INHERITED PERIPHERAL NEUROPATHIES: MUTATIONAL ANALYSIS OF CYTOPLASMIC DYNEIN-DYNACTIN GENES AND THE IDENTIFICATION OF A NOVEL AUTOSOMAL RECESSIVE GENE

TEY SHELISA

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

INHERITED PERIPHERAL NEUROPATHIES: MUTATIONAL ANALYSIS OF CYTOPLASMIC DYNEIN-DYNACTIN GENES AND THE IDENTIFICATION OF A NOVEL AUTOSOMAL RECESSIVE GENE

TEY SHELISA

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Tey Shelisa (I.C/Passport No:)

Matric No: MHA120065

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Inherited peripheral neuropathies: mutational analysis of cytoplasmic dynein-dynactin genes and the identification of a novel autosomal recessive gene.

Field of Study: Neurogenetics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: Azlina Ahmad Annuar

Designation: Doctor

ABSTRACT

Inherited peripheral neuropathies (IPNs) are a clinically and genetically heterogeneous group of disorders affecting the peripheral nervous system. IPNs can be classified into three main groups - Hereditary Motor Neuropathies (HMN), Hereditary Sensory Neuropathies (HSN) and Hereditary Motor and Sensory Neuropathies (HMSN) which is also known as Charcot-Marie-Tooth disease (CMT), depending on the type of nerves affected. Over 80 causal genes have been reported for IPNs, yet there are many cases that remain genetically unsolved. The main objective of this study was to identify new genes associated with IPNs. A candidate gene approach was taken to screen eighteen genes within the cytoplasmic dynein-dynactin complex in 136 genetically undetermined IPN patients, whereby ten genes encoding the cytoplasmic dynein complex (*DYNC1H1*, *DYNC1I1*, *DYNC1I2*, *DYNC1LI1*, *DYNC1LI2*, *DYNLL1*, *DYNLL2*, *DYNLT1*, *DYNLT3*, *DYNLRB1*); and eight genes encoding the dynactin complex (*DCTN1*, *DCTN2*, *DCTN3*, *DCTN4*, *DCTN5*, *DCTN6*, *ACTR1A*, *ACTR1B*) were screened. In total, 25 variants were identified; however segregation analysis excluded them from being the likely cause of the disease in the respective families. In the second part of this PhD study, a consanguineous family with CMT, named CMT861, was investigated further to identify the causal gene. A combination of linkage analysis, homozygosity mapping and whole exome sequencing provided strong genetic evidence towards one candidate variant (c.A118C; p.T40P) located in *AHNAK2*. Gene expression analysis of *AHNAK2* in cultured fibroblasts from the affected brothers showed a significant decrease at the mRNA and protein levels. Spatiotemporal expression analysis of *Ahnak2* in mouse neuronal and muscle tissues suggested that this gene is a good candidate for the CMT phenotype in the family as it is expressed in the tissues that are affected in peripheral neuropathy. AHNAK2 is known to form heterodimers with

PRX, in which *PRX* is a reported autosomal recessive CMT gene. *PRX* encodes for a protein called periaxin, which is required for myelin maintenance. We therefore hypothesise that mutations in *AHNAK2* may disrupt the AHNAK2-PRX pathway and affect nerve myelination which is the most commonly affected pathway in IPN. In conclusion, although no pathogenic mutation was identified in the cytoplasmic dynein-dynactin genes, variants identified in this study could be useful as reference for other research groups in evaluating the pathogenicity of variants identified in their cohorts. Furthermore, we report the identification of *AHNAK2* as a novel candidate gene for autosomal recessive CMT.

ABSTRAK

Penyakit neuropati periferi yang diwarisi (ataupun dikenal sebagai '*inherited peripheral neuropathies*', IPNs) adalah sekumpulan penyakit yang mempunyai pelbagai ciri-ciri klinikal dan usul-usul genetik. IPN boleh dikelaskan kepada tiga kumpulan utama, iaitu neuropati saraf motor yang diwarisi ('*hereditary motor neuropathy*', HMN), neuropati saraf sensorik yang diwarisi ('*hereditary sensory neuropathy*', HSN) dan penyakit Charcot-Marie-Tooth (CMT). Menurut pangkalan data mutasi IPN, terdapat lebih daripada 80 gen yang telah dikenalpasti sebagai punca penyakit IPN. Walaubagaimanapun, masih terdapat sekumpulan besar kes-kes yang belum dikenalpasti punca penyakitnya di peringkat DNA. Objektif utama penyelidikan tesis ini adalah untuk mengenalpasti gen-gen baru yang mungkin boleh mengakibatkan IPN. Untuk mencapai objektif ini, lapan belas gen di dalam sebuah kompleks protin yang dikenali sebagai 'kompleks dynein-dynactin sitoplasmik' telah diperiksa dalam 136 pesakit IPN di mana punca penyakitnya masih lagi belum dikenalpasti. Sebanyak sepuluh gen kompleks sitoplasmik dynein ini (*DYNC1H1, DYNC1I1, DYNC1I2, DYNC1LI1, DYNC1LI2, DYNLL1, DYNLL2, DYNLT1, DYNLT3, DYNLRB1);* dan sebanyak lapan gen untuk kompleks dynactin (*DCTN1, DCTN2, DCTN3, DCTN4, DCTN5, DCTN6, ACTR1A, ACTR1B)* telah disiasat secara terperinci. Sebanyak 25 perubahan jujukan ('varian') telah dikesan, tetapi analisa mengikut corak segregasi penyakit di dalam keluarga-keluarga terbabit telah mengecualikan kesemua varian ini daripada menjadi punca-punca penyakit tersebut. Pada bahagian kedua tesis ini, kajian genetik ke atas pesakit CMT daripada sebuah keluarga di mana ibubapanya adalah sepupu telah dilaksanakan. Gabungan analisa daripada data secara WES, analisa pautan genetik dan pemetaan rantau homozigot telah memberi bukti yang kukuh bahawa satu varian pada gen *AHNAK2*, (c.A118C, p.T40P) mungkin memainkan peranan dalam penyakit CMT di dalam keluarga ini. Lanjutan itu, kajian tahap ekspresi mRNA dan

protin *AHNAK2* pada sel-sel fibroblast untuk kedua-dua adik-beradik ini didapati lebih rendah berbanding dengan sel-sel fibroblast kawalan. Analisasi tahap spatio-temporal ekspresi *Ahnak2* dalam tisu saraf dan tisu otot tikus telah mengenalpastikan kewujudan gene ini dimana tisu-tisu tersebut biasanya dipengaruh dalam pesakit neuropati periferi. AHNAK2 didapati boleh membentuk struktur heterodimer sesame PRX, dimana *PRX* adalah satu autosomal resesif gen yang diketahui bagi CMT. *PRX* mengekod protein periaxin, dimana ia memainkan peranan penting dalam penyelenggaran myelin. Kami mengandaikan bahawa mutasi dalam *AHNAK2* boleh menganggu system AHNAK2-PRX dan menjejaskan fungsi myelinasi saraf, dimana ia merupakan fungsi yang paling serang diganggu dalam penyakit IPN. Secara kesimpulan, walaupun tiada varian pathogenic dikesan dalam gen-gen kompleks dynein-dynactin sitoplasmik, varian-varian yang dilaporkan dalam kajian ini boleh diguna oleh para penyelidik sebagai rujukan dalam penilaian sifat pathogenic varian-varian yand dikesani dalam kohort mereka. Kami juga berjaya mengenal pastikan *AHNAK2* sebagai gene baru bagi penyakit autosomal resesif CMT.

STEPS TAKEN TO INVESTIGATE THE DISEASE-CAUSING GENE IN CMT861 AND THE RATIONALE BEHIND IT

At first, known genes that have been reported to cause IPN were screened in CMT861 and were shown to be negative. This indicated that a new disease gene is likely to be the cause of the disease in this family. To identify the disease locus, we performed a genome-wide linkage analysis on CMT861 using genotype data obtained from the Golden gate linkage V panel. This SNP array consists of 6056 highly informative SNPs that are distributed evenly across the genome. This analysis allowed us to identify potential disease loci; this also means that we managed to exclude other regions on the genome which did not achieve suggestive LOD scores. Through genome-wide linkage analysis, we identified 8 suggestive linkage regions. As the genomic distance between the SNPs for linkage analysis was quite sparse, certain recombinations that occurred between these SNPs might not have been detected. Therefore, part of these suggestive linkage regions could represent false positive regions. Microsatellite markers are generally more informative compared to SNP markers. To verify our data, we saturated these regions with microsatellite markers and performed linkage analysis on the 8 suggestive linkage regions. This approach allowed us to further refine the potential disease locus to only 2 suggestive linkage regions, which were on chromosome 12 and chromosome 14 respectively. To determine the disease locus between these 2 regions, we performed homozygosity mapping. For this analysis, the genotype data extracted from the whole exome sequencing (WES) data of five of the family members were used. Homozygosity mapping is based on the principle that the affected children of a consanguineous marriage are highly likely to be homozygous for the causal variant, which has arisen in a common ancestor and has been transmitted by both parents who are related. The alleles that are in linkage disequilibrium with the causal variant are therefore likely to be transmitted in the same haplotype block. By

looking for regions with 'runs of homozygosity' (ROH) in the genome, the locus for a causal variant may be identified. Through this approach, we managed to identify 1 ROH that is shared by the two affected brothers but not present in the unaffected family members. By combining the linkage and homozygosity mapping data, the genomic coordinate of the ROH is overlapping with the suggestive linkage interval on chromosome 14. This interval is located on chromosome 14q32.2-q32.33 and spans 7.48Mb. The next approach was to identify the candidate variants that fall within this interval. By performing a series of WES data filtering according to the disease segregation pattern, 8 candidate variants were identified within the chromosome 14 interval. Among these variants, two were in the coding region and one of them was a novel variant. This novel variant falls on the gene AHNAK2. We recognise that WES is only able to detect variants within the coding regions and we might have failed to detect potential causal variant if it is located within the non-coding region. However, based on the statistic that approximately 85% of the disease-causing mutations occur within the coding region for Mendelian disorders, our approach by first targeting the coding region is reasonable. Our data was further supported by several functional analyses. Gene expression analysis showed that AHNAK2 expression was reduced by at least 8-fold in the primary fibroblast cells of the two affected brothers as compared to the controls. Immunofluorescence studies on the fibroblast cells of one of the affected brothers also showed drastic reduction of the AHNAK2 protein. Molecular and functional analyses have supported our hypothesis that AHNAK2 is the novel autosomal recessive CMT gene that causes the disease in CMT861.

ACKNOWLEDGEMENTS

Undertaking and completing my PhD project has been a tough endeavor at times, but it was an amazing and fruitful experience. This journey would not have been possible without the constant guidance and support from the following people:

Dr. Azlina Ahmad Annuar, my main supervisor; I am forever grateful for the opportunities you have given me as your student, to perform research overseas and to attend international conferences. The consultant of my PhD project, Associate Professor Marina Kennerson; I want to thank you for hosting me in your laboratory and making me feel like home during my time in Sydney. My co-supervisor, Professor Nortina Shahrizaila; I want to thank you for helping me with the clinical side of the project and for providing us invaluable samples for the research. Thank you my supervisors and consultant for providing me constant encouragement and advice throughout this project, and enhancing my thesis with helpful suggestions.

To all the lovely members from Neuroscience Laboratory, Department of Biomedical Science, University of Malaya, and Northcott Neuroscience Laboratory, ANZAC Research Institute; thank you for your support and friendship. Dr. Alexander Drew, thank you for your help and guidance with the bioinformatics analysis of NGS data; I have learnt a lot from you. To all my friends, thank you for your encouragement and good times. Big thanks to all the family members who participated in this study.

To my lovely family; thank you for encouraging me, believing in me and loving me throughout the years. Finally, to my beloved boyfriend Janko; without your loving support and patience, this thesis would have been even harder.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

%	:	Percent
(v/v)	:	Volume per volume
(w/v)	:	Weight per volume
˚C	:	Degree Celsius
aa	:	Amino acid
ACRF	:	Australian Cancer Research Foundation
AD	:	Autosomal dominant
ALS	:	Amyotrophic lateral sclerosis
AR	:	Autosomal recessive
bp	:	Base pairs
cDNA	:	Complementary DNA
CEU	:	Utah Residents (CEPH) with Northern and Western Ancestry
Chr	:	Chromosome
cM	:	Centimorgan
CMT	:	Charcot-Marie-Tooth
CMTX	:	X-linked CMT
CNS	:	Central nervous system
CNV	:	Copy number variation
CO2	:	Carbon dioxide
dbSNP	:	SNP database
dHMN	:	Distal hereditary motor neuropathies
DMEM	:	Dulbecco's modified eagle medium
DNA	:	Deoxyribonucleic acid
DNM2	:	Dynamin
dNTP	:	Deoxynucleotide triphosphate
dNTP	:	Deoxyribonucleotide triphosphate
dsDNA	:	Double stranded DNA
e.g.	:	Example
EDTA	:	Ethylenediaminetetraacetic acid
ELOD	:	Expected LOD
ENCODE	:	ENCyclopedia of DNA Elements
ExAC	:	Exome Aggregation Consortium
Exome	:	Exonic sequence of the genome
FTD	:	Frontotemporal dementia
g	:	Gravitational force
GIH	:	Gujarati Indian from Houston, Texas
HGP	:	Human genome project
HMN	:	Hereditary motor neuropathy
HMSN	:	Hereditary motor and sensory neuropathy
HNPP	:	Hereditary neuropathy with liability to pressure palsies
HRM	:	High resolution melt
HSF	:	Human Splicing Finder
HSN	:	Hereditary sensory neuropathy
IGV	:	Integrative Genomics Viewer
Indel	:	Insertions and deletion
IPN	:	Inherited peripheral neuropathy

kb	:	Kilo base pairs
kDa	:	Kilo Dalton
LOD	:	Logarithm of the odds
LVP	:	Linkage V panel
MAF	:	Minor allele frequency
max.	:	Maximum
Mb	:	Megabase
MDa	:	Mega Dalton
mg	:	Milligram
min	:	Minutes
mL	:	Milliliter
mM	:	Millimolar
mm	:	Millimeter
mM	:	Millimolar
mRNA	:	Messenger RNA
MT	:	Mutant
NCBI	:	National Center for Biotechnology Information
NCV	:	Nerve conduction velocity
ng	:	Nanogram
NGS	:	Next generation sequencing
PBS	:	Phosphate buffered saline
PCR	:	Polymerase chain reaction
PNS	:	Peripheral nervous system
PRX	:	Periaxin
RBC	:	Red blood cells
RFLP	:	Restriction fragment length polymorphism
RNA	:	Ribonucleic acid
rpm	:	Revolutions per minute
RT	:	Reverse transcription
SD	:	Standard deviation
SMA	:	Spinal muscular atrophy
SMA-LED	:	Spinal muscular atrophy with lower extremities predominance
SNP	:	Single nucleotide polymorphism
Ta	:	Annealing temperature
TAE	:	Tris-acetate-EDTA buffer
Tm	:	Melting temperature
UCSC	:	University of California, Santa Cruz
UMMC	:	University of Malaya Medical Center
UTR	:	Untranslated region
WES	:	Whole exome sequencing
WGS	:	Whole genome sequencing
WT	:	Wild type
X	:	Times
θ	:	Recombination fraction
µg	:	Microgram
µL	:	Microliter

LIST OF APPENDICES

CHAPTER 1: LITERATURE REVIEW AND GENERAL INTRODUCTION

1.1 Introduction

Inherited peripheral neuropathies (IPNs) are a clinically and genetically heterogeneous group of disorders. Even though more than 80 genes have been identified to cause IPNs, approximately 50% of cases remain genetically unsolved (Rossor et al., 2015; Timmerman et al., 2014). The aim of this PhD was to explore potential causative genes in IPNs to add to the knowledge of the genes and pathways that contribute to these diseases.

In the first part of this project, a candidate gene approach was used to screen genes within the cytoplasmic dynein-dynactin complex. The cytoplasmic dynein-dynactin complex is a microtubule-associated motor protein involved in retrograde transport along axons and is composed of multiple subunits (Pfister et al., 2006; Schiavo et al., 2013). Mutations in the dynein heavy chain (*DYNC1H1*) and dynactin 1 (*DCTN1*) genes have been reported to cause Charcot-Marie-Tooth disease (Weedon *et al.*, 2011) and distal hereditary motor neuropathy (dHMN) (Puls *et al.*, 2005) respectively. We postulated that mutations in other subunits of the dynein-dynactin complex may also play a role in causing neuropathies.

In the second part of the study, a positional cloning approach based on linkage analysis, homozygosity mapping and whole exome sequencing (WES) were used to map the disease-causing gene in a family (CMT861) with genetically unclassified Charcot-Marie-Tooth disease (CMT). All the known IPN genes were excluded (including all known autosomal recessive genes) and this suggested that a novel IPN gene may be responsible for the disease in the CMT861 family.

1.2 Hypothesis

1. Mutations in genes encoding cytoplasmic dynein-dynactin complex can cause IPNs.
2. The CMT phenotype in CMT861 is caused by a novel IPN gene.

1.3 Aims

1. To screen genes within the cytoplasmic dynein-dynactin complex for mutations in IPN patients through a combination of HRM and WES analysis.
2. To identify the causative gene in the CMT861 family using a positional cloning approach through a combination of genome-wide linkage analysis and WES for homozygosity mapping and variant filtering.

1.4 Literature review

1.4.1 Peripheral neuropathies

The peripheral nervous system (PNS) consists of motor and sensory nerves that send signals back and forth between the peripheral organs and the central nervous system for voluntary and involuntary responses. Peripheral neuropathies present with a range of symptoms including numbness and tingling in the hands and feet, shooting pain, sensory loss, muscle weakness and loss of motor function (Cioroiu & Brannagan, 2014). Peripheral neuropathies can be acquired by direct injury to the nerves or due to secondary causes such as autoimmune diseases (Guillain-Barre syndrome) and diabetes mellitus. Inherited forms of peripheral neuropathy (IPN) can be divided into three groups based on the involvement of motor neurons (hereditary motor neuropathy), sensory neurons (hereditary sensory neuropathy) or both motor and sensory neurons (Charcot-Marie-Tooth disease).

1.4.2 Hereditary motor neuropathy (HMN)

Hereditary motor neuropathy, also known as distal spinal muscular atrophy, is predominantly characterised by the loss of function of the distal motor neurons (Harding, 1993). The signs and symptoms of HMN include foot deformities with loss of extensor function of the toes and feet, before progressively affecting the upper limbs. Additional clinical features can be observed in some HMN families including the predominance of muscle wasting in the hands as well as paralysis of the vocal cord and the diaphragm (Harding, 1993; Irobi et al., 2004a).

Distal HMNs (dHMNs) were initially classified into seven subtypes based on the causal genes, mode of inheritance and disease phenotype (Harding, 1993). With the introduction of next generation sequencing (NGS), the identification of new genes has expanded the clinical phenotype of dHMN including dHMN-Jerash, dHMN-ALS and

dHMN-X (X-linked) (Chen et al., 2004; Kennerson et al., 2010; Li et al., 2015). Genes responsible for dHMN are genes within the *HSP* (heat shock protein) family, followed by rarer mutations in *GARS*, *DSMA3*, *BSCL2*, *DCTN1* and *SIGMAR1* (Dubourg et al., 2006; Li et al., 2015; Puls et al., 2005; Viollet et al., 2004; Windpassinger et al., 2004). Table 1.1 gives an overview of the subtypes and genes for dHMN.

Table 1.1: Classification of distal HMN based on mode of inheritance, phenotype and causative gene (Irobi et al., 2004a; Timmerman et al., 2014)

Type	Causal gene	Mode of inheritance
dHMN1	Unknown	Autosomal dominant
dHMN2	*HSPB1, HSPB3, HSPB8*	Autosomal dominant
dHMN3	*DSMA3*	Autosomal recessive
dHMN4	Unknown	Autosomal recessive
dHMN5	*GARS, REEP1*, *BSCL2*	Autosomal dominant
dHMN6	*IGHMBP2*	Autosomal recessive
dHMN7	*DCTN1*, *SLC5A7*	Autosomal dominant
dHMN-Jerash	*SIGMAR1*	Autosomal recessive
dHMN-ALS	*SETX*	Autosomal dominant
dHMN-X	*ATP7A*	X-linked

1.4.3 Hereditary sensory neuropathy (HSN)

Hereditary sensory neuropathy (HSN), or hereditary sensory and autonomic neuropathy (HSAN), is characterised by a progressive loss of sensory neurons. Most patients are not aware of the symptoms until they present with open or infected wound due to the loss of pain sensation or slower healing rates of foot ulcers (Auer-Grumbach et al., 2003; Reilly & Shy, 2009). Certain subtypes of HSN may also include defects in the autonomic neurons which can lead to gastroesophageal reflux, anhydrosis, constipation and irregular heart rhythms (Mroczek et al., 2015). HSN is relatively rare amongst the inherited peripheral neuropathies with an estimated incidence of 1 in 25,000 (Davidson et al., 2012; Mroczek et al., 2015).

HSN is classified into five subtypes (HSN1-5) based on the causative gene, the mode of inheritance and the age of disease onset (Auer-Grumbach et al., 2003; Houlden et al., 2004). To date, 12 genes have been reported to cause the various subtypes of HSN (Table 1.2).

Table 1.2: Classification of HSN based on the mode of inheritance, phenotype and causative gene (Mroczek et al., 2015).

Type	Causal gene	Mode of inheritance
HSN1	*SPTLC1*, *SPTLC2, ATL1*, *DNMT1*, *RAB7*	Autosomal dominant
HSN2	*WNK1, WNK2, FAM234B*, *SCN9A*	Autosomal recessive
HSN3	*IKBKAP*	Autosomal recessive
HSN4	*NTRK1*	Autosomal recessive
HSN5	*NGFB*	Autosomal dominant/recessive

1.4.4 Charcot-Marie-Tooth (CMT) disease

Charcot-Marie-Tooth disease, a peripheral neuropathy involving both the motor and sensory nerves, is named after the three physicians who first described this disorder; Jean-Martin Charcot, Pierre Marie and Howard Henry Tooth (Charcot & Marie, 1886; Tooth, 1886). The worldwide prevalence of CMT is approximately 1 in 2,500 people. It is the most common form of inherited peripheral neuropathy (Brennan et al., 2015b; Skre, 1974).

Clinically, CMT patients present with distal muscle weakness, sensory loss, foot deformities (hammer toe, pes cavus, and high arch) and the absence of reflexes. The classification of CMT has traditionally been based on electrophysiological recordings and a careful examination of the mode of inheritance in the affected families (Harding & Thomas, 1980).

In general, CMT2 is less common than CMT1 (Reilly & Shy, 2009). The phenotypes of CMT1 and CMT2 often overlap and are indistinguishable in the absence of electrophysiological recordings. Patients with nerve conduction velocities lower than 38 m/s are classified as CMT type 1 (CMT1) or 'demyelinating' CMT (Brennan et al., 2015b; Harding & Thomas, 1980), as the myelin sheath is primarily affected. Segmental demyelination, remyelination and 'onion bulb' formations are commonly observed in nerve biopsies of CMT1 patients. Patients with nerve conduction velocities greater than 38 m/s are classified as CMT type 2 (CMT2) or 'axonal' CMT. These patients show axonal degeneration of the neurons with intact myelin sheaths. A third category, an intermediate form of CMT is now recognised, in which affected individuals in the same family have a range of nerve conduction velocities that overlap both CMT1 and CMT2 (Nicholson & Myers, 2006).

All patterns of inheritance are observed in Charcot-Marie-Tooth disease including autosomal dominant, autosomal recessive or X-linked inheritance. The identification of numerous CMT genes has allowed this disorder to be more accurately sub-categorised based on the clinical classification and the mode of inheritance. Figure 1.1 gives an overview of the subtypes of CMT and the reported genes.

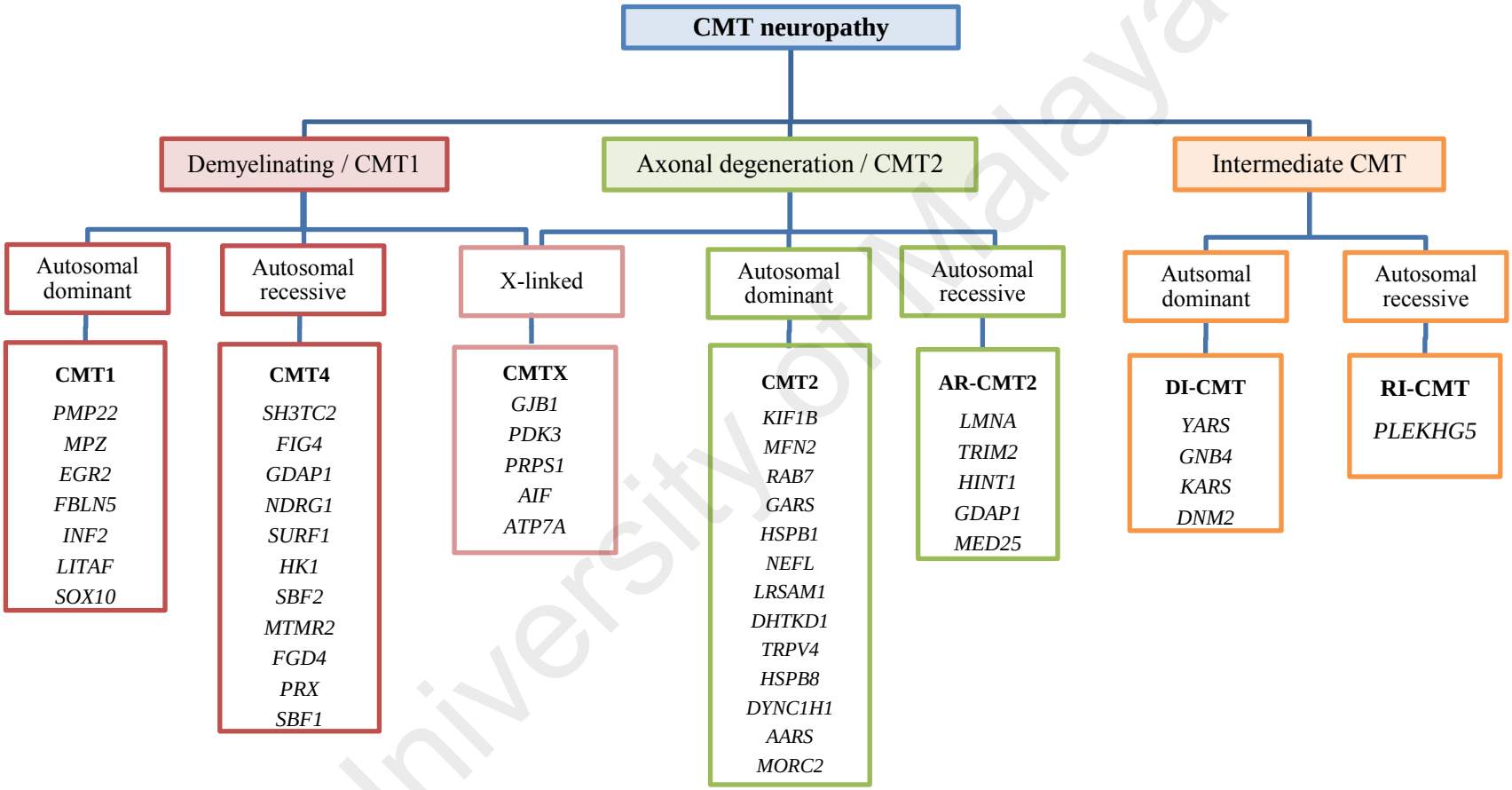


Figure 1.1: Classification of CMT subtypes.

CMT is classified based on nerve conduction velocities and the mode of inheritance. Charcot-Marie-Tooth type 4 (CMT4) is a group of CMT neuropathies distinguished from the other forms of CMT by autosomal recessive inheritance with a demyelinating phenotype. Abbreviation: DI = Dominant intermediate; RI = Recessive intermediate.

1.5 Genetically unclassified IPNs - a potential resource for new gene discovery

In the past two decades, the pace of gene discovery in IPNs was slow due to the limitations in genetic testing technologies at that time. Earlier mapping efforts used positional cloning (detailed in section 1.7.2), which was particularly successful in identifying disease loci in multi-generational families with several affected family members. However, small pedigrees remained genetically unclassified, due to difficulties in achieving significant linkage.

The overall contribution of the known genes in IPNs is approximately 50%, with 90% of these genetically diagnosed cases harbouring mutations in the four common genes, *PMP22*, *MPZ*, *GJB1* and *MFN2* (Rossor et al., 2015; Saporta et al., 2011). Mutations in the other known IPN genes are relatively rare and only account for a limited number of cases. Approximately 80% of CMT1 cases have been genetically diagnosed, however the majority of families with CMT2, HSN and HMN remain genetically unclassified (Rossor et al., 2015). Mutations in the known genes account for only 20% and 15% of the HSN and HMN families, respectively (Dierick et al., 2008; Mroczek et al., 2015; Rossor et al., 2015). A genetic study from the United Kingdom has provided a comprehensive overview of the genetic classification for IPN subtypes (Figure 1.2).

Our laboratory has shown that the frequency of mutations in the four common genes (*PMP22*, *MPZ*, *GJB1*, and *MFN2*) in our Malaysian CMT cohort is similar to the overall global pattern. Approximately 65% of our cohort has been genetically classified, with 35% of the cases remaining genetically unsolved (Shahrizaila et al., 2014). These genetically unclassified cases represent a potential resource for the discovery of new genetic loci for IPN. With the advancement of NGS technologies, it is now possible to revisit the families which are currently genetically unclassified.

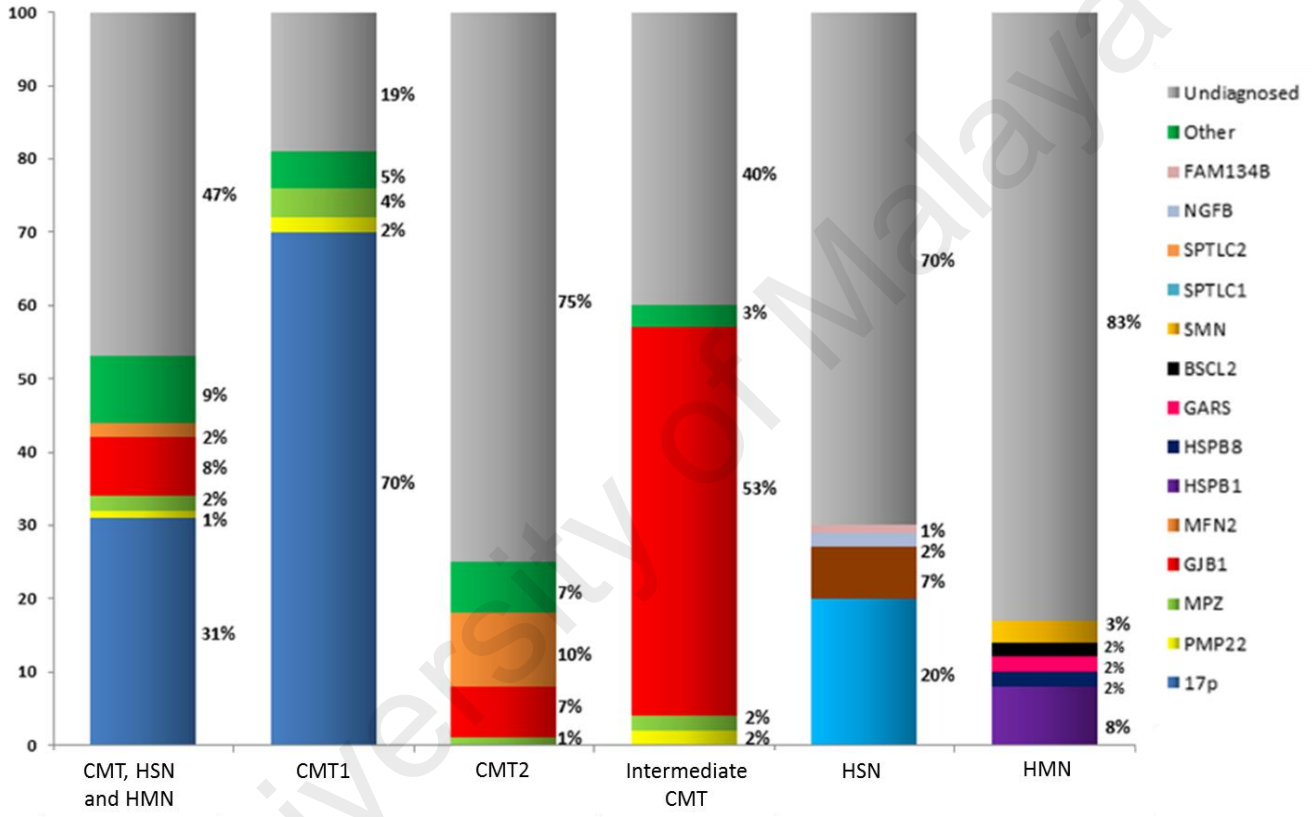


Figure 1.2: Percentage of genetically diagnosed IPNs.

This figure shows the percentage of CMT, HSN and HMN patients that are genetically diagnosed and unclassified in the United Kingdom. Most of the CMT1 patients have been genetically diagnosed; while the majority of CMT2, HSN and HMN patients remain genetically unsolved. Figure adapted from Rossor et al. (Rossor et al., 2015).

1.6 Cellular pathways accounting for peripheral neuropathies

The pathological hallmark of peripheral neuropathies is length-dependent axonal degeneration, a process characterised by the gradual dying back of axons at the neuromuscular junction towards the cell body (Wang et al., 2012). A wide range of biological pathways have been reported to contribute to the pathology of peripheral neuropathies, these include enzyme metabolism (Antonellis et al., 2003; Penno et al., 2010), mitochondria function (Zuchner et al., 2004), protein folding (Evgrafov et al., 2004; Houlden et al., 2008; Irobi et al., 2004b), lipid metabolism (Windpassinger et al., 2004), RNA splicing mechanisms (Cartegni et al., 2006; Prior et al., 2009), growth factors (Baxter et al., 2002; Cuesta et al., 2002; Einarsdottir et al., 2004; Nelis et al., 2002), DNA repair and gene transcription regulation (Albulym et al., 2016), cooper transportation (Kennerson et al., 2010), axonal transportation (Harms et al., 2012; Scoto et al., 2015) and many others. Some of these genes are myelin or neuronal specific, while some of the genes are ubiquitously expressed (Timmerman et al., 2014). Yet, the vast majority of IPN genes converge along common cellular pathways, mainly those involved in maintaining the structure and function of the Schwann cells and axons (Saporta & Shy, 2013).

Schwann cells play a vital role in myelinating the axons of the PNS. The myelin sheath is responsible for the maintenance and regeneration of axons, and to increase the propagation speed of nerve impulses by salutatory conduction (Bhatheja & Field, 2006; Hartline & Colman, 2007). The first IPN gene identified was *PMP22* for a form of CMT1 (CMT1A), and it encodes the peripheral myelin protein 22 (PMP22) which is involved in myelin maintenance (Matsunami et al., 1992; Patel et al., 1992; Timmerman et al., 1992; Valentijn et al., 1992). Duplications of *PMP22* account for approximately 70% of CMT1 cases (Brennan et al., 2015b; Rossor et al., 2015; Saporta et al., 2011). Deletions of this gene can cause hereditary neuropathy with liability to pressure palsy

(HNPP) (Chance et al., 1993; Nicholson et al., 1994). This observation indicated that regulation of *PMP22* gene dosage is essential for proper nerve functions. Apart from that, mutations in *EGR2*, a transcription factor that regulates the expression of genes involved in the formation and maintenance of the myelin sheath also leads to CMT (Warner *et al.*, 1998). Several other mutations in myelin specific genes such as *MPZ*, *SOX10*, *NDRG1*, *PRX* and *GJB1* were also found to cause demyelinating CMT, thus providing insights into the various mechanisms that are involved in myelin membrane homeostasis (Bondurand et al., 2001; Guilbot et al., 2001; Hayasaka et al., 1993; Kalaydjieva et al., 2000; Matsunami et al., 1992; Timmerman et al., 1992; Valentijn et al., 1992; Warner et al., 1998).

The axons for motor and sensory neurons can be over a meter long, reaching from their cell body to the sensory receptors in muscles or skin and to their neuromuscular junctions in the muscles (Gentil & Cooper, 2012). These long nerves require intensive trafficking of factors essential for nerve maintenance between the cell body and the axon terminals (Azzedine et al., 2012). Axonal transport can be disrupted due to ineffective trafficking of motor proteins along the axon or alteration of the cytoskeletal architecture within the axons, leading to axonal degeneration (Lloyd, 2012). Mutations in *DYNC1H1*, an important subunit of the cytoplasmic dynein complex for axonal retrograde transport, have been reported to cause CMT2 (Poirier et al., 2013; Strickland et al., 2015; Weedon et al., 2011). In addition, *NEFL* [neurofilament protein (Jordanova *et al.*, 2003)], *LMNA* [laminin (De Sandre-Giovannoli *et al.*, 2002)], and *HSPB1* [small heat shock protein (Evgrafov *et al.*, 2004)], are involved in neurofilament homeostasis and the architectural organisation of the axonal cytoskeleton, which when mutated, affect axonal transport and cause CMT (De Sandre-Giovannoli et al., 2002; Jordanova et al., 2003).

