Whole Exome Sequencing to Identify Disease-Causing Mutations in Lower Motor Neuron Disease and Peripheral Neuropathy

Justin Wagner

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the
MSc degree in Biochemistry
Collaborative Program in Human and Molecular Genetics

Department of Biochemistry, Microbiology and Immunology
Faculty of Medicine
University of Ottawa

Abstract

Lower motor neuron diseases and peripheral neuropathies are two groups of diseases that include multiple rare disorders where many causes are unknown and definitive treatments are unavailable. Understanding the molecular etiology of these genetic diseases provides an opportunity for rapid diagnosis, preconception genetic counseling and, in a subset, direction for the development of future treatment options. The recent introduction of whole exome sequencing (WES) marks a new era in Mendelian genetic disease research as the majority of the coding region of the genome can be sequenced in a timely and cost-effective manner. In this study, WES was used to investigate the molecular etiology of a cohort of 37 patients presenting with lower motor neuron disease or peripheral neuropathy. A molecular diagnosis was determined for seven patients informing the diagnostic utility of WES. Novel phenotypes were found for three genes originally associated with a different disorder. Finally, the foundation has been laid, through the use of functional studies and large scale data-sharing, to identify novel disease-causing genes for lower motor neuron disease and peripheral neuropathy.

Acknowledgments

First and foremost, I would like to thank my supervisors, Dr. Dennis Bulman and Dr. Kym Boycott, for accepting me as a student in their lab. They have provided me with the guidance, support and training, enabling me to complete my program. They are both great role models in how they conduct themselves in a professional manner at work, but also in how they make their family their number one priority. Both Dr. Bulman and Dr. Boycott care so much for the staff and students working under them and have always been willing to write that letter of recommendation, or provide advice on how to reach my career goals.

I owe a huge thanks, also, to every member of the Bulman/ Boycott lab, as they have made each day a memorable experience. They have all offered me so much help and advice in how to conduct my research, and we have become very close friends through our journey together.

Combining the leadership of both of my supervisors, along with my daily interactions with all the lab members, I could not have asked for a better place to conduct my Master's.

I would also like to acknowledge each clinician I worked with, in particular Dr. David Dyment, Dr. Sarah Sawyer and Dr. Jodi Warman. All of whom I worked very closely with, and were paramount to the completion of my project. My advisory committee, formed by Dr. Martin Holcik and Dr. Alex Mackenzie, also provided much advice into the direction my research should go.

I also could not have completed this project without the support of my family. My mom had incredible home cooking prepared for me every time I made the trip home, and my dad spent countless hours proofreading various assignments and reports of mine. My three brothers have always been my three closest friends and offered their help and support whenever they could. In particular, my youngest brother Tyson typed up Table AA1 (the

largest table in my thesis) under the promise that I would mention his name in the Acknowledgments. I would also like to extend my love and gratitude to my fiancée, Sami, for all the love and support she has given me throughout my Master's, as we look forward to our wedding day in less than a year.

Finally, I would like to acknowledge my Lord and Savior, Jesus Christ, who has blessed me with the ability and resources to complete this program, but more importantly, without his sacrifice, makes my life and this research pointless.

Table of Contents

Page

List of Abbreviations

µg	microgram
µl	microliter
µM	micromolar
AD	Autosomal Dominant
ADM	Abductor Digiti Minimi
AH	Abductor Hallicus
ALS	Amyotrophic Lateral Sclerosis
AMP	Ampicillin
APB	Abductor Pollicus Brevis
ATP	Adenosine Tri-phosphate
BAG3	BCL2-associated Athanogene 3
bp	Base Pair
BICD2	Bicaudal D2
C4R	Care4Rare
CA	California
CCDS	Consensus Coding Sequence
cDNA	Complementary Deoxyribonucleic Acid
CDS	Coding Sequence
CHEO	Children's Hospital of Eastern Ontario
CHN	Congenital Hypomyelinating Neuropathy
CK	Creatine Kinase
CMAP	Compound Motor Action Potential

CMT	Charcot-Marie-Tooth
CMT1	Charcot-Marie-Tooth Type 1
CMT2	Charcot-Marie-Tooth Type 2
CMT3	Charcot-Marie-Tooth Type 3
CMT4	Charcot-Marie-Tooth Type 4
CMTX	Charcot-Marie-Tooth X-linked
CNV	Copy Number Variant
CV	Conduction Velocity
dbSNP	Single Nucleotide Polymorphism Database
DM	Diabetes Mellitus
DML	Distal Onset Motor Latency
DNA	Deoxyribonucleic Acid
DNM2	Dynamin 2
DSD	Dejerine-Sottas Disease
EDB	Extensor Digitorum Brevis
EGR2	Early Growth Response 2
EMG	Electromyography
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium
FBXO38	Fork-box Only Protein 38
FH	Forkhead associated domain
FIG4	S. Cerevisiae Homolog of FIG4
FVC	Forced Vital Capacity
g	Force of Gravity

GARS	Glycyl-tRNA Synthetase
Gbp	Giga-base pair
GDAP1	Ganglioside-induced Differentiation Assoicated Protein 1
GEDI	Gaffa Evented Data Interface
GEM.app	Genomes Management Application
GERP	Genomic Evolutionary Rate Profiling
GJB1	Gap Junction Protein Beta-1
GPCR	G-protein Coupled Receptor
HDL	High Density Lipoprotein
HEK	Human Embryonic Kidney
HMN	Hereditary Motor Neuropathy
HMSN	Hereditary Motor and Sensory Neuropathy
HNPP	Hereditary Neuropathy with Liability to Pressure Palsies
HPO	Human Phenotype Ontology
HR2	Heptad Repeat 2 domain
HSN	Hereditary Sensory Neuropathy
HSN2C2	Hereditary Sesory Neuropathy type IIC
HSP	Hereditary Spastic Paraplegia
HSPB1	Heat-shock Protein 27
HSPB8	Heat-shock Protein 22
ID	Intellectual Disability
IGHMBP2	Immunoglobin mu-binding Protein 2
kDa	Kilodalton
KIF1A	Kinesin Family Member 1A

KIFs	Kinesin family of proteins
LDL	Low Density Lipoproteins
LHSC	London Health Sciences Centre
LITAF	Lipopolysaccharide-induced Tumor Necrosis Factor-Alpha Factor
LMNA	Lamin A/C
LMND	Lower Motor Neuron Disease
Mb	Megabase
MCV	Motor Conduction Velocity
MERRF	Myoclonic Epilepsy associated with Ragged-red Fibers
MFN2	Mitofusin 2
ml	Milliliter
MPZ	Myelin Protein Zero
MRI	Magnetic Resonance Imaging
mRNA	Messenger Ribonucleic Acid
MSL	Multiple Symmetric Lipomatosis
NCBI	National Center for Biotechnology Information
NCS	Nerve Conduction Studies
NCV	Nerve Conduction Velocity
NEFL	Neurofilament Protein, Light Polypeptide
ng	Nanogram
NGS	Next Generation Sequencing
NIH	National Institutes of Health
NR	No Response
OMIM	Online Mendelian Inheritance in Man

PBS	Phosphate Buffered Saline
PC	PhenomeCentral
PCR	Polymerase Chain Reaction
PL	Peak Onset Latency
PMP22	Peripheral Myelin Protein 22
PRX	Periaxin
QC	Québec
qPCR	Quantitate Polymerase Chain Reaction
RAB7A	RAS Associated Protein RAB7
RBC	Red Blood Cell
RI	Research Institute
RNA	Ribonucleic Acid
RRF	Ragged Red Fibers
SH3TC2	SH3 Domain and Tetratricopeptide Repeat Domain 2
SIFT	Sorting Tolerant From Intolerant
SIGMAR1	Sigma non-opioid Intracellular Receptor 1
SMA	Spinal Muscular Atrophy
SMARD1	Spinal Muscular Atrophy with Respiratory Distress Type 1
SNAP	Sensory Nerve Action Potential
SNCV	Slowed Nerve Conduction Velocity
SNP	Single Nucleotide Polymorphism
SOX8	SRY-Box 8
SOX9	SRY-Box 9
SOX10	SRY-Box 10

SPTLC1	Serine Palmitoyltransferase, Long-chain Base Subunit
TBS-T	Tris-Buffered Saline
TLC	Total Lung Capacity
T_M	Melting Temperature
TRPV4	Transient Receptor Potential Cation Channel Subfamily V, Member 4
UCSC	University of California Santa Cruz
UK	United Kingdom
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WT	Wild-type

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Impact of Rare Diseases

Rare diseases are defined as those affecting fewer than 1:2000 individuals and are estimated to affect at least 1:50 individuals world-wide (Orphanet). Eighty percent of rare diseases have their onset in childhood and it is estimated that 30% of infants with a rare disorder succumb within their first year of life (Dodge et al., 2011); those that survive longer experience higher mortality rates over their lifetime (Dye et al., 2011; Yoon et al., 1997). Both children and adults, with rare diseases, have a disproportionate number of hospital admissions that are both longer and more costly than other patients (Yoon et al., 1997). In addition to hospital-related costs, there are tremendous costs associated with outpatient care (e.g. rehabilitation, drug regimens), direct non-medical services (e.g. educational support, assistance with activities of daily living) and the indirect loss of productivity experienced by the patient and family members. There are also incalculable emotional effects on patients and their family, not only due to suffering from a disease, but living with a disease where often very little is known about prognosis and a definitive treatment is lacking. Thus, with an estimated one million Canadians affected with a rare monogenic disease (Baird et al., 1988; Carter, 1977), it is easy to see that rare disease has a significant impact on our society. Genetic lower motor neuron diseases (LMND) and peripheral neuropathies are two groups of rare diseases associated with significant morbidity, in addition to clinical and genetic heterogeneity, resulting in a significant impact on diagnostic assessment and patient management.

1.2 Lower Motor Neuron Diseases

Lower motor neurons are located in the anterior gray column of the spinal cord or brain stem and have axons that travel by way of the cranial or peripheral nerves to the motor end plates of the muscles. It is through these nerves that neural impulses reach the muscle, resulting in contraction, either driven by reflex activity and/or by voluntary actions. Physical damage to these nerves results in the loss of both voluntary and reflex movements (flaccid paralysis) as well as severe muscle atrophy. Clinically, LMND is characterized by muscle weakness, atrophy, fasciculations and muscle cramps. There is a spectrum of LMNDs, ranging from life threatening and severely debilitating such as amyotrophic lateral sclerosis (ALS) and childhood-onset spinal muscular atrophy (SMA), to more mild disorders such as segmental SMA. Depending on the cause of the LMND, age of onset can vary from prenatal onset to 80 years of age (Van Den Berg-Vos et al., 2003). These diseases can progress in a rapid or slow manner or, in some cases, begin rapidly followed by slow progression, or even spontaneous arrest in disease course (Van Den Berg-Vos et al., 2003). Different LMNDs can be inherited in an autosomal dominant, autosomal recessive or X-linked manner and can affect either the arms or legs, or both in an asymmetrical or symmetrical manner. Mode of inheritance, age of onset, disease progression and localization of muscle weakness are important features to facilitate patient diagnosis and inform long-term prognosis.

1.3 Peripheral Neuropathies

The peripheral nervous system consists of the nerves located outside of the brain and spinal cord. Peripheral neuropathies can be sub-divided into hereditary sensory neuropathy (HSN) and hereditary motor neuropathy (HMN) depending on which nerves, sensory or motor, are affected. The clinical features of peripheral neuropathies differ based on which nerves are

affected, but generally result in distal weakness, numbness and pain. If both sensory and motor nerves are affected, the group of rare disorders is referred to as Charcot-Marie-Tooth (CMT) disease. The classic CMT phenotype involves length-dependent weakness and sensory loss that begins in the feet and progresses to the knees prior to symptoms appearing in the hands. Clinically, demyelinating and axonal subtypes of CMT are distinguished by nerve conduction studies (NCS) of the median/ulnar nerve. Patients with nerve conduction velocities (NCV) <38 m/s are classified as having demyelinating disease, and patients with NCV >38 m/s but with a reduced compound muscle action potential have axonal disease. Intermediate CMT can have velocities anywhere in the range from 25 to 45 m/s. Nerve biopsies from patients with demyelinating CMT show segmental de/re-myelination forming the pathological finding of "onion bulbs", whereas patients with axonal disease show axonal loss with no evidence of de/re-myelination. CMT can be further classified into subtypes based on inheritance and demyelinating/axonal patterns; CMT type 1-4. CMT1 refers to autosomal dominant demyelinating types of CMT and CMT4 the autosomal recessive and demyelinating forms. CMT2 refers to axonal degeneration and can be dominant or recessive. Intermediate CMTs have both demyelinating and axonal features and can be inherited in a dominant or recessive manner. CMTX is caused by mutations of X-linked genes. CMT3 is a term not used anymore but historically represented Dejerine-Sottas disease, an early-onset demyelinating neuropathy that has now been grouped within CMT1 and 4.

Diagnosing the specific type of peripheral neuropathy can be difficult as there is much variability in symptoms between different neuropathies. A peripheral neuropathy can be syndromic, being part of complex, multisystem disorders such as spinocerebellar ataxia, porphyria, or disorders of lipid or mitochondrial metabolism. CMT, HSN and HMN refer to

nonsyndromic forms of peripheral neuropathies, as the predominant feature is the neuropathy itself.

There are at least 51 genes known to cause CMT, with an additional 20 genes known to cause HMN and 17 causing HSN (Table AA1), but ~50% of all cases have no molecular diagnosis (Murphy et al., 2012; Rossor et al., 2015). CMT1 is the most common form of CMT, representing over two-thirds of affected individuals, of which CMT1A, caused by a duplication of the 17p region, which contains the *PMP22* gene, is the predominant cause. CMT2A, caused by mutations in *MFN2*, is the most common form of CMT2, but represents only 10% of CMT2 patients. Fifty-three percent of all intermediate CMT cases are caused by mutations in *GJB1*. The genes most commonly responsible for HSN and HMN are *SPTLC1* and *HSPB1*, respectively. A recent study from the UK aiming to determine the frequencies of known CMT genes, determined that 19% and 75% of all CMT1 and CMT2 patients, respectively, have no known molecular diagnosis (Rossor et al., 2013). Even with >80 genes known to cause CMT and related disorders, it is estimated that there are dozens of genes yet to be discovered (Rossor et al., 2013).

1.4 Standard-of-Care Testing for Lower Motor Neuron Diseases and Peripheral Neuropathies

Each province in Canada has its own standard-of-care testing for LMNDs and peripheral neuropathies. The majority of LMND testing is performed outside of Canada. In Ontario, the clinical approach for an inherited peripheral neuropathy is first to identify the mode of inheritance (dominant, recessive, X-linked) and review NCV studies to determine if the neuropathy is axonal or demyelinating. If demyelinating features are present, the patient will undergo 17p duplication testing followed by a CMT1/Intermediate CMT gene panel. If the

NCV studies show an axonal type of CMT, depending on whether the patient has sensory, motor or sensorimotor involvement, a HSN, HMN or CMT2/ Intermediate CMT panel, respectively, can be performed. At the time patients were enrolled in this study, a CMT1 panel contained two genes; *PMP22* and *MPZ*. *GJB1* was sequenced for CMTX patients and *MFN2* for CMT2 patients. Panels have since expanded for Ontario patients as testing available from London Health Sciences Centre (LHSC; http://www.lhsc.on.ca/) includes sequencing of *PMP22, MPZ, LITAF, EGR2, NEFL, GDAP1, GJB1, FIG4, PRX* and *SH3TC2* for patients with a demyelinating subtype of CMT. The current axonal gene panel from LHSC contains *MFN2, RAB7A, LMNA, TRPV4, GARS, NEFL, GDAP1, HSPB1, MPZ, HSPB8* and *GJB1*. Patients with a negative panel can be offered participation in research studies to facilitate novel gene discovery.

1.5 Identification of Rare Disease Genes

Using Online Mendelian Inheritance in Man (OMIM; http://omim.org/) (McKusick, 2007), a database of human genes and their associated genetic disease, and Orphanet (http://www.orpha.net/) (Ayme et al., 1998), a comprehensive online database for information on rare diseases, one can estimate that there are between 6,000 and 7,000 rare genetic diseases. At present, approximately 3,500 genes have been discovered. Most of these genes have been identified through the labour- and resource-intensive processes of linkage mapping and candidate gene analysis. However, in the past few years next-generation sequencing (NGS) has changed the landscape of rare genetic disease research, accelerating the rate of gene discovery (Boycott et al., 2013). Two NGS methods, whole-genome sequencing (WGS) and whole-exome sequencing (WES), offer unbiased and powerful approaches for identifying rare disease-causing genes. WGS aims to sequence all

the DNA of an individual, coding and non-coding, while WES only sequences the coding regions of the genome. For a gene to be judged as novel disease-causing, additional families with the same disorder and a deleterious-appearing variant(s) in the same gene must be identified, or functional studies must provide additional experimental evidence of pathogenicity. Many current practices see a combination of WES, to identify candidate rare-disease-causing genes in a single family, and large-scale data-sharing methods to identify a second family, to support the pathogenesis of the gene as disease-causing.

1.6 Whole-Exome Sequencing

The exome refers to the ~1% of the genome that codes for proteins. Whole-exome sequencing allows for the majority of the coding genome to be sequenced in a timely and cost-effective manner. Compared to WGS, WES generates a proportionally greater amount of relevant and interpretable data, has fewer non-specific variants and is less expensive. Since the first use of WES in 2010 until December of 2012, >180 novel disease genes were discovered using this technology (Boycott et al., 2013). It is estimated that by 2020, most of the remaining undiscovered genes for the ~7,000 monogenic rare diseases will be identified (Boycott et al., 2013). The principle behind WES is target enrichment through shearing and hybridizing genomic DNA to oligonucleotide probes designed to target the coding region of the genome. Available commercial enrichment kits attempt to maximize coverage of the 176,266 exons from 18,409 genes found in the NCBI Consensus CDS database (Pruitt et al., 2009). Comparison of the sequencing data from any individual to the reference genome reveals on average 20,000 variants. Using WES data from two databases (Exome Variant Server, http://evs.gs.washington.edu/; 1000 Genomes Project, http://www.1000genomes.org/) as controls, and allele frequencies from dbSNP

(http://www.ncbi.nlm.nih.gov/SNP/index.html), <500 variants per exome can be classified as rare (occurring at ≤1% allele frequency). More recently, the Exome Aggregation Consortium (ExAC; http://exac.broadinstitute.org/), a WES dataset of >60,000 unrelated individuals not affected by severe pediatric disease, is also now being utilized to determine allele frequencies for candidate variants. Filtering techniques, to identify disease-causing variants, are then adapted to each disorder and change based on the particular samples being analyzed and presumed mode of inheritance. In-silico predictors of genetic variants, SIFT (Ng and Henikoff, 2003), Polyphen-2 (Adzhubei et al., 2010), Mutation Taster (Schwarz et al., 2010) and GERP (Cooper et al., 2005), provide an estimate into the pathogenicity/ conservation of specific variants (Table 1.1). SIFT scores predict the pathogenicity of a non-synonymous variant by predicting the degree of conservation of amino acid residues in closely related sequences. Polyphen-2 and Mutation Taster predict pathogenicity through sequence and structure based predictors. GERP scores are based upon the evolutionary conservation of the substituted amino acid.

1.7 Data Sharing

As the price of NGS continues to decrease, more research programs and clinical laboratories around the world are utilizing WES to determine the cause of genetic disorders. This has created a need for centralized databases where variant data, candidate genes and phenotypic data can be shared among researchers to identify other families with the same rare disease. PhenomeCentral (PC) and Genomes Management Application (GEM.app) are two such databases designed to facilitate new gene discovery through data-sharing. PhenomeCentral is a "repository for secure data sharing targeted to clinicians and scientists working in the

Table 1.1. In silico predictors that facilitate the interpretation of variant pathogenicity.

In Silico Predictors	Deleterious Threshold*
SIFT	>0.95
Polyphen-2	>0.5
Mutation Taster	>0.5
	Conservation Threshold**
GERP	>2

*Values obtained from Li et al. (Li et al., 2014).
**Value generally accepted, as stated by the National Institutes of Health (NIH; http://www.nih.gov/)

rare disorder community…" and "…encourages global scientific collaboration while respecting privacy of patients profiled…" (https://phenomecentral.org/). Genomes Management Application "is an evolving set of tools that facilitates the storage, annotation and analysis of genome-scale variant data" (https://genomics.med.miami.edu/) with a focus on neuromuscular disorders. PC emphasizes both detailed clinical phenotype information as well as genetic variants, while the main focus of GEM.app is genetic variants with high level disease descriptors. PC interface also connects to other data-sharing sites, such as GeneMatcher (https://genematcher.org/), a data-sharing site where clinicians and scientists can upload candidate genes for their patients with a certain disorder. Both GEM.app and PC have been shown to be useful data-sharing tools to enable discovery.

1.8 Care for Rare

Care for Rare (C4R) is a collaborative team of clinicians, informaticians, scientists and researchers that utilizes NGS and data-sharing to identify rare disease-causing genes. The goal of

C4R is to, "improve clinical care for patients and families affected by rare diseases" (www.care4rare.ca) through the completion of the following four objectives:

1. Use next-generation DNA sequencing to identify disease-causing genetic changes underlying rare genetic disorders affecting Canadian families.
2. Facilitate the integration of this type of testing into clinical care.
3. Initiate a low cost and rapid computer and laboratory experiment-based exploration of repurposing of clinically approved agents for novel indications.

4. Promote and contribute to the agenda of rare diseases in Canada through a role in Orphanet-Canada and as a facilitator and stakeholder of a Canadian Plan for Rare Diseases.

1.9 Rationale

WES is emerging as a clinical diagnostic tool in several jurisdictions. Traditional diagnostic assessments for patients with rare diseases are often lengthy, expensive and futile. It is estimated that 50% of patients with rare diseases never receive a molecular diagnosis (Shashi et al., 2014), so definitive treatment options are usually limited and future research on their particular disease is impacted. WES has the potential to not only be more cost-effective than conventional diagnostic approaches but also to dramatically shorten the diagnostic process.

1.10 Hypothesis and Specific Aims

My hypothesis is that WES can be used to facilitate our understanding of the molecular etiology of rare diseases characterized by lower motor neuron disease or peripheral neuropathy. To test my hypothesis, using the framework set up by C4R, I accomplished the following specific aims:

Specific Aim 1: Evaluate the diagnostic utility of WES with respect to lower motor neuron disease / peripheral neuropathy.

Specific Aim 2: Identify novel diseases for known lower motor neuron disease / peripheral neuropathy causing genes.

Specific Aim 3: Identify novel candidate disease genes for lower motor neuron disease / peripheral neuropathy for further study.

Chapter 2: Materials and Methods

2.1 Patient Cohort

Thirty-seven patients from across Canada (Table 2.1), who were diagnosed with LMND or peripheral neuropathy, and were negative for the standard-of-care molecular testing available within their respective province, were included in this study. Institutional research ethics board approval was obtained (Alberta Children's Hospital, McMaster University Medical Centre, Hospital for Sick Children, Children's Hospital of Eastern Ontario, IWK Health Centre) and the patients and family members were enrolled with informed consent (Table 2.1).

