Rodriguez, W., Schiffmann, S.N., Pirson, I., and Abramowicz, M.J. (2011). Progressive myoclonic epilepsy-associated gene KCTD7 is a regulator of potassium conductance in neurons. *Mol Neurobiol* *44*, 111-121.
- Badhwar, A., Berkovic, S.F., Dowling, J.P., Gonzales, M., Narayanan, S., Brodtmann, A., Berzen, L., Caviness, J., Trenkwalder, C., Winkelmann, J., *et al.* (2004). Action myoclonus-renal failure syndrome: characterization of a unique cerebro-renal disorder. *Brain* *127*, 2173-2182.
- Blumkin, L., Kivity, S., Lev, D., Cohen, S., Shomrat, R., Lerman-Sagie, T., and Leshinsky-Silver, E. (2012). A compound heterozygous missense mutation and a large deletion in the KCTD7 gene presenting as an opsoclonus-myoclonus ataxia-like syndrome. *J Neurol* *259*, 2590-2598.
- Boustany, R.M. (2013). Lysosomal storage diseases--the horizon expands. *Nat Rev Neurol* *9*, 583-598.
- Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* *14*, 681-691.
- Buono, R.J. (2013). Genome wide association studies (GWAS) and common forms of human epilepsy. *Epilepsy Behav* *28 Suppl 1*, S63-65.
- Chan, E.M., Young, E.J., Ianzano, L., Munteanu, I., Zhao, X., Christopoulos, C.C., Avanzini, G., Elia, M., Ackerley, C.A., Jovic, N.J., *et al.* (2003). Mutations in NHLRC1 cause progressive myoclonus epilepsy. *Nat Genet* *35*, 125-127.
- de Araujo Furtado, M., Rossetti, F., Chanda, S., and Yourick, D. (2012). Exposure to nerve agents: from status epilepticus to neuroinflammation, brain damage, neurogenesis and epilepsy. *Neurotoxicology* *33*, 1476-1490.

- Deng, H., Xiu, X., and Song, Z. (2014). The molecular biology of genetic-based epilepsies. *Mol Neurobiol* 49, 352-367.
- Dibbens, L.M., Karakis, I., Bayly, M.A., Costello, D.J., Cole, A.J., and Berkovic, S.F. (2011). Mutation of SCARB2 in a patient with progressive myoclonus epilepsy and demyelinating peripheral neuropathy. *Arch Neurol* 68, 812-813.
- Dibbens, L.M., Michelucci, R., Gambardella, A., Andermann, F., Rubboli, G., Bayly, M.A., Joensuu, T., Vears, D.F., Franceschetti, S., Canafoglia, L., *et al.* (2009). SCARB2 mutations in progressive myoclonus epilepsy (PME) without renal failure. *Ann Neurol* 66, 532-536.
- Fisher, R.S., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., and Engel, J., Jr. (2005). Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* 46, 470-472.
- Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., *et al.* (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216-220.
- Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.
- Gentry, M.S., Worby, C.A., and Dixon, J.E. (2005). Insights into Lafora disease: malin is an E3 ubiquitin ligase that ubiquitinates and promotes the degradation of laforin. *Proc Natl Acad Sci U S A* 102, 8501-8506.
- Girard, J.M., Turnbull, J., Ramachandran, N., and Minassian, B.A. (2013). Progressive myoclonus epilepsy. *Handb Clin Neurol* 113, 1731-1736.
- Joensuu, T., Lehesjoki, A.E., and Kopra, O. (2008). Molecular background of EPM1-Unverricht-Lundborg disease. *Epilepsia* 49, 557-563.
- Kalviainen, R., Khyuppenen, J., Koskenkorva, P., Eriksson, K., Vanninen, R., and Mervaala, E. (2008). Clinical picture of EPM1-Unverricht-Lundborg disease. *Epilepsia* 49, 549-556.
- Katchanov, J., and Birbeck, G.L. (2012). Epilepsy care guidelines for low- and middle- income countries: From WHO mental health GAP to national programs. *BMC Med* 10, 107.
- Kim, S.J., Zhang, Z., Lee, Y.C., and Mukherjee, A.B. (2006). Palmitoyl-protein thioesterase-1 deficiency leads to the activation of caspase-9 and contributes to rapid neurodegeneration in INCL. *Hum Mol Genet* 15, 1580-1586.
- Kousi, M., Anttila, V., Schulz, A., Calafato, S., Jakkula, E., Riesch, E., Myllykangas, L., Kalimo, H., Topcu, M., Gokben, S., *et al.* (2012). Novel mutations consolidate KCTD7 as a progressive myoclonus epilepsy gene. *J Med Genet* 49, 391-399.

Krabichler, B., Rostasy, K., Baumann, M., Karall, D., Scholl-Burgi, S., Schwarzer, C., Gautsch, K., Spreiz, A., Kotzot, D., Zschocke, J., *et al.* (2012). Novel mutation in potassium channel related gene KCTD7 and progressive myoclonic epilepsy. *Ann Hum Genet* 76, 326-331.

Mole, S.E., Williams, R.E., and Goebel, H.H. (2005). Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses. *Neurogenetics* 6, 107-126.

Newton, C.R., and Garcia, H.H. (2012). Epilepsy in poor regions of the world. *Lancet* 380, 1193-1201.

Pennacchio, L.A., Bouley, D.M., Higgins, K.M., Scott, M.P., Noebels, J.L., and Myers, R.M. (1998). Progressive ataxia, myoclonic epilepsy and cerebellar apoptosis in cystatin B-deficient mice. *Nat Genet* 20, 251-258.

Radke, J., Stenzel, W., and Goebel, H.H. (2015). Human NCL Neuropathology. *Biochim Biophys Acta*.

Ramachandran, N., Girard, J.M., Turnbull, J., and Minassian, B.A. (2009). The autosomal recessively inherited progressive myoclonus epilepsies and their genes. *Epilepsia* 50 Suppl 5, 29-36.

Schmiedt, M.L., Bessa, C., Heine, C., Ribeiro, M.G., Jalanko, A., and Kyttala, A. (2010). The neuronal ceroid lipofuscinosis protein CLN5: new insights into cellular maturation, transport, and consequences of mutations. *Hum Mutat* 31, 356-365.

Serratos, J.M., Gomez-Garre, P., Gallardo, M.E., Anta, B., de Bernabe, D.B., Lindhout, D., Augustijn, P.B., Tassinari, C.A., Malafosse, R.M., Topcu, M., *et al.* (1999). A novel protein tyrosine phosphatase gene is mutated in progressive myoclonus epilepsy of the Lafora type (EPM2). *Hum Mol Genet* 8, 345-352.

Staropoli, J.F., Karaa, A., Lim, E.T., Kirby, A., Elbalalesy, N., Romansky, S.G., Leydiker, K.B., Coppel, S.H., Barone, R., Xin, W., *et al.* (2012). A homozygous mutation in KCTD7 links neuronal ceroid lipofuscinosis to the ubiquitin-proteasome system. *Am J Hum Genet* 91, 202-208.

Tator, C., Bray, G., and Morin, D. (2007). The CBANCH report--the burden of neurological diseases, disorders, and injuries in Canada. *Can J Neurol Sci* 34, 268-269.

Van Bogaert, P., Azizieh, R., Desir, J., Aeby, A., De Meirleir, L., Laes, J.F., Christiaens, F., and Abramowicz, M.J. (2007). Mutation of a potassium channel-related gene in progressive myoclonic epilepsy. *Ann Neurol* 61, 579-586.

Walus, M., Kida, E., and Golabek, A.A. (2010). Functional consequences and rescue potential of pathogenic missense mutations in tripeptidyl peptidase I. *Hum Mutat* 31, 710-721.

Worby, C.A., Gentry, M.S., and Dixon, J.E. (2006). Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates. *J Biol Chem* 281, 30412-30418.

Zeigler, M., Meiner, V., Newman, J.P., Steiner-Birmanns, B., Bargal, R., Sury, V., Mengistu, G., Kakhlon, O., Leykin, I., Argov, Z., *et al.* (2014). A novel SCARB2 mutation in progressive myoclonus epilepsy indicated by reduced beta-glucocerebrosidase activity. *J Neurol Sci* 339, 210-213.

Chapter 5 - Known phenotype, novel gene: loss of a novel synaptic protein, *TMTC3*, in patients with nocturnal seizures with developmental delay.

5 Study rationale

Four siblings, three females and one male, were born to a healthy non-consanguineous Pakistani couple and are equally affected with nocturnal seizures and developmental delay, which is a previously described phenotype. The four affected individuals did not demonstrate significant loci that mapped to any of the known genes associated with nocturnal frontal epilepsy, which suggests that nocturnal seizures with developmental delay, is likely caused by novel genomic loci. The family agreed to participate in genetic testing and the project was funded by the FORGE Canada Consortium as well as the Rare Diseases: Models & Mechanisms Network.

Following gene identification, we collaborated with Dr. Jamie Kramer and Dr. Michael Poulter to study the function of *TMTC3* using model organisms. Our goal here is to apply comprehensive neuro-phenotyping of the *Drosophila melanogaster TMTC3* ortholog, which may reveal novel functions for this protein in the central nervous system. Furthermore, we intend on applying specialized assays to model seizures and developmental delay in a model organism, which may recapitulate clinical features of the four children, and also further characterize the role of *TMTC3* in seizure development.

Chapter 5 was prepared as an independent study: “Loss of *TMTC3*, a synaptic protein, in patients with nocturnal seizures and intellectual disability”, and has been submitted for publication (Farhan et al., 2016).

5.1 Overview

Herein, we provide evidence that loss of the transmembrane and tetratricopeptide repeat containing 3 (*TMTC3*) protein causes nocturnal seizures and intellectual disability (ID) in a family with four affected children. Using exome sequencing, we identified compound

heterozygous variants (p.Arg71His and p. Leu729ThrfsTer6) in the TMTC3 gene in this family. All four affected siblings had this combination of mutations demonstrating autosomal recessive inheritance. Expression analysis using patient-derived cells confirmed reduced TMTC3 transcript levels and loss of the TMTC3 protein when compared to cell lines taken from both parents and an additional healthy control. As TMTC3 function is currently unexplored in brain, we gathered support for a neurobiological role for TMTC3 by generating flies with postmitotic neuron-specific knockdown of the highly conserved *Drosophila melanogaster* TMTC3 ortholog, CG4050. Knockdown of this gene in flies showed increased susceptibility to mechanically induced seizures, a phenotype that was ameliorated by reintroducing the human TMTC3 protein into fly neurons. Furthermore, we show that TMTC3 is localized to pre-synaptic terminals and boutons in GABAergic inhibitory neurons of the rat hypothalamus and piriform cortex. By demonstrating that loss of TMTC3 leads to seizure susceptibility in both humans and flies, and by characterizing the synaptic localization of this protein, our results uncover a novel role for TMTC3 in the regulation of synaptic function and neuronal physiology.

5.2 Introduction

Epilepsy, defined by recurrent unprovoked seizures, is a clinically diverse disorder with a complex etiology (Farhan et al., 2014). Early genetic studies in families with a history of epilepsy revealed a role for synaptic ion channels, such as the nicotinic acetylcholine receptors (nAChRs) (Aridon et al., 2006; De Fusco et al., 2000; Steinlein et al., 1995). This led to the hypothesis that dysregulation of ion channels is an important mechanism underlying epilepsy. More recent studies using massively parallel sequencing indeed support the role of ion channels in the etiology of epilepsy, but also reveal a heterogeneous group of genes with other functions that contribute to various types of epilepsy (Myers and Mefford, 2015; Thomas and Berkovic, 2014).

Some forms of epilepsy are characterized by nocturnal seizures, which occur almost exclusively during sleep. Nocturnal seizures are seen in association with certain syndromes, including focal epilepsy with speech disorder (FESD - OMIM 245570), and autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE - OMIM 600513, 603204, 605375, 610353, 615005) (Bazil, 2004). FESD, a clinically diverse disorder that is often comorbid with intellectual

disability, is caused by mutations in *GRIN2A* (Lemke et al., 2013; Lesca et al., 2013; Venkateswaran et al., 2014). *GRIN2A* is a subunit of the NMDA receptor, a ligand- and voltage-gated calcium channel. Similar to FESD, *ADNFLE* is also often comorbid with speech and psychiatric disorders, or intellectual disability (Nobili et al., 2014). *ADNFLE* has been recurrently associated with multiple loci encoding components of the synaptic nAChRs (Becchetti et al., 2015), which are ligand-gated non-selective ion channels. Causative variants have been identified in *CHRNA4* (Steinlein et al., 1995), *CHRNA2* (De Fusco et al., 2000), and *CHRNA2* (Aridon et al., 2006). Recently, *ADNFLE* was also linked to mutations in *KCNT1*, encoding a sodium-gated potassium channel (Heron et al., 2012). The effects of these variants are further exemplified as murine models harboring the same variants causal of *ADNFLE*, recapitulate the epileptic seizures observed in the human disease state (Lim et al., 2016; Shiba et al., 2015). Thus, in general, epilepsy syndromes that are characterized by nocturnal seizures appear to be predominantly associated with defects in ion channel function (Kim et al., 2014; Lim et al., 2016).

Here, we have ascertained a family with four affected siblings who presented with nocturnal seizures and mild to moderate intellectual disability. Through whole exome sequencing, we identified compound heterozygous mutations in the *transmembrane and tetratricopeptide repeat containing 3 (TMTC3)* gene in this family. *TMTC3* mutations have not been previously identified as a cause of human disease. *Tmtc3* deficient mice exhibit neonatal lethality with only a few significantly smaller mice surviving for up to three weeks (Yun and Vu, 2012), however, there is no knowledge regarding the role of *TMTC3* in nervous system development or function. In this report, we demonstrate a neuron-specific role for *TMTC3* in mechanically induced seizures in *Drosophila*, and provide evidence for pre-synaptic localization of *TMTC3* in the mammalian brain. Taken together, our results strongly support that *TMTC3* defects can cause nocturnal seizures and ID in humans, and uncover a novel and evolutionarily conserved role for *TMTC3* in synaptic function and seizure susceptibility.

5.3 Materials and methods

5.3.1 Ethics

As described in Chapter 2, subsection 2.3.1, ethics approval was obtained from the parents on behalf of all six members of the family.

5.3.2 Patients and biological materials

Four children of a healthy non-consanguineous Pakistani couple were clinically ascertained because they were experiencing nocturnal seizures and also had developmental delay. Blood and tissue (skin-derived fibroblast cells) samples were collected from all four affected children and the two parents.

5.3.3 DNA isolation

DNA was isolated from all six family members as described in Chapter 2, subsection 2.3.3.

5.3.4 Exome sequencing

Exome sequencing was performed using gDNA from two affected individuals (II-1 and II-4) using the protocols described in Chapter 2, subsection 2.3.8.

5.3.5 Sequence alignment

Sequence alignment was performed using the same procedures described in Chapter 2, subsection 2.3.9.

5.3.6 Variant calling and annotation

Variant calling and annotation were performed using the same procedures described in Chapter 2, subsection 2.3.10.

5.3.7 Variant discovery

5.3.7.1 Prioritization of autosomal recessive, non-synonymous, and rare variants

We investigated whether there were any autosomal recessive variant (homozygous or compound heterozygous) non-synonymous, rare variants in the exomes of the patients. We scanned for variants in genes known to cause seizures and/or intellectual disability based on literature and OMIM, as well as all genes.

5.3.7.2 *In silico* analyses

To determine the predicted biological effect of non-synonymous variants on protein function, candidate variants (*TMTC3* p.Arg71His and p.Leu728fs*6) were assessed using the same tools described in Chapter 2, subsection 2.3.11.2.

5.3.8 Variant validation

To determine whether p.Arg71His and p.Leu728fs*6 in *TMTC3* segregated with disease status in the pedigree, we genotyped family members using standard PCR and Sanger sequencing methods as described in Chapter 2, subsection 2.3.12.1 and 2.3.12.2.

5.3.8.1 PCR

Forward and reverse primers specific to amplify g.11018G>A in *TMTC3* were: 5'-CAGTGATGAGCTTTTGTACCC-3' and 5'-CACACTACTCTTGTGTCC-3', respectively; and primers for *TMTC3* g.52793insA were: 5'-AAGCCGACTTCCGAAGTGCT-3' and 5'-TTCCTTGCACATTGCTTGG-3'. The PCR conditions for both primer pairs were as follows: initial denaturation at 95°C for 5 minutes; 30 cycles of denaturation at 95°C for 30 seconds; annealing at 58°C for 30 seconds; extension 72°C for 30 seconds and a final extension at 72°C for 7 minutes.

5.3.8.2 Imaging, purifying, and sequencing of PCR products

Imaging, purifying, and sequencing of PCR products were performed using the same procedures described in Chapter 2, subsection 2.3.12.2.

5.3.9 Cell culture

Skin fibroblasts from patients, both parents, and a healthy control were cultured as described in Chapter 2, subsection 2.3.14.

5.3.10 Protein isolation

Protein isolation was performed as described in Chapter 3, subsection 3.3.13.

5.3.11 Immunoblotting

Protein expression was measured using immunoblotting as described in Chapter 2, subsection 2.3.15.

5.3.12 Reverse transcriptase (RT)-PCR

RT-PCR was performed as described in Chapter 3, subsection 3.3.15.

5.3.13 Model organisms

In collaboration with Dr. Jamie Kramer, we are studying the functional effect of *TMTC3* deficiency in *Drosophila melanogaster*.

5.3.13.1 Gateway cloning technology

This step was performed by Dr. Heleen Arts and myself.

A pTW destination vector was obtained from the Drosophila Genomics Resource Center (Bloomington, Indiana, USA) for fly rescue experiments. Once the UAS site was sequence verified, the vector was sent to the SPARC BioCentre at The Hospital for Sick Children (Toronto, Ontario, Canada) for insertion of full length *TMTC3* cDNA (Internal ID 53914 of the hORFeome V5.1 of the CCSB Human Orfeome Collection and Entrez Gene ID 160418) through regular Gateway cloning (Life Technologies, Carlsbad, CA). Following growth of LB agar plates with colonies resulting from LR reaction, we picked several colonies for mini cultures in LB medium with ampicillin (50 ug/ml) and incubated these overnight in a shaker at 37C°. Plasmid DNA was isolated with a DNA isolation kit according to the manufacturer's instructions (NucleoSpin Plasmid Easy Pure, Machery Nagel, Düren, Germany). Clones were validated

through restriction analysis with *ApaLI* (New England Biolabs, Ipswich, MA, USA). The insert and flanking sequences were analyzed via Sanger sequencing as previously described.

A verified construct was retransformed in DH5 α bacteria (Subcloning Efficiency DH5 α Competent Cells, Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. Bacteria containing the pTW *TMTC3* vector were selected on LB plates with ampicillin (50 ug/ml). 100 ml of LB medium with antibiotics was inoculated directly from a single colony and grown overnight. A Plasmid Midi Kit (Qiagen, Hilden, Germany) was used for DNA isolation according to the manufacturer's instructions. Vector DNA was validated through restriction analysis with *ApaLI* and sent for generation of transgenic flies.

5.3.13.2 *Drosophila* stocks

These experiments were performed in Dr. Kramer's lab.

Fly lines expressing transgenic RNAi encoding inverted repeats for the *CG4050* gene (33248 and 110199, here referred to as UAS-*tmtc3*-IR1 and UAS-*tmtc3*-IR2, respectively) and the appropriate genetic background control lines (60000 and 60100, respectively) were obtained from the Vienna *Drosophila* Resource Center (Vienna, Austria). *elav-Gal4; UAS-dcr2* was a gift from A. Schenck. *UAS-dcr2* (24644 and 24650), *Rut-Gal4* (48667) *UAS-mcD8::GFP* (5130) and *477-Gal4, UAS-mcD8::GFP* (8746) were obtained from the Bloomington *Drosophila* Stock Center (Indiana, USA). For creation of transgenic flies with inducible expression of human *TMTC3* (*UAS-hTMTC3-1* and *UAS-hTMTC3-2*), *TMTC3* cDNA (Internal ID 53914 of the hORFeome V5.1 of the CCSB Human Orfeome Collection and Entrez Gene ID 160418) was inserted into the pTW vector (*Drosophila* Genomics Resource Center, Bloomington, Indiana, USA) at the SPARC BioCentre at The Hospital for Sick Children (Toronto, Ontario, Canada) using standard Gateway cloning (Life Technologies, Carlsbad, CA). The resulting pTW-*TMTC3* plasmid was validated by restriction digest and Sanger sequencing of the insert and flanking regions. Creation of transgenic flies using the validated plasmid was performed by BestGene Inc. (California, USA) using standard procedures.

5.3.13.3 RT-qPCR

Quantitative Real-Time PCR (RT-qPCR) was performed to evaluate the potency of UAS-*tmtc3*-IR1 and UAS-*tmtc3*-IR2. RNAi lines and the appropriate genetic background controls were

crossed the ubiquitous *Actin-Gal4* driver line and third instar larvae were selected for analysis. RNA was extracted using the RNeasy Lipid Tissue Mini Kit (Qiagen). DNase digestion was carried out using the RNase-free DNase Set (Qiagen) and cDNA was made using the iScript cDNA synthesis kit (BioRad). RT-qPCR was performed using the BioRad CFX 384 and the SYBRgreen mastermix (Biorad) with primers directed against *CG4050*, and the reference genes, *betacop*, *eIF2B*, and *rpIII40*. All primer sets were validated for efficiency according to standard protocols. *CG4050* expression was compared in control and RNAi genotypes using a Student's t-test.

5.3.13.4 Bang sensitivity

Bang sensitivity assays were performed using 5-9 day old male flies, raised at 29 °C. The following genotypes were analyzed: *elavGal4/+; UAS-dcr2/+* (Control), *elavGal4/+; UAS-dcr2/UAS-tmtc3-IR1*, *elavGal4/UAS-hTMTC3-1; UAS-dcr2/UAS-tmtc3-IR1*, and *elavGal4/UAS-hTMTC3-2; UAS-dcr2/UAS-tmtc3-IR1*. Male flies were collected in groups of ten to twelve and aged for several days. On the testing day 5-9 day old flies were transferred to empty food vials without the use of CO₂. After a five-minute acclimatization period, the vials were vortexed at the highest setting for 20 seconds. Immediately after vortexing, the behavioural response was video recorded. Videos were observed for seizure-like responses, including complete immobilization on the back and uncontrollable movement. The time for each fly to right itself was recorded. Outliers (defined by righting times of greater than five standard deviations from the mean and accounting for 7 of 538 data points with equal distribution between genotypes) were removed and a Dunn's test for multiple comparisons using the Benjamini-Hochberg correction was used to compare the average righting time of each genotype. The differences in the proportion of flies that demonstrated seizure-like behavior in each genotype were compared using a two-proportion z-test.

5.3.13.5 Immunohistochemistry, image acquisition, and analysis

In collaboration with Dr. Michael Poulter, we are studying the localization of *Tmtc3* in rat brains.

Rat brains: sections were washed with 1x PBS with 0.2% Triton X-100 three times for 5 min and blocked in 10% goat serum and 10% donkey serum in 1x PBS with 0.025% Triton X-100 and 1% bovine serum albumin (BSA) for 1h. The primary antibodies were diluted in 1x PBS

with 1% BSA and 0.025% Triton X-100. The sections incubated with primary antibodies, anti-mouse VGat antibody (1:200, Synaptic Systems, 131011) and anti-rabbit TMTC3 (1:500, Abcam ab81473), for 24 h at 4°C. Sections were then washed two times with 1x PBS with 0.2% Triton X-100 for 5 min. They were incubated with secondary antibodies for 1h at room temperature; Alexa Fluor 647 donkey anti-rabbit IgG (A31573) and Alexa Fluor 488 donkey anti-mouse IgG antibody (Molecular Probes, A-21202) diluted 1:1000 in PBS with 1% BSA and 0.025% Triton X-100. Sections were then washed three times with PBS with 0.2% Triton X-100 for 10 min and incubated in 1% Sudan Black B (Sigma-Aldrich, S2380) in 70% ethanol for 5 min. Sections were rinsed twice with 70% ethanol for 1 min, followed by two 5 min washes with 1x PBS. Sections were mounted on glass slides with glass cover slips using Prolong Gold Antifade Reagent with DAPI mounting medium (Molecular Probes, P36935). Confocal images were taken on an Olympus IX 60 inverted microscope outfitted with a Perkin Elmer spinning disk confocal attachment with either a 60 × (Numerical aperture = 1.4) immersion oil, 40 × (N.A. = 0.75), 20 × (N.A. = 0.5), or 10 × (N.A. = 0.3) objective. The microscope was equipped with a Hamamatsu Orca ER CCD camera (1300 × 1030 pixels; pixel size 6.5 μm) and images were acquired using Velocity software. Each image shown represents a stack of 10 images spaced 0.2 μm apart in the z-plane, for each wavelength.

Fly multidendritic neurons: The following genotypes were dissected as third instar larvae and fixed in 4% paraformaldehyde: *477-Gal4, UAS-mcD8::GFP/+* (control) and *+/+; 477-Gal4, UAS-mcD8::GFP/UAS-tmtc3-IR1*. Tissue was incubated in PBT (PBS with 0.3% triton-X 100) for 40 minutes and blocked in PBT supplemented with 5% normal goat serum. Larvae were incubated for two nights with anti-GFP (1:25, mouse, developmental studies hybridoma bank [DSHB]) and anti-mCD8 (1:100, rat, Life Technologies) primary antibodies diluted in PBT with 5% NGS. After washing in PBT, larvae were incubated overnight with secondary antibodies anti-mouse-DyLight488 (1:400, goat, ThermoFisher) and anti-rat-DyLight488 (1:400, goat, ThermoFisher). Dorsal type IV multidendritic neurons were imaged using a Zeiss LSM 510 DUO Vario confocal microscope. The entire neuron was captured by combining several fields of view using the tile scan function. Image stacks were imported to Neuron Studio and traced (Wearne et al., 2005). Centrifugal edge labeling was performed and analyzed using the LMeasure program to quantify various features of the neurons (Scorcioni et al., 2008). A student's t-test was used to determine if any significant differences were present between the

number of bifurcations, number of branches, number of tips, and overall length of the control and *UAS-tmtc3-IR1* Type IV multidendritic neurons.

Fly neuromuscular junction (NMJ): the genotypes *elavGal4/+; UAS-dcr2/+* (control) and *elavGal4/+; UASdcr2/UAS-tmtc3-IR1*, were processed as described above, with the following changes. Primary antibodies used were: anti-HRP (1:500, rabbit, Jackson Immuno Research) and anti-brp (nc82, 1:15, mouse, DSHB). Secondary antibodies used were anti-rabbit-DyLight488 (1:400, goat, ThermoFisher) and anti-mouse-DyLight594 (1:400, goat, ThermoFisher). To determine whether the knockdown of the *tmtc3* gene had any effect on the morphology of the *Drosophila* NMJ, images of the NMJ were loaded into the open-source image analysis software, Fiji (Schindelin et al., 2012). Using the algorithm, ‘Drosophila_NMJ_Morphometrics’ (Nijhof et al., 2016) features of the NMJ were quantified for analysis. A Mann-Whitney test for non-parametric data was used to determine if any significant differences were present between the number of branches, active zones, and the overall length and area of the control and *UAS-tmtc3-IR1* NMJ.

Fly mushroom bodies: The mushroom body specific *rut-Gal4* driver line (Jenett et al., 2012) was used to express *tmtc3* RNAi constructs and GFP. Adult male fly brains were dissected from the genotypes *UAS-dcr2/mcD8::GFP; rut-Gal4/UAS-tmtc3-IR1*, and *UAS-dcr2/mcD8GFP; rut-Gal4/+*, and fixed for 1 hour in 4% paraformaldehyde. GFP labeled mushroom bodies were imaged using a Zeiss LSM 510 DUO Vario confocal microscope.

5.3.14 Antibodies

Anti-TMTC3 (1:100, Santa Cruz Biotechnology, Dallas, Texas) was used as a primary antibody. For a positive control, α -tubulin (1:10,000, Sigma-Aldrich, Saint Louis, Missouri) was used. Anti-mouse (1:10,000, Jackson ImmunoResearch Laboratories, Inc) was used as a secondary antibody.

5.3.15 Statistics

Statistics were performed as described in Chapter 2, subsection 2.3.19.

5.4 Results

5.4.1 Clinical presentation of patients with nocturnal seizures with developmental delay

The content in this section was compiled and transcribed by me and extensively reviewed by Dr. Siu, Dr. Rupar, and Dr. Hegele.

Four siblings, three females and one male, were born to a healthy non-consanguineous Pakistani couple and are equally affected with nocturnal seizures and developmental delay (Figure 5.4.1). The history of all four sibs is quite similar (Table 5.4.1). They display developmental delay, crowding of the dentition with anterior overbite, and have a history of nocturnal seizures with an onset at two years of age. Nocturnal seizures in these children last from a few seconds up to one minute and consist of stiffening of the limbs, trembling, and rolling of the eyes. The frequency of seizures ranges between once daily to once or twice per month with the two younger children having more frequent seizures ranging from one to two per week. The children have no staring spells and have never had diurnal seizures. Furthermore, all four children displayed cognitive delays at about two years of age however, there has not been any regression of skills in any of the children. Some of the children have behavioral problems consisting of aggression, agitation, irritability, and self-harm.

Although the clinical presentation observed in these children is quite consistent, there is variability in the severity of clinical features. Individual II-I (DOB, 1987) has been treated with anticonvulsants, carbamazepine and valproic acid, which have decreased the frequency of her seizures to only one-two times per year. Individual II-II (DOB, 1989) is the most cognitively functional of all four children but still quite delayed. His seizures have occurred only once every three years and he is not on any medications at this time. II-III (DOB, 1993) and II-IV (DOB, 1994) seem to be more affected, displaying a lower level of cognitive functioning. Like her siblings, II-III presents with a constellation of nocturnal, probably complex partial seizures, intellectual disability, and behavior problems. Specifically, she awakes at night and cries, her arms are stiff and flexed, her hands are clenched and there may be a coarse tremor of the upper limbs lasting about a minute. Her seizures are more frequent occurring every six months with anticonvulsants, carbamazepine and nitrazepam. Her gross and fine motor functions are normal.

Her facial appearance is different from her sibs, and consists of midfacial hypoplasia with a very low nasal bridge, mild synophrys, and small anteverted nares. II-IV was a very active child however, she did display cognitive delay. Her nocturnal seizures occurred 1-2 times per week but have reduced to once every four months following treatment with valproic acid and nitrazepam.

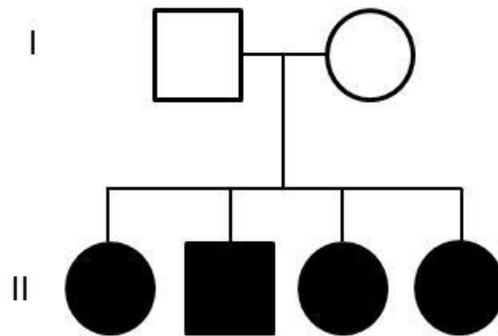


Figure 5.4.1 Pedigree with four children affected with nocturnal seizures with developmental delay.

Squares depict males; circles depict females; affected individuals are shaded.

Table 5.4.1 Clinical features of patients with nocturnal seizures with ID.

Clinical Features	Affected individuals			
	II-1	II-2	II-3	II-4
Gestation			41 weeks	
Delivery complications			Breech, caesarean section	Repeat caesarean section
Sex	F	M	F	F
Karyotype	46,XX	46,XY	46,XX	46,XX
Birth length (cm)				
Birth weight (g)			3,855	
Age at last assessment (years)	30	28	23	21
Head circumference (cm)	53.5 (2nd %ile)	55.8 (25-50th %ile)	53.7 (5th %ile)	53.3 (25th %ile)
Height (cm)	164 (50th %ile)	168 (10th %ile)	150 (<3rd %ile)	159.5 (20th %ile)
Weight (kg)	54.4 (25-50th %ile)	66.5 (25-50th %ile)	46.2 (4th %ile)	64.8 (60th %ile)
Hand length (cm)	17 (25th %ile)	18.5 (75th %ile)	15.8 (3rd %ile)	17 (25th %ile)
Middle finger length (cm)	7.2 (25th %ile)	7.8 (75th %ile)	6.8 (10th %ile)	6.8 (12th %ile)
Ear length (cm)	5.5 (25th %ile)	6 (50th %ile)	5.8 (25-50th %ile)	5.2 (10th %ile)
Inner canthal distance (cm)	3.4 (75th %ile)	3 (50th %ile)	3.3 (75th %ile)	3.3 (75th %ile)
Outer canthal distance (cm)	8.3 (10th %ile)	8.3 (25rd %ile)	8.5 (25th %ile)	8.5 (25th %ile)
Foot length (cm)	23.2 (25th %ile)	25.3 (25th %ile)	21.2 (3rd %ile)	23.3 (25-50th %ile)
Associated diagnoses				
Age to walk	9 months	< 1 year	< 1 year	< 1 year
Age to talk	1 year	< 1 year	4 years	< 1 year
Able to read	+	+	-	-
Age at onset of nocturnal seizures (years)	2	5	3	3
Enamel hypoplasia	+	-	+	+
Palate	N	N	High arched	N
Anterior Overbite	+	-	+	+
Medial deviation of mandibular dentition	+	-	-	+
Tongue hyperpigmentation	+, single location	-	+, extensive	-
Flexion contracture of 5th PIP	+	-	+	+
Hair follicle hyperpigmentation	-	+	+	-
Lymphaedema	-	-	+	-
3+ DTRs	+	+	+	+
Hallux Valgus	+	+	+	+

5.4.2 Compound heterozygous *TMTC3* variants in patients with nocturnal seizures with developmental delay

As the family pedigree is suggestive of an autosomal recessive disorder (Figure 5.4.2 A), a recessive disease model was applied to determine the genetic cause of nocturnal seizures with developmental delay. Through whole exome sequencing, we observed two novel compound heterozygous variants in the affected individuals: c.432G>C, p.Arg71His; and c.2404insA, p.Leu728fs*6 in *TMTC3*, which encodes the transmembrane and tetratricopeptide repeat containing 3 protein. These alleles have not been reported in publically accessible databases of genomic variation, including NCBI dbSNP, 1000 Genomes, NHLBI Exome Variant Server, and the Exome Aggregation Consortium (ExAC). Overall, there are no homozygous *TMTC3* protein truncating variants found in ExAC, which further supports candidate pathogenicity of the *TMTC3* variants.