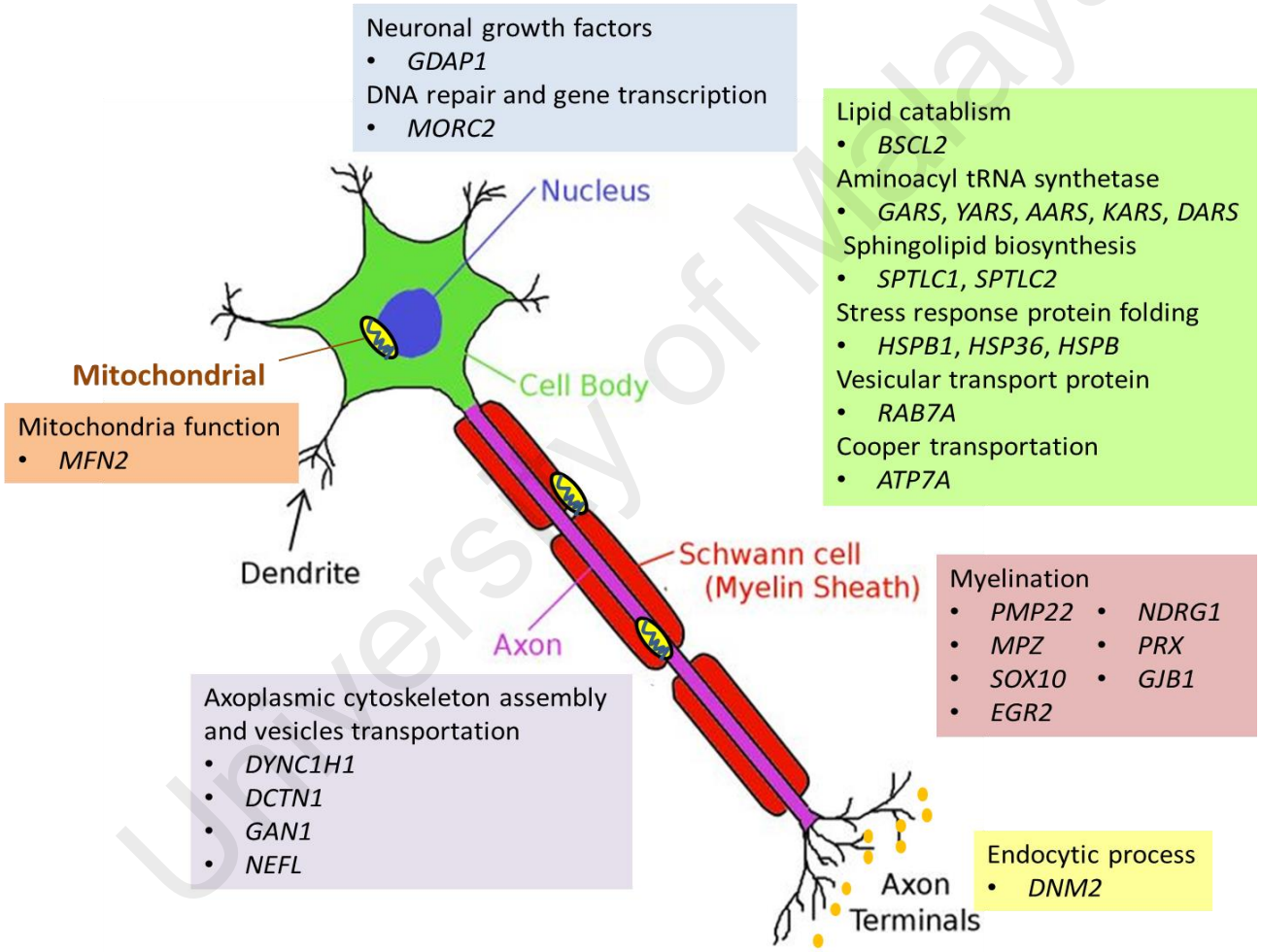
The transportation of mitochondria is dependent on fully operational axonal transport, and is particularly important because mitochondria are needed along the axon (Saxton & Hollenbeck, 2012). Mitofusin 2 is a protein that is localised in the endoplasmic reticulum and is involved in promoting mitochondria fusion (de Brito & Scorrano, 2008; Eura et al., 2003). The fusion is important for mitochondrial transport via the microtubule-based kinesin and dynein transport system (Misko et al., 2010). Mutations in mitofusin 2 *(MFN2)* have been shown to compromise mitochondrial function and dynamics leading to axonal CMT (Baloh et al., 2007; Kijima et al., 2005; Zuchner et al., 2004). At present, *MFN2* is the most commonly affected gene in CMT2 (Barisic et al., 2008; Rossor et al., 2015; Zuchner et al., 2004).

The diversity of molecular and cellular functions of proteins that are associated with different forms of IPN has provided insights into numerous pathomechanisms of the disease and the physiology of the peripheral nervous system. Figure 1.3 shows the cellular pathways that are affected in IPNs.

Figure 1.3: Cellular pathways accounting for peripheral neuropathies.

Genes with diverse cellular and molecular functions have been associated with peripheral neuropathies. These genes play a role in: myelination [eg: *PMP22*, *MPZ*, *SOX10*, *EGR2*, *NDRG1*, *PRX* and *GJB1* (Bondurand et al., 2001; Guilbot et al., 2001; Hayasaka et al., 1993; Kalaydjieva et al., 2000; Matsunami et al., 1992; Timmerman et al., 1992; Valentijn et al., 1992; Warner et al., 1998)]; mitochondria function [eg: mitofusin 2 (*MFN2*) (Zuchner et al., 2004)]; stress response protein folding [eg: small heat shock proteins (*HSPB1*, *HSP36* and *HSPB8*) (Evgrafov et al., 2004; Houlden et al., 2008; Irobi et al., 2004b; Kolb et al., 2010; Tang et al., 2005)]; cellular metabolism enzymes [eg: aminoacyl tRNA synthetase (*GARS*, *YARS*, *AARS*, *KARS*, *DARS*) (Stum *et al.*, 2011) (Antonellis *et al.*, 2003), serine palmitoyltransferase 1 and 2 (*SPTLC1*, *SPTLC2*) (Penno *et al.*, 2010)]; lipid catabolism [eg: Seipin (*BSCL2*) (Windpassinger et al., 2004)]; cooper transportation [eg: ATPase cooper transporting alpha (*ATP7A*) (Kennerson et al., 2010)]; neuronal growth factors [eg: ganglioside induced differentiation associated protein 1 (*GDAP1*) (Baxter et al., 2002; Nelis et al., 2002; Niemann et al., 2005)]; DNA repair and gene transcription regulation [eg: *MORC2* (Albulym et al., 2016)]; axoplasmic cytoskeleton assembly [eg: Neurofilament light chain (*NEFL*) (Jordanova et al., 2003; Mersiyanova et al., 2000), giant axonal neuropathy 1 (*GAN1*) (Zemmouri *et al.*, 2000)]; cellular vesicle transportation [eg: cytoplasmic dynein heavy chain 1 (*DYNC1H1*) (Harms et al., 2012; Weedon et al., 2011), dynactin subunit 1 (*DCTN1*) (Puls et al., 2005), Ras-related protein 7A (*RAB7A*) (Verhoeven, K. et al., 2003)].

Figure 1.3 (continued)



1.7 Strategies for gene discovery

Multiple strategies can be performed in gene discovery, such as the candidate gene approach, positional cloning and more recently NGS technologies. Often, results from these approaches are combined to increase the chance of disease gene identification and to provide supportive evidence towards the result obtained. In this study, these approaches were undertaken with the aim to discover new genes in genetically undiagnosed IPN samples.

1.7.1 Candidate gene screening

In a candidate gene approach, the genes selected for analysis typically have a known biological function that is directly or indirectly involved in the particular phenotype (Zhu & Zhao, 2007). Once a disease gene has been identified for a trait, other genes that are functionally relevant or structurally interact with the disease gene are good candidates to be screened for mutations, as they are probably involved in the same cellular pathway.

For example, mutations in *HSPB8*, a gene that belongs to the small heat shock protein (HSP) family, was identified to cause dHMN via positional cloning (Irobi et al., 2004a). Following the identification of *HSPB8*, the screening of other genes encoding HSP proteins identified mutations in *HSPB1* and *HSPB3* as a cause of axonal CMT and dHMN respectively (Evgrafov et al., 2004; Irobi et al., 2004b; Kolb et al., 2010). Similarly, GARS mutations were reported in families with CMT2 and dHMN (Antonellis et al., 2003) and the candidate gene approach led to the identification of mutations in other aminoacyl tRNA synthetase genes such as *AARS*, *YARS* and *KARS*, which cause CMT2, intermediate CMT and dHMN respectively (Jordanova et al., 2006; McLaughlin et al., 2012; McLaughlin et al., 2010).

1.7.2 Positional cloning strategy

Positional cloning is a powerful method to identify causative genes for Mendelian disorders. This strategy starts with the identification of a disease locus based on linkage analysis in families with multiple affected individuals across several generations, followed by mutational analysis of candidate genes in the affected individuals (Puliti et al., 2007). For this approach, information of genetic markers genotyped in the family members is used for linkage analysis in order to define a chromosomal region where the disease gene is likely to be located.

The principle of linkage analysis relies on recombination events that occur between homologous chromosomes during meiosis and the co-segregation of the genetic markers with the disease trait within families in which the pattern of inheritance is known. Recombination events can be detected by genotyping genetic markers and analysing the haplotype blocks that segregate in the family. The genetic markers are DNA sequence variations in the genome such as microsatellite and single nucleotide polymorphisms (SNPs). Highly informative genetic markers allow marker phase to be assigned, and recombinant and non-recombinant meiosis to be identified in family individuals. To quantitate the informativeness of a marker, the heterozygosity value is calculated based on the number and frequency of alleles at the locus. This value can be interpreted as the probability that an individual randomly selected from the population will be heterozygous at the locus. Microsatellite markers are short sequences of 2-6 base pairs in length which are tandemly repeated. The number of tandem repeats can vary between individuals making them useful to genotype individuals and profile their genomes. SNP markers are less informative than microsatellites, as there are only 2 alleles at a marker locus. However, due to the ease of genotyping SNPs, they are routinely used for linkage analysis as power can be gained by haplotyping groups of SNPs for the linkage analysis. The availability of SNP-genotyping arrays have allowed

high density or high throughput genotyping of SNPs across the genome to detect haplotype blocks for linkage analysis (Purcell *et al.*, 2007).

A statistical test generating a logarithm of odds (LOD) score is used to determine if the disease locus is linked to the markers tested or whether the linkage has occurred by chance (Morton, 1955). The LOD score is a function of the distance between the marker and the disease allele. The genetic distance is expressed in centimorgan (cM), and is dependent on the recombination fraction (θ) (Morton, 1955). The recombination fraction is the probability of a crossover occurring between two markers. If the genetic distance between two loci is one centimorgan, there is (by definition) a 1% probability of a crossover occurring between the two loci and this is equivalent to a recombinant fraction of 0.01 (θ = 0.01) (Kosambi, 1943; Morgan, 1915; Morton, 1955). Free recombination (50%) occurs when two loci are far away from each other on one chromosome or situated on different chromosomes. This is equivalent to a recombination fraction of 0.5. Recombination values less than 50% suggest that the loci are on the same chromosome at a genetic distance equivalent to the θ value. The LOD score (Z value) calculates the ratio of the likelihood of two assumptions; 1) the loci are linked ($0 < \theta < 0.5$) and 2) the loci are not linked ($\theta = 0.5$) (Morton, 1955). Genotype information from a pedigree is used to estimate the recombination frequency for the calculation of a LOD score, using the equation below:

$$\text{LOD score} = Z = \log_{10} \frac{\text{probability of linkage}}{\text{probability of no linkage}} = \log_{10} \frac{(1-\theta)^{NR} \text{ X } \theta^{R}}{0.5^{(NR+R)}}$$

R denotes the number of recombinant offspring observed in the pedigree, NR denotes the number of non-recombinant offspring observed in the pedigree and θ is a nominated recombinant fraction. An example for the identification of recombinant events in a pedigree is shown in Figure 1.4. The Z score is calculated at a range of

recombination fractions (θ = 0 to 0.5), giving the best estimation of distance between the putative disease locus and a known genetic marker. By convention, a LOD score greater than or equal to +3.3 indicates a high evidence of linkage; the odds are 1000:1 in favour of linkage (Lander & Kruglyak, 1995). A LOD score ≤ -2 indicates evidence for the exclusion of linkage; the odds are 100:1 against linkage (Morton, 1984). LOD scores between these limits are inconclusive. However, LOD scores of more than 1 are considered suggestive of linkage and warrant further investigation (Lander & Kruglyak, 1995; Pericak-Vance, 2001).

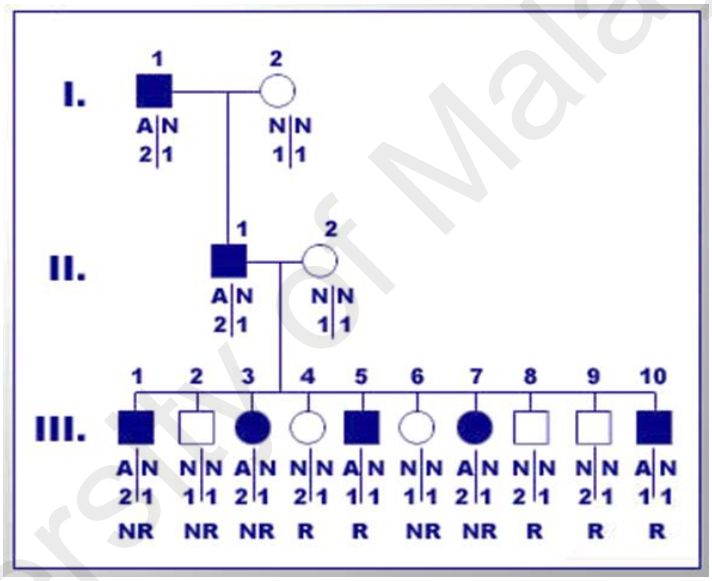


Figure 1.4: Recombinant and non-recombinant meiotic events in a pedigree.

The family segregating a disorder has been genotyped for a marker with 2 alleles. The 'A' and 'N' represent the disease and normal allele respectively. Individual II:1 is considered to be phase known. Therefore, in generation III, the recombination events can be unequivocally assigned. In this case, 5 of the offspring are recombinant, and another 5 are non-recombinant. This example is adapted from the online linkage course of the University of Miami (http://hihg.med.miami.edu/).

Before the completion of the genomic sequencing project, positional cloning involved the isolation of overlapping DNA segments from genomic libraries, progressively cloned and sequenced, to identify candidate genes within the linkage region (Royer-Pokora et al., 1986). Since the completion of human genome sequencing

in 2001, linkage analysis can be performed at a greater resolution (Venter et al., 2001). Genes within a linkage interval can be identified by browsing through web portals such as the UCSC Genome Browser (University of California, Santa Cruz) and the National Center of Biotechnology Information (NCBI) Genome Browser. As there are likely to be many genes within a linkage interval, candidate genes with relevant function related to the disease are often prioritised for mutational analysis.

1.7.3 Next generation sequencing

Next generation sequencing technologies enable rapid massively parallel sequencing of DNA samples and has been used extensively to elucidate the genetic aetiology of various Mendelian disorders, as reviewed elsewhere (Bamshad et al., 2011; Gonzaga-Jauregui et al., 2012; Ku et al., 2011). In the context of IPNs, gene discoveries using NGS have increased by 40% in the last few years (Timmerman et al., 2014). NGS strategies include whole genome sequencing (WGS), whole exome sequencing (WES), targeted gene panel sequencing and transcriptome sequencing, depending upon the study's framework.

NGS technologies have uncovered new possibilities in gene discovery. They offer the opportunity to study small families and isolated patients which has previously not been possible. Prior to the use of NGS, linkage analyses on small families usually produced multiple suggestive linkage loci with maximum LOD scores below the value of significant linkage, indicating that all but one of these loci were false positive results of the analysis. These loci with low LOD score were time- and cost- consuming to be followed up with conventional Sanger sequencing, leaving the small families genetically unsolved (Ott et al., 2015). The use of NGS in families with low LOD scores now facilitates the follow up of variants within suggestive linkage regions which can be readily queried and examined. Also, the ability of NGS to sequence the genome

of a person allows the identification of mutations that are specific to an individual, and through sequencing trios (parents and the affected child), *de novo* mutations can be identified.

Whole exome sequencing is used to assess the coding region of the genome (Teer & Mullikin, 2010). The justification of using WES as a gene discovery strategy is based on the finding that the majority of mutations identified in Mendelian disorders result in the disruption of protein-coding sequences (Choi et al., 2009; Stenson et al., 2009). Furthermore, in comparison to WGS, WES has a lower computational and analysis burden. In spite of that, it remains challenging to identify causative mutations amongst the massive amount of variants detected via WES. As NGS has become more cost effective, many genome or exome sequences are being deposited in the public repositories such as the dbSNP (Sherry *et al.*, 2001), the HapMap (International HapMap Project, 2003), the Exome Variant Server (EVS) (Exome Variant Server), the 1000 genomes (1000 Genome Project Consortium, 2010) and the Exome Aggregation Consortium (ExAC) databases. This allows researchers to exchange information and query the variants of interest; and the vast majority of common variants can be excluded from being pathogenic. Conversely, the massive amount of sequence variants deposited in the public repositories has made the justification for ruling out variants becoming more difficult, especially for autosomal recessive disorders in which a variant is not pathogenic in the heterozygous state and will only be deleterious in the homozygous state (Brunham & Hayden, 2013). As a result, many studies for autosomal recessive disorders have set a frequency threshold for variants found in these database, where only variants with minor allele frequency < 1% are considered as candidate for pathogenicity investigation (Bamshad et al., 2011; McDonald et al., 2012).

There are several advantages for using WES to study families with IPN. First of all, IPN is a genetically heterogeneous group of disorders (Skre, 1974). Patients with mutations in the same gene can show different severities of the phenotype (allelic heterogeneity); while patients with the same phenotypes can have mutation in different genes (locus heterogeneity). Therefore, the strategy of combining WES data from unrelated patients to find the common causative gene mutation is not feasible. The factor of genetic heterogeneity that can confound gene identification can be solved by performing family-based heuristic WES variant filtering. This approach assumes that the affected family members carry the exact same mutation and the unaffected family members will be the best genetic control for common variant elimination. Secondly, using the genotype information extracted from the WES data of a family, linkage analysis can be performed (Abecasis et al., 2002; Purcell et al., 2007). This helps to focus analysis of the variants located within the suggestive linkage regions that have been identified by family-based heuristic WES variant filtering.

As new NGS platforms continue developing and sequencing is becoming more affordable, the shift in research towards analysing small families or isolated patients is now a highly plausible approach for novel gene discovery. The incorporation of linkage analysis into NGS is a powerful tool for gene mapping especially in the studies of Mendelian's disorders. In this study, both the NGS and positional cloning approach were applied with the aim to discover novel disease-causing gene in a cohort of genetically unclassified IPN patients.

CHAPTER 2: GENERAL MATERIALS AND METHODS

This chapter details the general materials and methods used throughout this project. Further details of specific protocols are detailed in the relevant chapters.

2.1 Patient consent

Patient ascertainment and collection of blood samples for DNA analysis in Malaysia was performed according to the protocols approved by the University of Malaya Medical Centre (UMMC) Ethics committee. In Australia, patient ascertainment and collection of blood samples for DNA analysis protocols were approved by the Sydney Local Health District, Human Ethics Committee, Concord Hospital, Australia. Written consent was obtained from all patients and control individuals involved in this study.

2.2 DNA extraction

Patient and control DNA samples collected at the Neurogenetics Laboratory, Faculty of Medicine, University of Malaya (Malaysia) were extracted using the QIAamp DNA blood Maxi kit (Qiagen). The genomic DNA samples collected at the Molecular Medicine Laboratory, Concord Repatriation General Hospital (Sydney, Australia) were extracted by staff using the Gentra Puregene blood kit (QIAGEN).

2.2.1 DNA extraction using the QIAamp DNA blood Maxi Kit (QIAGEN)

In brief, 3 mL of whole blood was added to 500 µL of Qiagen protease in a 50 mL centrifuge tube and 1X PBS was added to a final volume of 5 mL. Cells were lysed by adding 6 mL of lysis buffer (Buffer AL provided), mixed vigorously for 1 min followed by 10 min incubation at 70°C in a water bath. DNA was precipitated by adding 5 mL of absolute ethanol to the sample and mixed to homogeneity. The DNA precipitate was collected by pouring the sample onto the membrane of the QIAamp maxi column, centrifuged for 3 min at 1850 x g. The column membrane containing the DNA was

washed with 5 mL of Buffer AW1 and centrifuged at 4000 x g for 1 min followed by another wash with 5 mL of Buffer AW2 and centrifuged at 4000 x g for 15 min. To elute the DNA, 600 µL of distilled water was added onto the membrane, incubated at room temperature for 5 min, followed by centrifugation at 4000 x g for 2 min. The DNA was collected and stored at 4°C.

2.3 Whole exome sequencing (WES)

In this study, 25 µL of DNA with concentration of 200 ng/µl was prepared for each sample that was sent for WES. WES was out-sourced to Axeq Technologies (Seoul, Korea). The service provided by Axeq Tehnologies includes sequencing library preparation, exome enrichment and sequencing using the Illumina HiSeq 2000 platform (Illumina, Inc.). A brief description of protocols used by the company is described in the following sections.

2.3.1 Sequencing library preparation

The sequence libraries were prepared using the Illumina TrueSeq DNA Sample Prep system (Illumina, Inc.) following the manufacturer's protocols. In brief, the DNA was fragmented into small pieces and the Illumina adapters were ligated to the DNA fragments. DNA fragments within the size of 300-400 base pairs were selected for PCR amplification to enrich the adaptor ligated DNA fragments. The Bioanalyzer (Agilent Technologies) was used to assay the concentration and size of DNA fragments.

2.3.2 Exome capture, sample enrichment and sequencing

The sequence libraries were pooled for enrichment using the Illumina TrueSeq Exome Enrichment system (Illumina, Inc.) following manufacturer's protocols. In brief, DNA libraries were hybridised to capture probes bound to streptavidin bead. The bound libraries were washed three times to remove non-specific binding fragments. The enriched libraries were eluted from the capture probes and PCR amplified to further

enrich the adaptor ligated DNA fragments. Automated sequencing was carried out using the Illumina HiSeq 2000.

2.3.3 Sequence analysis using bioinformatics methods

Initial bioinformatics was provided by Axeq Technologies includes sequence read-mapping and variant annotation. The sequence data was aligned to the human genome reference (GRCh37/hg19) assembly using BWA software (Li & Durbin, 2010). Duplicated reads were discarded using Picard (http://picard.sourceforge.net/). Sequence variants including SNP or indel (insertion or deletion) were called using SAMTOOLS (Li et al., 2009) and SNPs were annotated with dbSNP 144, and the 1000 Genomes call release 2013 using ANNOVAR (Wang et al., 2010).

2.4 Primer design

Primers were designed using the LightScanner primer design software version 1.0 (Idaho Technology Inc.) or the Primer3 online software (Rozen & Skaletsky, 2000). The sequences of the reference genes were obtained from the UCSC Genome Browser. The primers were designed so that the melting temperature ranged between 55°C to 65°C and primer lengths were between 17 to 30 base pairs (bp). For HRM analysis, the primers were designed so that the amplicon size ranged from 70 to 350 bp; for HRM SNP genotyping, the amplicon size ranged from 70 to 110 bp. The selected primers were re-examined using the BLAT tool in the UCSC Genome Browser to ensure that the primers target specifically to the region of interest. All primer sequences are listed in Appendix section A.

2.5 Polymerase chain reaction (PCR)

For HRM analysis purposes, double-stranded DNA (dsDNA) was amplified using HRM master mix (Idaho Technology); for Sanger sequencing or microsatellite marker size fragmentation purposes, dsDNA was amplified using Immomix master mix

(Bioline). PCR was carried out in 10 µL reaction volumes containing 4 µL of HRM master mix or 5 µL Immomix master mix, with 1 µL of 10 µM of forward and reverse primers, and 10 ng of DNA. When the PCR produced non-specific amplification, or the amplified product was faint, 1 µL of PCR enhancer (Invitrogen) was added to increase the efficiency of the PCR.

PCR thermocycling was performed on either Eppendorf vapo.protect Mastercycler pro (Eppendrof), Mastercycler ep Gradient S (Eppendorf) or Veriti Thermocycler (Applied Biosystems). The thermal cycling protocol includes 3 steps: initial denaturation at 95°C for 10 min followed by 30 cycles of 95°C for 30 secs; optimal annealing temperature for 30 secs; and a final extension at 72°C for 5 min. Alternatively, a touchdown PCR protocol was used for optimisation. The touchdown PCR protocol includes: initial denaturation at 95°C for 10 min; 10 cycles of 95°C for 30 secs, 72°C for 30 secs with a 1.5°C decrease in annealing temperature per cycle and 72°C 30 secs; 25 cycles of 95°C for 30 secs, 55°C for 30 secs and 72°C for 30 secs; and a final extension at 72°C for 5 min.

Different sets of primers were optimised under a gradient of annealing temperatures ranging from 50 to 72°C. The optimal annealing temperature was determined by comparing the brightness of the bands after agarose gel electrophoresis. PCR products were size fractioned on 1-2% (w/v) agarose gels. The agarose gels were prepared by dissolving the agarose powder (Molecular grade, Vivantis) into 1X TAE buffer (Amresco) with heating using a microwave oven. Molten agarose was allowed to cool to 65°C before adding SYBR Safe DNA gel stain (Invitrogen), then poured into a horizontal gel tray and allowed to solidify. Gel electrophoresis was performed at 40 Volts/cm for 35-40 min, and visualised using the Safe Imager 2.0 Blue-Light

Transilluminator. All samples were run with a size standard using either HyperLadder I or HyperLadder IV (Bioline).

2.6 High resolution melting (HRM) analysis for mutation screening and SNP genotyping

For mutation screening using HRM, primers for amplification were designed flanking the exons of genes, with maximum amplicon size not more than 350 bp. For HRM SNP genotyping, primers were designed flanking the variant of interest with an amplicon size between 70 to 110 bp. DNA samples were amplified in a 96-wells HRM plate (Bio-Rad) using HRM master mix. Each sample was duplicated. Prior to scanning in the LightScanner (Idaho Technology), the PCR products were overlaid with 10 µl mineral oil to avoid evaporation during the melting procedure. The melting temperatures ranged from 62°C to 98°C with an incremental rate of 0.1°C/s. The HRM master mix contains the fluorescent dye LC Green Plus (Idaho Technology), which binds to dsDNA. In principle, each PCR sample is identified by its unique melting temperature (Tm), which corresponds to the temperature at which half of the dsDNA is denatured. HRM relies on detecting the formation of heteroduplexes in a heterozygous sample (Gundry et al., 2003; Palais et al., 2005; Wittwer et al., 2003). Heteroduplexes are formed when wild-type DNA strand anneals with the mutant DNA strand. Essentially the presence of heteroduplex DNA changes the stability of the DNA resulting in a different Tm when compared to the wild-type DNA.

The melt profile data were analysed using the small amplicon option in the LightScanner analysis call-IT 2.0 software (Version 2.0.0.1331; Idaho Technology). The acquired melt curves were displayed as derivative fluorescent curve plots. The fluorescence data was normalised by temperature shifting on either side of the melt transition. By shifting the sample melt curves along the X-axis, this ensured all the

curves passed through a common temperature point and eliminated slight temperature variations between wells and facilitated grouping of the melt curves. After normalising, the melt curves were converted into normalised derivative fluorescent curve plots. Amplicons with differential melt profile were sent for Sanger sequencing at Australian Cancer Research Foundation (ACRF), Garvan Institute, Sydney. The sequence data were then aligned to the human genome reference sequence (GRCh37/hg19) to identify possible variants.

2.7 DNA Sanger sequencing and analysis

2.7.1 Sanger sequencing

In Australia, Sanger sequencing was outsourced to the Australian Cancer Research Foundation (ACRF), Garvan Institute; in Malaysia, Sanger sequencing was outsourced to MyTACG Bioscience Enterprise.

2.7.2 Sequence analysis

Sequencing traces were provided in ab1 format. The primer sequences and reference sequences were converted into seq files using the EditSeq software (version 5.08; DNAStar). The human genome reference sequence (GRCh37/hg19) from USCS genome browser was used for sequence alignment. The trace files and seq files were aligned to the reference sequence using either SeqMan II version 5.08 (DNASTAR) and Sequencher version 5.2.4 (Genecode).

CHAPTER 3: INVESTIGATING THE CYTOPLASMIC DYNEIN-DYNACTIN COMPLEX FOR A PATHOGENIC ROLE IN INHERITED PERIPHERAL NEUROPATHIES

In this study, a candidate gene screening approach was undertaken to investigate whether genes encoding the cytoplasmic dynein-dynactin complex have a pathogenic role in inherited peripheral neuropathies (IPNs). In neurons, this complex plays a specific role for the retrograde transport of membrane-bound vesicles along microtubules (Hirokawa, 1998). Animal models and human studies have implicated several genes within this complex, namely the cytoplasmic dynein heavy chain gene (*DYNC1H1*) and dynactin subunit 1 (*DCTN1*) for a range of neurodegenerative phenotypes including axonal Charcot-Marie-Tooth disease, spinal bulbar muscular atrophy, and intellectual disability (Harms et al., 2012; Weedon et al., 2011; Willemsen et al., 2012). However, *DYNC1H1* and *DCTN1* are only two subunits within this large multi-subunit protein complex and there are not many studies on the other subunits. These other subunits act together with *DYNC1H1* and *DCTN1* to maintain neuronal survival and function by ensuring proper transport of essential trophic and signaling factors.

3.1 Introduction

The cytoplasmic dynein and dynactin complexes interact with microtubules and play a diverse range of cellular functions, including mitotic spindle and centrosome assembly, nuclear positioning during mitosis and the transportation of membrane bound organelles such as endosomes, lysosomes, autophagosomes and mitochondria (Allan, 2011; Egan et al., 2015; Eschbach & Dupuis, 2011; Pfister et al., 2006). In neurons, the dynein-dynactin complex is involved in neuronal migration and transporting cargo such as signalling factors within axons and dendrites (Hirokawa, 1998; Schiavo et al., 2013).

The dynactin complex also stabilises and enhances the motor dynamics of the dynein motor complex through added binding to microtubules (Cantalupo et al., 2001; Culver-Hanlon et al., 2006; King & Schroer, 2000).

The cytoplasmic dynein is a ~1.5 MDa protein complex composed of 10 subunits including the heavy chain (*DYNC1H1*), intermediate chains (*DYNC1I1*, *DYNC1I2*), light intermediate chains (*DYNC1LI1*, *DYNC1LI2*) and light chains (*DYNLL2*, *DYNLL3*, *DYNLT1*, *DYNLT3*, *DYNLRB1*), (Figure 3.1), (Pfister et al., 2006). The dynactin complex is a ~1.2 MDa protein complex composed of 8 subunits including Dynactin 1-6 (*DCTN1*-6) and the Arp1 complex which includes the Arp1 filaments (*ACTR1A* and *ACTR1B*), (Figure 3.1), (Eckley et al., 1999; Schafer et al., 1994).

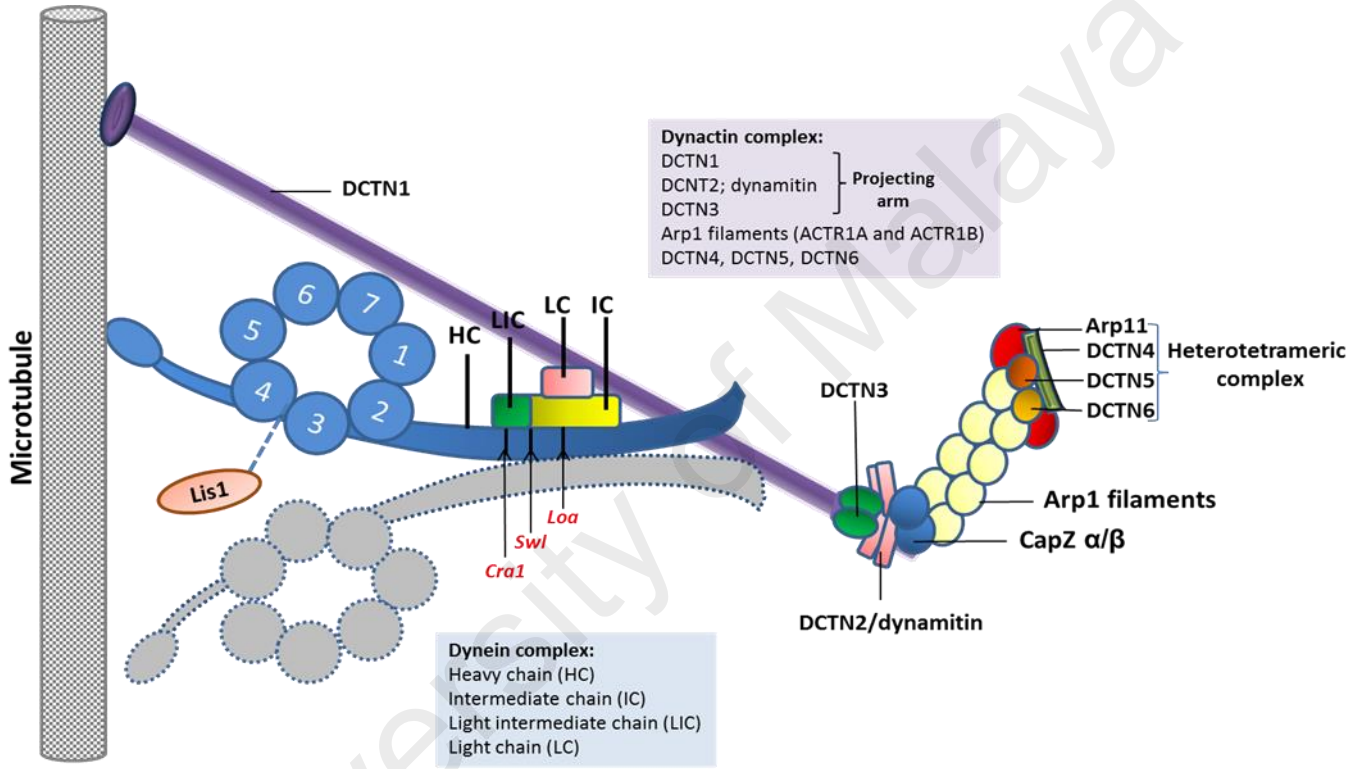


Figure 3.1: Schematic of the cytoplasmic dynein-dynactin complex.

The cytoplasmic dynein complex is composed of a heavy chain (HC) dimer, intermediate chains (IC), light intermediate chains (LIC) and light chains (LC). Locations of mutations identified in mutant mice (*Cra1*, *Swl* and *Loa*) with peripheral neuropathy phenotypes are annotated in red. The heavy chain has two globular heads with an elongated stalk which bind to microtubules for dynamic movements. The intermediate and light intermediate chains bind to the homodimerisation region of the heavy chains, and the light chains bind to intermediate chains. The dynactin complex is structurally divided into 2 parts: the projecting side arm consisting of the DCTN1, DCTN3 and DCTN2 (dynamitin) subunits, and the Arp1 rod which includes the Arp1 filaments (ACTR1A and ACTR1B) and the heterotetrameric complex composed of DCTN4, DCTN5, DCTN6 and Arp11. The dynactin complex and LIS1 are the binding partners of the cytoplasmic dynein complex. Dynein and dynactin complexes bind to each other via the coiled-coil domain of the DCTN1 subunit and the cytoplasmic dynein intermediate chains.

3.1.1 Evidence for a pathogenic role of genes encoding the dynein-dynactin complex in neurodegenerative diseases

Several lines of evidence have clearly shown that mutations in genes within the dynein-dynactin complex can result in neurodegenerative phenotypes. The majority of this evidence comes from studies on a selected number of genes as discussed below.

3.1.1.1 Heavy chain (*DYNC1H1*)

The cytoplasmic dynein heavy chain (DYNC1H1) forms a dimer, that folds into two globular heads comprising of six ATPase rings and two microtubule-interacting stalks at the C-terminal. These structures form the motor domain that drives the entire complex and its cargos along the microtubules, (Figure 3.1), (Kikkawa, 2013; Neely et al., 1990). The dimerization of heavy chains occurs at the N-terminal, and the intermediate chains and light intermediate chains bind to this region, (Figure 3.1), (Habura et al., 1999; Tynan et al., 2000). The light chains and dynactin 1 bind to the intermediate chains, (Figure 3.1), (Lo et al., 2007; Makokha et al., 2002; Susalka et al., 2002; Vaughan & Vallee, 1995).

Homozygous *Dync1h1* knockout mice are embryonically inviable and display distinct disorganisation of the Golgi complex and improper distribution of lysosomes and endosomes. This indicates that Dync1h1 is an essential protein for housekeeping functions including mitotic spindle and centrosome assembly, nuclear positioning, transportation of endosomes and Golgi complex maintenance (Harada *et al.*, 1998). Surprisingly, heterozygous *Dync1h1* knockout mice do not have a gross phenotype. (Harada et al., 1998).

Evidence of a role in neurodegenerative diseases came from several animal models with mutations within the *Dync1h1* homodimerisation domain including Legs at Odd Angles (*Loa*), Cramping 1 (*Cra1*) and Sprawling (*Swl*) mice. Homozygote $Dync1h1^{Loa}$

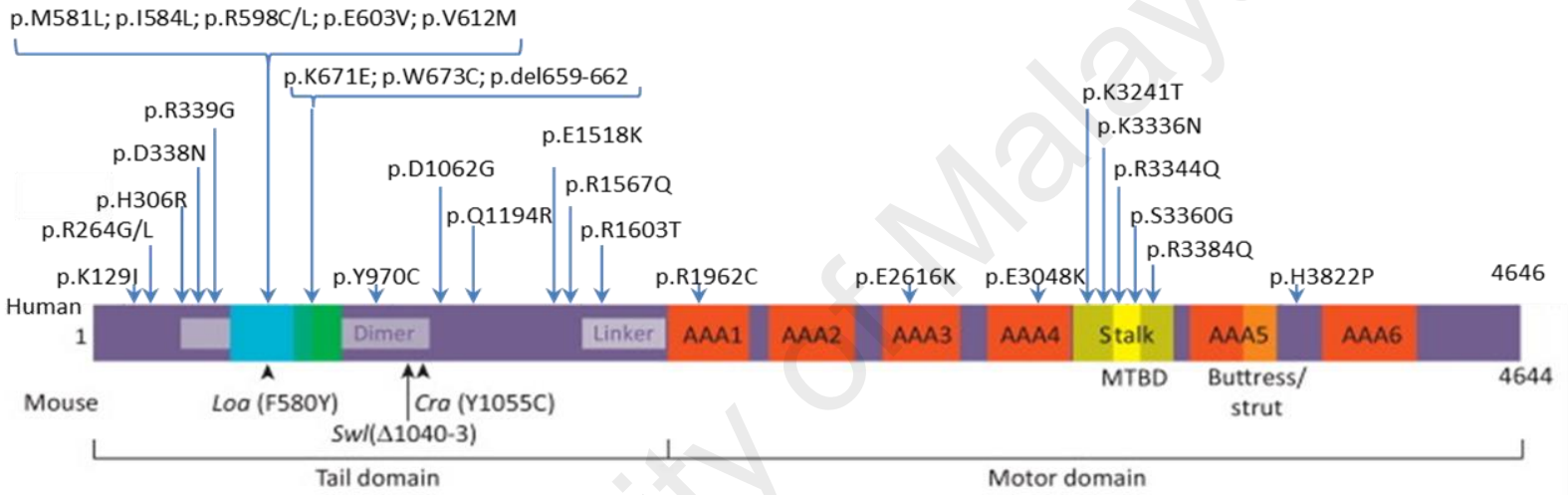
and *Dync1h1*^{Cra1} mice (with point mutations) do not survive after birth (Hafezparast *et al.*, 2003) and homozygote *Dync1h1*^{Swl} mice (with 9-bp deletion) die at early gastrulation stages (Duchen, 1974). Heterozygote *Loa*, *Cra1* and *Swl* mice are viable but display progressive sensory and motor neuron degeneration (Chen et al., 2007; Dupuis et al., 2009; Hafezparast et al., 2003). The phenotype observed in the heterozygote *Loa*, *Cra1* and *Swl* mice is not observed in the heterozygous knockout mice possibly because the heterozygous knockout mice were studied before neurodegeneration was apparent.