2.2 DNA Extraction

Whole blood was collected from patients using purple top ACD Vacutainer tubes. The blood was lysed by adding 9 ml of RBC Lysis Solution (QIAGEN) to 3 ml of blood. The solution was briefly mixed by vortex and left at room temperature for ten minutes. The solution was then centrifuged at 400 x g for 15 minutes followed by the supernatant being poured out. The pellet was re-suspended in 200 µl of 1x PBS and genomic DNA was then extracted using the QIAamp® DNA Mini and Blood Mini Kit (QIAGEN) following the Blood Spin Protocol. In brief, 20 µl proteinase K was added, followed by a quick mix by vortex. Buffer AL (200 µl) was then added to the sample followed by 15 seconds of vortexing. Samples were incubated at 56°C for ten minutes, followed by the addition of 200 µl of 100% ethanol, a 15 second vortex and a brief centrifugation. The sample was added to the QIAamp Mini spin column and centrifuged at 6,000 x g for one minute. The filtrate was discarded, the spin column was placed in a clean collection tube, and 500 µl Buffer AW1 was added. Following

Table 2.1. Patients recruited at Care4Rare sites.

Clinic Location	Number of Patients
Alberta Children's Hospital	3
McMaster University Medical Centre	14
Hospital for Sick Children	3
Children's Hospital of Eastern Ontario	16
IWK Health Centre	1

a 6000 x g centrifugation for one minute and the discarding of the filtrate, 500 µl Buffer AW2 was added to the column and centrifuged for three minutes at 20,000 x g. To ensure no carryover of Buffer AW2 occurred, the column was placed in a new collection tube and centrifuged at full speed for one minute. Pre-warmed Buffer AE (100 µl) was added to the column and left to sit at room temperature for 30 minutes followed by a one minute spin at 6,000 x g to elute the DNA. Concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific).

2.3 Whole-Exome Sequencing and Data Analysis

Exome capture and high-throughput sequencing of DNA was performed at McGill University and Genome Quebec Innovation Centre (Montreal, Canada) from total genomic DNA extracted from blood. Target enrichment for each sample was performed using the Agilent SureSelect 50Mb (V5) All ExonKit. Sequencing (Illumina HiSeq 2000 Systems, San Diego, CA) generated approximately 6 Gbp of 100 base pair, paired-end reads per sample. Read alignment, variant calling (single nucleotide variants and insertions/deletions), and annotation were performed as previously described (Wang et al., 2010). The mean sequencing coverage was in the range of 80-120X. Each sample had >95% of bases in the consensus coding sequences (CCDS) covered by at least 20 reads.

All rare and nonsynonymous rare coding variants were examined. Rare variants are defined as those with a mapping allele frequency less than 1% in the databases *Exome Variant Server* (EVS; http://evs.gs.washington.edu/EVS/) and *1000 Genomes Project* (http://www.1000genomes.org/). The suspected inheritance pattern of the disorder in each patient facilitated filtering for either dominant (heterozygous) or recessive (homozygous; multiple heterozygotes) alleles. Heterozygous variants for clearly dominant disorders were

filtered out if they were also seen >5 times in ExAC. Statistical evaluations such as the GERP scores were utilized to show conservation of an affected residue, and SIFT, Polyphen-2 and Mutation Taster provided a prediction regarding the pathogenicity of the variant. Genes known to cause CMT and related disorders (Table AA1) were examined first, followed by a panel of all known disease-causing genes as documented by OMIM (~3500). If these analyses were unrevealing, candidate genes were selected based on gene expression pattern, predicted pathogenicity, conservation and functional insight based on published studies performed using animal models. Variants in candidate disease-causing genes were confirmed using Sanger sequencing in both the patient and available family members.

2.4 Primer Design and PCR

For each candidate variant being considered, the genomic DNA sequence (assembly GRC Ch37/hg19) was obtained from the University of California Santa Cruz (UCSC) Genome Browser website (http://genome.ucsc.edu/). Primers were designed using the Helmholtz Centre Munich, ExonPrimer tool (https://ihg.gsf.de/ihg/ExonPrimer.html) with an optimal primer size of 20 bp and an optimal melting temperature (T_M) of 60°C. DNA amplification was performed by PCR using GoTaq® Green Master Mix (Promega #M712) using primers specific to each variant (Table 2.2).

Table 2.2. The forward and reverse primers that were used to PCR amplify variants of interest.

Gene (variant)	**Forward Primer** (5`-3`)	**Reverse Primer** (5`-3`)
BICD2 (c.A2195G)	CAGAGCACCTCTAGGCTGAC	GAGGCTAAGGGTTCCAGAGG
DNM2 (c.G1051A)	TAAACCCTGGCTTGACTTGG	TTGAGACCTTATTGCCTGGG
BAG3 (c.C1516T)	AAAGTGGAAGCCATCCTGG	TGCATGCAACTTAAAATTCCC
FBXO38 (c.3050-4T>G)	CTGCAGAATTTGGCTCTGTC	CATTTAACACTGCACTGCTCC
GARS (c.T1573C)	TTGTTCTCAGGCTCATTTTCTG	CTTTCAGAACCTGGCAGATG
SIGMAR1 (c.T194A)	GTGAGCGCAAAGCCTCAG	GCCAAACATCAGAAAGGAGG
SIGMAR1 (c.18delC)	GTGAGCGCAAAGCCTCAG	GCCAAACATCAGAAAGGAGG
PRX (c.G1936A)	AGTGCAGCTTCCAAAAGTCC	ACGTGATGGGGACTCTGC
PRX (c.A1604G)	CTGAAGTGAAGCTCCCCAAG	TTCATCTCTGGGACTTTCGG
MFN2 (c.C2119T)	ACACACCCCAACTGGGTCCCT	CCAGCCTCACCTGAGCAGCTT
KIF1A (c.G430T)	CATCTCCTCACTCAGGGGTC	GAGGTGAAGGGGCTTCCTC
IGHMBP2 (c.2601_2604del)	GACCAGCCTGATCTGAGGAC	CATGCTCTGAGTGCTGCCTA
SOX8 (c.422+5G>C)	TGAAGCGGCCCATGAACGCAT	AGCACCTCCAGAGAGCCGCAT
SOX8 (c.583_584insC)	CTGGTCAGAGCCTCCTGG	CCTTCTCAGCCAGAAACCC

2.5 Sanger Sequencing

Sanger sequencing reactions were run at the McGill University and Génome Québec Innovation Centre (Montreal, QC). To prepare for sequencing, 8 µl of unpurified PCR product diluted in 20 µl of dH_2O was sent to the Innovation Centre along with 13 µl of forward and reverse primer at a concentration of 5 µM. The sequencing products were then retrieved from the Innovation Centre Portal, Nanuq, and aligned to a reference sequence for analysis. Sequence alignment and variant analysis were both performed using Geneious® (version 8.1.4) software from Biomatters Ltd.

2.6 RNA Extraction and cDNA Synthesis

RNA was extracted from both human lymphoblast and fibroblast cells using the RNeasy® Mini Kit (QIAGEN), Total RNA from Animal Cells Using Spin Technology protocol, as outlined in the accompanying handbook. Trypsinized fibroblasts, or lymphoblasts, were centrifuged for five minutes at 300 x g to form a pellet. One volume (350 µl) of Buffer RLT and one volume of 70% ethanol was added to the pellet and mixed well by pipetting. The sample was transferred to an RNeasy Mini spin column, centrifuged for 15 seconds at ≥8,000 x g and the flow-through was discarded. Buffer RW1 (350 µl) was then added followed by centrifuging the sample for 15 seconds at ≥8,000 x g. To perform the on-column DNase digestion, 80 µl DNase I incubation mix was added directly to the RNeasy column membrane and left at room temperature for 15 minutes. This was followed by adding 350 µl Buffer RW1 and centrifuging for 15 seconds at ≥8,000 x g. Buffer RPE (500 µl) was then added to the column and centrifuged for 15 seconds at ≥8,000 x g followed by adding Buffer RPE (500 µl) and centrifuging for two minutes at ≥8,000 x g. The column was dried by placing it in a new collection tube and centrifuging at full speed for one minute. To elute the

RNA, 40 µl of RNase-free water was added to the column and centrifuged for one minute at ≥8,000 x g. Concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific).

Complementary DNA (cDNA) was synthesized using iScript™ Advanced cDNA Synthesis Kit for RT-qPCR (BIO-RAD). Each reaction contained 4 µl of 5x iScript advanced reaction mix, 1 µl iScript advanced reverse transcriptase, ~5 µg RNA template and enough nuclease-free water to bring the total reaction volume to 20 µl. Once all reagents were added, the sample underwent the following incubations:

Step	Temperature (°C)	Time
1	42	30 minutes
2	85	5 minutes
3	4	∞

Samples were stored in a -20°C freezer.

2.7 Quantitative PCR

A quantitative PCR (qPCR) was required for two patients in the study: CH0082 and 120879P. RNA was isolated from a lymphoblast cell line established from CH0082 using Qiagen RNeasy Mini Kit and reverse transcribed into cDNA with iScript Advanced cDNA Synthesis Kit (BIO-RAD) as described in Chapter 2.6. The qPCR was performed with iQ™ SYBR® Green Supermix (BIO-RAD) using the *IGHMBP2* specific primers 5'-GAAGACCCTGGTGGAGTATTT-3' and 5'-CTGGGAGTTCTCATGGGAATAG-3' for transcript NM_002180.2, and 5'-CTGCTGAAGGCCAGAAAG-3' and 5-'CCAGGGATGTGTCTACAGATTG-3' for transcript XM_005273976.1. Each reaction

contained 10 µl of 2x iQ™ SYBR® Green Supermix,0.4 µl of 5 µM forward primer, 0.4 µl of 5 µM reverse primer, ~125 ng cDNA template, and enough dH_2O to bring the total reaction to 20 µl. *GAPDH* and *HPRT1* were used as house-keeping genes with a healthy individual as a control sample. The qPCR was performed using a CFX96™ Real-Time System with the following program:

Step	Temperature (°C)	Time
1	95	3 minutes
2	95	10 seconds
3	60	10 seconds
4	72	30 seconds
5	Plate read	
6	Return to step 2 – 39X	

For patient 1208798P, RNA was isolated from a fibroblast cell line also using Qiagen RNeasy Mini Kit and reverse transcribed into cDNA with iScript Advanced cDNA Synthesis Kit (Bio-Rad) as described in Chapter 2.6. Each reaction contained 1 µl of 1x Taqman® Gene Expression Assay, 10 µl of 1x TaqMan® Gene Expression Master Mix, 100 ng of template cDNA, and 5 µl of RNase-free water. A *SOX8* specific TaqMan® Gene Expression Assay (Thermo Fisher Scientific# Hs00232723_m1) was used with the following program:

Step	Temperature (°C)	Time
1	95	10 minutes
2	95	15 seconds
3	60	1 minute
5	Plate read	
6	Return to step 2 – 39X	

CFX™ Manager v3.0 (BIO-RAD) was used to interpret qPCR data. *GAPDH* and *HPRT1* were used as house-keeping genes with a healthy individual as a control sample.

2.8 Protein Extraction

Protein was extracted using RIPA buffer (25 ml of 1M Tris-HCl, pH 7.4; 5 ml of 1% NP-40; 2.5g of 0.5% sodium deoxycholate; 0.5g of 0.1% SDS; 15ml of 5M NaCl; 2ml of 0.5M EDTA; 1.05g of NaF; final volume of 500ml) for both lymphoblast and fibroblast cell lines. RIPA solution was prepared by mixing one Protease Inhibitor Cocktail Tablet (Roche) in 10 ml of RIPA buffer. Lymphoblasts were centrifuged for five minutes at 300 x g to form a pellet. Fibroblasts were grown in 10 cm plates, washed using 1x PBS, scraped from the plate in 2 ml of 1x PBS and then centrifuged for three minutes at 300 x g to form a pellet. For both lymphoblasts and fibroblasts, the supernatant was removed and 200 µl RIPA solution was added to the pellet, and mixed by pipetting. The sample was left on ice for 30 minutes and then centrifuged at maximum speed at 4°C for ten minutes. The supernatant, which contained the protein, was transferred to a new container and stored at -80°C.

2.9 Western Blot

Western blots were performed for two patients in the study: CH0082 and 120879P. Protein was isolated from patient (CH0082) and control lymphoblast cell lines, according to the protocol described in Chapter 2.8, to determine IGHMBP2 expression. Protein concentrations were determined by Bradford Assay (BIO-RAD) using the Spectra Max 340 PC plate-reader (Molecular Devices) with the Soft Max Pro 4.7.1 software (Molecular Devices). Reaction mixes, totalling 35 µl and containing 1x NuPAGE LDS Sample Buffer (Life Technologies), 1x NuPAGE Reducing Agent (Life Technologies) and 200 µg protein,

were prepared in dH_2O. The protein was loaded into a NuPAGE 4-12% Bis-Tris Gel (Life Technologies), along with 15 µl Novex Sharp Pre-stained Protein Standard (Life Technologies) and ran in 1x NuPAGE MOPS SDS Running Buffer (Life Technologies) at 150 volts for ~one hour and 15 minutes. The protein was transferred to a 0.2 µm nitrocellulose membrane (Bio-Rad) for one hour at 100 volts in 80% transfer buffer (14.4g/L glycine, 3g/L Tris base, in deionized water) and 20% methanol. The membrane was blocked for one hour in 5% skim milk in TBS-T (200ml Tris, pH 8.0; 175.32g NaCl, 10ml Tween 20; final volume of 1000ml; dilute 50ml in 950ml ddH_2O to 1X). The primary antibody, a monoclonal mouse antibody (Millipore), was diluted (1:1000) in 5% skim-milk TBS-T and incubated overnight at 4°C. After five 10-minute washes in TBS-T, the membrane was incubated, on a shaker, at room temperature for one hour in diluted (1:5000) HRP-goat anti-mouse (Zymed; Santiago, Chile) secondary anti-body in 5% skim-milk TBS-T. After another five 10-minute washes in TBS-T, Clarity Western ECL Substrate (Bio-Rad; Hercules, USA) and Amersham Hyperfilm ECL (GE Healthcare Limited; Buckinghamshire, UK) was used to visualize the protein. Protein levels were quantified using ImageJ (http://imagej.nih.gov/ij/).

Protein was isolated from patient (120879P) and control fibroblast cell lines, according to the protocol described in Chapter 2.8, to determine SOX8 expression. Other than the antibodies, the same materials and methods, as described above, were used. The primary antibody was rabbit polyclonal anti-SOX8 (Abcam# ab76196; Cambridge, USA) and diluted (1:1000) in 5% skim milk. The secondary antibody (HRP-goat anti-rabbit; Bethyl Laboratories; Montgomery, USA) was also diluted (1:2000) in 5% skim-milk.

2.10 SOX8 Bacterial Clone Isolation and Transfection into HEK 293 Cells

A human clone of the gene, *SOX8* (BC031797), subcloned into pCMV-SPORT6, was ordered from The Centre for Applied Genomics in Toronto, ON, and was shipped as an LB stab. It was streaked on an LB + ampicillin (AMP) agar plate and grown overnight at 37°C. Single colonies were picked and grown in 3 ml LB plus 3 µl AMP solution overnight in a shaking 37°C incubator. A 1 ml sample of the culture was centrifuged for three minutes at 6,800 x g and the supernatant was discarded. The vector was isolated using the QIAprep® Spin Miniprep Kit protocol as described below. The pellet was re-suspended in 250 µl Buffer P1. Buffer P2 (250 µl) was then added and the sample was mixed by inverting the tube four to six times. Buffer N3 (350 µl) was added and the tube was inverted another four to six times followed by centrifuging for ten minutes at 17,900 x g. The supernatant was applied to a QIAprep spin column and centrifuged for one minute. After discarding the flow-through, 500 µl Buffer PB was added, followed by a one minute centrifugation at 17,900 x g, to wash the sample. The sample was further washed by adding 750 µl Buffer PE followed by centrifugation for one minute at 17,900 x g. The spin column was placed in a new collection tube and centrifuged for one minute at 17,900 x g to remove residual wash buffer. The vector was eluted by adding 50 µl Buffer EB to the column, letting it stand for one minute at room temperature and then centrifuging it for one minute at 17,900 x g. The insert was verified using Sanger sequencing.

The vector containing SOX8 was transfected into Human Embryonic Kidney (HEK) 293 cells using TurboFect™ (Thermo Scientific). HEK 293 cells were grown to 70-90% confluency in a six-welled plate. A mixture containing 3 µg vector, 6 µl TurboFect™ reagent and enough DMEM to total 200 µl was prepared and left to sit for 20 minutes at room temperature. The mixture was added to the cells which were then incubated at 37°C for 48

hours. Cells containing the vector were selected for by using G418 (Life Technologies) at a concentration of 500 μg/ml. Cells were harvested after three days of selection. Protein extraction was performed according to the protocol for fibroblasts described in Chapter 2.8.

Chapter 3: Results

3.1 Evaluating the Diagnostic Utility of WES

WES was performed on the entire cohort of 37 undiagnosed patients. To evaluate the diagnostic utility of WES for CMT and related disorders, the data was analyzed for mutations in genes associated with CMT and related disorders (Table AA1). Mutations in the genes *BICD2*, *DNM2*, *BAG3*, *FBXO38*, *GARS*, *SIGMAR1* and *PRX* were identified in patients with the same phenotypes as those previously associated with these genes (Table 3.1).

3.1.1 BICD2

Patient 48579 was seen at the Children's Hospital of Eastern Ontario (CHEO) with distal wasting and bilateral clubbed feet with prenatal onset. Both the patient's mother and daughter were affected by the same disorder suggesting a dominant inheritance pattern. There were 224 rare variants identified by WES, of which two were in known CMT and related disorder causing genes. Only the disorder associated with bicaudal D2 (*BICD2*), characterized by distal muscle weakness, muscle wasting and foot deformities, corresponded to the clinical presentation seen in our patient. At the time of this finding, mutations in *BICD2* were only very recently reported to cause autosomal dominant, lower extremity predominant SMA2 (SMALED2; MIM# 615290) (Oates et al., 2013; Peeters et al., 2013). WES revealed a novel heterozygous missense variant (c.A2195G; p.E732G) with a high conservation score (GERP= 4.79) and predicted to be pathogenic (SIFT= 1; Polyphen-2= 0.983; Mutation Taster= 0.997). This variant had not been reported in any in-house or external database. Sanger sequencing showed that all three affected family members carried

Table 3.1. Diagnoses by WES in patients with peripheral neuropathy or lower motor neuron disease

Patient ID	Gene	Mode of Inheritance	Variant(s)	Disorder (MIM #)
48579	*BICD2*	Dominant	c.A2195G; p.E732G/ WT	Spinal muscular atrophy, lower extremity-predominant, 2, Autosomal Dominant (615290)
53645	*DNM2*	Dominant	c.G1051A; p.V351M/ WT	Charcot-Marie-Tooth disease, axonal, type 2M (606482)
Subject 16	*BAG3*	Dominant	c.C1516T; p.Q506X/ WT	Myopathy, myofibrillar, 6 (612954)
Subject 19	*FBXO38*	Dominant	c.3050-4T>G/ WT	Neuronopathy, distal hereditary motor, type IID (615575)
Subject 21	*SIGMAR1*	Recessive	c.T194A; p.L65Q/ c.18delC; p.G6fs	Amyotrophic lateral sclerosis 16, juvenile (614373)
Subject 22	*PRX*	Recessive	c.A1604G; p.E535G/ c.G1936A; p.E646K	Charcot-Marie-Tooth disease, type 4F (614895)
Subject 25	*GARS*	Dominant	c.T1573C; p.C525R/ WT	Charcot-Marie-Tooth disease, type 2D (601472)

the variant in *BICD2* but the patient's unaffected father and unaffected husband did not (Figure 3.1).

3.1.2 DNM2

Patient 53645 was seen at the Alberta Children's Hospital with a distal axonal sensory / motor neuropathy. The patient's father, uncle and paternal grandmother were affected with the same condition suggesting a dominant pattern of inheritance. WES identified 190 rare variants in this patient of which two were in known CMT and related disorder causing genes. Mutations in Dynamin 2 (*DNM2*) are known to cause CMT2M (MIM# 606482) (Fabrizi et al., 2007); a dominant, axonal sensory / motor neuropathy as was seen in this patient. WES revealed a novel heterozygous missense variant (c.G1051A; p.V351M) with a high conservation score (GERP= 4.44) and predicted to be pathogenic by two of the three in silico prediction programs (SIFT= 0.99; Polyphen-2= 0.302; Mutation Taster= 0.998). This variant has not been reported in any in-house or external database. Sanger sequencing confirmed this variant to be present in the patient (Figure 3.2).

3.1.3 BAG3

Subject 16, was seen at the McMaster University Medical Centre with a peripheral neuropathy causing muscle weakness and pes cavus. The patient's now deceased mother was also affected by this disorder, suggesting autosomal dominant inheritance. There were 200 rare variants identified by WES of which one was in a known CMT and related disorder causing genes, *IGHMBP2*. The phenotype and inheritance pattern associated with *IGHMBP2* mutations does not correspond to that of our patient so the entire OMIM list was searched for genes that could cause a peripheral neuropathy matching the clinical description

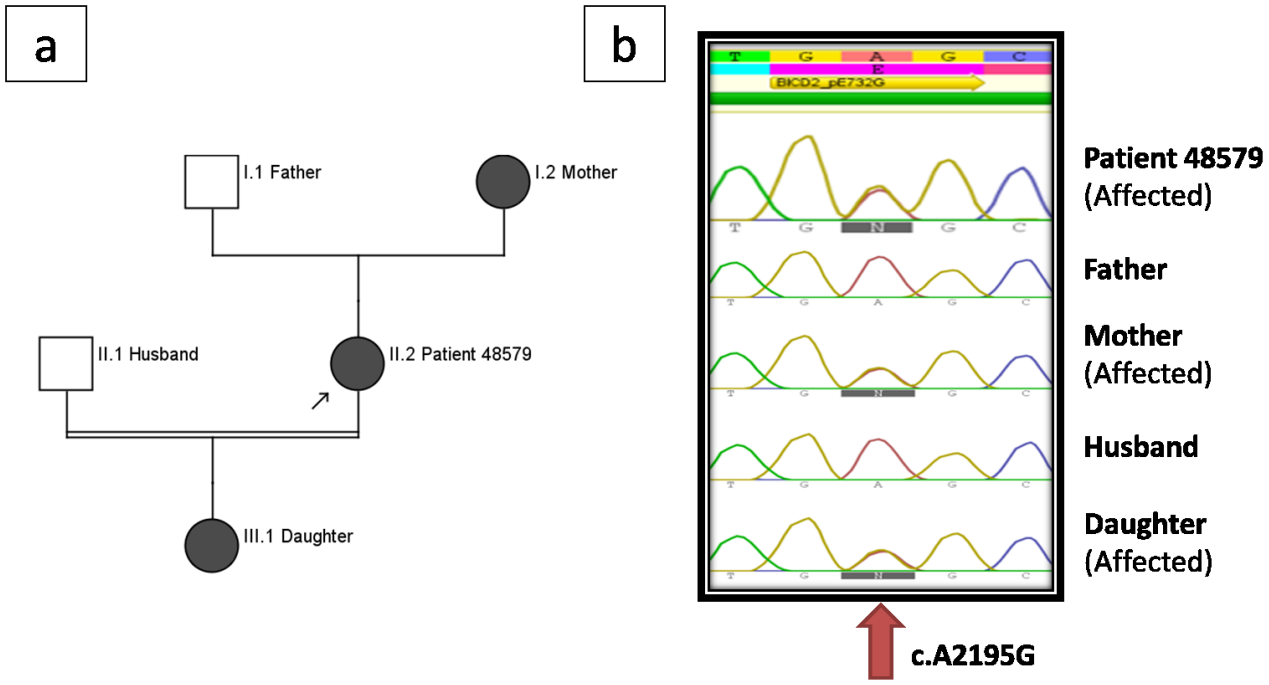


Figure 3.1. a) A pedigree showing the proband and the proband's affected mother and daughter, but unaffected husband and father. b) Sanger sequencing of the heterozygous missense variant in the three affected individuals, and the wildtype unaffected individuals.