TMTC3 contains 914 amino acids and is characterized by 9 transmembrane domains and 9 tetratricopeptide repeat (TPR) domains (Figure 5.4.2 B). Multiple *in silico* analyses predicted the p.Arg71His mutation to be detrimental to protein function. This amino acid is conserved across many mammalian species, zebrafish, and *Drosophila*, which further supports the probable pathogenicity of this variant. The p.Leu728fs mutation is predicted to cause truncation of the protein leading to complete loss of the final three TPR domains (Figure 5.4.2 B). TPR domains are critical for protein interactions and are known to be functional only in groups of 3 or 4. These domains represent the main functional component of *TMTC3*, and therefore their loss is likely to have a dramatic impact on the protein. Furthermore, both *TMTC3* variants segregated with disease status in the family (Figure 5.4.2 A). All four children are heterozygous for both variants, whereas the father is only heterozygous for p.Leu728fs and the mother is heterozygous for p.Arg71His.

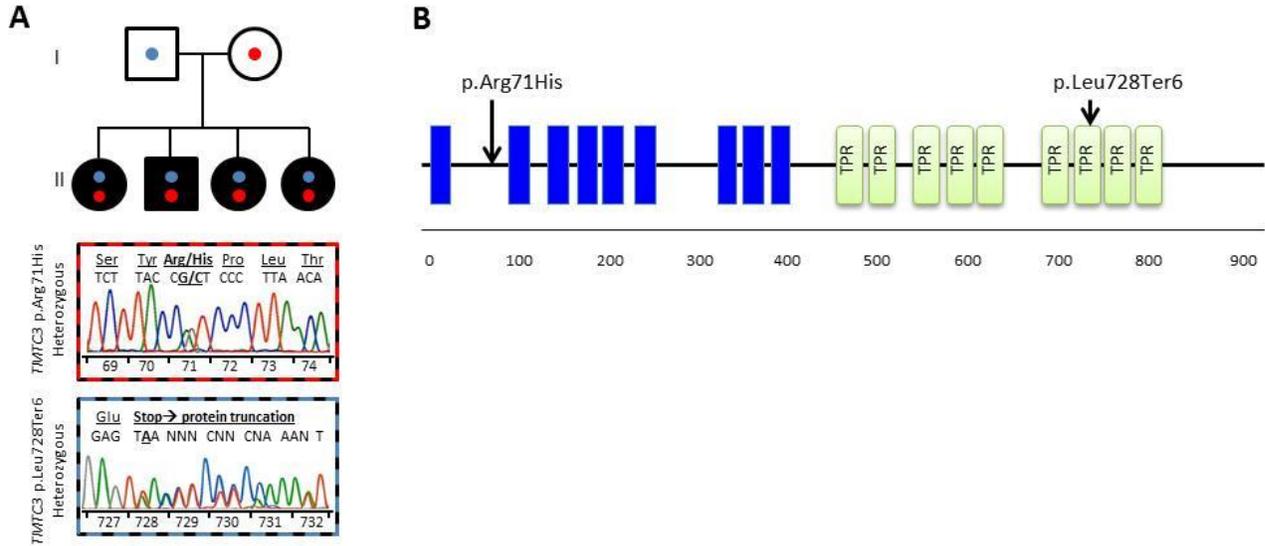


Figure 5.4.2 Compound heterozygous *TMTC3* variants in patients with nocturnal seizures with developmental delay.

(A) Pedigree and electropherograms of the mutation identified in individuals with nocturnal seizures with developmental delay. Different coloured circles in pedigrees represent different mutations. (B) Protein structure and associated domains of *TMTC3* with respective mutations identified.

5.4.3 Loss of TMTC3 in patients with nocturnal seizures and developmental delay

To determine the effect of the *TMTC3* variants on protein and transcript expression, we evaluated *TMTC3* expression levels via immunoblotting and RT-PCR, respectively, using patient derived fibroblast cells. We observed complete loss of *TMTC3* protein in fibroblast cells extracted from two affected individuals relative to both parents and a healthy *TMTC3* wild type control (Figure 5.4.3 A). *TMTC3* transcript expression was detected in patient cells but at significantly lower levels than both parents and a control ($P < 0.01$) (Figure 5.4.3 B). Since the *TMTC3* antibody used here is directed against the C-terminal of the protein, we cannot rule out the production of a truncated protein product from the p.Leu728fs*6 allele, however, our data does demonstrate that any protein product from the p.Arg71His missense mutation is degraded. These findings suggest that the observed *TMTC3* mutations likely underlie nocturnal seizures and developmental delay.

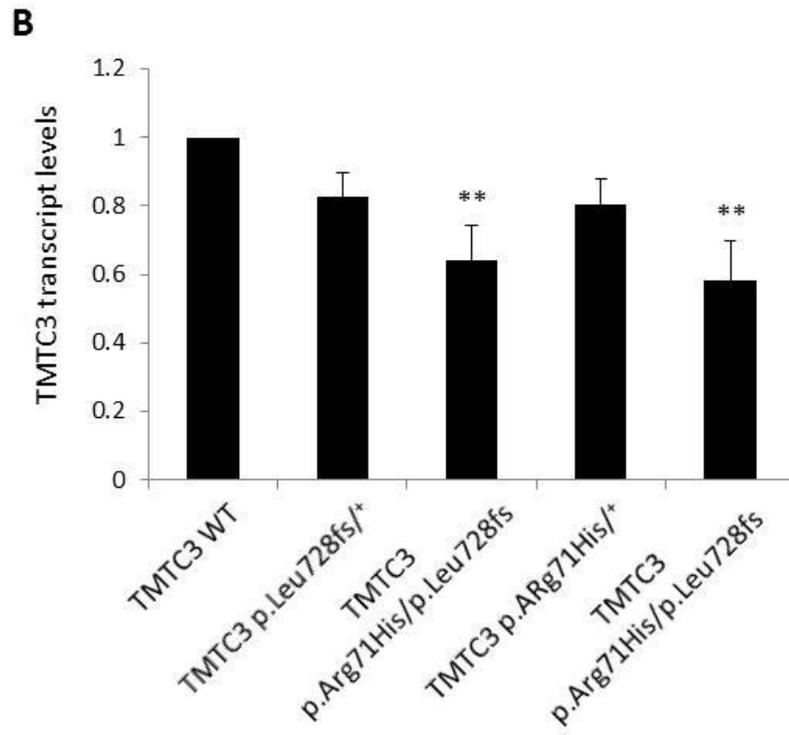
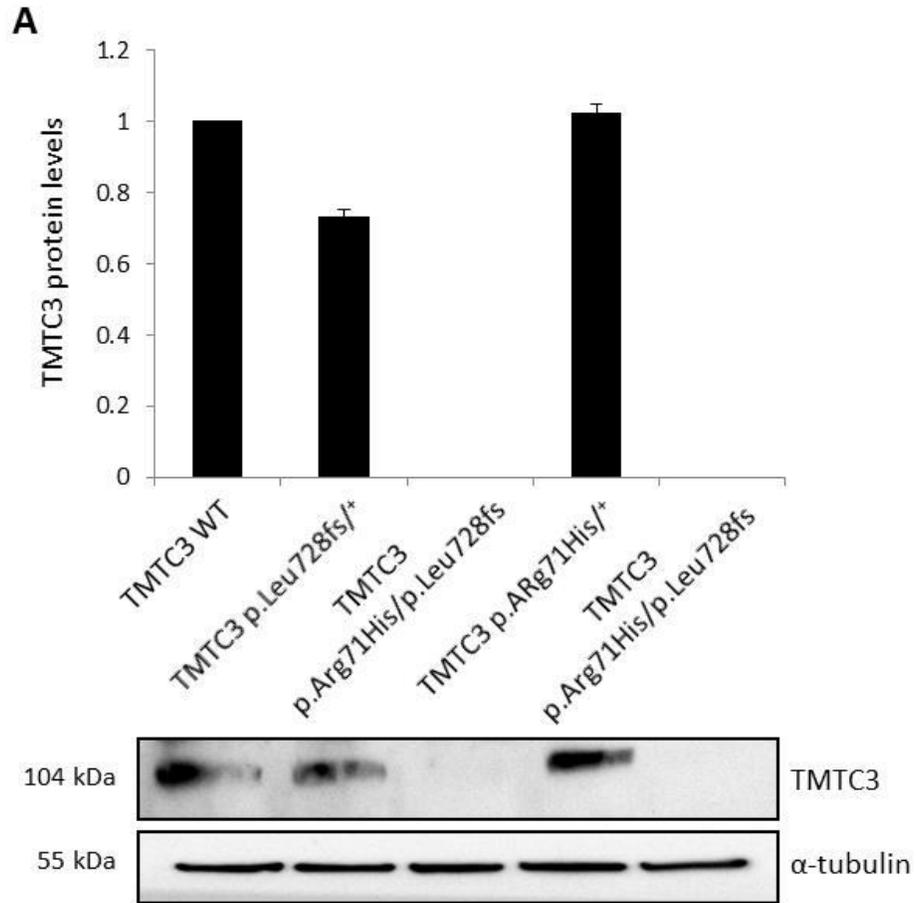


Figure 5.4.3 Depleted TMTC3 protein and reduced transcript expression in cells of patients affected by nocturnal seizures with developmental delay.

(A) Immunoblotting analysis showing depleted TMTC3 protein expression in fibroblast cells of patients. The upper blot shows protein expression of TMTC3, relative to TMTC3 wild type (WT), TMTC3 p.Arg71His/+ (parent 1) and TMTC3 p.Leu728fs/+ (parent 2). The lower blot shows the constitutive expression of α -tubulin in both experiments. (B) RT-PCR show *TMTC3* transcript expression in patients, both parents, and a healthy control. Bar graphs indicate means \pm standard deviations. ** represents *P*-value <0.01.

5.4.4 Neuronal knockdown of *Drosophila tmtc3* causes increased susceptibility to mechanically induced seizures

Having observed loss of the TMTC3 protein in individuals with a neurodevelopmental disorder characterized by nocturnal seizures, we sought to understand the function of TMTC3 in neurons. The TMTC3 protein is highly conserved (60% amino acid similarity and 46% identity) and displays a one-to-one orthology with the *Drosophila* protein encoded by *CG4050*, here referred to as *tmtc3*. We tested two inducible RNAi transgenes encoding inverted repeats (IR) homologous to *Drosophila tmtc3*, *UAS-tmtc3-IR1* and *UAS-tmtc3-IR2*. When RNAi expression was induced using the ubiquitous actin promoter there was a 70% reduction in *tmtc3* mRNA with *UAS-tmtc3-IR1* and only a 20% reduction with *UAS-tmtc3-IR2* (Figure 5.4.4 A), therefore our analysis focuses on *UAS-tmtc3-IR1*.

We investigated the function of *tmtc3* in flies using an established seizure paradigm known as bang sensitivity (Ganetzky and Wu, 1982; Howlett et al., 2013). In this assay seizure like behaviors are induced by mechanical disturbance (vortexing). The response is measured as the amount of time that individual flies are immobilized. To test for a role in mechanically induced seizures, we used flies with a neuron-specific RNAi knockdown of *tmtc3* (*elav-Gal4 + UAS-tmtc3-RNAi*) and compared these to a genetic background control (*elav-Gal4*). Knockdown flies and controls were exposed to an initial round of vortexing (bang 1) followed by a second round (bang 2) after 10 minutes, and the time to recovery was recorded at both time points for each individual fly. Notably, the proportion of flies experiencing immobilization in response to vortexing was increased by more than 2-fold upon neuron specific knockdown of *tmtc3* (Figure 5.4.4 B). Importantly we also tested two rescue genotypes that express human TMTC3 (*UAS-TMTC3-1* and *UAS-TMTC3-2*) in the presence of *UAS-tmtc3-IR1*. Expression of both human transgenes did cause a reduction in the proportion of flies immobilized in response to vortexing (Figure 5.4.4 B). This trend was statistically significant for *UAS-TMTC3-2* (bang 1: $P=0.029$; bang 2: $P=0.004$), and showed a trend approaching statistical significance for *UAS-TMTC3-1* (bang 1: $P=0.062$; bang 2: $P=0.12$). Therefore, it appears that *tmtc3* knockdown in neurons increases susceptibility to mechanically

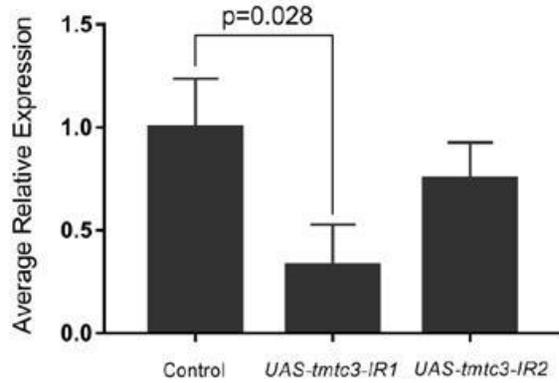
induced seizure in *Drosophila*, and this phenotype can be recovered by transgenic rescue in a humanized fly.

In addition to seizure susceptibility at the population level, we also examined the duration of immobilization in response to vortexing for each individual fly. On average, *tmtc3* knockdown flies had a significantly longer recovery time after bang 2 when compared to controls ($P=0.011$), and showed a trend towards a longer recovery time after bang 1 ($P=0.079$) (Figure 5.4.4 C). The total average recovery time is very short, in the range of 1-2 seconds, because a large number of tested flies in all genotypes (70% to 90%) were not immobilized at all in response to vortexing (Figure 5.4.4 B). When considering only flies that were immobilized, the average recovery time is in the range of 8-12 seconds and there is no significant difference between *tmtc3* knockdown flies, controls, and humanized rescue genotypes. In general, however, controls and humanized rescue genotypes all showed a distinct trend towards decreased recovery time from bang 1 to bang 2, while knockdown flies showed no difference in recovery time between bang 1 and 2 (Figure 5.4.4 D). Therefore, the overall difference in recovery time between *tmtc3* knockdown flies, control flies, and humanized rescue flies (Figure 5.4.4 C) is not due to the severity of immobilization, but rather the higher proportion of knockdown flies that are immobilized in response to vortexing (Figure 5.4.4 B), and their lack of change in recovery time between bang 1 and bang 2 (Figure 5.4.4 D). Taken together, these results suggest that neuron-specific *tmtc3* knockdown affects susceptibility to mechanically induced seizures, and the post-seizure recovery period between bang 1 and bang 2.

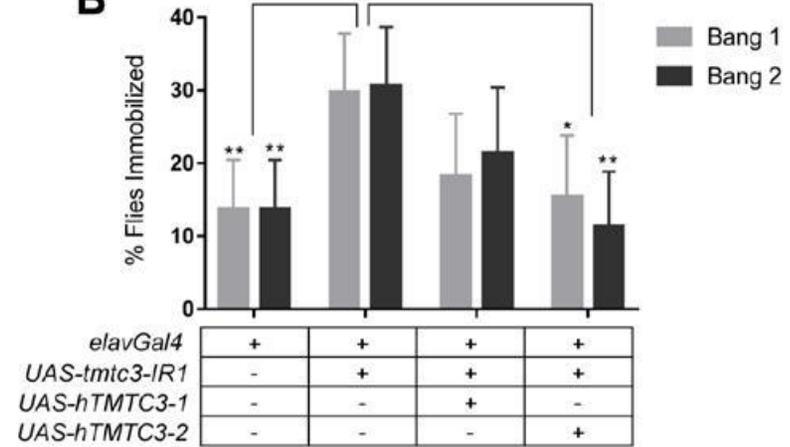
To further elucidate the underlying mechanisms of the observed bang sensitivity phenotypes, we investigated whether the knockdown of *Drosophila tmtc3* may cause changes in neuronal morphology. We identified no morphological differences upon examination of several different neuronal structures in *tmtc3* knockdown flies; including *gros* mushroom body morphology (Figure 5.4.5 A and 5.4.5 B), dendrite arborization in type 4 multidendrite neurons (Figure 5.4.5 C-G), and synaptic morphology at the neuromuscular junction (Figure 5.4.5 H-L). These structures represent model systems for the analysis of synapse, dendrite, and axon morphogenesis, respectively. This suggests that *Drosophila tmtc3* does not have a widespread role in the regulation of neuronal

morphogenesis, and likely influences the response to mechanical stimulation through a role in the regulation of neuronal physiology.

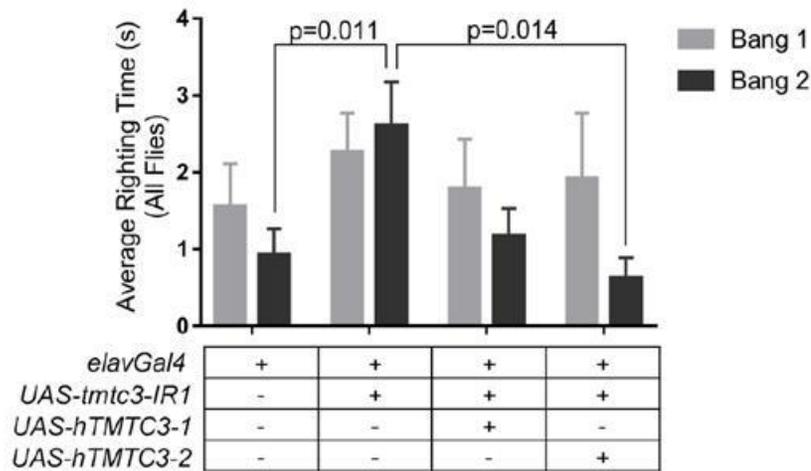
A



B



C



D

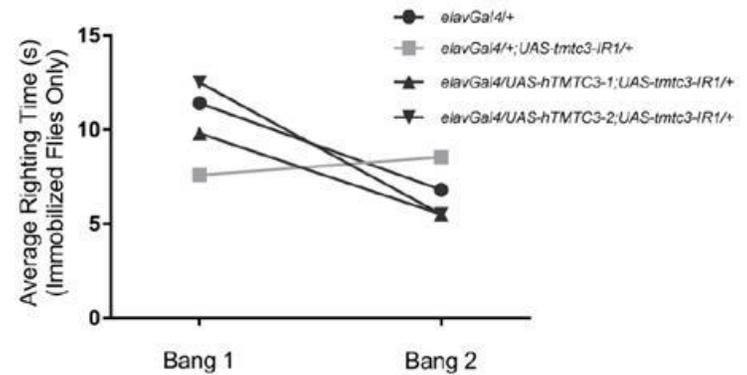


Figure 5.4.4. *Tmtc3* deficiency confers susceptibility to seizures in *Drosophila melanogaster*.

(A) Quantitative PCR results showing the average transcript levels of the *CG4050* gene (\pm SEM) in flies expressing the UAS *tmtc3*-RNAi relative to their respective controls. UAS *tmtc3*-RNAi 1 (M=0.34, SEM=0.06) showed significantly lower expression of the *CG4050* gene than UAS *tmtc3*-RNAi 2 (M=0.82, SEM=0.11) as determined by a Student's t-test ($p < 0.05$). (B) Average time of seizure in affected flies. The UAS *tmtc3*-RNAi 1 flies showed an increase in average seizure time from bang 1 to bang 2 whereas the control and human rescue flies displayed an overall decrease in average seizure time from bang 1 to bang 2. (C) Average righting time (\pm SEM) for male flies of each genotype. The average righting time for the UAS *tmtc3*-RNAi 1 knockdown during bang 2 was significantly longer than that of the control and UAS *tmtc3*-RNAi 1 + UAS *tmtc3* human 2 rescue ($p = 0.011$ and $p = 0.014$, respectively) as determined by a Benjamini-Hochberg corrected Dunn's Test. The UAS *tmtc3*-RNAi 1 + UAS *tmtc3* human 1 genotype demonstrated a decreasing trend in average righting time during bang 2 when compared to the UAS *tmtc3*-RNAi 1 genotype ($p = 0.141$). (D) Percentage of flies that demonstrated seizure-like behavior. For both bang 1 and bang 2, the UAS *tmtc3*-RNAi 1 genotype had a significantly higher proportion of flies that demonstrated seizure-like behavior than the control and UAS *tmtc3*-RNAi 1 + *tmtc3* human 2 genotypes as determined by a two-proportion z-test. The UAS *tmtc3*-RNAi 1 + UAS *tmtc3* human 1 genotype demonstrated a trend-like decrease in seizure-like behavior when compared to the UAS *tmtc3*-RNAi 1 genotype (bang 1: $p = 0.062$, bang 2: $p = 0.115$). * P -value < 0.05 , ** P -value < 0.01 .

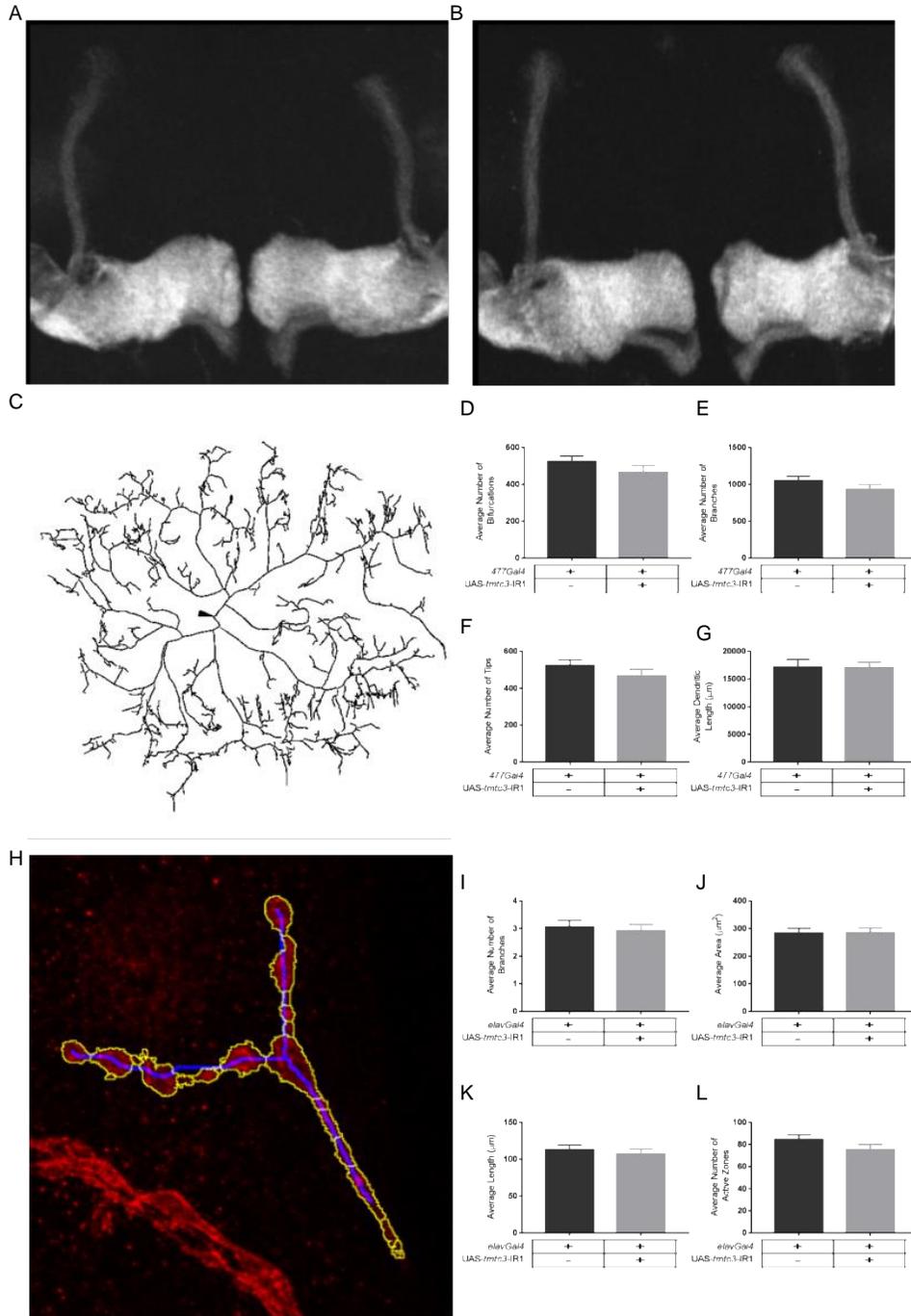


Figure 5.4.5 *Tmtc3* is not implicated in neuromorphogenesis.

(A) Mushroom body of the control fly. (B) Mushroom body of UAS-*tmtc3*-IR1 expressed under the *rutGal4* driver. Expression of UAS-*tmtc3*-IR1 does not affect the development of the mushroom body. (C) Image of a Type IV multidendritic neuron traced in Neuron Studio. (D) Average (\pm SEM) number of bifurcations in the Type IV multidendritic neuron. There is no significant difference between the number of bifurcations in the control flies and the UAS-*tmtc3*-IR1 flies. (E) Average (\pm SEM) number of branches in the Type IV multidendritic neuron. There is no significant difference between the number of branches in the control flies and the UAS-*tmtc3*-IR1 flies. (F) Average (\pm SEM) number of tips in the Type IV multidendritic neuron. There is no significant difference between the number of branches in the control flies and the UAS-*tmtc3*-IR1 flies. (G) Average (\pm SEM) dendritic length of the Type IV multidendritic neuron. There is no significant difference between the average dendritic length in the control flies and the UAS-*tmtc3*-IR1 flies. (H) Image of the neuromuscular junction (NMJ) prepared by the *Drosophila*-_NMJ_Morphometrics FIJI algorithm. (I) Average (\pm SEM) number of branches in the NMJ. There is no significant difference between number of branches in the control flies and the UAS-*tmtc3*-IR1 flies. (J) Average (\pm SEM) area of the NMJ. There is no significant difference between the areas of control NMJ and the UAS-*tmtc3*-IR1 NMJ. (K) Average (\pm SEM) length of the NMJ. There is no significant difference between the lengths of control NMJ and the UAS-*tmtc3*-IR1 NMJ. (L) Average (\pm SEM) number of active zones in the NMJ. There is no difference in the number of active zones between the control NMJ and the UAS-*tmtc3*-IR1 NMJ.

5.4.5 *TMTC3* is localized at presynaptic terminals in rat brains

TMTC3 is a predicted membrane protein containing 9 transmembrane domains and has been shown to partially co-localize with an endoplasmic reticulum marker in cultured odontoblasts (Racape et al., 2011). Considering that TMTC3 has a functional role in *Drosophila* neurons and appears to play a conserved role in seizure susceptibility in both flies and humans, we sought to investigate the localization of TMTC3 at synapses in the mammalian brain, using rat brain sections. We used an anti-TMTC3 antibody directed against the human protein, which is predicted to interact with rat *Tmtc3* based on a 90% amino acid identity between the two proteins. In the rat brain TMTC3 was detectable in the piriform cortex and hypothalamus, but not

the hippocampus (Figure 5.4.6). We performed colocalization analysis with the vesicular GABA transporter (VGAT), a presynaptic marker for inhibitory synapses. VGAT and TMTC3 are colocalized in puncta surrounding cell bodies along the third ventricle (Figure 5.4.6 A-D) and in the piriform cortex (Figure 5.4.6 E). These data show that TMTC3 is localized in a subpopulation of GABAergic synaptic terminals and boutons.

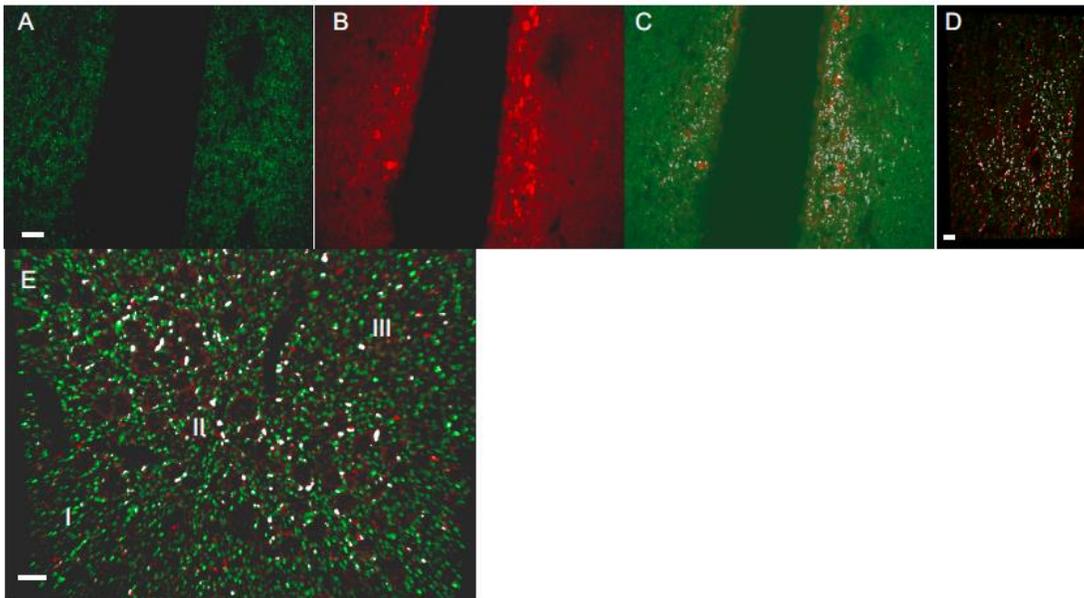


Figure 5.4.6. TMTC3 co-localizes with vesicular GABA transporter (VGAT) in hypothalamus and piriform cortex of rat.

(A) VGAT staining along the third ventricle. (B) TMTC3 immunoreactivity is shown. (C) TMTC3 and VGAT are co-localized in puncta (shown in white) surrounding cell bodies (20X mag scale bar is 20 μm). (D) high magnification co-localization of TMTC3/VGAT at 60 X magnification (scale bar is 10 μm). (E) TMTC3/VGAT co-localize around cell bodies in layer II and III of the piriform cortex.

5.5 Discussion

Here, we provide evidence that disruption of the *TMTC3* gene causes a neurological disorder characterized by nocturnal seizures and ID. Previously, *TMTC3* has not been implicated in any human disease and there is very limited knowledge about its function in the human, *Drosophila*, or rat nervous system. Here, we demonstrate a neuron specific role for *tmtc3* in *Drosophila* seizure susceptibility, and show that *Tmtc3* is localized at synapses in the rat brain. This data is in keeping with the current state of knowledge regarding the role of synaptic ion channels in the etiology of nocturnal epilepsy.

Nocturnal seizures are most commonly described in the context of ADFLE and a spectrum of other disorders falling under the classification of FESD, which include Landau-Kleffner syndrome (LKS), continuous spike and waves during slow-wave sleep syndrome (CSWSS), and Rolandic epilepsy (Bazil, 2004). Our current knowledge indicates that these disorders are caused primarily by dominant heterozygous inherited or *de novo* mutations in components of gated synaptic ion channels (Aridon et al., 2006; De Fusco et al., 2000; Steinlein et al., 1995). Here, we describe an autosomal recessive inheritance pattern for nocturnal seizures resulting in loss of the *TMTC3* protein. The identified variants have not been reported in any publicly accessible database of genetic variation and our whole exome analysis did not reveal any other plausible candidate genetic variants in the four affected siblings. The ExAC database does not report a single example of homozygous loss of function or damaging missense mutations in *TMTC3*, indicating that disruption of the *TMTC3* protein is likely to be detrimental to cell function. Together, these data suggest that we have identified a novel, rare, autosomal recessive disorder characterized by nocturnal epilepsy and ID.

In addition to the compelling genetic and molecular evidence, we have functionally implicated *TMTC3* in seizure susceptibility in *Drosophila*. The “bang sensitivity” test employed here is a classic assay for mechanically induced seizure, characterized by immobilization in response to a strong mechanical stimulus (Song and Tanouye, 2008). Loss of *Drosophila tmtc3* specifically in post-mitotic neurons nearly doubles the rate of immobilization in response to vortexing, suggesting that loss of *tmtc3* renders neurons susceptible to seizure. Here, we have created a humanized fly model and

demonstrated that fly seizure phenotypes resulting from *tmtc3* knockdown can be eliminated by expression of a human *TMTC3* transgene. Therefore, although the seizure paradigm employed here is not directly comparable to the human condition, we can conclude that the human protein is functionally relevant in the context of mechanically induced seizures in *Drosophila*. These data suggest that *TMTC3* has an evolutionarily conserved role in the regulation of neuronal physiology and that disruption of this protein family leads to seizure susceptibility.

Genetic and functional studies have shown that disruption or dysregulation of synaptic ion channels is an important mechanism in the etiology of nocturnal seizures (Becchetti et al., 2015; Lemke et al., 2013; Lesca et al., 2013; Nobili et al., 2014; Venkateswaran et al., 2014). Here, we show that *TMTC3* is localized at inhibitory GABAergic synapses in the hypothalamus and piriform cortex of the rat brain. Interestingly, the piriform cortex is predicted to play an important role in facilitating and amplifying seizures in focal epilepsy (Vaughan and Jackson, 2014). It is possible that *TMTC3* is required at these inhibitory synapses in the piriform cortex to prevent uncontrolled neuronal excitation that is a hallmark of epilepsy, but its molecular role at the synapse remains unclear. *TMTC3* contains 9 transmembrane domains, and 9 tetratricopeptide (TPR) domains. TPR-domains mediate protein-protein interactions and are involved in a wide array of biological functions (D'Andrea and Regan, 2003).

Interestingly, the *TMTC3* paralogs, *TMTC1* and *TMTC2*, have been predicted to act as adaptors that nucleate the formation of large protein complexes involved in the regulation of calcium transport across the ER membrane (Sunryd et al., 2014). Based on the similar domain structure seen in *TMTC3*, it is possible that this protein plays a similar role at the synapse, and that disruption of *TMTC3* would affect neuronal physiology by disruption of normal ion transport mechanisms. This possibility is supported by protein-protein interaction data for *TMTC3* from several proteomic studies (Gupta et al., 2015; Hein et al., 2015; Huttlin et al., 2015; Rolland et al., 2014), which are annotated through the BioGRID database (Chatr-Aryamontri et al., 2015). Of the 21 *TMTC3* interactors listed in BioGRID, six are involved in gated ion channel activity (*CHRND*, *HTR3C*, *HTR3A*, *ZACN*, *GABRE*, and *SCN3B*) and five are localized at the synapse (*CHRND*, *HTR3A*, *NRG1*, *GABRE*, and *DAB1*) (Annotations from PANTHER) (Mi et al., 2005).

Thus, in keeping with the known genetic etiology of nocturnal seizures, there is a body of evidence, including the protein localization demonstrated here, that supports a role for TMTC3 in the regulation of synaptic ion channels.

5.6 Conclusion

In summary, we have implicated loss of TMTC3 as novel cause of autosomal recessive nocturnal seizures and ID, and have revealed a previously unknown and conserved function for this family of proteins in seizure susceptibility. Our data, as well as publicly available proteomic data, support a role for TMTC3 in synaptic ion channel activity, providing a basis for further investigation of the role of this protein family in neuronal physiology.