Next generation sequencing facilitated the discovery of *DYNC1H1* mutations in a family with mental retardation and gait abnormalities (Vissers et al., 2010). Since then, several other *DYNC1H1* mutations have been reported in families with diseases such as severe intellectual disability with malformation of cortical development, axonal CMT and spinal muscular atrophy predominantly affecting the lower extremities, (Figure 3.2), (Fiorillo et al., 2014; Harms et al., 2012; Peeters et al., 2015; Poirier et al., 2013; Punetha et al., 2015; Scoto et al., 2015; Strickland et al., 2015; Tsurusaki et al., 2012; Weedon et al., 2011; Willemsen et al., 2012). Mutations that span both the tail and the motor domains of *DYNC1H1* can cause different phenotypes with central and/or peripheral nervous system involvement, indicating a broad spectrum of roles of the cytoplasmic dynein-dynactin complex (Figure 3.2). Based on the reported phenotypes, mutations in the motor domain are likely to cause more severe phenotypes.

Figure 3.2: A schematic diagram of the DYNC1H1 protein domains and the reported mutations up to 2015.

DYNC1H1 is encoded by 4646 amino acids (aa), divided into two domains; the tail domain (aa 1-1867) and the motor domain (aa 1868-4646). The tail domain consists of the homodimerisation region (aa 300-1140), the intermediate chains binding site (aa 448-703), and the light intermediate chains binding site (aa 651-802). The motor domain consists of six ATPase rings – AAA1 (aa 1868-2099), AAA2 (aa 2180-2452), AAA3 (aa 2556-2805), AAA4 (aa 2899-3168), the stalk (aa 3189-3532), AAA5 (aa 3552-3782) and AAA6 (aa 4005-4221). The reported human mutations on the heavy chain are illustrated above the schematic structure while the reported mouse mutants with peripheral neuropathy are shown below. Thirty mutations have been reported to date which cause phenotypes ranging from cortical development malformation to peripheral neuropathies. Abbreviation: MCD = malformation of cortical development; SMA = spinal muscular atrophy; LED = lower extremities predominance; CMT = Charcot-Marie-Tooth neuropathy. Adapted from Schiavo *et al*., 2013 (Schiavo et al., 2013) and includes reported mutations up to 2015.

Figure 3.2 (continued)



***DYNC1H1* mutations**	**Phenotypes**	**Key references**
p.K129I; p.R1962C; p.K3241T; p.R3344Q	MCD	Poirier et al., 2013
p.D338N; p.M581L; p.I584L; p.R598C; p.R598L; p.E603V; p.V612M; p.K671E; p.W673C; p.E2616K; p.S3360G	SMA-LED	Harms et al., 2012; Peeters et al., 2015; Scoto et al., 2015; Strickland et al., 2015
p.R264L	Axonal CMT	Peeters et al., 2015
p.D1062G	Spastic tetraplegia	Strickland et al., 2015
p.R264G; p.R339G; p.Y970C; p.Q1194R; p.E1518K; p.R1567Q; p.R1603T; p.E3048K; p.K3336N; p.R3384Q; p.H3822P	MCD and SMA-LED	Vissers et al., 2010; Harms et al., 2012; Willemsen et al., 2012; Poirier et al., 2013; Fiorillo et al., 2014; Scoto et al., 2015
p.H306R	MCD; SMA-LED; Axonal CMT	Weedon et al., 2011; Tsurusaki et al., 2012
p.del659-662	MCD and spastic tetraplegia	Poirier et al., 2013

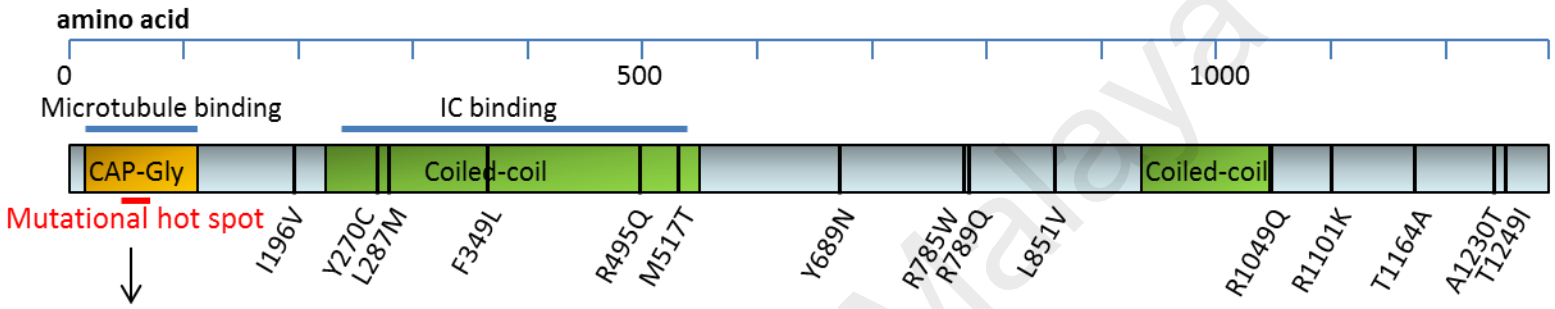
3.1.1.2 Dynactin 1 (*DCTN1*)

Dynactin 1 (DCTN1) is the largest subunit of the dynactin complex. It is composed of a CAP-Gly domain and two coiled-coiled domains, (Figure 3.3), (King, Stephen J. et al., 2003). The CAP-Gly domain binds directly to the microtubules and also interacts with several microtubule-binding proteins (EB1 and CLIP-170) to regulate microtubule dynamics, (Berrueta et al., 1999; Goodson et al., 2003; Ligon et al., 2003). The first coiled-coil domain of DCTN1 is the interaction site for the cytoplasmic dynein intermediate chains (King et al., 2003; Schroer, 2004).

A mutation in *DCTN1* (p.G95S) was first described in a family with distal spinal and bulbar muscular atrophy (Puls *et al.*, 2003). Since then, *DCTN1* has been associated with amyotrophic lateral sclerosis (ALS), frontal temporal dementia (FTD), Parkinson's disease (PD), supranuclear palsy-like phenotypes and Perry syndrome (Araki et al., 2014; Caroppo et al., 2014; Farrer et al., 2009; Munch et al., 2005; Munch et al., 2004; Newsway et al., 2010; Puls et al., 2003; Puls et al., 2005), (Figure 3.3). Interestingly, despite the phenotypic heterogeneity, all the reported pathogenic variants are located in the CAP-Gly domain of DCTN1 (Figure 3.3).

DCTN1 binds to Rab7 proteins, which regulate endocytic membrane trafficking (Cantalupo et al., 2001; Johansson et al., 2007; Jordens et al., 2001) and the trafficking of the TrkA nerve growth factor and peripherin, which are crucial for neurite outgrowth and axonal repair (Cogli et al., 2013; Deinhardt et

al., 2006; Saxena et al., 2005). Mutations in *RAB7* have been reported to cause Charcot-Marie-Tooth type 2B, where the underlying pathomechanism is still under debate (Cogli et al., 2013; Spinosa et al., 2008; Verhoeven, K. et al., 2003).



Mutation	Phenotype	Key reference
F52L	Late onset parkinsonism and frontotemporal atrophy	**Araki et al. 2014**
G59S	Distal spinal and bulbar muscular atrophy	**Puls et al. 2003**
G71R/E/A T72P Q74P	Perry Syndrome	**Farrer et al. 2009**
G71E K56R	Progressive supranuclear Palsy-like phenotype (PSP)	**Caroppo et al. 2014, Gustavsson et al. 2015**

Figure 3.3: A schematic diagram of the DCTN1 protein domains and reported mutations up to 2015.

The CAP-Gly domain (orange block; aa 29-95) of DCTN1 is the binding site for microtubules and also appears to be a mutational hot spot (listed in the box). The first coiled-coil domain (green block; aa 217-548) is the binding site for cytoplasmic dynein intermediate chains (IC), the second coiled-coil domain is located from aa 926-1049. Non-pathogenic variants previously reported are shown in black.

3.1.1.3 Dynactin 2 (*DCTN2*)

One of the first lines of evidence linking the dynein-dynactin complex with motor neuron degeneration came from a study of mice overexpressing *Dctn2*. These mice showed late-onset progressive motor neuron degeneration and skeletal muscle atrophy (LaMonte et al., 2002). Excess Dctn2 causes the dissociation of the dynactin complex, and destabilises the links between the projecting arm (Dctn1, Dctn2 and Dctn3) and the Arp1 rod (Melkonian et al., 2007), which renders the dynactin complex non-functional. Recently, a *DCTN2* mutation was identified in an intermediate CMT family, associating this gene in human disorders for the first time (Braathen *et al.*, 2015). In addition, the DCTN2 subunit is an adaptor protein for Bicaudal D (BICD1 and BICD2) and mutations in *BICD2* have been reported in congenital spinal muscular atrophy and hereditary spastic paraplegia (Hoogenraad et al., 2001; Neveling et al., 2013; Oates et al., 2013).

3.1.1.4 Intermediate chain 1 (*DYNC1I1*)

Recently, *DYNC1I1* mutations were reported in families with split hand-split foot malformation and intellectual disability (Tayebi *et al.*, 2014). Structurally, the cytoplasmic dynein intermediate chains bind to the heavy chain within the homodimerisation region, and act as an adaptor for other subunits such as the dynein light chains and the dynactin complex (Susalka et al., 2002; Vaughan & Vallee, 1995). In *Drosophila*, *Dic* mutations cause larval lethality and an abnormal crawling phenotype which is likely due to axonal transport abnormalities (Boylan et al., 2002). Similar to DCTN1, dynein intermediate chains also interact with Rab7 protein in neurons (Mitchell *et al.*, 2012). As mutations in *RAB7* have been reported to cause CMT2 (Verhoeven, Kristien et al., 2003), this suggests that there may be valid reasons for investigating genes encoding dynein intermediate chains in CMT and other peripheral neuropathies.

3.1.1.5 Light intermediate chains (*DYNC1LI1* and *DYNC1LI2*)

Dync1li1 mutant mice display increased anxiety and have abnormal neuronal morphology coupled with a significant reduction of myelinated sensory nerve recordings (Banks et al., 2011). Although the *Dync1li1* mutant mice do not show obvious defects in the developing peripheral nerves, it is possible that the mutation may give rise to a late-onset peripheral neuropathy which was not examined in these mice (Banks et al., 2011). Similar to the dynein intermediate chains and *DCTN1*, the light intermediate chains are also the binding partners of the Rab protein family including Rab4a (Bielli *et al.*, 2001). In human, RAB4A interacts with NDRG1 protein, which is encoded by the *NDRG1* gene, a reported CMT gene (Hunter *et al.*, 2003). This network of interconnected pathways involving dynein light intermediate chains and the *RAB* family of genes is an interesting avenue to further explore for IPNs.

3.1.1.6 The other subunits: Light chains (*DYNLL, DYNLT* and *DYNLRB*), Dynactin subunit 3-6 (*DCTN3, DCTN4, DCTN5, DCTN6*) and Arp1 filaments

There is little known about the other subunits in the complex, although there is evidence that some subunits are involved in neuronal function such as Dynll1 and Dynlt1 in neurite outgrowth (Chuang et al., 2001; Phillis et al., 1996). In humans, the DCTN4 subunit interacts with the copper transporter ATP7B, which is involved in copper homeostasis (Lim et al., 2006). P-type ATPases have been linked to a number of diseases of copper dysregulation including Menkes disease (*ATP7A* mutations) and Wilson's disease (*ATP7B* mutation), (Bull et al., 1993; Tanzi et al., 1993; Vulpe et al., 1993). Recently, *ATP7A* mutations were reported in an X-linked distal hereditary motor neuropathy which causes impaired copper trafficking (Kennerson et al., 2010). It is therefore plausible that mutations in *DCTN4* may disrupt copper homeostasis in neurons as well.

3.1.2 Hypothesis

Based on the role of the cytoplasmic dynein-dynactin complex in retrograde axonal transport in neurons, we hypothesise that mutations in genes encoding components of the complex will cause peripheral neuropathy.

3.1.3 Aims

The overall aim is to identify mutations in the cytoplasmic dynein-dynactin complex. The specific aims will be to perform systematic mutation screening of genes encoding components of the cytoplasmic dynein-dynactin complex (dynein heavy chain, dynein intermediate chains, light intermediate chains, light chains, dynactin subunits 1-6, actin-related protein homolog A and B) in a genetically unsolved IPN cohort using a combination of whole exome sequencing (WES) and high resolution melting (HRM) technologies.

3.2 Methodology

3.2.1 Patient selection

IPN families were recruited through the Neurogenetics Clinic in Concord Hospital, Sydney, and University of Malaya Medical Centre (UMMC), Kuala Lumpur according to ethics protocols approved by the respective Human Ethics Committees. Index patients from 136 families were diagnosed with different types of hereditary peripheral neuropathies based on clinical presentation and included 76 CMT patients, 29 HSN patients and 31 HMN patients, all of whom had multiple family members who were affected. The index patients were further sub-classified into: demyelinating CMT (CMT1); axonal degenerative CMT (CMT2); congenital or juvenile onset CMT (CMT3); autosomal recessive CMT (CMT4); CMTX and HMNX families which were classified as 'query X-linked' since there was no male to male transmission and the females had a less severe phenotype; HMN with pyramidal signs (HMNP); adult-onset HSN (HSN1); and young-onset HSN patients (HSN2). All index individuals were excluded for mutations in the most common IPN genes (*PMP22*, *MPZ*, *MFN2* and *GJB1*). WES was performed on 106 patients and data was queried to exclude mutations in each family for the remaining known IPN genes. The screening strategy was designed to query the WES variant data for all the dynein-dynactin genes from the 106 index cases, and to screen the non-exomed index patients for the same genes using HRM analysis. The numbers of patients in each subtype are listed in Table 3.1.

Table 3.1: Cohort of IPN families for cytoplasmic dynein-dynactin genes screening.

A total of 106 patients were screened through WES, and 30 patients were screened with HRM analysis. CMT: Charcot-Marie-Tooth disease; HSN: Hereditary Sensory Neuropathy; HMN; Hereditary Motor Neuropathy; X-: query X-linked.

Peripheral Neuropathy	**Patient Number and Mutation Screening Method**	
	WES	**HRM**
CMT (n=76)		
CMT1 (4)	4	-
CMT2 (43)	43	-
CMT4 (1)	-	1
CMT5 (3)	2	1
CMTX (25)	25	-
HSN (n=30)		
HSN1 (28)	1	27
HSN2 (1)	-	1
Other (n=31)		
HMN (14)	14	-
HMNP (15)	15	-
HMNX (2)	2	-

3.2.2 Mutation screening

Genomic DNA was isolated from whole blood using the PureGene DNA isolation kit (Qiagen) and QIAamp DNA blood Maxi kit (Qiagen) according to the manufacturer's recommendations (Chapter2, section 2.2).

Sixteen genes (Table 3.2) including intermediate chains (*DYNC1I1*, *DYNC1I2*), light intermediate chains (*DYNC1LI1*, *DYNC1LI2*), light chains (DYNLL1, *DYNLL2*, *DYNLT1*, *DYNLT3*, *DYNLRB1*), dynactin subunits 2-6 (*DCTN2*, *DCTN3*, *DCTN4*, *DCTN5*, *DCTN6*) and actin-related protein (*ACTR1A*, *ACTR1B*) underwent mutation screening by either WES or HRM. Due to the large size of the *DYNC1H1* and *DCTN1* they were only examined in samples that had undergone WES.

Table 3.2: Genes encoding the components of the cytoplasmic dynein-dynactin complex.

This table shows the chromosome position, reference gene ID and the exon number of genes encoding the cytoplasmic dynein-dynactin complex.

Name	Symbol	Chromosome position	Gene ID	No. of exons
Cytoplasmic dynein complex				
Cytoplasmic dynein heavy chain 1	*DYNC1H1*	14q32	NM_001376.4	78
Cytoplasmic dynein intermediate chain 1	*DYNC1I1*	7q21.3-22.1	NM_004411	17
Cytoplasmic dynein intermediate chain 2	*DYNC1I2*	2q31.1	NM_001378.1	18
Cytoplasmic dynein light intermediate chain 1	*DYNC1LI1*	3p22.3	NM_016141.3	13
Cytoplasmic dynein light intermediate chain 2	*DYNC1LI2*	16q22.1	NM_006141.2	13
Cytoplasmic dynein light chain 1 (LC8 type 1)	*DYNLL1*	12q24.23	NM_003746	3
Cytoplasmic dynein light chain 2 (LC8 type 2)	*DYNLL2*	17q22	NM_080677.2	3
Cytoplasmic dynein light chain Roadblock 1	*DYNLRB1*	20q11.21	NM_014183	4
Cytoplasmic dynein light chain Tctex 1	*DYNLT1*	6q25.2	NM_006519.2	5
Cytoplasmic dynein light chain Tctex 3	*DYNLT3*	Xp21	NM_006520	6
Dynactin complex				
Dynactin 1 (p150)	*DCTN1*	2p13	NM_004082	32
Dynactin 2 (p50/dynamitin)	*DCTN2*	12q13.3	NM_006400	14
Dynactin 3 (p22/p24)	*DCTN3*	9p13	NM_024348	6
Dynactin 4 (p62)	*DCTN4*	5q31-32	NM_001135643.1	13
Dynactin 5 (p25)	*DCTN5*	16p12.2	NM_032486	6
Dynactin 6 (p27)	*DCTN6*	8p12-p11	NM_006571	7
Actin-related protein 1A	*ACTR1A*	10q24.32	NM_005736	11
Actin-related protein 1B	*ACTR1B*	2q11.1-11.2	NM_005735	11

3.2.3 Whole exome sequencing (WES) analysis

Whole exome sequencing was outsourced to Axeq Technologies in South Korea. Sequencing library preparation, sequence data alignment and variant calling of WES data is described in Chapter 2 (section 2.3). Variants identified by WES were validated by Sanger sequencing.

3.2.4 High resolution melting (HRM) analysis for variant detection

Primers flanking individual exons were designed using the LightScanner primer design software 1.0 (Idaho Technology Inc.), (Chapter 2, section 2.4). HRM amplicons were designed (< 350 bp) to obtain optimum melt curve profiles. Primer information for each exon can be found in the Appendix (section A). The HRM melting and analysis protocols are previously described in Chapter 2 (section 2.6). Samples with distinct melt profiles were sent for Sanger sequencing for variant identification. The protocol for sequence analysis after Sanger sequencing is described in Chapter 2 (section 2.7).

3.2.5 Segregation analysis to evaluate the transmission of a variant within a pedigree

Non-synonymous variants or indels within the coding regions that had a minor allele frequency (MAF) < 1% were selected for segregation analysis in additional family members using Sanger sequencing. All common and synonymous variants were not tested for segregation. If the variant was absent in one of the other affected family member(s) or present in unaffected family member(s), it was considered unlikely to be pathogenic.

3.3 Results

In total, 14 variants were identified within the cytoplasmic dynein subunits and 11 variants were identified within the dynactin subunits in the screened cohort.

3.3.1 Variants in *DYNC1H1*

Four variants were detected in *DYNC1H1* in which two were synonymous and two were non-synonymous variants (Table 3.3). The synonymous variants were predicted to be likely non-pathogenic and further validation was not performed. The first non-synonymous variant is a novel variant (p.D2566A) detected in the heterozygous state in a CMT2 patient. This variant was located within the AAA3 ATPase ring of the motor domain. Family segregation analysis showed that the novel variant is absent in three affected family members, and is therefore likely to be a rare non-pathogenic polymorphism. The second non-synonymous variant is a reported common SNP, rs10129889 (p.H4029Q), (Table 3.3). This variant was detected in 7 CMT and 4 HMN families in the heterozygous state. Segregation studies were not performed for the common SNPs.

3.3.2 Variants in *DYNC1I1* and *DYNC1I2*

Four variants were identified in *DYNC1I1* and none in *DYNC1I2*. Two of the *DYNC1I1* variants were reported synonymous SNPs, (rs3757697 and rs1048666), (Table 3.3). As these variants were common and did not change the encoded amino acids, they were deemed unlikely to be pathogenic. Heterozygous non-synonymous variants were identified in a CMTX (rs369627655; p.T176M) and a CMT5 (rs35314029; p.L199M) patient, (Table 3.3). The p.T176M variant is located in the dynein intermediate chain dimerization domain while the p.L199M variant resides in the postulated binding domains for DYNLT (Tctex dynein light chain) and DYNLL (LC8 dynein light chain). These variants have a low MAF of 0.01% and 0.97%

respectively. However, segregation analysis excluded the variants as likely to be pathogenic. An example of variant detection using HRM analysis is shown in Figure 3.4.

3.3.3 Variants in *DYNC1LI1* and *DYNC1LI2*

Four nonsynonymous variants in *DYNC1LI1* were identified in the heterozygous state. One variant rs2303857 is a common polymorphism, detected in CMT, HSN and HMN families (Table 3.3). The three remaining variants (rs143775988, rs138677120 and rs116075738) are reported SNPs with a MAF less than 1.0%, (Table 3.3). The variant rs143775988 (p.G77R) was detected in a HSN1 patient. The variant rs138677120 (p.V139I) was detected in CMT2, CMTX, HSN1 and HMNP patients. The rs116075738 (p.V146A) was detected in a HMN patient. Segregation analysis showed that rs116075738 and rs138677120 did not segregate with the disease phenotype in the respective families and therefore deemed non-pathogenic. For rs143775988, segregation analysis was not possible as additional DNA samples from family members were not available.

One variant was identified in *DYNC1LI2*, (rs10990). The variant was synonymous and predicted to be non-pathogenic (Table 3.3).

(a) (b)

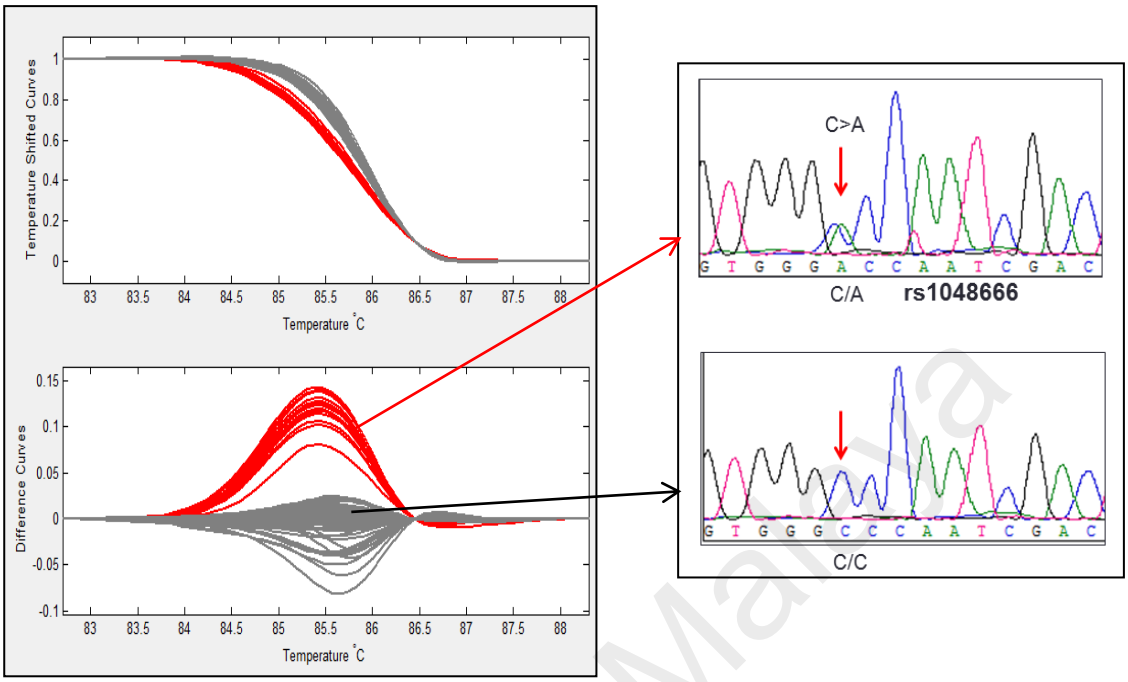


Figure 3.4: Representative high resolution differential melt curve profiles for rs1048666 SNP in exon 14 of *DYNC1I1*.

(a) Two distinct melt curves are shown in red and grey. Samples were screened in duplicates. Normalised temperature shifted curves (top panel) and difference curves (bottom panel) are shown. (b) Sequence chromatogram of representative sample from each melt profile group. The sequence analysis identified a reported SNP in the Human Genome (hg19) database. The samples that grouped in the grey melt profile were homozygous for the reference C allele and the red group represented the heterozygous C/A genotype.

3.3.4 Variants in *DYNLL1, DYNLL2, DYNLRB1, DYNLT1, DYNLT3*

Cytoplasmic dynein light chains are composed of 5 genes; *DYNLL1*, *DYNLL2*, *DYNLRB1*, *DYNLT1* and *DYNLT3*. Among these 5 genes, only one synonymous variant (rs11771) was identified in *DYNLT3* (Table 3.3). It was not pursued further.

3.3.5 Variants in *DCTN1*

Two rare heterozygous nonsynonymous variants in *DCTN1* were identified, rs72466496 (p.R495Q) and rs17721059 (p.T1249I), (Table 3.3). The p.R495Q variant, located within the binding site between dynactin and the cytoplasmic dynein intermediate chains, was detected in CMT2 and HMN patients. The p.T1249I variant, which was identified in a CMT1 patient, did not lie within any domain on DCTN1. Segregation analysis in the respective families excluded these variants from being pathogenic.

3.3.6 Variant in *DCTN2*

A variant was found in *DCTN2* in a CMT1 patient (rs2292656; p.A255A), (Table 3.3) but was excluded from further investigation as it was a common polymorphism.

3.3.7 Variants in *DCTN3*

A rare nonsynonymous variant was detected in *DCTN3*, (rs200578089; p.P139A) with a MAF of 0.05% (Table 3.3). This variant was present in the heterozygous state in a HMN patient and did not segregate when tested in additional family members.

3.3.8 Variants in *DCTN4*

Three variants were identified in *DCTN4* (Table 3.3). One variant was a common polymorphism, rs11954652 (p.F349L) identified in CMT, HSN and HMN patients. The second variant, rs781156279 (p.I233T) was a rare SNP (MAF of 0.006%), detected in a CMTX patient. The third variant, rs35772018 (p.Y270C) had a MAF of 0.52%. This

variant was detected in the heterozygous state in CMT and HMN patients. Segregation analysis showed that the variants did not segregate with the disease in the respective families.

3.3.9 Variants in *DCTN5* and *DCTN6*

A common synonymous variant (rs11545871; p.A10A) was identified in the *DCTN5* gene (Table 3.3). No variants were detected in *DCTN6*.

3.3.10 Variants in *ACTR1A* and *ACTR1B*

No variant was identified in *ACTR1A.* Three variants were detected in *ACTR1B*, including a synonymous variant, rs1042705 (p.V268V), and two common nonsynonymous variants, rs79616802 (p.M45V) and rs11692435 (p.A143V), (Table 3.3). Segregation analysis was not performed as these variants were either common or had no effect on the encoded amino acids.

Table 3.3: Summary of variants identified in the cytoplasmic dynein-dynactin genes.

This table shows variants that have been reported in other disease cohorts are marked with an asterisk.

Gene	SNP ID	Nucleotide change	Amino acid change	Segregation analysis	MAF (db144)	No. of individuals with the variant in CMT families	No. of individuals with the variant in HSN families	No. of individuals with the variant in HMN families
DYNC1H1	rs3818188	c.G624A	p.P208P	Not tested (Synonymous variant)	18.77%	1/75	-	-
	rs17541158	c.C7449T	p.I2483I	Not tested (Synonymous variant)	2.85%	1/75	-	-
	Novel	c.A7697C	p.D2566A	No segregation	-	1/75	-	-
	rs10129889	c.C12087A	p.H4029Q	Not tested (Common polymorphism)	11.8%	7/75	-	2/31
DYNC1I1	rs3757697	c.A465G	p.Q155Q	Not tested (Synonymous variant)	10.13%	-	5/30	1/31
	rs369627655	c.C527T	p.T176M	No segregation	0.01%	1/75	-	-
	rs35314029	c.C595A	p.L199M	No segregation	0.97%	1/75	-	-
	rs1048666	c.C1491A	p.G497G	Not tested (Synonymous variant)	24.02%	-	6/30	-
DYNC1LI1	rs143775988	c.G229A	p.G77R	Unavailable for testing	0.16%	-	1/30	-
	rs138677120	c.G415A	p.V139I	No segregation	0.96%	3/76	3/30	1/31
	rs116075738	c.T437C	p.V146A	No segregation	0.06%	-	-	1/31
	rs2303857	c.A830G	p.Q277R	Not tested (Common polymorphism)	5.2%	12/75	4/30	5/31

Table 3.3 (continued)

DYNC1LI2	rs10990	c.A1236G	p.V412V	Not tested (Synonymous variant)	1.18%	-	2/30	-
DYNLT3	rs11771	c.C183T	p.A61A	Not tested (Synonymous variant)	33.35%	1/75	-	-
DCTN1	rs17721059	c.G1484A	p.R495Q*	No segregation	1.51%	4/75	-	2/31
	rs72466496	c.C3746T	p.T1249I*	No segregation	0.36%	2/75	-	1/31
DCTN2	rs2292656	c.A765C	p.A255A	Not tested (Synonymous variant)	8.26%	1/75	-	-
DCTN3	rs200578089	c.C415G	p.P139A	No segregation	0.05%	-	-	1/31
DCTN4	rs781156279	c.T668C	p.I223T	No segregation	0.006%	1/75	-	-
	rs35772018	c.A809G	p.Y270C*	No segregation	0.52%	2/75	-	1/31
	rs11954652	c.C1047G	p.F349L*	Not tested (Common polymorphism)	27.01%	6/75	1/30	2/31
DCTN5	rs11545871	c.A306T	p.A10A	Not tested (Synonymous variant)	9.37%	-	2/30	1/31
ACTR1B	rs79616802	c.A133G	p.M45V	Not tested (Common polymorphism)	14.43%	12/75	4/30	4/31
	rs11692435	c.C428T	p.A143V	Not tested (Common polymorphism)	4.97%	14/75	7/30	3/31
	rs1042705	c.C804G	p.V268V	Not tested (Synonymous variant)	33.1%	1/75	12/30	1/31

3.4 Discussion

This study used a candidate gene approach to test the hypothesis that mutations in the cytoplasmic dynein-dynactin complex causes an IPN. The complex is encoded by 18 genes, representing a high burden for mutation scanning using the conventional Sanger sequencing method. Thus, a combination of WES and HRM was used to screen the genes in 136 IPN patients.

The WES dataset was examined to identify variants in the genes encoding the cytoplasmic dynein-dynactin. The WES platform used in this study was the Illumina HiSeq 2000 which generated 100 bp paired end reads with a median depth of coverage of 50X. The sequence for 106 IPN patients achieved a mean read depth of > 30X for the targeted regions (Drew et al., 2015). Thus, we estimate that our WES data had > 90% sensitivity to detect heterozygous variants within the targeted exome regions (Choi et al., 2009). Samples that were not sent for WES were screened using HRM analysis. The HRM method enables the detection of minor changes in the melting temperature due to mismatching of heteroduplexes when the dsDNA undergoes melting through a range of temperatures. This technique effectively reduced the sequencing load by only selecting samples for sequencing in which differential melt profiles were observed. In this study, all of the HRM primers were designed with an amplicon size of less than 350 bp in order to achieve 100% sensitivity and specificity in detecting heterozygous variants and 96% of the homozygous variants (Reed et al., 2007).

In total, twenty-five variants were detected in the cytoplasmic dynein-dynactin genes. Four of the variants have been previously reported in other disease cohorts and have been excluded from a pathogenic role in our IPN cohort.

There are over 30 mutations reported thus far in *DYNC1H1* but none of these reported mutations were detected in our IPN cohort. A novel nonsynonymous variant

(p.D2566A) which was located in the AAA3 ATPase ring was identified in this study. The AAA3 and AAA4 ATPase rings are the binding sites for the Lis1 protein, which can also binds to microtubules (Huang et al., 2012). Lis1 increases the binding affinity of the dynein complex to microtubules to enhance the motor processivity along the microtubules (Huang et al., 2012). Mutations in *LIS1* cause classical lissencephaly which is a severe brain development disorder in humans (Dobyns et al., 1993; Lo Nigro et al., 1997; Saillour et al., 2009). Although a mutation nearby (p.E2616K) was reported to cause SMA-LED with cognitive impairment (Scoto et al., 2015), the novel variant identified in our study was excluded from a pathogenic role in IPN as it did not segregate with the disease in the respective families. Perhaps, this domain contains a specific motif that is responsible for the LIS1 protein binding and the p.D2566A variant is not within this motif.

There are over nine mutations reported in *DCTN1* thus far, but none of these were found in our cohort. Two previously reported variants (p.R495Q and p.T1249I) were identified. The p.R495Q variant was first reported in a candidate gene screening approach of *DCTN1* in a multiple sclerosis case-control study with 200 participants. Its pathogenicity was excluded as it was observed equally in cases and controls (Munch *et al.*, 2007). In our study, the pathogenicity of this variant was excluded as it did not segregate with the phenotype in the patients' family. The second *DCTN1* variant, p.T1249I was first reported in a patient with amyotrophic lateral sclerosis (ALS) with no family history of motor neuron disease (Munch et al., 2004). Several candidate gene screening studies of *DCTN1* in the cohort of multiple sclerosis (Munch et al., 2007), Parkinson's diseases (PD) and frontotemporal lobar degeneration (FTLD) (Vilarino-Guell *et al.*, 2009) have excluded its pathogenicity in a total number of 1186 patients due to the lack of segregation with the phenotype in the families. Similarly, we ruled out pathogenicity for the variant in our families. These variants together with the other

reported variants on *DCTN1* (Figure 3.3) show no pathogenic effect probably because they are located outside of the CAP-Gly domain of *DCTN1* which is the mutational hotspot.

It is interesting to note that mutations within the CAP-Gly domain can manifest in different disease phenotypes including parkinsonism and frontotemporal atrophy, distal spinal and bulbar muscular atrophy, supranuclear palsy-like phenotypes and ALS (Araki et al., 2014; Caroppo et al., 2014; Farrer et al., 2009; Munch et al., 2005; Munch et al., 2004; Newsway et al., 2010; Puls et al., 2003; Puls et al., 2005). Possibly, this domain contains several motifs that are responsible for different cellular roles. For example, the mutation responsible for distal spinal and bulbar muscular atrophy (p.G59S) is likely to affect microtubule binding affinity and retrograde transport (Puls et al., 2005). Evidence in *Drosophila* with the p.G38S mutation (orthologue of p.G59S in human) show accumulation of transmembrane proteins and endosomes at the neural tips indicating retrograde transport defects (Lloyd et al., 2012). Mutations reported in Perry Syndrome (p.G71A/E/R, p.T72P, p.Q74P) fall within the highly conserved GKNDG motif which is involved in protein-protein interaction (Moughamian & Holzbaur, 2012). Mutations in this motif could potentially disrupt the interaction of dynactin with proteins such as CLIP170 and EB which are the linkers between vesicles and microtubules (Weisbrich et al., 2007).

In *DCTN4*, the two variants (p.Y270C and p.F349L) identified have been previously reported in a study on *Pseudomonas aeruginosa* infection in cystic fibrosis patients (Emond *et al.*, 2012). DCTN4 acts as a cargo adaptor, linking ubiquinated protein aggregates to the autophagic machinery along the microtubules to be degraded by autophagy (Pankiv et al., 2007). Emond and group hypothesised that isoforms of *DCTN4* might inhibit or decrease the activity of autophagic machinery in the airway,

making cystic fibrosis patients with *DCTN4* mutations more susceptible to *Pseudomonas aeruginosa* infections (Emond et al., 2012). In addition, DCTN4 was found to co-localise with many ubiquinated protein aggregates in neurodegenerative diseases (Nixon, 2006), and it interacts with one of the copper-transporting ATPases, ATP7B (Lim et al., 2006). Interestingly, mutations in another copper-transporting ATPase, *ATP7A,* have been found to cause X-linked HMN (Kennerson et al., 2010). In this study, both p.F349L and p.Y270C variants were excluded from being pathogenic in our cohort.

Amongst the five genes that encode the cytoplasmic light chains, only one synonymous variant in *DYNLL1* was detected in our cohort. The function of the light chain has not been fully explored; however a role in neurodegenerative diseases has been suggested in animal models. Mutations in the *Dlc* (*LC8*) gene were found to disrupt the sensory neuron projections in *Drosophila* and neurite outgrowth was reported to be disrupted by *Dynlt1* mutations in rats (Chuang et al., 2001; Phillis et al., 1996).