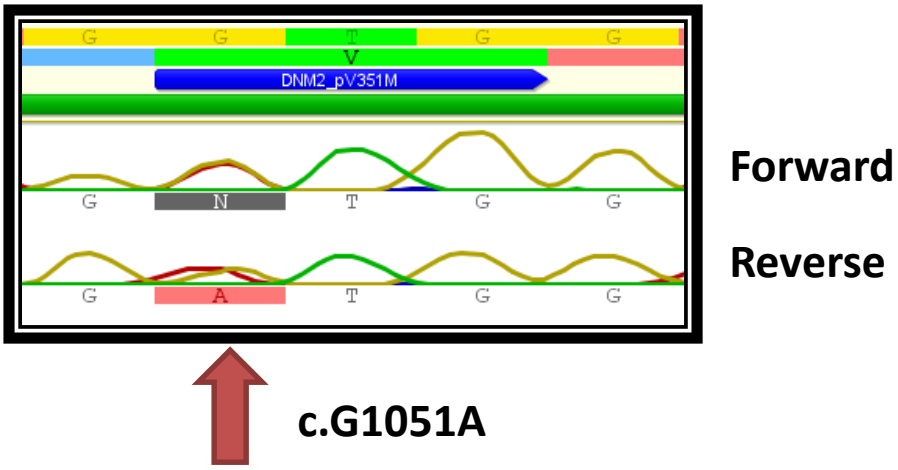


Figure 3.2. Sanger sequencing confirming the presence of the c.G1051A variant in the gene *DNM2,* in the patient.

as our patient. Mutations in BCL2-associated athanogene 3 (*BAG3*) are known to cause Myofibrillar myopathy 6 (MIM# 612954) (Selcen et al., 2009), a dominant muscle disease also characterized by an axonal and demyelinating peripheral neuropathy and causing muscle weakness and pes cavus , consistent with the clinical presentation of this patient. WES revealed a novel heterozygous nonsense variant (c.1516T; p.Q506X) in *BAG3* with a high conservation score (GERP= 5.68) and predicted to be pathogenic by two of three in silico prediction programs (SIFT= 0.904; Polyphen-2= 0.735; Mutation Taster= 1). Sanger sequencing confirmed this variant to be present in the patient (Figure 3.3).

3.1.4 FBXO38

Subject 19 was seen at McMaster University Medical Centre with a neuropathy characterized by mild wasting in the shins, pes cavus, hammer toes and carpal tunnel syndrome. The patient's mother was also predicted to be affected by this disorder, suggesting autosomal dominant inheritance. WES identified 192 rare variants of which two were in known CMT and related disorder causing genes. Mutations in Fork-box only protein 38 (*FBXO38*) are known to cause distal, hereditary, motor neuronopathy type IID (MIM# 615575), which is a dominant, progressive disorder known to cause weakness in the limbs and pes cavus (Sumner et al., 2013). WES revealed a heterozygous variant (c.3050-4T>G) in the *FBXO38* gene that was predicted to affect splicing through the mutation of the splice acceptor site. Sanger sequencing confirmed this variant to be present in the patient (Figure 3.4).

3.1.5 SIGMAR1

Subject 21 was seen at McMaster University Medical Centre with an axonal neuropathy and distal weakness. The patient had no reported family history of this disorder. There were 228

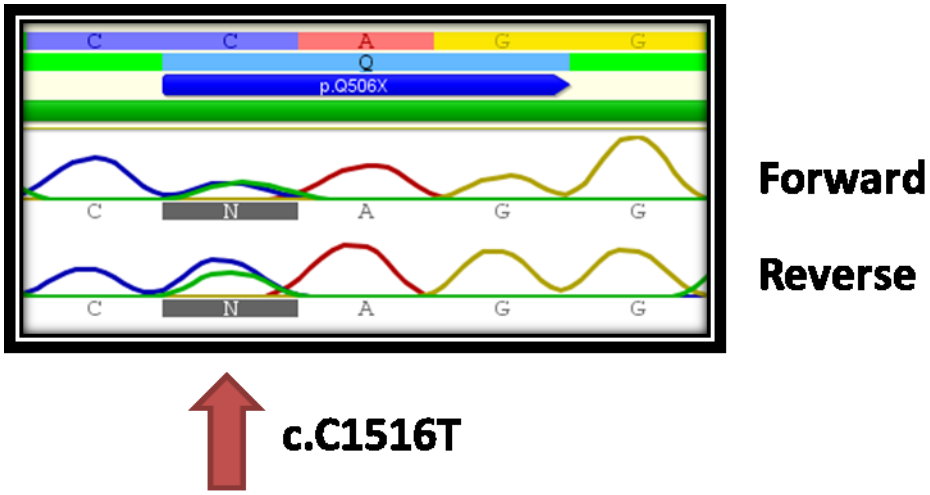


Figure 3.3. Sanger sequencing confirming the presence of the c.C1516T variant in the gene *BAG3,* in the patient.

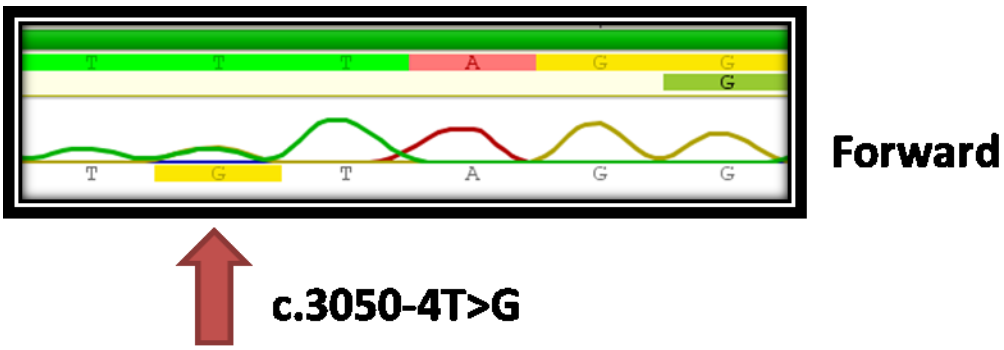


Figure 3.4. Sanger sequencing confirming the c.3050-4T>G variant in the gene *FBXO38,* in the patient.

rare variants identified through WES of which one was in a known CMT and related disorder causing gene, *KIF1A*. Our patient did not match the clinical description associated with mutations in *KIF1A*, so the entire OMIM gene list was searched for genes that could cause a peripheral neuropathy matching the clinical description of our patient. With no family history, both dominant and recessive variants had to be analyzed. WES revealed multiple heterozygous variants (c.18delC; p.G6fs and c.T194A; p.L65Q) in the gene Sigma non-opioid intracellular receptor 1 (*SIGMAR1*). Recessive mutations in *SIGMAR1* are known to cause ALS16 (MIM# 614373) (Al-Saif et al., 2011), a juvenile-onset, progressive distal muscle weakness with nerve degeneration. The novel c.18delC; p.G6fs variant is considered damaging as it causes a frameshift mutation. The c.T194A; p.L65Q variant has been reported once before in EVS and has a high conservation score (GERP= 4.45). Pathogenicity scores suggested this variant was deleterious (SIFT= 1; Polyphen-2= 0.997; Mutation Taster=0.995). Both variants were confirmed to be present in the patient by Sanger sequencing (Figure 3.5).

3.1.6 PRX

Subject 22 was seen at the McMaster University Medical Centre with severe distal wasting and bilateral foot drop caused by a peripheral neuropathy. The patient had no reported family history of this disorder. There were 233 rare variants identified through WES of which two were in a known CMT and related disorder causing gene. WES revealed two novel multiple heterozygous variants in the gene, Periaxin (*PRX*), c.A1604G; p.E535G and c.G1936A; p.E646K. Mutations in *PRX* are known to cause CMT4F (MIM# 614895), a recessive demyelinating CMT consistent with this patient's clinical presentation (Guilbot et al., 2001). The c.A1604G; p.E535G variant was conserved (GERP= 3.72) and predicted to

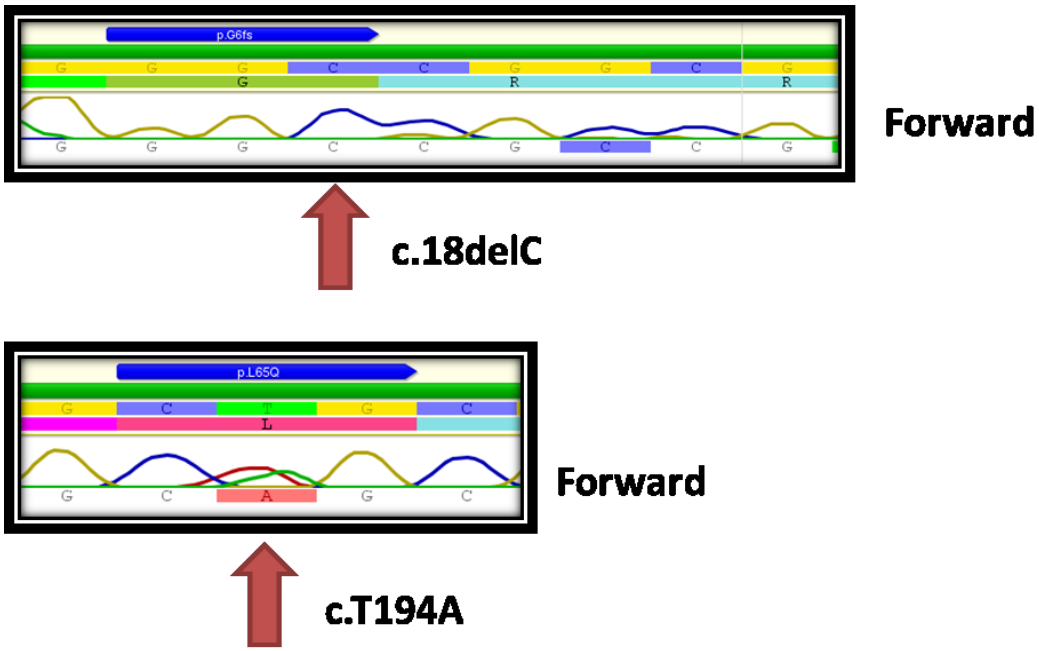


Figure 3.5. Sanger sequencing confirming the presence of both the c.18delC and the c.T194A variants in the gene *SIGMAR1,* in the patient.

be pathogenic by one of the three in silico prediction programs (SIFT= 0.83; Polyphen-2= 0.27; Mutation Taster= 0.665). The c.G1936A; p.E646K variant was also conserved (GERP score= 4.4) and predicted to be pathogenic by one of the three in silico prediction programs (SIFT= 0.67; Polyphen-2= 0.08; Mutation Taster score= 0.630). Both variants were confirmed to be present in the patient by Sanger sequencing (Figure 3.6).

3.1.7 GARS

Subject 25 was seen at McMaster University Medical Centre with a sensory axonal neuropathy and mild wasting in the legs. The patient had no reported family history of this disorder. WES revealed 200 rare variants of which two were in a known CMT and related disorder causing gene. A heterozygous variant (c.T1573C; p.C525R) was found in the gene Glycyl-tRNA synthetase (*GARS*). Mutations in *GARS* are known to cause CMT2D (MIM# 601472) (Ionasescu et al., 1996); a dominant axonal peripheral neuropathy, consistent with this patient's clinical presentation. The variant was conserved (GERP= 2.67) and predicted to be pathogenic by two of the three in silico prediction programs (SIFT= 0.96; Polyphen-2= 0.002; Mutation Taster= 1). This variant has been seen once before in ExAC, at a frequency of $8.82X10^{-4}$%, but not in any other database. Sanger sequencing confirmed this variant to be present in the patient (Figure 3.7).

3.1.8 Evaluating the Diagnostic Utility of WES: Summary

There was a ~19% diagnostic rate for the patients with CMT or related disorders, corresponding to a molecular diagnosis for seven of the 37 patients. The remaining 30 patients went on to the next stage of the study (Chapter 3.2) to identify novel diseases associated with known disease-causing genes.

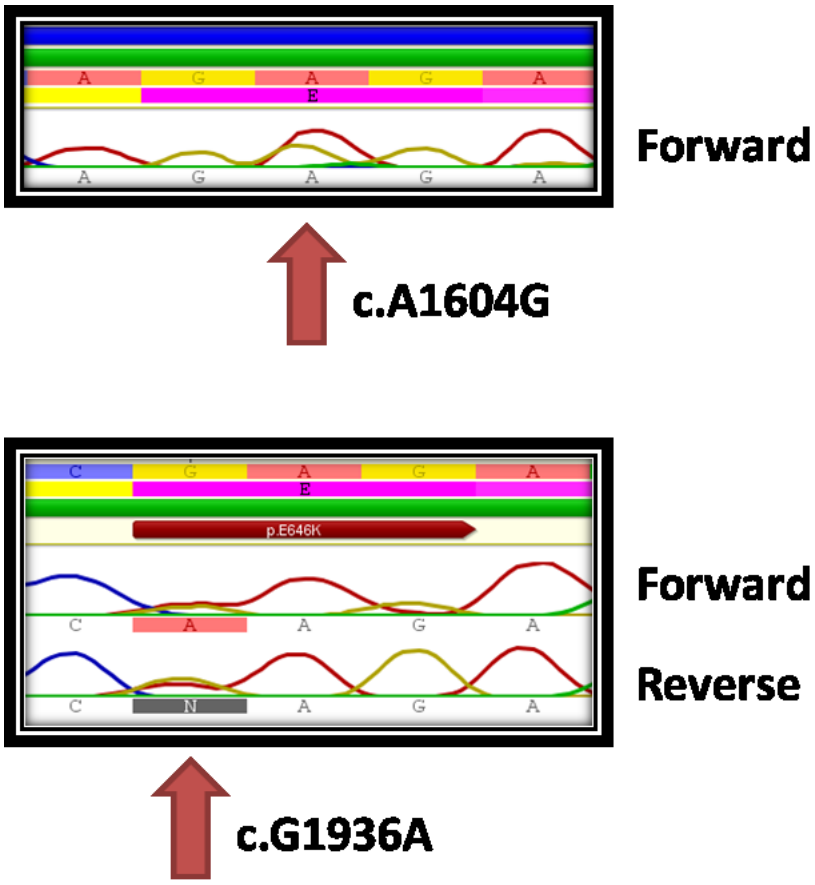


Figure 3.6. Sanger sequencing confirming the presence of both the c.A1604G and the c.G1936A variants in the gene *PRX,* in the patient.

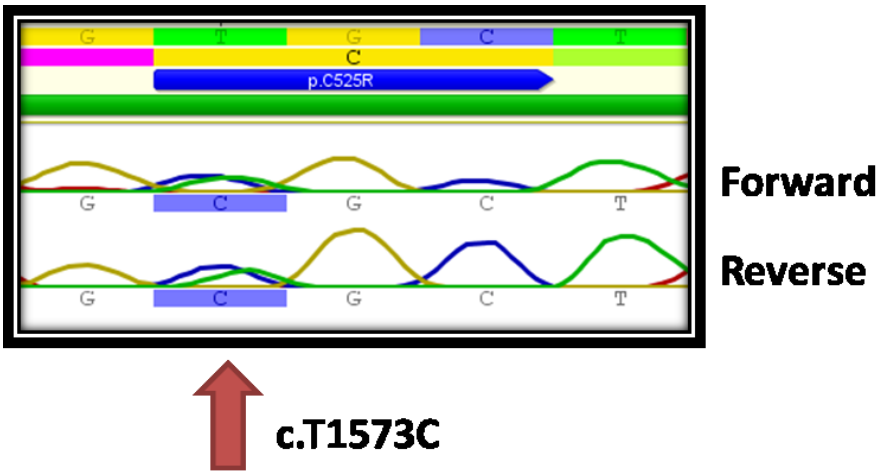


Figure 3.7. Sanger sequencing confirming the presence of the c.T1573C variant in the gene *GARS,* in the patient.

3.2 Novel Diseases Associated with Known Disease-Causing Genes

Thirty patients diagnosed with lower motor neuron disease or peripheral neuropathy that had undergone WES, and in whom a mutation in a gene known to cause CMT or a related disorder was not identified as definitively causative, were further evaluated. The expanded OMIM panel of 3500 genes was examined to identify potentially deleterious variants in genes where the phenotype was not necessarily completely in keeping with what was reported as associated with mutations in a particular gene. Of the 31 patients, four were found to have new diseases associated with a clinically distinct presentation caused by mutations in *MFN2*, *KIF1A* and *IGHMBP2* (Table 3.2). These three genes are already known to cause peripheral neuropathy or LMND, but have not previously been implicated with the specific, novel, clinical presentations described in the following sections.

3.2.1 MFN2

Two affected adult siblings, 130451 and 130105S, presented with an axonal motor and sensory peripheral neuropathy and multiple symmetric lipomatosis (MSL, MIM# 151800); 130451 was evaluated at CHEO and his brother, 130105S, was evaluated at the Alberta Children's Hospital. MSL is characterized by the accumulation of brown fat around the upper arms, neck and shoulder areas (additional clinical information for the patients can be found in Appendix B). There was a remote history of consanguinity. WES revealed 102 rare variants shared between the two siblings, of which two were consistent with a multiple heterozygous inheritance (two mutations in the same gene) and one was homozygous. The two patients shared a homozygous variant (c.C2119T; p.R707W) in the gene, Mitofusin 2 (*MFN2*), a known-disease-causing gene in which heterozygous mutations cause CMT2A (Zuchner et al., 2004).

Table 3.2. New diseases, associated with known disease genes, identified using WES.

Patient ID	Gene	Previously Reported Disorder	New/ Additional Phenotypes
130451 & 130105S	*MFN2*	Charcot Marie Tooth disease, Type 2A (MIM# 609260)	Multiple Symmetric Lipomatosis
120915T	*KIF1A*	ID, HSN or HSP	ID, HSN and HSP
CH0082	*IGHMBP2*	SMARD1 (MIM# 604320)	CMT2

ID, Intellectual disability; HSN, Hereditary Sensory Neuropathy; HSP, Hereditary Spastic Paraplegia; SMARD1, Spinal Muscular Atrophy with Respiratory Distress Type 1

The variant was located at a conserved residue (GERP= 3.67) and was predicted to be pathogenic (SIFT= 0.953; Polyphen-2= 0.976; Mutation Taster= 1). Sanger sequencing confirmed that the homozygous variant was found in each affected sibling while their two unaffected siblings were wild-type (Figure 3.8); parents were deceased and thus not available for testing. Previous to this study, mutations in *MFN2* had not been implicated in MSL. Functional studies were performed in a collaborating lab (Dr. Robert Screaton, CHEO RI) providing evidence that the c.C2119T; p.R707W variant causes the MSL, in addition to the peripheral neuropathy, in our patients (Sawyer et al., 2015a).

3.2.2 KIF1A

Patient 120915T and his monozygotic twin presented at CHEO with intellectual disability (ID), HSN and hereditary spastic paraplegia (HSP) (additional clinical information for both siblings can be found in Appendix C). WES of the proband revealed 338 rare variants, of which 58 were associated with known disease-causing genes. Forty-six were heterozygous variants leaving 12 that were homozgyous or multiple heterozygous. A novel heterozygous variant (c.G430T; p.V144F) was identified in the Kinesin family member 1A (*KIF1A*) that was predicted to be deleterious (SIFT= 1; Polyphen-2= 1). Sanger sequencing confirmed the variant to be found in both affected individuals but not present in either unaffected parent, consistent with de novo inheritance (Figure 3.9). Mutations in *KIF1A* have been implicated to cause autosomal dominant (AD) syndromic intellectual disability (ID) (Hamdan et al., 2011), recessive HSN (Riviere et al., 2011) and recessive HSP (Erlich et al., 2011), as discrete clinical presentations in separate individuals. The novelty of our family was the presence of all three phenotypic presentations in both twins. We contacted Dr. Jacques Michaud at the University of Montreal, who reported on mutations in *KIF1A* causing AD ID,

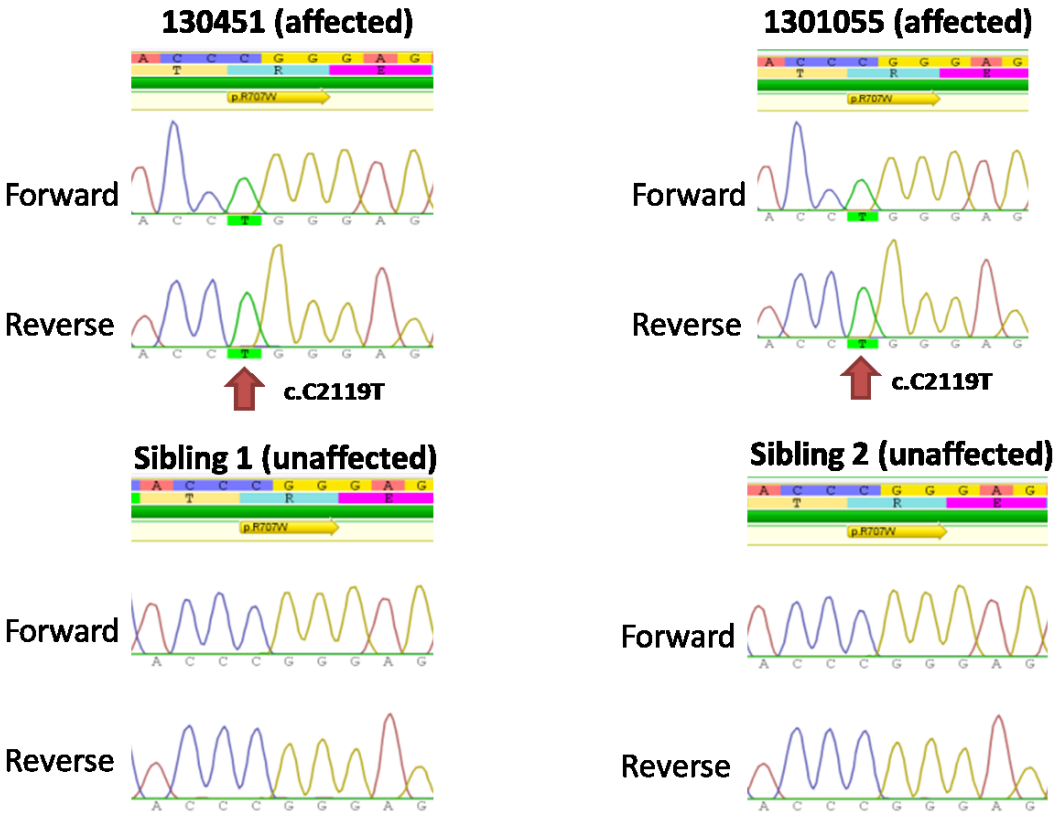


Figure 3.8. Sanger sequencing confirming the presence of the homozygous c.C2119T variant in the two affected siblings in the gene *MFN2*, but not present in the unaffected siblings.