5.7 References

- Aridon, P., Marini, C., Di Resta, C., Brilli, E., De Fusco, M., Politi, F., Parrini, E., Manfredi, I., Pisano, T., Pruna, D., *et al.* (2006). Increased sensitivity of the neuronal nicotinic receptor alpha 2 subunit causes familial epilepsy with nocturnal wandering and ictal fear. *Am J Hum Genet* 79, 342-350.
- Bazil, C.W. (2004). Nocturnal seizures. *Semin Neurol* 24, 293-300.
- Becchetti, A., Aracri, P., Meneghini, S., Brusco, S., and Amadeo, A. (2015). The role of nicotinic acetylcholine receptors in autosomal dominant nocturnal frontal lobe epilepsy. *Front Physiol* 6, 22.
- Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., *et al.* (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 43, D470-478.
- D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem Sci* 28, 655-662.
- De Fusco, M., Becchetti, A., Patrignani, A., Annesi, G., Gambardella, A., Quattrone, A., Ballabio, A., Wanke, E., and Casari, G. (2000). The nicotinic receptor beta 2 subunit is mutant in nocturnal frontal lobe epilepsy. *Nat Genet* 26, 275-276.
- Farhan, S.M., Murphy, L.M., Robinson, J.F., Wang, J., Siu, V.M., Rupa, C.A., Prasad, A.N., Consortium, F.C., and Hegele, R.A. (2014). Linkage analysis and exome

sequencing identify a novel mutation in KCTD7 in patients with progressive myoclonus epilepsy with ataxia. *Epilepsia* 55, e106-111.

Ganetzky, B., and Wu, C.F. (1982). Indirect Suppression Involving Behavioral Mutants with Altered Nerve Excitability in *DROSOPHILA MELANOGASTER*. *Genetics* 100, 597-614.

Gupta, G.D., Coyaud, E., Goncalves, J., Mojarad, B.A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J.M., Cheung, S.W., *et al.* (2015). A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* 163, 1484-1499.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., *et al.* (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712-723.

Heron, S.E., Smith, K.R., Bahlo, M., Nobili, L., Kahana, E., Licchetta, L., Oliver, K.L., Mazarib, A., Afawi, Z., Korczyn, A., *et al.* (2012). Missense mutations in the sodium-gated potassium channel gene KCNT1 cause severe autosomal dominant nocturnal frontal lobe epilepsy. *Nat Genet* 44, 1188-1190.

Howlett, I.C., Rusan, Z.M., Parker, L., and Tanouye, M.A. (2013). *Drosophila* as a model for intractable epilepsy: gilgamesh suppresses seizures in para(bss1) heterozygote flies. *G3 (Bethesda)* 3, 1399-1407.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., *et al.* (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162, 425-440.

Jenett, A., Rubin, G.M., Ngo, T.T., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B.D., Cavallaro, A., Hall, D., Jeter, J., *et al.* (2012). A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep* 2, 991-1001.

Kim, G.E., Kronengold, J., Barcia, G., Quraishi, I.H., Martin, H.C., Blair, E., Taylor, J.C., Dulac, O., Colleaux, L., Nabbout, R., *et al.* (2014). Human slack potassium channel mutations increase positive cooperativity between individual channels. *Cell Rep* 9, 1661-1672.

Lemke, J.R., Lal, D., Reinthaler, E.M., Steiner, I., Nothnagel, M., Alber, M., Geider, K., Laube, B., Schwake, M., Finsterwalder, K., *et al.* (2013). Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nat Genet* 45, 1067-1072.

Lesca, G., Rudolf, G., Bruneau, N., Lozovaya, N., Labalme, A., Boutry-Kryza, N., Salmi, M., Tsintsadze, T., Addis, L., Motte, J., *et al.* (2013). GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat Genet* 45, 1061-1066.

- Lim, C.X., Ricos, M.G., Dibbens, L.M., and Heron, S.E. (2016). KCNT1 mutations in seizure disorders: the phenotypic spectrum and functional effects. *J Med Genet* *53*, 217-225.
- Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., *et al.* (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res* *33*, D284-288.
- Myers, C.T., and Mefford, H.C. (2015). Advancing epilepsy genetics in the genomic era. *Genome Med* *7*, 91.
- Nijhof, B., Castells-Nobau, A., Wolf, L., Scheffer-de Gooyert, J.M., Monedero, I., Torroja, L., Coromina, L., van der Laak, J.A., and Schenck, A. (2016). A New Fiji-Based Algorithm That Systematically Quantifies Nine Synaptic Parameters Provides Insights into *Drosophila* NMJ Morphometry. *PLoS Comput Biol* *12*, e1004823.
- Nobili, L., Proserpio, P., Combi, R., Provini, F., Plazzi, G., Bisulli, F., Tassi, L., and Tinuper, P. (2014). Nocturnal frontal lobe epilepsy. *Curr Neurol Neurosci Rep* *14*, 424.
- Racape, M., Duong Van Huyen, J.P., Danger, R., Giral, M., Bleicher, F., Foucher, Y., Pallier, A., Pilet, P., Tafelmeyer, P., Ashton-Chess, J., *et al.* (2011). The involvement of SMILE/TMTC3 in endoplasmic reticulum stress response. *PLoS One* *6*, e19321.
- Rolland, T., Tasan, M., Charlotiaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., *et al.* (2014). A proteome-scale map of the human interactome network. *Cell* *159*, 1212-1226.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* *9*, 676-682.
- Scorcioni, R., Polavaram, S., and Ascoli, G.A. (2008). L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nat Protoc* *3*, 866-876.
- Shiba, Y., Mori, F., Yamada, J., Migita, K., Nikaido, Y., Wakabayashi, K., Kaneko, S., Okada, M., Hirose, S., and Ueno, S. (2015). Spontaneous epileptic seizures in transgenic rats harboring a human ADNFLE missense mutation in the beta2-subunit of the nicotinic acetylcholine receptor. *Neurosci Res* *100*, 46-54.
- Song, J., and Tanouye, M.A. (2008). From bench to drug: human seizure modeling using *Drosophila*. *Prog Neurobiol* *84*, 182-191.
- Steinlein, O.K., Mulley, J.C., Propping, P., Wallace, R.H., Phillips, H.A., Sutherland, G.R., Scheffer, I.E., and Berkovic, S.F. (1995). A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. *Nat Genet* *11*, 201-203.

Sunryd, J.C., Cheon, B., Graham, J.B., Giorda, K.M., Fissore, R.A., and Hebert, D.N. (2014). TMTC1 and TMTC2 are novel endoplasmic reticulum tetratricopeptide repeat-containing adapter proteins involved in calcium homeostasis. *J Biol Chem* 289, 16085-16099.

Thomas, R.H., and Berkovic, S.F. (2014). The hidden genetics of epilepsy-a clinically important new paradigm. *Nat Rev Neurol* 10, 283-292.

Vaughan, D.N., and Jackson, G.D. (2014). The piriform cortex and human focal epilepsy. *Front Neurol* 5, 259.

Venkateswaran, S., Myers, K.A., Smith, A.C., Beaulieu, C.L., Schwartzentruber, J.A., Consortium, F.C., Majewski, J., Bulman, D., Boycott, K.M., and Dymment, D.A. (2014). Whole-exome sequencing in an individual with severe global developmental delay and intractable epilepsy identifies a novel, de novo GRIN2A mutation. *Epilepsia* 55, e75-79.

Wearne, S.L., Rodriguez, A., Ehlenberger, D.B., Rocher, A.B., Henderson, S.C., and Hof, P.R. (2005). New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience* 136, 661-680.

Yun, E.J., and Vu, T.H. (2012). mSmile is necessary for bronchial smooth muscle and alveolar myofibroblast development. *Anat Rec (Hoboken)* 295, 167-176.

Chapter 6 - The ONDRISeq panel: custom designed next generation sequencing of genes related to neurodegeneration as part of the Ontario Neurodegenerative Disease Research Initiative

6 Study rationale

The next two Chapters describe the application of next generation sequencing in identifying the genetic basis of complex diseases including familial and sporadic forms of neurological diseases.

As described in Chapter 1, within ONDRI, ONDRI Genomics focuses on identifying genetic variants that may contribute to disease onset and progression. As a member of the ONDRI Genomics team, which is led by Dr. Hegele, I was directly involved in the development of the ONDRISeq panel, which is a custom designed next generation sequencing based on neurodegeneration genes specific to the ONDRI diseases. This included the development, optimization, and integration of ONDRISeq in the ONDRI study. Chapter 6 was prepared as an independent study to describe the methodologies used within ONDRI genomics and the manuscript: “The ONDRISeq panel: custom designed next generation sequencing of genes related to neurodegeneration” has been accepted for publication in *npj Genomic Medicine*. Also, the methods describing ONDRISeq have been successfully filed with the United States Patent and Trademark Office in 2016: ‘Neurodegeneration targeted resequencing panel’. Inventors: Michael Strong, John Robinson, Sali Farhan, Robert Hegele, and Tom Mikkelsen. United States Provisional Patent Application ID: 62/333,369.

6.1 Overview

The Ontario Neurodegenerative Disease Research Initiative (ONDRI) is a multimodal, multi-year, prospective observational cohort study to characterize five diseases: 1)

Alzheimer's disease (AD) or amnesic single or multidomain mild cognitive impairment (aMCI) (AD/MCI); 2) amyotrophic lateral sclerosis (ALS); 3) frontotemporal dementia (FTD); 4) Parkinson's disease (PD); and 5) vascular cognitive impairment (VCI). The ONDRI Genomics subgroup is range of genes associated with neurodegeneration for their contribution to each of these disease processes, and most specifically to the occurrence of neuropsychological dysfunction. We have developed a custom next generation sequencing based panel, ONDRISeq that targets 80 genes known to be associated with neurodegeneration. We processed DNA collected from 216 individuals diagnosed with one of the five diseases, on ONDRISeq. All runs were executed on a MiSeq and subjected to rigorous quality control assessments. We also independently validated a subset of the variant calls using NeuroX (a genome-wide array for neurodegenerative disorders), TaqMan allelic discrimination assay, or Sanger sequencing. ONDRISeq consistently generated high quality data and all targets were sufficiently covered with a mean coverage of 76x. We also observed 100% concordance for the variants identified via ONDRISeq and validated by other genomic technologies. We were successful in detecting known as well as novel rare variants in 72.2% of cases although not all variants are disease-causing. Using ONDRISeq, we also found that the APOE E4 allele had a frequency of 0.167 in these samples. Our optimized workflow highlights next generation sequencing as a robust tool in elucidating the genetic basis of neurodegenerative diseases by screening multiple candidate genes simultaneously.

6.2 Introduction

Dementia encompasses a heterogeneous group of neurodegenerative diseases characterized by a progressive decline in cognitive function, language deficiency, and in some cases, motor impairment and behavioural anomalies. Currently, dementia has a global prevalence of 47.5 million cases and an incidence of 7.7 million new cases annually (2008; 2015; Robinson et al., 2015). While today there are no direct treatments available to alter the progressive disease course, early diagnosis has been one of the best predictors of disease outcome (Hebert et al., 2013; Robinson et al., 2015). Further

understanding of the molecular basis of dementia can lead to earlier diagnosis and the eventual development of targeted and efficacious treatment modalities.

Genetics is an important risk factor for neurodegenerative disease. Approximately 5-10% of cases with neurodegenerative diseases are familial and can be attributed to variants in several genes (Guerreiro et al., 2014; Renton et al., 2014; Rohrer et al., 2015). However, it is likely we are underestimating the incidence of familial cases based on clinical ascertainment, as the death of pre-symptomatic individuals may be due to other medical or extra-health incidents prior to the development of the neurodegenerative syndrome. Furthermore, genetic testing is not universally recommended in the clinical management guidelines of neurodegenerative diseases (Goldman et al., 2011; Grimes et al., 2012; Jack et al., 2011; Miller et al., 2009a, b; Strong et al., 2009). As such, most neurologists, if they choose to pursue genetic testing, only screen for a small subset of genes and often choose to genotype their patients for highly penetrant and known variants rather than agnostically sequencing all neurodegenerative disease genes. Together, these common clinical ascertainment as well as the high costs associated with genetic testing skew the incidence rates to significantly less than what is perhaps biologically accurate. The five neurodegenerative disorders under study could partly be caused by single, rare, pathogenic variants (monogenic) or multiple, small effect variants acting synergistically to mediate disease expression (oligogenic).

Advancements in next generation sequencing (NGS) have allowed for efficient genetic variant detection at reduced costs. Currently, there are three main types of NGS applications including: 1) whole genome sequencing (WGS); 2) whole exome sequencing (WES); and 3) targeted gene panels (Farhan and Hegele, 2014). WGS is an indiscriminate approach that evaluates the genetic information in an individual's entire genome. In contrast, WES targets only the protein-coding regions of the genome as disease-associated variants are significantly over-represented in coding regions (Farhan and Hegele, 2014). Consequently, WES has been one of the most widely used NGS approaches, however it still presents with several challenges. First, the cost of WES with adequate coverage (i.e. minimum 30x) still remains high at approximately \$700. This makes the cumulative cost for studies with a large sample size prohibitively expensive. Secondly, the amount of genetic variation generated from the exome is excessive and

often overwhelming for many researchers and more so for clinicians who may require the patient's genetic diagnosis to determine if any genotype specific treatments are available. Thirdly, WES can generate secondary findings unrelated to the disease of interest, which should be reported to the patient's primary healthcare provider, in accordance with the guidelines proposed by the American College of Medical Genetics (Richards et al., 2015). Thus, in both a clinical and research application, WGS or WES data are still often reduced to focus on likely pathogenic disease specific loci. In contrast, the use of a targeted gene panel that is clinically focused on the genes underlying the disease(s) of interest, overcomes these issues that often arise when sifting through WGS and WES data.

6.3 Materials and methods

6.3.1 Design of ONDRISeq

Using multiple databases, we catalogued the literature pertaining to neurodegeneration genetic studies. We surveyed 25 content experts (professors, scientists, and clinicians within ONDRI as well as my input) in molecular genetics of neurodegeneration, and used their consensus opinions to select 80 genes within the human genome that were involved in one or more of the five neurodegenerative disorders under study. Most genes were selected based on being implicated in neurodegeneration from human genetic studies; however, some of the genes were added based on pathway analysis.

We designed a composition for detecting variants in the protein-coding regions of 80 genes summing to 1,649 targets. The 80 genes selected have a total target size of 972,388 base pairs. Using the NGS chemistry Nextera Rapid Custom Capture (Illumina, San Diego, CA), we designed a total of 14,510 target specific probes that are each ~80 base pairs in length. For regions that were difficult to sequence, we incorporated additional probes to ensure sufficient coverage during sequencing thereby producing fewer false discoveries. Chromosome scaffold coordinates were obtained from the University of California Santa Cruz Genome Browser using the February 2009

GRCh37/hg19 genome build (19) and were submitted to the Illumina Online Design Studio (Illumina, San Diego, CA).

Table 6.3.1 Genes associated with amyotrophic lateral sclerosis, frontotemporal dementia, Alzheimer’s disease, Parkinson’s disease, or vascular cognitive impairment as represented on the ONDRISeq targeted resequencing panel.

Gene	Chromosomal location	Affected protein	Associated phenotype	Mode of inheritance	OMIM numbers (locus, phenotype)
Amyotrophic lateral sclerosis/frontotemporal dementia					
<i>ALS2</i>	2q33.1	Alsin	ALS2	AR (HZ), juvenile onset	606352, 205100
<i>ANG</i>	14q11.2	Angiogenin	ALS9	ADm, late onset	105850, 611895
<i>ARHGEF28</i>	5q13.2	Rho guanine nucleotide exchange factor 28	ALS and FTD	AR (HZ) and ADm, late onset	612790, PMID: 23286752 (phenotype not updated on OMIM)
<i>ATXN2</i>	12q24.12	Ataxin 2	ALS13	ADm, late onset	601517, 183090
<i>CENPV</i>	17p11.2	Centromere protein V	ALS	Genetic association, late onset	608139, PMID: 22959728 (phenotype not updated on OMIM)
<i>CHMP2B</i>	3p11.2	CHMP family member 2B	ALS17, FTD	ADm, late onset	609512, 614696
<i>DAO</i>	12q24.11	D-amino acid oxidase	ALS, schizophrenia	ADm, late onset	124050, 105400, 181500
<i>DCTN1</i>	2p13.1	Dynactin 1	ALS, HMN7B, Perry syndrome	ADm, late onset	601143, 105400, 607641, 168605
<i>FIG4</i>	6q21	FIG4 homolog, SAC1 lipid phosphatase domain containing	ALS11, CMT disease, YV syndrome	ADm, late onset; AR (HZ and CH), infantile onset; AR (HZ and CH), infantile onset	609390, 612577, 611228, 216340
<i>FUS</i>	16p11.2	Fused in sarcoma	ALS6, FTD, HET4	AR (HZ), ADm, late onset	137070, 608030, 614782
<i>GRN</i>	17q21.31	Granulin precursor	FTD, NCL	ADm, late onset; AR (HZ), juvenile onset	138945, 607485, 614706
<i>HNRNPA1</i>	12q13.13	Heterogeneous nuclear ribonucleoprotein A1	ALS20, inclusion body myopathy with early-onset Paget disease with/without FTD 3	ADm, late onset; ADm, early onset	164017, 615426, 615424
<i>HNRNPA2B1</i>	7p15.2	Heterogeneous nuclear ribonucleoprotein A2/B1	Inclusion body myopathy with early-onset Paget disease with/without FTD 2	ADm, early onset	600124, 615422
<i>MAPT/STH</i>	17q21.31	Microtubule-associated protein tau	ALS, FTD with parkinsonism, PD, AD, Pick disease, supranuclear palsy, tauopathy	ADm, late and early onset	157140, 105400, 600274, 168600, 104300, 172700, 601104, 260540
<i>NEFH</i>	22q12.2	Neurofilament protein, heavy	ALS1	ADm, late onset	162230, 105400

		polypeptide			
<i>OPTN</i>	10p13	Optineurin	ALS12, glaucoma	AR (HZ) and AD, early onset	602432, 613435, 606657
<i>PFN1</i>	17p13.2	Profilin 1	ALS18	ADm, earlier onset	176610, 614808
<i>PNPLA6</i>	19p13.2	Patatin-like phospholipase domain-containing protein 6	Spastic paraplegia, Boucher-Neuhauser syndrome	AR (HZ and CH), early onset	603197, 612020, 215470
<i>PRPH</i>	12q13.12	Peripherin	ALS1	ADm, late onset	170710, 105400
<i>SETX</i>	9q34.13	Senataxin	ALS4, spinocerebellar ataxia 1	ADm and AR, juvenile onset	608465, 602433, 606002
<i>SIGMAR1</i>	9p13.3	Sigma nonopioid intracellular receptor 1	ALS16, FTD	AR (HZ); ADm, early onset	601978, 614373, 105550
<i>SOD1</i>	21q22.11	Superoxide dismutase 1	ALS1	AR (HZ and CH), ADm, age of onset varies from 6-94 years old	147450, 105400
<i>SQSTM1</i>	5q35.3	Sequestosome 1	Paget disease of bone	ADm, late onset	601530, 167250
<i>TAF15</i>	17q12	TAF15 RNA polymerase II, TATA box binding protein associated factor	Chondrosarcoma		601574, 612237
<i>TARDBP</i>	1p36.22	Tar DNA-binding protein	ALS10, FTD	ADm, late onset	605078, 612069
<i>UBQLN2</i>	Xp11.21	Ubiquilin 2	ALS15, FTD	X-linked, juvenile and late onset	300264, 300857
<i>UNC13A</i>	19p13.11	Unc-13 homolog A (<i>C. elegans</i>)	ALS	Genetic association, late onset	609894, PMID: 22921269 (phenotype not updated on OMIM)
<i>VAPB</i>	20q13.33	Vesicle-associated membrane protein (VAMP)-associated protein B and C	ALS, spinal muscular atrophy (Finkel type)	ADm, early and late onset	605704, 608627, 182980
<i>VCP</i>	9p13.3	Valosin-containing protein	ALS14, FTD, inclusion body myopathy with early-onset Paget disease with/without FTD 1	ADm, early onset	601023, 613954, 167320
Alzheimer's disease/mild cognitive impairment					
<i>ABCA7</i>	19p13.3	ATP-binding cassette, subfamily a, member 7	AD	Genetic association, late onset	605414, 104300
<i>APOE</i>	19q13.32	Apolipoprotein E	AD2, lipoprotein glomerulopathy, sea-blue hystiocyte disease, macular degeneration	ACD, ADm, AR (HZ and CH), late onset	107741, 104310, 611771, 269600, 603075
<i>APP</i>	21q21.3	Amyloid beta A4 precursor protein	AD 1, cerebral amyloid angiopathy	ADm and AR (HZ), early and late onset	104760, 104300, 605714
<i>BINI</i>	2q14.3	Bridging integrator 1	AD	Genetic association, late onset	601248, PMID: 25365775 (phenotype not updated on OMIM)

<i>CD2AP</i>	6p12.3	CD2-associated protein	AD	Genetic association, late onset	604241, PMID: 25092125 (phenotype not updated on OMIM)
<i>CD33</i>	19q13.41	CD33 antigen	AD	Genetic association, late onset	159590, PMID: 23982747 (phenotype not updated on OMIM)
<i>CLU</i>	8p21.1	Clusterin	AD	Genetic association, late onset	185430, PMID: 25189118 (phenotype not updated on OMIM)
<i>CRI</i>	1q32.2	Complement component receptor 1	AD	Genetic association, late onset	120620, PMID: 25022885 (phenotype not updated on OMIM)
<i>CSF1R</i>	5q32	Colony-stimulating factor 1 receptor	HDLS with dementia	ADm, early and late onset	164770, 221820
<i>DNMT1</i>	19p13.2	DNA methyltransferase 1	HSN1E with dementia	ADm, early onset dementia	126375, 614116
<i>ITM2B</i>	13q14.2	Integral membrane protein 2B	Dementia	ADm, early and late onset	603904, 176500, 117300
<i>MS4A4E</i>	11q12.2	Membrane-spanning 4-domains, subfamily A, member 4E	AD	Genetic association, late onset	608401, PMID: 21460840 (phenotype not updated on OMIM)
<i>MS4A6A</i>	11q12.2	Membrane-spanning 4-domains, subfamily A, member 6A	AD	Genetic association, late onset	606548, PMID: 21460840 (phenotype not updated on OMIM)
<i>PICALM</i>	11q14.2	Phosphatidylinositol-binding clathrin assembly protein	AD	Genetic association, late onset	603025, PMID: 24613704 (phenotype not updated on OMIM)
<i>PLD3</i>	19q13.2	Phospholipase D family, member 3	AD19	Genetic association, late onset	615698, 615711
<i>PSEN1</i>	14q24.2	Presenilin 1	AD3, dilated cardiomyopathy, FTD, Pick disease, acne inversa	ADm, early onset	104311, 607822, 613694, 600274, 172700, 613737
<i>PRNP</i>	20p13	Prion protein	Dementia	ADm, early onset	176640, 606688
<i>PSEN2</i>	1q32.13	Presenilin 2	AD4, dilated cardiomyopathy	ADm, early onset	600759, 606889, 613697
<i>SORL1</i>	11q24.1	Sortilin-related receptor	AD	ADm, combined gene burden, late onset	602005, 104300; PMID: 25382023 (phenotype not updated on OMIM)
<i>TREM2</i>	6p21.1	Triggering receptor expressed on myeloid cells 2	AD Nasu-Hakola disease (dementia and psychotic symptoms)	Genetic association, late onset	605086, PMID: 25596843 (phenotype not updated on OMIM), 221770
<i>TYROBP</i>	19q13.12	Tyrosine protein tyrosine kinase-binding protein	Nasu-Hakola disease (dementia and psychotic symptoms)	AR (HZ), juvenile onset	604142, 221770

Parkinson's disease					
<i>ADHIC</i>	4q23	Alcohol dehydrogenase 1C, gamma polypeptide	PD, alcohol dependence protection	Genetic association, late onset	103730, 168600, 103780
<i>ATP13A2 (PARK9)</i>	1p36.13	ATPase, type 13A2	PD, ceroid lipofuscinosis, dementia	Genetic association, early onset and late onset	610513, 606693
<i>DNAJC13</i>	3q22.1	DNAJ/HSP40 homolog, subfamily C, member 13	PD	ADm, late onset	614334, PMID: 25330418 (phenotype not updated on OMIM)
<i>EIF4G1</i>	3q27.1	Eukaryotic translation initiation factor 4-gamma	PD18	ADm, late onset	600495, 614251
<i>FBXO7</i>	22q12.3	F-box only protein 7	PD15	AR (HZ and CH), early onset	605648, 260300
<i>GAK</i>	4p16.3	Cyclin G-associated kinase	PD	Genetic association, late onset	602052, PMID: 21258085 (phenotype not updated on OMIM)
<i>GCH1</i>	14q22.2	GTP cyclohydrolase I	PD, dystonia	Genetic association, early onset	600225, 128230
<i>GIGYF2</i>	2q37.1	GRB10-interacting GYP protein 2	PD11	Genetic association, early and late onset	612003, 607688
<i>HTRA2</i>	2p13.1	HTRA serine peptidase 2	PD13	ADm and genetic association, early and late onset	606441, 610297
<i>LRRK2</i>	12q12	Leucine-rich repeat kinase 2	PD8	ADm and genetic association, early and late onset	609007, 607060
<i>MC1R</i>	16q24.3	Melanocortin 1 receptor	PD; melanoma, UV induced skin damage	Genetic association, late onset	155555, 613099, 266300, 168600
<i>NR4A2</i>	2q24.1	Nuclear receptor subfamily 4, group A, member 2	PD	Genetic association, late onset	601828, 168600
<i>PANK2</i>	20p13	Pantothenate kinase 2	Neurodegeneration	AR (HZ and CH), early onset	606157, 234200
<i>PARK2 (PRKN)</i>	6q26	Parkin	PD2	AR (HZ and CH), juvenile onset; heterozygotes have late onset	602544, 600116
<i>PARK7(DJ1)</i>	1p36.23	Oncogene DJ1	PD7	AR (HZ and CH), early onset	602533, 606324
<i>PARL</i>	3q27.1	Presenilin-associated rhomboid-like protein	PD (based on biological mechanisms, no linkage confirmed)	NA	607858, PMID: 21355049 (phenotype not updated on OMIM)
<i>PINK1</i>	1p36.12	Pten-induced putative kinase 1	PD6	AR (HZ and CH), ADm, early onset	608309, 605909
<i>PLA2G6</i>	22q13.1	Phospholipase A2, group VI	PD14, NBIA2A, NBIA2B	AR (HZ and CH), early and late onset	603604, 612953, 256600, 610217
<i>PM20D1</i>	1q32	Peptidase M20 domain containing 1	PD16	Genetic association, late onset	Locus ID not available on OMIM, 613164
<i>RAB7L1</i>	1q32.1	RAB7-like 1	PD	Genetic association, late onset	603949, PMID: 25040112 (phenotype not updated on OMIM)

<i>SNCA</i>	4q22.1	Alpha-synuclein	PD1, PD4, LBD	ADm, early onset	163890, 168601, 605543, 127750
<i>UCHL1</i>	4p13	Ubiquitin carboxyl-terminal esterase L1	PD5, neurodegeneration with optic atrophy	ADm, AR (HZ), juvenile-onset	191342, 613643, 615491
<i>VPS35</i>	16q11.2	Vacuolar protein sorting 35	PD17	ADm, early and late onset	601501, 614203
Vascular cognitive impairment					
<i>ABCC6</i>	16p13.11	ATP-binding cassette, subfamily C, member 6	Arterial calcification; pseudoxanthoma elasticum; pseudoxanthoma elasticum forme fruste	AR (HZ), infantile onset; AR; ADm	603234, 614473, 264800, 177850
<i>COL4A1</i>	13q34	Collagen type IV, alpha-1	Angiopathy, brain small vessel disease, porencephaly 1, intracerebral haemorrhage susceptibility	ADm, infantile onset	120130, 611773, 607595, 175780, 614519
<i>COL4A2</i>	13q34	Collagen type IV, alpha-2	Porencephaly 2, intracerebral haemorrhage susceptibility	ADm, infantile onset	120090, 614483, 614519
<i>HTRA1</i>	10q26.13	HTRA serine peptidase 1	CARASIL syndrome, macular degeneration	AR (HZ), early onset	602194, 600142, 610149
<i>NOTCH3</i>	19p13.12	Notch homology protein 3	Infantile myofibromatosis 2, CADASIL	ADm, early onset	600276, 615293, 125310
<i>SAMHD1</i>	20q11.23	SAM domain and HD domain 1	Aicardi-Goutieres syndrome 5, Chilblain lupus 2	AR (HZ and CH), AD, infantile onset	606754, 612954, 614415
<i>TREX1</i>	3p21.31	3-prime repair exonuclease 1	Aicardi-Goutieres syndrome 1, Chilblain lupus, Vasculopathy, retinal, with cerebral leukodystrophy	AR (HZ and CH), juvenile onset, AD, AD	606609, 225750, 610448, 192315

Abbreviations are as follows: ALS, amyotrophic lateral sclerosis; FTD, frontotemporal dementia; HMN7B, neuropathy, distal hereditary motor, type VIIB; CMT disease, Charcot-Marie-Tooth disease; YV syndrome, Yunis-Varon syndrome; HET4, hereditary essential tremor, 4; NCL, neuronal ceroid-lipofuscinoses; PD, Parkinson's disease; AD, Alzheimer's disease; HDLS, leukoencephalopathy, diffuse hereditary, with spheroids; HSN1E, hereditary sensory neuropathy type 1E; NBIA2A, neurodegeneration with brain iron accumulation 2A; NBIA2B, neurodegeneration with brain iron accumulation 2B; LBD, Lewy body dementia; CARASIL syndrome, cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy;

CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; AR, autosomal recessive; HZ, homozygous; ADm, autosomal dominant; CH, compound heterozygous; ACD, autosomal co-dominant; OMIM, Online Mendelian Inheritance in Man; PMID, PubMed identification. Age of onset was classified as 'late onset' if greater than 65 years of age.

6.3.2 Sample collection and DNA isolation

The following steps were completed by Adam McIntyre with assistance from Allison Dilliott at the Hegele lab.

Blood samples were collected from 216 subjects following appropriate and informed consent in accordance with the Research Ethics Board at Parkwood Hospital (London, Ontario, Canada); London Health Sciences Centre (London, Ontario, Canada); Sunnybrook Health Sciences Centre (Toronto, Ontario, Canada); University Health Network-Toronto Western Hospital (Toronto, Ontario, Canada); St. Michael's Hospital (Toronto, Ontario, Canada); Centre for Addiction and Mental Health (Toronto, Ontario, Canada); Baycrest Centre for Geriatric Care (Toronto, Ontario, Canada); Hamilton General Hospital (Hamilton, Ontario, Canada); McMaster (Hamilton, Ontario, Canada); Elizabeth Bruyère Hospital (Ottawa, Ontario, Canada); and The Ottawa Hospital (Ottawa, Ontario, Canada).

All clinical diagnoses were supplied by each patient's healthcare provider in accordance with the criteria from the general ONDRI protocol (Farhan et al., 2015; in revision). DNA was isolated from 4-8 ml of blood collected from every participant as described in Chapter 2, subsection 2.3.3. DNA quality and concentration were initially measured by NanoDrop-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and followed by subsequent serial dilutions to obtain ~5 ng/ μ l. Qubit 2.0 fluorometer technology (Invitrogen, Carlsbad, CA) was then used to measure lower concentrations of DNA at a higher resolution.

6.3.3 Library preparation

The following steps were completed by Dr. Henian Cao. I performed the library preparation for three runs.

Libraries were prepared in house using the Nextera Rapid Custom Capture Enrichment kit in accordance with manufacturer's instructions. DNA samples were processed in sets of 12. DNA samples were fragmented followed by ligation of Nextera

Custom Enrichment Kit-specific adapters, amplified via PCR using unique sample barcodes, equimolar pooled, and hybridized to target probes (two cycles of 18 hr each). Samples were then amplified again to ensure specificity and greater DNA yield. A small aliquot of each library was analyzed using the Agilent 2100 BioAnalyzer (Agilent Technologies, Palo Alto, CA) to ensure adequate yield. The quantity and quality of the final libraries were measured using the KAPA quantitative PCR library quantification kit (KAPA Biosystems, Woburn, MA) using the ViiA 7 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA).

6.3.4 Next generation sequencing

All samples were sequenced on the Illumina MiSeq Personal Genome Sequencer (Illumina, San Diego, CA, USA) using the MiSeq Reagent Kit v3 in accordance with manufacturer's instructions. Indexed samples were pooled in equimolar ratios of 500 ng. Once combined, 16 pM of denatured pooled library was loaded on to a standard flow-cell on the Illumina MiSeq Personal Sequencer using 2×150 bp paired-end chemistry. Viral PhiX DNA was added as a positive control to ensure sequencer performance. Sequencing quality control was assessed using multiple parameters in Illumina MiSeq Reporter and visualized either in Illumina BaseSpace or locally using Illumina Sequencing Analysis Viewer.

6.3.5 Sequence alignment

After demultiplexing and adapter trimming, FASTQ files were aligned to the consensus human genome sequence build GRCh37/hg19 using a customized workflow within CLC Bio Genomics Workbench v6.5 (CLC Bio, Aarhus, Denmark). Software default settings of two built-in protocols were used to enhance read mapping including a local realignment for insertion-deletion variation and removal of duplicates introduced during PCR similar to Picard tools described in Chapter 2, subsection 2.3.9.

6.3.6 Variant calling

A quality-based variant detection tool was used to call sequence variants according to a minimum 10-fold coverage and 20% read frequency. These thresholds, while low were selected to minimize false negatives by calling variants in low-coverage regions, which can later be validated using Sanger sequencing. These parameters however can produce a higher false positive call rate. In addition, for each ONDRI sample, target region summary statistics were generated and variant reports were exported in VCF files for subsequent analysis. Variants identified in target regions and within 10 bp bordering targets, were functionally annotated.

6.3.7 Variant annotation

Variant annotation was performed using ANNOVAR (Wang et al., 2010), a publically available functional annotation pipeline using the authors' recommended guidelines and automated scripts. Additional databases such as ExAC (Lek et al., 2015), CADD (Kircher et al., 2014), HGMD (release 2015.1.), ClinVar (Landrum et al., 2014), and our own in house databases were integrated using customized scripts optimized by Eric Liang, a former undergraduate student in the Hegele lab.