There were no variants detected in *ACTR1A* while three variants which are common or synonymous were detected in *ACTR1B.* These genes encode Arp1 filaments of the dynactin complex. The Arp1 filaments are able to bind several isoforms of βIII-spectrin of various cell membranes, bringing different membrane cargos in close proximity to Rab7 (Holleran et al., 1996). The Arp1 filaments, together with DCTN1, form the Arp1-Rab7-DCTN1-RILP-ORP1L-βIII spectrin cargo-motor complex for intracellular trafficking of late endosomes to the microtubule minus end (Johansson et al., 2007). Although Arp1 interacts with DCTN1, in which the *DCTN1* is a known gene for several neurodegenerative diseases, no mutations have been reported for the Arp1 complex.

This is the first systematic gene screening of cytoplasmic dynein-dynactin complex in an IPN cohort. Although there were no pathogenic mutations detected in our cohort, publication of these findings (Tey et al., 2014, 2016) will aid in evaluating variants that may be identified through next generation sequencing approaches in other studies. We have excluded the coding regions of cytoplasmic dynein-dynactin genes as the cause of disease in our 136 IPN cohort. However, there remains the possibility that mutations within the non-coding regions of the genes are responsible which would necessitate the use of whole genome sequencing. The latter was beyond the scope of the current thesis but we would recommend future work to be done to address this. Given the functional roles of the cytoplasmic dynein-dynactin complex and the evidence of human disorders with *DYNC1H1*, *DCTN1* and *DCTN2* mutations, the genes encoding the subunits of this complex remain as attractive candidates for neurological disorders and warrant future investigation.

CHAPTER 4: MAPPING A NEW LOCUS FOR AUTOSOMAL RECESSIVE CHARCOT-MARIE-TOOTH NEUROPATHY IN A MALAYSIAN FAMILY

4.1 Introduction

4.1.1 A family with CMT: Clinical review and preliminary genetic investigation

Previously, we screened our Malaysian cohort for mutations in the common CMT genes (*PMP22*, *MPZ*, *MFN2* and *GJB1*), and a genetic diagnosis was achieved in 65% of cases (Shahrizaila et al., 2014). Of the remaining 35% that were genetically unsolved, one male proband (IV:7) was the offspring of a consanguineous marriage (Figure 4.1). His younger brother (IV:9) was also affected with CMT. Both parents (III:9 and III:10), their older sister (IV:6) and younger brother (IV:8) were unaffected. Hence this family with two affected family members represented an opportunity for further studies. The family was designated with the ID CMT861.

Clinically, both affected siblings had a similar phenotype with the onset of neuropathy symptoms during adolescence. At the time of examination, the proband was 24 years old, while the younger brother was 17. Nerve conduction studies indicated a demyelinating form of CMT (Appendix, Table B.2). Clinical examination revealed reduced strength in finger abduction and ankle dorsiflexion. The ankle reflexes were absent but both knee reflexes were brisk which is atypical for CMT. The proband progressively developed fixed plantar flexion deformities and increased muscle tone in his lower limbs; but these phenotypes were not observed in the younger brother although they may manifest later in age. Further details of the clinical examination for both brothers are shown in Appendix (section B).

The pedigree of CMT861 suggested the pattern of inheritance to likely be autosomal recessive. However, given that the sons were affected and there was no male to male

transmission, the possibility of a recessive X-linked inheritance could not be excluded. The proband was subsequently sent for WES, where all the known IPN genes (Appendix section E), including the known autosomal recessive and X-linked genes, were excluded as the causal gene. This suggested that a novel CMT gene might be the cause of the disease in the CMT861 family.

4.1.2 Mapping disease loci using linkage analysis

The linkage mapping approach was first used to map the disease locus for Huntington's disease (Gusella *et al.*, 1983) and it has been a powerful method for positional cloning strategies in large multi-generational families. Although identifying disease genes has been expedited by the use of NGS, the use of parametric linkage analysis in families with multiple affected individuals remains a powerful method to map a disease locus with statistical confidence. Parametric linkage analysis allows the calculation of a 'likelihood of odds' (LOD) that a genetic marker co-segregates with the disease allele according to a Mendelian mode of inheritance (Morton, 1984). The LOD scores provide statistical evidence for the disease locus which could then be further investigated for mutations in candidate genes. A LOD score $\geq +3.3$ indicates a high evidence of linkage; while a LOD score ≤ -2 indicates evidence for the exclusion of linkage (Lander & Kruglyak, 1995; Morton, 1984). LOD scores between these limits are inconclusive. However, LOD scores of > 1 are suggestive of linkage and warrant further investigation (Pericak-Vance, 2001).

In the context of autosomal recessive CMT (AR-CMT), the use of linkage analysis has helped to identify the genetic locus for a number of AR-CMT genes such as *FIG4*, *GDAP1*, *NDRG1* and *HK1* (Baxter et al., 2002; Chow et al., 2007; Hantke et al., 2009; Kalaydjieva et al., 2000). In X-linked CMT, linkage analysis has successfully identified several disease genes including *ATP7A* (Kennerson et al., 2010; Takata et al., 2004),

PRPS1 (Kim *et al.*, 2005; Kim *et al.*, 2007), *PDK3* (Kennerson et al., 2013) and *AIFM1* (Rinaldi *et al.*, 2012). Furthermore, a large insertion within the previously mapped CMTX3 locus on chromosome Xq26.3-q27.3 has recently been identified to cause CMT in two large families (Brewer et al., 2016).

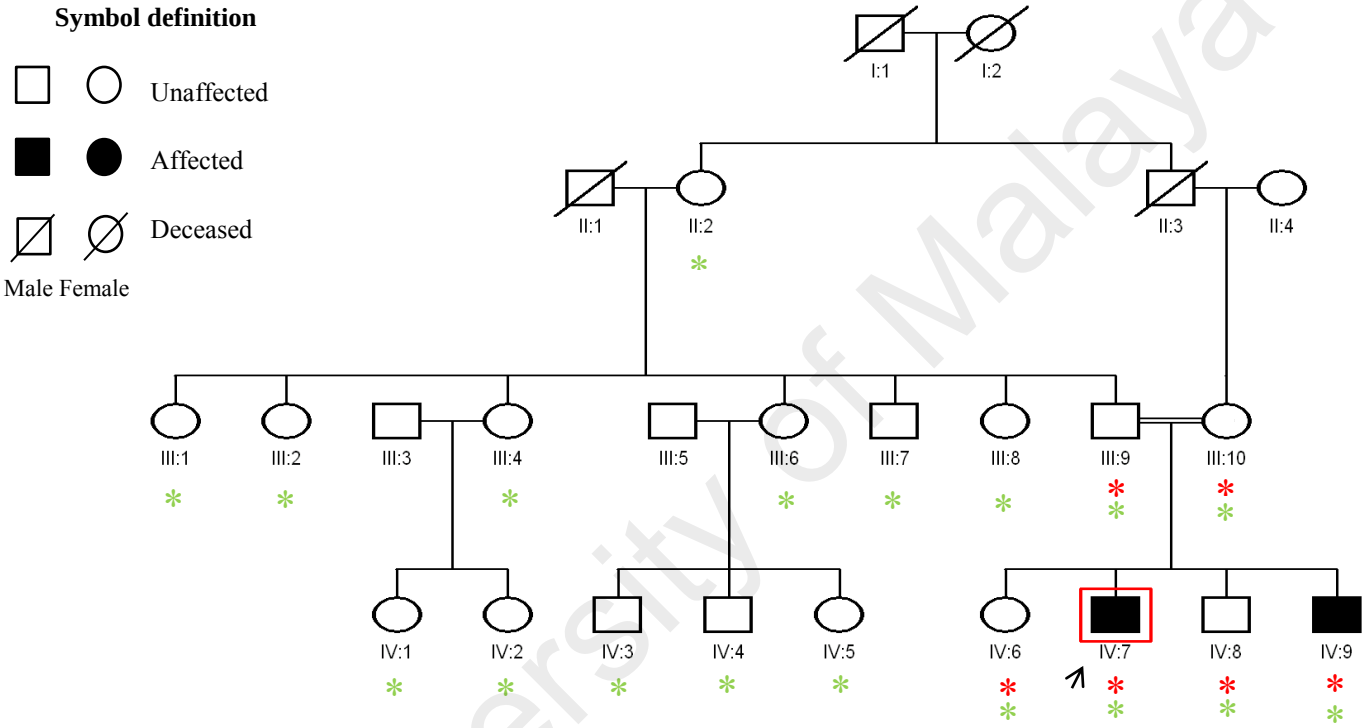


Figure 4.1: The CMT861 pedigree.

Individual III:9 and III:10, who are first cousins, have two sons affected with CMT (IV:7 and IV:9). The daughter (IV:6) and the other son (IV:8) are unaffected. The proband IV:7 (arrow) was previously sent for WES and was negative for all the known IPN gene mutations. DNA samples from 18 family members (asterisks) were collected for genetic investigation in this study. Individuals labelled with a red asterisk (*) were genotyped with the Golden Gate Human Linkage V panel for genome-wide linkage scan. Individuals labelled with a green asterisk (*) were genotyped using microsatellite markers for fine mapping of the suggestive linkage peaks identified.

4.1.3 Hypothesis

Mutation in a novel CMT gene is responsible for the autosomal recessive neuropathy in the CMT861 family.

4.1.4 Aims

To map the disease locus in CMT861 by:

1. Performing genome-wide linkage analysis using a SNP-genotyping array (Golden Gate Human Linkage V Panel).
2. Validating and refining the candidate loci identified by genome-wide linkage analysis using microsatellite markers.

4.2 Methodology

4.2.1 Patient samples collection

Blood samples from 18 family members were collected according to protocols approved by the University of Malaya Medical Centre Ethics committee. The protocol for DNA extraction is described in Chapter 2, section 2.2.

4.2.2 Estimation of theoretical LOD score using the FastSLink Simulation Program

The two-point LOD score simulation program, FastSLink (version 2.51) from the easyLINKAGE Plus package (version 5.02) was used to estimate the theoretical maximum two-point LOD score for CMT861. The simulation modelled autosomal recessive inheritance with a disease allele frequency of 0.001. The simulation of the theoretical LOD scores using a four-allele and a two-allele linked marker with equal allele frequency was performed. For statistical purposes the pedigree information was simulated for N=1000 replicates.

4.2.3 Genome-wide SNP genotyping using the Golden Gate Human Linkage V Panel

DNA samples from the two affected brothers (IV:7 and IV:9) as well as the four unaffected family members (III:9, III:10, IV:6 and IV:8) were sent to Macrogen (South Korea) for SNP genotyping using the Golden Gate Human Linkage V Panel (LVP), (Figure 4.1). This linkage panel is designed to genotype 6056 highly informative SNP markers distributed evenly across the genome with an average genetic distance of 0.62 cM and an average physical distance of 470 kb. Output files from the LVP genotyping were obtained in text file format.

4.2.4 Processing genotype data from LVP using LINKDATAGEN

LINKADATAGEN uses the Perl script linkdatagen.pl (release 20130704) to create input files for MERLIN linkage analysis using the genotype data generated from Illumina's LVP (Abecasis et al., 2002; Bahlo & Bromhead, 2009; Smith et al., 2011). These files include the pedigree file (pedfile), the genotype file (callFile), the order of individual's genotype file according to the pedigree (whichSampleslist) and the LVP annotation file. The LVP annotation file was downloaded from the Illumina website and used for the genotype-calling. The SNP population frequency was annotated according to the CEU population by default. CEU population is the Utah Residents with Northern and Western Ancestry. A binsize of zero "0" was set to maximise the number of SNP markers included in the analysis. To run parametric linkage analysis using MERLIN, a text file describing the disease model was prepared according to MERLIN's protocol. The text file includes the disease allele frequency, the disease penetrance with the probabilities for an individual to be affected with 0, 1 and 2 copies of disease allele and a label for the disease model. Autosomal recessive inheritance was modelled with a disease allele frequency of 0.001 and complete penetrance. The chromosome-X version of MERLIN is called MINX. MINX assumes that males scored as homozygous in the input file are hemizygous for the X chromosome. X-linked inheritance was modelled with a disease allele frequency of 0.001 and 95% penetrance. Scripts for the linkage analysis using MERLIN are shown in the Appendix (Section C.1).

4.2.5 Fine mapping the disease locus using microsatellite markers

Microsatellite markers were selected from the Rutgers Map Browser (Build 37) (Matise *et al.*, 2007). The browser integrates the physical position of the markers in base pairs (bp) with the genetic maps in centimorgan (cM). Microsatellite markers (dinucleotide and trinucleotide repeats) with a minimum heterozygosity of 0.6 within the region of interest were selected. For this study most markers had a heterozygosity

value of > 0.8. Primer sequences for the markers were obtained through the UCSC Genome Browser and are listed in the Appendix (Table A.2).

4.2.5.1 Microsatellite marker genotyping and data analysis

Eighteen family members were genotyped (Figure 4.1). Microsatellite markers were amplified by PCR with a 6-carboxyfluorescein (FAM) labelled forward primer using the protocols described in Chapter 2, section 2.5. Size fractionation of the amplified products was outsourced to the Australian Cancer Research Foundation (ACRF) at the Garvan Institute of Medical Research (Darlinghurst, NSW Australia). For marker allele scoring, the Gene Marker software, version 1.60 (SoftGenetics LLC) was used. The Cyrillic program, version 2.1 (Cherwell Scientific Publishing Ltd) was used to input pedigree information and to manage microsatellite marker genotype data. The pedigree information and genotype data was exported from Cyrillic for linkage analysis using the MLINK format option for two-point (MLINK) and multipoint (LINKMAP) linkage analysis.

4.2.6 Two-point and multipoint linkage analysis using the LINKAGE program

The LINKAGE package incorporates the five analysis programs ILINK, MLINK, LINKMAP, CILINK and CMAP (Lathrop et al., 1984). The input files for the LINKAGE program were exported from the Cyrillic pedigree program which include a pedigree information file (mlink.pre) as well as a file describing the disease locus, marker order and marker allele frequencies (mlink.dat). The MLINK and LINKMAP programs were used for two-point and multipoint linkage analysis respectively. Briefly, mlink.pre pedigree files were processed through MAKEPED to produce mlink.ped files with added identifiers specific for the analysis programs. The linkage control program (LCP) was then used to generate a command file with parameters to perform the analysis using either the MLINK or the LINKMAP program. The linkage report

program (LRP) was used to generate reports for both the two-point and multipoint linkage analyses.

4.3 Results

4.3.1 The maximum expected LOD score for CMT861 is below +3.3

Simulation studies were performed to predict the linkage power of CMT861 using microsatellites (four-allele marker) and SNPs (two-allele marker) genotype information. The pedigree information of 18 family members underwent 1000 times simulation for a linked marker, which generated the average expected LOD score (ELOD) and the maximum ELOD for a range of recombination fractions (mean ± SD; N=1000 replicates) as shown in Table 4.1. This information provides a reference for the LOD score to be expected from a fully informative linked marker for a range of recombination fractions. The maximum two-point ELODs obtained from the four-allele and two-allele marker simulation analysis are 1.75 and 1.45 at $\theta=0$ respectively (Table 4.1). The family did not have the power to independently achieve statistically significant linkage (LOD score of $\geq +3.3$), however, the maximum ELOD achieved was in the range of suggestive linkage (Lander & Kruglyak, 1995).

Table 4.1: The mean and maximum ELOD obtained from simulation analysis of CMT861.

Two-point LOD scores are based on the simulation of four-allele and two-allele linked marker systems with equal allele frequency. N=1000 replicates.

	Four-allele linked marker		Two-allele linked marker	
Recombination fraction (θ)	Mean ELOD ± SD	Maximum ELOD	Mean ELOD ± SD	Maximum ELOD
0.00	1.28 ± 0.49	1.75	0.73 ± 0.47	1.45
0.050	1.13 ± 0.44	1.59	0.64 ± 0.42	1.33
0.100	0.99 ± 0.39	1.42	0.55 ± 0.37	1.21
0.150	0.84 ± 0.35	1.24	0.46 ± 0.33	1.08
0.200	0.70 ± 0.30	1.06	0.38 ± 0.29	0.95

4.3.2 Genome-wide linkage analysis identified seven loci with LOD scores > 2

Under an autosomal recessive model, the genome-wide linkage scan achieved a maximum multipoint LOD score of 2.05 for several chromosomal loci. The multipoint LOD scores plotted against the genetic distance of the SNP markers for each chromosome are shown in the Appendix (section D.1). Seven suggestive linkage peaks with LOD scores > 2 were identified on six chromosomes: 3, 11, 12, 14, 19 and 20, with two peaks identified on chromosome 11. There was also a suggestive peak on chromosome 7 with a multipoint LOD score of almost 2 (LOD =1.94). The information of these suggestive linkage peaks is detailed in Table 4.2.

Under a recessive X-linked model, the maximum LOD score achieved on chromosome X was 0.9 (Appendix, Figure D.1b). Suggestive LOD scores were not obtained on this chromosome therefore it is unlikely that the mode of inheritance in CMT861 is X-linked or the causative gene maps to the chromosome X.

Table 4.2: Suggestive linkage regions with maximum multipoint LOD scores in CMT861.

Chromosome	Chromosomal position of suggestive linked region	Flanking SNPs	Size (Mb)	Maximum LOD score
3	190782523 – 197306512	rs2048417 – rs1030576	6.52Mb	2.05
7	878767 – 2803510	rs1127460 – rs798485	1.92Mb	1.94
11	288505 – 6280248	rs741737 – rs906895	5.99Mb	2.03
	63776265 – 71930207	rs562865 – rs514933	8.15Mb	2.05
12	72620122 – 92705204	rs1484828 – rs337663	20.09Mb	2.05
14	91557911 –107349540	rs4904745 – end of chromosome 14	15.79Mb	2.05
19	33231786 – 46894786	rs977708 – rs741233	13.66Mb	2.05
20	19341331 – 35946239	rs728481 – rs7272911	16.60Mb	2.05

4.3.3 Examination of suggestive loci using microsatellite markers linkage analysis

Microsatellite marker linkage analyses were performed to examine the eight suggestive linkage peaks obtained from the genome-wide linkage analysis. Forty microsatellite markers within the suggestive loci were genotyped in the 18 family members. Two-point and multipoint linkage analyses were performed using the microsatellite marker genotype information. Extended haplotypes were also constructed to test the co-segregation of the candidate chromosomal regions with the disease phenotype.

4.3.3.1 Exclusion of linkage peaks on chromosome 3, 7, 11, 19 and 20

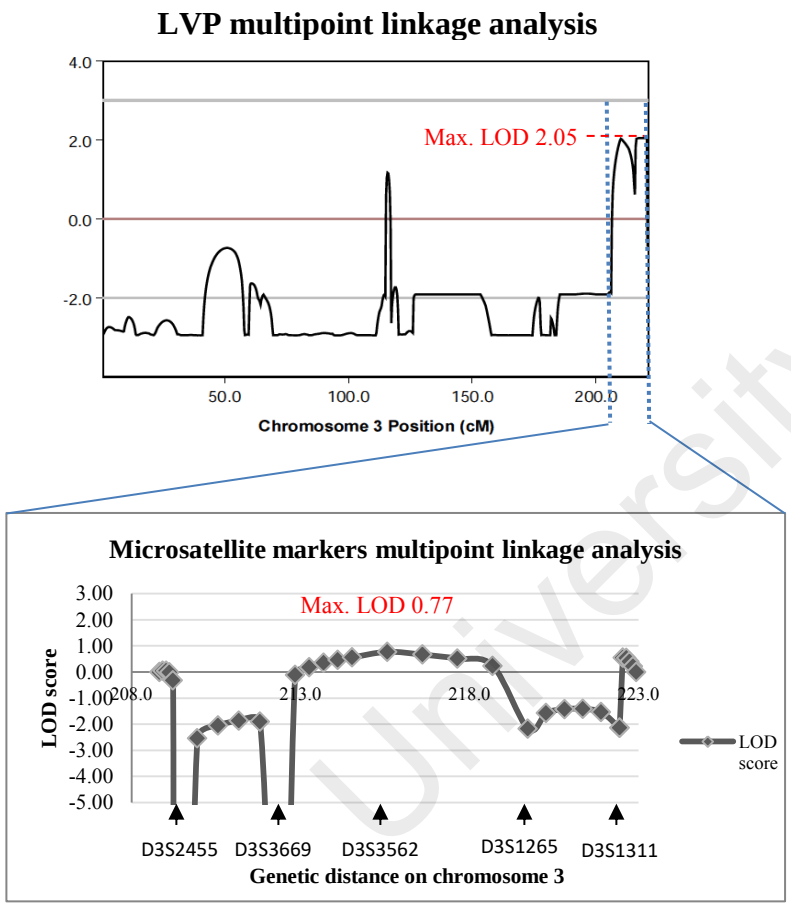
The suggestive loci identified on chromosome 7 and 11 were excluded as the multipoint LOD scores reached an exclusion threshold value of -2 across the majority of the likelihood curves (Figure 4.2). The loci on chromosome 3, 19 and 20 are unlikely to be the disease locus, as the multipoint LOD scores for the respective regions were below the simulated ELOD score of 1.75 and LOD scores of -2 were obtained at some locations across the interval (Figure 4.2). Therefore, these loci were given low priority for further analysis.

Figure 4.2: Likely exclusion of suggestive linkage regions identified through the genome-wide linkage analysis.

Multipoint LOD score likelihood curves for microsatellite marker fine mapping are shown below the original multipoint plots from the LVP genome-wide scan. The multipoint LOD score likelihood curves for the candidate interval on a) chromosome 3, b) chromosome 7, c) chromosome 11 peak 1 and peak 2, d) chromosome 19 and e) chromosome 20, do not support the suggestive linkage peaks identified from the LVP linkage analysis.

Figure 4.2 (continued)

a)



b)

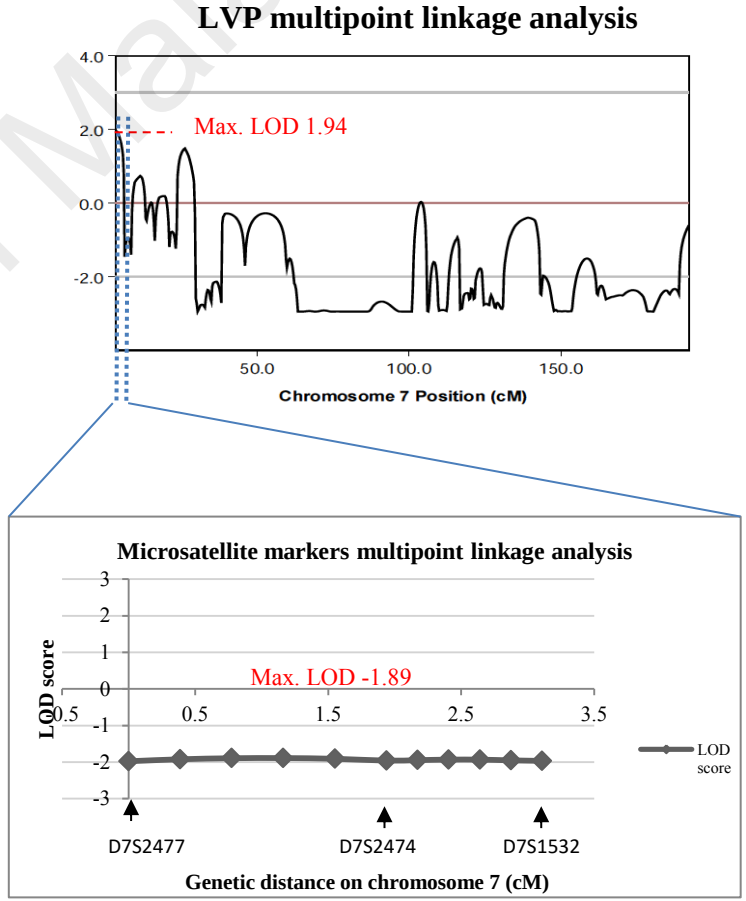


Figure 4.2 (continued)

c)

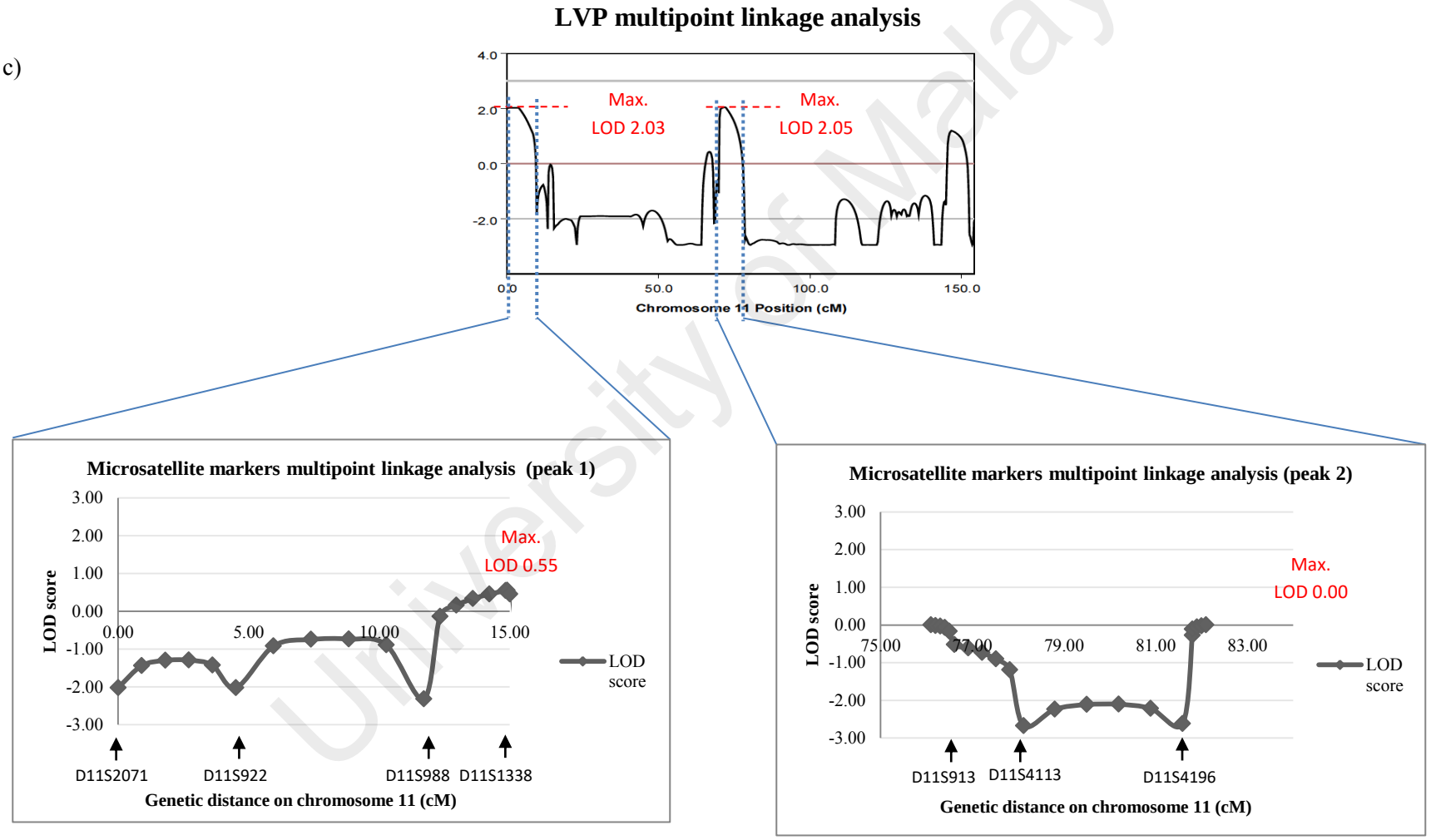
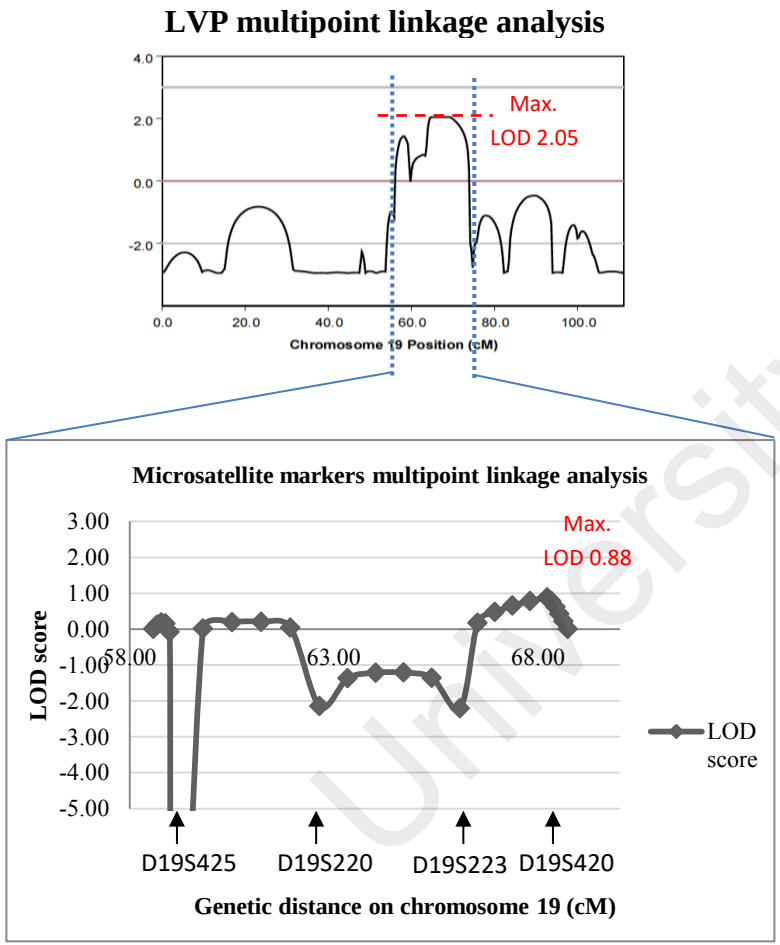
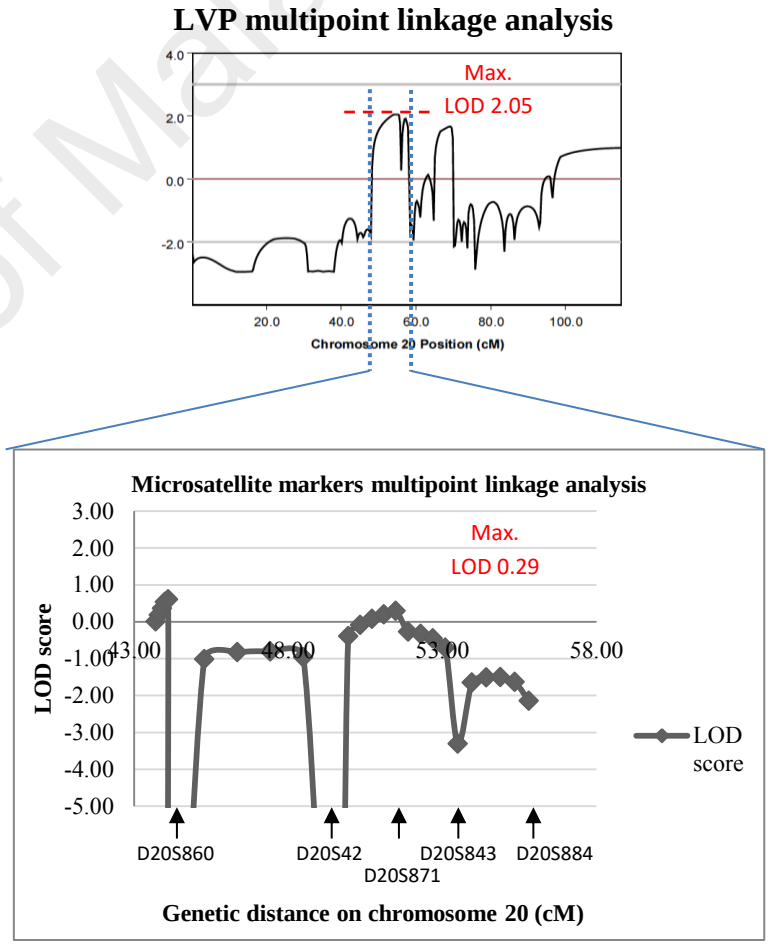


Figure 4.2 (continued)

d)



e)



4.3.3.2 The suggestive linkage region on chromosome 12 remains a candidate locus for CMT861

A maximum multipoint LOD score of 2.05 was obtained from the LVP genome-wide scan on chromosome 12q21.1-q22, which spans a 20.1 Mb interval (Table 4.2). To further examine this suggestive linkage peak, six microsatellite markers (D12S299, D12S1684, D12S2074, D12S2068, D12S88, and D12S838) were selected for analysis in the 18 family members (Figure 4.3) with the markers D12S299 (proximal) and D12S838 (distal) flanking the suggestive linkage region.

Two-point linkage analysis achieved positive LOD scores for the marker D12S1684, D12S2074, D12S2068 and D12S88, with a maximum LOD score of 1.61 at zero recombination (θ=0) for marker D12S1684. Similarly, multipoint linkage analysis gave a maximum LOD score of 1.74 for the likelihood curve between these markers.

Haplotype analysis shows that the parents (III:9 and III:10) are carriers of the haplotype (2-4-2-2-2-4). This haplotype been inherited by the two affected sons in a homozygous state (Figure 4.4). Two recombination events occurred in the proband (IV:7) at D12S299 and D12S838. The other affected brother (IV:9) also showed a recombination event at the marker D12S838 (Figure 4.4). The combined recombination data from both of the brothers defined D12S299 and D12S838 as the proximal and distal flanking markers which now correspond to a 19.3 Mb candidate interval. Furthermore, runs of homozygosity in between marker D12S1684 to D12S88 were observed in the two affected brothers and were absent in the unaffected family members.

Fine mapping of the chromosome 12 linkage interval supported the initial suggestive linkage region identified via LVP linkage analysis. Even though the maximum multipoint LOD score (LOD 1.74) obtained from fine mapping did not reach the initial

LOD score achieved in LVP linkage analysis (LOD 2.03), this region could not be excluded as it approached the maximum simulated ELOD (ELOD 1.75). The recombination events identified in the affected brothers defined the potential genetic interval flanked by marker D12S299 and D12S838. Therefore, the 19.3 Mb linkage interval between these markers on chromosome 12 remains a candidate for the disease locus. Further investigation of this locus is described in Chapter 5.

Chromosome 12 (chr12:72,620,122-92,705,204; size 20.1 Mb)

Flanking SNP markers: rs1484828 rs337663

Markers	Genetic distance (cM)	Physical position (bp)	Two-point linkage analysis LOD score at recombination fraction				
			0.0	0.1	0.2	0.3	0.4
D12S299	89.51	73169831	-infini	0.35	0.41	0.34	0.2
D12S1684	92.26	77263237	1.61	1.23	0.86	0.52	0.24
D12S2074	96.81	80431301	1.03	0.83	0.62	0.42	0.22
D12S2068	98.54	82985249	1.58	1.21	0.84	0.5	0.22
D12S88	100.19	86371458	0.71	0.49	0.28	0.12	0.02
D12S838	103.89	92495129	-2.15	-0.61	-0.27	-0.16	-0.09

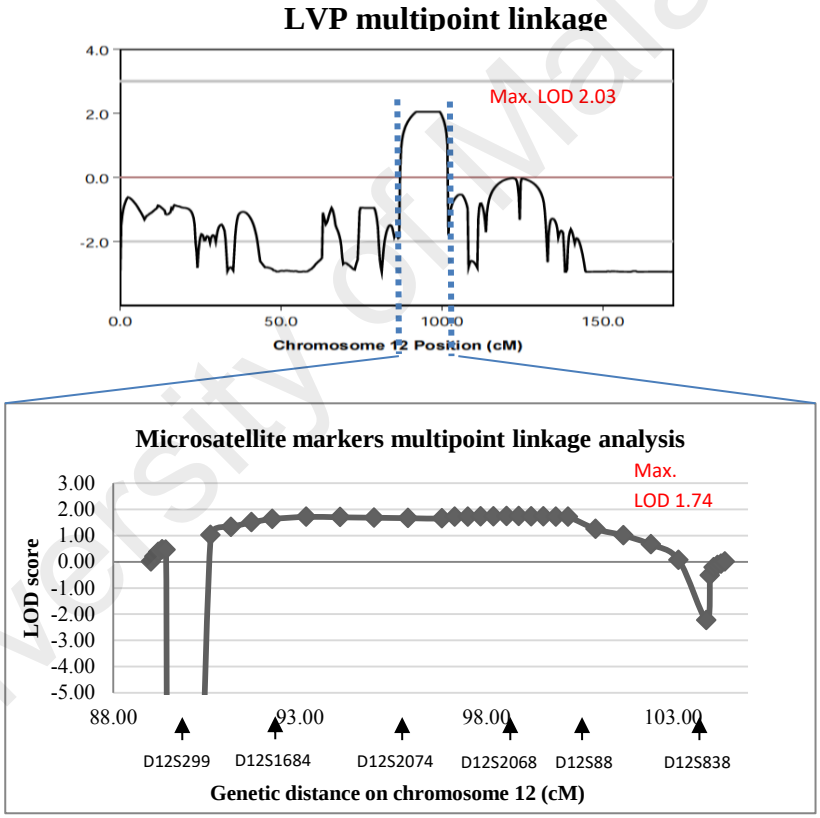


Figure 4.3: Investigation of the suggestive linkage peak on chromosome 12q21.1-q22.

The dotted blue line designates the suggestive linkage region (LOD 2.03) identified in the genome-wide linkage scan based on the flanking marker rs1484828 and rs337663 which have recombined with the disease locus in the family. The two-point LOD scores for the microsatellite markers selected within the linkage region are tabulated. Multipoint LOD scores are plotted against the chromosome position in cM.