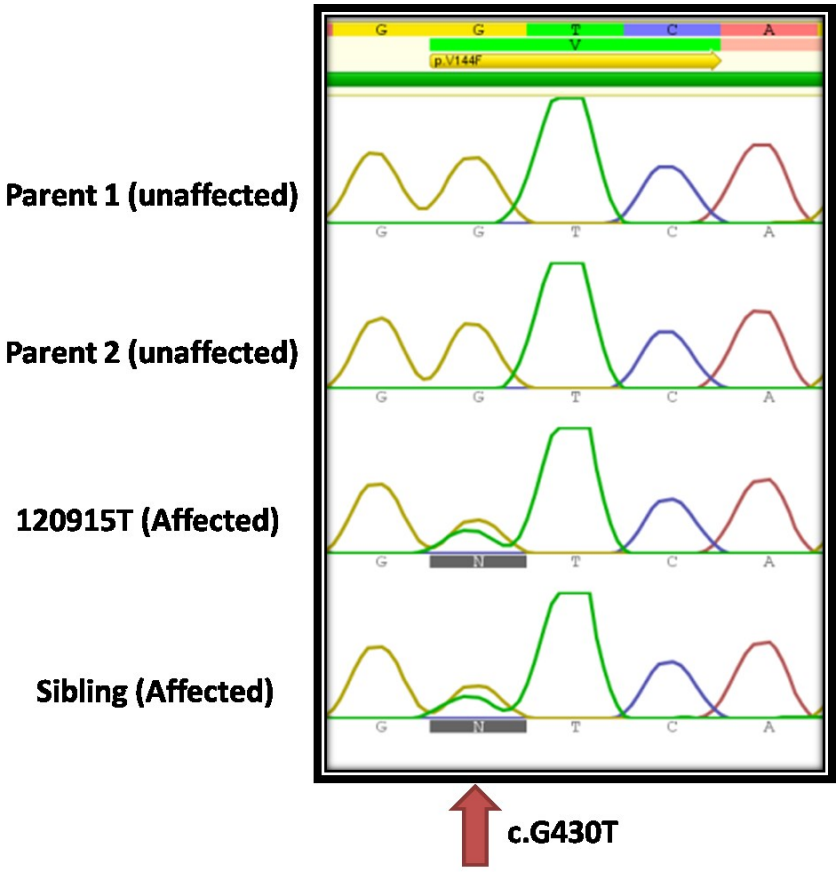


Figure 3.9. Sanger sequencing confirming the presence of the heterozygous c.G430T variant in the two affected monozygotic twins in the gene *KIF1A*, and absence in the unaffected parents, consistent with a de novo mutation.

to inquire about this gene. He had an unpublished cohort of similarly affected patients to ours, all with de novo mutations in *KIF1A*. Our patients were added to a cohort of 12 additional patients with the same novel disorder. Functional studies were performed in Dr. Michaud's laboratory, which provided further evidence that de novo dominant mutations in *KIF1A* cause a syndrome of ID, HSN and HSP (Lee et al., 2015).

3.2.3 IGHMBP2

Patient CH0082 was evaluated at CHEO with childhood-onset distal wasting in the upper and lower limbs, areflexia and decreased sensation, consistent with a CMT phenotype. The patient's younger sibling was also affected with this same disorder. They were of Syrian descent and while her parents reported no consanguinity, her maternal and paternal grandparents were from the same village. Detailed clinical information on both siblings, including patient photos (Figure AD1) and results of nerve conduction studies (Table AD1), can be found in Appendix D. WES of the patient revealed a homozygous c.2601_2604delAAAA; p.Lys868Profs*109 in the gene Immunoglobin mu-binding protein 2 (*IGHMBP2*). No statistical analysis were required for this 4 bp deletion that causes a frameshift leading to a premature stop codon and a protein truncated from 993 to 976 amino acids in length. This variant had never been seen in our internal database nor in 1000 Genomes Project or EVS. The variant had been seen in ExAC at an allele frequency of $1.803X10^{-3}$% but had never been seen in a homozygous state. Using Sanger sequencing, both patients were confirmed to carry the homozygous 4 bp deletion in exon 13 of *IGHMBP2* and both parents were heterozygous carriers of the deletion (Figure 3.10).

Until recently, mutations in *IGHMBP2* were exclusively associated with spinal muscular atrophy with respiratory distress type 1 (SMARD1; MIM# 604320), a rare

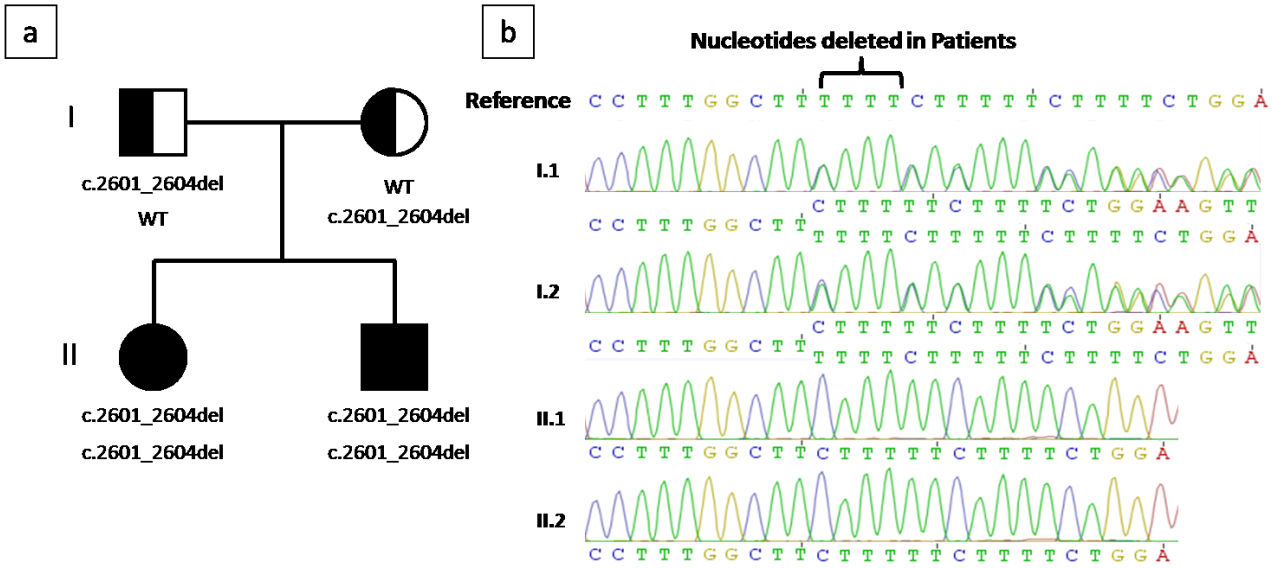


Figure 3.10. a) A pedigree showing the two patients and their carrier parents. b) Sanger sequencing of the homozygous deletion in the patients and the heterozygous deletion in both parents.

autosomal recessive disorder that typically presents within the first two months of life and is characterized by respiratory failure, caused by diaphragmatic paralysis, and a severe infantile axonal neuropathy (Grohmann et al., 2001). Here we describe two siblings with an axonal polyneuropathy in keeping with a recessive axonal CMT but with no signs of respiratory impairment.

Quantitative PCR was performed to compare *IGHMBP2* transcript (NM_002180.2) levels in the patient compared to a control (n=4). The results showed that the patient had ~65% the amount of *IGHMBP2* transcript (NM_002180.2) compared to a control (Figure 3.11). To determine whether the patient's relatively mild CMT phenotype, compared to the neuropathy and respiratory impairment seen in SMARD1 patients, could be due to alternate isoforms, a qPCR was also performed on a predicted-to-exist (2015) isoform (XM_005273976.1). Sanger sequencing of a product showed that the isoform XM_005273976.1 was present, but the qPCR did not show any significant difference in transcript levels, though with overlapping confidence intervals.

A Western blot, using an anti-IGHMBP2 antibody (Millipore #MABE162, Billerica, USA), was performed on protein isolated from patient and control lymphoblast cell lines. The Western blot analysis showed that the protein levels in the patient cell line were reduced to 40% compared to that of the control (Figure 3.12).

3.2.4 Novel Diseases Associated with Known Disease-Causing Genes: Summary

Four of the 30 (~13%) patients were diagnosed with a mutation for a novel disease in a known disease-causing gene. In total, 11/37 (~30%) of patients received a molecular diagnosis and the remaining 26 went on to the next stage of the study (Chapter 3.3) to identify novel disease genes.

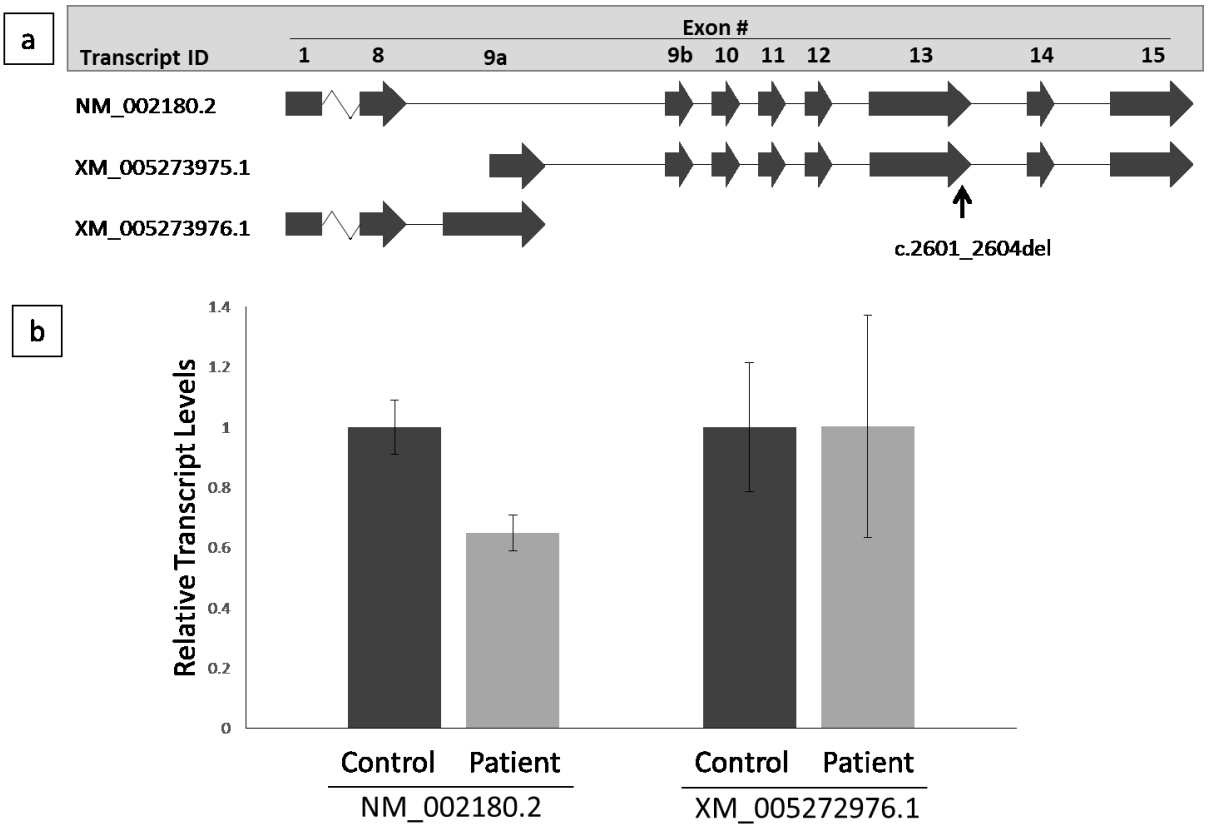


Figure 3.11. a) Three transcripts of *IGHMBP2* and the location of the CMT-causing mutation in our patients. b) The relative transcript levels for the isoforms NM_002180.2 and XM_005273976.1 in the patient compared to controls (n=4).

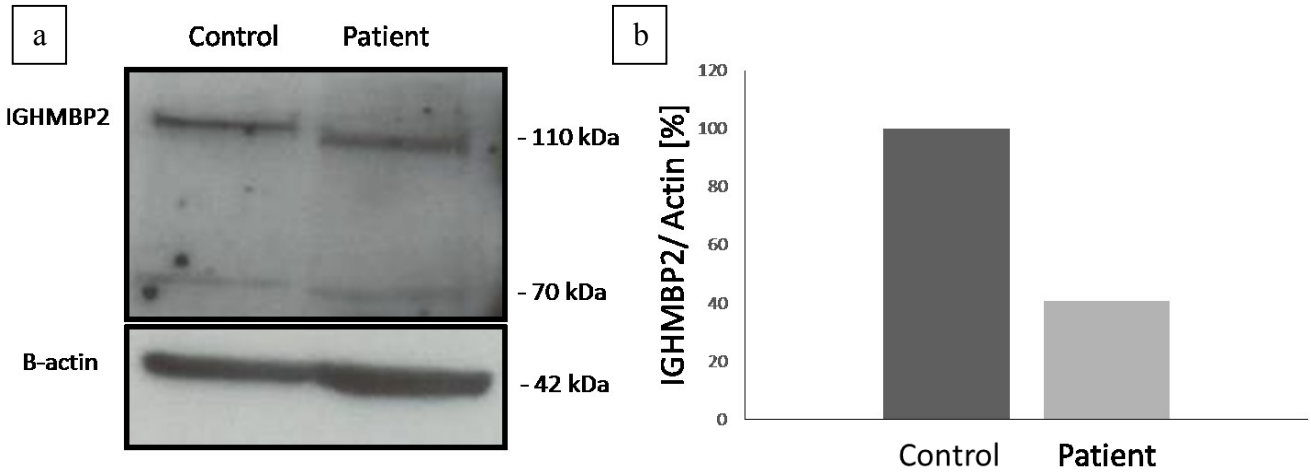


Figure 3.12. a) A Western blot of IGHMBP2 protein from control and patient lymphoblastoid cell lines. (n=1) b) IGHMBP2 protein levels normalized against β-actin in control and patient cell lines.

3.3 Identification of Novel Disease Genes

For a gene to be classified as novel disease-causing, a second family with the same disorder and a variant predicted to be deleterious in the same gene must be identified, or functional studies must provide additional experimental evidence of pathogenicity. The remaining 26 patients, with no known disease-causing genes as good candidates, were next analyzed to identify novel candidate genes. Additional supporting evidence was then sought through subsequent functional analyses and/ or large-scale data sharing to identify a second family.

3.3.1 Functional Studies of SOX8

Patient 120879P was evaluated at CHEO with congenital myopathy and lower motor neuron disease as well as skeletal dysplasia, contractures, short stature, amenorrhea and possible cataracts. WES was performed on the patient and her two unaffected parents. With no family history, both dominant and recessive variants were analyzed. WES revealed a total of 275 rare variants in the patient, of which there were three heterozygous variants not seen in either parent. There were also multiple heterozygous variants, that were inherited in *trans*, in two genes. None of these genes were known to be disease-causing so each was analyzed as a possible novel candidate (Table 3.3). The top candidate gene was *SOX8*, a transcription factor that plays a role in central nervous system, limb and facial development (Hong and Saint-Jeannet, 2005). Two multiple heterozygous variants were identified in the gene with one predicted to affect splicing through the loss of a splice donor site (c.422+5G>C) and the other variant a frameshift insertion causing a premature stop codon (c.583_584insC; p.H11fs). The frameshift truncates the protein from a molecular mass of 47 kDa to a predicted mass of 22 kDa. *SOX8*, being the most attractive candidate gene after initial analysis, was subjected to further functional evaluation.

Table 3.3. Prioritization of candidate genes from analysis of WES data for Patient 120879P.

Gene	Inheritance	Variant(s)	Excluding Factor
MUC12	Heterozygous	c.C2525T; p.T842M	Cell surface protein involved in epithelial protection; unlikely to be associated with patient phenotype
RIPK2	Heterozygous	c.1343+2delTAAAA	Involved in cells innate response to pathogens; unlikely to be associated with patient phenotype
PDPR	Heterozygous	c.C2542A; p.R848S	387 variants seen at this location in ExAC
SOX8	Compound heterozygous	c.422+5G>C and c.583_584insC; p.H11fs	Top candidate gene
SHANK1	Compound heterozygous	c.C4771G:p.P1591A and c.532-5C>T	Amplification of cDNA showed no splice effect from c.532-5C>T variant

WES, Whole Exome Sequencing; ExAC, Exome Aggregation Consortium

A qPCR was performed (n=2) on patient and control fibroblasts and identified a 3x increase in *SOX8* expression in the patient compared to a control (Figure 3.13a). PCR amplification, of reverse transcribed RNA from patient fibroblasts, using primers flanking the heterozygous c.583_584insC variant revealed only the mutant allele, consistent with the splice mutation causing nonsense mediated decay. A subsequent Western blot showed equal mass (~50 kDa) bands in both the patient and control, with no other bands of any size (Figure 3.13b). Using *ImageJ* software and β-actin as a control, the patient was shown to have 2x more protein than the control. As a size control marker, *SOX8* was overexpressed in human embryonic kidney (HEK) cells and also immunostained on this Western blot resulting in a larger band, of ~55 kDa, than those of the patient and control.

3.3.2 Large-scale Data Sharing

Two international databases have been used for large scale data-sharing, GEM.app and PC, to facilitate the identification of second patients/families with deleterious-appearing variants in the same gene and an overlapping phenotype. Twenty-six patients are unsolved and of these, we have consent for large-scale data sharing for ten of them.

GEM.app: WES data from the ten consented patients, who all had CMT2 diagnoses, were entered into GEM.app. Top novel candidate genes were chosen based on rarity, conservation scores, pathogenicity scores and the tissue types in which the genes were expressed. The GEM.app database was filtered for patients, from other sites, that share rare variants in the same candidate gene. The facilitators of GEM.app were then contacted to determine the patient's clinical diagnosis and if they had been solved (Table 3.4). Genes were excluded

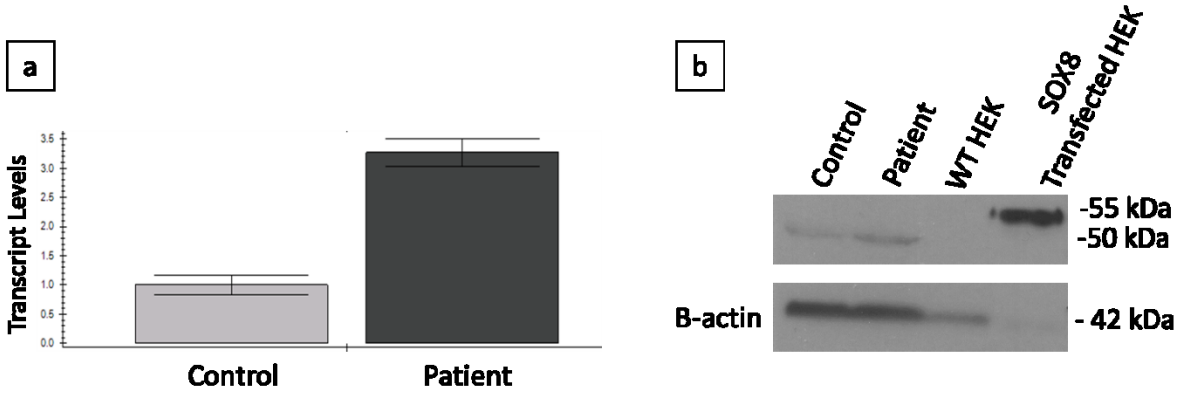


Figure 3.13. a) A qPCR comparing *SOX8* transcript levels in the patient compared to a control. b) A Western blot comparing SOX8 protein levels in the patient compared to controls. *SOX8* was over-expressed in HEK cells and used as a size control marker and test for antibody specificity. B-actin was used as a loading control.

Table 3.4. CMT2 candidate genes and possible matches in GEM.app.

Patient ID	Gene	Disorder of Matched Patient(s)	Reason for Exclusion
145460	*CACFD1*	CMT2; ALS	Matched patients have different phenotypes
156651	*NUBP2*	HSP	Matched patients have different phenotypes
55200	*KIAA1024*	CMT2	Variant too common
57606	*CTTNBP2*	Okinawan proximal HMSN	Matched patients have different phenotypes
54976	*PHKA2*	ALS	Matched patients have different phenotypes
54976	*NDFIP1*	Ataxia	Matched patients have different phenotypes

CMT2, Charcot Marie Tooth Disease Type 2; ALS, Amyotrophic Lateral Sclerosis; HSP, Hereditary Spastic Paraplegia; HMSN, Hereditary Motor and Sensory Neuropathy

from analysis if the high level clinical description of the patients did not match (e.g. ALS vs CMT vs HSP). No compelling matches were identified to justify further functional studies.

PhenomeCentral: Nine patients with promising novel disease gene candidates have been entered into PC to find gene / patient matches (Table 3.5); a subset of these patients were entered into GEM.app. No matches have been made within PC, but this is not surprising since C4R data encompasses a significant contribution to the data currently in PC and we would know about a match earlier in the informatics pipeline as we filter variants using our existing FORGE and C4R datasets. However, PC is our central repository for all consented data and it now connects to other matchmaking databases, such as GeneMatcher. This will increase the likelihood of further leads for these unsolved patients (see Discussion). One promising match has been made through GeneMatcher in the gene *IARS* for patient 53411; further functional work to support *IARS* as a disease-causing gene is being performed in Calgary.

Table 3.5. Summary of patient data entered into PhenomeCentral.

Patient ID	Phenotype	Gene(s)	Solved* or Candidate(s)
48579	Distal wasting; club feet; fused cervical vertebrae; Age of onset at one year	*BICD2*	Solved
53645	Distal sensory/ motor neuropathy	*DNM2*	Solved
130451	CMT2A; multiple symmetric lipomatosis	*MFN2*	Solved
130105S	CMT2A; multiple symmetric lipomatosis	*MFN2*	Solved
120915T	Intellectual disability; spastic paraplegia; hereditary sensory neuropathy	*KIF1A*	Solved
53473	Distal sensory/ motor neuropathy; gait ataxia; fine postural tremor; visual impairment; onset in 40s	*CHRM5; TRIML2*	Candidates
57606	Axonal sensory/ motor neuropathy; severe; childhood onset	*CTTNBP2*	Candidate
130206E	Distal sensory/ motor neuropathy; atrophy; distal weakness; axonal type	*TMEFF2; PPFIA2; ANKFY1*	Candidates
55200	Neuropathy; No myopathy; Kyphoscoliosis; Childhood onset	*OCSTAMP*	Candidate
156651 & 156652	Profound neuropathy; pain insensitivity; areflexia; hammertoes; severe developmental delay; autistic features	*NUBP2*	Candidate
54976	Hereditary neuropathy	*NDFIP1*	Candidate
53411	Motor/ sensory wasting; Hammer toes; axonal subtype	*IARS*	Candidate
87514	Hereditary sensory autonomic neuropathy; developmental delay	*SUGT1*	Candidate
Subject 16	Distal weakness; pes cavus	*BAG3*	Solved
Subject 19	Mild wasting in shins; pes cavus; hammer toes; superimposed carpal tunnel syndrome	*FBXO38*	Solved
Subject 21	Pes cavus; distal weakness; high IQ; axonal CMT pattern	*SIGMAR1*	Solved
Subject 22	Severe distal wasting; bilateral foot drop; hand wasting; axonal neuropathy	*PRX*	Solved
Subject 25	Numbness and tingling sensation associated with pain in the feet; mild muscle wasting in the leg; Sensory axonal neuropathy	*GARS*	Solved
CH0082	Distal weakness; decreased sensation; wasting in upper and lower limbs; areflexia; childhood onset	*IGHMBP2*	Solved
120879P	Congenital myopathy/ lower motor neuron disease; skeletal dysplasia; contractures; short stature; amenorrhea; possible cataracts	*SOX8*	Candidate

*Included in PC as a control dataset for algorithm improvement

Chapter 4: Discussion

4.1 Using WES as a Diagnostic Tool

A study performed by a CMT clinic in the UK showed that 47% of all patients with a CMT or related disorder have no known molecular diagnosis, and 44% of all cases are caused by 17p duplications, or mutations in *PMP22*, *MPZ*, *GJB1* or *MFN2* (Murphy et al., 2012). The remaining 9% have a molecular diagnosis in known, but less common, CMT and related disorders-causing genes. All patients enrolled in this study were negative for the standard-of-care molecular testing available to them, which included assessment for 17p duplications and sequencing for mutations in *PMP22*, *MPZ*, *GJB1* and *MFN2*, as indicated, suggesting that these patients have either a mutation in a known but less common gene (9% of CMT patients), or a novel disease gene (47% of CMT patients), thus representing 56% of all CMT patients. In Specific Aim# 1 of this study, I determined the efficacy of using WES as a diagnostic tool for patients that corresponded to the 9% with mutations in less common genes. I expected a diagnostic rate of ~16% (9 / 56) but a molecular diagnosis was found for 7/37 (~19%) of patients. Diagnoses were made for variants in the genes: *BICD2, DNM2, BAG3, FBXO38, GARS, SIGMAR1* and *PRX*. The concordance between the expected and actual diagnostic rates provides evidence that WES is a useful diagnostic tool for this subset of CMT patients.