6.3.8 Variant classification and prioritization

In general, I followed the guidelines for the interpretation of sequence variants proposed by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (Richards et al., 2015). I screened for rare variants, which in were considered here to be variants with MAF <1% based on 1000 Genomes, NHLBI Exome Sequencing Project, and the ExAC databases. Among rare variants, I investigated whether there were any non-synonymous changes (nucleotide substitutions, insertions, or deletions) that resulted in missense, nonsense, splicing, or frameshift variation. Variants were also assessed *in silico* using a compilation of prediction programs: PolyPhen-2, SIFT, and CADD. HGMD and ClinVar were also integrated to determine the novelty or

recurrence of any genetic variation with a specific disease state. More specifically, I was interested in determining how many variants were previously deposited into disease databases. In the study, variants were marked as clinically relevant if they were rare, exerted non-synonymous changes, were previously observed in individuals with the same disease state, and had values consistent with ‘disease-causing’ based on prediction outcomes of PolyPhen-2, SIFT, and CADD, as recommended by ACMG Standards and Guidelines. Importantly, I grouped the variants according to the categories set forth by the ACMG Standards and Guidelines. Alternatively, variants with uncertain clinical significance were variants that were not reported in disease databases and observed in genes that are not typically causative of the disease in which the individual is diagnosed with, as represented on Table 6.3.1. Finally, all ONDRI samples were compared to each other to resolve whether any variants were observed in multiple individuals with the same disease diagnosis. I used this approach to also determine whether the same variant(s) was present in a large subset of ONDRI samples and therefore, was more likely to be an artifact of sequencing or alignment.

6.3.9 *APOE* genotyping

Furthermore, using the ONDRISeq, in addition to screening all samples for variants within *APOE*, we genotype all individuals for the *APOE* risk alleles rs429358(CT) and rs7412(CT). The combination of both individual alleles determines the *APOE* genotype and is known to be one of the major genetic risk factors for late onset AD (Bertram et al., 2007). If there are no deletions at these loci, six potential *APOE* allele combinations are possible (2 alleles x 3 possible genotypes): 1) e2/e2; 2) e2/e3; 3) e2/e4; 4) e3/e3; 5) e3/e4; and 6) e4/e4, the latter of which is associated with up to an 11x increased risk in developing AD (Bertram et al., 2007; Ghebranious et al., 2005).

6.3.10 Variant validation

To validate variants detected by ONDRISeq, we used three independent genotyping techniques, namely 1) NeuroX, which is an array of specific genotypes that confer risk to several neurodegenerative disease phenotypes (Ghani et al., 2015), 2) TaqMan allelic discrimination, and 3) Sanger sequencing. We processed 115 samples on the NeuroX and the TaqMan allelic discrimination assay and determined the concordance rate between each assay and ONDRISeq. We also randomly selected ~10% of variants to genotype using Sanger sequencing to determine the occurrence if any, of false positives. To test for true negatives, we used DNA from four individuals who were diagnosed with ALS. These individuals were previously tested for genetic variation within SOD1 with no variants identified. Similarly, we did not identify any variants in SOD1 using ONDRISeq. The NeuroX genotyping, TaqMan allelic discrimination assay, and the prior SOD1 testing of ALS samples was performed independently without knowledge of the variants generated using ONDRISeq. Furthermore, different lab personnel completed each validation step to ensure objectivity when evaluating concordance of results.

6.3.10.1 Variant validation 1: NeuroX

The following steps were conducted by Dr. Mahdi Ghani and Christine Sato from Dr. Rogaeva's lab at the Tanz Institute for Neurodegenerative disease research at the University of Toronto.

DNA samples were genotyped on NeuroX exome array (Illumina, San Diego, CA, USA) according to manufacturer's instructions. NeuroX data were loaded to GenomeStudio (Illumina, San Diego, CA, USA) and all markers were clustered using the default Gen Call threshold (0.15); duplicate samples (N=2) revealed identical genotypes for all markers with available genotypes (N=268,399) (Ghani et al., 2015). Genotypes were converted to PLINK input files, and allele frequencies were calculated. In total, the 115 samples revealed 71,714 polymorphic autosomal markers including 43,129 exonic and 216 splicing variants; among them 39,390 polymorphisms were non-synonymous, as well as 423 stop-gain and 32 stop-loss variants, according to ANNOVAR analyses

(Ghani et al., 2015). Average sample call rate was 99.6%, indicating high genotype quality.

Next, 1,047 polymorphic markers, which included 252 exonic variants (229 nonsynonymous and 1 splicing) within the 80 genes of the ONDRISeq targeted sequencing panel, were further processed by removing all noncoding, synonymous, and common variants with frequency minor allele frequency >1% in any database of 1000Genomes (1000g2014oct_all), Exome Variant Server (esp6500si_all) and Exome Aggregation Consortium (ExAC). Variants overlapping segmental duplications were also excluded to avoid possible genotyping error. The remaining variants were filtered to those predicted to have a potential damaging effect on protein function, according to either PolyPhen-2 or SIFT analyses implemented in ANNOVAR.

6.3.10.2 Variant validation 2: TaqMan allelic discrimination assay

The following steps were completed by Allison Dilliot with guidance from Adam McIntyre and me.

APOE SNP genotyping was performed using the TaqMan allelic discrimination assay for 115 samples on the 7900HT Fast Real-Time PCR System (Life Technologies, Foster City, CA, USA), and genotypes were identified using automated software (SDS 2.3; Life Technologies). Two TaqMan assays were used to determine the *APOE* genotype, namely 1) C_3084793_20 (rs429358; *APOE* 112) and 2) C_904973_10 (rs7412; *APOE* 158).

6.3.10.3 Variant validation 3: Sanger sequencing

Variants within DNA samples from the ONDRI study were also validated using Sanger sequencing as described in Chapter 2, subsection 2.3.12.1 and 2.3.12.2.

6.3.10.3.1 PCR

For a complete list of the primers used with their respective PCR conditions, please see the table in Appendix B.

6.3.10.3.2 Imaging, purifying, and sequencing of PCR products

Imaging, purifying, and sequencing of PCR products were performed using the same procedures described in Chapter 2, subsection 2.3.12.2.

6.3.10.4 Variant validation 4: *SOD1* testing

Screening for genetic variants in the *SOD1* gene was performed by PCR followed by standard Sanger sequencing methods, on DNA from four individuals diagnosed with ALS. These steps were performed in other research laboratories prior to this study. Using ONDRISeq, we sequenced DNA from these four individuals to determine whether there were any *SOD1* genetic variants. This step allows us to evaluate any true/false negative discoveries.

6.3.11 *C9orf72* genotyping

The following steps were conducted by Dr. Ming Zhang and Christine Sato from Dr. Rogaeva's lab at the Tanz Institute for Neurodegenerative disease research at the University of Toronto.

All participants were genotyped for the G₄C₂-expansion in *C9orf72* using a two-step method: 1) amplicon length analysis and 2) repeat-primed PCR. Experimental procedures are described elsewhere (Xi et al., 2012).

6.3.12 Statistical analysis

The Student's t-test was used to determine the significance of the difference among quantitative patient characteristics within the different neurodegenerative disease cohorts, where appropriate. Chi-square analyses were used to compare discrete traits and allele frequencies. SAS version 9.2 was used for all statistical comparisons.

6.4 Results

6.4.1 Study subjects

We recruited 216 participants affected with one of the following disorders: 1) AD/MCI, N=40; 2) ALS, N=22; 3) FTD, N=21; 4) PD, N=56; and 5) VCI, N=77 as part of the ONDRI study (Table 6.4.1). The average + standard deviation age of our participants was 69.4 ± 7.8 years. Not surprisingly, individuals diagnosed with ALS were the youngest in our cohort with an average age of 61.9 ± 9.1 years. AD/MCI cases were the oldest patients (mean age of 74.5 ± 6.6 years). The youngest participant is a 40 year old male diagnosed with ALS; the oldest are four 85 year old participants (3 males, 1 female); two diagnosed with AD/MCI and two with VCI. In general, sex ratios showed an over representation of males (male:female, 1.8:1.0), which was largely driven by the PD and VCI cases (3.3:1.0 and 2.0:1.0, respectively) similar to the known sex distribution of these disorders in prior population studies. In contrast, in the AD/MCI, ALS, and FTD cohorts, the male:female ratios did not differ considerably (1.5:1.0, 1.2:1.0; and 0.9:1.0, respectively). The self-reported ethnicity of the participants was predominantly Caucasian (82.3%) with some admixture. Overall, participants did not have a family history of neurodegenerative disease and were considered sporadic cases here as determined by participant recall, which was confirmed by the participant's caregiver. Potential confounders such as age, sex, ethnicity, and family history did not affect our objectives or analysis.

Table 6.4.1 Patient demographics.

Disease ID	Cases	Mean age (years±SD)	Min age (years)	Max age (years)	Male:Female	Self-reported ethnicity as Caucasian (%)	Family history of neurodegeneration?
Total	216	69.4±7.8	40	85	140:76	82.3	Sporadic
Alzheimer's disease/mild cognitive impairment (AD/MCI)	40 (18.5%)	74.5±6.6	59	85	24:16	93.3	
Amyotrophic lateral sclerosis (ALS)	22 (10.2%)	61.9±9.1	40	77	12:10	67.9	
Frontotemporal dementia (FTD)	21 (9.8%)	68.8±6.6	55	79	10:11	82.6	
Parkinson's disease (PD)	56 (25.9%)	68.0±5.9	57	82	43:13	83.8	
Vascular cognitive impairment (VCI)	77 (35.6%)	70.2±7.4	55	85	51:26	84.0	

6.4.2 Quality assessment of ONDRISeq data

In total, 9 independent runs of 24 samples were processed on ONDRISeq (Table 6.4.2). All targets across the 216 DNA samples were sufficiently covered ($>30x$; mean coverage $76\pm 18x$) (Table 6.4.2). On average, 22.8 million of 29.8 million reads passed quality filter (PF) equating to 77%. With the exception of the poorest performance run, all ONDRISeq runs had reads PF of $>80\%$. Overall, 92.7% of all reads were mapped with 95% and 78% of reads mapped in the best and poorest performing runs, respectively. All other ONDRISeq runs had $>90\%$ of reads mapped. Of the mapped reads, 87.1% had a Phred quality score of >30 representing a base call accuracy of 99.9%. Similarly, with the exception of the poorest performing run, all ONDRISeq runs had $>85\%$ of reads with scores $>Q30$. Although the poorest performing run produced less than adequate results compared to the other 8 ONDRISeq runs, 100% of its targets were covered $>30x$ and were still analysed in our study.

Furthermore, an additional four DNA samples were extracted from brain tissue of deceased individuals. Post-autopsy, sections of the brain from all four individuals were frozen for over a decade. However, we were still able to generate adequate sequence calls. Among the four samples, 96% of reads were mapped and each sample had an average coverage of 71x.

Table 6.4.2. Quality control metrics for sequencing runs on ONDRISeq.

Parameters	Mean (\pm SD)	Best performance	Poorest performance
Cluster density ($\times 10^3/\text{mm}^2$)	1433.6 (± 165)	1320	1835
Target size (bp)	971,388	971,388	971,388
Total reads ($\times 10^6$)	29.8 (± 2.5)	29.1	35.6
Reads PF ($\times 10^6$)	22.8 (± 0.9)	24.1	22.1
Reads PF (%)	77 (± 5.8)	83	62
Targets bases ≥ 30 (%)	92.0%	95.3	84.9
Mean target coverage	76 (± 18)	NA	NA
Max target coverage	259	NA	NA
Min target coverage	0	NA	NA

Mean of 9 runs. NA, not applicable.

6.4.3 ONDRISeq is concordant with NeuroX, TaqMan allelic discrimination assay, and Sanger sequencing

Three independent genomic techniques, NeuroX, a genome-wide exome array for neurodegenerative disorders, TaqMan allelic discrimination assays, and Sanger sequencing were used to assess the concordance with ONDRISeq in variant detection. The NeuroX array captures known polymorphic variants within the genes represented on ONDRISeq; therefore, we evaluated whether ONDRISeq could detect the same variants as NeuroX. In doing so, we processed 115 DNA samples and ONDRISeq detected all 122 non-synonymous variants initially detected by NeuroX. Furthermore, Dr. Mahdi Ghani assessed rare and common, non-synonymous and synonymous variants called by the two platforms and observed 100% concordance between calls. Of note, there were variants detected by ONDRISeq but not included on the NeuroX array. However, there were no false negatives with ONDRISeq: all variants detected by NeuroX were also detected by ONDRISeq. Furthermore, we used a TaqMan allelic discrimination assay to genotype the same 115 DNA samples for *APOE*. Similarly, we observed 100% concordance between *APOE* genotyping calls on ONDRISeq and TaqMan.

To explore the rate of false positive variant calls by ONDRISeq, we performed an independent concordance study for ~10% (20) of randomly selected variants from samples that were called as variants by ONDRISeq using Sanger sequencing. Similar with the results of NeuroX and TaqMan allelic discrimination assay, we observed 100% concordance in variants initially detected by ONDRISeq and validated via Sanger sequencing. Thus, there were no false positives with ONDRISeq: all variants called as variants by ONDRISeq were also called as variants by validation using Sanger sequencing.

6.4.4 Genetic variation in patients with neurodegenerative disease

In these sections, I will discuss all genetic variation identified in patients within the study.

6.4.4.1 *C9orf72* hexanucleotide expansion in patients with neurodegenerative disease

All DNA samples were independently screened for a hexanucleotide expansion (G₄C₂) within *C9orf72*, a type of DNA variation that was not detectable by ONDRISeq or NeuroX. Of the 216 samples, only three (1.4%) carried an expansion within *C9orf72*, two were diagnosed with ALS and one with FTD (Table 6.4.4.1).

6.4.4.2 Genetic variation identified in patients with neurodegenerative disease using ONDRISeq

In total, only 60 out of 216 samples (27.8%) were completely free from rare potentially pathogenic variants (missense, nonsense, frameshift, in frame insertions and/or deletions, splicing, and minor allele frequency <1%) in ONDRISeq genes (Table 6.4.4.2.1). Of the remaining 156 cases, the AD/MCI and FTD cases had the highest variant rate based on ONDRISeq (>80%), although not necessarily disease causative. In the ALS and PD cases, we identified rare coding variants in 72.7% and 71.4% of individuals, respectively. The VCI disease cohort had the lowest number of variant carriers (65%) although still significantly higher than previous reports (Dwyer et al., 2013; Sun et al., 2015). While 76 (48.7%) of 156 individuals carried one variant, 57 (36.5%) carried two variants, and 23 (14.8%) carried three or more variants (Table 6.4.4.2.1).

Among the 156 cases with potentially pathogenic variants, a total of 266 unique non-synonymous, rare variants (Table 6.4.4.2.2) were identified, including 107 (40.2%) within genes known to cause the disease with which the patient had been diagnosed (e.g. variation in an AD gene in an AD patient). An additional 159 variants (59.8%) were found in genes that were not previously associated with the respective clinical phenotype of the patient, but within a gene responsible for another disease (e.g. variation in FTD gene in an AD patient). Of the 266 variants, 62 (23.3%) were previously reported in HGMD and/or ClinVar; while 204 (76.7%) were absent from disease databases and were thus classified here as novel variants (Table 6.4.4.2.2). The majority of novel variants detected were observed in FTD and PD cases (88.9% and 82.6%, respectively); while the

majority of known variants were observed in ALS and VCI cases (35.7% and 28%, respectively) (Table 6.4.4.2.2). On average, we observed 4 rare variants (minor allele frequency <1%) per individual; and 1 variant per individual that met criteria set by ACMG and was considered here, as candidate variants. More rare variants were observed in individuals of African descent (16 rare variants per individual; 2 variants that met ACMG guidelines, per individual). Individuals of South Asian and Chinese origin on average carried 4.5 and 4 rare variants; and 2.5 and 2 variants meeting ACMG guidelines, respectively. These observations are likely due to ascertainment bias in the databases as they typically contain significantly more individuals of European descent than any other ethnic cohort.

6.4.4.3 *APOE* genotypes in patients with neurodegenerative disease

Importantly, ONDRISeq is able to provide genotypes for *APOE*, which is not available by NeuroX and other arrays. In 216 cases, we did not identify a single case of *APOE* e2/e2 (Table 6.4.4.1). We identified 26 (12%) individuals who had an *APOE* e2/e3 genotype and 131 (60.6%) individuals who had an *APOE* e3/e3 genotype (Table 6.4.4.1). In total, 46 (21.3%) individuals were heterozygous for *APOE* e4 by possessing either an *APOE* e2/e4 or *APOE* e3/e4 genotype; while 13 (6.02%) individuals were homozygous for *APOE* e4 (Table 6.4.4.1). Not surprisingly, of the 13 *APOE* e4/e4 individuals, 7 (53.8%) were diagnosed with AD (Table 6.4.4.1).

Table 6.4.4.1 Other risk variants identified in a cohort of 216 disease cases.

Disease ID	<i>C9orf72</i> expansion carriers	<i>APOE</i> e2/e2 genotype	<i>APOE</i> e2/e3 genotype	<i>APOE</i> e2/e4 genotype	<i>APOE</i> e3e/3 genotype	<i>APOE</i> e3/e4 genotype	<i>APOE</i> e4/e4 genotype
Total (n=216)	3 (1.40%)	0 (0.00%)	26 (12.0%)	1 (0.46%)	131 (60.6%)	45 (20.8%)	13 (6.02%)
AD/MCI (n=40)	0 (0.00%)	0 (0.00%)	1 (2.50%)	0 (0.00%)	17 (42.5%)	15 (37.5%)	7 (17.5%)
ALS (n=22)	2 (9.09%)	0 (0.00%)	4 (18.2%)	0 (0.00%)	12 (54.5%)	6 (27.3%)	0 (0.00%)
FTD (n=21)	1 (4.76%)	0 (0.00%)	1 (4.76%)	0 (0.00%)	13 (61.9%)	5 (23.8%)	2 (9.52%)
PD (n=56)	0 (0.00%)	0 (0.00%)	10 (17.9%)	1 (1.79%)	39 (69.6%)	5 (8.90%)	1 (1.79%)
VCI (n=77)	0 (0.00%)	0 (0.00%)	10 (13.0%)	0 (0.00%)	50 (64.9%)	14 (18.2%)	3 (3.90%)

Table 6.4.4.2.1 Diagnostic yield of ONDRISeg in a cohort of 216 disease cases.

Disease ID	Individuals without any variants	Individuals with variants	Individuals with 1 variant	Individuals with 2 variants	Individuals with ≥ 3 variants
Total (n=216)	60 (27.8%)	156 (72.2%)	76 (48.7%)	57 (36.5%)	23 (14.8%)
AD/MCI (n=40)	7 (17.5%)	33 (82.5%)	18 (54.5%)	10 (30.3%)	5 (15.2%)
ALS (n=22)	6 (27.3%)	16 (72.7%)	6 (37.5%)	8 (50.0%)	2 (12.5%)
FTD (n=21)	4 (19.0%)	17 (81.0%)	9 (52.9%)	7 (41.2%)	1 (5.9%)
PD (n=56)	16 (28.6%)	40 (71.4%)	22 (55.0%)	13 (32.5%)	5 (12.5%)
VCI (n=77)	27 (35.1%)	50 (64.9%)	21 (42.0%)	19 (38.0%)	10 (20%)

Variant criteria were based on non-synonymous, rare variants (<1% in ExAC). The variants in Table 6.4.4.1 and Table 6.4.4.2.1 are the same but tabulated differently.

Table 6.4.4.2.2 Variants identified in a cohort of 216 disease cases as detected by ONDRISeq.

Disease ID	Individuals with variants	ONDRISeq variants	Variants in disease gene as diagnosed	Variants in other ONDRISeq disease genes	Variants in disease databases	Variants not found in disease databases
Total (n=216)	156 (72.2%)	266	107 (40.2%)	159 (59.8%)	62 (23.3%)	204 (76.7%)
AD/MCI (n=40)	33 (82.5%)	55	19 (34.5%)	36 (65.5%)	12 (21.8%)	43 (78.2%)
ALS (n=22)	16 (72.7%)	28	17 (60.7%)	11 (39.2%)	10 (35.7%)	18 (64.3%)
FTD (n=21)	17 (81.0%)	27	12 (44.4%)	15 (55.6%)	3 (11.1%)	24 (88.9%)
PD (n=56)	40 (71.4%)	63	31 (49.2%)	32 (50.8%)	11 (17.5%)	52 (82.6%)
VCI (n=77)	50 (64.9%)	93	28 (30.1%)	65 (69.9%)	26 (28.0%)	67 (72.0%)

‘ONDRISeq variants’ refers to the total number of variants identified in each disease cohort or the total number of neurodegenerative disease cases. ‘Variants in disease gene as diagnosed’ refers to variants in genes known to cause the disease the patient is diagnosed with. ‘Variants in other ONDRISeq disease genes’ refers to variants identified in genes that are not typically associated with the disease the patient is diagnosed with as categorized on the ONDRISeq gene panel. ‘Variants in disease databases’ were calculated by dividing the number of variants reported in HGMD or ClinVar by the total variants value for each disease. Similarly, ‘Variants not found in disease databases’ were classified as variants absent from HGMD or ClinVar. The variants in Table 5 and Table 6 are the same variants but tabulated differently.

6.4.4.4 Case study: strong evidence of pathogenicity for APP p.Ala713Thr in AD patient

We provide an example of a single neurodegenerative disease case to demonstrate the clinical utility of ONDRISeq and our complementary bioinformatics workflow.

The patient is a 73 year old male diagnosed with AD. We identified a heterozygous variant, namely g.11248G>A (c.2137G>A), resulting in a missense variant p.Ala713Thr in *APP*, a gene known to be associated with familial autosomal dominant AD (Figure 1A) (Bertram et al., 2007). The introduction of a polar amino acid within beta APP domain (amino acid residues 675-713) is predicted to effect protein function according to multiple *in silico* analyses and generated a CADD score of 5.483 (Figure 6.4.4.2 A, B). The affected codon is also highly conserved in evolution within the APP protein when aligned to a set of diverged species within the animal kingdom (Figure 6.4.4.2 C). The variant is very rare with a minor allele frequency of 0.006% according to ExAC and is absent from the 1000 Genomes database and the National Heart, Lung, and Blood Institute Exome Variant Server. Furthermore, the patient is the only carrier of p.Ala713Thr in *APP*, among the 216 samples in our study. However, the variant has been previously observed in AD cases as it is reported in both HGMD and ClinVar databases and has been previously reported in multiple publications (Armstrong et al., 2004; Bernardi et al., 2009; Rossi et al., 2004). Indeed the variant had sufficient coverage of 94x, nevertheless, we independently validated the presence of the variant using NeuroX and Sanger sequencing. The patient is also homozygous for *APOE* e3/e3.

A.

ID	Gene	cDNA change	Amino acid change	Type of variant	CADD	MAF (%)	Disease associated	ClinVar	HGMD	ONDRI cases	NeuroX	Sanger
1, AD	APP	c.2137G>A	p.A713T	Missense	5.483	0.006	AD	Yes	Yes	Single case	Yes	Yes



C. Mutation 670 KMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVITTVIVITLVMLKKKQYT 729
 Human 670 KMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 Monkey 670 KMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 Chicken 670 KMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 Mouse 670 KMDAEFGHDSGFEVRHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 Rat 670 KMDAEFGHDSGFEVRHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 Pig 670 KMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 Dog 670 KMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAATVIVITLVMLKKKQYT 729
 ***** **

D. 21:27264108C>T

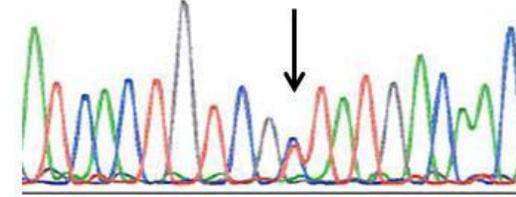

E. 21:27264108C>T

 Ala/Thr

Figure 6.4.4.4 APP variant in AD case

(A) Schematic of the gene and variant discovery process in a neurodegenerative disease case. AD, Alzheimer's disease, patient 1; *MAF was retrieved using ExAC database. (B) APP protein structure shown from N- to C-terminal, 1: amyloid A4 N-terminal heparin-binding domain; 2: copper-binding of amyloid precursor; 3: Kunitz/Bovine pancreatic trypsin inhibitor domain; 4: E2 domain of amyloid precursor protein; 5: beta-amyloid peptide domain; 6: beta-amyloid precursor protein C-terminus domain. The gold star represents the location of the missense variant. (C) Multiple alignments demonstrate high conservation of wild type amino acid residue p.Ala713 (in bold; the variant residue p.Thr173 is shown in red) across a set of species-specific APP homologs. The asterisks below indicate fully conserved residues. (D) The ONDRISeq output showing heterozygosity at the position of the genetic variant, 21:27264108G>A. ONDRISeq output produced 94x coverage. (E) An electropherogram showing the DNA sequence analysis of *APP* from a patient diagnosed with Alzheimer's disease. Our reported cDNA and amino acid positions are based on NM_000484.3 and NP_000475.1, respectively.

6.5 Discussion

Herein, we describe a NGS based custom-designed resequencing panel to assess genes related to neurodegenerative diseases and small vessel disease. ONDRISeq is a rapid and economical diagnostic approach that screens 80 neurodegenerative genes in parallel. We have processed a total of 216 samples on ONDRISeq in 9 runs with 24 batched samples and evaluated each run using highly stringent quality assessment criteria. With ONDRISeq, we have consistently generated high quality data and when coupled with our bioinformatics workflow, we have been able to identify rare genetic variants in >70% of patients diagnosed with one of five diseases: diseases: AD/MCI, ALS, FTD, PD, or VCI.

The ONDRISeq calls were highly reliable based on validation by three established genetic techniques: NeuroX, a rapid and economical genome-wide genotyping based neurodegeneration array, TaqMan allelic discrimination assay, and Sanger sequencing. While NeuroX is able to genotype >250,000 SNPs, the advantage of ONDRISeq is that it is sequencing based and able to detect novel variants (Ghani et al., 2015). This way, we can agnostically screen individuals for any novel or known variants within the 80 neurodegenerative genes. Furthermore, while the TaqMan allelic discrimination assay is a rapid genotyping approach, specific probes have to be designed for all SNPs of interest becoming ultimately costly and inefficient. Also, unlike Sanger sequencing, ONDRISeq is rapid, efficient, and economical. Following library preparation, we are able to analyze the genetic data for 24 samples in less than 30 hours.

We calculated the cost of sequencing 80 genes using standard Sanger sequencing. The total size of ONDRISeq is 971,388 base pairs, which can be processed via ~1,943 PCR reactions (estimation of 500 base pairs per reaction). Had we processed the sequencing reactions in bulk, the cost per sample for Sanger sequencing would have been \$38,860 USD per individual. Using NGS based approaches like WGS or WES with adequate coverage, the price still remains relatively high at \$1,400 and \$700 USD, respectively (prices based on The Centre for Applied Genomics, Ontario, CA; www.tcag.ca). Conversely, through strategic cost management we were able to bring our overall expenditures to a highly competitive price of \$340 per sample—a reduction of

>99% in cost of Sanger sequencing; a >75% reduction relative to WGS, and >50% reduction relative to WES.

Despite its efficiency and rapidity, there are still some limitations with ONDRISeq. First, it can only capture variants within the selected 80 genes, which prevents the discovery of novel disease loci. However, its custom design allows its genetic content to be altered to include novel genomic regions of interest. Second, ONDRISeq is unable to capture multi-nucleotide repeat expansions in genes, a limitation across all NGS platforms (Singleton, 2011). Many neurological diseases such as Huntington's disease, myotonic dystrophy, Friedreich's ataxia, Fragile X syndrome, and a subset of spinocerebellar ataxias arising due to multi-nucleotide repeat expansions cannot be detected with current NGS methodologies (La Spada et al., 1994; Stevens et al., 2013). More recently, a hexanucleotide (G₄C₂) repeat expansion in *C9orf72* has been observed in familial and sporadic ALS and FTD cases, and very rarely in PD cases (DeJesus-Hernandez et al., 2011; Renton et al., 2011b; Xi et al., 2012). Since its discovery in 2011, it has been one of the top investigated genes as both a diagnostic marker and a therapeutic target (DeJesus-Hernandez et al., 2011; Renton et al., 2011b). *C9orf72* alleles can range from 2-20 repeats, which are common in the healthy population and are likely benign; 20-few hundred, which confer risk; or >few hundred repeats and are pathogenic (Rohrer et al., 2015; Xi et al., 2015). As such, we independently examined all individuals in our cohort for the *C9orf72* expansion using: 1) an amplicon length PCR analysis and 2) a repeat primed PCR analysis. In doing so, we identified that 1.4% of the participants were carriers of a *C9orf72* repeat expansion, which is less than expected. However, when considering the ALS and FTD cohorts independently (n=22, n=21), *C9orf72* expansion carriers make up 9.09% and 4.76% of the cohorts, respectively; and 6.98% jointly. These observations are consistent with *C9orf72* expansions explaining up to 10% of sporadic ALS cases with/without features of FTD (Renton et al., 2014).

6.6 Conclusion

Despite these limitations and the complex heterogeneity in the neurodegenerative diseases, we were able to capture rare variants with a probable, but not certain disease

association, based on allele frequency in the general population and the predictive score of multiple *in silico* software in 72.2% of cases. As the etiology of neurodegenerative diseases is often heterogeneous and multiple factors such as genetics, dietary intake, traumatic brain injury, or serious infections including toxin exposure, can confer risk to disease onset, we intend to functionally validate the genetic variants, especially the novel variants, to determine their effect size and contribution to disease. Of particular interest are variants in genes with multiple disease associations as they may provide clues on the potential for development of therapy to treat symptoms common across all five neurodegenerative diseases.

6.7 References

(2015). WHO takes up the baton on dementia. *Lancet Neurol* 14, 455.

Armstrong, J., Boada, M., Rey, M.J., Vidal, N., and Ferrer, I. (2004). Familial Alzheimer disease associated with A713T mutation in APP. *Neurosci Lett* 370, 241-243.

Bernardi, L., Geracitano, S., Colao, R., Puccio, G., Gallo, M., Anfossi, M., Frangipane, F., Curcio, S.A., Mirabelli, M., Tomaino, C., *et al.* (2009). AbetaPP A713T mutation in late onset Alzheimer's disease with cerebrovascular lesions. *J Alzheimers Dis* 17, 383-389.

Bertram, L., McQueen, M.B., Mullin, K., Blacker, D., and Tanzi, R.E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 39, 17-23.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., *et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245-256.

Dwyer, R., Skrobot, O.A., Dwyer, J., Munafo, M., and Kehoe, P.G. (2013). Using Alzgene-like approaches to investigate susceptibility genes for vascular cognitive impairment. *J Alzheimers Dis* 34, 145-154.

Farhan, S.M., and Hegele, R.A. (2014). Exome sequencing: new insights into lipoprotein disorders. *Curr Cardiol Rep* 16, 507.

- Ghani, M., Lang, A.E., Zinman, L., Nacmias, B., Sorbi, S., Bessi, V., Tedde, A., Tartaglia, M.C., Surace, E.I., Sato, C., *et al.* (2015). Mutation analysis of patients with neurodegenerative disorders using NeuroX array. *Neurobiol Aging* 36, 545 e549-514.
- Ghebranious, N., Ivacic, L., Mallum, J., and Dokken, C. (2005). Detection of ApoE E2, E3 and E4 alleles using MALDI-TOF mass spectrometry and the homogeneous mass-extend technology. *Nucleic Acids Res* 33, e149.
- Goldman, J.S., Hahn, S.E., Catania, J.W., LaRusse-Eckert, S., Butson, M.B., Rumbaugh, M., Strecker, M.N., Roberts, J.S., Burke, W., Mayeux, R., *et al.* (2011). Genetic counseling and testing for Alzheimer disease: joint practice guidelines of the American College of Medical Genetics and the National Society of Genetic Counselors. *Genet Med* 13, 597-605.
- Grimes, D., Gordon, J., Snelgrove, B., Lim-Carter, I., Fon, E., Martin, W., Wieler, M., Suchowersky, O., Rajput, A., Lafontaine, A.L., *et al.* (2012). Canadian Guidelines on Parkinson's Disease. *Can J Neurol Sci* 39, S1-30.
- Guerreiro, R., Bras, J., Hardy, J., and Singleton, A. (2014). Next generation sequencing techniques in neurological diseases: redefining clinical and molecular associations. *Hum Mol Genet* 23, R47-53.
- Hebert, L.E., Weuve, J., Scherr, P.A., and Evans, D.A. (2013). Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* 80, 1778-1783.
- Jack, C.R., Jr., Albert, M.S., Knopman, D.S., McKhann, G.M., Sperling, R.A., Carrillo, M.C., Thies, B., and Phelps, C.H. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7, 257-262.
- Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46, 310-315.
- La Spada, A.R., Paulson, H.L., and Fischbeck, K.H. (1994). Trinucleotide repeat expansion in neurological disease. *Ann Neurol* 36, 814-822.
- Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research* 42, D980-985.
- Miller, R.G., Jackson, C.E., Kasarskis, E.J., England, J.D., Forshe, D., Johnston, W., Kalra, S., Katz, J.S., Mitumoto, H., Rosenfeld, J., *et al.* (2009a). Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 73, 1218-1226.

- Miller, R.G., Jackson, C.E., Kasarskis, E.J., England, J.D., Forshe, D., Johnston, W., Kalra, S., Katz, J.S., Mitsumoto, H., Rosenfeld, J., *et al.* (2009b). Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: multidisciplinary care, symptom management, and cognitive/behavioral impairment (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* *73*, 1227-1233.
- Renton, A.E., Chio, A., and Traynor, B.J. (2014). State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* *17*, 17-23.
- Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., *et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* *72*, 257-268.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *et al.* (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* *17*, 405-424.
- Robinson, L., Tang, E., and Taylor, J.P. (2015). Dementia: timely diagnosis and early intervention. *BMJ* *350*, h3029.
- Rohrer, J.D., Isaacs, A.M., Mizielińska, S., Mead, S., Lashley, T., Wray, S., Sidle, K., Fratta, P., Orrell, R.W., Hardy, J., *et al.* (2015). C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *Lancet Neurol* *14*, 291-301.
- Rossi, G., Giaccone, G., Maletta, R., Morbin, M., Capobianco, R., Mangieri, M., Giovagnoli, A.R., Bizzi, A., Tomaino, C., Perri, M., *et al.* (2004). A family with Alzheimer disease and strokes associated with A713T mutation of the APP gene. *Neurology* *63*, 910-912.
- Singleton, A.B. (2011). Exome sequencing: a transformative technology. *Lancet Neurol* *10*, 942-946.
- Stevens, J.R., Lahue, E.E., Li, G.M., and Lahue, R.S. (2013). Trinucleotide repeat expansions catalyzed by human cell-free extracts. *Cell Res* *23*, 565-572.
- Strong, M.J., Grace, G.M., Freedman, M., Lomen-Hoerth, C., Woolley, S., Goldstein, L.H., Murphy, J., Shoesmith, C., Rosenfeld, J., Leigh, P.N., *et al.* (2009). Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* *10*, 131-146.
- Sun, J.H., Tan, L., Wang, H.F., Tan, M.S., Li, J.Q., Xu, W., Zhu, X.C., Jiang, T., and Yu, J.T. (2015). Genetics of Vascular Dementia: Systematic Review and Meta-Analysis. *J Alzheimers Dis* *46*, 611-629.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164.