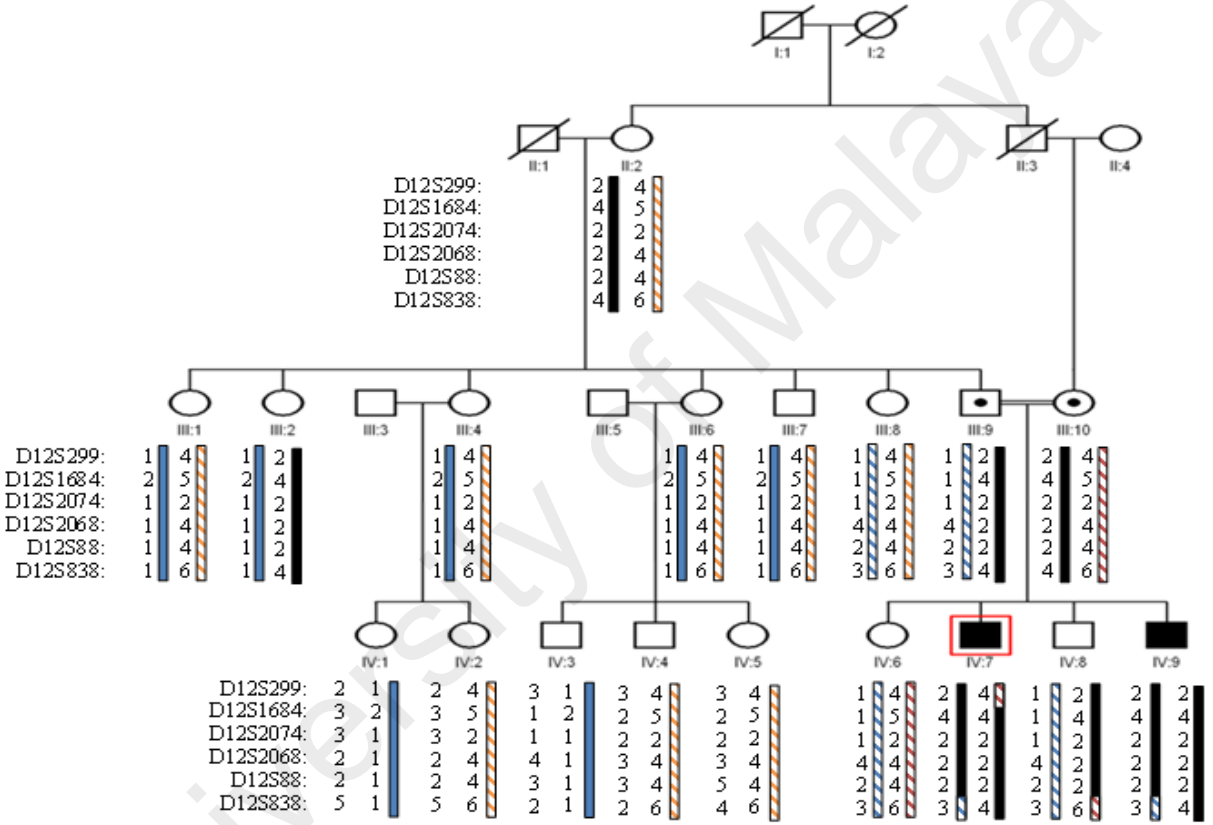


Figure 4.4: Haplotype analysis of microsatellite markers spanning chromosome 12q21.1-q22.

Females and males are represented by circles and squares respectively. Solid symbols indicate affected individuals. Symbols with a black dot indicate carriers. Marker alleles are listed below the individuals and ordered from centromere to telomere. The haplotypes are indicated by bars. The solid black haplotype segregates with the disease. The recombination events in individuals IV:7 and IV:9 define the candidate interval to a 19.3 Mb region flanked by the marker D12S299 and D12S838.

4.3.3.3 Suggestive linkage region on chromosome 14 is further supported as a candidate disease locus for CMT861

A suggestive linkage region with a maximum multipoint LOD score of 2.05 was identified on chromsome 14q32.11-q32.33 by LVP genome-wide linkage scan (Table 4.2). This region was flanked by SNP marker rs4904645 at the proximal end of the interval and a likelihood curve with maximum LOD score of 2.05 extended across the interval to marker rs1989750, which is the last marker available on chromsome 14 in the LVP array (Figure 4.5). This suggestive linkage region corresponds to a 15.79 Mb interval.

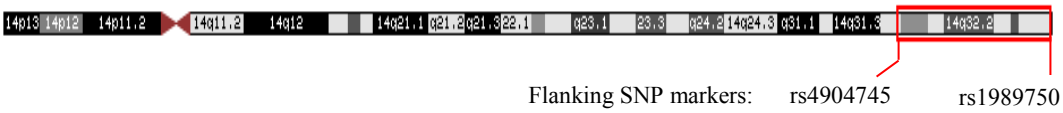
To further investigate the suggestive linkage peak, seven microsatellite markers and five SNP markers within the interval were analysed in 18 family members (Figure 4.5). The SNP markers were selected from the WES data of the proband and were either novel variants or had a MAF $\leq$ 0.001. For linkage analysis, the alleles of the microsatellite markers observed in CMT861 were assigned with equal frequency, while alleles of SNP markers were assigned with their respective reported MAF. Novel SNP markers were given a MAF of 0.001 based on screening of 1000 chromosomes of neurologically normal individuals.

Two-point linkage analysis showed that the marker D14S291, D14S1050, D14S62 and D14S267 have recombined with the disease locus (Figure 4.5). This was supported by the multipoint linkage analysis in which negative multipoint LOD scores for the likelihood curves were observed between the marker D14S291 and D14S267 (Figure 4.5). Based on this information, the proximal part of the suggestive linkage interval has been statistically excluded and the intial 15.79 Mb interval was now refined to 8.12 Mb.

The suggestive linkage interval between the marker D14S267 and D14S1007 showed positive two-point LOD scores. The five SNP markers showed a maximum two-point LOD score of 2.03 at $\theta = 0$, which exceeded the maximum ELOD (LOD = 1.45 at zero recombination) simulated using two-allele marker with equal frequency (Figure 4.5). A positive multipoint LOD score of 2.06 was also achieved across the genetic interval from marker D14S267 to the distal end of the long arm of chromosome 14, with the marker D14S1007 being the most distally mapped marker on chromsome 14q. Therefore, the distal flanking marker could not be defined and the distal end of this suggestive linkage interval could not be further refined.

Haplotype analysis showed that the parents (III:9 and III:10) are carriers of the haplotype (2-2-1-1-2-2-2-4). This haplotype has been inherited by the two affected son in a homozygous state (Figure 4.6). The runs of homozygosity from marker SNP1 to D14S1007 were observed in the two affected brothers and were absent in the unaffected family members. This region therefore remained a strong candidate locus for further investigation.

Chromosome 14 (position chr14:91,557,911-107,349,540; 15.79 Mb)



Markers	Genetic distance (cM)	Physical position (bp)	Microsatellite markers two-point linkage analysis LOD score at recombination fraction (θ)				
			0.0	0.1	0.2	0.3	0.4
D14S291	91.63	91270020	-infini	-0.28	0.02	0.08	0.06
D14S1050	94.44	92915550	-infini	-0.24	0.05	0.1	0.07
D14S62	100.89	95989592	-2.21	0.09	0.12	0.07	0.02
D14S267	108.97	99224198	-2.15	0.59	0.6	0.46	0.24
SNP1	112.95	100800159	2.03	1.67	1.27	0.86	0.43
SNP2	113.00	101003741	2.03	1.67	1.27	0.86	0.43
D14S1006	113.31	101179611	0.43	0.28	0.16	0.07	0.02
D14S1010	119.64	104214349	1.75	1.42	1.07	0.71	0.35
SNP3	121.66	105236365	2.03	1.67	1.27	0.86	0.43
SNP4	122.00	105419045	2.03	1.67	1.27	0.86	0.43
SNP5	122.1	105423842	2.03	1.67	1.27	0.86	0.43
D14S1007	123.36	105977978	1.06	0.84	0.64	0.43	0.22

Figure 4.5: Investigation of the suggestive linkage peak on chromosome 14q32.11-q32.33.

The dotted blue line designates the suggestive linkage region (LOD 2.05) identified in the genome-wide linkage scan based on the flanking marker rs4904745 and rs1989750. The two-point LOD scores for the microsatellite markers and SNP markers selected within the linkage region are tabulated. Multipoint LOD scores are plotted against the chromosome position in cM.

Figure 4.5 (continued)

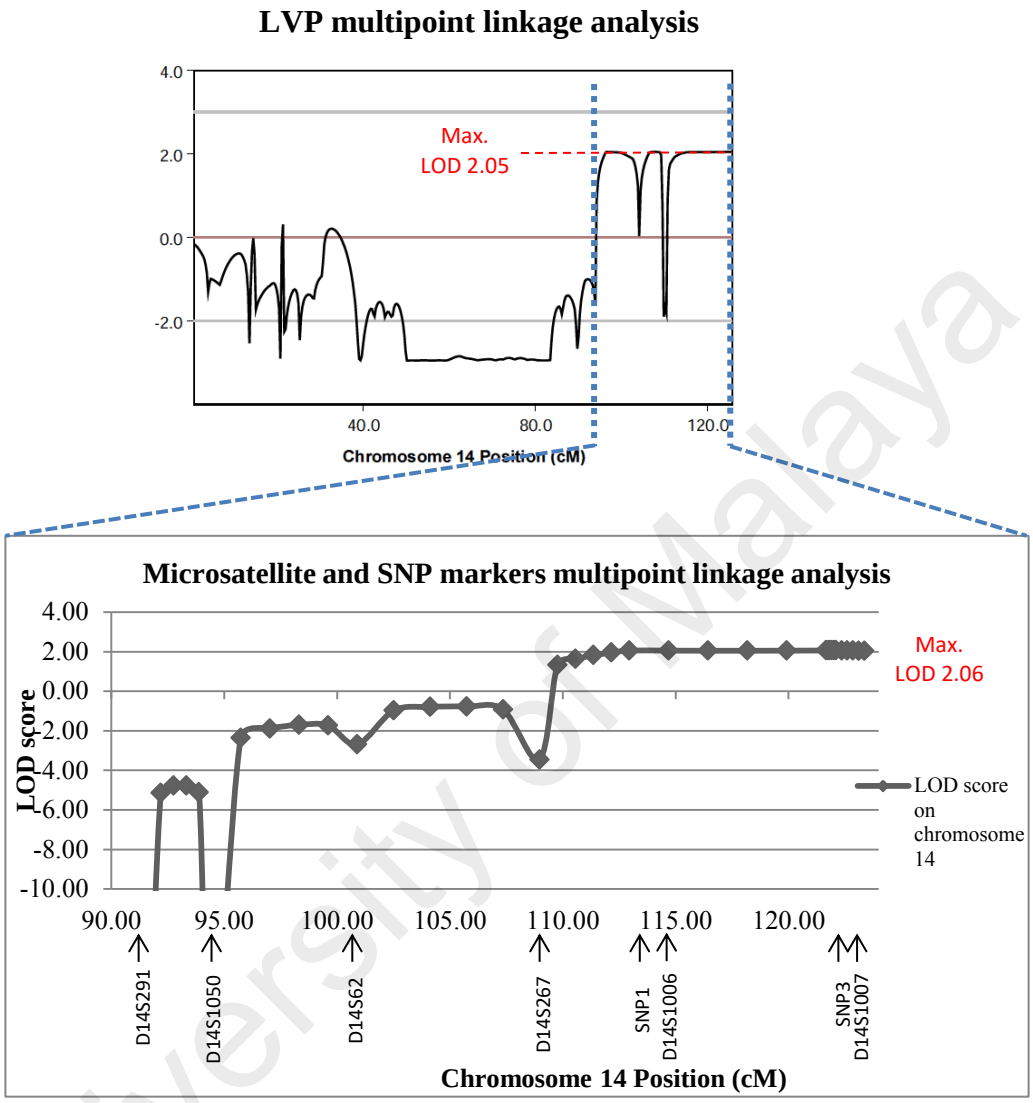
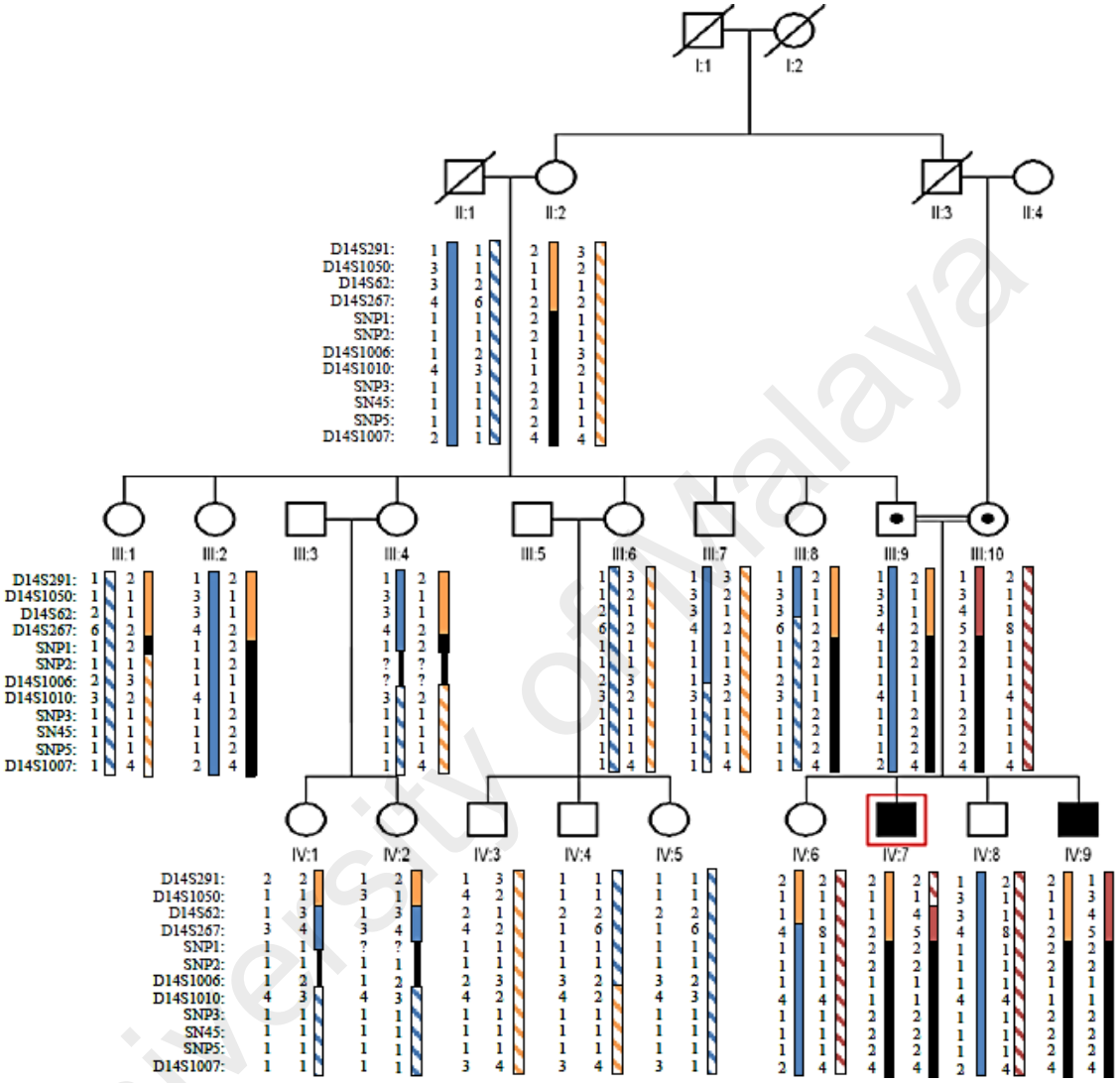


Figure 4.6: Haplotype analysis of microsatellite markers spanning chromosome 14q32.11-q32.33.

Females and males are represented by circles and squares respectively. Solid symbols indicate affected individuals. Symbols with a black dot indicate carriers. Marker alleles are listed below the individuals and ordered from centromere to telomere. The haplotypes are indicated by bars. The solid black haplotype segregates with the disease. Haplotype analysis shows that the parents are carriers of the disease haplotype (2-2-1-1-2-2-2-4). A recombination event identified in the affected brothers between marker D14S267 and SNP1 defined D14S267 as the proximal flanking marker. There is no recombination between the disease allele and marker D14S1007. The suggestive linkage region was therefore successfully refined to the interval between marker D14S267 and the distal end of chromosome 14q. This corresponds to an 8.12 Mb interval on chromosome 14q32.3-q32.33.

Figure 4.6 (continued)



4.4 Discussion

In this chapter, we used a linkage mapping strategy to identify the genetic locus for autosomal recessive CMT in the CMT861 family.

The maximum two-point ELOD for a simulated four-allele and two-allele linked marker with equal allele frequency on 18 family members showed the maximum ELOD of 1.75 and 1.45 respectively at zero recombination. This simulated data represents the maximum linkage power that could be achieved in CMT861 with the currently available family structure and phenotype. Simulation studies were also performed on the six members of the nuclear family (III:9, III:10, IV:6; IV:7, IV:8 and IV:9) plus the grandmother (II:2). Interestingly, a similar maximum ELOD of 1.75 and 1.45 was obtained (Appendix, Table D.2). This demonstrates that the majority of LOD score information came from the family members within the consanguinity loop (Figure 4.7).

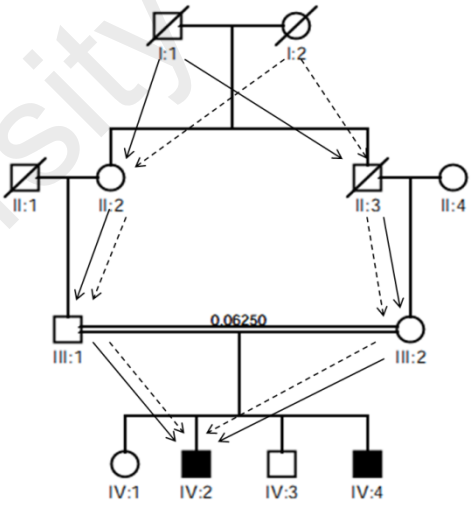


Figure 4.7: Consanguinity loop of family CMT861.

The consanguinity loop is the pathway that linked the parents of the proband (IV:7) to a common ancestor. The solid line shows the pathway that links the parents to the great-grandfather (I:1); the dotted lines shows the pathway that links the parents to the great-grandmother (II:2).

Due to multiple testing between the disease locus and LVP markers, multiple suggestive linkage peaks are likely to be identified (Risch 1991). Therefore a portion of the linkage peaks identified in CMT861 may represent false positive linkage regions. Ideally, a LOD score of 3.3 is required to achieve significant linkage so that the probability for false-positive results can be reduced to < 5% (Morton, 1955; Kruglyak et al., 1995). However, significant linkage values could not be achieved with CMT861 due to limited pedigree information. Yet, the simulation studies we performed provided an idea of the expected LOD scores achievable if a tested marker segregates with the disease. Genome-wide scan linkage analysis using an autosomal recessive model in CMT861 gave maximum multipoint LOD scores of 2.05 on several autosomes. Further LOD score analysis exceeded the maximum simulated ELOD possible in this family, thus, supporting an autosomal recessive mode of inheritance in CMT861. The maximum LOD score of 0.9 achieved on chromosome X was not further supported with the likelihood curve across the region giving values approaching -2. This further suggested that the mode of inheritance was unlikely to be X-linked.

Examining the suggestive linkage peaks was aided by testing additional markers for these regions in the expanded family. Although no additional affected individuals were available among the extended family members, the genotype information from the grandmother (II:2) was important to determine the linkage phase of alleles inherited in the affected brothers. Inclusion of the unaffected uncles, aunts and cousins allows for non-pathogenic markers to be excluded. The fine mapping with microsatellite markers excluded six of the suggestive linkage loci initially identified through the LVP linkage scan while the loci on chromosome 12 and 14 remain possible candidates. Although the maximum multipoint LOD score obtained on chromosome 12 (LOD 1.74) is lower compared to chromosome 14 (LOD 2.06), this linkage interval remains a candidate as it is close to the maximum ELOD for a fully informative four-allele marker. On

chromosome 14, the SNP markers (SNP1 to SNP5) gave higher two-point LOD scores compared to the microsatellite markers within the linkage interval. This is because the SNP markers tested in the family were assigned the actual allele frequency reported in dbSNP144 or based on screening of 1000 control chromosomes (MAF = 0.001). In contrast, the microsatellite markers were assigned with equal allele frequency. To test if the assigned allele frequencies will change the value of a two-point LOD score, two-point linkage analysis on the SNP markers assigned with equal allele frequency was performed (MAF = 0.5), assuming these SNPs are common. Two-point LOD scores ranging from 1.37 to 1.43 were observed in this analysis at $\theta = 0$ (Appendix, section D.3, Table D.3), corresponding to the maximum ELOD of 1.45 simulated from a fully informative two-allele linked marker with equal allele frequency. Therefore, the linkage signal obtained when SNP markers assigned with the actual allele frequency is true. For multipoint linkage analysis, the allele frequency does not have an obvious effect on the LOD scores. Multipoint linkage analysis across the SNP markers on chromosome 14 gave similar results when an equal allele frequency (LOD 2.02) or the actual allele frequency (LOD 2.06) was assigned (Appendix, section D.3, Figure D.3).

The main limitation of using linkage analysis on small nuclear families is that a significant LOD score may not be achieved. Small families such as CMT861 are more commonly seen in clinics than large multi-generational families and they can still represent a rich resource for gene discovery. This study has shown linkage analysis in smaller families is useful to suggest linked regions on the genome and importantly, to exclude non-pathogenic regions. CMT861 has provided simulation and experimental linkage data that has identified and supported suggestive disease loci for AR-CMT in this family. However, additional approaches are needed to facilitate these linkage studies to give confidence to the results obtained. Given that there are > 150 genes in each of the suggestive loci on chromosome 12 and 14, an effective strategy is required

to prioritise the loci for further analysis. Chapter 5 describes the use of WES with homozygosity mapping to further explore the suggestive linkage regions on chromosome 12 and 14, and variant filtering of WES data to identify the gene responsible for the AR-CMT phenotype in CMT861.

CHAPTER 5: WHOLE EXOME SEQUENCING FOR HOMOZYGOSITY MAPPING AND AR-CMT CANDIDATE GENE IDENTIFICATION IN CMT861

5.1 Introduction

In Chapter 4, genome-wide linkage analysis and fine mapping in CMT861 identified two suggestive linkage intervals on chromosome 12 and 14. As these intervals were still large (19.3 and 8.12 Mb respectively), we employed additional experimental approaches to exclude and narrow the candidate regions further. In this chapter, homozygosity mapping was performed to identify regions of homozygosity shared between the two affected brothers. Filtering of WES data from several family members was then performed to identify candidate variants within the suggestive loci.

5.1.1 Homozygosity mapping

Homozygosity mapping is an approach taken to identify stretches of homozygous sequences within the genome. This method is often used to map autosomal recessive loci in inbred populations or in consanguineous families (Lander & Botstein, 1987). Homozygosity mapping is based on the principle that the affected children of a consanguineous marriage are likely to be homozygous for the causal variant, which has arisen in a common ancestor and has been transmitted by both parents who are related (Lander & Botstein, 1987). The alleles that are in linkage disequilibrium with the causal variant are therefore likely to be transmitted in the same haplotype block. By looking for regions with 'runs of homozygosity' (ROH) in the genome, the locus for a causal variant may be identified.

Homozygosity mapping can be performed using genotype information from polymorphic markers such as microsatellites or bi-allelic SNP markers. SNPs are the markers of choice for homozygosity mapping as large SNP numbers can be rapidly

screened using array based systems (Alkuraya, 2010). The approach of homozygosity mapping for identifying autosomal recessive disease genes in consanguineous families has been successful for a number of autosomal recessive CMT genes including *HINT1*, *PRX*, *GDAP1*, *SH3TC2*, and *LMNA* (Cuesta et al., 2002; De Sandre-Giovannoli et al., 2002; Delague et al., 2000; LeGuern et al., 1996; Zimon et al., 2012).

As the segregation of the disease phenotype in CMT861 from a consanguineous marriage suggests autosomal recessive inheritance, homozygosity mapping would be a useful tool to identify ROH that support the suggestive linkage intervals previously identified in CMT861. The ROH overlapping with the suggestive linked loci provides a rationale for prioritising the identified loci for further investigation. In this study, homozygosity mapping was performed in family CMT861 using the genotype data from WES.

5.1.2 Whole exome sequencing for candidate gene discovery

Next generation sequencing (also known as massive parallel sequencing technology) refers to the high-throughput technology for sequencing the whole genome (known as whole genome sequencing; WGS) or the coding exons (known as whole exome sequencing; WES). The use of WES in this project was chosen for various reasons. Although the exome only accounts for ~1% of the genome, approximately 85% of the mutations for Mendelian disorders are reported in the exons (Choi et al., 2009; Stenson et al., 2009). WES has therefore become the favoured option when studying Mendelian diseases in small nuclear families with only a few affected family members, in which the power for linkage analysis is lacking. Also, the management of WES data is much more convenient and affordable as it generates only a fraction of the sequence data as compared to WGS.

One of the first studies using WES to identify genes for Mendelian disorders was a study on the Freeman-Sheldon syndrome, which successfully identified *MYH3* mutations (Ng et al., 2009). The identification of new genes for CMT also substantially increased with the introduction of WES which included the discovery of mutations in *DYNC1H1*, *HARS*, *SCN11A* that cause autosomal dominant CMT (Leipold et al., 2013; Vester et al., 2013; Weedon et al., 2011) and *TRIM2* in autosomal recessive CMT (Ylikallio *et al.*, 2013).

On average, WES generates datasets of over 80,000 annotated variants for a single exome (Feng *et al.*, 2013; Linhares *et al.*, 2014) and the challenge is to decipher the pathogenic from non-pathogenic variants. One way to reduce the number of candidate variants is to perform heuristic filtering of the data to determine the variants shared by the affected individuals and exclude those that are common with the unaffected family members, which would likely be non-pathogenic (Brunham & Hayden, 2013). An additional step to further reduce the number of candidate variants is to exclude common variants (MAF > 5%) that are present in the normal population. This can be achieved using data mined from multi-ethnic public repositories that catalogue genetic variations. These repositories include the HapMap project (International HapMap Project, 2003), the 1000 Genomes project (1000 Genome Project Consortium, 2010), the dbSNP (Sherry et al., 2001), the NHLBI Exome Variant Server (Exome Variant Server) and the Exome Aggregation Consortium (ExAC), (Lek *et al.*, 2015).

In this chapter, data from previous linkage studies was combined with homozygosity mapping and WES variant filtering to help prioritising the suggestive linkage peaks and reducing the number of putative pathogenic candidate variants. Combining WES with homozygosity mapping and linkage analysis has been a successful strategy to identify new genes for autosomal recessive non-syndromatic

hearing loss (Walsh *et al.*, 2010), oculocutaneous albinism and neutropenia (Cullinane *et al.*, 2011), Knobloch Syndrome (Haghighi *et al.*, 2014) and spastic ataxia (Dalal et al., 2015). Therefore, the use of homozygosity mapping in this project is likely to add additional mapping information for the disease locus in CMT861.

5.1.3 Hypothesis

WES and homozygosity mapping will prioritise the suggestive linkage intervals for CMT861 and identify the causative gene for AR-CMT in the CMT861 family.

5.1.4 Aims

The overall aim is to identify the causative gene causing AR-CMT in the CMT861 family. The specific aims are to perform:

1. Homozygosity mapping using variants identified by WES to identify ROH that overlap with the suggestive linkage peaks identified in CMT861.
2. Variant filtering of WES data to identify candidate variants localised within the overlapping suggestive linkage interval and ROH.

5.2 Methodology

5.2.1 Whole exome sequencing data analysis

In this study, five individuals from CMT861 including the two affected brothers (IV:7 and IV:9), the unaffected brother (IV:8) and the parents (III:9 and III:10) were sent for WES. The WES was outsourced to Axeq Technologies (South Korea). Sequencing library preparation, sequence data alignment and variant calling of WES data is described in Chapter 2 (section 2.3). The SNPs and indels were annotated with dbSNP 144 and the 1000 Genomes project database release 2010 using ANNOVAR (Wang et al., 2010). The variant reports were viewed in Excel files for ease of data management. BAM files were used with the Integrative Genomics Viewer (IGV) to view sequence alignments and examine the coverage for each nucleotide and the variant annotation.

5.2.2 Homozygosity mapping using WES data

LINKDATAGEN was used to prepare the input files required for PLINK from WES data. The Perl script vcf2linkdatagen.pl (release 20130704) was used to create a BRLMM genotype file from the VCF files which were generated from the BAM files using SAMTOOLS (Smith et al., 2011). This BRLMM file contains the genotype data for all individuals in a pedigree file (pedfile). Annotation files were prepared with HapMap II CEU genotypes. A binsize of zero '0' was set in order to maximise the number of SNPs included in the analysis. SNPs that were missing, not annotated (-missingness 0) or had a read depth of less than 10 (-mindepth 10) were discarded. The Perl script linkdatagen.pl (release 20130704) was used to generate input files for homozygosity mapping using PLINK (Bahlo & Bromhead, 2009; Purcell et al., 2007). To perform this analysis, various files were required: genotype data (-data m for massive parallel sequencing data), pedigree file, the order of individual genotypes

according to the pedigree (whichSamplesFile), the BRLMM genotype file (callFile), and the annotation file (HapMap II CEU genotypes).

The assigned parameters in PLINK for determining runs of homozygosity (ROH) scanned the genome with a window of 50 SNPs consecutively. One heterozygous SNP was allowed in each window and the overlap between the sliding windows was set at 5%. The maximum gap allowed in between SNPs was 1000 kb and only regions with a minimum of 100 consecutive homozygous SNPs were selected. Overlapping homozygous segments with at least 95% similarity between any of the family members were pooled. Finally, the homozygous segments shared between the two affected brothers but absent in the other unaffected family members were chosen for further analysis. Scripts for the homozygosity mapping performed can be found in the Appendix section C.3.

5.2.3 Whole exome sequencing variant filtering

A series of steps were taken to reduce the list of candidate variants. Using tools provided by the Galaxy web-based platform (http://galaxyproject.org/), the variant reports were uploaded for variant filtering (Blankenberg et al., 2010; Goecks et al., 2010). Figure 5.1 shows a flow chart describing the rationale for the variant filtering (refer to Appendix section F for full details of the protocol used for Galaxy analysis).

Firstly, WES datasets from the two affected brothers (IV:7 and IV:9) were compared and all shared variants were retained under the assumption that the two brothers would have the same mutation (Dataset 1). To filter out variants that were present within the family but not associated with the disease, Dataset 1 was compared with the WES data of the unaffected family members (III:9, III:10 and IV:8) and only variants unique to the affected brothers and absent in the other unaffected family members were retained for investigation (Dataset 2). To filter out common variants that

are present in unrelated neurologically normal control individuals, Dataset 2 was compared to the WES data of twenty healthy Caucasian and one ethnically matched healthy Indian male. Variants that were shared between the two affected brothers and were absent in unaffected family members and unrelated healthy controls were retained for investigation (Dataset 3).

After combining the results of the linkage analysis and homozygosity mapping, variants that are localised in the overlapping suggestive linkage interval and ROH were prioritised for further investigation (Dataset 4). Due to the autosomal recessive inheritance of the disease and the consanguineous marriage between the parents, homozygous variants with MAF below 1% in Dataset 4 were selected for further analysis (Dataset 5). The MAF cut off value of 1% was determined based on the prevalence of CMT (1 in every 2500 people) and has been applied in other autosomal recessive studies including autosomal recessive CMT (Bamshad et al., 2011; Higuchi et al., 2016).

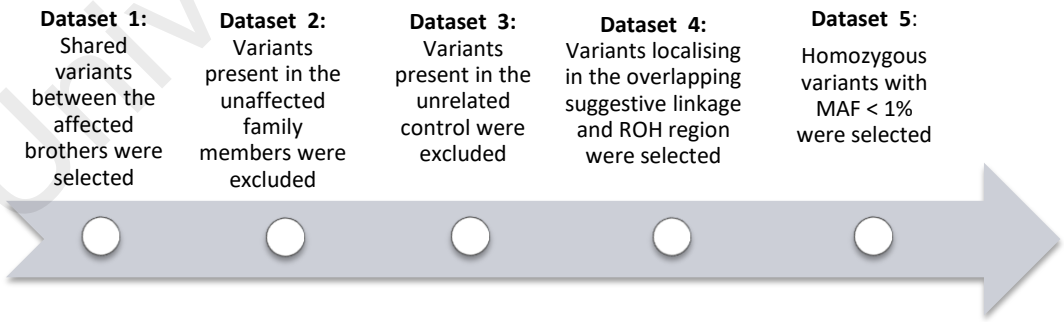


Figure 5.1: Variant filtering strategy to identify candidate pathogenic variants in CMT861.

5.2.4 Validation and segregation analysis of candidate variants

Candidate variants detected by WES were validated by Sanger sequencing and assessed for segregation with the disease phenotype in CMT861. The protocols for primer design, PCR and Sanger sequencing data analysis are described previously (Chapter 2 section 2.4, 2.5 and 2.7 respectively). The primers used for variant validation and segregation analysis in this chapter are listed in the Appendix (Table A.3).

5.2.5 Screening additional controls for the candidate variants with high resolution melting analysis

The initial WES data filtering was performed using the WES data from a limited number of controls. To further assess the candidate variants, additional controls were screened to provide supporting evidence for the likely pathogenicity of the variants. DNA samples of neurologically normal individuals were obtained from a DNA repository at the ANZAC Research Institute, Northcott Neuroscience Laboratory (760 Caucasian controls) and from a DNA bank at the University of Malaya (140 Indian controls). HRM technique was used to screen the variants (details of HRM protocols can be found in Chapter 2, section 2.6). The HRM primers used for SNP genotyping are listed in the Appendix (Table A.4).

5.2.6 Conservation analysis of the candidate variants

The conservation of nucleotides and amino acid residues affected by the candidate variants was evaluated using the GERP, PhastCons and PhyloP tracks within the UCSC Genome Browser. Genomic evolutionary rate profiling (GERP) assesses the conservation of a sequence across different species using a multiple alignment scoring method (Cooper *et al.*, 2005). The GERP score ranges from -12.3 to 6.17, with 6.17 indicating the most conserved sequence. PhastCon estimates the conservation of a sequence based on the multiple alignments of the genes across species and identifies

runs of conservation elements (Siepel & Haussler, 2004). The degree of conservation is described with a number between 0 and 1, with a higher score indicating higher conservation. PhyloP predicts the conservation and the rate of evolution under natural drift (Siepel et al., 2006). The sites predicted to be conserved are assigned with positive scores while sites predicted to be fast-evolving and less conserved are assigned with negative scores.

5.2.7 Predicting pathogenicity of the candidate variants

Candidate variants were evaluated for potential pathogenicity using several online pathogenicity prediction programs including Sorted Intolerant From Tolerant (SIFT) (Kumar et al., 2009), Polyphen-2 (Adzhubei *et al.*, 2010), Protein Variation Effect Analyzer (PROVEAN) (Choi et al., 2012), MutationAssessor (Reva et al., 2011), MutationTaster (Schwarz et al., 2010), Functional Analysis Through Hidden Markov Models (FATHMM) (Shihab *et al.*, 2013) and Screening for Non Acceptable Polymorphisms (SNAP) (Bromberg & Rost, 2007). The range of scores for each prediction program is included in Table 5.4 (results section below).

5.3 Results

The following sections describe the results for homozygosity mapping using WES data and heuristic filtering of WES data to identify the likely causal variant. The candidate variants identified that are localised within the loci supported by homozygosity mapping and linkage analysis were prioritised for further investigation.

5.3.1 Two regions of homozygosity were detected in CMT861

Using the default homozygosity mapping parameters of PLINK, 55635 markers annotated using the HapMap II CEU population annotation file were extracted from the WES data of the proband (IV:7), the affected brother (IV:9), the unaffected brother (IV:8) and the carrier parents (III:9 and III:10). Two ROHs in the affected brothers were identified on chromosome 11 and 14 (Table 5.1).

Whilst linkage analysis had excluded chromosome 11 (multipoint LOD score < -2) as a candidate disease locus (Chapter 4, section 4.3.3.1, Figure 4.2), PLINK analysis identified a ROH on chromosome 11q13.1-q13.2, spanning approximately 2.63 Mb (chr11:65,294,223-67,924,904). The 2.63 Mb ROH that is partially shared between the affected brothers and the carrier parents is located between the markers rs7107453 and rs2286163 (Table 5.1, Figure 5.2). The ROHs exclusively present in the affected brothers included a proximal interval located between the markers rs1622368 and rs7107453 (0.25 Mb) and a distal interval located between the markers rs2286163 and rs7121441 (0.13 Mb), (Figure 5.2).

The second interval of ROH shared between the affected brothers was identified on chromosome 14q32.2-q32.33 and defined a 7.48 Mb interval flanked by the markers rs2124843 and rs4983409 (Table 5.1, Figure 5.3). This ROH was shared between the affected brothers and was absent in the carrier parents and the unaffected brother.

Table 5.1: Regions with ROH detected in CMT861 using variants extracted from WES data.

ROH shared between the affected brothers (indicated by an asterisk). ROH = runs of homozygosity; SNP1 = first marker of the ROH; SNP2 = last marker of the ROH; BP1 = genomic location of SNP1; BP2 = genomic location of SNP2; NSNP = number of homozygous SNPs analysed within the ROH; Size = interval of DNA in Mb; CON = contig, the overlapping homozygous DNA segment shared between both the affected brothers and any other family members with ROH fulfilling the PLINK criteria.

Chromosome	ID	SNP1	SNP2	BP1	BP2	Size (Mb)	NSNP
11	III:10	rs658524	rs2286163	65647260	67793509	2.15	112
	III:9	rs7107453	rs1790736	65545306	67226249	1.68	100
	IV:7*	rs1622368	rs7121441	65294223	67924904	2.63	142
	IV:9*	rs1622368	rs7121441	65294223	67924904	2.63	142
	CON	rs658524	rs1790736	65647260	67226249	1.58	90
14	IV:7*	rs2124843	rs4983409	98435696	105916420	7.48	157
	IV:9*	rs2124843	rs4983409	98435696	105916420	7.48	157
	CON	rs2124843	rs4983409	98435696	105916420	7.48	157

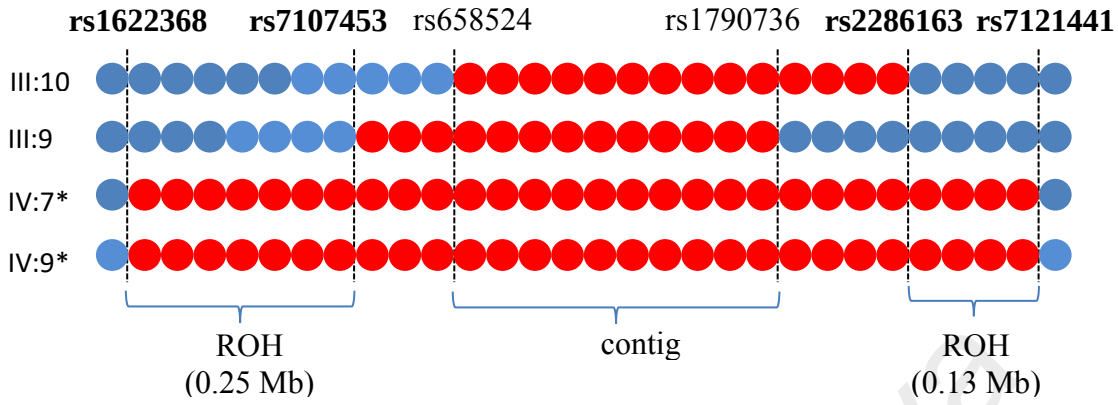


Figure 5.2: Runs of homozygosity (ROH) on chromosome 11q13.1-q13.2.