4.1.1 The Diagnostic Utility of WES vs Gene Panels

Before WES can be adopted into clinical practice, a cost analysis must be performed to determine if the benefits outweigh the costs. GeneDx (http://www.genedx.com/) currently offers panel testing for CMT. Duplication testing of 17p costs $500 USD, and from there

panels specific to axonal or demyelinating CMT, as well as HMN panels, range from $4,200 to $4,780 USD. An analysis of the variants and clinical reporting is included in that cost. In 2011, Baylor College of Medicine (https://www.bcm.edu/) started offering clinical WES for pediatric patients at a cost of $9,000 USD, including analysis. WES, in a research setting (primarily consumables), costs ~$1,000 CDN and that number continues to drop as sequencing technology improves. The future clinical approach to CMT may include a combination of panels and WES, for example a CMT1 panel for those with a demyelinating presentation of CMT, followed by WES if the panel is negative. For patients with an axonal presentation of CMT, where 75% of cases are without a molecular diagnosis, perhaps WES is a better first line approach. A cost analysis will partly inform whether WES may eventually replace panels, or if there will be a combination of panels and WES used, and tailored to clinical indication.

Even though panels to test for CMT are becoming more encompassing, it seems unlikely that they will ever cover all the genes that cause CMT and related disorders given the rate at which they are being discovered. WES enables researchers and clinicians to assess for mutations found in the entire coding genome. At present, there are 88 genes known to cause CMT and related disorders (Rossor et al., 2015), and even if there was a panel containing all of these genes, molecular diagnoses will still be missed. CMT is associated with significant clinical heterogeneity with most subtypes having overlapping clinical features. For instance, two patients that were enrolled in this study with a diagnosis of CMT, Subject 16 and Subject 21, were found to have mutations in *BAG3* and *SIGMAR1*, respectively, genes that are not found on the CMT and related disorders gene list as they cause similar, but distinct disorders (Myofibillar myopathy 6 and ALS16, respectively). In keeping with this finding, a recent review showed that ~25% of patients that are undiagnosed

on a molecular level have an atypical clinical presentation that was unrecognizable to the clinician (Sawyer et al., 2015b). The results of this study have shown that WES is a valuable diagnostic tool as we are able to not only find a diagnosis for ~19% of CMT patients, but were also able to establish a molecular diagnosis for patients in genes that were not even initially considered, highlighting the advantages of an unbiased approach.

4.1.2 Limitations of WES

There are current limitations to WES technologies which can result in failure to diagnosis a patient at the molecular level (Summarized in (Biesecker and Green, 2014)). On average, WES provides adequate coverage for 85 to 95% of the exome. The remaining portion is not covered sufficiently to make an accurate call regarding the presence or absence of a variant. Variant types poorly or not at all detected by WES including those found in repetitive DNA, copy-number variants (CNVs), insertions or deletions larger than ten nucleotides, translocations, aneuploidy and epigenetic alterations. There may also be disease-causing mutations found in introns and non-coding regions that are not assessed using WES. Limitations, such as the interpretation of variation on a large-scale such as generated by WES, can be improved over time, through broad national and international data-sharing initiatives. Also, as sequencing technology improves, the average coverage, particularly of disease-relevant genes, obtained through WES will also improve. There will be, however, some cases where WES will never be able to make a diagnosis and other options, such as WGS, will need to be utilized.

4.2 Novel Research Discoveries Using WES

This study has highlighted WES as a useful diagnostic tool but the advantage of subsequent use of the data for related research cannot be understated. Gene panels result in two scenarios: a molecular diagnosis or not. In contrast, WES can provide a molecular diagnosis, but in the case that one is not identified, WES enables future research to be conducted to, one hopes, ultimately find that elusive diagnosis. In addition, as highlighted by Patients 130451, 130105S, 120915T and CH0082, new disease associations can be identified for known disease-causing genes, thus WES enables new discoveries that the current standard-of-care molecular testing cannot. Through data-sharing, WES also enables novel gene discoveries and aids in future research by contributing to minor allele frequencies that are used to filter data and identify candidate mutations. Therefore, the advantages of WES when used as the diagnostic approach are significant and need to be considered when selecting the genetic investigations for any given patient.

4.2.1 Multiple Symmetric Lipomatosis is Caused by Mutations in MFN2

MFN2 plays a role in the formation and maintenance of mitochondrial filaments and networks by regulating mitochondrial fusion (Santel and Fuller, 2000). Homozygous *Mfn2* knockout mice die in mid-gestation, but mouse embryonic fibroblast cultures from *Mfn2*-deficient mice show reduced mitochondrial mobility (Chen et al., 2003a), an element that is important to the functional health of axons in the peripheral nervous system. Mutations in *MFN2* are known to cause CMT2A (Zuchner et al., 2004) by affecting mitochondrial transport and network formation.

MSL is a rare disorder of brown fat, and a mitochondrial etiology was first implicated in 1991 when ragged red fibers (RRF) were identified on muscle biopsies of patients with

this clinical condition (Berkovic et al., 1991). Subsequently, the mitochondrial mutation, m.8344A>G, typically associated with myoclonic epilepsy associated with ragged-red fibers (MERRF, MIM# 545000), was also associated with MSL (Klopstock et al., 1997). Our patient was tested and found to be negative for known MERRF mutations, suggesting another genetic etiology for his clinical presentation.

I identified a homozygous *MFN2* mutation, p.R707W, in two siblings that presented with MSL and an axonal form of peripheral neuropathy. MFN2 is localized to the outer membrane of mitochondria, and encodes a GTPase of the dynamin family that mediates the homotypic interaction and fusion of mitochondria (Chen et al., 2003b). Using RNA interference to knock down MFN2 and also by expressing MFN2 with the p.R707W mutation in fibroblasts and comparing them to overexpressed wild-type MFN2 fibroblasts, a recent study on these patients showed that the mitochondrial capacity to form networks by tubulating is reduced, and also, that the mitochondria were more likely to aggregate in a perinuclear fashion when mutated (Sawyer et al., 2015a). The p.R707W mutation is located within the heptad repeat 2 (HR2) domain of MFN2, the domain that was thought to allow homotypic interactions and heterotypic interactions with MFN1. Our collaborators determined that the p.R707W mutation prevented MFN2 from forming homo-oligomers, but it did not prevent MFN2 and its homolog, MFN1, from forming hetero-oligomers (Sawyer et al., 2015a). It has been proposed that MFN1 is able to rescue the disease phenotype in most tissues in the body (Detmer and Chan, 2007), but MFN1 is nearly undetectable in the brown fat, the tissue implicated in the pathogenesis of MSL. This data supports that *MFN2* can lead to mitochondrial dysfunction in brown fat (Plummer et al., 2013), and that the homozygous p.R707W mutation can cause an axonal peripheral neuropathy in association with MSL, thus identifying a novel cause of MSL.

4.2.2 Novel Disease Associated with Mutations in the Motor Domain of KIF1A

KIF1A is a neuron-specific motor protein and is part of the Kinesin family of proteins (KIFs); microtubule-dependent motor proteins that play an important role in neuron function by transporting cellular proteins, macromolecules and organelles along neurites (Hirokawa and Tanaka, 2015). KIF1A is composed of an N-terminal motor domain, a forkhead associated domain (FH) and a pleckstrin homology domain (PH) (Shin et al., 2003) and is responsible for the transport of synaptic vesicle precursors and postsynaptic proteins along axons (Okada et al., 1995; Shin et al., 2003).

Previous reports have associated mutations in KIF1A with multiple different neurodegenerative disorders. Both homozygous and compound heterozygous truncating mutations that are located downstream of the N-terminal motor domain have been shown to cause hereditary sensory neuropathy type IIC (HSN2C2; MIM# 614213) (Riviere et al., 2011). Recessive mutations located at the distal part of the N-terminal motor domain of KIF1A are known to cause hereditary spastic paraparesis (HSP)/ spastic paraplegia type 30 (SPG30; MIM# 610357) (Erlich et al., 2011). Certain mutations in KIF1A have also been reported to cause dominant ID and cerebellar atrophy (Hamdan et al., 2011). I identified a *de novo* heterozygous variant (p.V144F), which was predicted to be pathogenic in the N-terminal motor domain of KIF1A in monozygotic twin boys that causes a phenotype characterized by all three of HSN, HSP and ID.

Our patients were added to a cohort of 12 other patients, all of whom also had *de novo* mutations in the motor domain of *KIF1A* (Lee et al., 2015). The mutations in KIF1A in the 14 individuals were: p.S58L, p.T99M, p.G102D, p.V144F, p.R167C, p.A202P, p.S215R, p.R216P, p.L249Q, p.E253K and p.R316W, and all had SIFT and Polyphen-2 scores equal to 1. After ATP hydrolysis in the motor domain of KIF1A, the ɣ-phosphate is released through

a specialized structure called the back door (Nitta et al., 2004). Our collaborators showed, through modelling, that the mutation in our patients, p.V144F, is likely to disrupt a nearby helical structure, converting it to a loop structure, which could affect ATP binding or ɣ-phosphate release (Lee et al., 2015). Lee *et al*. (2015) further characterized the impact of mutations in the N-terminal motor domain of KIF1A on transport by visualizing EGFP-tagged motor domain constructs of p.T99M, p.A202P, p.S215R, p.R216P and p.E253K KIF1A proteins expressed in cultured hippocampal neurons (Lee et al., 2015). KIF1A in wild-type cells accumulated in the distal regions of the neurites while those with point mutations were located mostly in the cell bodies with dramatically less peripheral accumulation (Lee et al., 2015). These data support that mutations in the motor domain of *KIF1A* cause a novel dominant disease characterized by HSN, HSP and ID.

4.2.3 IGHMBP2, a Novel CMT-Causing Gene

IGHMBP2 mutations have been classically linked to SMARD1 with severe respiratory distress and neonatal death (Grohmann et al., 2001). Less severe phenotypes were then reported with later-onset respiratory symptoms (Messina et al., 2012) and/or survival into adolescence / adult life (Eckart et al., 2012; Pierson et al., 2011). Most recently, our patients, and those from two recent reports, have linked *IGHMBP2* to a classic axonal CMT phenotype without respiratory involvement (Cottenie et al., 2014; Schottmann et al., 2015). Missense and truncating mutations have been identified throughout the gene and have been linked to both SMARD1 and CMT (Cottenie et al., 2014; Grohmann et al., 2003) and the genotype-phenotype correlation is unclear. Mutations resulting in the milder CMT presentation have also been linked to the severe SMARD1 end of the clinical spectrum, though the specific combination of genotypes have differed. For example, homozygous

IGHMBP2 mutations (c.1488C>A; p.Cys496*) (Grohmann et al., 2003) have been linked to a severe SMARD1 presentation, while compound heterozygous *IGHMBP2* mutations (c.1488C>A; p.Cys496* + c.238A>G; p.Ser80Gly) with only the CMT presentation (Cottenie et al., 2014). In this example the missense substitution resulted in the milder phenotype and yet there are several examples where homozygous missense and homozygous stops result in both CMT and SMARD1 (Cottenie et al., 2014; Grohmann et al., 2003). Thus, further work is required to understand how these mutations are impacting IGHMBP2 function.

Our functional studies suggest that the genotype-phenotype correlation may be related, at least in part, to the amount of residual IGHMBP2 protein. Our patient (c.2601_2604del; p.Lys868Profs*109) had 40% IGHMBP2 protein compared to a control, which is predicted to be a result of nonsense-mediated decay. This finding is consistent with a recent report that showed overall significantly higher protein levels in IGHMBP2-related CMT compared to the classic SMARD1 phenotype in five individuals from three families (Cottenie et al., 2014). Cottenie *et al.* (2014) suggested that residual truncated protein may result in a CMT presentation, rather than SMARD1 (Cottenie et al., 2014). The same group also identified a 70 kDa fragment on Western blots of the CMT patients with specific mutations (c.138T>A; p.Cys46* + c.2911_2912delAG; p.Arg971Glufs*4) but not in controls or SMARD1 patients. They suggested that lower levels of functioning protein could activate a feedback mechanism that preserved residual truncated protein (Cottenie et al., 2014). However, using the same antibody, we also identified the 70kDa fragment in both our healthy control and patient lymphocyte cell lines, indicating that the 70kDa protein fragment does not easily explain the genotype-phenotype correlation and the 70kDa fragment is not likely the residual and truncated protein. Of note, the observation of higher residual

IGHMBP2 protein levels associated with the CMT presentation was not supported in the recent work by Schottmann *et al*. (2015), where no residual IGHMBP2 protein was observed in their patient's fibroblasts (Schottmann et al., 2015). This was observed in two individuals from one family that were homozygous for an early frameshift mutation (c.449+1G>T; p.Lys150Asnfs*0) (Schottmann et al., 2015). One of the individuals tested had no respiratory involvement despite also having no residual protein as demonstrated by Western blot. Another hypothesis is that alternate transcripts of *IGHMBP2,* that do not contain the disease-causing mutations, may exist and in part "rescue" the phenotype. A recent study proposed that exon skipping, or alternative splicing, may be a contributing factor in genetic pleiotropy (mutations in the same gene resulting in different phenotypes) for multiple diseases (Drivas et al., 2015). While I did show the presence of an alternate transcript (XM_005272976.1), I was not able to show a protein associated with this transcript from cultured lymphocytes, but my experiment was of course limited by available tissue and time-point. Modifying genes may also play an important role in *IGHMBP2* expression and protein function. For example, variable expression has been reported in a sibling pair where one sibling had SMARD1 while the other had CMT with no respiratory insufficiency (Schottmann et al., 2015). Each sibling was homozygous for c.449+1G>T; p.Lys150Asnfs*0. Studies have demonstrated that modifiers on mouse chromosomes 9, 10, and 16 were able to reduce the severity of symptoms caused by a mutated *Ighmbp2* and increase longevity in mice (Maddatu et al., 2005). It is possible that individual mutations may make specific genotypes more amenable to "rescue" by the effects of homologous isoforms or genetic modifiers allowing the clinical presentation of SMARD1 to become the less severe CMT in such cases.

Our study demonstrates that a novel homozygous frameshift mutation in *IGHMBP2* causes a CMT phenotype. Our data supports the finding of Cottenie *et al.* (2014) that residual protein level is associated with CMT rather than SMARD1 with respiratory impairment. It is also possible that modifier genes or alternative isoforms of *IGHMBP2* may contribute to the milder clinical presentation of CMT. There is a need for a greater understanding of disease mechanism for this spectrum of conditions; such insight could have significant therapeutic implications for patients suffering from these diseases.

4.2.4 SOX8 as a Candidate Gene for Syndromic Lower Motor Neuron Disease

The gene, *SOX8*, is in the Sox protein family of transcription factors, and more specifically, belongs to the subgroup E, along with *SOX9* and *SOX10*. SOX8 protein shares 47% amino acid similarity with the other SoxE members and is identical to the mouse Sox8 protein, except for one amino acid. SoxE members are characterized by their shared 79 amino acid high-mobility-group (HMG) box domain which docks to the minor groove of specific DNA sequences. These proteins are expressed in many cell types throughout the body and direct many processes including gliogenesis, sex determination and neural crest development (Barrionuevo and Scherer, 2010; Haldin and LaBonne, 2010; Stolt and Wegner, 2010). The neural crest develops into craniofacial cartilage and bone, smooth muscle and peripheral neurons, so mutations in SoxE genes may have effects on many different tissue types.

Mutations in SoxE genes, specifically *SOX9* and *SOX10*, have been associated with different disorders. Mutations in *SOX9* are known to cause campomelic dysplasia (MIM# 114290) which is an autosomal dominant disorder causing primarily bone malformations, but can also include hearing loss, developmental delay and male-to-female sex reversal (Foster et al., 1994). *SOX10* mutations are associated with an autosomal dominant peripheral

demyelinating neuropathy, central dysmyelination, Waardenburg syndrome features and Hirschsprung disease (PCWH; MIM# 609136) (Inoue et al., 2004). A 2 Mb deletion, which encompassed the *SOX8* gene, was shown to cause alpha-thalassemia/ mental retardation syndrome, type 1 (ATR-16 syndrome; MIM# 141750), which is primarily characterized by ID (Pfeifer et al., 2000). Further studies investigated the function of *SOX8* by generating Sox8-deficient mice (Sock et al., 2001). The mice were viable and the only recognizable phenotype different from wild-type mice was a substantial reduction in weight. It was predicted that the functional redundancy between SoxE proteins was able to rescue the Sox8-deficient mice from a more severe phenotype.

This study presented a patient with congenital myopathy and lower motor neuron disease in addition to skeletal dysplasia, contractures, short stature, amenorrhea and possible cataracts. WES revealed multiple heterozygous variants (c.422+5G>C and c.583_584insC; p.H11fs) in *SOX8*. *SOX8* is a good candidate for this rare clinical presentation based on disorders associated with other SoxE genes, what is understood about the function of *SOX8*, and the nature of the mutations. Mutations in *SOX9* and *SOX10* result in bone abnormalities and neuropathies, defects with the same tissue distribution as this patient. *SOX8* is also a good candidate due to its function in directing the neural crest to develop into smooth muscle, bone and parts of the nervous system. Also, given that *SOX8* is composed of only three exons, a c.422+5G>C variant that is predicted to affect splicing through the loss of a splice donor site could have detrimental effects on the remaining protein by removing, or adding, an exon. Finally, the frameshift insertion seen in our patient, p.H11fs, is predicted to truncate the protein from its wild-type size mass of 47 kDa to less than half its mass at 22 kDa.

Functional studies, qPCR and Western blot, were used to provide evidence of the pathogenicity of the variants seen in our patient. The qPCR resulted in a surprisingly 3x increase in transcript levels in the patient compared to a wild-type control. Perhaps a reduction in SOX8 protein results in concomitant upregulation of its transcript. The amplification of only the frameshift insertion allele from patient cDNA suggests that the mRNA containing the splice variant is being degraded through non-sense mediated decay, while the mRNA containing the frameshift allele is overexpressed. A Western blot was performed to determine if increased *SOX8* transcript levels translated into a change in protein levels. The patient did have 2x more protein than control, but the size of the protein was not as expected. The Western blot showed a band with a mass of ~50 kDa in both the patient and control with no other bands of any size. If the mRNA containing the splice variant was degraded, only the truncated mRNA would be translated, resulting in a protein half the size of the WT. However, even the band seen in the control did not match the expected size of the protein identified in the lane containing protein from *SOX8* over-expressed HEK cells. The size difference suggests that the antibody in use may not be specific to SOX8 and instead, may be recognizing SOX9 or SOX10, which are similar in size and share homology with SOX8. A second antibody was used (Abnova H00030812-M01; Taipei City, Taiwan) but was unable to identify any protein from the *SOX8* over-expressed HEK cells, so this antibody was also judged to be not useful.

WES revealed additional multiple heterozygous candidates in the gene, *SHANK1*, for this patient. The two variants in the gene were a missense c.C4771G; p.P1591A and one predicted to affect splicing (c.532-5C>T) through the loss of a splice acceptor site. SHANK1 interacts with the protein DNM2 (Okamoto et al., 2001) which, when mutated, can cause congenital contractures, myopathy and neuropathy (Bitoun et al., 2005; Koutsopoulos

et al., 2013; Zuchner et al., 2005). A series of primers were used to amplify cDNA at various points on either side of the c.532-5C>T variant to evaluate its potential impact on splicing. Sanger sequencing showed that there was no detectable change in splicing. The other *de novo* variants in the patient, found in *MUC12*, *RIPK2* and *PDPR*, were excluded as candidate genes as their function was not compelling as causative of the disease phenotype, or in the case of *PDPR*, the variant has been reported ~400 times in ExAC. Thus, only one gene, *SOX8*, remains as a good candidate to cause this patient's phenotype. At this time, however, there are limitations to the functional investigations that can be performed, due to the lack of quality, commercially available antibodies specific to SOX8, to further support it as disease-causing. Identification of an additional and unrelated patient with an overlapping phenotype and recessive deleterious-appearing variants in this gene would provide much needed additional evidence and enable discovery of *SOX8* as a novel disease gene.

4.3 Functional Studies vs Large-scale Data Sharing

For a gene to be defined as novel disease-causing, either functional studies must provide substantial evidence supporting variant pathogenicity and impact on a biological pathway sufficient to cause the disease phenotype, or, ideally and the current gold-standard, through the identification of a second family with the same disorder and mutated gene. Each method has its strengths, but also its limitations. For example, some genes are highly variable, with the majority of variants being common SNPs and presumed not having any detrimental effect on the protein's function. Therefore, sharing of such variants amongst individuals with the same disorder, does not implicate them as disease-causing. Functional data must also be interpreted correctly, and with any *in vitro* assay, the question remains whether the results can be trusted as reproducible, or applicable to the *in vivo* biological system. Begley and

Ellis (2015) performed an analysis of the reproducibility of landmark papers in cancer research, and determined that the findings in only 11% (6/53) have been confirmed (Begley and Ellis, 2012).