Xi, Z., van Blitterswijk, M., Zhang, M., McGoldrick, P., McLean, J.R., Yunusova, Y., Knock, E., Moreno, D., Sato, C., McKeever, P.M., *et al.* (2015). Jump from pre-mutation to pathologic expansion in C9orf72. *Am J Hum Genet* 96, 962-970.

Xi, Z., Zinman, L., Grinberg, Y., Moreno, D., Sato, C., Bilbao, J.M., Ghani, M., Hernandez, I., Ruiz, A., Boada, M., *et al.* (2012). Investigation of c9orf72 in 4 neurodegenerative disorders. *Arch Neurol* 69, 1583-1590.

Chapter 7 - Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia

7 Study rationale

In addition to the neurodegeneration samples collected through ONDRI, we have clinically ascertained three independent families with amyotrophic lateral sclerosis (ALS) in which several family members had also been assessed for the presence of neuropsychological deficits. Family 1 includes individuals presenting with exclusively ALS. Family 2 includes individuals presenting with a spectrum of neurodegenerative phenotypes including ALS, frontotemporal dementia (FTD), myelopathy, Alzheimer's disease (AD), and Parkinson's disease (PD). Family 3 includes individuals presenting with ALS and a behavioural variant of FTD (bvFTD).

Interestingly, all three families harbour a hexanucleotide repeat expansion in the *C9orf72* gene. The expansion size among the three families did not differ yet the inter- and intrafamilial clinical presentation varied significantly. We sought to determine whether individuals with atypical disease presentation carried additional genetic variation. Specifically, I hypothesized that the variation in clinical presentation among the patients can be explained by additional genetic variants that are modifying the neurodegenerative syndrome. In addition to performing *C9orf72* expansion testing, we also processed DNA of affected individuals on the ONDRISeq to potentially identify additional genetic variation using protocols described in Chapter 6.

In this Chapter, I describe our approach of identifying the genetic variation that may in part explain the variability in phenotypes across the three families. We performed genetic analysis consisting of *C9orf72* expansion, *ATXN2* expansion, and ONDRISeq to identify whether there was any coding variation in neurodegeneration genes.

Chapter 7 was prepared as an independent study: "Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia" and is currently submitted for publication (Farhan et al., 2016).

7.1 Overview

We have ascertained three families affected with familial amyotrophic lateral sclerosis (ALS) in which the same hexanucleotide repeat expansion in the *C9orf72* gene has been identified, but in which the clinical manifestations vary significantly. The inter- and intrafamilial disparity in clinical phenotype ranged from a much earlier age of onset, ALS combined with a frontotemporal dementia (FTD), or ALS with no additional features. Given this marked phenotypic heterogeneity, we screened for additional genetic variation. We performed genetic testing consisting of *C9orf72* expansion, *ATXN2* expansion, and targeted next generation sequencing of 80 genes known to be associated with neurodegenerative diseases namely, ALS, FTD, Alzheimer's disease, Parkinson's disease, or vascular cognitive impairment. We observed the presence of an *ATXN2* expansion, *ARHGEF28* p.Lys280Met>fs40Ter, or an *OPTN* p.Met468Arg in patients with a *C9orf72* expansion and who exhibit an atypical disease course. These genetic profiles segregated with distinct clinical presentations in all three families and are likely modifying the ALS phenotype. We suggest that the variability in the clinical expression of *C9orf72* expansion carriers is driven by the presence of these additional genetic mutations and provide further evidence for biologically relevant oligogenic inheritance in ALS.

7.2 Introduction

Based on both clinical and genetic studies, it has been proposed that amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are part of the same neurodegenerative disease continuum (Hardiman et al., 2011; Woolley and Strong, 2015). ALS is a progressive adult-onset neurodegenerative disorder in which the loss of both upper and lower motor neurons leads to a relentless weakening of limb, bulbar, and respiratory muscles. Death usually ensues within 3–5 years of onset (Renton et al., 2014). In contrast, FTD is characterized by progressive neuropsychological deficits, including cognitive and behavioural dysfunction. Survival is generally on the order of five to ten years (Rohrer et al., 2015).

There is increasing recognition that ALS and FTD can occur in the same individual and that both disorders are syndromic with multiple potential etiologies, many of which are genetic in origin (Hardy and Rogaeva, 2014; Ittner et al., 2015). In general, approximately 15% of ALS or FTD cases will progress to develop both phenotypes (Hardy and Rogaeva, 2014; Ittner et al., 2015). This observation underlies the current conceptualization that ALS and FTD are two points on a continuum of a single disorder. The discovery that a large proportion of ALS and FTD cases harbour a pathogenic hexanucleotide expansion in *C9orf72* provides some support for this hypothesis (DeJesus-Hernandez et al., 2011; Renton et al., 2011). However, this does not explain the significant clinical heterogeneity among patients with similar *C9orf72* expansion profiles (Rohrer et al., 2015). Families in which similar *C9orf72* expansions are present, but in whom there is clinical heterogeneity both within families and between pedigrees, can provide insight into this question. However, until recently the finding of a single genetic variant led to a cessation of the search for additional genetic variants. This reductionist approach has generally constrained the ascertainment of the full extent of genetic mutations in ALS and FTD. However, when such an approach has been used, it has supported the concept of digenic (hereafter termed oligogenic) mutation inheritance leading to atypical disease presentation (summarized in Table 7.5) (Bury et al., 2015; Chio et al., 2012; Kaivorinne et al., 2014; King et al., 2013; Lashley et al., 2014; Pottier et al., 2015; Tarlarini et al., 2015; van Blitterswijk et al., 2012a; van Blitterswijk et al., 2012b).

In this report, we describe three families that lie along this spectrum of ALS and FTD. While each family harbours a similar *C9orf72* expansion, each family varies significantly with respect to their mutation load. We used an agnostic approach by determining the families' *C9orf72* expansion profiles in addition to sequencing all known ALS-associated disease genes. We found that all three harbour additional genetic variants and postulate that the presence of oligogenic inheritance accounts for phenotypic variability.

7.3 Materials and methods

7.3.1 Ethics

All study subjects provided written and informed consent to participate in our study. All clinical, cognitive, and molecular testing were conducted in accordance with protocols approved by the Research Ethics Board (105522) at Western University, London, Ontario, Canada.

7.3.2 Sample collection clinical assessment

We have ~25 years of clinical documentation on the history of the families and their affected relatives. All study participants and their affected relatives, were independently evaluated by multiple neurologists including Dr. Strong. In our study, Dr. Strong conducted a neurological evaluation including the Montreal Cognitive Assessment (MoCA) on all individuals.

7.3.3 DNA isolation

DNA was isolated from 34 participants using methods described in Chapter 2, subsection 2.3.3.

7.3.4 *C9orf72* testing

Study participants were tested for the *C9orf72* expansion using: i) amplicon length analysis and ii) repeat-primed PCR. Additionally, post-autopsy brain sections were collected from deceased individuals and immunohistochemistry on brain sections was performed to verify the presence of the *C9orf72* dipeptide repeat proteins. Our results were validated by three independent research laboratories, and some samples were validated using laboratories with clinical laboratory improvement amendments (CLIA)

certification. DNA from positive and negative control samples as well as deceased individuals were also tested for the *C9orf72* expansion via Southern immunoblotting.

7.3.4.1 Amplicon length analysis

The G₄C₂ hexanucleotide repeat in *C9orf72* was PCR amplified using one fluorescently labeled primer: 5'-FAM-CAAGGAGGGAAACAACCGCAGCC-3' and a regular RVS primer: 5'-GCAGGCACCGCAACCGCAG-3'. The PCR reaction was performed using a mixture containing 1M betaine solution, 1% dimethylsulfoxide (DMSO) and 7-deaza-2-deoxy GTP in substitution for dGTP in accordance with previously tested protocols with slight modifications (DeJesus-Hernandez et al., 2011; Renton et al., 2011). PCR conditions were as follows: (1) initial denaturation: 98°C for 5 min; (2) denaturation: 97°C for 30 sec; annealing: 65-56°C for 30 sec; extension: 68°C for 1.5 min. This entire cycle was repeated 10x and with every cycle, the annealing temperature was decreased by 1°C. Another set of cycles are programmed: (3) denaturation: 97°C for 30 sec; annealing: 55°C for 30 sec; extension: 68°C for 1.5 min. This entire cycle was repeated 24x. A final extension for 68°C for 10 min was next. Samples were diluted 1:10 and analyzed using fragment length analysis on an ABI3730 DNA Analyzer. Allele identification and scoring was performed using GeneMapper v4.0 software.

7.3.4.2 Repeat-primed PCR

To assess the presence of an expanded (G₄C₂)_n hexanucleotide repeat in *C9orf72*, we performed repeat primed PCR using three primers: (1) FAM-TGTAACGACGGCCAGTCAAGGAGGGAAACAACCGCAGCC-3'; (2) 5'-CAGGAAACAGCTATGACC-3'; and (3) 5'-CAGGAAACAGCTATGACCGGGCCCGCCCCGACCACGCCCCGGCCCCGGCCCCGG-3'. Primer concentrations added were in a 2:1:2 ratio, respectively. The PCR reaction was performed in a mixture containing 1M betaine solution, 5% DMSO and 7-deaza-2-deoxy GTP in substitution for dGTP. PCR conditions were as follows: (1) initial denaturation: 95°C for 5 min; (2) denaturation: 95°C for 30 sec; annealing: 70-56°C for

30 sec; extension: 72°C for 8 min. This entire cycle was repeated 8x with every cycle, the annealing temperature was decreased by 2°C. Another set of cycles are programmed: (3) denaturation: 95°C for 30 sec; annealing: 56°C for 30 sec; extension: 72°C for 8-16 min. This entire cycle was repeated 26x and with every cycle the extension time was increased by 20 sec. A final extension for 72°C for 7 min was next. PCR products were analyzed on an ABI3730 DNA Analyzer and visualized using GeneMapper software.

7.3.4.3 Immunohistochemistry

Dr. Tania Gendron and Dr. Leonard Petrucelli at the Mayo Clinic, Jacksonville, Florida, performed these procedures.

To confirm that the hexanucleotide expansion of *C9orf72* was pathogenic in the affected individuals, we immunostained cerebellar tissue from each case for poly(GP), poly(GA) or poly(GR) proteins. Five-micron thick cerebellar slices were cut from formalin-fixed, paraffin-embedded blocks and mounted on glass slides. After drying, slides were deparaffinized and rehydrated in xylene and alcohol washes before being steamed for 30 min in deionized water for antigen retrieval. All slides were processed on a Dako Autostainer with the Dako EnVision™ system and 3,3'-diaminobenzidine chromogen. After staining with anti-GP (Rb5823; 1:5,000), anti-GA (Rb9880, 1:50,000) or anti-GR (Rb7810, 1:2,500), slides were counterstained with Lerner hematoxylin and coverslipped with Cytoseal permanent mounting media. Cerebellar sections were examined using an Olympus BX41 microscope and images captured with cellSens Standard 1.5 software.

7.3.4.4 Southern immunoblotting

In addition to verifying the results of fluorescent fragmental length analysis using protocols prepared by other research and clinical laboratories, we performed Southern immunoblotting on positive and negative control samples. We also performed Southern immunoblotting using DNA from frozen brain tissue of deceased individuals.

Unfortunately, we did not have sufficient quantities of DNA from study subjects to

perform Southern immunoblotting. A 241 base pair digoxigenin (DIG)-labeled probe was made from genomic DNA (gDNA) of a healthy control by PCR using FWD primer: 5'-AGAACAGGACAAGTTGCC-3' and RVS primer: 5'-AACACACACCTCCTAAACC-3'. The PCR DIG probe synthesis kit was used in accordance with manufacturer's instructions (Sigma-Aldrich, St.Louis, MO, USA). A total of 10 µg of gDNA was digested with *XbaI* at 37 °C for ~20 hours and electrophoresed. DNA was transferred to a positively charged nylon membrane by capillary blotting and cross-linked by UV irradiation. Post prehybridization in DIG EasyHyb solution (Sigma-Aldrich) at 47°C for 3h, hybridization was at 47°C overnight. The membranes were then washed in 2X standard sodium citrate (SSC), 0.1% sodium dodecyl sulfate (SDS) at room temperature for 5 min and in 0.1x SSC, 0.1% SDS at 68°C for 15 min each. An anti-DIG antibody (1:10,000; Sigma-Aldrich) in blocking solution was added for 30 mins at room temperature. Membranes were then washed using standard wash buffer. Detection of the hybridized probe DNA was achieved using the CDP-star chemiluminescent substrate and signals were visualized on LI-COR.

7.3.5 *ATXN2* expansion testing

Study participants were screened for the *ATXN2* CAG expansion using fluorescent fragment length analysis similar to the *C9orf72* amplicon length analysis protocol. Initially, a laboratory with CLIA certification detected an *ATXN2* expansion in gDNA of patient II-3 from family 2. We tested the other members of family 2: II-2, II-5, and III-5, using protocols developed at the Rogaeva lab. Dr. Ming Zhang performed these analyses at the Rogaeva lab.

7.3.6 Next generation sequencing

DNA of family 1, III-2 and III-5; family 2, II-3; and family 3, III-5 and III-7 were subjected to a next generation sequencing. All family members were screened for genetic variants identified in the index patients, using standard PCR and Sanger sequencing

protocols as described in subsequent sections. Samples were sequenced at the London Regional Genomics Centre on the Illumina MiSeq Personal Genome Sequencer (Illumina, San Diego, CA, USA) using the MiSeq Reagent Kit v3 in accordance with manufacturer's instructions as described in Chapter 6, subsection 6.3.4.

7.3.7 Sequence alignment

Sequence alignment was performed as described in Chapter 6, subsection 6.3.5.

7.3.8 Variant calling

Variant calling was performed as described in Chapter 6, subsection 6.3.6.

7.3.9 Variant annotation

Variant annotation was performed as described in Chapter 6, subsection 6.3.7.

7.3.10 Variant classification and prioritization

Variant classification and prioritization were performed as described in Chapter 6, subsection 6.3.8. We previously identified *ARHGEF28* p.Lys280Met>fs40Ter in patient III-5 in family 1. In addition, we identified *OPTN* p.Met468Arg in patient II-3 in family 2 and patients III-5 and III-7 in family 3.

7.3.11 Variant validation

To determine whether *ARHGEF28* p.Lys280Met>fs40Ter and *OPTN* p.Met468Arg segregated with disease status in the respective pedigree, we genotyped family members using standard Sanger sequencing methods as described in Chapter 2, subsection 2.3.12.1 and 2.3.12.2.

7.3.11.1 PCR

Forward and reverse primers for *ARHGEF28* p.Lys280Met>fs40Ter are: 5'-GGATCATTTTGCCAGTTGGA-3' and 5'-TGATGCCTGCATTCTCTGAA-3', respectively. Forward and reverse primers for *OPTN* p.Met468Arg are 5'-CTGCTATCGGAATGTACCTGG-3' and 5'-ATGCTGATGTGAGCTCTGGG-3', respectively. The PCR conditions for both primer sets were as follows: initial denaturation at 95°C for 5 minutes; 30 cycles of denaturation at 95°C for 30 seconds; annealing at 60°C for 30 seconds; extension 72°C for 30 seconds and a final extension at 72°C for 7 minutes.

7.3.11.2 Imaging, purifying, and sequencing of PCR products

Imaging, purifying, and sequencing of PCR products were performed using the same procedures described in Chapter 2, subsection 2.3.12.2.

7.4 Results

7.4.1 Clinical description

The content in this section was compiled and transcribed by me and extensively reviewed by Dr. Strong and Dr. Hegele. The table containing all genetic and clinical information on all individuals within the study can be found in Appendix C.

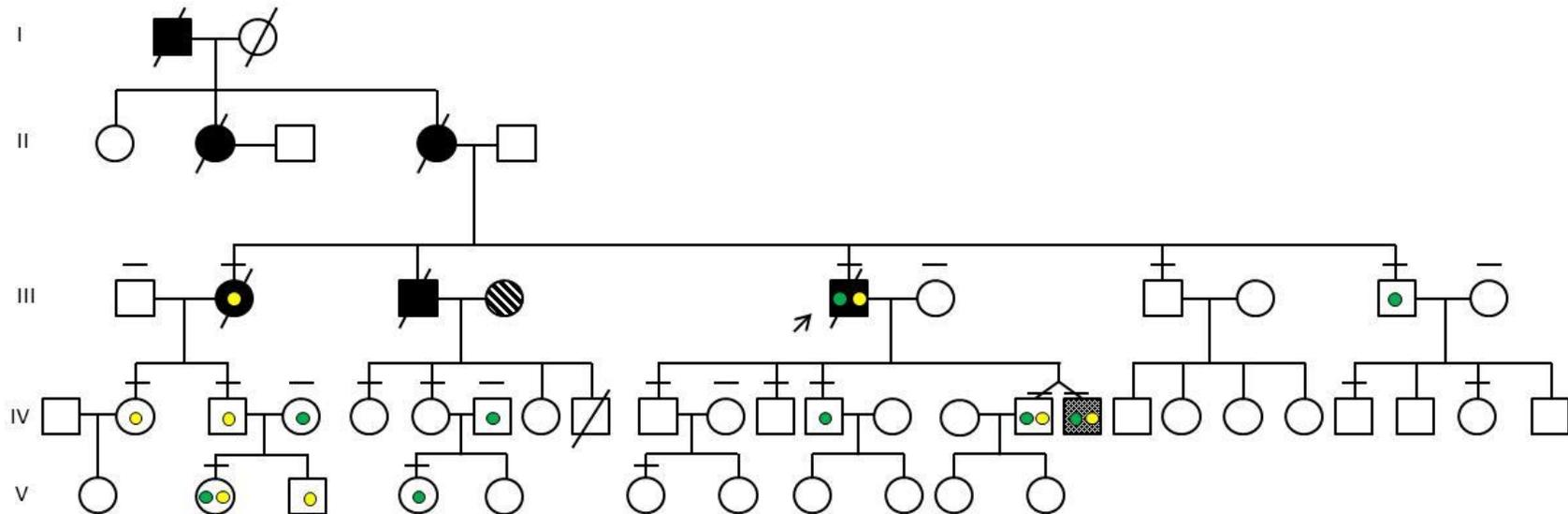
In family 1, patient III-2 initially presented with left upper extremity weakness at age 60 and later developed features of bulbar dysfunction, diffuse limb wasting and weakness, and pathologically brisk reflexes. Given the multigenerational history of ALS in the family (I-1, II-2, and II-4), the clinical diagnosis was definite familial ALS using the El Escorial criteria (Brooks, 1994). The patient survived 4 years from the time of symptom onset. The post mortem neuropathology confirmed the diagnosis of ALS in the absence of frontotemporal lobar degeneration.

Her brothers, III-3 and III-5, were also diagnosed with definite ALS but initially presented at a much younger age of 47 and 49, respectively. On examination, III-3 presented with bilateral upper extremity weakness as well as limb wasting. His total disease duration was 65.2 months). An autopsy was not conducted on III-3. Similar to III-2, the first symptoms observed in III-5 were left upper extremity weakness. He later developed symptoms of bulbar dysfunction, limb wasting, and pyramidal weakness. The total disease duration was >86 months. We conducted neurological and cognitive examinations on the two remaining siblings, III-7 and III-9. Neither had any features suggesting motor neuron degeneration or dementia. Both individuals are much older (70 and 63 years) than the average age for disease onset in their family, which may suggest that they are unlikely to develop ALS. To our knowledge, there is currently no single individual affected with ALS in the subsequent generation; however, all are younger than the age of onset of their affected parent. We have examined all members of generation IV and found no evidence of motor neuron disease or dementia (all members received MoCA scores within the normal range).

Family 2 consists of individuals with multiple, distinct neurodegenerative phenotypes. Patient II-3 initially presented at age 53 with bulbar dysfunction and subsequently developed limb wasting, pyramidal weakness, and cognitive impairment (previously described in Strong et al, 1999). The total disease duration was 22.3 months. His brother, II-5 presents with a perplexing clinical syndrome. While some of his symptoms overlap with those of II-3, he does not meet the diagnostic criteria for ALS and has a communicating hydrocephalus, significant cervical spondylolytic pathology and a previous history of head injury from a motor vehicle accident resulting in a loss of consciousness. He is currently 70 years of age and displays progressive cognitive decline with both short and long-term memory impairment. He is unable to recall people's names and gets lost in familiar environments. He scored 10/30 on MoCA testing. He is currently wheelchair bound with significant spasticity. Their sister, II-7, was diagnosed during life with behavioral variant FTD (bvFTD) and at autopsy was found to also have motor neuron disease (MND). Their eldest sister, II-2, has no deficit on either neurological examination or on MoCA testing.

Family 3 is characterized by ALS (I-2, II-3, II-8, II-13), bvFTD and historical reports of dementia described as Alzheimer's disease (II-4 and II-17), although there are no records that would allow differentiation of Alzheimer's from FTD in these latter cases. At the age of 56, III-3 began displaying inappropriate behaviour consisting of apathy, agitation, reduced verbal output, poor financial judgement, sexual indiscretion, obsessive-compulsive behaviour, and repetitive movements. Within two years, she also developed impairment in executive function with deficits in attention and concentration, language, visuospatial construction, memory and facial recognition. Dysarthria and dysphagia had become prominent. Her examination demonstrated ocular impersistence with preserved optokinetic nystagmus, prominent primitive reflexes, and a generalized loss of muscle bulk with diffuse, severe spasticity. Electrophysiological studies were unremarkable with the exception of an incidental median entrapment neuropathy. Within two years, she had become preservative, inattentive, whispered when speaking and, in addition to diffuse motor neuron dysfunction, had mild cogwheeling. She did not consent to further electrophysiological studies, and died at age 56 years. An autopsy was not performed. Her siblings, III-5 and III-7 had a very similar disease course. III-5, a 53 year old accountant, first developed difficulty with organizational skills, word finding difficulties, an inability to do the daily household banking, an increased appetite, and an obsessive behaviour with increasing stubbornness. Within 4 months, he had developed choking episodes, dysarthria and reduced speech volume. His clinical and electrophysiological examinations at that point were consistent with definite ALS. This was confirmed by neuropathological examination at the time of his death 16 months following initial symptom onset, as was evidence of a frontotemporal lobar degeneration. Case III-7 presented with first episode psychosis at age 56. He began incessant smoking and became homeless. One year later he developed aspiration pneumonia and was placed on full ventilator support. His course thereafter was typical of ALS with complete anarthria, diffuse muscle wasting with fasciculations, pathologically brisk reflexes, and electrophysiological evidence of multisegmental acute and chronic motor unit remodelling. Neuropathological examination confirmed the presence of frontotemporal lobar degeneration with ALS including TDP-43 immunoreactive inclusions within cortical and spinal motor neurons.

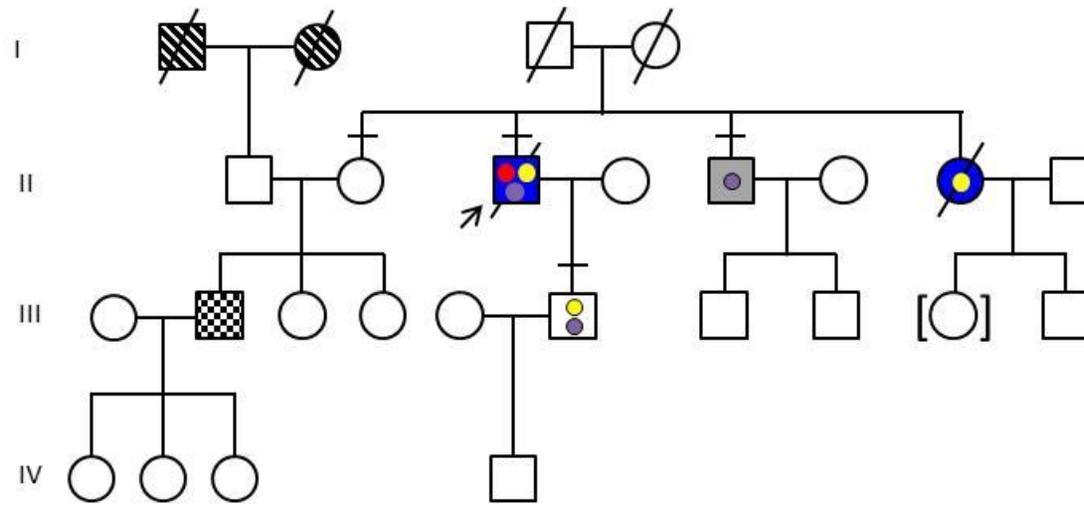
Family 1: primarily ALS phenotype



Legend:

- *C9orf72* expansion
- *ARHGEF28* p.Lys280Met>fs40Ter
- Amyotrophic lateral sclerosis
- ▨ Cerebral palsy

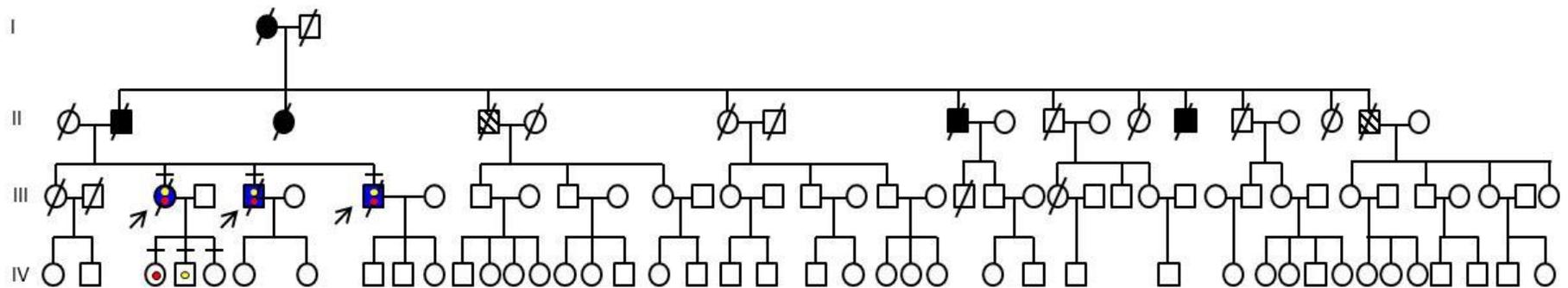
Family 2: mixed neurodegeneration



Legend:

- *C9orf72* expansion
- *OPTN* p.Met468Arg
- *ATXN2* expansion
- Amyotrophic lateral sclerosis with frontotemporal dementia
- Alzheimer's disease
- Parkinson's disease
- Myelopathy

Family 3: ALS and bvFTD phenotype



Legend:

- *C9orf72* expansion
- *OPTN* p.Met468Arg
- Amyotrophic lateral sclerosis
- Amyotrophic lateral sclerosis with frontotemporal dementia
- ▨ Alzheimer's disease

Figure 7.4.1. Pedigrees of three families affected with ALS or ALS-FTD.

Family 1 is affected primarily with ALS without features of cognitive impairment. Family 2 is affected with mixed neurodegenerative phenotypes. Family 3 is affected with bvFTD and ALS. Small dashes above family members in the pedigrees indicate those individuals who were available for DNA analysis. Squares and circles denote male and female family members, respectively. Coloured symbols represent affected family members as described on the legend, and open symbols represent unaffected family members. Diagonal lines indicate deceased family members, and arrows indicate the proband(s) in each family. Coloured small circles within individuals represent their carrier status of genetic variants identified on the legend.

7.4.2 Variants identified in patients with ALS

Patients and participating family members were screened for a hexanucleotide (G₄C₂) expansion in *C9orf72* using amplicon length analysis coupled with repeat-primed PCR as previously described (DeJesus-Hernandez et al., 2011; Xi et al., 2012). Positive and negative control samples as well as affected individuals with a *C9orf72* expansion were also independently validated via Southern immunoblotting using a DIG labelled probe. *C9orf72* genotypes are provided in supplementary table 1 as well as figures 3 and 4. In families 1, 2, and 3, we identified eight, three, and four *C9orf72* expansion carriers, respectively. However, all expansions were less than 100 repeats and are within the ‘uncertain clinical significance’ classification (Figure 7.4.1, Table 7.4.4). Of the carriers identified in family 1, two (III-2 and III-5) died with a diagnosis of ALS. It is unclear whether the other six carriers in family 1 (IV-2, IV-3, IV-16, IV-17, V-2 and V-3) are disease free or are currently pre-symptomatic as no neurologic abnormalities were observed when examined by a neurologist with expertise in ALS (Dr. Strong). Similarly in family 2, we identified three carriers (II-3, II-7, and III-6) who are also neurologically intact. II-3 and II-7 were diagnosed with ALS-FTD and are deceased. The third carrier, III-6 is the son of an ALS-FTD affected individual. In family 3, we identified four *C9orf72* expansion carriers (III-3, III-5, III-7, and IV-5); three carriers were diagnosed initially with bvFTD, which manifested into bvFTD-ALS. The living carrier, IV-5, is currently unaffected or presymptomatic.

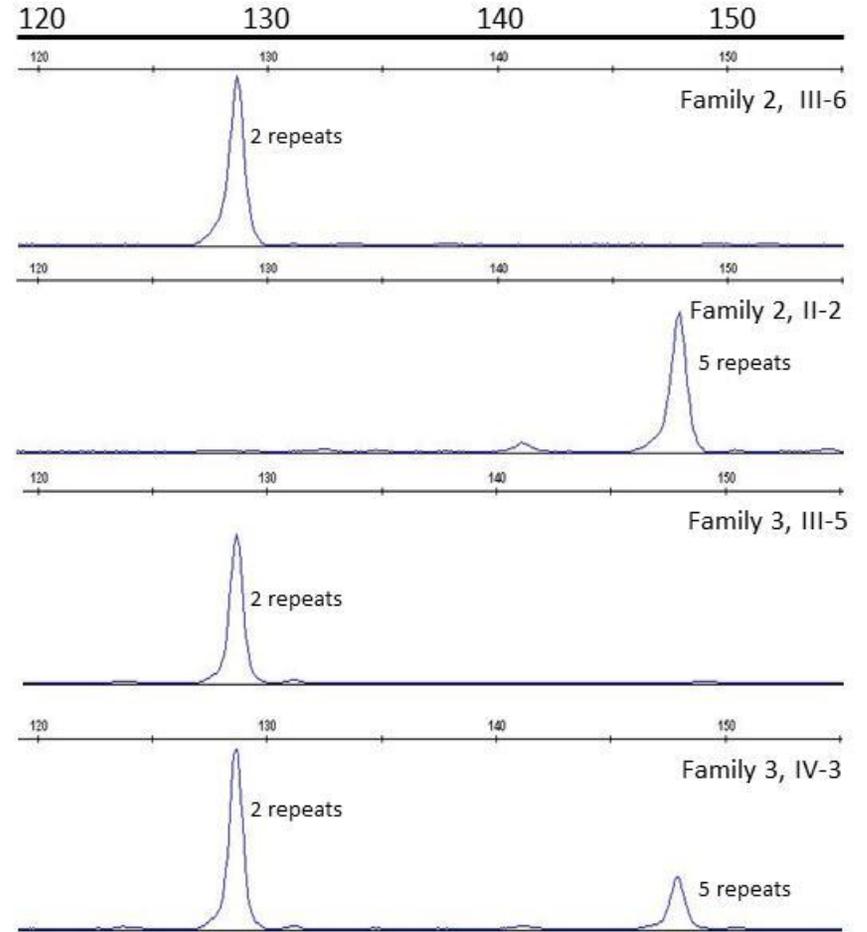
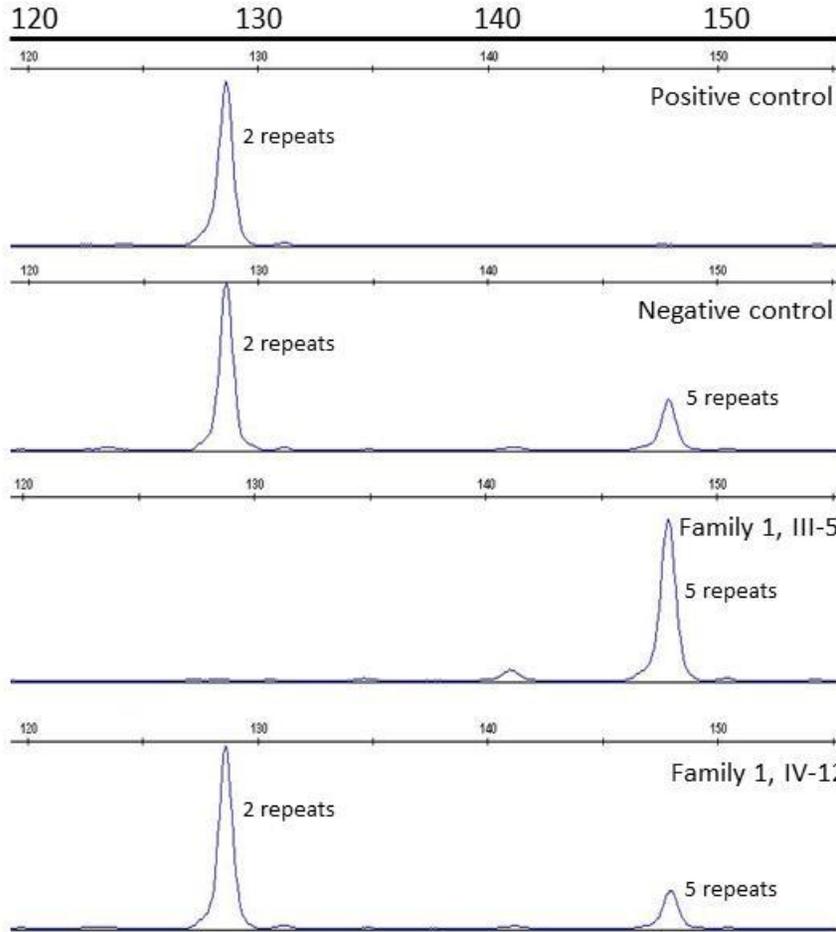


Figure 7.4.2.1 Genetic analyses of *C9orf72* genotypes of representative individuals from each family.

Results from fluorescent fragment length analyses. The presence of two individual peaks corresponds to non-*C9orf72* expansion carriers. Repeat numbers are beside each peak and are calculated by subtracting 117 from the number determined using the horizontal axis. Single peaks represent either homozygosity or heterozygosity for the corresponding peak. Single peak profiles can also be indicative of a *C9orf72* expansion that cannot be detected using this technique. Here, positive and negative controls are a *C9orf72* expansion carrier and a non-*C9orf72* expansion carrier, respectively. The other panels show a positive and negative *C9orf72* expansion carrier in each family. All participants were evaluated and all results were repeated three times with the same findings across all three experiments. The vertical and horizontal axes represent fluorescent intensity and product size, respectively.