The red circles indicate the regions of homozygosity. The affected brothers are marked with an asterisk (*). The ROHs shared by the affected brothers correspond to the genomic intervals between rs1622368 and rs7107453, as well as rs2286163 and rs7121441.

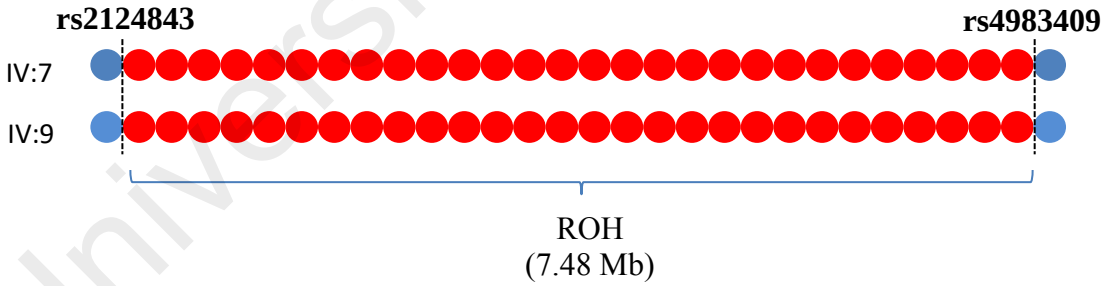


Figure 5.3: Runs of homozygosity (ROH) on chromosome 14q32.2-q32.33.

The red circles indicate the regions of homozygosity. This ROH shared between the affected brothers (IV:7 and IV:9) spans a 7.48 Mb interval between the markers rs2124843 and rs4983409. This ROH was absent in the carrier parents and the unaffected brother.

5.3.2 Linkage and homozygosity mapping data support chromosome 14q32.2-q32.33 as the candidate locus

Merging of linkage and homozygosity mapping data suggests that a new disease locus for AR-CMT is likely to be located on chromosome 14q32.2-q32.33. The genomic co-ordinates defining the ROH identified on chromosome 14 overlap with the suggestive linkage peak previously identified from genome-wide linkage analysis (Table 5.2, Figure 5.4). Importantly, the homozygosity mapping refined the suggestive linkage interval on chromosome 14 from 8.12 Mb to 7.48 Mb and this will aid the identification of possible causal variants when querying the WES data.

The segment of ROH on chromosome 11 is short and is likely to have occurred by chance. Moreover, linkage analysis excluded chromosome 11 as a possible disease locus. Although a suggestive linkage region (multipoint LOD score of 1.74) was identified on chromosome 12 (discussed in Chapter 4), no ROH was identified for this chromosome. The absence of ROH adds additional evidence suggesting that the locus is unlikely to harbour the causal gene responsible for AR-CMT in this family (Table 5.2).

Table 5.2: Summary of merging linkage and homozygosity mapping data.

The ROH detected on chromosome 11 has previously been statistically excluded (LOD scores < -2) from being linked to the disease locus by microsatellite markers linkage analysis. ROH was absent within the suggestive linked locus on chromosome 12. Therefore, it is unlikely that the causal variant is located within these two loci. The ROH identified on chromosome 14 overlaps with the suggestive linkage region on chromosome 14q32.2-32.33 (Figure 5.4).

Chr.	LOD scores	Regions of ROH	Conclusion after merging linkage and homozygosity mapping results
11	• Linkage excluded (LOD < -2)	• Short segments of ROH detected	• Unlikely to be the disease locus
12	• Maximum multipoint LOD score of 1.74	• No ROH detected	• Unlikely to be the disease locus
14	• Maximum multipoint LOD score of 2.06	• ROH detected	• Supported and remains a strong candidate locus.

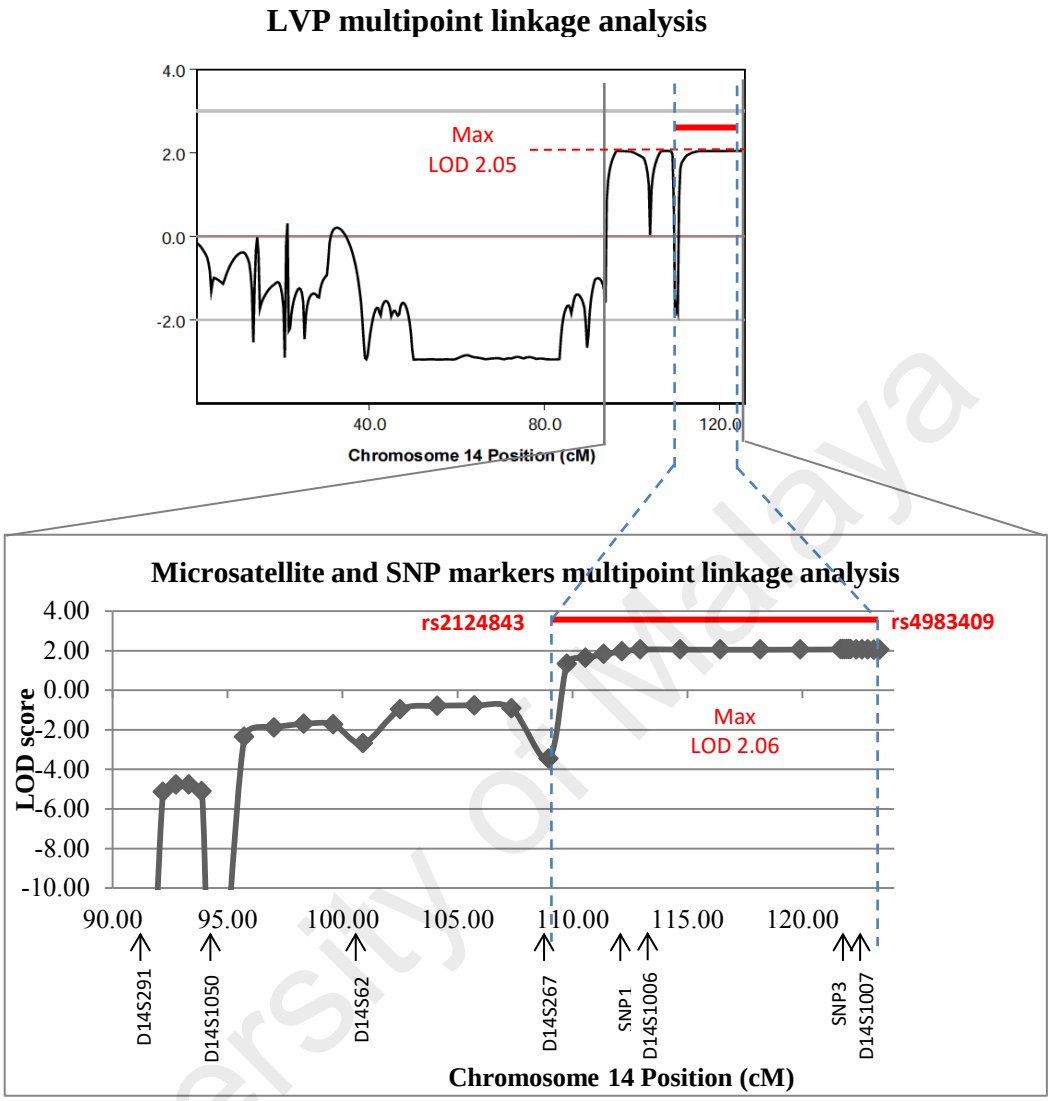


Figure 5.4: Combined LOD scores and ROH data support a region on chromosome 14 as a candidate locus for AR-CMT in CMT861.

The multipoint LOD score likelihood curve defining the suggestive linkage region on chromosome 14q32.2-q32.3 spans ~8.12 Mb. The red bar represents the ROH, spanning a 7.48 Mb interval between marker rs2124843 and rs4983409. The blue dotted lines represent the overlapping interval of suggestive linkage and ROH.

5.3.3 WES data filtering identifies seven candidate genes

Variant filtering of WES data from five family members (III:9, III:10, IV:7, IV:8, and IV:9) was performed. An average of 80,000 variants was detected for each WES dataset. A series of filtering steps were applied to reduce the list of probable candidate gene variants (Figure 5.5) and were based on the following assumptions: 1) the affected brothers share the same causal variant, 2) the causal variant is in a homozygous state in the affected individuals, 3) the variant is unreported or has a MAF < 1% in the heterozygous state and 4) the unaffected family members or unrelated control individuals do not carry the causal variant in the homozygous state.

Comparison of the variant dataset in the two affected brothers identified a total of 51,653 shared variants (Dataset 1). Dataset 1 was then compared with the datasets of the three unaffected family members and common variants were excluded, enabled a substantial reduction of variants (160-fold), to 3,278 (Dataset 2). Further reduction in the number of candidate variants was not achieved when WES datasets of twenty Caucasian controls were used to remove common polymorphisms from Dataset 2. These controls are the married-in partner of our patients with neuropathy. In contrast, the use of an ethnically matched Indian control further reduced the number of candidate variants to 2,757 (Dataset 3). Within the interval where the linkage and ROH region overlap on chromosome 14, there were 65 variants including one heterozygous variant and 64 homozygous variants (Dataset 4). Eight of the 64 homozygous variants had a MAF < 1% (Dataset 5). Annotation of the eight variants showed that two were exonic and six were situated in non-coding regions (Figure 5.5, Table 5.3).

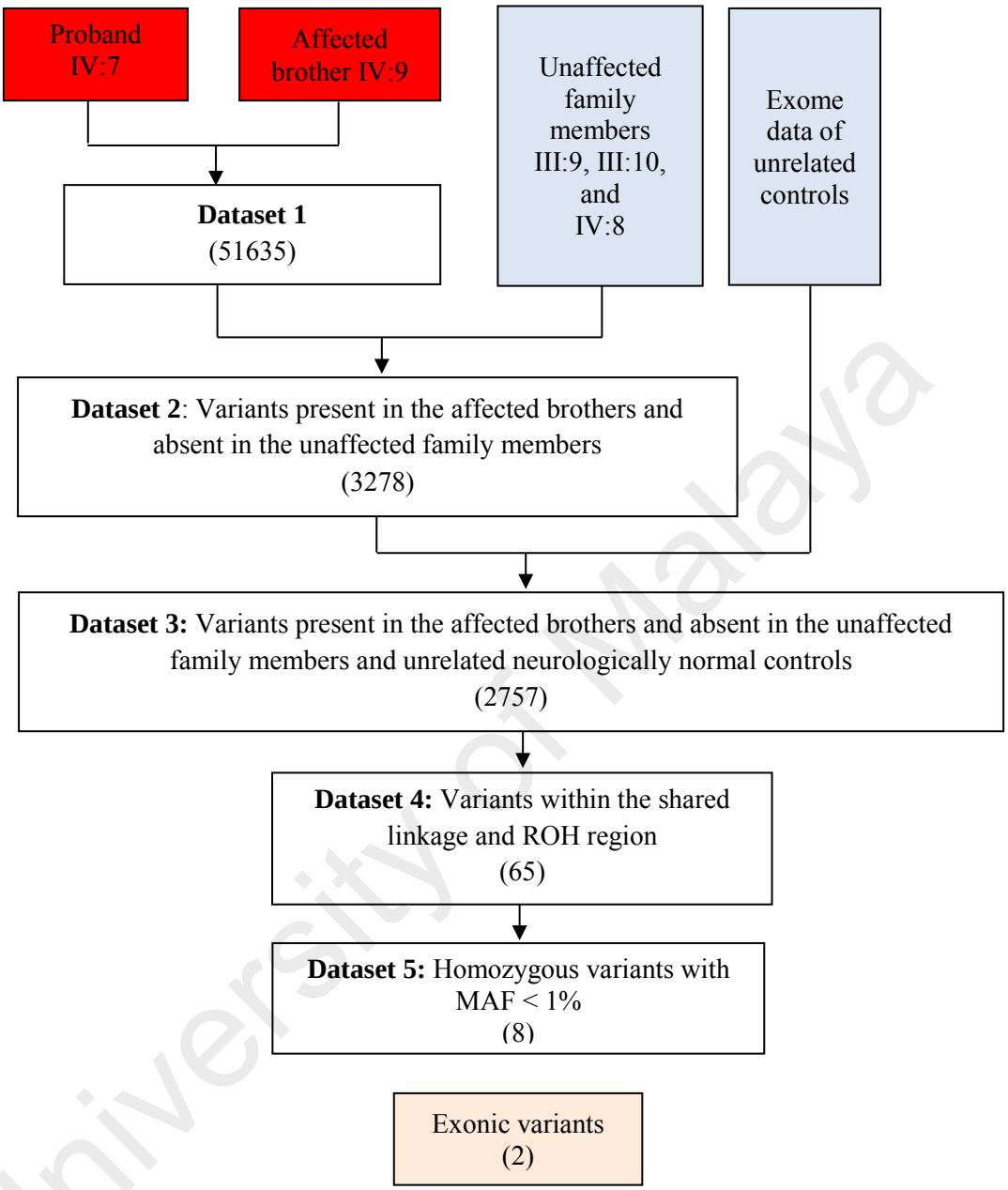


Figure 5.5: Strategy to identify candidate variants on chromosome 14q32.2-q32.33 by variant filtering of WES data.

The number of variants after each filtering step is shown in brackets.

5.3.4 Evaluating variants within the chromosome 14 candidate interval

For the non-coding variants, one is located in the 3' UTR of the *WARS* gene, which encodes the protein tryptophanyl-tRNA synthetase, one of the aminoacyl-tRNA synthetase (ARS) family members (Fleckner et al., 1991). Mutations in other ARS genes including *YARS*, *KARS*, *AARS*, and *GARS* have been reported to cause CMT (Antonellis et al., 2003; Jordanova et al., 2006; Latour et al., 2010; McLaughlin et al., 2010). However, those mutations changed the coding sequence of amino acids while the variant detected in this study was in the 3' UTR.

The other non-coding variants were detected in the 3' UTR of the *BEGAIN*, *AKT1*, and *ZNF839* genes. *BEGAIN* encodes for the brain-enriched guanylate kinase-associated protein, which is localised at the dendrites and synapses of neurons, suggesting a role in synaptogenesis (Yao et al., 2002). *AKT1* encodes for a serine-threonine protein kinase which acts as a mediator of growth factors in nervous system and is also an important factor regulating the apoptotic machinery (Franke et al., 1997; Romashkova & Makarov, 1999). *ZNF839* encodes for the zinc finger protein 839 and its function is still unknown (Franke et al., 1997; Romashkova & Makarov, 1999). To date, none of these genes have been reported to have a role in neurological diseases.

Intronic non-coding variants were identified for the *MOK* and *KLC1* genes. *MOK* plays a role in cell differentiation in the intestinal membrane (Chen et al., 2013), while *KLC1* encodes the kinesin light chain 1 protein. Although mutations in *KLC1* have not been reported in CMT or other disease, its role as part of the kinesin anterograde transport motor complex (Hirokawa et al., 2009) makes the intronic change a possible pathogenic candidate variant.

Two exonic variants were identified in the *AHNAK2* gene. *AHNAK2* encodes for AHNAK nucleoprotein 2, which may play a role in regulating calcium channels in

cardiac muscles and in establishing costameric networks that interact with cytoskeleton proteins in skeletal muscles (Komuro et al., 2004; Marg et al., 2010). The protein is also involved in the FGF1 secretion pathway under stress-induced environments (Kirov et al., 2015). Interestingly, *AHNAK2* is a homologue of the periaxin gene (*PRX*) (Han & Kursula, 2014) and mutations in *PRX* have been reported to cause autosomal recessive CMT4F (Boerkoel et al., 2001; Guilbot et al., 2001).

Although variants within non-coding DNA can have a pathological impact on gene expression, gene transcription and DNA methylation (Mercer et al., 2009) and have been reported in cancers (Fang & Fullwood, 2016), epilepsy (Lee et al., 2015a) and Alzheimer's disease (Faghihi et al., 2008; Lee et al., 2015b), the exonic variants in the *AHNAK2* gene were prioritised for further investigation. Exonic variants resulting in non-synonymous amino acid changes represent a potential direct impact on the protein structure and function. AHNAK2 is able to form heterodimers with PRX via their PDZ domain (Han & Kursula, 2014) and *PRX* mutations are known to cause CMT4F, suggesting *AHNAK2* as a strong candidate gene underlying the neuropathy in CMT861.

Table 5.3: Candidate variants within the 7.48 Mb interval on chromosome 14q32.2-32.33.

	Gene	Nucleotide change	Amino acid change	Location of variant	SNP ID	MAF (dbSNP142)
Coding variant	*AHNAK2* (NM_138420)	c.A118C	p.T40P (nonsynonymous)	Exon 3	Novel	-
	AHNAK2 (NM_138420)	c.C2743T	p.H915Y (nonsynonymous)	Exon 7	rs375578054	0.1%
Non-coding variant	*WARS* (NM_004184.3)	c.*1053C>T	-	3' UTR	rs568623036	0.44%
	BEGAIN (NM_001159531.1)	c.*565 G>A	-	3' UTR	Novel	-
	MOK (NM_001272011.1)	c.892-28G>A	-	Intronic	rs550281723	0.34%
	ZNF839 (NM_018335.4)	c.*40_*44delTGGAG	-	3' UTR	rs57469946	NA
	KLC1 (NM_001130107.1)	c.492+4A>G	-	Intronic	Novel	-
	AKT1 (NM_005163.2)	c.*313T>C	-	3' UTR	rs770697761	NA

5.3.5 Investigating the *AHNAK2* exonic variants

The variants reported in *AHNAK2* have been annotated using the accession sequence number NM_138420 (Table 5.3). A transition variant (c.A118C) in exon 3 is predicted to cause a threonine to proline (p.T40P) substitution. This is a novel variant and was absent in the current online databases including dbSNP 144, EVS, ExAC and 1000 Genomes release 2013. The second exonic variant in exon 7, (c.C2743T; p.H915Y) is a reported SNP (rs375578054) with a MAF of 0.1%.

5.3.5.1 Sequence validation and segregation analysis of the c.A118C and c.C2743T variants

Sanger sequencing confirmed the presence of the homozygous variants (Figure 5.6). Segregation analysis was performed using all available family members (Figure 5.7). The parents and maternal grandmother (III:9, III:10 and II:2) and two aunts (III:2, III:8) were heterozygous for both the c.A118C and c.C2743T variants. The homozygous state for the alternate allele was only present in the affected brothers (IV:7 and IV:9).

a)

Family members of CMT861 who carry the candidate variants	Variant: c.A118C; p.T40P (Chr. coordinate: 14:105423842)			Variant: c.C2743T; p.H915Y (Chr. coordinate: 14:105419045)		
	Total reads	**Alternate reads**	**Zygosity**	**Total reads**	**Alternate reads**	**Zygosity**
III: 9 (Father)	97	42	heterozygous	68	36	heterozygous
III:10 (Mother)	95	52	heterozygous	84	29	heterozygous
IV:7 (Proband)	48	47	homozygous	69	65	homozygous
IV:9 (Affected brother)	45	45	homozygous	112	112	homozygous

b)

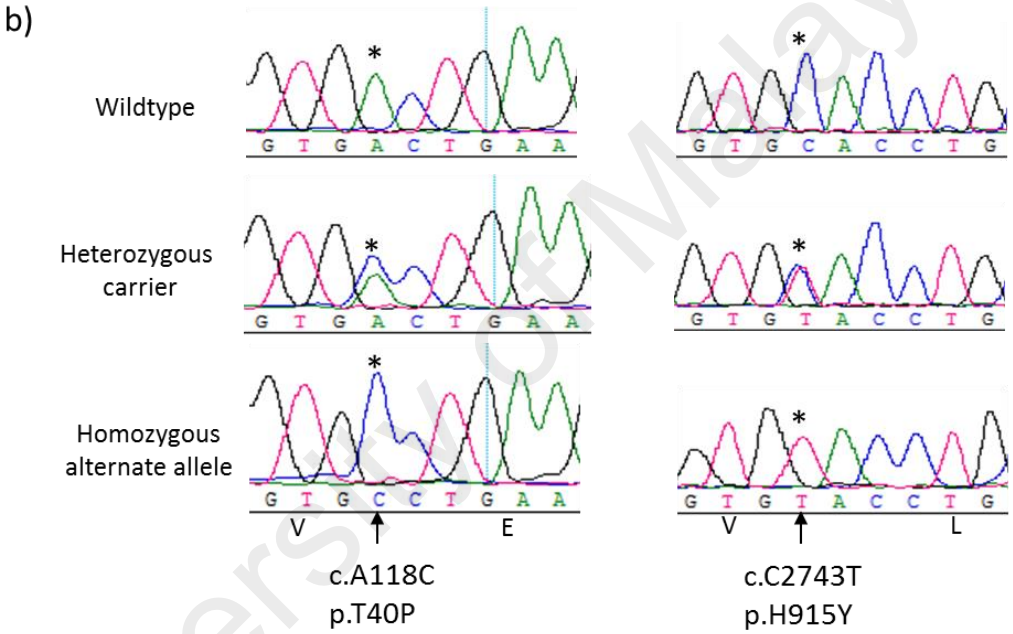


Figure 5.6: Validation of *AHNAK2* variants, c.A118C and c.C2743T.

a) Whole exome sequencing variant report for the c.A118C (p.T40P) and c.C2743T (p.H915Y) variants in CMT861. The total and alternate reads for the nucleotide at the position of base change indicate that the parents are the carriers and the affected brothers are homozygous for the alternate allele. b) Sequence traces showing the affected brothers, the carrier parents and wildtype controls for the p.T40P and p.H915Y variants. An asterisk denotes the base change resulting in missense mutations that segregate with the disease phenotype in the family. The GenBank sequences NM_138420 and NP_612429.2 were used as reference sequences for *AHNAK2* and the AHNAK2 protein.

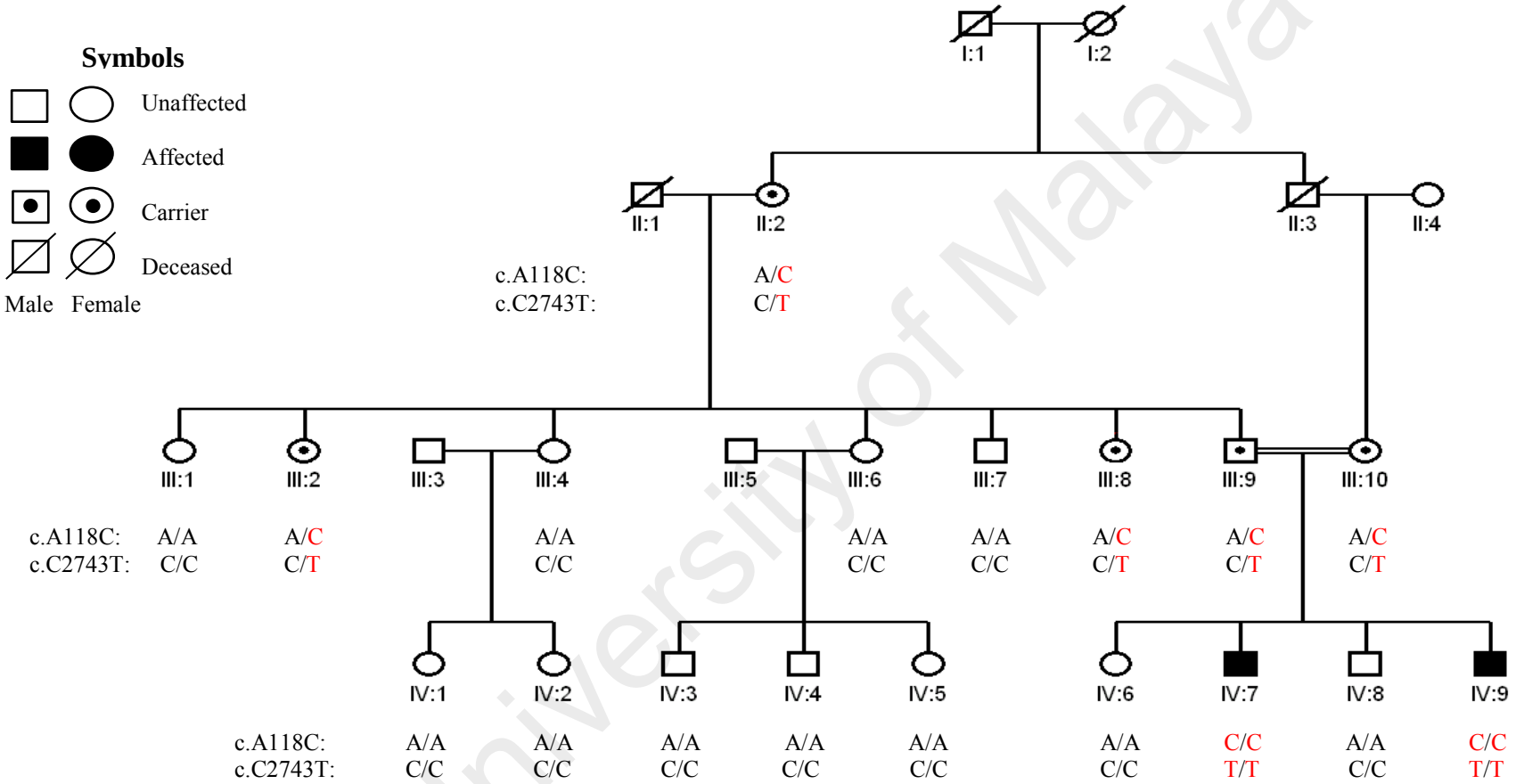


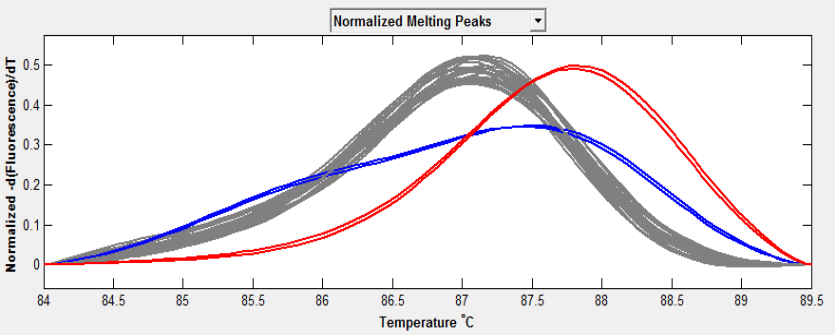
Figure 5.7: Segregation analysis of the *AHNAK2* variants, c.A118C and c.C2743T for CMT861.

5.3.5.2 The *AHNAK2* variants are rare and only present in the homozygous state in the affected individuals of family CMT861

The c.A118C variant was absent in 1772 chromosomes after additional screening of neurologically normal controls, including 264 chromosomes from individuals with Indian ethnicity (Figure 5.8).

The second variant c.C2743T was absent in 400 chromosomes of neurologically normal individuals, including 110 chromosomes from individuals with Indian ethnicity (Figure 5.9). This variant was reported in the heterozygous state in 5 out of 5008 (0.1%) chromosomes in the 1000 Genomes database, and 19 out of 120144 (0.01%) chromosomes in the ExAC database. No genotype in the homozygous state was reported for the c.C2743T variant.

a)



b)

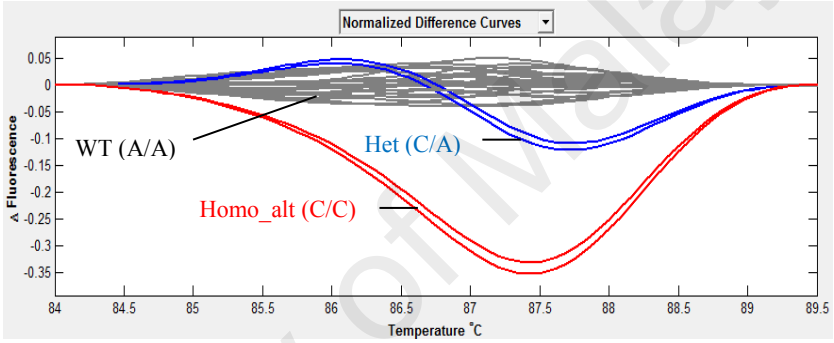
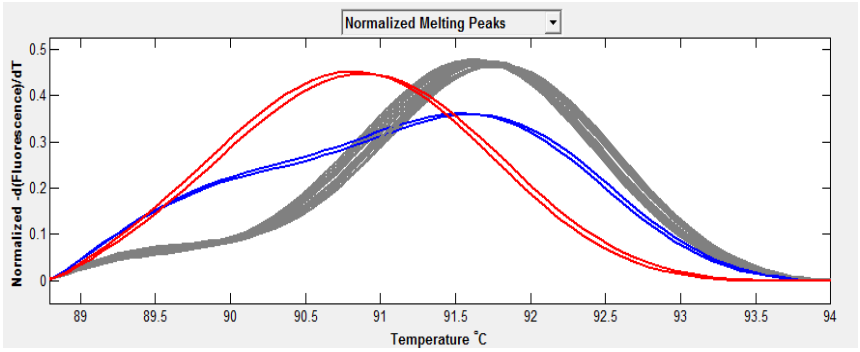


Figure 5.8: Genotyping the *AHNAK2* c.A118C variant using HRM analysis.

a) Normalised derivative fluorescent melt curves and b) normalised difference curves for controls (A/A), carriers (A/C) and homozygote patients (C/C) genotypes. Red curves: Homozygous for the alternate allele (Homo_alt); blue curves: Heterozygous (Het); grey curves: Homozygous wildtype allele (WT).

a)



b)

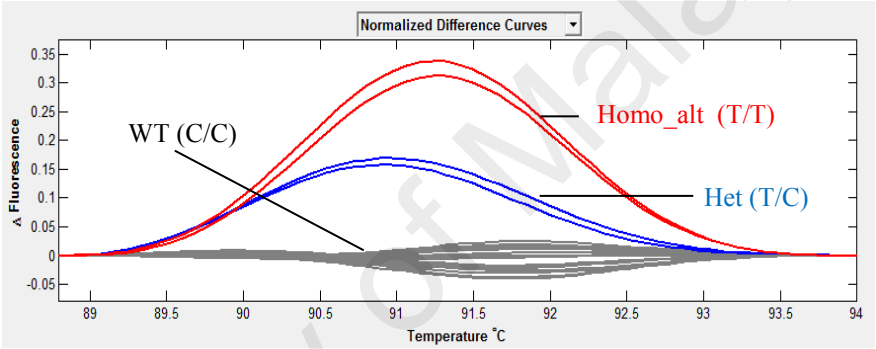


Figure 5.9: Genotyping the *AHNAK2* c.C2743T variant using HRM analysis.

a) Normalised derivative fluorescent melt curves and b) normalised difference curves for controls (C/C), carriers (T/C) and homozygote patients (T/T) genotypes. Red curves: Homozygous for the alternate allele (Homo_alt); blue curves: Heterozygous (Het); grey curves: Homozygous wildtype alelle (WT).

5.3.5.3 The c.A118C variant is predicted to be more damaging compared to the c.C2743T variant

Several prediction programs were used to assess the potential pathogenicity of the two *AHNAK2* variants (Table 5.4). For the c.A118C variant (p.T40P), the wild type threonine residue has a GERP score of 3.21, a PhastCons score of 0.99 and a PhyloP score of 2.63, indicating a high level of conservation. The wildtype histidine residue for the c.C2473T variant (p.H915Y) has a GERP score of 0.69, a PhastCons score of 0 and a PhyloP score of 0.39, indicating a low level of conservation.

The c.A118C variant was predicted to be 'damaging' by SIFT, Polyphen-2, and FATHMM, while it was deemed to have a neutral effect by PROVEAN, MutationTaster and MutationAssessor. In SNAP, this variant scored +35, which indicates that it has an intermediate effect on the protein function. The prediction tools indicated a milder impact of the other variant, c.C2743T as only Polyphen-2 predicted a possible 'damaging' effect. Additionally, SNAP scores indicated a milder negative impact on the protein function of this variant compared to the c.A118C variant.

Table 5.4: Conservation and pathogenicity prediction of the *AHNAK2* variants.

	Programs	*AHNAK2*, c.A118C; p.T40P	*AHNAK2*, c.C2743T; p.H915Y	Range of test scores
Conservation	**PhastCons**	0.99	0	Between 0 to 1; high score = more conserved
	PhyloP	2.63	0.39	-20 to 7.53; high score = more conserved
	GERP	3.21	0.69	-12.3 to 6.17; high score = more conserved
Pathogenicity prediction	**SIFT**	0.0 (Damaging)	0.065 (Tolerated)	Between 0 to 1; <0.05 is predicted damaging
	PolyPhen-2	0.998 (Probably damaging)	0.995 (Probably damaging)	Between 0 to 1; >0.9 is predicted damaging
	PROVEAN	-2.1 (Neutral)	-1.6 (Neutral)	Between -13 to 4; <-2.5 is predicted damaging
	MutationTaster	Polymorphism	Polymorphism	Disease-causing = probably deleterious; polymorphism = probably harmless
	MutationAssessor	1.545 (Low)	1.87 (Low)	Neutral, low, medium, high
	FATHMM	-4.86 (Damaging)	-0.79 (Tolerated)	<-3 is predicted damaging
	SNAP	35	15	Severity of effects: between -100 (Neutral) to 100 (Severe)

5.4 Discussion

Homozygosity mapping using WES data from CMT861 was performed as a complementary approach to provide further genetic evidence for the suggestive linkage loci identified earlier. The merging of linkage and homozygosity mapping data helped to refine the likely genetic locus for AR-CMT in CMT861 to a 7.48 Mb region on chromosome 14q32.2-q32.33. The WES data was screened for probable candidate variants within this interval using a series of intuitive filtering strategies and successfully identified several candidate variants.

The PLINK program was used for homozygosity mapping as it has been reported to have a better sensitivity in detecting ROH compared to BEAGLE or GERMLINE which are the alternative tools (Howrigan et al., 2011). This program uses SNP markers to identify continuous stretches of homozygous variants in the genome. The higher the density of SNP markers, the better the accuracy of ROH that can be acquired (Ferencakovic et al., 2013). This study used SNP markers extracted from the WES data using the HapMap II annotation file, which had ~4 million SNP markers. The HapMap III annotation file was not selected as it contained only ~1.5 million SNP markers.

In this study, the suggestive genetic linkage interval on chromosome 14 was supported by the homozygosity mapping. Taken together the results suggest that this locus represents the most likely region to harbour the mutation causing AR-CMT in CMT861. In addition, the size of the ROH identified on chromosome 14 (7.48 Mb) fell within the strongly correlated size range of ROHs (estimated to be between 4.5 to 45 Mb) seen in autosomal recessive CMT patients whose parents are first cousins (Zimon *et al.*, 2015). This locus was therefore prioritised for the identification of possible disease-causing variants.

Filtering the WES data of family members according to the disease phenotype reduced the number of variants by 95% which highlights the power of variant filtering by including the genotype data of family members. Filtering the WES data with unrelated normal controls demonstrates the power of matching the ethnicity of the control cohort. The Caucasian controls data did not reduce the number of variants as effectively as the Indian control. Filtering against a single neurologically normal Indian control reduced the number of variants further by 15%.

Eight candidate variants in seven genes were identified through WES analysis. The nonsynonymous variants in *AHNAK2* were prioritised for further investigation, due to the greater predicted pathogenic functional impact. The two variants identified in the *AHNAK2* are likely to be in linkage disequilibrium with each other, thereby confounding the possibility of eliminating one variant over the other through segregation studies. Conservation and pathogenicity predictions suggest that the novel variant in exon 3 (c.A118C) has more functional impact on the AHNAK2 protein than the variant in exon 7 (c.C2743T). However, we do not discount the possibility that both variants could be acting synergistically in a compound homozygous state.

Whilst mining the genome databases, a variant in the same p.T40 amino acid residue was identified, which in contrast to the p.T40P variant identified in this study, resulting in a missense change from threonine to isoleucine (c.C119T; p.T40I), (Figure 5.10). The p.T40I variant was reported in a heterozygous state in the ExAC database (1 out of 120,848 chromosomes), and was not detected in our own 1772 control chromosomes. Notably, there was no report of our p.T40P variant in the 120,848 chromosomes in the ExAC database. It is likely that this rare p.T40I variant is non-pathogenic in the heterozygous state and is similar to the phenotype observed in the unaffected, heterozygote p.T40P carriers in CMT861. Pathogenicity prediction using

SIFT and Polyphen-2 suggest that the p.T40I variant is probably damaging, similar to the prediction for the p.T40P variant. In the heterozygous state, the wildtype allele might be sufficient to compensate for the mutant allele. However, an individual who is homozygous for the p.T40I alternate allele may manifest the disease due to loss-of-function of the protein. The presence of rare heterozygous pathogenic variants in the general population is relatively common especially in the case of autosomal recessive disorders, in which the carriers do not show the disease phenotype (Brunham & Hayden, 2013; Feder et al., 1996). Due to the low probability that the rare heterozygous variants being transmitted together when unrelated individuals have offspring, the pathogenicity of these variants in the homozygous state is often under recognised. In our family, due to consanguinity, rare heterozygous variants such as the ones in *AHNAK2* have a higher chance of being transmitted to the affected sons in the homozygous state and would fit an autosomal recessive model of inheritance in the family.