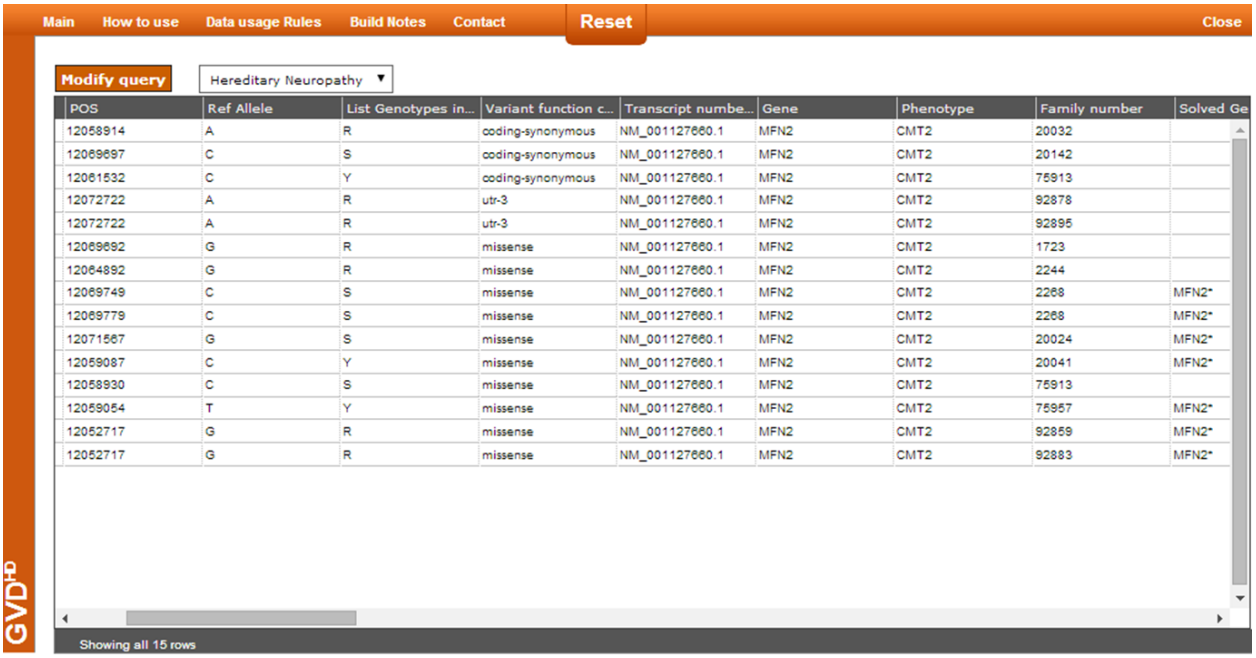
Western blots are one of the most common functional assays performed to determine whether a variant is impacting the protein, but recent studies suggest that the majority of antibodies are either nonspecific or are very specific but to the wrong protein. Michel *et al*. (2009) tested the specificity of 49 commercially available antibodies that targeted G-protein-coupled receptors (GPCRs) and found that most bound to more than one protein (Michel et al., 2009). Much like the SoxE family of proteins, GPCRs share large sequences of homology potentially making it difficult to design an antibody specific to just one. An analysis of 246 antibodies specific to histone-modifiers showed that at least 25% were nonspecific while four antibodies were specific but to the wrong protein (Egelhofer et al., 2011). There are many antibody-dependent assays, Western blots only being one that utilize / rely on protein specific antibodies for results. For example, another study looked at ~20,000 commercial antibodies and determined that only 50% were able to effectively identify protein distribution in preserved slices of tissue (Berglund et al., 2008). Clearly, functional insight into novel candidate disease genes is limited by the current tools and resources available for such studies.

Given the challenges outlined above, it is not surprising that functional studies require a large investment of both time and money. In the investigation of rare diseases, many gene candidates have not been widely studied given that much effort and resources must be spent validating antibodies and assays before research on the variant of interest can be conducted. Therefore, the emerging trend is to put more emphasis on large-scale data sharing to identify candidate genes that are shared amongst families, prior to investing

heavily in functional studies. Databases must be able to effectively manage large amounts of sequencing data and allow user friendly ways to search and interpret the variants. Two databases currently exist that contain suitable unsolved exomes for matching with the unsolved patients from this study, GEM.app and PC, and I highlight their strengths and weaknesses in the following sections.

4.4 Utility of GEnomes Management Application (GEM.app)

GEnomes Management Application is a software tool for large-scale collaborative genome / exome analysis (Gonzalez et al., 2013b). GEM.app offers researchers a user-friendly tool to annotate, manage, visualize and analyze large datasets. VCF files are uploaded to the database, and the GEDI (Gaffa evented data interface) module annotates all the variants, segregates variants within families, and performs minor allele frequency calculations. GEDI annotates the variants by using the SeattleSeq annotation server (http://snp.gs.washington.edu/), which includes statistical evaluations such as GERP, Phastcons and Polyphen-2, as well obtaining variant frequency data from dbSNP, EVS, OMIM, ExAC and many others. The affection status, individual ID, family ID and predicted inheritance patterns are imported along with the VCF file, to aid in variant filtering. After each new VCF file is uploaded, the system re-calculates allele frequencies so the filters are always up-to-date. The interface allows many filtering options and ways to search the database, such as genomic positions, conservation scores, quality scores, inheritance patterns or for known disease-causing genes for a specific rare disease. The graphical interface has a row for each variant, with sortable columns, so specific information can be easily targeted (Figure 4.1). The search engine performs at a high level with all queries typically finishing in a matter of seconds. GEM.app has proven to be a successful tool as multiple novel



POS	Ref Allele	List Genotypes in...	Variant function c...	Transcript numbe...	Gene	Phenotype	Family number	Solved Ge
12058914	A	R	coding-synonymous	NM_001127660.1	MFN2	CMT2	20032	
12069697	C	S	coding-synonymous	NM_001127660.1	MFN2	CMT2	20142	
12061532	C	Y	coding-synonymous	NM_001127660.1	MFN2	CMT2	75913	
12072722	A	R	utr-3	NM_001127660.1	MFN2	CMT2	92878	
12072722	A	R	utr-3	NM_001127660.1	MFN2	CMT2	92895	
12069692	G	R	missense	NM_001127660.1	MFN2	CMT2	1723	
12064892	G	R	missense	NM_001127660.1	MFN2	CMT2	2244	
12069749	C	S	missense	NM_001127660.1	MFN2	CMT2	2268	MFN2*
12069779	C	S	missense	NM_001127660.1	MFN2	CMT2	2268	MFN2*
12071567	G	S	missense	NM_001127660.1	MFN2	CMT2	20024	MFN2*
12059087	C	Y	missense	NM_001127660.1	MFN2	CMT2	20041	MFN2*
12058930	C	S	missense	NM_001127660.1	MFN2	CMT2	75913	
12059054	T	Y	missense	NM_001127660.1	MFN2	CMT2	75957	MFN2*
12052717	G	R	missense	NM_001127660.1	MFN2	CMT2	92859	MFN2*
12052717	G	R	missense	NM_001127660.1	MFN2	CMT2	92883	MFN2*

Figure 4.1. A screen shot of the online interface of the gene search function in GEM.app.

disease-causing genes have already been identified through the database (Gonzalez et al., 2013a; Martin et al., 2013; Montenegro et al., 2012).

There are many benefits to using GEM.app that can benefit rare disease research. One significant strength is the number of exomes that are in the database. Currently, there are >4000 exomes and that number continues to grow as more researchers start sharing their data. Therefore, GEM.app has become a great tool for the study of rare neuromuscular disorders, such as HSP and CMT, as the majority of the exomes in the database are from patients diagnosed with neuromuscular disorders. However, the database is not only for neuromuscular disorders as clinically relevant variants have been identified for inherited forms of deafness and dilated cardiomyopathies (Diaz-Horta et al., 2012; Norton et al., 2012). A large number of exomes also allows for more accurate minor allele frequencies and more controls to filter out variants. Additional strengths include the database's simplicity, speed of performance and the number of analysis that are performed on the variants. It is relatively easy to filter through large amounts of data in a matter of seconds, as the filter criteria is presented in a step-by-step format. Follow-up of candidate variants is also made easier through GEM.app, as links specific for each variant are found for OMIM, UCSC genome browser (https://genome.ucsc.edu/), NCBI (http://www.ncbi.nlm.nih.gov/) and others.

Though the database is continuing to be updated with new and improved features, there are currently some weaknesses. If a variant in a gene of interest is identified, the next step is to contact the clinician of that patient to determine if the patients have similar clinical presentations. This is performed by contacting the facilitators of GEM.app, who act as a liaison between the two parties. This process can be cumbersome and a direct link to clinician contact would make the process much more efficient. Secondly, there are >4000

exomes in the database, but not all of that data is visible. Restrictions on data sharing capabilities from certain REBs have made some data private. The variants are still used to calculate minor allele frequencies, but cannot be seen by other researchers trying to find shared variants in candidate genes. In addition, there is lack of updated patient information from each clinician. For example, when all the variants in a specific gene are called-up, there is a column that identifies if that patient has been solved, and what gene has been identified to cause their disorder. On multiple occasions, clinicians have been contacted about a potential gene of interest to this study and they report back that their patient has already been solved with a different gene. Time and effort could be saved if clinicians kept their submissions in the database up-to-date. Lastly, the greatest weakness of the database is the lack of specific phenotypic information. Patients are presented in the database with a high-level clinical diagnosis, such as CMT2 or HSP, but no other information. Neuromuscular disorders are highly-heterogeneous and can be associated with other clinical features. In addition, NCV results would be very useful as these distinguish the type of CMT. However, there is a need for a standard method of reporting NCV values before this can be incorporated into the database. Standard terminology does exist for reporting most other clinical features (Human Phenotype Ontology (HPO); http://human-phenotype-ontology.org/ (Kohler et al., 2014)), but GEM.app is not designed to receive such data at this time.

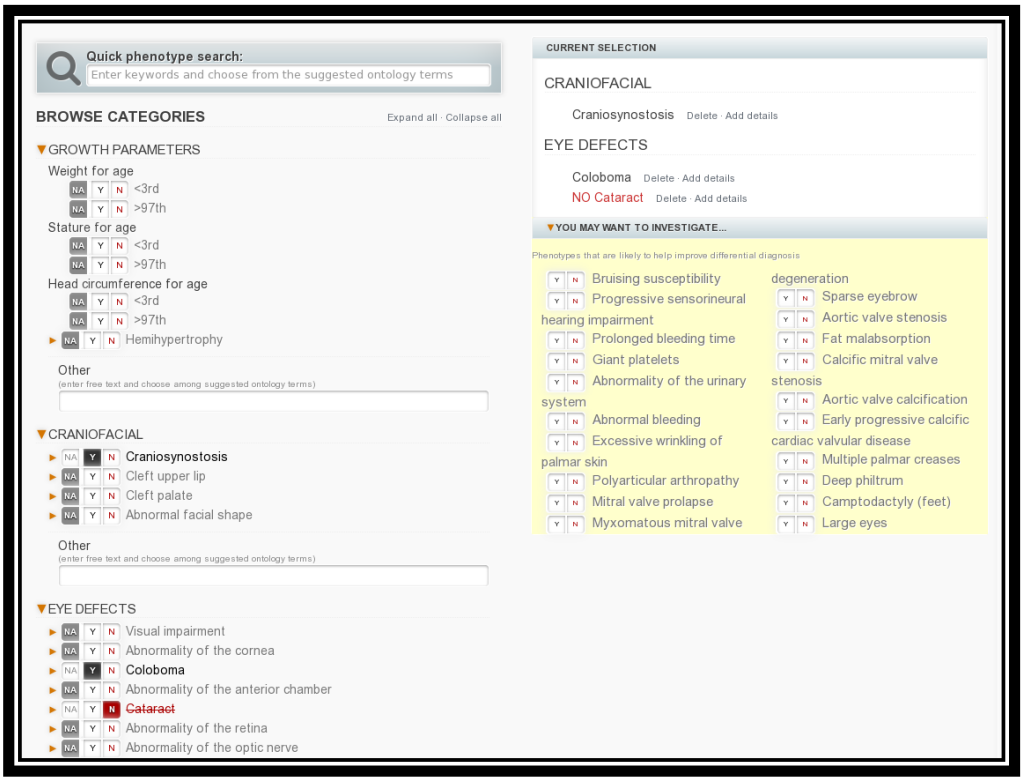
GEM.app has a proven track-record of success which has been built on their high-performance processing and user-friendly interface. However, there is significant room for improvement by streamlining the process of contacting clinicians, and by adding comprehensive clinical features using standard terminology for the patients in the database. As C4R already has in place a system of annotating and filtering through variants, I have

found it unnecessary to upload all patient VCF files into GEM.app, but instead just use its search functions to identify other patients that share variants in genes of interest. To date, no additional patients have been identified using GEM.app with variants in the same gene and with a similar clinical presentation; however, I am confident this will change as the database continues to expand.

4.5 Utility of PhenomeCentral (PC)

PhenomeCentral, a repository for secure data sharing, was developed to encourage scientific collaboration and allow clinicians and scientists to find other patients with similar disease manifestations as patients of their own. Registered users are able to contribute new cases, discover other patients, establish collaborations and share scientific information while protecting the privacy of their patients. PhenomeCentral relies on PhenoTips software (http://phenotips.org/), a free web-based application tool for collecting and analyzing phenotypic information for patients with genetic disorders, and uses standardized vocabulary defined by HPO, to enable relevant and uniformly reported patient descriptions (Figure 4.2a). Similar cases are automatically reported, on the web interface, for clinicians to review and follow-up (Figure 4.2b). The system also allows users to directly initiate contact without disclosing any actual contact information. On top of matching based on phenotype, PC also matches based on genotype. Top gene candidates that have been identified by WES or other methods can be manually entered into the patient's profile. VCF files can also be uploaded to PC, where an analysis is performed by Exomiser (https://www.sanger.ac.uk/resources/software/exomiser/), an online tool that annotates variants obtained from WES and prioritizes them based on pathogenicity, quality, inheritance pattern and model organism data. PhenomeCentral also matches patients based on top

a



b

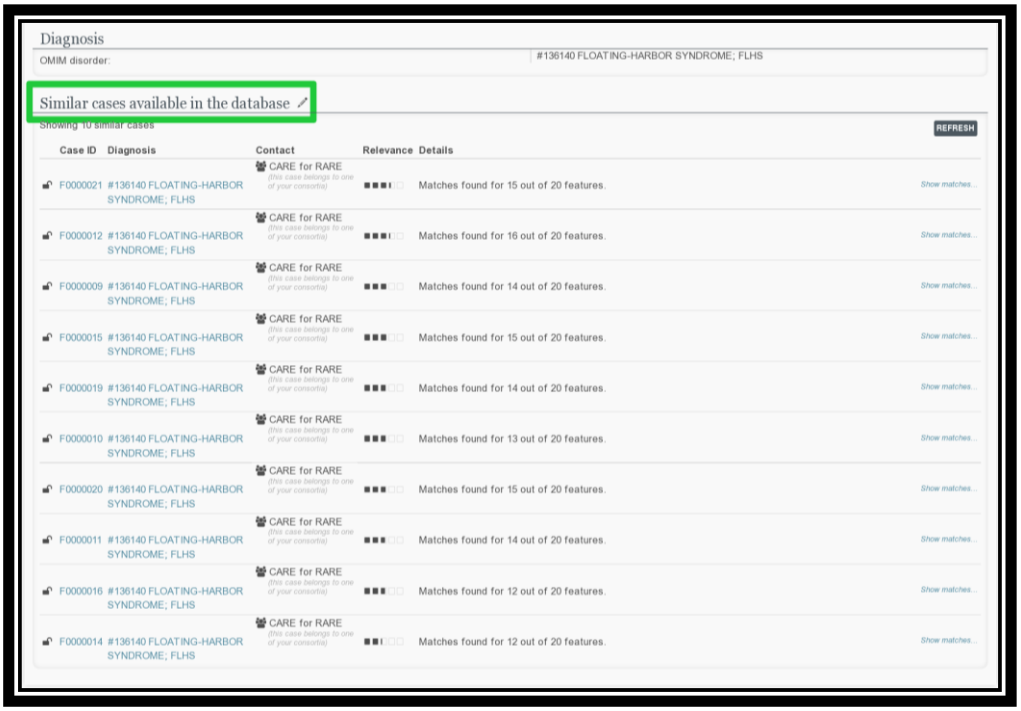


Figure 4.2. Sample screen shots from the PhenomeCentral database. a) A sample interface for entering the clinical presentation of a patient using standardized Human Phenotype Ontology (HPO) terms. b) A sample interface showing how PhenomeCentral matches patients with similar phenotypes and reports it for further follow-up.

results reported by Exomiser. In addition, PC connects with GeneMatcher (GM; https://genematcher.org/) via an application programming interface. GeneMatcher is a database where novel candidate genes have been entered for various disorders whereby the software notifies clinicians if there is a gene match.

PhenomeCentral has many strengths and only a few weaknesses. One of the main strengths is the detailed and searchable clinical information that can be captured in the database. In-depth family history can also be uploaded to the database with easy visualization, such as pedigree information. The standardized vocabulary, using HPO terms, allows matching to be as precise as possible. The matching algorithm is quite innovative as it ranks HPO terms based on similarity and rarity of the phenotype. Clinicians have full control of who views their data, to protect patient privacy, as they can set it as "Public" (viewable and matchable by all other PC users), "Matchable" (but not viewable) or "Private" (only the clinician and manually selected collaborators can view data) based on the type of consent obtained from the patient. One of the weaknesses of PC is in the utilization of the program Exomiser to prioritize candidate genes. Currently, Exomiser does not filter out variants based on minor allele frequency so, often, genes with common variants are highly ranked. This problem will be solved with implementation of future updates of Exomiser. One other weakness of PC is the smaller number of patients that have been entered into the database; currently <1,600. However, the number continues to grow as more clinicians and scientists around the world enter into collaborations with PC.

Patient consent must be obtained before uploading detailed patient information into PC. Some of my patients were consented to undergo WES before there was a push for large-scale sharing of this type of data. As a result, I am only permitted to describe these patients with their high level clinical presentation (e.g. CMT) and list of candidate genes. I am not

able to upload VCF files or share information about specific variants in PC. For this subset of patients, only those for whom I have identified top gene candidates have been entered, but no matches have been made to date, however I am confident this will change as the database continues to expand.

4.6 Final Conclusions

I hypothesized that WES could be used to understand the molecular etiology of rare disorders characterized by lower motor neuron disease or peripheral neuropathy. This hypothesis was evaluated with the following specific aims:

4.6.1 Specific Aim 1: Evaluate the diagnostic utility of WES with respect to lower motor neuron disease / peripheral neuropathy

WES data was analyzed from a cohort of 37 patients diagnosed with lower motor neuron disease or peripheral neuropathy, but were negative for the standard-of-care molecular testing available to them (a summary of the entire cohort can be found in Appendix E). A molecular diagnosis was found for seven patients in the genes *BICD2, DNM2, BAG3, FBXO38, GARS, SIGMAR1* and *PRX*. This diagnostic solve rate of ~19% (7/37) was as expected, based on previous studies, suggesting that WES is a valuable diagnostic tool.

4.6.2 Specific Aim 2: Identify novel diseases for known lower motor neuron / peripheral neuropathy causing genes

For 30 patients, after the initial analysis of the WES data did not identify a molecular diagnosis, data was re-analyzed to identify mutations in known disease-causing genes, but where the patient presents with a different disease or significantly expanded phenotype than

that previously reported. The gene *MFN2* had been previously associated with dominant CMT2A, but in two siblings we show that homozygous mutations can cause axonal neuropathy with MSL. *KIF1A* had been associated with ID, HSN and HSP, each as distinct diseases, but this study has demonstrated that they can also be present in the same individual through mutations localized to the motor domain of the protein. *IGHMBP2* had been associated with SMARD1, but this study demonstrated that variants could also cause CMT2, with no respiratory distress.

4.6.3 Specific Aim 3: Identification of novel disease genes for lower motor neuron disease / peripheral neuropathy for further study.

Twenty-six patients remained without a molecular diagnosis after completion of Specific Aims 1 and 2. A strong candidate gene was identified for one of these patients and functional studies were performed on the top candidate gene, *SOX8*, for a patient diagnosed with congenital myopathy, lower motor neuron disease and skeletal dysplasia. The results were inconclusive for definitive support of pathogenicity and highlighted the significant need for large-scale data sharing in rare disease research.

The prioritized candidate genes for the unsolved patients in this study have been entered into two databases, GEM.app and PhenomeCentral, laying the groundwork for the future identification of novel disease-causing genes when additional patients / families are identified with mutations in the same gene and overlapping clinical features. I conducted an evaluation on the strengths and weaknesses of these online data-sharing centres, based on my experience using these databases to enable novel gene discovery for this cohort of patients.

Whole-exome sequencing is a relatively new technology but its popularity is ever-increasing for the investigation of rare diseases. I have shown how useful WES can be, not only in determining the molecular diagnosis of patients with complex and highly heterogeneous disorders, but also in additional applications for novel research discoveries. In a cohort of 37 peripheral neuropathy or LMND patients, seven received a molecular diagnosis, four were identified to have a novel disease in a known disease-causing gene and the groundwork has been laid to solve the remaining patients using large-scale data-sharing. There are plans to transition WES into clinical practice at CHEO but currently, WES is not routinely being used as a diagnostic tool in patient care in Canada. That day may soon be approaching.

References

(2015). PREDICTED: Homo sapiens immunoglobulin mu binding protein 2 (IGHMBP2), transcript variant X5, mRNA (http://www.ncbi.nlm.nih.gov/nuccore/XM_005273976.1: NCBI).

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.

Al-Saif, A., Al-Mohanna, F., and Bohlega, S. (2011). A mutation in sigma-1 receptor causes juvenile amyotrophic lateral sclerosis. Annals of neurology *70*, 913-919.

Ayme, S., Urbero, B., Oziel, D., Lecouturier, E., and Biscarat, A.C. (1998). [Information on rare diseases: the Orphanet project]. Rev Med Interne *19 Suppl 3*, 376S-377S.

Baird, P.A., Anderson, T.W., Newcombe, H.B., and Lowry, R.B. (1988). Genetic disorders in children and young adults: a population study. American journal of human genetics *42*, 677-693.

Barrionuevo, F., and Scherer, G. (2010). SOX E genes: SOX9 and SOX8 in mammalian testis development. Int J Biochem Cell Biol *42*, 433-436.

Begley, C.G., and Ellis, L.M. (2012). Raise standards for prelcinical cancer research. Nature *483*, 531-533.

Berglund, L., Bjorling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szigyarto, C.A., Persson, A., Ottosson, J., Wernerus, H., Nilsson, P.*, et al.* (2008). A genecentric Human Protein Atlas for expression profiles based on antibodies. Mol Cell Proteomics *7*, 2019-2027.

Berkovic, S.F., Andermann, F., Shoubridge, E.A., Carpenter, S., Robitaille, Y., Andermann, E., Melmed, C., and Karpati, G. (1991). Mitochondrial dysfunction in multiple symmetrical lipomatosis. Ann Neurol *29*, 566-569.

Biesecker, L.G., and Green, R.C. (2014). Diagnostic clinical genome and exome sequencing. N Engl J Med *370*, 2418-2425.

Bitoun, M., Maugenre, S., Jeannet, P.Y., Lacene, E., Ferrer, X., Laforet, P., Martin, J.J., Laporte, J., Lochmuller, H., Beggs, A.H., *et al.* (2005). Mutations in dynamin 2 cause dominant centronuclear myopathy. Nature genetics *37*, 1207-1209.

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. Nature reviews Genetics *14*, 681-691.

Carter, C.O. (1977). Monogenic disorders. J Med Genet *14*, 316-320.

Chen, H., Detmer, S.A., Ewald, A.J., Griffin, E.E., Fraser, S.E., and Chan, D.C. (2003a). Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. J Cell Biol *160*, 189-200.

Chen, H., Detmer, S.A., Ewald, A.J., Griffin, E.E., Fraser, S.E., and Chan, D.C. (2003b). Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. J Cell Biol *160*, 189-200.

Cooper, G.M., Stone, E.A., Asimenos, G., Program, N.C.S., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome research *15*, 901-913.

Cottenie, E., Kochanski, A., Jordanova, A., Bansagi, B., Zimon, M., Horga, A., Jaunmuktane, Z., Saveri, P., Rasic, Vedrana M., Baets, J., *et al.* (2014). Truncating and Missense Mutations in IGHMBP2 Cause Charcot-Marie Tooth Disease Type 2. American journal of human genetics *95*, 590-601.