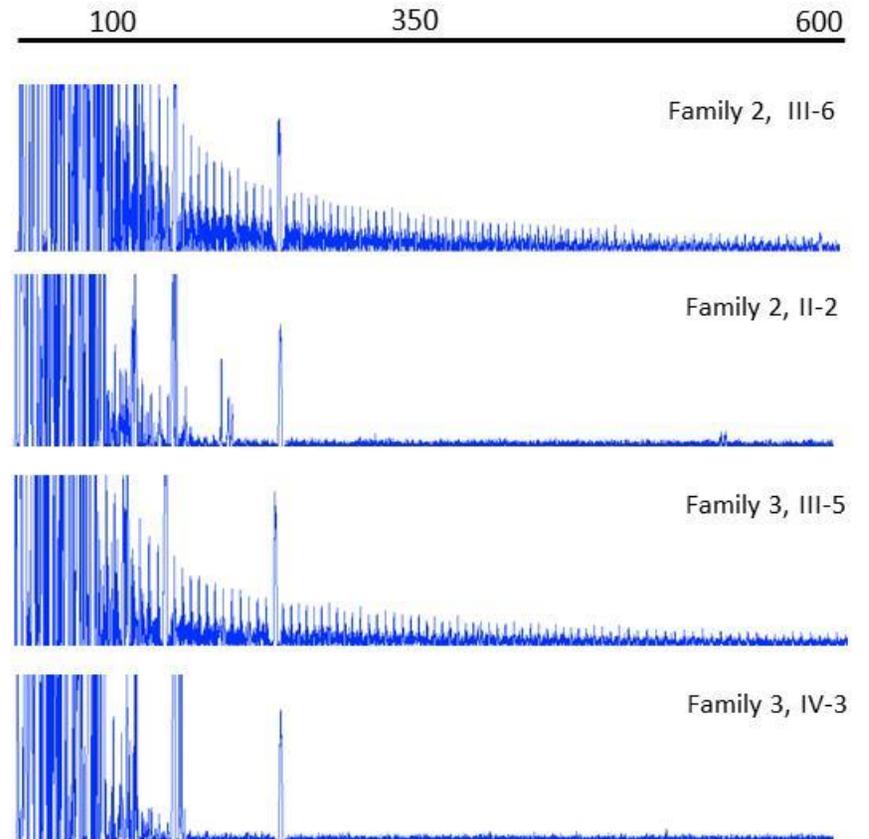
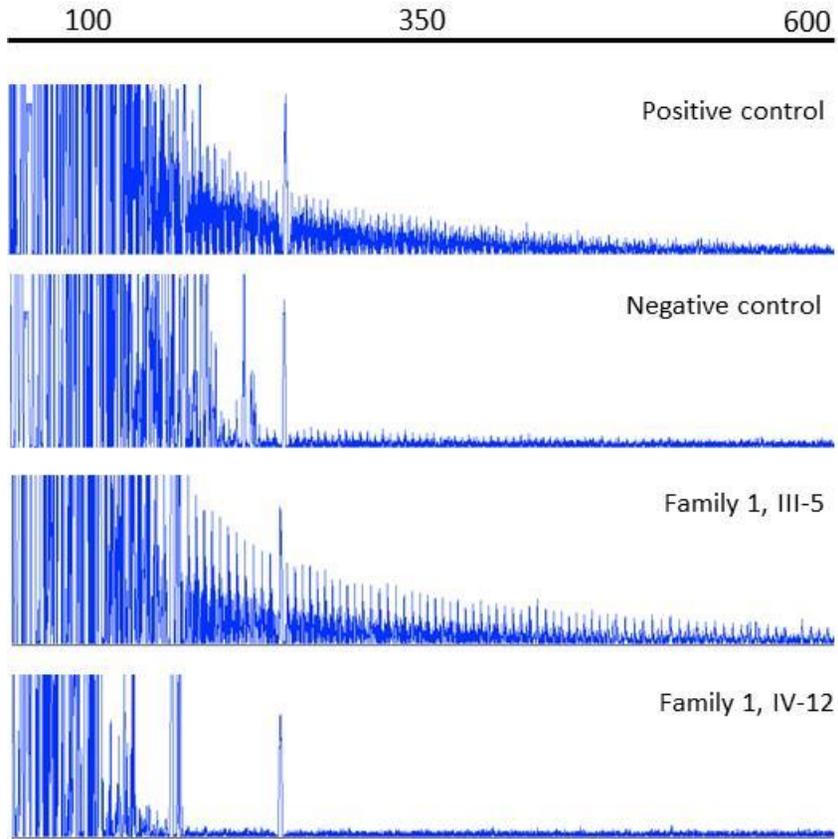


Figure 7.4.2.2 Genetic analyses of *C9orf72* expansion profiles of representative individuals from each family.

Results from fluorescent fragment analysis traces show the *C9orf72* expansion profiles. The ‘saw-tooth’ slopping pattern is indicative of a *C9orf72* expansion. Here, positive and negative controls are a *C9orf72* expansion carrier and a non-*C9orf72* expansion carrier, respectively. The other panels show a positive and negative *C9orf72* expansion carrier in each family. All participants were evaluated and all results were repeated three times with the same findings across all three experiments. The vertical and horizontal axes represent fluorescent intensity and product size, respectively.

We also screened participants for CAG expansions within *ATXN2*. In family 2, II-3, II-5, and III-6 are heterozygous carriers of ~30 repeats (Figure 7.4.1, Fig 7.4.2.3, Table 7.4.4). Interestingly, the *ATXN2* variation seems to segregate with spasticity as non-carriers are unaffected and are older than the typical age of onset with the exception of III-6, who may possibly be presymptomatic as he possesses both *C9orf72* and *ATXN2* expansions.

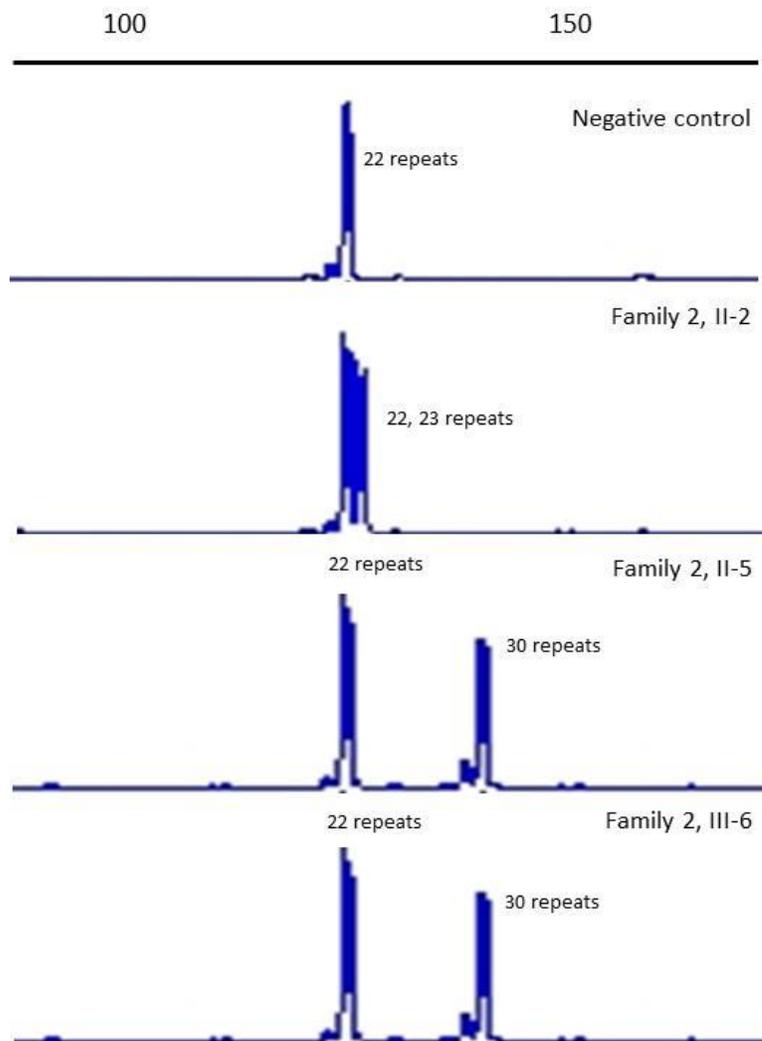


Figure 7.4.2.3 Genetic analysis of *ATXN2* genotyping profiles of family 2.

Results from fluorescent fragment length analyses. Repeat numbers are beside each peak and are calculated by subtracting 54 from the number determined using the horizontal axis. A negative control shows homozygosity for repeat length <30. The other panels show *ATXN2* expansion carrier status in family 2. The vertical and horizontal axes represent fluorescent intensity and product size, respectively.

Concurrently, we used a targeted neurodegeneration sequencing panel to identify any additional genetic variation in the three families. Previously, members of the Strong lab have identified *ARHGEF28* p.Lys280Met>fs40Ter (Droppelmann et al., 2013) in III-5. As a result, we subsequently genotyped additional family members to determine if they are carriers of *ARHGEF28* p.Lys280Met>fs40Ter. In doing so, we identified an additional eight *ARHGEF28* p.Lys280Met>fs40Ter heterozygous carriers (II-9, IV-4, IV-7, IV-13, IV-16, IV-17, V-2, and V-4) (Figure 7.4.1, Figure 7.4.2.4, Table 7.4.4). We also identified *ARHGEF28* p.Lys280Met>fs40Ter in two unrelated spouses (IV-4 and IV-7), which may be explained by common ancestral origins as both individuals are from the same Northern Netherlands region as other members of the family. Based on family history, there are no known consanguineous unions in family 1.

In family 2 and family 3, we identified a novel *OPTN* mutation (p.Met468Arg) (Figure 7.4.2.2). Variants in *OPTN* have been previously reported to be associated with ALS and FTD (Cirulli et al., 2015; Maruyama et al., 2010; Pottier et al., 2015). We subsequently genotyped other members of the families and in total, observed *OPTN* p.Met468Arg in four individuals (family 2, II-3; family 3, III-3, III-5, and III-7) with ALS-FTD who are also *C9orf72* carriers and one who is also an *ATXN* expansion carrier (family 2, II-3) (Figure 7.4.1, Table 7.4.4). We also observed the variant in an individual (IV-4) who is not a *C9orf72* carrier and is currently free of any disease symptoms.

ARHGEF28 p.Lys280Met>fs40Ter and *OPTN* p.Met468Arg are both very rare variants based on their absence from public databases archiving genetic information on individuals free of neurodegenerative disease ExAC as well as diseased individuals, including Human Gene Mutation Database (HGMD), ClinVar, and ALS Online genetics Database (ALSoD). Based on the location of the mutation (amino acid 280 of 1731) leading to the eventual termination of the RGNEF polypeptide, it is likely to disrupt protein function. *OPTN* p.Met468Arg is harboured within a highly conserved ubiquitin-binding domain and is predicted to exert a damaging effect on *OPTN* function based on a compilation of *in silico* software.

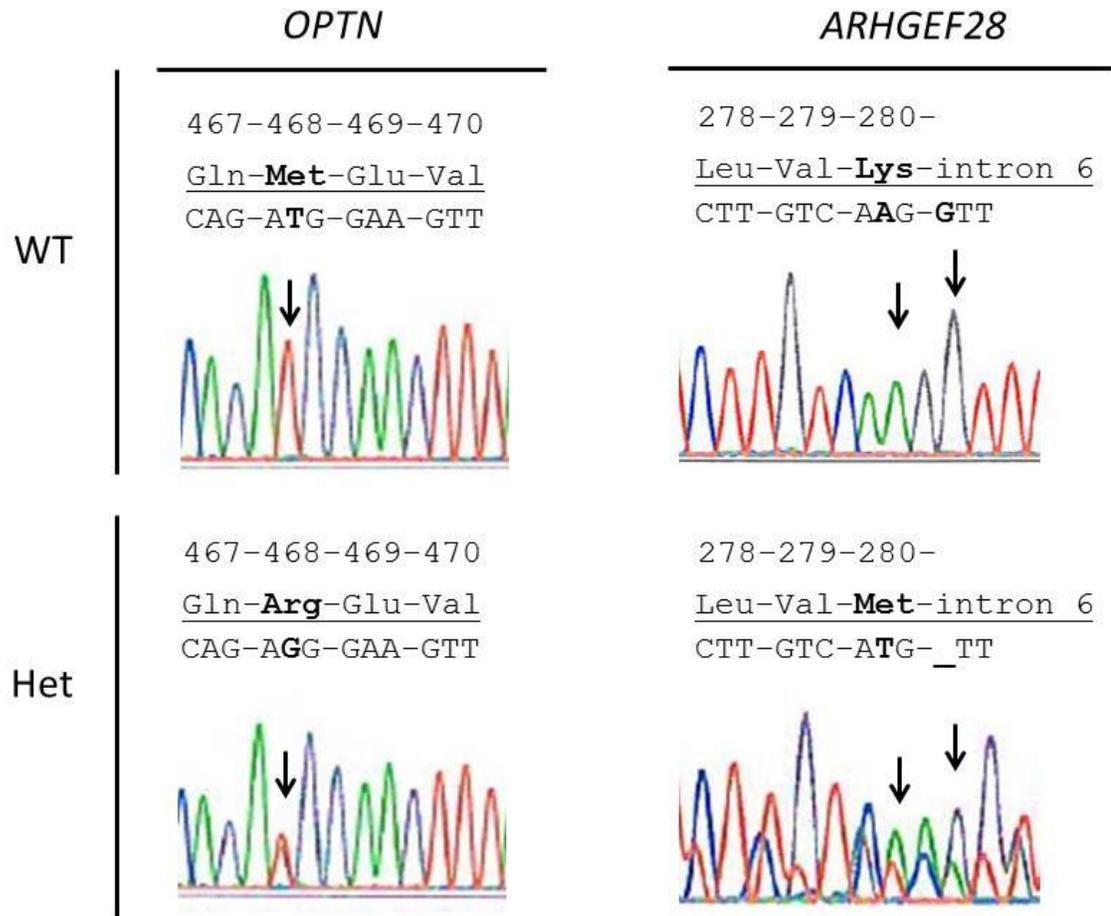


Figure 7.4.2.4 Validation of variants in *OPTN* and *ARHGEF28*.

Sanger sequencing electropherograms of *OPTN* p.Met468Arg and *ARHGEF28*

p.Lys280Met>fs40Ter in wild type and heterozygous carriers as observed in the families.

7.4.3 *C9orf72* dipeptide immunostaining

Cerebellar sections from each case were obtained from archived neuropathological specimens and immunostained for dipeptide repeat proteins [poly(GP), poly(GA), and poly(GR)] produced from *C9orf72* G₄C₂ expansions through repeat associated non-ATG (RAN) translation. In each instance, dipeptide repeat proteins were observed (Figure 7.4.3).

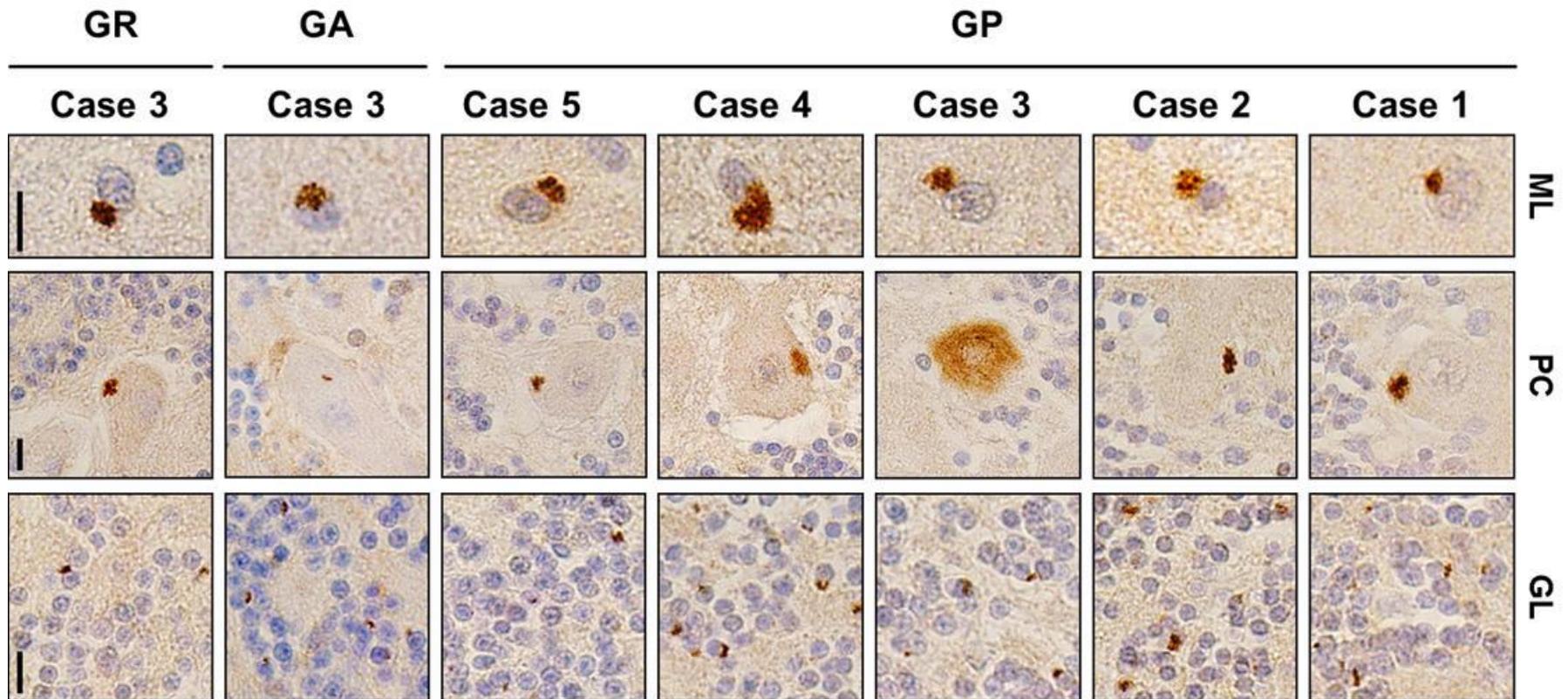


Figure 7.4.3 Cerebellar dipeptide repeat protein pathology.

Characteristic neuropathological lesions immunopositive for poly(GP), poly(GA) or poly(GR) proteins in Purkinje cells (PC), the granule cell layer (GL) or the molecular layer (ML) of the cerebellum. Scale bar 10 μ m. Case 1 is family 2, II-3; cases 2 and 3 are family 1, III-2 and III-5, respectively; and cases 4 and 5 are family 3, III-5 and III-7, respectively.

7.4.4 Oligogenic inheritance in patients

While the relationship between *C9orf72* expansion length and phenotype remains unclear, the expansions observed are unlikely to explain the inter- and intrafamilial clinical heterogeneity as all three families possessed expansions of similar size (Hardy and Rogaeva, 2014; Rohrer et al., 2015). It is more likely that the *C9orf72* expansion carriers within this range are at a greater risk and the burden of additional variants in other ALS causing genes further modulates the expression of *C9orf72* or disrupts the same pathway.

In family 1, the index patient (III-5) carried a *C9orf72* expansion as well as *ARHGEF28* p.Lys280Met>fs40Ter and presented with ALS 12 years earlier than his sister (III-2) who carried only a *C9orf72* expansion. Their brother presented with ALS at age 47, two years earlier than the index patient and 14 years earlier than his sister. However, it is uncertain whether he carried any genetic variation, as his DNA was unavailable. All members of the subsequent generation are currently unaffected and are younger than the typical age of onset. However, it is likely that *C9orf72* expansion carriers are currently presymptomatic and may develop ALS related symptoms when they become ~60 years of age. Carriers of *C9orf72* expansions as well as *ARHGEF28* p.Lys280Met>fs40Ter may develop ALS much earlier than individuals who only carry a *C9orf72* expansion.

In family 2, the index patient (II-3) carried a *C9orf72* expansion, *ATXN2* expansion, and *OPTN* p.Met468Arg. He presented with ALS at age 53 and was found to have frontotemporal dysfunction on neuropsychological testing. His disease duration was less than 2 years. His sister's death was also due to FTD with ALS being found at autopsy. While she was documented as harbouring a *C9orf72* expansion, this was done elsewhere and we were unable to obtain DNA for further analysis. II-5 carries an *ATXN2* expansion and has myelopathy, spasticity, and progressive cognitive decline. The proband's son (III-6) carries both a *C9orf72* expansion and an *ATXN2* expansion and is likely presymptomatic as he is currently younger than the typical age of onset within this pedigree.

The clinical presentation was consistent in all affected individuals in family 3 in that all three patients initially presented with features of bvFTD and then progressed to either ALS or a diffuse motor neuron disease marked by prominent spasticity. Death occurred within 1 to 4 years following symptom onset. All affected individuals were carriers of a *C9orf72* expansion as well as *OPTN* p.Met468Arg. Members of the subsequent generation currently do not display features suggestive of ALS and were screened for both variations. We observed two individuals who each carried one variant, but no individual carried both variants. Based on the effect of the *C9orf72* expansion, it is likely that the individual with this variant is at risk for developing ALS or FTD. However, whether the individual carrying only *OPTN* p.Met468Arg will develop ALS or bvFTD remains uncertain.

Table 7.4.4 Summary of clinical and genetic information.

ID	Family 1	Family 2	Family 3
N, affected	6	2	8
N, relatives affected with other phenotypes	1, cerebral palsy	1, myelopathy; 1 Parkinson's disease	2, Alzheimer's disease
Primary diagnosis	ALS	ALS	bvFTD
Diagnosis subcategory		FTD	ALS
Site of onset	limb	bulbar	mixed, primarily behavioural changes
Female:male	3:3	1:1	3:5
Mean age of onset (years±SD)	52 (±6)	54.4 (±2)	55 (±8)
Minimum age of onset (years)	47		35
Maximum age of onset (years)	60		64
Average disease duration (months±SD)	66 (±16)	29 (±7)	25 (±8)
No. of <i>C9orf72</i> expansion carriers	8	4	3
No. of <i>ATXN2</i> expansion carriers	0	3	0
No. of <i>OPTN</i> p.Met468Arg carriers	0	1	4
No. of <i>ARHGEF28</i> p.Lys280Met>fs40Ter carriers	9	0	0

7.5 Discussion

We have identified three families with multiple generations affected with ALS or ALS-FTD. All three families are carriers of a *C9orf72* expansion of similar repeat size but present with distinct clinical features. Although the length of the pathological expansion falls within the equivocal range with respect to predicting disease occurrence, neuropathological examination confirmed the presence of inclusions formed of dipeptide repeat proteins within cerebellar tissue. Thus, the *C9orf72* expansions reported here are pathogenic.

Given the remarkable phenotypic heterogeneity observed here, we next explored whether additional genetic variations contributed to this heterogeneity. We used a neurodegeneration targeted next generation sequencing gene panel to sequence affected individuals and their families. We identified two additional mutations that may confer additional disease risk. In family one, the index patient carried both a *C9orf72* expansion and an *ARHGEF28* variant, and developed ALS 12 years younger than his sister who possessed only the *C9orf72* expansion. In family two, the proband carried *C9orf72* and *ATXN2* expansions and a variant in *OPTN*, which collectively may have contributed to his earlier onset of ALS and FTD. Finally, in family three, the disease course in individuals III-3, III-5, and III-7 was very similar and all carried a *C9orf72* expansion in addition to an *OPTN* variant. Therefore, the genetic variants are segregating with distinct clinical presentations in all three families and a greater mutation burden is likely mediating an atypical ALS or ALS-FTD disease course.

In addition to our small sample size, one clear limitation of validating our findings is the inherent complexity involved in quantifying qualitative traits such as abnormal behaviour or progressive cognitive decline. However, our findings highlight the importance of agnostically screening all at risk patients for all known neurodegenerative disease genes. Although today this approach remains relatively costly, recent targeted next generation sequencing approaches like the ONDRISeq neurodegeneration gene panel facilitates these approaches on a rapid, efficient, and economical basis. Clinicians faced with patients who have atypical disease presentation are encouraged to consider additional mutations in related genes to identify possible oligogenic interactions. Table

7.5 summarizes findings of studies that conducted additional genetic testing in patients (Bury et al., 2015; Chio et al., 2012; Kaivorinne et al., 2014; King et al., 2013; Lashley et al., 2014; Pottier et al., 2015; Tarlarini et al., 2015; van Blitterswijk et al., 2012a; van Blitterswijk et al., 2012b). By identifying additional genetic variation, complex phenotypes can be explained by the additive effect of variants with large or small, yet significant effects that converge on the same disease pathway.

Table 7.5 ALS and/or FTD patients with multiple genetic variants.

Study	Diagnosis	Other symptoms	Site of onset	Sex	Age at onset (years)	Disease duration (months)	Ethnic origin	<i>C9orf72</i> expansion	Other variation	
Chio et al., 2012	ALS	FTD	Bulbar	M	43		Sardinia	Yes	<i>TARDBP</i> p.A382T	
	ALS	Mild FTD	Spinal	F	69	43	Sardinia	Yes	<i>TARDBP</i> p.A382T	
	ALS	Flail arm variant	Upper limb	M	35		Sardinia	Yes	<i>TARDBP</i> p.A382T	
van Blitterswijk et al., 2012	ALS						NL	Yes	<i>VAPB</i> p.V234I	
van Blitterswijk et al., 2012	ALS		Cervical	M	42	>91	NL	Yes	<i>TARDBP</i> p.N352S	
	ALS		Cervical	F	47	>15	NL	Yes	<i>TARDBP</i> p.N352S	
	ALS		Cervical	M	61	8	NL	No	(1) <i>TARDBP</i> p.N352S; (2) <i>ANG</i> p.K17I	
	ALS		Lumbosacral	F	73	47	NL	No	(1) <i>TARDBP</i> p.N352S; (2) <i>ANG</i> p.K17I	
	ALS		Cervical	M	54	52	NL	No	(1) <i>TARDBP</i> p.N352S; (2) <i>ANG</i> p.K17I	
	ALS		Cervical	F	51	77	NL	Yes	<i>SOD1</i> p.D90A	
	ALS		Lumbosacral	F	53	>24	NL	Yes	(1) <i>FUS</i> p.R521C; (2) <i>ANG</i> p.K17I	
	ALS		Lumbosacral	M	58	25	NL		<i>FUS</i> p.Q210H	
	King et al., 2013	FTD		Behavioural	F	57	~144		Yes	<i>MAPT</i> p.A239T
		ALS	FTD	Lower limb	M	60	~24		Yes	<i>MAPT</i> p.A239T
ALS		FTD	Lower limb	M	51	~24		Yes	<i>MAPT</i> p.A239T	
Kaivorinne et al., 2014	FTD		Behavioural	M	54	>120	Finland	Yes	<i>TARDBP</i> p.S292del	
	FTD		Behavioural	F	47	>36	Finland	Yes	<i>TARDBP</i> p.S292del	
Lashley et al., 2014	ALS	FTD		F	59	~84	UK	Yes	<i>GRN</i> p.C31fs	
	ALS	FTD		F	46	~36	UK	Yes	<i>GRN</i> p.C31fs	
Bury et al., 2015	ALS	FTD					UK	Yes	<i>OPTN</i> p.E322K	
Tarlarini et al., 2015	ALS	FTD	Bulbar	F	68	38	Italian	Yes	<i>FUS</i> p.R491C	

Pottier et al., 2015	FTD	agPPA	M	68	~48	No	(1) <i>OPTN</i> p.Gly538Glufs*27; (2) <i>TBKI</i> p.Arg117*
-------------------------	-----	-------	---	----	-----	----	---

Abbreviations are as follows: M, male; F, female; NL, Netherlands; UK, United Kingdom; agPPA, progressive non fluent/agrammatic variant of primary progressive aphasia.

7.6 Conclusion

In this Chapter, we describe three families affected with ALS with or without FTD among other neurodegenerative symptoms. While each family harbours a similar *C9orf72* expansion, each family varies significantly with respect to their mutation load. We used an agnostic approach by determining the families' *C9orf72* and *ATXN2* expansion profiles in addition to sequencing all known ALS-associated disease genes. We found that all three families carry additional genetic variation and attribute the phenotypic variability to oligogenic inheritance. It is likely that across all three families, the *C9orf72* expansion is the main driver of disease expression as all affected individuals carry the expansion. However, there are presymptomatic *C9orf72* expansion carriers and we intend to follow these study participants as they age and continue to visit Dr. Strong's clinic. If they never go on to develop ALS and/or FTD, it is likely that the presence of a *C9orf72* expansion is not sufficient to cause disease in these pedigrees. From our results, we would expect all *C9orf72* expansion carriers to develop ALS however, we also expect that individuals who carry additional genetic variants, which are likely acting as genetic modifiers, will develop ALS and/or FTD earlier or will likely present with an atypical disease course.

7.7 References

- Brooks, B.R. (1994). El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J Neurol Sci* 124 Suppl, 96-107.
- Bury, J.J., Highley, J.R., Cooper-Knock, J., Goodall, E.F., Higginbottom, A., McDermott, C.J., Ince, P.G., Shaw, P.J., and Kirby, J. (2015). Oligogenic inheritance of optineurin (OPTN) and C9ORF72 mutations in ALS highlights localisation of OPTN in the TDP-43-negative inclusions of C9ORF72-ALS. *Neuropathology*.
- Chio, A., Restagno, G., Brunetti, M., Ossola, I., Calvo, A., Canosa, A., Moglia, C., Floris, G., Tacconi, P., Marrosu, F., *et al.* (2012). ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations. *J Neurol Neurosurg Psychiatry* 83, 730-733.

Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.F., Wang, Q., Krueger, B.J., *et al.* (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 347, 1436-1441.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., *et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245-256.

Droppelmann, C.A., Wang, J., Campos-Melo, D., Keller, B., Volkening, K., Hegele, R.A., and Strong, M.J. (2013). Detection of a novel frameshift mutation and regions with homozygosity within ARHGEF28 gene in familial amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener* 14, 444-451.

Hardiman, O., van den Berg, L.H., and Kiernan, M.C. (2011). Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat Rev Neurol* 7, 639-649.

Hardy, J., and Rogaeva, E. (2014). Motor neuron disease and frontotemporal dementia: sometimes related, sometimes not. *Exp Neurol* 262 Pt B, 75-83.

Ittner, L.M., Halliday, G.M., Kril, J.J., Gotz, J., Hodges, J.R., and Kiernan, M.C. (2015). FTD and ALS--translating mouse studies into clinical trials. *Nat Rev Neurol* 11, 360-366.

Kaivorinne, A.L., Moilanen, V., Kervinen, M., Renton, A.E., Traynor, B.J., Majamaa, K., and Remes, A.M. (2014). Novel TARDBP sequence variant and C9ORF72 repeat expansion in a family with frontotemporal dementia. *Alzheimer Dis Assoc Disord* 28, 190-193.

King, A., Al-Sarraj, S., Troakes, C., Smith, B.N., Maekawa, S., Iovino, M., Spillantini, M.G., and Shaw, C.E. (2013). Mixed tau, TDP-43 and p62 pathology in FTLD associated with a C9ORF72 repeat expansion and p.Ala239Thr MAPT (tau) variant. *Acta Neuropathol* 125, 303-310.

Lashley, T., Rohrer, J.D., Mahoney, C., Gordon, E., Beck, J., Mead, S., Warren, J., Rossor, M., and Revesz, T. (2014). A pathogenic progranulin mutation and C9orf72 repeat expansion in a family with frontotemporal dementia. *Neuropathol Appl Neurobiol* 40, 502-513.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., *et al.* (2010). Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* 465, 223-226.

Pottier, C., Bieniek, K.F., Finch, N., van de Vorst, M., Baker, M., Perkersen, R., Brown, P., Ravenscroft, T., van Blitterswijk, M., Nicholson, A.M., *et al.* (2015). Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. *Acta Neuropathol* 130, 77-92.

Renton, A.E., Chio, A., and Traynor, B.J. (2014). State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 17, 17-23.

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., *et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257-268.

Rohrer, J.D., Isaacs, A.M., Mizielinska, S., Mead, S., Lashley, T., Wray, S., Sidle, K., Fratta, P., Orrell, R.W., Hardy, J., *et al.* (2015). C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *Lancet Neurol* 14, 291-301.

Tarlarini, C., Lunetta, C., Mosca, L., Avemaria, F., Riva, N., Mantero, V., Maestri, E., Quattrini, A., Corbo, M., Melazzini, M.G., *et al.* (2015). Novel FUS mutations identified through molecular screening in a large cohort of familial and sporadic amyotrophic lateral sclerosis. *Eur J Neurol* 22, 1474-1481.

van Blitterswijk, M., van Es, M.A., Hennekam, E.A., Dooijes, D., van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., *et al.* (2012a). Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet* 21, 3776-3784.

van Blitterswijk, M., van Es, M.A., Koppers, M., van Rheenen, W., Medic, J., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., Veldink, J.H., and van den Berg, L.H. (2012b). VAPB and C9orf72 mutations in 1 familial amyotrophic lateral sclerosis patient. *Neurobiol Aging* 33, 2950 e2951-2954.

Woolley, S.C., and Strong, M.J. (2015). Frontotemporal Dysfunction and Dementia in Amyotrophic Lateral Sclerosis. *Neurol Clin* 33, 787-805.

Xi, Z., Zinman, L., Grinberg, Y., Moreno, D., Sato, C., Bilbao, J.M., Ghani, M., Hernandez, I., Ruiz, A., Boda, M., *et al.* (2012). Investigation of c9orf72 in 4 neurodegenerative disorders. *Arch Neurol* 69, 1583-1590.

Chapter 8 - Discussion

8 Summary

The studies presented in this PhD Thesis have identified genetic variants underlying Mendelian diseases and genetic variation that may predispose individuals ascertained in these studies, to complex diseases, primarily neurodegenerative disorders. The context of each study findings, the strengths, implications, as well as limitations of my dissertation, and future goals, will be discussed in subsequent sections.

8.1 Context of study findings

In this section, I will highlight the conclusions and limitations from each chapter.

8.1.1 Chapter 2: *NFS1*, a novel disease gene underlying a novel metabolic disease

The findings of Chapter 2 include the discovery of a novel disease gene, *NFS1* as the underlying genetic basis of a novel mitochondrial disease, infantile mitochondrial complex II/III deficiency (IMC23D) (Farhan et al., 2014b). Variants in genes encoding Fe-S proteins can disrupt Fe-S cluster assembly, which may result in various disorders such as Friedreich's ataxia (Koeppen, 2011) or ISCU myopathy (Kollberg et al., 2009). In Chapter 2, I describe for the first time, the clinical presentation of patients affected by IMC23D. The application of a classic method, autozygosity mapping, and a more modern sequencing approach, exome sequencing, accelerated gene discovery. Using these experiments, I observed p.Arg72Gln in *NFS1*, in a candidate homozygous region on chromosome 20. *NFS1* encodes a cysteine desulfurase, which is a Fe-S protein involved in Fe-S cluster biogenesis, iron maintenance, and DNA repair (Schmucker et al., 2011). Collectively, the results of Chapter 2 describe the first case of *NFS1* deficiency in

humans and further demonstrate the importance of sufficient NFS1 expression in human physiology.