To investigate if there are other families with *AHNAK2* mutations, we screened the WES datasets of 115 genetically unclassified IPN patients in the ANZAC Neuroscience Laboratory's exome database. Furthermore, we performed HRM analysis on additional 30 index IPN patients with suggestive autosomal recessive CMT, in which WES data was unavailable. Over 3000 homozygous and heterozygous variants were detected in *AHNAK2* and all these variants were located in exon 7 (Figure 5.10). There were no variants detected in exons 1-6. Shared results from our collaboration with another CMT research group also indicated that there were several *AHNAK2* variants in exon 7 in their AR-CMT families, and none in exons 1-6. Two of these variants (p.I4729M and p.R448W) in exon 7 were found not segregating with the disease in the respective families (Unpublished data from our collaborator). Collectively, these data show that the exon 7 is highly polymorphic and suggest that the variant c.C2743T (p.H915Y) is a rare non-pathogenic polymorphism, which is in linkage disequilibrium with p.T40P. On

the other hand, exons 1-6 are highly conserved and the novel variant in exon 3 (c.A118C; p.T40P) is highly suggestive as the pathogenic mutation.

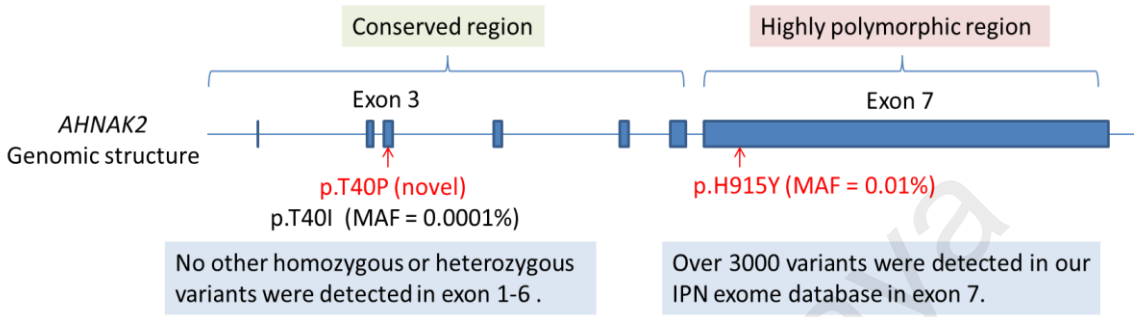


Figure 5.10: Likely exclusion of the p.H915Y (c.C2743T) variant due to the highly polymorphic nature of exon 7.

Schematic diagram of the genomic structure of *AHNAK2*. This gene is composed of seven exons with each exon represented by a blue solid bar. The novel variant, p.T40P, identified in this study and the p.T40I variant, reported on the ExAC database are located within a conserved region of the gene while p.H915Y is located within exon 7, which is highly polymorphic.

In this chapter, the strategy of using homozygosity mapping in combination with linkage analysis has defined a new locus for AR-CMT in CMT861. WES analysis and variant filtering strategies have identified *AHNAK2* as a likely candidate gene for autosomal recessive CMT in family CMT861. The data presented in this chapter support c.A118C (p.T40P) as the likely pathogenic variant. Chapter 6 investigates the impact of the p.T40P mutation in more detail with respect to the *AHNAK2* expression profile in the patient's fibroblasts.

CHAPTER 6: ASSESSING *AHNAK2* AS A CANDIDATE GENE FOR AUTOSOMAL RECESSIVE CMT

6.1 Introduction

Using a combination of linkage and homozygosity mapping, a genetic locus for the AR-CMT phenotype in CMT861 was mapped on chromosome 14. Whole exome sequencing identified a putative causative variant in the *AHNAK2* gene (c.A118C; p.T40P) which is localised within the chromosome 14 interval. In this chapter, the effects of the p.T40P variant are explored by examining the levels of *AHNAK2* expression in patient fibroblasts. The spatiotemporal expression profile of the *Ahnak2* gene is also investigated in mice to determine if *Ahnak2* is expressed in neuronal tissues affected in the AR-CMT phenotype.

6.1.1 AHNAK2 protein structure and known functions

The *AHNAK2* gene encodes a 7 exons transcript for the AHNAK nucleoprotein 2 (AHNAK2), a large 616 kDa protein composed of 5795 amino acids (Komuro et al., 2004). This gene was first described in a study of AHNAK1 in which the AHNAK1 antibody detected another large protein, which was later designated as AHNAK2 (Komuro et al., 2004). Subsequent work showed that the genomic structure of these two genes was similar, consisting of a PDZ domain at the N-terminal and a large central region of long repetitive motifs (Komuro et al., 2004; Shtivelman et al., 1992), (Figure 6.1). However, the two proteins share only 14% identity for the entire amino acid sequence (data obtained from Ensembl).

The functional role of AHNAK2 is poorly understood. In mice, Ahnak2 was found to co-localise with Ahnak1 at the Z-band region of cardiac sarcomeres and the costameres of skeletal muscles (Komuro et al., 2004; Marg et al., 2010). Based on

similar domains, structure and cellular co-localisation between the two proteins, Ahnak2 is predicted to have a similar functional role to Ahnak1 in regulating calcium channels and establishing costameric networks interacting with multiple cytoskeleton proteins such as dysferlin, annesin 2, myoferlin, actin and dystrophin (Hohaus et al., 2002; Huang et al., 2007; Komuro et al., 2004; Marg et al., 2010). Interestingly, homozygous *Ahnak1* knockout mice exhibit no obvious phenotype (Komuro et al., 2004). It was hypothesised that the absence of a distinct phenotype in the *Ahnak1*$^{-/-}$ mice could be due to a compensatory role of *Ahnak2* (Komuro et al., 2004).

A study recently found that AHNAK2 is associated with the secretion of fibroblast growth factor-1 (FGF1) under stressed cellular environments (Kirov et al., 2015). Following the knock-down of endogeneous *Ahnak2* in mouse fibroblasts, a tenfold decrease in Fgf1 secretion under heat-shocked conditions was observed, indicating a role of Ahnak2 in Fgf1 export (Kirov et al., 2015). FGF1 belongs to the fibroblast growth factor family and regulates a wide range of cellular events including cell proliferation, growth, migration, differentiation, survival, tissue repair and angiogenesis (Raju *et al.*, 2014). In neurons, FGF1 was proposed to be an important growth factor for axon regeneration and has been used to enhance the growth of artificial nerve grafts in patients with nerve injury (Kirchmair et al., 2007; Kuffler, 2013; Laird et al., 1995; Suzuki et al., 2016).

Recently, a study using crystal structures of the PDZ domain of human AHNAK2 and periaxin (PRX) has shown that these proteins have similar domains and are able to form heterodimers (Han & Kursula, 2014). The PDZ domains of both AHNAK2 and PRX belong to the same subfamily (cd00992; reported on NCBI), and are highly similar to each other with 57% amino acid sequence similarity (Han & Kursula, 2014). The PDZ domains shared by AHNAK2 and PRX suggest that both proteins may have a

similar functional role in protein-protein interaction for protein targeting and protein complex assembly (Hung & Sheng, 2002).

PRX is essential for myelin sheath maintenance (Gillespie et al., 1994) and mutations in *PRX* have been reported to cause demyelinating Charcot-Marie-Tooth disease type 4F and Dejerine-Sottas disease (CMT3), (Boerkoel et al., 2001; Guilbot et al., 2001). As PRX is known to form heterodimers with AHNAK2, it is possible that AHNAK2 may also be involved in myelination processes and mutations in *AHNAK2* might disrupt this interaction, eventually leading to peripheral neuropathy.

6.1.2 Hypothesis

The p.T40P mutation identified in *AHNAK2* will alter gene and protein expression levels.

6.1.3 Aims

1. To examine the expression of *AHNAK2* in patient fibroblasts.
2. To determine the spatial and temporal gene expression profile of *Ahnak2* in mouse neural and muscle tissues at different stages of development.

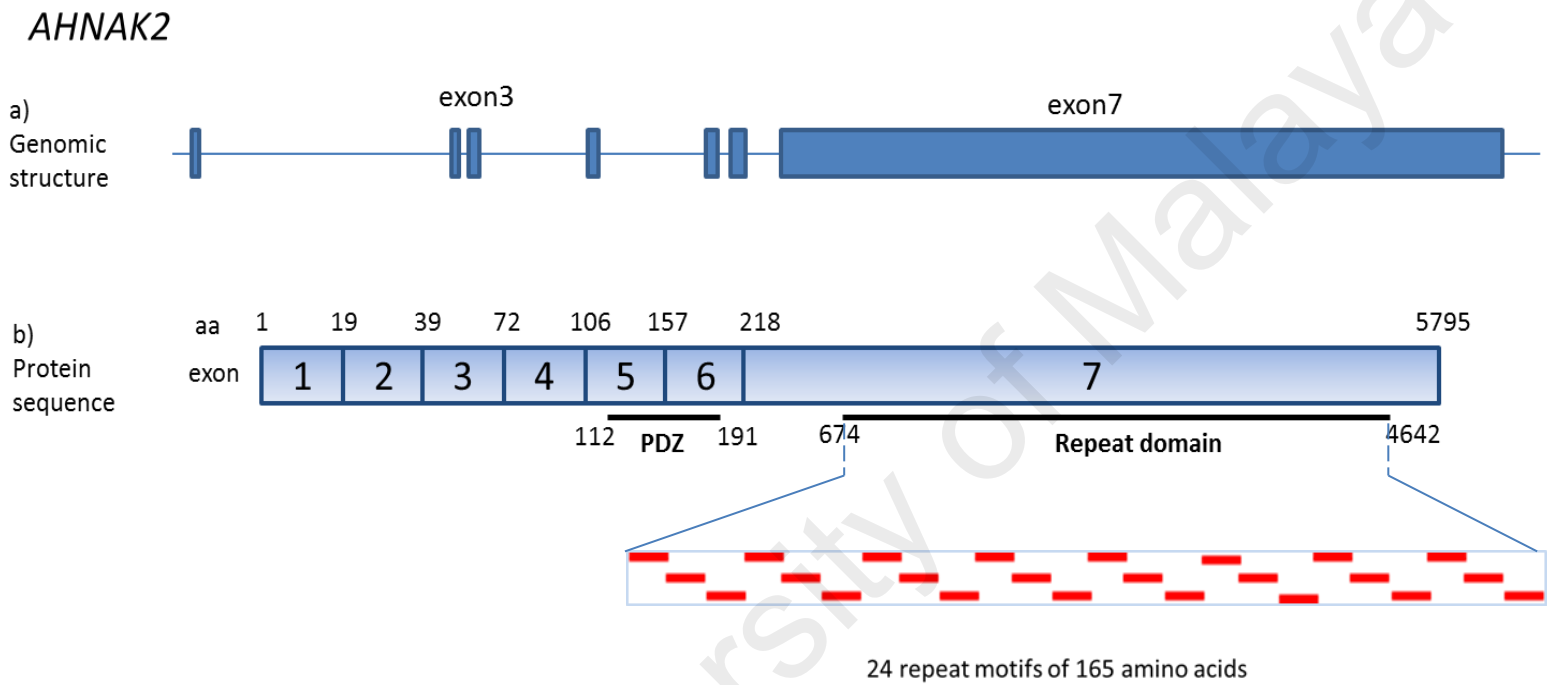


Figure 6.1: Schematic diagram of *AHNAK2* genomic and protein structures.

a) Genomic structure of *AHNAK2*. The blue blocks represent the exons. b) Protein sequence of *AHNAK2*. Each blue box represents an exon. The numbering of amino acids (aa), defining the beginning and the end of each exon are illustrated above. The PDZ domain is located in exon 5 and 6 between amino acids 112-191. The repeat domain is located in exon 7 and spans amino acids 674-4642, with 24 repeat motifs composed of 165 amino acids.

6.2 Methodology

6.2.1 Fibroblast cell culture

Ethical approval was obtained from the Medical Ethnic Committee, University of Malaya. Informed written consent was obtained from five healthy individuals and the two affected brothers for skin biopsy. A 3 mm diameter skin biopsy was obtained using a skin puncher and placed in pre-warmed Dulbecco's modified eagle medium (DMEM), (Gibco). The skin biopsy was repeatedly washed using the following steps: 15 s in 70% (v/v) ethanol, 30 s in 1X phosphate-buffered saline (PBS) and a final 2 min in 1X PBS. The skin was then transferred to a T25 cell culture flask, where it was covered with 200 µL of 1 mg/mL collagenase type I (Sigma) and cut coarsely using a sterile scalpel. The tissue was incubated in a 37°C incubator with 5% CO_2 for approximately 3 h. Then, 5 mL of pre-warmed FDMEM was added which contained 10% (v/v) inactivated fetal bovine serum (FBS), (Gibco), 200 mM L-glutamine (Gibco), and 5000 U/mL Penicilin/Streptomycin (Gibco) to inactivate the collagenase activity. The tissue was then cut into finer pieces with a small scalpel whilst making grooves in the plastic with the scalpel to encourage attachment of the tissue. The tissue was incubated at 37°C with 5% CO_2 for 2-3 weeks. The existing media was replenished with fresh media every few days or when necessary by removing half of the existing media from the flask and replacing with fresh FDMEM.

When the fibroblast cells were approximately 70% confluent, the culture media was removed and the cells were washed three times with 1X PBS. To passage in T25 flasks, 0.5 mL of 0.25% (v/v) pre-warmed Trypsin-EDTA (Invitrogen) was added followed by vigorous agitation of the flask. To enhance cell detachment, the flask was incubated at 37°C for approximately 30 s. Five mL of FDMEM was then added to the flask by pipetting the media down the wall of the flask to wash the cells. The cell suspension was collected in a 15 mL Falcon tube and centrifuged for 5 min at 160 x g at room

temperature. The supernatant was discarded and the pellet re-suspended with 5 mL of fresh FDMEM. To passage into a T25 flask, 4 mL of pre-warmed FDMEM and 0.5 mL of the cell suspension was added and incubated at 37°C with 5% CO_2.

6.2.2 RNA extraction from fibroblast cells

Fibroblast cells grown to 70% confluency were trypsinised, collected in a 15 mL Falcon tube with FDMEM and centrifuged at 160 x g (using 5417C centrifuge, Eppendorf) for 5 min at room temperature. The supernatant was discarded and the pellet was re-suspended with 1 mL of 1X PBS, transferred into a 1.5 mL microcentrifuge tube and centrifuged at 160 x g for 5 min at room temperature. The supernatant was discarded. The RNA extraction was performed using the NuclearSpin RNA kit (MACHEREY-NAGEL) according to the manufacturer protocols. Briefly, the pellet was re-suspended with 350 µL of Buffer RA1 and 3.5 µL of β-mercaptoethanol followed by vortexing to homogenise the mixture. The lysate was further homogenised by passing the cell suspension through a 0.9 mm needle for 5 times. The lysate was filtered through the NucleoSpin filter by centrifugation for 1 min at 11,000 x g. The filter was discarded and 350 µL 70% (v/v) ethanol was added to the homogenised lysate and mixed by pipetting douncing. The lysate was transferred to the NucleoSpin RNA Column provided and centrifuged for 30 s at 11,000 x g. The lysate was discarded and 350 µL membrane desalting buffer (MDB) provided was added to the membrane of the column which was then centrifuged at 11,000 x g for 1 min. DNase reaction mixture (95 µL) was applied directly onto the centre of the column membrane and incubated at room temperature for 15 min. Three subsequent washing steps were performed using 200 µL Buffer RAW2, 600 µL Buffer RA3, and 250 µL Buffer RA3 respectively, with 30 s centrifugation steps at 11,000 x g in between each wash, followed by a final spin at 11,000 x g for 2 min to dry the membrane completely. The RNA was eluted in 60 µL of RNase-free water and collected by centrifugation at 11,000 x g for 1 min.

6.2.3 Preparation of cDNA template

RNA (500 ng to 1 μg) was used for cDNA synthesis using an iScript cDNA synthesis kit (Bio-Rad). The protocol was performed as recommended by the manufacturer instructions. For each reaction, 4 μL of 5x iScript reaction mix, 1 μL of iScript reverse transcriptase, RNA template and nuclease-free water were combined to make up the final volume of 20 μL. The reaction mixture was incubated in a thermal cycler using the following protocol: 5 min at 25˚C, 20 min at 46˚C, 1 min at 95˚C and a final incubation at 4˚C.

6.2.4 Quantitative real-time PCR (RT-PCR)

Taqman gene expression assays were used to determine the expression levels of the human *AHNAK2* (Hs00292832_M1) and mouse *Ahnak2* (Mm01331631_m1), and a housekeeping gene for the human *HPRT1* (Hs02800695_m1) and mouse *Tbp* (Mm00446971_m1).

The relative PCR efficiencies of the target and the housekeeping assay amplifications were estimated from the slope of the standard curve. A 100 ng sample was prepared for a serial dilution with a factor of five, to generate five amounts of cDNA for the standard curve: 100 ng, 25 ng, 6.25 ng, 1.56 ng, and 0.39 ng. Three replicates of each standard curve point were included for each amplicon. A standard curve slope of -3.3 indicated a PCR with 100% efficiency (Pfaffl, 2001; Rasmussen, 2001). The $2^{-\Delta\Delta Ct}$ method was used to analyse the relative changes in the gene expression (Livak & Schmittgen, 2001). For the comparative C_t ($\Delta\Delta C_t$) calculation to be valid, similar PCR efficiencies of the target gene and housekeeping gene were required. To determine the efficiency, the delta C_t (C_t target – C_t housekeeping) was calculated for each standard curve point and plotted against the log cDNA amount to generate a semi-log regression line. If the slope

of the line was < 0.1, the efficiencies of the two PCR were considered to be equal (Livak & Schmittgen, 2001).

The PCR were set up in a MicroAmp Fast Optical 96-Well reaction plate. Reactions (20 µL) were prepared with TaqMan assay probe (1 µL), master mix (10 µL) (Applied Biosystem), cDNA (2 µL) and nuclease-free H_2O(7 µL). Samples were run in triplicate and a housekeeping gene was included for normalisation of the target gene expression. The reactions were performed on either a StepOnePlus or ViiA 7 real-time PCR system (Applied Biosystems). The real time PCR were cycled with an initial denaturation step at 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min with a final temperature of 25°C.

6.2.5 Immunofluorescent staining for protein expression analysis

Fibroblast cells were fixed on the cover slips with 4% (w/v) paraformaldehyde for 15 min at room temperature. The paraformaldehyde was removed and the cover slips were washed 2 times with 1X PBS. Blocking was performed with 0.1% (v/v) Triton X-100 (CAS 9002-93-1, Santa Cruz Biotechnology) for 5 min followed by 5% (w/v) bovine serum albumin (BSA) for 30 min at room temperature. The blocking media was removed and the cells were incubated with the rabbit anti-human carboxyl terminal AHNAK2 primary antibody (HPA004145, Atlas Antibodies) at a dilution of 1:100 (in 1% (w/v) BSA) at room temperature for 3 hour, followed by washing twice with 1X PBS. Subsequently, Alexa 488 labelled anti-rabbit IgG antibody (A-11070, Molecular Probes) was used as the secondary antibody at a dilution of 1:200 (in 1X PBS) and incubated at room temperature for 1 hour, followed by washing for 5 times with a 2 min gap in between. The cells were stained with Hoechst stain (H3570, Molecular Probes) for 5 min, followed by 2 times washing. The cover slips were mounted onto a glass slide using the DAKO fluorescent mounting medium (S3023, DAKO). Cells were

viewed with a Leica TCS SP5 II confocal microscope and images were captured with LAS AF software.

6.2.6 Secondary RNA structure stability prediction

The stability of the secondary RNA structure was predicted using the mfold web server (Zuker, 2003) which calculates the minimum free folding energy generated from the mRNA sequence. The RNA secondary structure with the minimum free folding energy is likely to be more stable *in vivo*. The free folding energy was calculated for the wild type and mutant sequences for a 50 or 100 nucleotides window upstream and downstream of the variant of interest. For the analysis, the first 50/100 nucleotides upstream of the variant were selected for the first window and each window thereafter was moved by 10 nucleotides towards the downstream of the variant (Park et al., 2013; Supek et al., 2014). In total, 6 sets of data were obtained for the 50 nucleotides window and 11 sets were obtained for the 100 nucleotides window, for both the wild type and mutant mRNA sequences. The mean free folding energy for each window (i.e 50 and 100 nucleotides) was compared between the wild type (ΔG_{WT}) and the mutant (ΔGM_{T}) mRNA sequence.

6.2.7 Harvesting the mouse tissues

All animal work was approved by the Sydney Local Health District Animal Welfare Committee, Research and Development Office, Royal Prince Alfred Hospital, Camperdown, Australia under Protocol No. 2011/028C. Tissue samples were harvested from wild type mice (C57BL/J6) at three different time points: early postnatal period (postnatal day 2-3), post-weaning period (postnatal day 21) and adulthood (postnatal day 93-105). Sample collection included four neural tissues (the brain, spinal cord, dorsal root ganglion (DRG) and sciatic nerve) as well as the gastrocnemius muscle. The

tissues were snap frozen in liquid nitrogen once collected and stored at -80˚C prior to RNA extraction.

6.2.8 RNA extraction from mouse tissues

Tissues collected from the mice were extracted using the Trizol method. Briefly, frozen tissue was transferred into 3 mL pre-cooled Trizol (Ambion), and immediately homogenised with an aggregate homogeniser (Polytron, Kinematica). The homogenised tissue was incubated at 25˚C for 5 min, followed by centrifugation at 3000 x g (Beckman Coulter SX 4750 centrifuge) for 10 min at 4˚C. After centrifugation, the supernatant was poured into a tube containing 600 µL of chloroform and shaken vigorously for approximately 15 s. The mixture was incubated at room temperature for 2-3 min followed by another centrifugation at 3000 x g, 15 min at 4˚C. After the second centrifugation, two layers were observed in the tube. Using a blunt pipette tip, the top layer was transferred into a new 15 mL Falcon tube and 1.5 mL of isopropanol was added and mixed by gentle inversion. The samples were incubated at 25˚C for 10 min, followed by centrifugation at 3000 x g for 10 min at 4˚C. The supernatant was discarded and 3 mL of 75% (v/v) ethanol was added to the pellet. The pellet was vortexed and centrifuged at 3000 x g for 10 min at 4˚C, after that the ethanol was discarded and the pellet was air-dried for approximately 15-30 min. According to the size of the pellet, the pellet was re-suspended with 50 uL to 500 uL of RNase free water and stored at -80˚C until further processing.

6.2.9 Statistical analysis

The Student t-test was used to determine significant differences in the gene expression studies and the predicted mRNA stability analysis between two comparative groups.

6.3 Results

To investigate the potential pathogenic effect of the *AHNAK2* variant (c.A118C, p.T40P) identified in family CMT861, gene expression analysis at the mRNA and protein level was assessed using the primary fibroblast cells of the affected brothers. In addition, as there were limited studies on the *AHNAK2* gene, a spatiotemporal gene expression analysis was performed on several neural mouse tissues to determine the expression profile of *Ahnak2* in the murine nervous system.

6.3.1 Mutant *AHNAK2* shows reduced expression in AR-CMT patient fibroblasts

The expression of *AHNAK2* at the mRNA level was examined in the two affected brothers (IV:7 and IV:9) using primary fibroblasts harbouring the *AHNAK2* variant (c.A118C; p.T40P). As the skin biopsies of these brothers were collected at different time points during the PhD study, the gene expression studies were performed in separate centres and the analyses for each patient were done separately. However, for each assay performed, three controls were included with Control 1 being common to both of the experiments.

The *AHNAK2* expression in the proband's fibroblast was reduced significantly by at least 9-fold (p-value < 0.0001) compared to Control 1 (Figure 6.2a). There was no significant difference in the expression levels of Control 2 and 3 compared to Control 1 (Figure 6.2a). A similar result was observed in the other affected brother (IV:9), with a significant 8-fold decrease (p-value < 0.0001) in the fibroblast's *AHNAK2* expression compared to Control 1 (Figure 6.2b). Also, there was no significant difference in the expression levels of Control 4 and 5 compared to Control 1 (Figure 6.2b). Control 1, 2, 4 and 5 are fibroblasts from Caucasian individuals of different ages while Control 3 are fibroblast from an individual who is ethnically and age-matched to the proband. The

similar expression levels observed in the controls suggests that there are no ethnic or age-specific effects in *AHNAK2* gene expression. The data provides compelling evidence for the significant reduction in the *AHNAK2* mRNA levels in primary fibroblasts of the affected brothers.

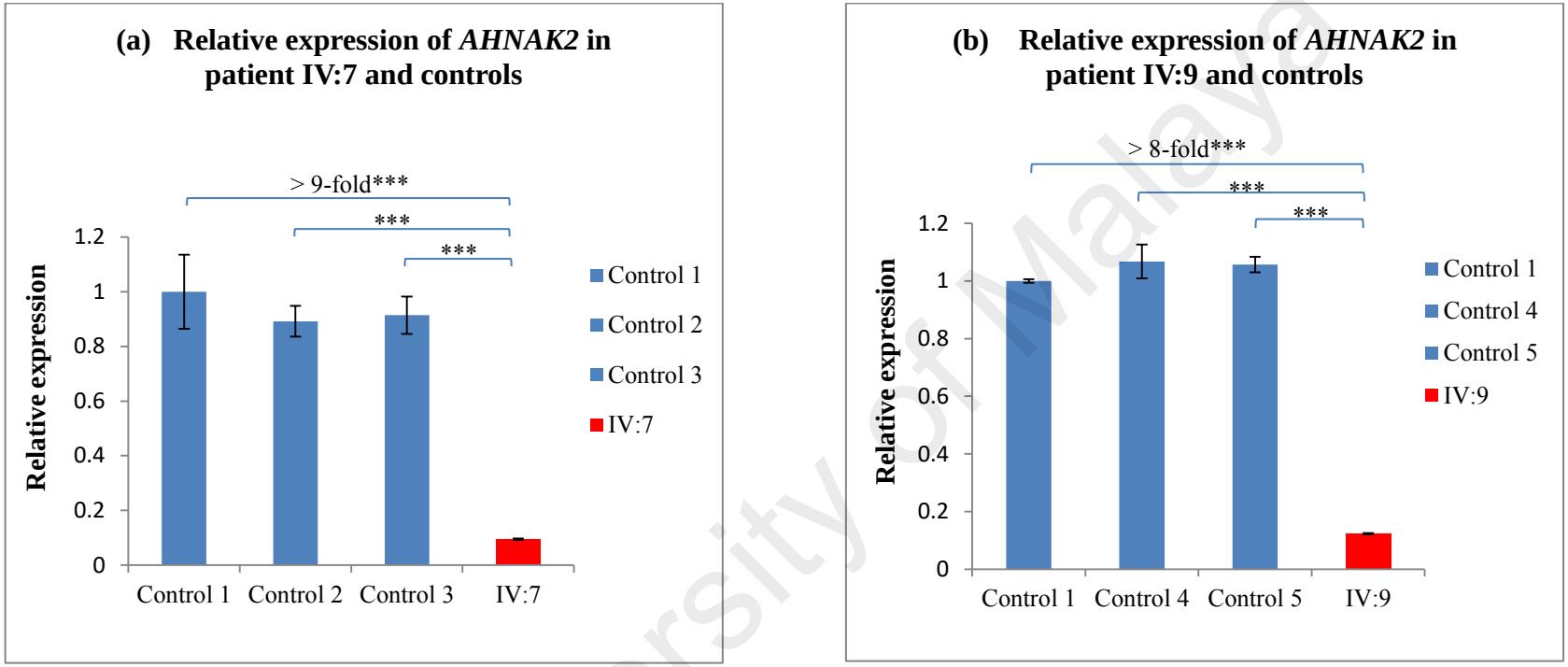


Figure 6.2: Relative expression of *AHNAK2* in the fibroblasts of patient IV:7 and IV:9 compared to the fibroblasts of three neurologically normal controls

Each cell line was grown in triplicates. Bars show the mean mRNA levels (±SEM; error bars) relative to Control 1, which has been set to +1. A Student t-test was performed comparing the patient to each control (*** indicates $p < 0.001$). (a) A 9-fold decrease in *AHNAK2* expression was observed in the fibroblast cells from patient IV:7 in comparison to Control 1, (b) An 8-fold decrease in *AHNAK2* expression was observed in the fibroblast cells from patient IV:9 in comparison to Control 1.

6.3.2 AHNAK2 protein expression is reduced in patient fibroblasts

To further determine the effect of the *AHNAK2* variant on the protein level, immunofluorescent staining of the fibroblasts of one of the patients (IV:9) and controls was performed. The immunofluorescence analysis showed that AHNAK2 protein expression in the patient was dramatically reduced and this result was reproducible (N=4), (Figure 6.3). Although preliminary, this result supports the reduced mRNA levels observed with the TaqMan gene expression analysis seen in the fibroblasts of both brothers.

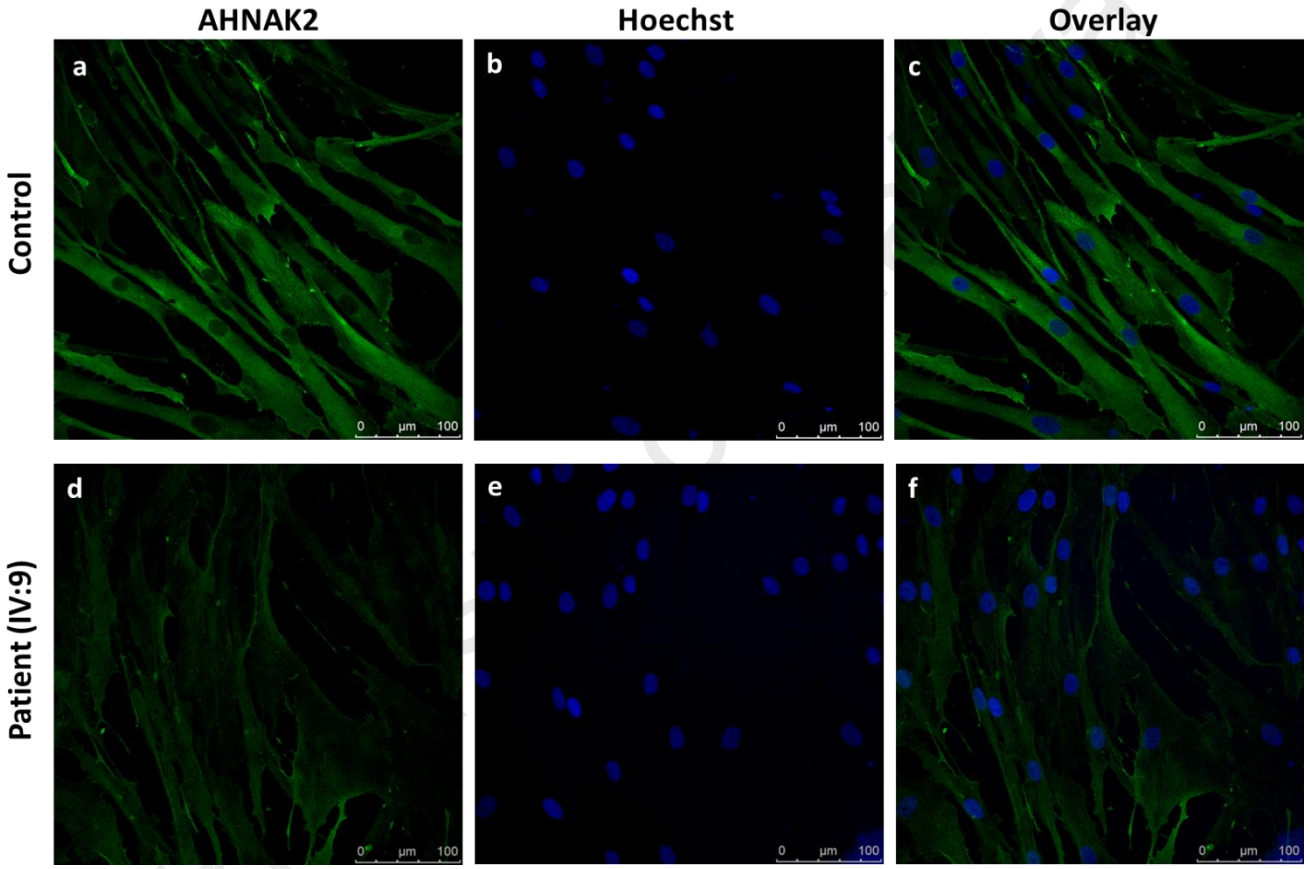


Figure 6.3: Immunofluorescent staining of AHNAK2 in fibroblast cells.

Representative images of control (panel a, b and c) and patient (IV:9) (panel d, e and f) fibroblasts stained for AHNAK2 (green) and Hoechst nuclear stain (blue). Panel c and f show the overlay of AHNAK2 and nucleus staining for the control and the patient respectively. AHNAK2 has clearly reduced staining in the patient fibroblasts compared to the control fibroblasts. The images were taken with a 40X objective lens using a Leica TCS SP5 II confocal microscope. The scale bar represents 100 µm.

6.3.3 The *AHNAK2* c.A118C; p.T40P variant does not affect mRNA stability

To explore the potential mechanism that causes the reduction of *AHNAK2* mRNA expression, a bioinformatic analysis was performed to predict if the c.A118C; p.T40P variant identified in *AHNAK2* affects the stability of the mature mRNA transcript. This analysis was performed using the mfold program, which predicts the stability of an mRNA transcript by inferring the free folding energy of the secondary RNA structure. The free folding energy was calculated and compared between the wild type and mutant transcripts. mRNA transcripts with a lower free folding energy (ΔG) value would be considered more stable than a transcript with a higher free folding energy value.

The mean ΔG for the mutant using a 50 and 100 nucleotides window was -15.3 kcal/mol and -36.8 kcal/mol respectively, which are both more negative than the wild type (-14.8 kcal/mol and -36.5 kcal/mol, for the 50 and 100 nucleotides window respectively), (Figure 6.4). As the ΔG for the mutant and wild type mRNA free folding energy showed no significant differences ($p > 0.05$), the p.T40P variant is not likely to affect the mRNA stability. This finding suggests that the low *AHNAK2* expression levels observed in the fibroblast cells of the affected brothers are unlikely to be due to mRNA decay.

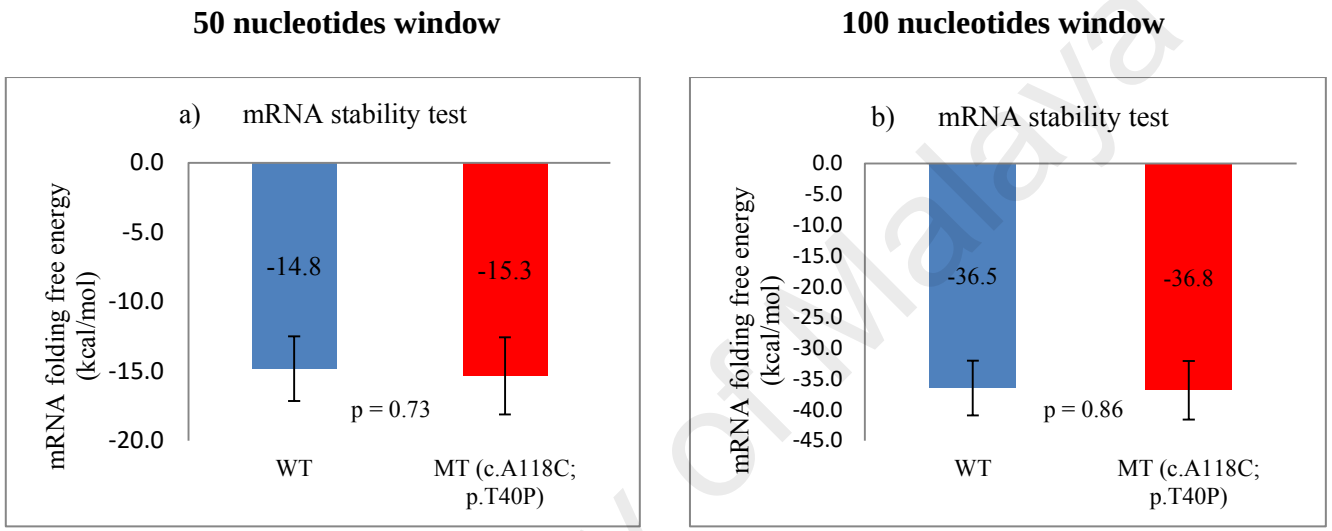


Figure 6.4: RNA structure stability prediction for the *AHNAK2* c.A118C; p.T40P variant.

The graph shows the predicted mean mRNA free folding energy using a 50 (a) and 100 (b) nucleotides window analysis for the c.A118C; p.T40P variant. The mRNA free folding energy for each window was compared between the wildtype (WT) and the mutant (MT) using the Student t-test and no significant difference was identified. The p-value is shown in each graph.

6.3.4 *Ahnak2* is expressed in the selected mouse neuronal tissues

To establish the expression profile of *Ahnak2* in tissues that are affected in CMT, gene expression studies were performed on mouse neuronal and muscle tissues at different stages of development. Both tissues show pathological changes in patients with CMT.

The experiment was repeated three times using three different sets of dissected mouse tissues for each time point, except for the sciatic nerves. For each experimental replicate, the sciatic nerves were collected from multiple mice and analysed as one sample as the RNA yield obtained from the sciatic nerves of a single mouse was low. A summary of the expression profiles for the different tissues is summarised in Figure 6.5.

Comparing the expression of *Ahnak2* across different ages of the mice which correspond to an early postnatal period (P2-P3), post-weaning period (P21) and adulthood (P92-P105), the results show:

1. *Ahnak2* is expressed in all the tissues studied (brain, spinal cord, DRG, sciatic nerve, gastrocnemius muscle) and at all the different time points. However, the expression levels differ in between the tissues and different developmental time points.
2. Brain: Compared to the other tissues studied, the brain has the lowest expression levels and these low expression levels remain constant across the three studied time points.
3. Spinal cord: The expression of *Ahnak2* begins at a low level in P2-P3 mice, after which it significantly increases at P21 ($p < 0.05$), and P92-P105 ($p < 0.01$).
4. Gastrocnemius muscle: The expression of *Ahnak2* is low across all three time points compared to the neural tissues studied, but it is higher than in the brain.

The expression level increases significantly between P2-P3 and P92-P105 ($p < 0.05$).