Detmer, S.A., and Chan, D.C. (2007). Complementation between mouse Mfn1 and Mfn2 protects mitochondrial fusion defects caused by CMT2A disease mutations. J Cell Biol *176*, 405-414.

Diaz-Horta, O., Duman, D., Foster, J., 2nd, Sirmaci, A., Gonzalez, M., Mahdieh, N., Fotouhi, N., Bonyadi, M., Cengiz, F.B., Menendez, I., *et al.* (2012). Whole-exome sequencing efficiently detects rare mutations in autosomal recessive nonsyndromic hearing loss. PLoS One *7*, e50628.

Dodge, J.A., Chigladze, T., Donadieu, J., Grossman, Z., Ramos, F., Serlicorni, A., Siderius, L., Stefanidis, C.J., Tasic, V., Valiulis, A., *et al.* (2011). The importance of rare diseases: from the gene to society. Arch Dis Child *96*, 791-792.

Drivas, T.G., Wojno, A.P., Tucker, B.A., Stone, E.M., and Bennett, J. (2015). Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. Sci Transl Med *7*, 291ra297.

Dye, D.E., Brameld, K.J., Maxwell, S., Goldblatt, J., and O'Leary, P. (2011). The impact of single gene and chromosomal disorders on hospital admissions in an adult population. J Community Genet *2*, 81-90.

Eckart, M., Guenther, U.P., Idkowiak, J., Varon, R., Grolle, B., Boffi, P., Van Maldergem, L., Hubner, C., Schuelke, M., and von Au, K. (2012). The natural course of infantile spinal muscular atrophy with respiratory distress type 1 (SMARD1). Pediatrics *129*, e148-156.

Egelhofer, T.A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A.A., Cheung, M.S., Day, D.S., Gadel, S., Gorchakov, A.A.*, et al.* (2011). An assessment of histone-modification antibody quality. Nat Struct Mol Biol *18*, 91-93.

Erlich, Y., Edvardson, S., Hodges, E., Zenvirt, S., Thekkat, P., Shaag, A., Dor, T., Hannon, G.J., and Elpeleg, O. (2011). Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. Genome research *21*, 658-664.

Fabrizi, G.M., Ferrarini, M., Cavallaro, T., Cabrini, I., Cerini, R., Bertolasi, L., and Rizzuto, N. (2007). Two novel mutations in dynamin-2 cause axonal Charcot-Marie-Tooth disease. Neurology *69*, 291-295.

Foster, J.W., Dominguez-Steglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I.D., Goodfellow, P.N.*, et al.* (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature *372*, 525-530.

Gonzalez, M., Nampoothiri, S., Kornblum, C., Oteyza, A.C., Walter, J., Konidari, I., Hulme, W., Speziani, F., Schols, L., Zuchner, S.*, et al.* (2013a). Mutations in phospholipase DDHD2 cause autosomal recessive hereditary spastic paraplegia (SPG54). European journal of human genetics : EJHG *21*, 1214-1218.

Gonzalez, M.A., Lebrigio, R.F., Van Booven, D., Ulloa, R.H., Powell, E., Speziani, F., Tekin, M., Schule, R., and Zuchner, S. (2013b). GEnomes Management Application (GEM.app): a new software tool for large-scale collaborative genome analysis. Human mutation *34*, 842-846.

Grohmann, K., Schuelke, M., Diers, A., Hoffmann, K., Lucke, B., Adams, C., Bertini, E., Leonhardt-Horti, H., Muntoni, F., Ouvrier, R.*, et al.* (2001). Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. Nature genetics *29*, 75-77.

Grohmann, K., Varon, R., Stolz, P., Schuelke, M., Janetzki, C., Bertini, E., Bushby, K., Muntoni, F., Ouvrier, R., Van Maldergem, L.*, et al.* (2003). Infantile spinal muscular atrophy with respiratory distress type 1 (SMARD1). Annals of neurology *54*, 719-724.

Guilbot, A., Williams, A., Ravise, N., Verny, C., Brice, A., Sherman, D.L., Brophy, P.J., LeGuern, E., Delague, V., Bareil, C.*, et al.* (2001). A mutation in periaxin is responsible for CMT4F, an autosomal recessive form of Charcot-Marie-Tooth disease. Human molecular genetics *10*, 415-421.

Haldin, C.E., and LaBonne, C. (2010). SoxE factors as multifunctional neural crest regulatory factors. Int J Biochem Cell Biol *42*, 441-444.

Hamdan, F.F., Gauthier, J., Araki, Y., Lin, D.T., Yoshizawa, Y., Higashi, K., Park, A.R., Spiegelman, D., Dobrzeniecka, S., Piton, A.*, et al.* (2011). Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. American journal of human genetics *88*, 306-316.

Hirokawa, N., and Tanaka, Y. (2015). Kinesin superfamily proteins (KIFs): Various functions and their relevance for important phenomena in life and diseases. Exp Cell Res *334*, 16-25.

Hong, C.S., and Saint-Jeannet, J.P. (2005). Sox proteins and neural crest development. Semin Cell Dev Biol *16*, 694-703.

Inoue, K., Khajavi, M., Ohyama, T., Hirabayashi, S., Wilson, J., Reggin, J.D., Mancias, P., Butler, I.J., Wilkinson, M.F., Wegner, M.*, et al.* (2004). Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. Nature genetics *36*, 361-369.

Ionasescu, V., Searby, C., Sheffield, V.C., Roklina, T., Nishimura, D., and Ionasescu, R. (1996). Autosomal dominant Charcot-Marie-Tooth axonal neuropathy mapped on chromosome 7p (CMT2D). Human molecular genetics *5*, 1373-1375.

Klopstock, T., Naumann, M., Seibel, P., Shalke, B., Reiners, K., and Reichmann, H. (1997). Mitochondrial DNA mutations in multiple symmetric lipomatosis. Mol Cell Biochem *174*, 271-275.

Kohler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C., Brown, D.L., Brudno, M., Campbell, J.*, et al.* (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic acids research *42*, D966-974.

Koutsopoulos, O.S., Kretz, C., Weller, C.M., Roux, A., Mojzisova, H., Bohm, J., Koch, C., Toussaint, A., Heckel, E., Stemkens, D.*, et al.* (2013). Dynamin 2 homozygous mutation in humans with a lethal congenital syndrome. European journal of human genetics : EJHG *21*, 637-642.

Lee, J.R., Srour, M., Kim, D., Hamdan, F.F., Lim, S.H., Brunel-Guitton, C., Decarie, J.C., Rossignol, E., Mitchell, G.A., Schreiber, A.*, et al.* (2015). De novo mutations in the motor domain of KIF1A cause cognitive impairment, spastic paraparesis, axonal neuropathy, and cerebellar atrophy. Human mutation *36*, 69-78.

Li, Q., Liu, X., Gibbs, R.A., Boerwinkle, E., Polychronakos, C., and Qu, H.Q. (2014). Gene-specific function prediction for non-synonymous mutations in monogenic diabetes genes. PLoS One *9*, e104452.

Maddatu, T.P., Garvey, S.M., Schroeder, D.G., Zhang, W., Kim, S.Y., Nicholson, A.I., Davis, C.J., and Cox, G.A. (2005). Dilated cardiomyopathy in the nmd mouse: transgenic rescue and QTLs that improve cardiac function and survival. Human molecular genetics *14*, 3179-3189.

Martin, E., Schule, R., Smets, K., Rastetter, A., Boukhris, A., Loureiro, J.L., Gonzalez, M.A., Mundwiller, E., Deconinck, T., Wessner, M.*, et al.* (2013). Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. American journal of human genetics *92*, 238-244.

McKusick, V.A. (2007). Mendelian Inheritance in Man and its online version, OMIM. American journal of human genetics *80*, 588-604.

Messina, M.F., Messina, S., Gaeta, M., Rodolico, C., Salpietro Damiano, A.M., Lombardo, F., Crisafulli, G., and De Luca, F. (2012). Infantile spinal muscular atrophy with respiratory distress type I (SMARD 1): an atypical phenotype and review of the literature. European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society *16*, 90-94.

Michel, M.C., Wieland, T., and Tsujimoto, G. (2009). How reliable are G-protein-coupled receptor antibodies? Naunyn Schmiedebergs Arch Pharmacol *379*, 385-388.

Montenegro, G., Rebelo, A.P., Connell, J., Allison, R., Babalini, C., D'Aloia, M., Montieri, P., Schule, R., Ishiura, H., Price, J.*, et al.* (2012). Mutations in the ER-shaping protein reticulon 2 cause the axon-degenerative disorder hereditary spastic paraplegia type 12. J Clin Invest *122*, 538-544.

Murphy, S.M., Laura, M., Fawcett, K., Pandraud, A., Liu, Y.T., Davidson, G.L., Rossor, A.M., Polke, J.M., Castleman, V., Manji, H.*, et al.* (2012). Charcot-Marie-Tooth disease: frequency of genetic subtypes and guidelines for genetic testing. Journal of neurology, neurosurgery, and psychiatry *83*, 706-710.

Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research *31*, 3812-3814.

Nitta, R., Kikkawa, M., Okada, Y., and Hirokawa, N. (2004). KIF1A alternatively uses two loops to bind microtubules. Science *305*, 678-683.

Norton, N., Robertson, P.D., Rieder, M.J., Zuchner, S., Rampersaud, E., Martin, E., Li, D., Nickerson, D.A., Hershberger, R.E., National Heart, L.*, et al.* (2012). Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circ Cardiovasc Genet *5*, 167-174.

Oates, E.C., Rossor, A.M., Hafezparast, M., Gonzalez, M., Speziani, F., MacArthur, D.G., Lek, M., Cottenie, E., Scoto, M., Foley, A.R.*, et al.* (2013). Mutations in BICD2 cause dominant congenital spinal muscular atrophy and hereditary spastic paraplegia. American journal of human genetics *92*, 965-973.

Okada, Y., Yamazaki, H., Sekine-Aizawa, Y., and Hirokawa, N. (1995). The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric motor for antergrade axonal transport of synaptic vesicle precursors. Cell *81*, 769-780.

Okamoto, P.M., Gamby, C., Wells, D., Fallon, J., and Vallee, R.B. (2001). Dynamin isoform-specific interaction with the shank/ProSAP scaffolding proteins of the postsynaptic density and actin cytoskeleton. The Journal of biological chemistry *276*, 48458-48465.

Orphanet Prevalence of rare disease: Bibliographic data. Orphanet Reports Series 2014; Rare Diseases collection.

Peeters, K., Litvinenko, I., Asselbergh, B., Almeida-Souza, L., Chamova, T., Geuens, T., Ydens, E., Zimon, M., Irobi, J., De Vriendt, E.*, et al.* (2013). Molecular defects in the motor adaptor BICD2 cause proximal spinal muscular atrophy with autosomal-dominant inheritance. American journal of human genetics *92*, 955-964.

Pfeifer, D., Poulat, F., Holinski-Feder, E., Kooy, F., and Scherer, G. (2000). The SOX8 gene is located within 700 kb of the tip of chromosome 16p and is deleted in a patient with ATR-16 syndrome. Genomics *63*, 108-116.

Pierson, T.M., Tart, G., Adams, D., Toro, C., Golas, G., Tifft, C., and Gahl, W. (2011). Infantile-onset spinal muscular atrophy with respiratory distress-1 diagnosed in a 20-year-old man. Neuromuscular disorders : NMD *21*, 353-355.

Plummer, C., Spring, P.J., Marotta, R., Chin, J., Taylor, G., Sharpe, D., Athanasou, N.A., Thyagarajan, D., and Berkovic, S.F. (2013). Multiple Symmetrical Lipomatosis--a mitochondrial disorder of brown fat. Mitochondrion *13*, 269-276.

Pruitt, K.D., Harrow, J., Harte, R.A., Wallin, C., Diekhans, M., Maglott, D.R., Searle, S., Farrell, C.M., Loveland, J.E., Ruef, B.J.*, et al.* (2009). The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome research *19*, 1316-1323.

Riviere, J.B., Ramalingam, S., Lavastre, V., Shekarabi, M., Holbert, S., Lafontaine, J., Srour, M., Merner, N., Rochefort, D., Hince, P.*, et al.* (2011). KIF1A, an axonal transporter of synaptic vesicles, is mutated in hereditary sensory and autonomic neuropathy type 2. American journal of human genetics *89*, 219-230.

Rossor, A.M., Evans, M.R., and Reilly, M.M. (2015). A practical approach to the genetic neuropathies. Pract Neurol *15*, 187-198.

Rossor, A.M., Polke, J.M., Houlden, H., and Reilly, M.M. (2013). Clinical implications of genetic advances in Charcot–Marie–Tooth disease. Nature Reviews Neurology *9*, 562-571.

Santel, A., and Fuller, M.T. (2000). Control of mitochondrial morphology by a human mitofusin. Journal of cell science *114*, 867-874.

Sawyer, S.L., Cheuk-Him Ng, A., Innes, A.M., Wagner, J.D., Dyment, D.A., Tetreault, M., Care4Rare Canada, C., Majewski, J., Boycott, K.M., Screaton, R.A.*, et al.* (2015a). Homozygous mutations in MFN2 cause multiple symmetric lipomatosis associated with neuropathy. Human molecular genetics.

Sawyer, S.L., Hartley, T., Dyment, D.A., Beaulieu, C.L., Schwartzentruber, J., Smith, A., Bedford, H.M., Bernard, G., Bernier, F.P., Brais, B.*, et al.* (2015b). Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clinical genetics.

Schottmann, G., Jungbluth, H., Schara, U., Knierim, E., Morales Gonzalez, S., Gill, E., Seifert, F., Norwood, F., Deshpande, C., von Au, K.*, et al.* (2015). Recessive truncating IGHMBP2 mutations presenting as axonal sensorimotor neuropathy. Neurology *84*, 523-531.

Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nature methods *7*, 575-576.

Selcen, D., Muntoni, F., Burton, B.K., Pegoraro, E., Sewry, C., Bite, A.V., and Engel, A.G. (2009). Mutation in BAG3 causes severe dominant childhood muscular dystrophy. Annals of neurology *65*, 83-89.

Shashi, V., McConkie-Rosell, A., Rosell, B., Schoch, K., Vellore, K., McDonald, M., Jiang, Y.H., Xie, P., Need, A., and Goldstein, D.B. (2014). The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med *16*, 176-182.

Shin, H., Wyszynski, M., Huh, K.H., Valtschanoff, J.G., Lee, J.R., Ko, J., Streuli, M., Weinberg, R.J., Sheng, M., and Kim, E. (2003). Association of the kinesin motor KIF1A with the multimodular protein liprin-alpha. The Journal of biological chemistry *278*, 11393-11401.

Sock, E., Schmidt, K., Hermanns-Borgmeyer, I., Bosl, M.R., and Wegner, M. (2001). Idiopathic weight reduction in mice deficient in the high-mobility-group transcription factor Sox8. Mol Cell Biol *21*, 6951-6959.

Stolt, C.C., and Wegner, M. (2010). SoxE function in vertebrate nervous system development. Int J Biochem Cell Biol *42*, 437-440.

Sumner, C.J., d'Ydewalle, C., Wooley, J., Fawcett, K.A., Hernandez, D., Gardiner, A.R., Kalmar, B., Baloh, R.H., Gonzalez, M., Zuchner, S.*, et al.* (2013). A dominant mutation in FBXO38 causes distal spinal muscular atrophy with calf predominance. American journal of human genetics *93*, 976-983.

Van Den Berg-Vos, R.M., Van Den Berg, L.H., Visser, J., de Visser, M., Franssen, H., and Wokke, J.H. (2003). The spectrum of lower motor neuron syndromes. Journal of neurology *250*, 1279-1292.

Wagner, J.D., Huang, L., Tetreault, M., Majewski, J., Boycott, K.M., Bulman, D.E., Care4Rare Canada, C., Dyment, D.A., and McMillan, H.J. (2015). Autosomal recessive axonal polyneuropathy in a sibling pair due to a novel homozygous mutation in IGHMBP2. Neuromuscular disorders : NMD.

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

Yoon, P.W., Olney, R.S., Khoury, M.J., Sappenfield, W.M., Chavez, G.F., and Taylor, D. (1997). Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. Arch Pediatr Adolesc Med *151*, 1096-1103.

Zuchner, S., Mersiyanova, I.V., Muglia, M., Bissar-Tadmouri, N., Rochelle, J., Dadali, E.L., Zappia, M., Nelis, E., Patitucci, A., Senderek, J.*, et al.* (2004). Mutations in the mitochondrial GTPase mitofusin 2 cause Charcot-Marie-Tooth neuropathy type 2A. Nature genetics *36*, 449-451.

Zuchner, S., Noureddine, M., Kennerson, M., Verhoeven, K., Claeys, K., De Jonghe, P., Merory, J., Oliveira, S.A., Speer, M.C., Stenger, J.E.*, et al.* (2005). Mutations in the pleckstrin homology domain of dynamin 2 cause dominant intermediate Charcot-Marie-Tooth disease. Nature genetics *37*, 289-294.

Appendix A: All Known Genes which cause CMT and Related Disorders

Table AA1. All genes known to cause CMT and related disorders. Modified from Rossor *et al*. 2015.

Type (MIM# number)	Gene	Phenotype
Autosomal dominant CMT1		
CMT1A (118220)	17p dup. (*PMP22*) *PMP22* point mutation	Classic CMT1; Classic CMT1, , DSD, CHN (rarely recessive)
CMT1B (118200)	*MPZ*	CMT1, DSD, CHN, CMT2 (rarely recessive)
CMT1C (601098)	*LITAF*	Classic CMT1
CMT1D (607678)	*EGR2*	Classic CMT1, DSD, CHN
CMT1F (607734)	*NEFL*	CMT2 but can have slow MCV in the CMT1 range (rarely recessive)
CMT1 plus (614434)	*FBLN5*	Macular degeneration, cutis laxa, HMN, slow NCV
SNCV/CMT1 (608236)	*ARHGEF10*	Asymptomatic slow conduction velocities
Hereditary neuropathy with liability to pressure palsies		
HNPP (162500)	17p del. (*PMP22*) *PMP22* point mutation	Typical HNPP; Typical HNPP
Autosomal recessive CMT1 (CMT4)		
CMT4A (214400)	*GDAP1*	CMT2, usually severe early onset; Vocal cord and diaphragmatic paralysis described
CMT4B1 (601382)	*MTMR2*	Severe CMT1, facial, bulbar, focally folded myelin
CMT4B2 (604563)	*SBF2*	Severe CMT1, glaucoma, focally folded myelin
CMT4B3 (615284)	*SBF1*	CMT1, focally folded myelin
CMT4C (601596)	*SH3TC2*	Severe CMT1, scoliosis, cytoplasmic inclusions
CMT4D or HMSNL (601455)	*NDRG1*	Severe CMT1, gypsy, deafness, tongue atrophy
CMT4E (605253)	*EGR2*	CMT1, DSD, CHN phenotype
CMT4F (614895)	*PRX*	CMT1, predominantly sensory, focally folded myelin

CMT4G or HMSN Russe (605285)	*HK1*	Severe early-onset CMT1, gypsy
CMT4H (609311)	*FGD4* (Frabin)	Classic CMT1
CMT4J (611228)	*FIG4*	CMT1, predominantly motor, progressive
CCFDN (604168)	*CTDP1*	CMT1, gypsy, cataracts, dysmorphic features
CMT4	SURF-1	CMT1, encephalopathy, ataxia, reduced life span, Leigh's syndrome
Autosomal dominant CMT2		
CMT2A (609260)	*MFN2*	CMT2, progressive, optic atrophy (rarely recessive)
CMT2B or HSAN1B (600882)	*RAB7*	CMT2 with sensory complications (ulcero mutilating)
CMT2C (606071)	*TRPV4*	CMT2, vocal cord paralysis
CMT2D (601472)	*GARS*	CMT2 with predominant hand wasting
CMT2E (607684)	*NEFL*	CMT2 but can have nerve conduction velocities in the CMT1 range (rarely recessive)
CMT2F (606595)	*HSPB1*	Motor-predominant CMT2
CMT2I (607677)	*MPZ*	Late-onset CMT2
CMT2J (607736)	*MPZ*	CMT2 with hearing loss and pupillary abnormalities
CMT2K (607831)	*GDAP1*	Late-onset CMT2 (dominant), severe CMT2 (recessive)
CMT2L (608673)	*HSPB8*	Motor-predominant CMT2
CMTDIB or CMT2M (606482)	*DNM2*	ntermediate CMT or CMT2, cataracts, ophthalmoplegia
CMT2N (613287)	*AARS*	Classic CMT2
CMT2P (614436)	*LRSAM1*	Mild sensory-predominant CMT2 (dominant and recessive)
CMT2Q (615025)	*DHTKD1*	CMT2
HMSNP (604484)	*TFG*	CMT2 with proximal involvement
CMT2	*MARS*	Late-onset CMT2
CMT2	*HARS*	CMT2
CMT2	*VCP*	CMT2
SPG10 (604187)	*KIF5A*	CMT, hereditary spastic paraplegia

CMT2	*MT-ATP6*	CMT2, pyramidal signs, relapsing
Autosomal recessive CMT2		
CMT2B1 (605588)	*LMNA*	CMT2 rapid progression
CMT2B2 (605589)	*MED25*	Classic CMT2
NMAN (137200)	*HINT1*	Neuromyotonia and axonal neuropathy, motor predominant
CMT2R (615490)	*TRIM2*	Infantile-onset CMT2
AR-CMT2	*IGHMBP2*	CMT2
AR-CMT2	*HSJ1*	CMT2
X-linked CMT		
CMTX1 (302800)	*GJB1*	Males CMT1 (patchy NCV); females CMT2
CMTX4 or Cowchock'o syndrome (310490)	*AIFM1*	CMT2, infantile onset, developmental delay, deafness, learning difficulties
CMTX5 (311070)	*PRPS1*	CMT2, deafness, optic atrophy
CMTX6 (300905)	*PDK3*	CMT2
Dominant intermediate CMT		
CMTDIB or CMT2M (606482)	*DNM2*	Intermediate CMT or CMT2, cataracts, ophthalmoplegia, ptosis
CMTDIC (608323)	*YARS*	Intermediate CMT
CMTDID (607791)	*MPZ*	Intermediate CMT
CMTDIE (614455)	*IFN2*	Intermediate CMT, focal segmental glomerulosclerosis, end-stage renal failure
CMTD1F (615185)	*GNB4*	Intermediate CMT
Recessive intermediate CMT		
CMTRIA (608340)	*GDAP1*	Intermediate CMT
CMTRIB (613641)	*KARS*	Intermediate CMT, learning difficulty, vestibular schwannoma
CMTRIC (615376)	*PLEKHG5*	Intermediate CMT, SMA
CMTRID (616039)	*COX6A1*	Intermediate CMT, onset first decade
Hereditary motor neuropathy		
HMN2A (158590)	*HSPB8*	Classical HMN, dominant
HMN2B (608634)	*HSPB1*	Classical HMN, dominant
HMN2C (613376)	*HSPB3*	Classical HMN, dominant
HMN2D (615575)	*FBXO38*	Classical HMN, dominant