While I conducted *ex vivo* experiments using patient cells to determine the effect exerted by NFS1 p.Arg72Gln, I did not recapitulate the results in an *in vitro* or an *in vivo* model. I leveraged previous experiments that consistently demonstrated Nfs1 deficiency leads to abnormal mitochondria morphology and function (Biederbick et al., 2006; Fosset et al., 2006). In patient cells, NFS1 p.Arg72Gln led to loss of function of the NFS1 protein. Accordingly, I inferred that the same molecular defects were likely also occurring in patients with NFS1 deficiency. However, these experiments would have provided further evidence in support of an NFS1-mediated mitochondrial disease. The clinical features observed in the patients were consistent with the features of other patients diagnosed with mitochondrial diseases caused by Fe-S protein deficiency, which converge on the same pathway as NFS1. Taken together, the genetic data, the overlapping clinical features in other patients affected with Fe-S protein diseases, and the known role of NFS1 in mitochondrial function, suggest that NFS1 deficiency is the underlying mechanism of IMC23D.

8.1.2 Chapter 3: Expanding the biological function of *EXT2*, a known disease gene in a novel neurodevelopmental disease

Previous to the findings from this Chapter, heterozygous variants in *EXT2* had been recurrently implicated only in autosomal dominant multiple exostoses (Busse-Wicher et al., 2014). In Chapter 3, I described the first instance of an autosomal recessive *EXT2*-mediated neurodevelopmental disorder classified as seizures-scoliosis-macrocephaly (SSM) syndrome (Farhan et al., 2015). Similar to the methods discussed in Chapter 2, given the known consanguineous union between the parents, I used autozygosity mapping to identify homozygous regions in the genome, which I subsequently applied exome sequencing, to expedite variant discovery. Because of the known association of *EXT2* in another Mendelian disorder, I studied the effect of *EXT2* variants on protein expression and function in patient cells, and also using an *in vitro* system to ensure I did not merely identify rare, benign variants in a disease gene. These experiments yielded

results consistent with *EXT2* loss of function as the probable mechanism leading to SSM syndrome. Thus, I have implicated *EXT2* in a novel disorder (now available on OMIM #616682) and have further illustrated the importance of heparan sulfate biosynthesis enzymes in neurodevelopment.

Despite performing functional assays to validate the impact of the *EXT2* variants on protein expression and function, I did not test whether there were other *EXT2* interactions that were also comprised. I also did not investigate whether integration of wild type *EXT2* variants are able to rescue these interactions. Finally, while I reviewed the literature for *EXT2* model organisms and discussed these in the manuscript, I did not perform any *in vivo* replication studies. These experiments would have strengthened the evidence of gene-based causality.

8.1.3 Chapter 4: Confirming the association of *KCTD7* in progressive myoclonus epilepsy

Several independent studies have implicated *KCTD7* variants in progressive myoclonus epilepsy (Azizieh et al., 2011; Blumkin et al., 2012; Kousi et al., 2012; Krabichler et al., 2012; Staropoli et al., 2012). In this study, I also used autozygosity mapping and exome sequencing to identify a novel homozygous variant, p.Tyr276Cys in *KCTD7* in three affected females (Farhan et al., 2014a). The results confirm previous reports describing the role of *KCTD7* in epilepsy. As I discussed in Chapter 4, *KCTD7* p.Tyr276Cys is now a genetic marker and can be used in genetic counseling for the siblings and their prospective mates and for other members of the same Mennonite community. Interestingly, following the publication of our manuscript, a local medical geneticist, Dr. Sharan Goobie, identified another family from the same Mennonite community with progressive myoclonus epilepsy. Affected individuals were immediately screened for the same variant, p.Tyr276Cys in *KCTD7* and were also homozygous carriers. Our initial observation and report allowed for the prioritization of the variant, which is likely also causing the epilepsy phenotype in her patients. In addition, as discussed in Chapter 4, prior to identification of the genetic variant, the patients received a different diagnosis.

Our findings helped guide clinicians to correctly diagnose the patients with progressive myoclonus epilepsy in keeping with previous reports.

Because *KCTD7* is a previously known epilepsy gene, I did not pursue any functional validation steps. This was also due to the specific brain expression of *KCTD7*, which limited us from obtaining any relevant tissues from the patients. However, future steps may include genetic engineering of *KCTD7* p.Tyr276Cys in neural specific cell lines to examine whether the variant exerts a loss of gain of protein function therefore, determining the mechanism of epileptogenesis.

8.1.4 Chapter 5: The application of a model organism to understand the function of a novel disease gene *TMTC3*, in neurodevelopment

While the results of Chapter 5 are not yet finalized, we have identified several key features of the disease. Firstly, I observed two compound heterozygous variants in *TMTC3*, which encodes transmembrane and tetratricopeptide repeat containing 3 protein, in all four affected individuals. When measuring *TMTC3* protein expression, we observed abolished expression in patient cells compared to the unaffected parents and a healthy, homozygous wild type control. Therefore, *TMTC3* loss of function is the most probable disease mechanism. We are also currently studying the effect of *TMTC3* deficiency in *Drosophila melanogaster* and *TMTC3* localization in collaboration with Dr. Jamie Kramer and Dr. Michael Poulter. Preliminary results from the Kramer lab suggest that neuronal knockdown of *Drosophila melanogaster Tmtc3* causes increased susceptibility to mechanically induced seizures. In addition, preliminary results from the Poulter lab suggest *TMTC3* is specifically localized at synapses in the rat brain. While *TMTC3* is a poorly characterized gene, loss of function in patients with nocturnal seizures with developmental delay emphasizes an important role for *TMTC3* in neurodevelopment.

8.1.5 Chapter 6: Developing a custom sequencing based method to study neurodegeneration

In Chapter 6, I discuss our implementation of a next generation sequencing based neurodegeneration gene panel that is currently being used by a large, multidisciplinary study: The Ontario Neurodegenerative Disease Research Initiative (ONDRI). The approach of selecting the 80 genes within ONDRISeq was based on their known association with neurodegeneration. In this study, we processed DNA from 216 patients diagnosed with one of the five diseases: 1) Alzheimer's disease/mild cognitive impairment; 2) amyotrophic lateral sclerosis; 3) frontotemporal dementia; 4) Parkinson's disease; or 5) vascular cognitive impairment. We also independently validated a subset of the variant calls using multiple, independent genomic approaches including: NeuroX, TaqMan allelic discrimination assay, or Sanger sequencing. We observed 100% concordance for the variants identified via ONDRISeq and screened by other genomic technologies. We were successful in detecting known as well as novel rare variants in a large fraction of cases. Importantly, not all the variants observed are causal, and many need further validation.

There are several important limitations in this Chapter. While I did not discuss any environmental factors, the ONDRI study coordinators documented all potential environmental influences that may have induced neurodegeneration in the study participants. Upon further evaluation of genetic variants, we intend on including any potential environmental influences in addition to the data from the other ONDRI platforms to further characterize the neurodegenerative diseases. In addition, we omitted the majority of human variation. Specifically, we did not investigate any variation outside of the 80 genes, which would have enabled us to identify novel neurodegeneration loci. We also did not assess the role of structural variation, copy number variation, non-coding variation, or epigenetic variation on disease. Examination of these omitted classes of variation may explain the genetically unsolved cases in the study cohort. However, we are in the early stages of implementing an algorithm to detect copy number variation using ONDRISeq. We are investigating whether we can use the uniformity of the targeted gene panel and normalize across the target regions to obtain baseline expression. This

way, we can determine whether a sample has copy number variation at certain regions by detecting any statistically significant deviation from baseline. These techniques have proven to be useful when assessing copy number variation in patients with hereditary hearing loss (Sommen et al., 2016).

8.1.6 Chapter 7: The effect of multiple genetic variants on neurodegenerative disease risk

In Chapter 7, I discuss the probable genetic basis of neurodegeneration in three families. The three families are affected with ALS and/or FTD and in which the same hexanucleotide repeat expansion in the *C9orf72* gene has been identified, but their clinical manifestations vary significantly. The inter- and intrafamilial disparity in clinical phenotype ranged from a much earlier age of onset, ALS combined with FTD, or ALS with no additional features. Following genetic testing, I observed the presence of an *ATXN2* expansion, *ARHGEF28* p.Lys280Met>fs40Ter, or an *OPTN* p.Met468Arg in patients with a *C9orf72* expansion and who exhibit an atypical disease course. These genetic profiles segregated with distinct clinical presentations in all three families and are likely modifying the ALS phenotype. Perhaps the variability in the clinical expression of *C9orf72* expansion carriers is driven by the presence of these additional genetic variants further supporting the observation of oligogenic inheritance in ALS.

Within this Chapter, there were several limitations that must be considered. First, I did not perform any experiments demonstrating an additive or synergistic deleterious effect on cell viability. In the literature, additional ALS genetic variants that may act as modifiers of *C9orf72*, have been previously reported, strengthening the oligogenic inheritance hypothesis (Bury et al., 2015; Chio et al., 2012; Kaivorinne et al., 2014; King et al., 2013; Lashley et al., 2014; Pottier et al., 2015; Tarlarini et al., 2015; van Blitterswijk et al., 2012a; van Blitterswijk et al., 2012b). Indeed, variation in all four genes (*C9orf72*, *ATXN2*, *ARHGEF28*, and *OPTN*) has been previously implicated in ALS or FTD however, the observation of multiple variants in genes implicated in the same disease is still not sufficient to establish that the presence of multiple genetic variants leads to an atypical disease course.

In addition, while we had access to clinical history and DNA from three large families, I was not able to genotype all affected and unaffected members of the family (past and present), which would have strengthened or countered the observation that multiple variants clustered in individuals with an atypical disease course. I also postulate that the genetic variants observed in the subsequent generation would likely manifest into the neurodegenerative phenotype. This assumption is based on the genetic profiles of deceased individuals.

Another limitation of this study is the absence of knowledge regarding each family's overall neurodegeneration risk. I assumed that the presence and accumulation of these genetic variants was inducing disease however, it is unclear whether these families already possess a heightened neurodegeneration risk (i.e. reside at the tail of distribution) and the presence of these additional variants further increases risk of neurodegenerative disease. Knowledge of the baseline neurodegeneration risk conferred by common genetic variation would allow for the investigation of all neurodegeneration genetic risk factors.

Finally, in this study, I did not fully account for other confounding factors originating from each individual's or family's environment. We documented each participant's educational attainment, whether they have ever experienced a head or neck injury, any potential toxin exposure, and their general lifestyle, however we did not measure the effect of genetic variants coupled with these insults, on cell viability. These variables are intrinsically difficult to quantify however, knowledge of their presence can aid in understanding why individuals who do not possess any genetic risk factors can nevertheless develop neurodegenerative phenotypes. In this study, all affected individuals carried at least one of the genetic variants discussed in our study.

8.2 Methodological considerations

There were several strengths and implications as well as general limitations in my studies, which I will discuss in depth here. I will also outline potential solutions to overcome these limitations in the future.

8.2.1 Study strengths and implications

The results of this PhD thesis have helped further annotate the human genome, provided guidance for effective analytical approaches for gene discovery, and also provided genetic diagnoses for the affected families. In this PhD thesis, I characterized genes not previously associated with a disease; or associated with a well-characterized disease but now also underlie a novel phenotype; or implicated in the same phenotype and is now associated with additional clinical features. By defining the role of genetic variation in this subset of genes using conservative causality criteria, we are able to understand their normal biological function in the human genome more clearly.

In addition, by applying next generation sequencing, I was able to rapidly and efficiently provide a genetic diagnosis for the patients and their families. In this context, a genetic diagnosis is defined as the observation of DNA variation in patients and in which the variation detected, are in accordance with conservative causality criteria. Knowledge of the DNA variation penetrance and each individual's carrier status can help during genetic counseling and reproductive planning. In Chapters 2-5, the variants observed likely have a large effect on gene function and when transmitted in an autosomal recessive inheritance pattern, will cause the disease. In contrast, the late-onset neurodegenerative diseases described in Chapters 6 and 7, do not necessarily follow classical Mendelian disease inheritance patterns. It is likely that a burden of genetic variation within genes increases predisposition to disease. Therefore, the presence of a genetic variant in a disease gene does not always result in a disease phenotype. This discrepancy in simple, monogenic diseases and complex, oligogenic or polygenic diseases, may in part, be explained by incomplete penetrance, which makes the interpretation of DNA variation and genetic counseling immensely challenging.

Furthermore, by deeply phenotyping the patients and performing gene-specific assays, we are able to guide during drug development. Based on my work, industry or translational scientists who are interested in the genes described here, will have a greater understanding of how loss or gain of function in this subset of genes, can affect patients. Also, we can provide information as to whether inhibition or activation of these genes will be the most applicable therapeutic intervention. Finally, knowledge of the patients'

genetic information can make them eligible for clinical trials. For example, if industry or translational scientists have developed a therapeutic agent targeting *LRRK2*, a known Parkinson's disease gene, they may be interested in recruiting patients with *LRRK2* variants in addition to non-*LRRK2* carriers, to assess the efficacy of the drug.

8.2.2 Additional study caveats

There are additional study limitations that must be considered during the interpretation of the results generated in this PhD Thesis. First, in Chapters 2-5, given the rarity of each disease, the sample sizes were small and additional pedigrees were not available. To overcome the lack of replication in human samples especially for novel disease genes, I implemented objective metrics including identifying regions with high LOD scores during autozygosity mapping, and multiple functional assays to recapitulate the disease phenotype. Unfortunately, these limitations are common in rare disease research. Accordingly, many researchers have come together to form the Matchmaker Exchange, a rare disease global alliance that facilitates gene discovery among researchers by data sharing (Philippakis et al., 2015). This initiative allows researchers to connect and collaborate on novel or unexplained phenotypes by increasing their sample sizes.

Second, during these studies I derived genetic information by targeting the exome, which contains protein-coding variation only. There are valid biological reasons to prioritize variation within the exome, which are in parallel with the study design. For example, the overwhelming majority of Mendelian phenotypes are caused by DNA variants harbored within the exome, making exome sequencing the preferred sequencing approach (Chong et al., 2015; Ng et al., 2009). However, the selection of exome sequencing was also motivated by practical considerations. When we initially planned the experiments in 2012, the cost of sequencing one exome or genome were ~\$4,000 and ~\$10,000, respectively. In addition to these expenses, the computational requirements to store the data, manipulate, and analyze the data, were extremely resource intensive. Because of these limitations, I implemented strategic approaches in selecting only a few samples, which were often the most severely affected patients, for exome sequencing. Without these barriers, which are beginning to be resolved as sequencing costs decrease

and computational approaches become less rigorous, I would have preferred to be unbiased and investigate the entire genome. Moreover, whole genome sequencing can detect indels, structural variation, non-coding variation, copy number variation, or multinucleotide expansions, which are often missed by exome sequencing (Belkadi et al., 2015; Singleton, 2011).

Third, one of the greatest limitations of working with human samples is the inaccessibility to all affected tissues. These challenges are unlikely to be resolved until we develop less invasive extraction methods. Using these theoretical methods, we may obtain single cells or tissues and perform sequencing assays to identify any variation in expression. This approach is in part, already in place within the Genotype-Tissue Expression (GTEx) project, a control dataset of multiple human tissues from donors who are densely genotyped, to assess genetic variation within their genomes (Consortium, 2013).

Finally, I did not systematically control for most environmental effects. A key assumption underlying the approach taken in Chapters 2-5 was that the disorders were more likely caused by monogenic variants, which are highly penetrant and are typically independent of the environment. Thus, the strong genetic effects in these instances made it less critical to account for environmental influences. Conversely, while the neurodegenerative disorders described in Chapters 6 and 7, can be modified or even caused by environmental factors such as traumatic brain injury or toxin exposure, I did not adequately measure these possibly confounding variables.

8.3 Future directions

While there is much advancement achieved through my PhD Thesis, there are still many future directions applicable to my own work, and some priorities that can be outlined for gene discovery research in general, which I will summarize here.

(1) Open access to control datasets

As sequencing costs decrease, more large-scale sequencing projects are pursued, which allows for a larger control dataset to develop. These control dataset act as reference

samples and are extremely beneficial in extracting allele frequencies for variants of interest. The database of Genotypes and Phenotypes (dbGaP) is one example of a repository to archive genetic information from a collection of studies with ethics approval (Mailman et al., 2007; Tryka et al., 2014). These studies range from array-based association studies or large-scale next generation sequencing studies, which investigated the relationship of genotypes and phenotypes in cases and controls. Another example is the United Kingdom Biobank, however this can be a costly endeavor and the genetic data are based on only 820,967 genetic markers on the Axion Array (Sudlow et al., 2015). While there are other publically available resources to determine variant allele frequencies such as 1000 Genomes (Genomes Project et al., 2015), NHLBI Exome Sequencing Project (Fu et al., 2013), or ExAC (Lek et al., 2015), individual level genetic data are not easily accessible. Furthermore, the phenotypes associated with each sample are not provided, which makes case-control studies difficult to conduct and interpret. Finally, in general these reference databases do not have adequate representation of all ancestries, especially African and Middle Eastern populations.

(2) Interpretation of genetic variants

Our ability to sequence human genomes has exceeded our ability to adequately interpret the wealth of genetic variation in a single genome. We are limited by our ignorance regarding the underlying mechanism of disease. How do we interpret variants that are observed in patients but are also observed in controls? How do we interpret variants that exert their effects later in life? These questions fall into the hands of members of the American College of Medical Genetics and Genomics (ACMG). ACMG members have the burden of developing standards and guidelines, which are forced to progress at the same rate as technological advancements to prevent any false interpretations in clinical genetic testing (Green et al., 2013; Hampel et al., 2015; Rehm et al., 2013; Richards et al., 2015). These initiatives serve the entire scientific community by developing and upholding standards in the interpretation of genetic variation. Despite the committee's best efforts, many challenges lay ahead. One such challenge is determining the effect size of a genetic variant and whether its presence in a disease-free individual renders it benign or incompletely penetrant.

(3) Standardization of *in silico* predictive programs

One of the most common approaches to interpreting the potential effect of genetic variation is the implementation of a series of computational (*in silico*) predictive programs. *In silico* tools use algorithms to assess the effect of non-synonymous nucleotide changes on genes and proteins. In general, *in silico* tools that assess the effect of missense changes rely on amino acid structural changes and conservation across species lineages. Alternatively, *in silico* tools that predict the effect of nucleotide variation on splicing depend on the proximity of nucleotide changes to the exon/intron boundaries and implement mathematical models to quantify the resulting entropy change.

While many predictive programs implement multiple variables to assess the impact of nucleotide variation on the protein function, they are not always consistent among each other and perform poorly when evaluating insertions or deletions. Evidently, the most common predictive tools in clinical laboratories are PolyPhen2 (Adzhubei et al., 2013), SIFT (Sim et al., 2012), and MutationTaster (Schwarz et al., 2010). Some of these tools use the same underlying criteria and are likely to yield similar results; therefore, they should be used in combination rather than as independent lines of evidence, as recommended by ACMG (Richards et al., 2015). In addition, the underlying assumption is that deviation from the conserved amino acid is predictive of deleteriousness. While this assumption is perhaps valid when evaluating metabolic traits, it may not be useful when studying neurodevelopmental or cognitive traits unique to humans. Finally, to the best of my knowledge, most *in silico* tools do not incorporate variant zygosity. Refining these tools to match the biological state of disease will likely resolve some of these matters although more research is needed to develop the exact criteria that should be implemented.

(4) Technological advancements will allow us to be unbiased in variant detection

As the cost of sequencing decreases, researchers are better equipped to study genetic diseases. Here, I focused primarily on the exome, which is a major limitation throughout my dissertation. Having access to the entire genome will allow researchers to conduct their studies comprehensively and investigate whether there are copy number variation or

non-coding variation that are modifying the human phenotype. Furthermore, researchers can implement multiple levels of evidence when evaluating the functional effects of genetic variation by performing tissue specific sequencing to identify any somatic variation.

8.4 Final conclusions

In this thesis, I have furthered our understanding of several human diseases and their genetic basis. Specifically, in Chapters 2-5, by extracting rare, non-synonymous variants within the exome, I was able to delineate the signals, which in this context are large effect, monogenic variants, from the noise generated by the voluminous amount of benign variation. Additionally, I used a smaller-scale genomic approach, which allowed for the detection of genetic variation within a preselected list of genes, for the neurodegenerative disease studies in Chapters 6 and 7. While I implemented multiple objective metrics to prevent false discoveries, and to directly link the newly discovered DNA variation to the human phenotypes, much remains to be learned about the altered protein function or dosage leading to disease pathogenesis. Nevertheless, the genetic and genomic discoveries made using next generation sequencing methods in this PhD Thesis as well as in other studies, have tremendously improved the diagnostic yield while minimizing time and cost.

8.5 References

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics* / editorial board, Jonathan L Haines [et al] *Chapter 7, Unit7* 20.

Azizieh, R., Orduz, D., Van Bogaert, P., Bouschet, T., Rodriguez, W., Schiffmann, S.N., Pirson, I., and Abramowicz, M.J. (2011). Progressive myoclonic epilepsy-associated gene KCTD7 is a regulator of potassium conductance in neurons. *Molecular neurobiology* *44*, 111-121.

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., Antipenko, A., Shang, L., Boisson, B., Casanova, J.L., and Abel, L. (2015). Whole-genome sequencing is more

powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America* 112, 5473-5478.

Biederbick, A., Stehling, O., Rosser, R., Niggemeyer, B., Nakai, Y., Elsasser, H.P., and Lill, R. (2006). Role of human mitochondrial *Nfs1* in cytosolic iron-sulfur protein biogenesis and iron regulation. *Molecular and cellular biology* 26, 5675-5687.

Blumkin, L., Kivity, S., Lev, D., Cohen, S., Shomrat, R., Lerman-Sagie, T., and Leshinsky-Silver, E. (2012). A compound heterozygous missense mutation and a large deletion in the *KCTD7* gene presenting as an opsoclonus-myoclonus ataxia-like syndrome. *Journal of neurology* 259, 2590-2598.

Bury, J.J., Highley, J.R., Cooper-Knock, J., Goodall, E.F., Higginbottom, A., McDermott, C.J., Ince, P.G., Shaw, P.J., and Kirby, J. (2015). Oligogenic inheritance of optineurin (*OPTN*) and *C9ORF72* mutations in ALS highlights localisation of *OPTN* in the TDP-43-negative inclusions of *C9ORF72*-ALS. *Neuropathology : official journal of the Japanese Society of Neuropathology*.

Busse-Wicher, M., Wicher, K.B., and Kusche-Gullberg, M. (2014). The exostosin family: proteins with many functions. *Matrix biology : journal of the International Society for Matrix Biology* 35, 25-33.

Chio, A., Restagno, G., Brunetti, M., Ossola, I., Calvo, A., Canosa, A., Moglia, C., Floris, G., Tacconi, P., Marrosu, F., *et al.* (2012). ALS/FTD phenotype in two Sardinian families carrying both *C9ORF72* and *TARDBP* mutations. *Journal of neurology, neurosurgery, and psychiatry* 83, 730-733.

Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., *et al.* (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *American journal of human genetics* 97, 199-215.

Consortium, G.T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature genetics* 45, 580-585.

Farhan, S.M., Murphy, L.M., Robinson, J.F., Wang, J., Siu, V.M., Rupar, C.A., Prasad, A.N., Consortium, F.C., and Hegele, R.A. (2014a). Linkage analysis and exome sequencing identify a novel mutation in *KCTD7* in patients with progressive myoclonus epilepsy with ataxia. *Epilepsia* 55, e106-111.

Farhan, S.M., Wang, J., Robinson, J.F., Lahiry, P., Siu, V.M., Prasad, C., Kronick, J.B., Ramsay, D.A., Rupar, C.A., and Hegele, R.A. (2014b). Exome sequencing identifies *NFS1* deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency. *Molecular genetics & genomic medicine* 2, 73-80.

- Farhan, S.M., Wang, J., Robinson, J.F., Prasad, A.N., Rupa, C.A., Siu, V.M., Consortium, F.C., and Hegele, R.A. (2015). Old gene, new phenotype: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses. *Journal of medical genetics* 52, 666-675.
- Fosset, C., Chauveau, M.J., Guillon, B., Canal, F., Drapier, J.C., and Bouton, C. (2006). RNA silencing of mitochondrial m-Nfs1 reduces Fe-S enzyme activity both in mitochondria and cytosol of mammalian cells. *The Journal of biological chemistry* 281, 25398-25406.
- Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., *et al.* (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216-220.
- Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., *et al.* (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* 15, 565-574.
- Hampel, H., Bennett, R.L., Buchanan, A., Pearlman, R., Wiesner, G.L., Guideline Development Group, A.C.o.M.G., Genomics Professional, P., Guidelines, C., and National Society of Genetic Counselors Practice Guidelines, C. (2015). A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: referral indications for cancer predisposition assessment. *Genetics in medicine : official journal of the American College of Medical Genetics* 17, 70-87.
- Kaivorinne, A.L., Moilanen, V., Kervinen, M., Renton, A.E., Traynor, B.J., Majamaa, K., and Remes, A.M. (2014). Novel TARDBP sequence variant and C9ORF72 repeat expansion in a family with frontotemporal dementia. *Alzheimer disease and associated disorders* 28, 190-193.
- King, A., Al-Sarraj, S., Troakes, C., Smith, B.N., Maekawa, S., Iovino, M., Spillantini, M.G., and Shaw, C.E. (2013). Mixed tau, TDP-43 and p62 pathology in FTL associated with a C9ORF72 repeat expansion and p.Ala239Thr MAPT (tau) variant. *Acta neuropathologica* 125, 303-310.
- Koeppen, A.H. (2011). Friedreich's ataxia: pathology, pathogenesis, and molecular genetics. *Journal of the neurological sciences* 303, 1-12.

Kollberg, G., Tulinius, M., Melberg, A., Darin, N., Andersen, O., Holmgren, D., Oldfors, A., and Holme, E. (2009). Clinical manifestation and a new ISCU mutation in iron-sulphur cluster deficiency myopathy. *Brain : a journal of neurology* *132*, 2170-2179.

Kousi, M., Anttila, V., Schulz, A., Calafato, S., Jakkula, E., Riesch, E., Myllykangas, L., Kalimo, H., Topcu, M., Gokben, S., *et al.* (2012). Novel mutations consolidate KCTD7 as a progressive myoclonus epilepsy gene. *Journal of medical genetics* *49*, 391-399.

Krabichler, B., Rostasy, K., Baumann, M., Karall, D., Scholl-Burgi, S., Schwarzer, C., Gautsch, K., Spreiz, A., Kotzot, D., Zschocke, J., *et al.* (2012). Novel mutation in potassium channel related gene KCTD7 and progressive myoclonic epilepsy. *Annals of human genetics* *76*, 326-331.

Lashley, T., Rohrer, J.D., Mahoney, C., Gordon, E., Beck, J., Mead, S., Warren, J., Rossor, M., and Revesz, T. (2014). A pathogenic progranulin mutation and C9orf72 repeat expansion in a family with frontotemporal dementia. *Neuropathology and applied neurobiology* *40*, 502-513.

Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., *et al.* (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature genetics* *39*, 1181-1186.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* *461*, 272-276.

Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., *et al.* (2015). The Matchmaker Exchange: a platform for rare disease gene discovery. *Human mutation* *36*, 915-921.

Pottier, C., Bieniek, K.F., Finch, N., van de Vorst, M., Baker, M., Perkersen, R., Brown, P., Ravenscroft, T., van Blitterswijk, M., Nicholson, A.M., *et al.* (2015). Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. *Acta neuropathologica* *130*, 77-92.

Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R., Lyon, E., *et al.* (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* *15*, 733-747.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *et al.* (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* *17*, 405-424.

- Schmucker, S., Martelli, A., Colin, F., Page, A., Wattenhofer-Donze, M., Reutenauer, L., and Puccio, H. (2011). Mammalian frataxin: an essential function for cellular viability through an interaction with a preformed ISCU/NFS1/ISD11 iron-sulfur assembly complex. *PloS one* 6, e16199.
- Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* 7, 575-576.
- Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research* 40, W452-457.
- Singleton, A.B. (2011). Exome sequencing: a transformative technology. *The Lancet Neurology* 10, 942-946.
- Sommen, M., Schrauwen, I., Vandeweyer, G., Boeckx, N., Corneveaux, J.J., van den Ende, J., Boudewyns, A., De Leenheer, E., Janssens, S., Claes, K., *et al.* (2016). DNA Diagnostics of Hereditary Hearing Loss: A Targeted Resequencing Approach Combined With a Mutation Classification System. *Human mutation*.
- Staropoli, J.F., Karaa, A., Lim, E.T., Kirby, A., Elbalalesy, N., Romansky, S.G., Leydiker, K.B., Coppel, S.H., Barone, R., Xin, W., *et al.* (2012). A homozygous mutation in KCTD7 links neuronal ceroid lipofuscinosis to the ubiquitin-proteasome system. *American journal of human genetics* 91, 202-208.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., *et al.* (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* 12, e1001779.
- Tarlarini, C., Lunetta, C., Mosca, L., Avemaria, F., Riva, N., Mantero, V., Maestri, E., Quattrini, A., Corbo, M., Melazzini, M.G., *et al.* (2015). Novel FUS mutations identified through molecular screening in a large cohort of familial and sporadic amyotrophic lateral sclerosis. *European journal of neurology* 22, 1474-1481.
- Tryka, K.A., Hao, L., Sturcke, A., Jin, Y., Wang, Z.Y., Ziyabari, L., Lee, M., Popova, N., Sharopova, N., Kimura, M., *et al.* (2014). NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic acids research* 42, D975-979.
- van Blitterswijk, M., van Es, M.A., Hennekam, E.A., Dooijes, D., van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., *et al.* (2012a). Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Human molecular genetics* 21, 3776-3784.
- van Blitterswijk, M., van Es, M.A., Koppers, M., van Rheenen, W., Medic, J., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., Veldink, J.H., and van den Berg, L.H. (2012b).

VAPB and C9orf72 mutations in 1 familial amyotrophic lateral sclerosis patient.
Neurobiology of aging 33, 2950 e2951-2954.

Appendices

Appendix A - Ethics approval



Research Ethics

Western University Health Science Research Ethics Board HSREB Annual Continuing Ethics Approval Notice

Date: September 23, 2015
Principal Investigator: Dr. Robert Hegele
Department & Institution: Science/Vascular Biology, Robarts Research Institute

Review Type: Expedited
HSREB File Number: 379
Study Title: Candidate gene sequencing, genetic and genomic analysis for identification of new genetic determinants of intermediate traits of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population. 07920E
Sponsor: Genome Canada
 Heart and Stroke Foundation of Canada
 Canadian Institutes of Health Research

HSREB Renewal Due Date & HSREB Expiry Date:
 Renewal Due -2016/09/30
 Renewal Due -2016/10/31
 Expiry Date -2016/11/03

The Western University Health Science Research Ethics Board (HSREB) has reviewed the Continuing Ethics Review (CER) Form and is re-issuing approval for the above noted study.

The Western University HSREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice (ICH E6 R1), the Ontario Freedom of Information and Protection of Privacy Act (FIPPA, 1990), the Ontario Personal Health Information Protection Act (PHIPA, 2004), Part 4 of the Natural Health Product Regulations, Health Canada Medical Device Regulations and Part C, Division 5, of the Food and Drug Regulations of Health Canada.

Members of the HSREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Ethics Officer to Contact for Further Information

Erika Basile ebasile@uwo.ca	Grace Kelly grace.kelly@uwo.ca	Mina Mekhail mmekhail@uwo.ca	Vikki Tran vikki.tran@uwo.ca
--------------------------------	-----------------------------------	---------------------------------	---------------------------------

This is an official document. Please retain the original in your files.

LAWSON HEALTH RESEARCH INSTITUTE**FINAL APPROVAL NOTICE**

RESEARCH OFFICE REVIEW NO.: R-11-522

PROJECT TITLE: Finding of Rare Disease Genes in Canada (FORGE Canada)

PRINCIPAL INVESTIGATOR: Dr. Victoria Mok Siu

DATE OF REVIEW BY CRIC: January 6, 2012

Health Sciences REB#: 18152

Please be advised that the above project was reviewed by the Clinical Research Impact Committee and the project:

Was Approved

PLEASE INFORM THE APPROPRIATE NURSING UNITS, LABORATORIES, ETC. BEFORE STARTING THIS PROTOCOL. THE RESEARCH OFFICE NUMBER MUST BE USED WHEN COMMUNICATING WITH THESE AREAS.

Dr. David Hill

V.P. Research

Lawson Health Research Institute

All future correspondence concerning this study should include the Research Office Review Number and should be directed to Sherry Paiva, CRIC Liaison, LHSC, Rm. C210, Nurses Residence, South Street Hospital.

cc: Administration

Office of Research Ethics



Reset Filters | Export To Excel

	File No	Project Title	Principal Investigator	Application Type	Status Snapshot	Workflow Message
	<input type="text"/>	<input type="text"/>	<input type="text"/>	All	<input type="text"/>	
View Clone Events	105522	Candidate gene sequencing, phenomics, genetic and genomic analysis for identification of new genetic determinants of amyotrophic lateral sclerosis, frontotemporal dementia, and related phenotypes	Dr. Michael Strong (Schulich School of Medicine and Dentistry/Pathology)	Delegated Health Sciences Form (Certification/Human Ethics)	Project Status: Active Workflow Status: Approval Decision Made	
View Clone Events	104915	The Ontario Neurodegenerative Disease Research Initiative	Dr. Michael Strong (Schulich School of Medicine and Dentistry/Pathology)	Full Board Health Sciences Form (Certification/Human Ethics)	Project Status: Active Workflow Status: Approval Decision Made	

Appendix B - Chapter 6: Primer list

Primer ID	Sequence (5'-->3')	T_m (°C)
<i>LRRK2_FWD</i>	GTGTGCTCTTGTATATGCTTTCC	58
<i>LRRK2_RVS</i>	CACTTTCAGCCACTTCAGGAG	59
<i>MAPT_FWD</i>	ATCTCAGCAATGTCTCCTCCACC	62
<i>MAPT_RVS</i>	TAACCGAACTGCGAGGAGCA	61
<i>CD2AP_FWD</i>	AGCTAGATTCTGAACAGCTGCC	60
<i>CD2AP_RVS</i>	TAAATGCACCACCGCTGTCCA	63
<i>NEFH_FWD</i>	GAAGAAACAAAGTCTCCCCC	56
<i>NEFH_RVS</i>	TCCTTTGCTGGAGACTTGGCCT	64
<i>DNAJC13_FWD</i>	GCATTGGCCTTGAAAACCTGG	61
<i>DNAJC13_RVS</i>	AAGTGGAACCTGAACACAGG	56
<i>APP_FWD</i>	GTCAAGTTTACCTACCTCCACC	58
<i>APP_RVS</i>	CTCATCCAAATGTCCCCTGCA	60
<i>HTRA2_FWD</i>	TGCTGACGTCAGGAACTTCTG	60
<i>HTRA2_RVS</i>	AGGATCTCGATATAGACCACGG	58
<i>MC1R_FWD</i>	ATCCCAGAGAAGACTTCTGGG	59
<i>MC1R_RVS</i>	AGCAGATGAAGCAGTACATGG	58
<i>EIF4G1_FWD</i>	GACATGAAAGTAGGCAGTGGGA	60
<i>EIF4G1_RVS</i>	AGTAGACAGATGCCAGCACA	61
<i>TREM2_FWD</i>	AATCTGGGGCTACAAAGGCC	60
<i>TREM2_RVS</i>	TGGCTAAGGGAGTTCCTGCTG	62

Appendix C - Chapter 7: Expanded genetic and clinical information on all individuals within the study

Patient ID	Family relationship	Disease status	Age at onset (years)	Current age (years)	Age at death (years)	Sex	MoCA (30)	<i>C9orf72</i> genotype	Other variants(s)
Family 1, amyotrophic lateral sclerosis									
III-1	Unrelated spouse	Unaffected		78		M	27	2,5	
III-2	Related, patient	ALS	61		64	F		2, exp.	
III-3	Related, patient	ALS	47		52	M			
III-4	Unrelated spouse	AD		80*		F			
III-5	Related, patient	ALS	49		58	M		2, exp.	
III-6	Unrelated spouse	Unaffected		70		F	28	2,5	
III-7	Related, brother of 3 affected individuals	Unaffected		70		M	28	2,18	
III-8	Unrelated spouse	Unaffected				F			
III-9	Related, brother of 3 affected individuals	Unaffected		63		M	29	2,18	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
III-10	Unrelated spouse	Unaffected		61		F	28	2,5	
IV-1	Unrelated spouse	Unaffected				M			
IV-2	Related, daughter of an affected individual	Unaffected or presymptomatic		50		F	26	2, exp.	
IV-3	Related, son of an affected individual	Unaffected or presymptomatic		48		M	28	2, exp.	
IV-4	Unrelated spouse	Unaffected		46		F	29	2,5	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
IV-5	Related, daughter of an affected individual	Unaffected		43		F	28	5,5	
IV-6	Related, daughter of an affected individual	Unaffected		46		F	28	5,5	
IV-7	Unrelated spouse	Unaffected		45		M	30	2,2	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
IV-8	Related, daughter of an affected individual	Unaffected				F			
IV-9	Related, son of an affected individual	Unaffected			20*	M			
IV-10	Related, son of an affected individual	Unaffected		49		M	30	5,5	
IV-11	Unrelated spouse	Unaffected		48		F	30	5,5	

IV-12	Related, son of an affected individual	Unaffected	48	M	30	2,5	
IV-13	Related, son of an affected individual	Unaffected	46	M	27	5,5	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
IV-14	Unrelated spouse	Unaffected		F			
IV-15	Unrelated spouse	MS		F			
IV-16	Related, son of an affected individual	Unaffected or presymptomatic	44	M	30	5, exp.	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
IV-17	Related, son of an affected individual	Unaffected or presymptomatic with ALS. Affected with CP.	44	M		5, exp.	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
IV-18	Related, son of an unaffected individual	Unaffected		M			
IV-19	Related, daughter of an unaffected individual	Unaffected		F			
IV-20	Related, daughter of an unaffected individual	Unaffected		F			
IV-21	Related, daughter of an unaffected individual	Unaffected		F			
IV-22	Related, son of an unaffected individual	Unaffected	36	M	28	5,18	
IV-23	Related, son of an unaffected individual	Unaffected		M			
IV-24	Related, daughter of an unaffected individual	Unaffected	29	F	28	2,2	
IV-25	Related, son of an unaffected individual	Unaffected		M			
V-1	Related, granddaughter of an affected individual	Unaffected		F			
V-2	Related, granddaughter of an affected individual	Unaffected or presymptomatic	21	F	29	5, exp.	<i>ARHGEF28</i> p.Lys280Met>fs40Ter
V-3	Related, grandson of an affected individual	Unaffected or presymptomatic		M		5, exp.	
V-4	Related, granddaughter of an affected individual	Unaffected	19	F	28	2,5	<i>ARHGEF28</i> p.Lys280Met>fs40Ter

V-5	Related, granddaughter of an affected individual	Unaffected		<18	F		
V-6	Related, granddaughter of an affected individual	Unaffected		19	F	30	5,5
V-7	Related, granddaughter of an affected individual	Unaffected		<18	F		
V-8	Related, granddaughter of an affected individual	Unaffected		<18	F		
V-9	Related, granddaughter of an affected individual	Unaffected		<18	F		
V-10	Related, granddaughter of an affected individual	Unaffected		<18	F		
V-11	Related, granddaughter of an affected individual	Unaffected		<18	F		
Family 2, mixed neurodegeneration							
I-1	Unrelated	AD			Deceased	M	
I-2	Unrelated	AD			Deceased	F	
I-3	Related	Unaffected			Deceased	M	
I-4	Related	Unaffected			Deceased	F	
II-1	Unrelated spouse, son of parents affected with AD	Unaffected		80		M	
II-2	Related, sister of 3 affected individuals	Unaffected		77		F	5,5
II-3	Related, patient	ALS and FTD	53		55	M	2, exp <i>ATXN2</i> exp (22,30), <i>OPTN</i> p.Met468Arg
II-4	Unrelated spouse	Unaffected				F	
II-5	Related, patient	Myelopathy and CI		70		M	10 2,2 <i>ATXN2</i> exp (22,30)
II-6	Unrelated spouse	Unaffected by ALS. Affected by cancer.				F	
II-7	Related, patient	FTD and ALS	56		59	F	exp [#]
II-8	Unrelated spouse	Unaffected				M	
III-1	Unrelated spouse	Unaffected				F	
III-2	Related, son of an unaffected	PD	49	52		M	

	individual						
III-3	Related, daughter of an unaffected individual	Unaffected	55	F			
III-4	Related, daughter of an unaffected individual	Unaffected	50	F			
III-5	Unrelated spouse	Unaffected		F			
III-6	Related, son of an affected individual	Unaffected or presymptomatic	41	M	29	2, exp	ATXN2 exp (22,30)
III-7	Related, son of an unaffected individual	Unaffected		M			
III-8	Related, son of an unaffected individual	Unaffected		M			
III-9	Biologically unrelated, adopted daughter of an affected individual	Unaffected		F			
III-10	Related, son of an affected individual	Unaffected		M			

Family 3, amyotrophic lateral sclerosis and behavioural variant frontotemporal dementia

I-1	Related	ALS	64	F			
I-2	Unrelated spouse	Unaffected	71	M			
II-1	Unrelated spouse	Unaffected	Deceased	F			
II-2	Related, son of an affected individual	ALS	59	M			
II-3	Related, daughter of an affected individual	ALS	35	F			
II-4	Related, son of an affected individual	AD	77	M			
II-5	Unrelated spouse	Unaffected					
II-6	Related, daughter of an affected individual	CVA	86	F			
II-7	Unrelated spouse	Unaffected	Deceased	M			
II-8	Related, son of an affected individual	ALS	60	M			
II-9	Unrelated spouse	Unaffected		F			
II-10	Related, son of an affected individual	Bone cancer	60	M			
II-11	Unrelated spouse	Unaffected		F			
II-12	Related, daughter of an affected individual	Pancreatic cancer	78	F			
II-13	Related, son of an affected individual	ALS	60	M			
II-14	Related, son of an affected individual	Unaffected	Deceased	M			

II-15	Related, daughter of an affected individual	Unaffected			F			
II-16	Related, daughter of an affected individual	Unaffected		Deceased	F			
II-17	Related, son of an affected individual	Senility/AD		74	M			
II-18	Unrelated spouse	Unaffected			F			
III-1	Related, daughter of an affected individual	Unaffected		Deceased	F			
III-2	Unrelated spouse	Unaffected		Deceased	M			
III-3	Related, patient	bvFTD, later developed ALS	56	59	F	exp		<i>OPTN</i> p.Met468Arg
III-4	Unrelated spouse	Unaffected		73	M			
III-5	Related, patient	bvFTD, later developed ALS and parkinsonism	53	54	M	8,exp		<i>OPTN</i> p.Met468Arg
III-6	Unrelated spouse	Unaffected		Deceased				
III-7	Related, patient	bvFTD, later developed ALS	54	56	M	5,exp		<i>OPTN</i> p.Met468Arg
III-8	Unrelated spouse	Unaffected						
IV-4	Related, daughter of an affected individual	Unaffected		44	F	30	2,5	<i>OPTN</i> p.Met468Arg
IV-5	Related, son of an affected individual	Unaffected		40	M	28	2,exp	
IV-6	Related, daughter of an affected individual	Unaffected		39	F	24	2,5	

*Information based on family recall, not confirmed through birth or death records. Some individuals were not genetically tested as they were not age of consent, were deceased, or declined to participate. Their clinical and family history was provided to us and was included in experimental planning. #Unfortunately, we did not have access to clinical records or tissue samples to conduct genetic analysis but through personal communication with her neurologist, we know she is a carrier of a *C9orf72* expansion.

Appendix D - Journal copyright approval

Note: All other publications included in this PhD thesis are open access publications that are reproducible with attribution to the original source.



The screenshot shows the Copyright Clearance Center RightsLink interface. At the top left is the Copyright Clearance Center logo. To its right is the RightsLink logo. On the top right, there are navigation buttons for Home, Account Info, Help, and a Live Chat icon. Below the navigation is a user login box showing the user is logged in as Sali Farhan with account number 3001038510 and a LOGOUT button. The main content area displays a journal cover for 'CURRENT Cardiology reports' and a list of metadata:

- Title:** Exome Sequencing: New Insights into Lipoprotein Disorders
- Author:** Sali M. K. Farhan
- Publication:** Current Cardiology Reports
- Publisher:** Springer
- Date:** Jan 1, 2014

At the bottom of the metadata, it says 'Copyright © 2014, Springer Science+Business Media New York'.

Order Completed

Thank you for your order.

This Agreement between Sali Farhan ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

[Get the printable license.](#)

License Number	3899060260640
License date	Jun 30, 2016
Licensed Content Publisher	Springer
Licensed Content Publication	Current Cardiology Reports
Licensed Content Title	Exome Sequencing: New Insights into Lipoprotein Disorders
Licensed Content Author	Sali M. K. Farhan
Licensed Content Date	Jan 1, 2014
Licensed Content Volume	16
Licensed Content Issue	7
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	1
Title of your thesis / dissertation	Gene discovery in Mendelian and complex diseases
Expected completion date	Jun 2016
Estimated size(pages)	300

Title: Genetics 101 for Cardiologists: Rare Genetic Variants and Monogenic Cardiovascular Disease
Author: Sali M.K. Farhan, Robert A. Hegele
Publication: Canadian Journal of Cardiology
Publisher: Elsevier
Date: Jan 1, 2013

Copyright © 2013, Elsevier

Logged in as:
 Sali Farhan
 Account #:
 3001038510

LOGOUT

Order Completed

Thank you for your order.

This Agreement between Sali Farhan ("You") and Elsevier ("Elsevier") consists of your order details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License number	Reference confirmation email for license number
License date	Jun 19, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Canadian Journal of Cardiology
Licensed Content Title	Genetics 101 for Cardiologists: Rare Genetic Variants and Monogenic Cardiovascular Disease
Licensed Content Author	Sali M.K. Farhan, Robert A. Hegele
Licensed Content Date	January 2013
Licensed Content Volume	29
Licensed Content Issue	1
Licensed Content Pages	5
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes

Curriculum Vitae

Sali M. K. Farhan

EDUCATION

- 2012-2016 **Doctor of Philosophy (with distinction) candidate**
 Supervisor: Dr. Robert A. Hegele
 Co-supervisor: Dr. Michael J. Strong
 Department of Biochemistry, Schulich School of Medicine and Dentistry
 Robarts Research Institute, Western University, London, Ontario, Canada
- 2008-2012 **Bachelor of Science, Honours Specialization in Genetics**
 Department of Biology, Western University, London, Ontario, Canada

AWARDS

Salary awards

- 2016 ALS Canada Tim E. Noël Postdoctoral Fellowship (\$55,000/year); funded for 3 years
- 2016 Canadian Institutes of Health Research CGS Michael Smith Foreign Study Supplement award (\$6,000)
- 2016 Doctoral Excellence Research Award (\$10,000)
- 2015-2016 Ontario Graduate Scholarship (\$15,000/year) (declined)
- 2015-present Canadian Institutes of Health Research, Frederick Banting and Charles Best CGS Doctoral Award (\$35,000/year); funded for 3 years; ranked in the top 1.75%
- 2014-2015 Ontario Graduate Scholarship (\$15,000/year)
- 2012-2014 Canadian Institutes of Health Research, STP Research Fellowship (\$12,000/year)

Other awards

- 2016 Western University PSAC Local 610 Academic Achievement Scholarship (\$500)
- 2015 5th International Research Workshop on Frontotemporal Dementia in Amyotrophic Lateral Sclerosis, first place presentation award (\$750)

- 2015 Western University PSAC Local 610 Academic Achievement Scholarship (\$500)
- 2014 Schulich School of Medicine and Dentistry Scholarship for Medical Research (\$2,000)
- 2014 Schulich School of Medicine and Dentistry: Norman E. Nixon Marie Rämö Nixon Award (\$1,500) (Awarded on June 2014)
- 2013 Schulich School of Medicine and Dentistry: Norman E. Nixon Marie Rämö Nixon Award (\$1,500) (Awarded on May 2014)
- 2014 Department of Pediatrics Research Day, Best Poster Presentation in the Basic Sciences (\$200)
- 2013 Canadian Institutes of Health Research, Institute of Human Development, Child and Youth Health Travel Award (\$1,000)
- 2013 Department of Medicine Research Day, Best Poster Presentation (\$500)
- 2013 2nd Annual Canadian Human and Statistical Genetics Meeting Best Oral and Poster Presentation Award (\$500)
- 2013 Western University PSAC Local 610 Community Scholarship (\$400)
- 2013 London Health Sciences Centre Research Day, Clinical Investigation and Metabolic Disease Best Poster Award (\$500)
- 2012-Present Western Graduate Research Scholarship (\$4,500/year)
- 2011 Lynn Fordham Award in Science and Engineering (\$2,500)
- 2011 Biology Graduate Research Forum Excellence in Scientific Communication, best poster award
- 2009-2012 Dean's Honour Roll
- 2008 Western University Scholarship of Distinction (\$1,500)
- 2008 Canadian Millennium Local Excellence Scholarship (\$4,000)
- 2008 Rotary Annual Citizenship Award (\$500)

RESEARCH TRAINING

2016 **Analytic and Translational Genetics Unit, Broad Institute of MIT and Harvard and Massachusetts General Hospital: Dr. Mark Daly and Dr. Daniel MacArthur**

Visiting PhD student for 6 months.

Project: Identifying putative disease-causing variants in assumed healthy controls within the Exome Aggregation Consortium database.

2011-Present **Blackburn Cardiovascular Genetics Laboratory, Western University: Dr. Robert A. Hegele and Dr. Michael J. Strong**

PhD candidate and former fourth year honours thesis student

Project: Gene discovery in Mendelian and complex diseases.

2010-2011 **Centre for Human Immunology, Western University: Dr. Joaquín Madrenas**

Research student

Project: Identifying gene expression profiles in monocyte-derived macrophages and dendritic cells in response to *S. aureus* peptidoglycan.

2009-2010 **Molecular Genetics Unit, Western University: Dr. Shiva M. Singh**

Research student

Project: Studying the effect of alcohol exposure on fetal development and cognition in a mouse model of Fetal Alcohol Spectrum Disorders.

Relevant technical skills

- Human ethics protocol preparation
- Genetic mapping, next generation sequencing analysis
- Bioinformatic and statistical tools
- Designing gene panels for human genetic diseases
- Next generation sequencing library preparation
- DNA, RNA, and protein isolation
- Sanger sequencing
- PCR, reverse transcriptase (RT) PCR, and quantitative PCR
- SNP genotyping (TaqMan[®] genotyping assay)
- Restriction enzyme digestion genotyping
- Isolation of PBMCs from blood and cell differentiation
- Molecular cloning
- Western blotting
- Southern blotting
- Experience with radioactivity
- Co-immunoprecipitation

- *in vitro* cell culture—primary culture and established cell lines
- Site-directed DNA mutagenesis and transfection methods
- Animal handling (murine developmental milestone tests, behavioural analysis, euthanasia, and dissections)

CERTIFICATES

- 2012 Teaching Assistant Training Program (TATP), Western University
 2012 Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, Western University

PROFESSIONAL ACTIVITIES

Research grants

- Assisted in co-writing microgrants for functional genomics studies.

- 2014 Rare disease foundation, microgrant program. “Using cellular and fruit fly models to study the effect of *FSD1* missense mutations in an inherited disorder that resembles Angelman syndrome”. (\$3,500).
 Principal applicant: Dr. Robert Hegele; co-applicants: Dr. Sharan Goobie, Dr. Heleen Arts, Dr. Jamie Kramer, and Sali Farhan.
- 2014 Rare disease foundation, microgrant program. “Determining the genetic cause of a unique congenital hydranencephaly syndrome”. (\$5,000).
 Principal applicant: Dr. Victoria Siu; co-applicants: Dr. Robert Hegele, Dr. Heleen Arts, Dr. C. Anthony Rugar, and Sali Farhan.
- 2014 CIHR, research catalyst network, rare diseases: models & mechanisms. “Using cellular and fruit fly methodologies to characterize the role of *TMTC3* in nocturnal seizures with developmental delay”. (\$25,000).
 Principal applicant: Dr. Robert Hegele; co-applicants: Dr. Jamie Kramer, Dr. Heleen Arts, Dr. Victoria Siu, and Sali Farhan.
- 2014 CIHR, research catalyst network, rare diseases: models & mechanisms. “Using cellular and fruit fly methodologies to characterize the role of *FSD1* in a disorder that resembles Angelman syndrome”. (\$25,000).
 Principal applicant: Dr. Robert Hegele; co-applicants: Dr. Jamie Kramer, Dr. Heleen Arts, Dr. Sharan Goobie, and Sali Farhan.

Teaching experience

- 2012-Present **Western University: Departments of Biochemistry, Biology**
 Teaching assistant and exam proctor, nominated for teaching award in 2013.

Course: Biochem 2280A, Biochemistry and Molecular Biology (2013-present)

Course: Biology 3592A, Principles of Human Genetics (2012-2013)

Course: Biochem 4463G, Biochemistry of Genetic Diseases (2012-present)

Mentorship of undergraduate students

- Trained students in wet lab experiments, data analysis, presentation skills, and edited their academic materials (proposals, progress reports, and theses).

2015-present Heba Almadhoun “Understanding the disease mechanism of Angelman-like syndrome”. Western University, Department of Biochemistry, fourth year honours thesis project. Co-mentorship with Dr. Heleen Arts.

2015-present Allison Dilliot “Identifying the cause of genetic diseases as part of the Care for Rare Consortium”. Western University, Department of Biochemistry, summer research student. “Identifying the genetic basis of neurodegeneration in 75 patients using a targeted resequencing panel”. Fourth year honours thesis project.

2014-2015 Grace Wang “Using LipidSeq, a targeted next-generation resequencing panel, in the genetic diagnosis of MODY”. Western University Scholar’s Elective program, third year project.

2014-2015 Rosettia Ho “Gene hunting: identifying the molecular basis of lipodystrophy through gene panels, whole-exome sequencing, and bioinformatics”. Western University, Department of Biology, fourth year honours thesis project.

2014-2015 Paula Romero “Identifying new genetic variants associated with familial hypercholesterolemia”. Western University, Department of Biochemistry, fourth year honours thesis project.

2014-2015 Jacqueline Dron “Elucidating the genetic determinants of extreme high-density lipoprotein phenotypes using next-generation sequencing”. Western University, Department of Biology, fourth year honours thesis project.

2013-2014 Grace Wang “Using genomic technologies to identify variants in a novel genetic disorder: ichthyosis-microcephaly in the Old Order Amish”. Western University Scholar’s Elective program, second year project.

2013-2014 Amanda Tong “Identifying pathogenic mutations in patients with atypical progeria”. Western University, Department of Biology, fourth year honours thesis project.

- 2013-2014 Jennifer Fu “Identifying the cause of a novel form of ataxia, dystonia and mental retardation”. Western University Scholar’s Elective program, third year project.
- 2013-2014 Erika Scott “The application of whole-exome sequencing to elucidate the genetic basis of a rare form of osteopetrosis”. Western University, Department of Biochemistry, fourth year honours thesis project.
- 2012-2013 Lisa Murphy “The identification of a novel variant in the potassium channel tetramerization domain-containing 7 gene, *KCTD7*, in a Mennonite family with an idiopathic seizure disorder”. Western University, Department of Biochemistry fourth, year honours thesis project.
- 2012 Melissa Loyzer “Sequencing *COL9A1* in patients with cleft palate syndrome”. Western University, summer research student.

Professional affiliations

- 2013-Present American Society of Human Genetics
 2014-Present Canadian Genetics and Genomics Society
 2014-Present Canadian Society of Molecular Biosciences

Conferences organized

- 2016 Chair, Gordon Research Seminars, Human Genetics and Genomics, July 8-9, 2017, Stowe, Vermont, USA.
- 2015 Discussion leader, “Moving Forward: Elucidating Functional Variation/ Mentorship Component: Publishing in a High Tier Journal”. Gordon Research Seminars, Human Genetics and Genomics, July 18-19 in Newport, Rhode Island, USA.
- 2015 Volunteer, Gordon Research Conference, Human Genetics and Genomics, lead organizer: Dr. Nancy J. Cox, Vanderbilt University, held on July 19-24 in Newport, Rhode Island, USA.
- 2015 Volunteer, 5th International Research Workshop on Frontotemporal Dementia in Amyotrophic Lateral Sclerosis, lead organizer: Dr. Michael J. Strong, Western University, held on June 7-10 in London, Ontario, Canada.

PEER REVIEWED PUBLICATIONS

1. **Farhan, SMK.**, Nixon, KCJ, Edwards, T., Everest, M., Long, S., Segal, D., Knip, MJ., Arts, HE., Wang, J., Robinson, JF., Rupa, CA., Siu, VM., Poulter, MO., FORGE Canada Consortium, Hegele, RA., Kramer, JM. TMT3 is a synaptic protein involved in seizure susceptibility and intellectual disability. Submitted.
2. **Farhan, SMK.**, Gendron, TF., Petrucelli, L., Strong, MJ., Hegele, RA. Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia. Submitted.
3. Badalato, L., **Farhan, SMK.**, Dilliot, AA., Care4Rare Canada Consortium, Bulman, DE., Hegele, RA., and Goobie, S. KMT2D p.Gln3575His segregating in a family with autosomal dominant choanal atresia strengthens the Kabuki/CHARGE connection. *Am J Med Genet A*. In revision. 2016.
4. **Farhan, SMK.**, ONDRI Investigators, Hegele, RA. The ONDRISeq panel: custom designed next generation sequencing of genes related to neurodegeneration. *npj Genomic Medicine*. Accepted. 2016.
5. **Farhan, SMK.**, ONDRI Investigators, Strong, MJ. The Ontario Neurodegenerative Disease Research Initiative (ONDRI). *Can J Neurol Sci*. Accepted. 2016.
6. **Farhan, SMK.**, and Hegele, RA. Antisenses working overtime in lipids. *Nat Rev Endocrinol*. 2015; 11(10):574-576.
7. **Farhan, SMK.**, Wang, J., Robinson, JF., Prasad, AN., Rupa, CA., Siu, VM., FORGE Canada Consortium, and Hegele, RA. Old gene, new phenotype: mutations in heparan sulfate synthesis enzyme, EXT2 leads to seizure and developmental disorder, no exostoses. *J Med Genet*. 2015; 52(10):666-675.
8. **Farhan, SMK.** Something old, something new: TBK1, a novel gene in known amyotrophic lateral sclerosis pathways. *Clin Genet*. 2015; 88(4):339-340.
9. **Farhan, SMK.**, Robinson, JF., McIntyre, AD., Marrosu, MG., Ticca, AF., Carboni, N., and Hegele, RA. A novel homozygous LIPE mutation in patients with familial partial lipodystrophy and muscular dystrophy. *Can J Cardiol*. 2014; Dec;30(12):1649-1654.
10. **Farhan, SMK.**, Murphy, LM., Robinson, JF., Wang, J., P., Siu, VM., Rupa, CA., Prasad, AN., and Hegele, RA. Linkage analysis and exome sequencing identify a novel mutation in KCTD7 in patients with progressive myoclonus epilepsy with ataxia. *Epilepsia*. 2014; 55(9):e106-111.
11. **Farhan, SMK.**, and Hegele, RA. Exome sequencing: new insights into lipoprotein disorders. *Curr Cardiol Rep*. 2014; 16(7):507-517.

12. **Farhan, SMK.**, and Hegele, RA. Sequencing: The Next Generation—what is the role of whole-exome sequencing in diagnosis of familial cardiovascular diseases. *Can J Cardiol.* 2013; 30(2):152-154.
13. **Farhan, SMK.**, Wang, J., Robinson, JF., Lahiry, P., Siu, VM., Prasad, C., Kronick, JB., Ramsay, DA., Rupa, CA., and Hegele, RA. Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster related disease, infantile mitochondrial complex II/III deficiency. *Mol Genet and Genomic Med.* 2013; 2(1):73-80.
14. **Farhan, SMK.**, and Hegele, RA. Genetics 101 for cardiologists: Rare genetic variants and monogenic cardiovascular disease. *Can J Cardiol.* 2012; 29(1):18-22.

Book chapters

1. **Farhan, SMK.**, and Prasad, AN. Exploring the Epilepsioime I: Genetics of Age dependent Epileptic Encephalopathies. *Pediatric Epilepsy, 4th edition, Chapter 7.*
2. Prasad, AN., and **Farhan, SMK.** Exploring the Epilepsioime II: Approaching the Complex Epilepsies. *Pediatric Epilepsy, 4th edition, Chapter 8.*

REVIEWED ARTICLES IN JOURNALS

Lancet, Diabetes and Endocrinology (1); *New England Journal of Medicine* (2); *Nature Communications* (1); *Nature Genetics* (1); *Alzheimer's and Dementia* (1); *Journal of Medical Genetics* (1); *Expert Review of Molecular Diagnostics* (1); *Canadian Journal of Cardiology* (1).

PATENTS

1. 2016 'Neurodegeneration targeted resequencing panel'. Inventors: Michael Strong, John Robinson, **Sali Farhan**, Robert Hegele, and Tom Mikkelsen,. United States Provisional Patent Application ID: 62/333,369.

CONFERENCES AND RESEARCH FORUMS ATTENDED

Oral presentations:

1. **Farhan SMK**, Liang E, McIntyre AD, Cao H, Robinson JF, Volkening K, Bulman DE, Rogaeva E, St. George-Hyslop P, Strong MJ, and Hegele RA, for the ONDRI Investigators. Tackling neurodegeneration using next-generation sequencing: identifying the genetics of five neurodegenerative disorders. Gordon Research Seminars, Human Genetics and Genomics. Newport, RI, USA. July 2015.

2. **Farhan SMK** and Prasad AN. Epilepsy-bedside to the bench in the era of genomic medicine. Clinical Neurological Sciences Grand Rounds, LHSC, Western University. London, Ontario. September 2014.
3. **Farhan SMK** and Siu VM. FORGE update: progress of rare disease studies in South-Western Ontario. Pediatric Grand Rounds, Children's Hospital of Western Ontario. London, Ontario. August 2013.
4. **Farhan SMK**, Wang J, Robinson JF, Lahiry P, Siu VM, Prasad, C, Kronick, JB, Ramsay, DA, Rupa, CA, and Hegele, RA. Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster related disease, infantile mitochondrial complex II/III deficiency. 2nd Annual Canadian Human and Statistical Genetics Meeting. Institute of Genetics. Estérel, Quebec. April 2013.
5. **Farhan SMK**, Wang J, Robinson JF, Siu VM, Prasad C, Rupa CA, and Hegele RA. Mapping the locus responsible for a rare novel autosomal recessive condition: Infantile Fatal Mitochondrial Disease. Ontario Biology Day. Laurentian University, Sudbury, ON., March 2012.
6. **Farhan SMK**, Sayedyahosseini S, and Madrenas J. Gene Expression Profiles in Monocyte-derived Macrophages and Dendritic Cells in Response to *Staphylococcus aureus* Peptidoglycan. Ontario-Quebec Undergraduate Immunology Conference, University of Toronto, Toronto, ON., April 2011.

Poster presentations:

1. **Farhan, SMK.**, Gendron, TF., Petrucelli, L., Strong, MJ., Hegele, RA. Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia. In preparation. American Society of Human Genetics, Baltimore, MD, USA. October 2015.
2. **Farhan SMK**, Liang E, McIntyre AD, Cao H, Robinson JF, Volkening K, Bulman DE, Rogaeva E, St. George-Hyslop P, Strong MJ, and Hegele RA, for the ONDRI Investigators. Tackling neurodegeneration using next-generation sequencing: identifying the genetics of five neurodegenerative disorders. Gordon Research Conference, Human Genetics and Genomics. Newport, RI, USA. July 2015.
3. **Farhan SMK**, Liang E, McIntyre AD, Cao H, Robinson JF, Volkening K, Bulman DE, Rogaeva E, St. George-Hyslop P, Strong MJ, and Hegele RA, for the ONDRI Investigators. Tackling neurodegeneration using next-generation sequencing: identifying the genetics of five neurodegenerative disorders. 5th International Research Workshop on Frontotemporal Dementia in Amyotrophic Lateral Sclerosis. London, ON., June 2015.
4. **Farhan SMK**, Liang E, McIntyre AD, Cao H, Robinson JF, Volkening K, Bulman DE, Rogaeva E, St. George-Hyslop P, Strong MJ, and Hegele RA, for the ONDRI

- Investigators. Tackling neurodegeneration using next-generation sequencing: identifying the genetics of five neurodegenerative disorders. 4th Annual Canadian Human and Statistical Genetics Meeting. Institute of Genetics. Vancouver, BC., April 2015.
5. **Farhan SMK**, Liang E, McIntyre AD, Cao H, Robinson JF, Volkening K, Bulman DE, Rogaeva E, St. George-Hyslop P, Strong MJ, and Hegele RA, for the ONDRI Investigators. Tackling neurodegeneration using next-generation sequencing: identifying the genetics of five neurodegenerative disorders. London Health Sciences Centre Research Day. London, ON., April 2015.
 6. **Farhan SMK**, Wang J, Robinson JF, Lahiry P, Siu VM, Rupar, CA, and Hegele, RA. The application of next-generation sequencing technology in identifying the genetic origins of unclassified familial disorders. 27th annual Paediatrics Research Day. Western University, London, ON., May 2014.
 7. **Farhan SMK**, Wang J, Robinson JF, Lahiry P, Siu VM, Rupar, CA, and Hegele, RA. The application of next-generation sequencing technology in identifying the genetic origins of unclassified familial disorders. 3rd Annual Canadian Human and Statistical Genetics Meeting. Institute of Genetics. Victoria, BC., May 2014.
 8. **Farhan SMK**, Wang J, Robinson JF, Lahiry P, Siu VM, Rupar, CA, and Hegele, RA. The application of next-generation sequencing technology in identifying the genetic origins of unclassified familial disorders. London Health Sciences Centre Research Day. London, ON., March 2014.
 9. **Farhan SMK**, Wang J, Robinson JF, Lahiry P, Siu VM, Prasad, C, Kronick, JB, Ramsay, DA, Rupar, CA, and Hegele, RA. Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster related disease, infantile mitochondrial complex II/III deficiency. American Society of Human Genetics meeting. Boston, MA., October 22 2013.
 10. **Farhan SMK**, Wang J, Robinson JF, Lahiry P, Siu VM, Prasad, C, Kronick, JB, Ramsay, DA, Rupar, CA, and Hegele, RA. Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster related disease, infantile mitochondrial complex II/III deficiency. Department of Medicine Research Day. Western University, London, ON., May 2013.
 11. **Farhan SMK**, Wang J, Robinson JF, Siu VM, Prasad C, Rupar CA, and Hegele RA. Mapping and Exome sequencing identify the loci associated with developing: Infantile Fatal Mitochondrial disease. London Health Sciences Research Day. Western University, London, ON., March 2013.
 12. **Farhan SMK**, Wang J, Robinson JF, Siu VM, Prasad C, Rupar CA, and Hegele RA. Mapping and Exome sequencing identify the loci associated with developing:

- Infantile Fatal Mitochondrial disease. Developmental Origins of Metabolic Syndrome. University of Michigan, Ann Arbor, MI., October 2012.
13. **Farhan SMK**, Wang J, Robinson JF, Siu VM, Prasad C, Rupar CA, and Hegele RA. Mapping and Exome sequencing identify the loci associated with developing: Infantile Fatal Mitochondrial disease. London Health Sciences Centre Research Day, Western University, London, ON., March 2012.
 14. **Farhan SMK**, Sayedyahosseini S, and Madrenas J. Gene Expression Profiles in Monocyte-derived Macrophages and Dendritic Cells in Response to *Staphylococcus aureus* Peptidoglycan. Biology Graduate Research Forum, Western University, London, ON., October 2011.
 15. **Farhan SMK**, Sayedyahosseini S, and Madrenas J. Gene Expression Profiles in Monocyte-derived Macrophages and Dendritic Cells in Response to *Staphylococcus aureus* Peptidoglycan. Infection and Immunity Research Forum, Western University, London, ON., November 2010.

ACADEMIC AND VOLUNTEER SERVICES

- | | |
|--------------|--|
| 2016 | Co-organizer, Medical Population Genetics Hackathon, Broad Institute of MIT and Harvard. |
| 2015 | Speaker on scholarship applications, CIHR division, Schulich School of Medicine and Dentistry. |
| 2015-present | Conducted experiments for elementary students, Canadian Medical Hall of Fame, London, Ontario. |
| 2015-present | Student representative, Menten Memorial Lecture Committee, Department of Biochemistry. |
| 2014-present | Student representative, Graduate Endowment Fund, Department of Biochemistry. |
| 2012-present | Scholarship reviewer, National Scholarship Program, Western University. |
| 2012-present | Student Representative, Annual Open House Event, Department of Biochemistry. |