HMN with pyramidal features or ALS4 (602433)	*SETX*	HMN with pyramidal signs, dominant
DSMA5 (614881)	*DNAJB2 (HSJ1)*	Classical HMN, recessive
HMN5A (600794) or SPG17 (270685)	*BSCL2*	Predominant hand wasting, Silver syndrome but can have sensory involvement as in CMT2D, dominant
HMN5A (600794)	*GARS*	Predominant hand wasting, dominant
HMN5B (614751) or SPG31 (610250)	*REEP1*	Predominant hand wasting, pyramidal signs, dominant
HMN6 or SMARD1 (604320)	*IGHMBP2*	Infantile onset, respiratory distress, recessive
MARD2 or SMAX	*LAS1L*	Infantile onset, respiratory distress, X-linked recessive
HMN7A (158580)	*SLC5A7*	Classical HMN, vocal cord palsy, dominant
HMN7B (607641)	*DCTN1*	HMN, bulbar and facial weakness, dominant
SMAX3 (300489)	*ATP7A*	Classical HMN, X-linked
SMALED (158600)	*DYNC1H1*	Congenital, contractures, lower-limb predominant, pyramidal signs, cortical migration defects, learning difficulties, dominant
SMALED2 (615290)	*BICD2*	Congenital, contractures, lower-limb predominant, pyramidal signs, dominant
PNMHH (614369)	*MYH14*	Typical HMN, distal myopathy, hoarseness, hearing loss, dominant
SPSMA (181405)	*TRPV4*	HMN, scapular winging, vocal cord palsy, dominant
HMN	*AARS*	HMN with neuromyotonia, recessive
Hereditary sensory neuropathy (also called Hereditary sensory and autonomic neuropathy (HSAN))		
HSAN1A (162400)	*SPTLC1*	HSN with sensory complications (ulcero mutilating), dominant
HSAN1C (613640)	*SPTLC2*	HSN with sensory complications (ulcero mutilating), dominant
CMT2B (600882)	*RAB7*	HSN with sensory complications (ulcero mutilating), dominant

HSN1D (613708) or SPG3A (182600)	*ATL1*	HSN with sensory complications (ulcero mutilating), spasticity, dominant
HSN1E (614116)	*DNMT1*	HSN, hearing loss, dementia, dominant
HSN1F (615632)	*ATL3*	HSN, bone destruction, dominant
HSAN2A (201300)	*WNK1*	HSN with sensory complications (ulcero mutilating), recessive
HSAN2B or HSAN1B (613115)	*FAM134B*	HSN with sensory complications (ulcero mutilating), recessive
HSN2C (614213) or SPG30 (610357)	*KIF1A*	HSN with sensory complications (ulcero mutilating), recessive
HSAN3, familial dysautonomia or Rileymia or Rileyt	*IKBKAP*	Ashkenazi Jewish, autonomic dysfunction, HSN, absent fungiform papillae, recessive
Insensitivity to pain (24300), paroxysmal extreme pain disorder (167400), primary erythermalgia (133020), small-fibre neuropathy	*SCN9A*	Recessive: insensitivity to pain Dominant: paroxysmal extreme pain disorder, primary erythermalgia, small fibre neuropathy
CIPA or HSAN4 (256800)	*NTRK1*	Congenital insensitivity to pain with anhydrosis, recessive
HSAN5 (608654)	*NGF-B*	Insensitivity to pain, recessive
HSAN6 (614653)	*DST*	Ashkenazi Jewish, autonomic dysfunction, HSN, absent fungiform papillae, death by age 2, recessive
HSAN7 (615548)	*SCN11A*	Congenital insensitivity to pain with hyperhidrosis and gastrointestinal dysfunction, dominant
HSAN and dementia	*PRNP*	Autonomic dysfunction, sensory loss, dementia, dominant
Hereditary sensory neuropathy with spastic paraplegia (256840)	*CCT5*	HSN with sensory complications (ulcero mutilating) and spastic paraplegia, recessive

CHN, congenital hypomyelinating neuropathy; CMT, Charcot–Marie–Tooth disease; DSD, Dejerine–Sottas disease; HMN, hereditary motor neuropathy; HNPP, hereditary neuropathy with liability to pressure palsies; HSN, hereditary sensory neuropathy; MCV, motor conduction velocity; NCV, nerve conduction velocity; SMA, spinal muscular atrophy; SNCV, slowed nerve conduction velocity

Appendix B: Additional Clinical Information for Patients 130451 and 130105S

This information is modified from Sawyer *et al*. (2015).

Patient 130451 and 130105S are brothers who developed cervical and thoracic lipomatosis in their mid 20's to mid 40's. Patient 130451 presented with two small lipomas on the back of his neck in his 20's. Over the course of the following 25 years he developed striking unencapsulated cervical and thoracic lipomatosis and progressive tongue hypertrophy. He has undergone multiple liposuction surgeries, which were only of temporary benefit before the lesions returned. He has swallowing difficulties due to the bulk of the cervical lesion and tongue hypertrophy. At age 53 years, he presented with significant peripheral neuropathy, paresthesias and weakness, and required a cane to ambulate by 58 years. Nerve conduction studies and EMG demonstrated length dependent axonal polyneuropathy with evidence of denervation/reinervation. He is on pregabalin for neuropathic pain. He was diagnosed with diabetes mellitus (DM) in his early 50's, and is well controlled by oral hypoglycemic medication. His glucose control was improved after liposuction; fasting glucose and HbA1C were within the normal range a few weeks postliposuction on his medications. Genetic testing for the MERRF mutation (m.8344A>G) was negative in a peripheral blood sample, as were other common mitochondrial mutations (m.3243A>G, m.3260A>G, m.3303C>T and m.8993T>G/C). Biochemical studies have repeatedly demonstrated an elevated lactate (5.02 and 4.7 mmol/L, upper limit of normal 2.2 mmol/L). His leptin was low at >0.6 ng/ml (0.7 to 5.3 ng/ml normal range) as was adiponectin at 3 mcg/ml (normal range 4-20 mcg/ml). His CK, cholesteral, triglycerides, LDL, homocysteine, apolipoprotein A were noral. His HDL was slightly low. A skin biopsy was sent for electron microscopy and mitochondrial structure was unremarkable. He has no history of alcohol abuse.

Patient 130105S had a history of 'bowed limbs' requiring casting in infancy, however was very healthy and physically active throughout early and mid-adulthood. He noted a change in his physique around age 45 with evidence of lipomatosis by his early 50s. He has a history of multiple encapsulated cervical and thoracic lipomas, with frequent surgical interventions. He has an approximately 15 year history of progressive neuropathy, and normal HbA1C. He had elevated lactate (3.0 mmol/L, upper limit of normal 2.2 mmol/L). He also had an elevated CK as high as 550 U/L. He had a muscle biopsy as part of his work up which reported marked variation in myofiber size and grouping by type, as well as diffuse expression of cd56 and esterase positive myofibers, which together is suggestive of chronic denervation. The biopsy also identified rare vacuoles, increased internalized nuclei, and rare RRFs, which are subtle myopathic features, though can also be seen secondary to denervation.

The brothers have two unaffected siblings. The parents were of Irish descent and are 5th degree cousins. Both parents are deceased, and neither was diagnosed with CMT or MSL. Both patients have children in their early 20's, none of whom have apparent MSL or gait abnormalities.

Appendix C: Additional Clinical Information for Patient 120915T and his Twin Brother

The patient information found here is modified from Lee *et al*. (2015).

Patient (120915T) and his monozygotic twin brother (120916R) are 14-years-old with mild-moderate intellectual disability, spastic paraparesis, axonal sensorimotor polyneuropathy, and mild optic atrophy. They were born after an unremarkable pregnancy at 38.5 weeks gestation by Caesarean section because 120915T was in a breech presentation. Delays were noted at 12 months of age. Though they sat at eight to nine months and walked with support at one year of age, the boys never walked completely independently until age five years as they had poor balance with frequent falls and preferred to crawl rather than walk. In addition, the twins were delayed in their fine motor skills and speech development. They are currently in grade eight working at an approximately grade one level, speak in sentences, and are easily understood. The issues with independent walking prompted further evaluation and the boys were diagnosed with hereditary spastic paraplegia at two years of age. At age nine years they had bilateral femoral derotation osteotomies and bilateral medial hamstring releases and are now able to walk fairly well with a cane; overall their spasticity has been fairly static over the years. Nerve conduction studies at age four years were reported as normal but at age 13 years they were found to have a severe sensorimotor polyneuropathy affecting sensory axons more motor axons, and the lower limbs more than the upper limbs. Optic atrophy was first noted at age 13 years and was mild for both twins. The boys also have strabismus and hyperopia. They suffer from dysfunctional voiding and constipation. They have normal hearing and have never had any clinical seizures. MRI at eight years of age showed a slightly enlarged pineal gland for both boys. 120916R had a developmental venous anomaly in the right putamen. Otherwise the MRIs were

unremarkable with normally appearing corpus callosum, cerebellum and cerebral hemispheres. Occipitofrontal diameter for both twins was less than the 2nd centile during childhood but moved to the normal curve between 10-11 years of age. At their last examination at 14 years of age, 120916R's height was 163 cm (50th centile), weight 74 kg (95th centile), and head circumference 54.5 cm (50th centile). 120915T's height was 157 cm (10-25th centile), weight 53 kg (50th centile), and head circumference was 53 cm (10th centile). Numerous genetic and metabolic investigations were unrevealing so the family was enrolled in C4R.

Appendix D: Additional Clinical Information for Patient CH0082 and her Affected Brother

Modified from (Wagner et al., 2015).

Patient CH0082 presented with leg and arm weakness. Onset was at three years of age with increased falling and slowly progressive distal leg weakness. Hand weakness became apparent at five years of age. At ten years of age she was able to walk wearing rigid ankle-foot orthosis. She had no proximal weakness and could climb stairs independently in an alternating manner. She had trouble writing and was unable to fasten buttons. She had no oculobular weakness and was cognitively normal. She was of Syrian descent and while her parents reported no consanguinity, her maternal and paternal grandparents were from the same village. Physical exam at ten years of age revealed normal cranial nerve testing. Marked wasting of her intrinsic hand, foot, and gastrocnemius muscles was noted as was pes cavus (Figure. AD1). Muscle weakness was noted as follows: wrist flexors 4+; first dorsal interossei 1, abductor pollicus brevis (APB) 1, abductor digiti minimi (ADM) 1, flexor pollicus longus 4, flexor digitorum profoundus 4-, tibialis anterior 0, tibilais posterior 0, peroneus longus 0, gastrocnemius 0, extensor hallicus longus 0. Proximal muscle strength was intact. Diffuse areflexia was noted. Vibration sense was absent at the toes and decreased at the thumbs. Pin-prick sensation was decreased below mid-thigh level and below her elbows. No contractures were evident. Respiratory exam revealed good air entry to bases on auscultation. Spine exam revealed mild hyperlordosis with no scoliosis or rigidity.

Nerve conduction studies at ten years of age revealed a severe sensorimotor polyneuropathy with predominantly axonal features (Table AD1). Pulmonary function testing at 14 years of age, revealed a normal forced vital capacity (FVC) of 3.73 L (93%

Figure AD1. Distal muscle wasting is shown in the upper and lower limbs for Patient CH0082 (a-b) and her affected brother (c-d). Gastrocnemius atrophy and pes cavus (b, d) is seen which was more severe in the brother.

predicted) and total lung capacity (TLC) of 4.76 L. Ultrasound confirmed normal motion of the diaphragm. She has remained clinically stable over the past four years. Biochemical testing revealed a serum creatine kinase of 234 U/L (normal < 175 U/L).

Extensive clinical genetic testing was performed and showed no mutation in the following genes: *MPZ*, *FGD4*, *CTDP1*, *FIG4*, *GDAP1*, *MTMR2*, *NDRG1*, *PRX*, *SBF2*, *SH3TC2*, *PMP22* (sequencing and duplication testing).

Her younger brother presented with similar, albeit more severe, symptoms. He took his first independent steps at 18 months old and has had gradual, progressive gait difficulty and distal weakness since that time. He requires ankle foot orthoses to ambulate. Physical examination at eight years of age revealed similar pattern of muscle wasting (Figure AD1), pes cavus, and weakness as his sister: extensor digitorum communis 3, flexor pollicus longus 4-, first dorsal interosseous 1, abductor pollicus brevis 1, tibialis anterior 0, tibialis posterior 0, peroneus longus 0, gastrocnemius 0, extensor hallicus longus 0. Proximal muscle strength was normal. He was areflexic. Sensation testing showed diminished vibration sense, light touch and pinprick.

Nerve conduction studies at eight years of age revealed a severe sensorimotor polyneuropathy with axonal features (Table AD1). Pulmonary function testing at 11 years ten months of age revealed a normal FVC of 3.00 L (93% predicted) and TLC of 3.81 L (89% predicted). Ultrasound confirmed normal motion of the diaphragm.

Table AD1. Nerve conduction studies for Patient CH0082 and her affected brother

	Normal*	CH0082		Brother
AGE:		At 10 yo		At 8 yo
		Right	Left	Right
MOTOR:				
Median nerve				
DML (wrist-APB)	≤ 4.0	**NR**	**5.7**	**NR**
CMAP (mV)	≥ 4.0		**0.1**	
CV (m/sec)	≥ 50		**13**	
Ulnar nerve				
DML (msec; wrist-ADM)	≤ 3.3	3.3		2.9
CMAP (mV)	≥ 6.0	**0.3**		**0.6**
CV (m/sec)	≥ 50	**38**		**39**
Tibial nerve				
DML (msec; ankle-AH)	≤ 5.8	**NR**		**NR**
CMAP (mV)	≥ 4.0			
CV (m/sec)	≥ 40			
Peroneal nerve				
DML (msec; ankle-EDB)	≤ 6.5	**NR**		
CMAP (mV)	≥ 2.0			
CV (m/sec)	≥ 40			
SENSORY:				
Median nerve				
PL (msec; wrist-digitII)	< 3.2	**NR**	**9.5**	**NR**
SNAP (μV)	≥ 20		**8.8**	
CV (m/sec)	≥ 50		**14**	
Ulnar nerve				
PL (msec; wrist-digitV)	< 3.3	**4.3**		**4.2**
SNAP (μV)	≥ 9	**3.6**		10
CV (m/sec)	≥ 40	**28**		**24**
Superficial peroneal				
PL (msec; ankle-foot)	< 3.8	**NR**		
SNAP (μV)	≥ 5			
CV (m/sec)	≥ 40			
Sural nerve				
PL (msec; calf-latmall)	< 4.2	**NR**		**NR**
SNAP (μV)	≥ 6			
CV (m/sec)	≥ 40			

Bold values are abnormal. All sensory responses are antidromic.
Legend: DML=distal onset motor latency; CMAP=compound motor action potential; CV=conduction velocity; PL=peak onset latency; SNAP=sensory nerve action potential; APB=abductor pollicus brevis; ADM=abductor digiti minimi; AH=abductor hallicus; EDB=extensor digitorum brevis; NR=no response.
***Normal reference values** from: Kang PB. Pediatric nerve conduction studies and EMG. In: Blum AS, Rutkove SB, eds. *The Clinical Neurophysiology Primer*. Totowa: Humana Press; 2007:369–389.

Appendix E: Summary of the 37 Patients in this Study

Patient ID	Clinician	Description	Investigations	Exome Status	Diagnosis/Candidates	Next Steps
48579	Warman	Distal wasting; club feet; fused cervical vertebrae; Age of onset at one year	Sanger validated	Solved	*BICD2*	
53645	Innes	Distal sensory/ motor neuropathy	Sanger validated	Solved	*DNM2*	
130451	Sawyer	CMT2A; multiple symmetric lipomatosis	Sanger validated; Functional analysis performed off-site	Solved	*MFN2*	
130105S	Sawyer	CMT2A; multiple symmetric lipomatosis	Sanger validated; Functional analysis performed off-site	Solved	*MFN2*	
120915T	Boycott	Intellectual disability; spastic paraplegia; hereditary sensory neuropathy	Sanger validated	Solved	*KIF1A*	
53473	Warman	Distal sensory/ motor neuropathy; gait ataxia; fine postural tremor; visual impairment; onset in 40s	Uploaded novel candidates into PC	Candidates	*CHRM5* *TRIML2*	Find additional family using data-sharing
57606	Warman	Axonal sensory/ motor neuropathy; severe; childhood onset	Uploaded novel candidate into PC	Candidate	*CTTNBP2*	Find additional family using data-sharing
145460	Yoon	Sensory/ motor polyneuropathy; Prominent axonal features		No clear candidates		Obtain patient consent to upload data into PC
130206E	Boycott	Distal sensory/ motor neuropathy; atrophy; distal weakness; axonal type	Uploaded novel candidates into PC	Candidates	*TMEFF2* *PPFIA2* *ANKFY1*	Find additional family using data-sharing

55200	Boycott	Neuropathy; No myopathy; Kyphoscoliosis; Childhood onset	Uploaded novel candidate into PC	Candidate	*OCSTAMP*	Find additional family using data-sharing
156651 (TR0004)	Yoon	Profound neuropathy; pain insensitivity; areflexia; hammer toes; severe developmental delays; autistic features	Sanger validated; Uploaded novel candidate into PC	Candidate	*NUBP2*	Find additional family using data-sharing
156652 (TR0003)	Yoon	Profound neuropathy; pain insensitivity; areflexia; hammer toes; severe developmental delays; autistic features	Sanger validated; Uploaded novel candidate into PC	Candidate	*NUBP2*	Find additional family using data-sharing
54976	Innes	Hereditary neuropathy	Uploaded novel candidate into PC	Candidate	*NDFIP1*	Find additional family using data-sharing
56761	Thomas	Mild CMT		No clear candidates		Obtain patient consent to upload data into PC
53411	Innes	Motor/ sensory wasting; Hammer toes; axonal subtype	Uploaded novel candidate into PC	Candidate	*IARS*	Calgary conducting further analysis; Find additional family using data-sharing
87514	Warman	Hereditary sensory autonomic neuropathy; developmental delay	Uploaded novel candidate into PC	Candidate	*SUGT1*	Find additional family using data-sharing
Subject 2	Baker	Hand amyotrophy; inverted legs; pes cavus; hammer toes; ankle weakness; plantar responses		No clear candidates		Obtain patient consent to upload data into PC
Subject 4	Baker	Slight weakness and burning/ pins and needles in feet and hands; mixed axonal and demyelinating signs; lower extremities predominant		No clear candidates		Obtain patient consent to upload data into PC
Subject 5	Baker	Lower extremity distal amyotrophy; pes cavus; hammer toes; axonal signs		No clear candidates		Obtain patient consent to upload data into PC

Subject 14	Baker	Upper and lower extremity amyotrophy; sensory loss; postural tremor		No clear candidates		Obtain patient consent to upload data into PC
Subject 16	Baker	Distal weakness; pes cavus	Sanger validated	Solved	*BAG3*	Find additional family using data-sharing
Subject 17 (60294)	Baker	Distal weakness; inverted legs; ankle/ knee jerks		No clear candidates		Obtain patient consent to upload data into PC
Subject 18	Baker	Foot drop; hammer toes; pes cavus; late onset		No clear candidates		Obtain patient consent to upload data into PC
Subject 19	Baker	Mild wasting in shins; pes cavus; hammer toes; superimposed carpal tunnel syndrome	Sanger validated	Solved	*FBXO38*	Find additional family using data-sharing
Subject 20	Baker	Distal weakness; moderate pes cavus; low hand dexterity; distal atrophy		No clear candidates		Obtain patient consent to upload data into PC
Subject 21	Baker	Pes cavus; distal weakness; high IQ; axonal CMT pattern	Sanger Validated	Solved	*SIGMAR1*	Find additional family using data-sharing
Subject 22	Baker	Severe distal wasting; bilateral foot drop; hand wasting; axonal neuropathy	Sanger Validated	Solved	*PRX*	Find additional family using data-sharing
Subject 23	Baker	Distal weakness; axonal neuropathy		No clear candidates		Obtain patient consent to upload data into PC
Subject 24	Baker	Axonal CMT; pes cavus; distal weakness		No clear candidates		Obtain patient consent to upload data into PC
Subject 25	Baker	Numbness and tingling sensation associated with pain in the feet; mild muscle wasting in the leg; Sensory axonal neuropathy	Sanger Validated	Solved	*GARS*	Find additional family using data-sharing
Subject 26	Warman	CMT2; Tremor; Reduced vibration sense		No clear candidates		Obtain patient consent to upload data into PC
Subject 27	Warman	CMT2; Tremor; Reduced reflexes; Myoclonic potentials		No clear candidates		Obtain patient consent to upload data into PC
Subject 28	Warman	Oculopharyngeal muscular dystrophy (OPMD); Ptosis; Distal weakness; facial dysplasia; Myotonic potentials		No clear candidates		Obtain patient consent to upload data into PC

Subject 29	Warman	Ptosis; CMT2		No clear candidates		Obtain patient consent to upload data into PC
Subject 30	Warman	Distal weakness; Shoulder weakness; Reduced reflexes; Reduced sensory nerves; Atrophy of hands and feet		No clear candidates		Obtain patient consent to upload data into PC
CH0082	Dyment	Distal weakness; decreased sensation; wasting in upper and lower limbs; areflexia; childhood onset	Sanger Validated; qPCR; Western Blot	Solved	*IGHMBP2*	
120879P	Boycott	Congenital myopathy/ lower motor neuron disease; skeletal dysplasia; contractures; short stature; amenorrhea; possible cataracts	Sanger validated; qPCR; Western Blot	Candidate	*SOX8*	Find additional family using data-sharing

PC, PhenomeCentral; CMT, Charcot-Marie-Tooth Disease

CURRICULUM VITAE

EDUCATIONAL BACKGROUND

- **Bachelor of Science, Honors Specialization in Biology and Genetics** **2013**
  University of Western Ontario, London, Ontario

ACADEMIC HIGHLIGHTS AND ACHIEVEMENTS

- Awarded the University of Ottawa Admission Scholarship- Masters **2013**
- Achieved Bachelor of Science with Honors **2013**
- Achieved Dean's Honor List (University of Western Ontario) **2010-2013**
- Awarded 1st place at the Keenan Research Summer Student Poster Competition **2012**
- Awarded the University of Western Ontario Fred Landon Continuing Admission Scholarship **2009-2012**

PUBLICATIONS

Wagner, J.D., Huang, L., Tetreault, M., Majewski, J., Boycott, K.M., Bulman, D.E., Care4Rare Canada, C., Dyment, D.A., and McMillan, H.J. (2015). Autosomal recessive axonal polyneuropathy in a sibling pair due to a novel homozygous mutation in IGHMBP2. Neuromuscular disorders *25*, 794-799.

Sawyer, S.L., Cheuk-Him Ng, A., Innes, A.M., Wagner, J.D., Dyment, D.A., Tetreault, M., Care4Rare Canada, C., Majewski, J., Boycott, K.M., Screaton, R.A.*, et al.* (2015). Homozygous mutations in MFN2 cause multiple symmetric lipomatosis associated with neuropathy. Human molecular genetics *24*, 5109-5114.

Lee, J.R., Srour, M., Kim, D., Hamdan, F.F., Lim, S.H., Brunel-Guitton, C., Decarie, J.C., Rossignol, E., Mitchell, G.A., Schreiber, A.*, et al.* (2015). De novo mutations in the motor domain of KIF1A cause cognitive impairment, spastic paraparesis, axonal neuropathy, and cerebellar atrophy. Human mutation *36*, 69-78.

Lin, S., Callaway, C.W., Shah, P.S., Wagner, J.D., Beyene, J., Ziegler, C.P., and Morrison, L.J. (2014). Adrenaline for out-of-hospital cardiac arrest resuscitation: a systematic review and meta-analysis of randomized controlled trials. Resuscitation *85*, 732-740.