5. Sciatic nerve: *Ahnak2* is expressed at relatively high levels in this tissue. The level increases gradually from P2-P3 to P21, and then decreases in adulthood (P92-P105) to a level lower than observed at infancy, although there was no significant difference in the expression level observed across the different time points.
6. DRG: *Ahnak2* is abundantly expressed in the DRG. There is a significant increase in *Ahnak2* expression between P2-P3 and P21 ($p < 0.05$). The expression decreases slightly in adulthood (P92-P105) although the level remains higher than during infancy (P2-P3) ($p < 0.05$).
7. The overall expression pattern of *Ahnak2* showed that it is expressed at low levels in the brain and gastrocnemius muscles, and is most highly expressed in the DRG where it is expressed 35-fold higher than in the brain at P2-P3. The spinal cord and the sciatic nerves had comparatively intermediate expression levels (Figure 6.5).

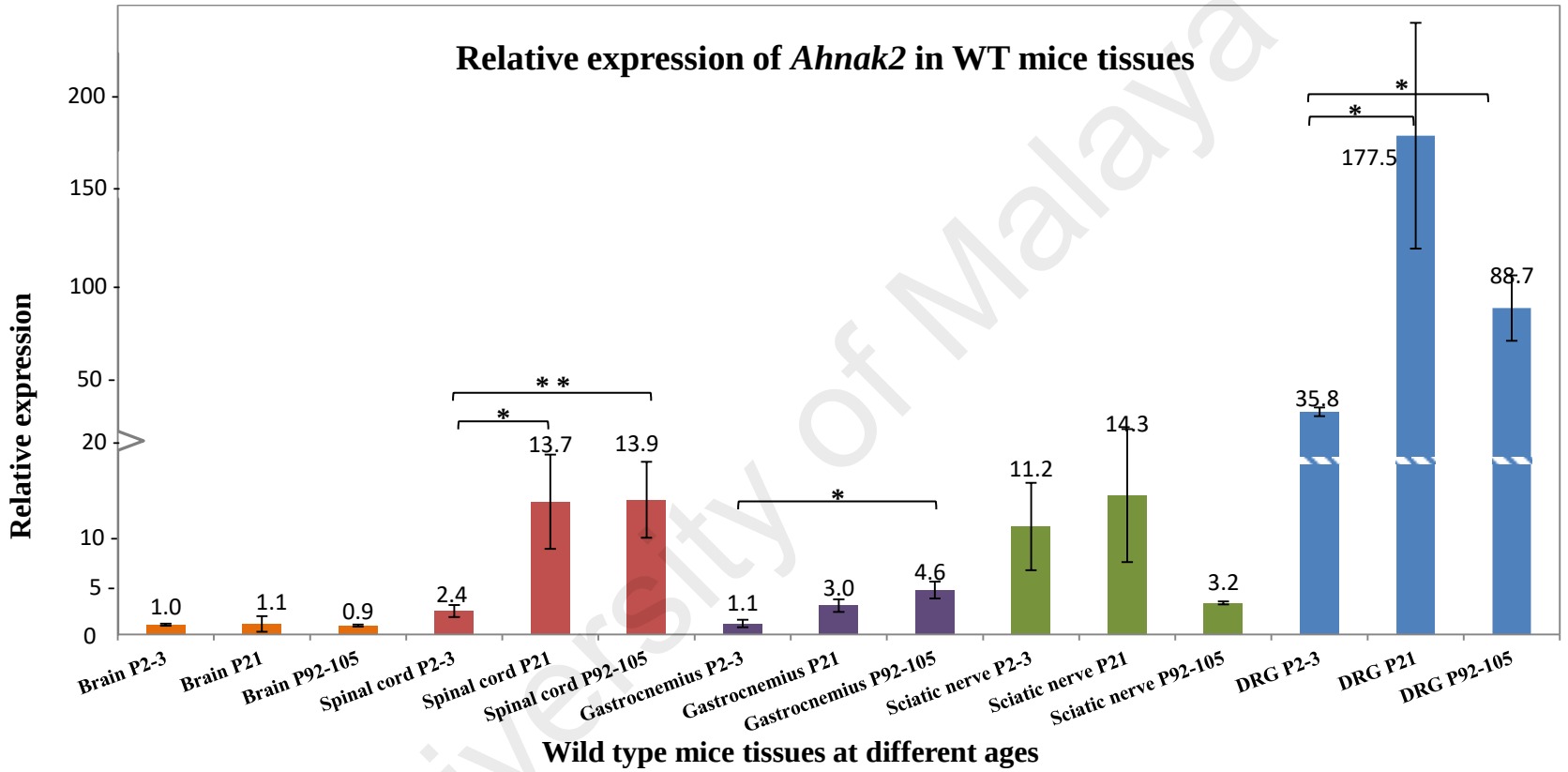


Figure 6.5: Relative expression profile of *Ahnak2* in neural and muscle tissues harvested from wild type mice at different ages.

The Y-axis represents the relative expression of the tissues in comparison to the brain tissue of the P2-3 mice (which has been set to a value of 1). The value above each column represents fold change in relation to the brain tissues of the P2-3 mice. The X-axis represents the wild type mice tissues at different ages. The student t-test was applied to compare the expression level at different time points of the same tissue (* indicates $p<0.05$; ** indicates $p<0.01$). The error bars represent the standard error of the mean (SEM).

6.4 Discussion

This chapter explored the possible pathogenic effect of the c.A118C; p.T40P variant in *AHNAK2* by investigating *AHNAK2* expression at the mRNA and protein level.

The key finding of this chapter was the reduced level of *AHNAK2* expression in the primary fibroblast cells of the affected patients. The *AHNAK2* mRNA expression in the control fibroblasts showed no significant difference in expression levels, suggesting that the low expression level is specific to the affected brothers with the c.A118C; p.T40P mutation. This result was recapitulated in the protein expression analysis whereby a reduced protein expression was also seen. This study provides preliminary evidence for a possible causative impact of the *AHNAK2* mutation on the protein function, suggesting the loss of function of AHNAK2 could lead to peripheral neuropathy.

The reason behind the reduced mRNA levels could be due to the instability of the mutant mRNA molecule and therefore a prediction of mRNA stability was performed. Assessing mRNA stability through calculating the mRNA secondary structure free folding energy has been applied in several studies (Gu et al., 2014; Supek et al., 2014; Trotta, 2016). An mRNA with low stability is likely to have a shorter half-life as it will be rapidly degraded while mRNAs with high stability will have a longer half-life, permitting prolonged translation periods (Ross, 1995). From this analysis, the c.A118C; p.T40P variant did not significantly alter the minimum free folding energy of the mRNA transcript, thus mRNA stability is unlikely to be the factor causing the low *AHNAK2* expression observed in the patients.

To further characterise *AHNAK2*, a spatiotemporal expression profile of *Ahnak2* in different mouse tissues was performed. Neural tissues from patients and healthy humans are not easily obtained other than at postmortem or by invasive nerve biopsies. The use of mouse tissues is an alternative approach to obtain information which could be used to

infer the human *AHNAK2* expression profile. As there is limited information regarding the function of the *AHNAK2* gene, examining the spatiotemporal gene expression may provide clues to its function (Alberts et al., 2002). The gene expression results indicate that this gene is a good candidate for the CMT phenotype in the family as it is expressed in the tissues that are affected in peripheral neuropathy.

The myelin sheath in the central nervous system (CNS) is produced by oligodendrocytes (Bradl & Lassmann, 2010); while in the PNS, myelin sheath is produced by Schwann cells (Bhatheja & Field, 2006). *Ahnak2* had the highest expression level in the sciatic nerve and DRG which are rich in Schwann cells. *Ahnak2* expression in these PNS tissues was generally higher than in the brain and the gastrocnemius muscle. This is interesting as the binding partner of AHNAK2, PRX is crucial for myelin sheath maintenance and is expressed abundantly in Schwann cells (Gillespie et al., 1994; Scherer et al., 1995). As AHNAK2 and PRX proteins can form heterodimers with each other, AHNAK2 and PRX may act together in this same pathway (Han & Kursula, 2014).

Further evidence of a possible role for AHNAK2 in myelination comes from studies showing that AHNAK1 is part of the dystroglycan-dystrophin complex in Schwann cells which promotes myelination (Salim et al., 2009; von Boxberg et al., 2014). Furthermore, AHNAK1 knock-down was shown to change Schwann cell morphology and increase the detachment of Schwann cells from the laminin basement membrane with a reduced level of β-dystroglycan (Salim et al., 2009).

Ahnak2 is expressed most abundantly in DRG neurons in P21 aged mice, which was approximately 170-fold higher than in the other tissues studied. *Ahnak2* expression in sciatic nerves increased markedly during early postnatal and post-weaning period, but at a lower level compared to DRG. The observation of high *Ahnak2* expression levels in

DRG and sciatic nerves in P2-P3 and P21 aged mice may correspond to the intense myelination occurring as neuronal networks are established during infancy to puberty (Masaki et al., 2003; Previtali et al., 2003).

DRGs are located in between the dorsal root and spinal cord, and contain the cell bodies of long afferent sensory nerves that relay sensory information from the periphery to the CNS (Sapunar et al., 2012). The high expression level of *Ahnak2* in the DRG indicates that *Ahnak2* is not only expressed in the myelin sheath surrounding the axons, but also expressed in the cell bodies of the peripheral nerves. Perhaps in the cell body, Ahnak2 might play a role in regulating the secretion of signaling factors for neuronal maintenance, as AHNAK2 is known to regulate FGF1 secretion (Kirov et al., 2015).

The myelin sheath in the CNS is produced by oligodendrocytes (Bradl & Lassmann, 2010). The observation of *Ahnak2* expression in spinal cord indicates that this gene is expressed at other cellular compartments apart from Schwann cells. In comparison to the DRG and sciatic nerves, the expression of *Ahnak2* in the spinal cord was lower during infancy even though the expression increased from P21 to P92-P105. As Ahnak1 and Ahnak2 were reported to co-localise in skeletal muscles (Marg et al., 2010), Ahnak2 might co-localise with Ahnak1 in the spinal cord. During spinal cord injury, *Ahnak1* expression increases significantly (von Boxberg et al., 2006) and is thought to form a protective barrier surrounding the lesion cavity while increasing neoangiogenesis at the injury site (von Boxberg et al., 2006). The high level of *Ahnak2* gene expression observed in the spinal cord of adolescent and adult mice may have a similar role as *Ahnak1*, to act as a protective precaution towards injury. Interestingly, AHNAK2 is necessary for FGF1 secretion under stress conditions, a protein which play a role in tissues repairing and neoangiogenesis (Kirov et al., 2015). In the brain, *Ahnak2* was expressed at very low levels at all ages studied. This observation is similar to the

expression of *Ahnak1* which was found to be restricted to the blood brain barrier endothelial cells of the brain (Gentil *et al.*, 2005).

In the skeletal muscle, both Ahnak1 and Ahnak2 interacts with the dysferlin complex, where it co-localises at the sarcomeres and redistributes to the cytoplasm for membrane repair and muscle regeneration (Huang et al., 2007). In the muscles of dysferlinopathy patients, dysferlin was unable to bind to AHNAK2 and a reduced interaction between dysferlin and AHNAK1 was observed (Huang et al., 2007). Although homozygous *Ahnak1* knockout mice demonstrate no obvious phenotype (Komuro et al., 2004), closer investigation revealed that *Ahnak1*$^{-/-}$ mice had transverse skeletal muscle stiffness and thinner unmyelinated sciatic nerves (Remak fibers) (Marg et al., 2010; von Boxberg et al., 2014). It was also hypothesised that the absence of a distinct phenotype in *Ahnak1*$^{-/-}$ mice could be due to a possible compensatory role of *Ahnak2*, however the extent to which this occurs is unknown (Komuro et al., 2004). In this study, the expression of *Ahnak2* in the gastrocnemius muscles was found to be expressed at lower levels when compared to the nervous tissues: spinal cord, sciatic nerve and DRG. AHNAK2 may play a different role in the skeletal muscle and in the PNS which warrant further investigation.

The spatiotemporal gene expression analysis of *Ahnak2* in mice, especially in the peripheral nervous tissues, has supported *AHNAK2* as a good candidate gene for peripheral neuropathy, although the extent to which the expression profile of this gene in mouse and human is comparable remains unknown. The data in this study show the expression profile of *Ahnak2* in neural and muscle tissues, suggesting the potential impact that *AHNAK2* mutations could have on IPN relevant tissues at critical developmental stages which could have pathogenic consequences if there is a loss of AHNAK2 function.

As several lines of evidence suggest that AHNAK2 plays a role in myelination process and through its interaction with PRX, mutation in *AHNAK2* is likely to be the cause of the demyelinating phenotype seen in CMT861. We hypothesise that, due to the specific binding of PRX to AHNAK2 but not AHNAK1 (Han & Kursula, 2014), the loss of function of AHNAK2 may not be compensated by AHNAK1. Further studies should investigate possible effects of *AHNAK2* knock down in Schwann cells.

In conclusion, by combining the results of gene mapping, WES, and gene and protein expression analysis, there is compelling evidence suggesting that *AHNAK2* is the causal gene for the AR-CMT phenotype in CMT861. Further studies are needed to investigate many other aspects of AHNAK2 in contribution to the disease which will be discussed in Chapter 7.

CHAPTER 7: GENERAL DISCUSSION

The aim of this PhD project was to investigate the genetic aetiology of inherited peripheral neuropathies in our cohort. Samples included in this study were negative for mutations in all known IPN genes and thus, they represented a potential resource for the identification of new genes. Identifying new genes may reveal novel pathways and mechanisms that contribute to axonal degeneration. In the first part of this project, a candidate gene approach was taken to investigate the genes encoding the cytoplasmic dynein-dynactin complex; while in the second part, linkage analysis, homozygosity mapping and variant filtering using WES data was performed to identify the causative gene in an AR-CMT family.

7.1 The first systematic candidate gene approach on the cytoplasmic dynein-dynactin complex

At the beginning of the project, there were already a number of studies performed on animal models with dynein-dynactin gene mutations; however, studies on humans were scarce. Valuable findings from the animal models, notably the *Loa*, *Cra1*, *Swl* and dynamitin overexpression mice, all clearly indicated a key role of the dynein-dynactin complex in neurodegenerative phenotypes. The large size of *DYNC1H1*, with 78 exons, might have hampered early screening methods and therefore, no human disorders had been linked to *DYNC1H1* until NGS was performed on the patients. Application of this method eventually led to the first report of *DYNC1H1* mutations in axonal CMT, which was soon followed by numerous studies linking *DYNC1H1* with various disorders including malformation of cortical development, SMA-LED and hereditary spastic paraplegia (Fiorillo et al., 2014; Harms et al., 2012; Peeters et al., 2015; Poirier et al., 2013; Punetha et al., 2015; Scoto et al., 2015; Strickland et al., 2015; Tsurusaki et al., 2012; Weedon et al., 2011; Willemsen et al., 2012).

While there have been a number of reports on *DYNC1H1* and *DCTN1* in neurodegenerative diseases, only few studies were looking at the other subunits of the dynein-dynactin complex, such as the intermediate chains, light intermediate chains, light chains and other dynactin subunits. To fill this gap of knowledge, we performed a systematic screening of these genes in a large IPN cohort. Through collaborations with Associate Professor Marina Kennerson and Professor Garth Nicholson at the Northcott Neuroscience Laboratory, ANZAC Research Institute, Sydney, we were able to access 136 families that were genetically unclassified and which had multiple affected family members. Even though no pathogenic variants were identified in this study, these genes remain good candidates for the screening of other IPN patients or individuals with other neurodegenerative disorders because of their critical role in neuronal function. In addition, sharing the information about the variants that we identified in these genes will help other researchers to evaluate the pathogenicity of these variants in their studies.

7.2 Family CMT861

The UMMC is a large teaching hospital which is the referral centre for many neurological disorders. As such, the neurologists have been able to gather over 40 probands with CMT from across Malaysia. This allowed our laboratory to gain access to these patient samples to perform the first genetic study on CMT in Malaysia (Shahrizaila et al., 2014). Out of over 40 samples collected, approximately 40% remained genetically unclassified.

Amongst the genetically unsolved cases, we were fortunate that one CMT family, named ‘CMT861’, was willing to participate in our study and gave full cooperation throughout this project by providing us with the family history, DNA samples and skin biopsies for genetic studies. The presence of two affected sons from a consanguineous

marriage suggested autosomal recessive CMT. All of the extended family members of CMT861 show no IPN symptoms or other neurological disorders. The two affected brothers developed the disease at around the same age and share similar phenotypes with the exception of the younger brother who has some cognitive impairment which may have been caused by a suffocation accident when he was a toddler.

7.3 Gene mapping of autosomal recessive CMT in family CMT861

Linkage analysis using a genome-wide SNP genotyping array was initially performed on six nuclear family members of CMT861. The suggestive linkage regions obtained were then further validated and refined using microsatellite markers on the nuclear family and 12 extended family members. Microsatellite marker linkage analysis eliminated several suggestive linkage intervals as the microsatellite markers were more informative compared to the SNP markers. Furthermore, the addition of the paternal grandmother's genotypes in the analysis improved the informativeness of the pedigree by allowing the determination of linkage phase for the segregated genotype.

The maximum multipoint LOD score that could be achieved in this family was 2.06 at zero recombination. The LOD scores did not significantly increase with the addition of genotypes from the unaffected uncles, aunts and cousins. This LOD score could only be further improved by adding genotypes from individuals within the consanguinity loop including the maternal grandfather, the great-grandparents (the founders of this family), or the offspring of the two affected brothers. However, it was not possible to include these samples mentioned above as they are not available.

Even though multiple studies have shown that linkage analysis can also be performed using genotype data obtained via WES, our attempt to define linkage using WES data was not conclusive as 35 peaks with a maximum LOD score of 0.7 was obtained. A similar study reported by Hildebrand and group using WES data of 4

individuals from a nuclear family with a consanguineous background also obtained similar LOD scores as in our study, where 36 regions with maximum LOD of 0.71 were identified (Hildebrand *et al.*, 2015). Several factors might have reduced the power of linkage analysis using WES data. Firstly, WES only captures variants within the exonic regions, resulting in big gaps between markers outside of the exons (Bamshad et al., 2011). This may cause the loss of informative markers and compromise the detection of recombination events. Secondly, this approach cannot detect linkage regions that lie outside the exome (Smith et al., 2011). Furthermore, genotypes obtained from massive parallel sequencing such as WES tend to have higher error rates compared to SNP genotyping arrays, affecting the linkage power (Schmitt *et al.*, 2012). In this study, linkage analysis using WES data was not suitable to define linkage regions in the CMT861 family.

To further support the linkage findings, a combined approach of genome-wide linkage analysis, homozygosity mapping and variant filtering using WES data was applied to explore the disease locus in CMT861 and to identify the potential causal mutation. In this study, the detection of common ROH in both the affected brothers within the linkage interval on chromosome 14 has further supported that the disease allele is in a homozygous state.

A recent study reported that not all the rare homozygous variants would cause loss-of-function of the proteins (Narasimhan *et al.*, 2016). The study screened over 3000 healthy individuals with high parental relatedness and no rare homozygous *AHNAK2* was detected. *AHNAK2* is a promising candidate gene, as AHNAK2 is known to interact with PRX (Han & Kursula, 2014) and mutations in *PRX* can cause CMT4F (a subtype of AR-CMT) or Dejerine-Sottas neuropathy (DSN), (Guilbot et al., 2001; Marchesi et al., 2010; Takashima et al., 2002). To date, over 12 mutations in *PRX* have

been identified (Choi et al., 2015; Marchesi et al., 2010). *PRX* mutations are thought to cause ~5% of autosomal recessive demyelinating CMT (CMT4), and its prevalence in the context of the whole CMT cohort is much lower (Baets *et al.*, 2011). *PRX*-associated neuropathies generally show an early-onset with slow clinical progression, prominent sensory involvement with almost absent sensory nerve action potentials and slow motor nerve conduction velocities (Marchesi et al., 2010). However, the demyelinating phenotype seen in the two affected brothers is less severe compared to patients with *PRX* mutations.

There are two transcripts being reported for *AHNAK2*. The longer transcript composed of 7 exons (exon 1 to exon 6) while the shorter transcript composed of only exon 7. For gene expression analysis, to make sure that we were targeting the longer transcript of *AHNAK2*, assay probe that spans exon 4 and exon 5 was used. In this study, spatial temporal gene expression on mice has shown that *Ahnak2* is expressed in peripheral nerves. Gene expression analysis of *AHNAK2* in primary fibroblast cell cultures showed a marked reduction in mRNA expression in both the affected brothers. Immunofluorescence studies on the fibroblast cell cultures of the affected brother (IV:9) also showed a drastic reduction in AHNAK2 protein expression. During the project, we experienced a power cut in our department which caused the loss of a vast amount of samples. This is the reason why we only managed to perform immunofluorescence analysis on one affected brother but not on the other. To quantify the reduction of AHNAK2 protein, western blot would be a useful method. However, during the time of this project, there was no commercial AHNAK2 antibody for Western blotting available that can bind specifically to the target. This approach is definitely worth to be done in the future when a specific AHNAK2 antibody is available.

Whilst both variants are present in these cells, we speculate that the novel variant in exon 3 (c.A118C; p.T40P) is more likely to exert an impact on the protein than the variant in exon 7 (c.C2743T; p.H915Y), although we could not rule out the latter through segregation analysis. Haplotype analysis on CMT861 showed that these variants segregate in the same haplotype block and they are likely to be in linkage disequilibrium with each other. One way to directly ascribe a pathogenic role for the p.T40P variant is to make gene constructs that contain only one variant and transfect them into cell lines to look for the effect of the variants separately. However, it may be technically challenging to clone the full-length *AHNAK2* cDNA construct as it is ~17 kbp in length with 5795 amino acids whereby exon 7 itself is ~16 kbp in length (5577 amino acids). These constructs may be difficult to transfect into cells. We were also unable to perform any protein structure predictions as presently, the only available crystal structure for AHNAK2 is limited to its PDZ domain, and the p.T40P variant lies outside this domain. Thus, the effect of the p.T40P variant on the protein structure and its impact on PRX binding could not be predicted.

The p.T40P variant does not appear to have any significant effect on the mRNA stability, therefore, the reduction of *AHNAK2* mRNA expression is unlikely caused by the degradation of mutant mRNA. Several other mechanisms have been proposed to regulate gene expression at the pre-translational level including histone acetylation, DNA methylation, and splicing machinery modulation (Baralle & Baralle, 2005; Bell et al., 2011). It is estimated that approximately 15% of point mutations which result in human diseases disrupt splicing (Krawczak et al., 1992). As the p.T40P variant is the fourth nucleotide in exon 3, and boundaries of exons are the hotspots for splicing regulatory elements (Fairbrother et al., 2004; Woolfe et al., 2010), we postulated that the p.T40P variant may cause splicing errors. The splicing consensus detecting program, Human Splicing Finder (HSF) predicted that the p.T40P variant alters the

consensuses exon splicing enhancer (ESE) binding motifs (data not shown). However, the full length AHNAK2 transcript was still detectable from the patients fibroblasts although at a very low level (data not shown). Thus it is not immediately indicative of splicing errors. Further experiments are needed to investigate the actual impact of the p.T40P on *AHNAK2* expression.

7.4 Role of AHNAK2 in myelination

There is limited information about the functional role of AHNAK2, however a recent publication on AHNAK2-PRX heterodimer formation has shed some light on the potential functional role of AHNAK2 in myelination (Han & Kursula, 2014). PRX and the dystrophin-related protein 2 (Drp2) form the Prx-Drp2-dystroglycan complex which is abundantly found at the Cajal bands that flanks the abaxonal surface of Schwann cell myelin sheath and the plasma membrane (Sherman et al., 2001; Sherman et al., 2012). In mice, defects in *Prx* or *Drp2* were shown to cause focal hypermyelination and concomitant demyelination in Schwann cells (Sherman et al., 2012). Recently, mutations in *DRP2* were reported to cause demyelinating CMTX (Brennan et al., 2015a). Given that AHNAK2 can bind to PRX, AHNAK2 might be part of the PRX-DRP2-dystroglycan multi-protein complex in Schwann cells which has not been discovered before. As mutations in both *PRX* and *DRP2* cause CMT, mutations in *AHNAK2* might be deleterious as well (Figure 7.1).

AHNAK1 is a known scaffolding protein that is involved in various biological pathways (Davis et al., 2014). In the PNS, AHNAK1 is strongly expressed at the external cytoplasm of myelinating Schwann cells and co-localises with the major laminin receptor, β-dystroglycan. This complex is involved in cell-cell adhesion, influences the cytoskeleton structure and motility, and supports the myelinating process in Schwann cells (Salim et al., 2009; von Boxberg et al., 2014). Based on the similarity

of the domain structure, AHNAK2 might be another scaffolding protein that plays a similar role as AHNAK1. It was interesting to note that both the AHNAK2-interacting proteins, AHNAK1 and PRX form protein complex with dystroglycan at the abaxonal membrane in Schwann cells and are involved in myelination. Therefore, the potential pathway of AHNAK2 in Schwann cell myelination via its interaction with PRX and AHNAK1 warrants further investigation (Figure 7.1).

In addition, AHNAK2 might act as the adaptor for FGF1, localizing it to the plasma membrane for eventual secretion (Kirov et al., 2015). Several lines of evidence demonstrated that FGF1 plays a neuroprotective role in both the PNS and CNS. FGF1 is expressed in neurons, particularly in motor and sensory neurons. FGF1 prevents cell death in motor neurons after axonal injury (Cuevas et al., 1995; Jacques et al., 1999). Loss of FGF1 accelerates motor neuron degeneration in patients with ALS (Kage et al., 2001). FGF1 is also known to promote proliferation of glial precursor cells (Engele & Bohn, 1992). A study on multiple sclerosis using the white matter lesion tissue of the patients revealed that FGF1 enhances myelination and re-myelination in the dissociated cultures (Mohan et al., 2014). In addition, neurotrophic effects of FGF1 are involved in a variety of signaling pathways in the neurons and play an important role against neurotoxins such as HIV-derived proteins and amyloid-beta protein in Alzheimer's disease (Hashimoto et al., 2002). Therefore, whether loss of function of AHNAK2 might compromise the neuroprotective activity of FGF1 in neurons and thus lead to axonal degeneration remains an interesting question to be further investigated (Figure 7.1).

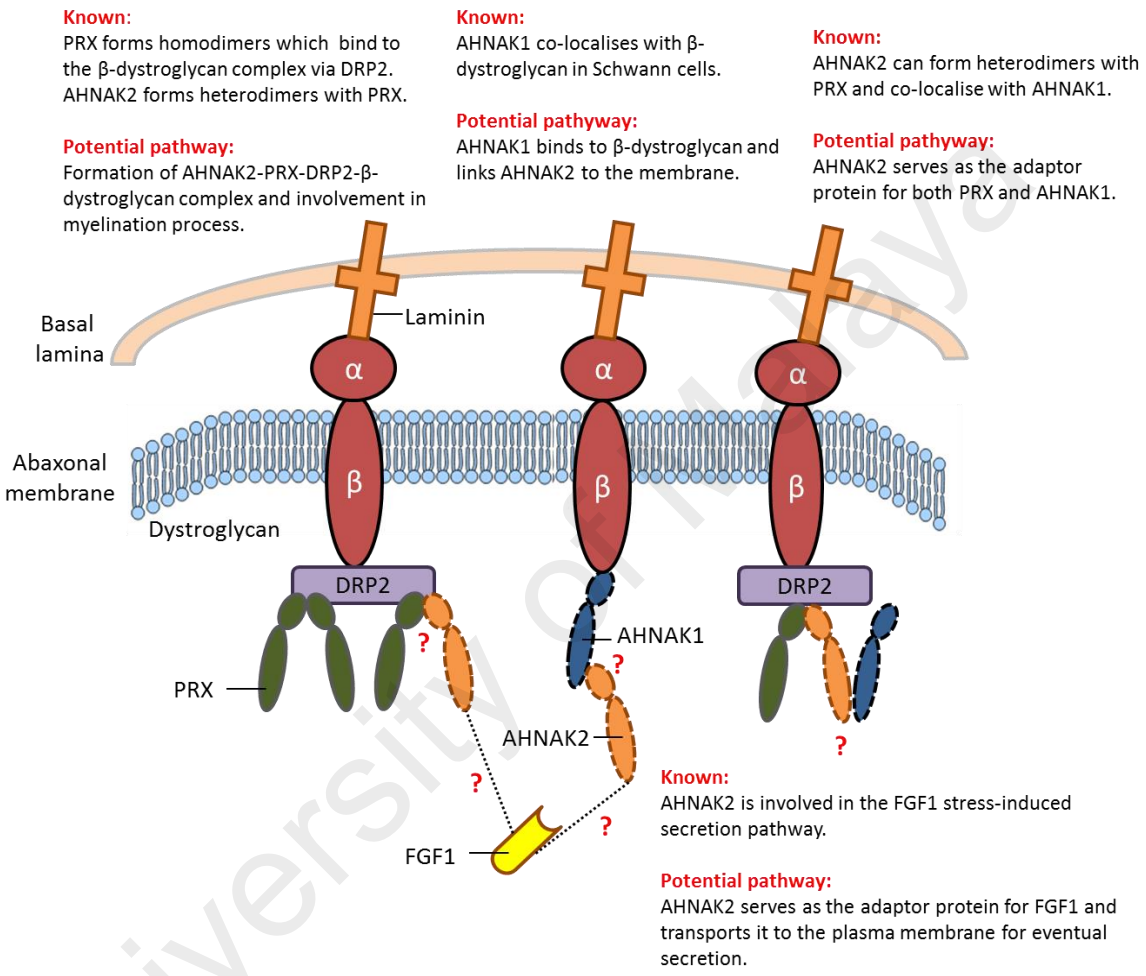


Figure 7.1: Potential functional pathways of AHNAK2 in the Schwann cells.

7.5 Future directions

7.5.1 Investigating the molecular mechanism of the p.T40P variant in causing *AHNAK2* expression downregulation

This study will continue to determine the underlying molecular mechanism of the p.T40P variant in affecting *AHNAK2* expression. As the splicing prediction program has predicted that the p.T40P variant may disrupt the binding motif of the ESE, a minigene splicing assay is a reliable experiment to investigate if this variant is causing a splicing defect in the *AHNAK2* pre-mRNA transcript (Pagani & Baralle, 2004). The mRNA transcript(s) produced from the minigene can be sequenced to determine the features of alternative (mutant) transcripts.

Missense mutations that abrogate an ESE can lead to exon skipping, which causes a large internal deletion. The premature termination codons that arise from the mutation induced exon skipping usually results in frameshift, which in turn triggers nonsense-mediated mRNA decay (NMD), (Cartegni et al., 2002; Colombo et al., 2013; Lejeune & Maquat, 2005; Pagani & Baralle, 2004). While mutations affecting splice sites have been reported in CMT, such as *GDAP1* (Kabzinska *et al.*, 2005), *MFN2* (Kotruchow et al., 2013), *LRSAM1* (Engeholm et al., 2014) and *MPZ* (Sabet et al., 2006), no reports have been published for splicing defects due to mutations in ESE motifs. Mutations in ESE motifs have been implicated in several other human diseases, such as spinal muscular atrophy (SMA), neurofibromatosis, breast cancer predisposition, colorectal cancer, Marfan syndrome and Becker's muscular dystrophy (Bottillo et al., 2007; Caputi et al., 2002; Cartegni et al., 2006; Colombo et al., 2013; Gorlov et al., 2003; Shiga et al., 1997; Stella et al., 2001).

In case that the p.T40P variant causes splicing defects, the aberrant mRNA transcribed produced would have been degraded via NMD *in vivo*, reducing mRNA and

protein expression in the affected brothers. Some leaky expression of the native mRNA transcript can still be detected, as was seen in the other studies of mutations in ESE binding sites (Bottillo et al., 2007; Cartegni et al., 2006).

To investigate the pathogenicity of p.T40P on a cellular level, mutations can be introduced into a healthy mammalian cell line to study the downstream effects using the CRISPR/Cas9 system. Using this system, we could also correct the mutation in the patient's fibroblast cells by introducing the wild type nucleotide to see if the functions of AHNAK2 can be restored.

Another investigation could be to study the impact of p.T40P variant on AHNAK2-PRX interaction and their localisation in Schwann cell. Given that PRX is produced primarily in Schwann cells, fibroblast cells might not be ideal for the study of AHNAK2-PRX interactions. The AHNAK2 construct containing p.T40P variant could be transfected into Schwann cell lines and proteins immunoprecipated from the cells to investigate if the mutant AHNAK2 can still bind to PRX. Using immunofluorescence, the expression and localisation of the wildtype AHNAK2 in Schwann cells can be explored. The impact of the p.T40P variant on AHNAK2 expression, localisation and interaction with PRX can also be determined with the transfected Schwann cell lines.

7.5.2 Transgenic *AHNAK2* animal models

In order to further characterise the cellular function of *AHNAK2*, transgenic knock-out animal models would be an ideal future direction to be taken. Transgenic mice allow close monitoring of the effects the mutant gene may have on the mouse from behavioural and phenotype development to the cellular level. Observing if *Ahnak2*-null mice develop a phenotype of peripheral neuropathy and if the disease development is similar to the patient with the *AHNAK2* mutation would be an important question to pursue. As sural nerve biopsy from patient is not easily accessible, the best alternative is

to obtain the nerve tissue from the transgenic mice, to examine the impact of *Ahnak2* knock-out on the nerve anatomy, such as myelin thickness as well as diameter and intermodal length of the axon. Furthermore, immunofluorescence study on the nerve biopsy can be performed, to examine the interaction of *Ahnak2* with the other myelin proteins in the nerve.

7.6 Final conclusion

The functional diversity of genes implicated in IPN illustrates the complex biological pathways involved in maintaining peripheral nerves physiology. The discoveries of new genes and pathways involved in peripheral neuropathies are essential to uncover the underlying mechanisms contributing to length-dependent axonal degeneration. In this study, a candidate gene, *AHNAK2*, was identified to cause the AR-CMT in CMT861. Although this gene has not been reported in causing human disorders and its functional implication is not well-established, we demonstrated that it is expressed in IPN-relevant mouse tissues, indicating its potential role in neuronal function. Literature studies on this gene suggested its role in myelination. We postulate that perturbation of AHNAK2 may disrupt the myelination process in peripheral nerves, thereby resulting in the demyelinating phenotype of the two affected brothers from CMT861. The potential molecular mechanisms of the p.T40P mutation on the functional mechanisms of AHNAK2 suggested by our findings warrants further investigation. The insights concerning the function of AHNAK2 in neurons maintenance might contribute to delineate the mechanism of length-dependent axonal degeneration. Ultimately, this information might help to develop therapeutic targets to ameliorate the pathogenic process of peripheral neuropathies.

References

1000 Genome Project Consortium. (2010). A map of human genome variation from population-scale sequencing. [10.1038/nature09534]. *Nature, 467*(7319), 1061-1073.

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet, 30*(1), 97-101.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat Methods, 7*(4), 248-249.

Alberts, B., Johnson, J., & Lewis, J. (2002). Molecular Biology of the Cell, Studying Gene Expression and Function Available from http://www.ncbi.nlm.nih.gov/books/NBK26818/

Albulym, O. M., Kennerson, M. L., Harms, M. B., Drew, A. P., Siddell, A. H., Auer-Grumbach, M., et al. (2016). MORC2 mutations cause axonal Charcot-Marie-Tooth disease with pyramidal signs. *Ann Neurol, 79*(3), 419-427.

Alkuraya, F. S. (2010). Homozygosity mapping: one more tool in the clinical geneticist's toolbox. *Genet Med, 12*(4), 236-239.

Allan, V. J. (2011). Cytoplasmic dynein. *Biochem Soc Trans, 39*(5), 1169-1178.

Antonellis, A., Ellsworth, R. E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S. Q., et al. (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am J Hum Genet, 72*(5), 1293-1299.

Araki, E., Tsuboi, Y., Daechsel, J., Milnerwood, A., Vilarino-Guell, C., Fujii, N., et al. (2014). A Novel DCTN1 mutation with late-onset parkinsonism and frontotemporal atrophy. *Mov Disord, 29: 1201-1204*.

Auer-Grumbach, M., De Jonghe, P., Verhoeven, K., & et al. (2003). Autosomal dominant inherited neuropathies with prominent sensory loss and mutilations: A review. *Archives of Neurology, 60*(3), 329-334.

Azzedine, H., Senderek, J., Rivolta, C., & Chrast, R. (2012). Molecular Genetics of Charcot-Marie-Tooth Disease: From Genes to Genomes. *Molecular Syndromology, 3*(5), 204-214.

Baets, J., Deconinck, T., De Vriendt, E., Zimoń, M., Yperzeele, L., Van Hoorenbeeck, K., et al. (2011). Genetic spectrum of hereditary neuropathies with onset in the first year of life. *Brain, 134*(9), 2664-2676.

Bahlo, M., & Bromhead, C. J. (2009). Generating linkage mapping files from Affymetrix SNP chip data. *Bioinformatics, 25*(15), 1961-1962.

Baloh, R. H., Schmidt, R. E., Pestronk, A., & Milbrandt, J. (2007). Altered Axonal Mitochondrial Transport in the Pathogenesis of Charcot-Marie-Tooth Disease from Mitofusin 2 Mutations. *J. Neurosci., 27*(2), 422-430.

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., et al. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. [10.1038/nrg3031]. *Nat Rev Genet, 12*(11), 745-755.

Banks, G. T., Haas, M. A., Line, S., Shepherd, H. L., AlQatari, M., Stewart, S., et al. (2011). Behavioral and Other Phenotypes in a Cytoplasmic Dynein Light Intermediate Chain 1 Mutant Mouse. *The Journal of Neuroscience, 31*(14), 5483-5494.

Baralle, D., & Baralle, M. (2005). Splicing in action: assessing disease causing sequence changes. *J Med Genet, 42*(10), 737-748.

Barisic, N., Claeys, K. G., Sirotkovic-Skerlev, M., Lofgren, A., Nelis, E., De Jonghe, P., et al. (2008). Charcot-Marie-Tooth disease: a clinico-genetic confrontation. *Ann Hum Genet, 72*(Pt 3), 416-441.

Baxter, R. V., Ben Othmane, K., Rochelle, J. M., Stajich, J. E., Hulette, C., Dew-Knight, S., et al. (2002). Ganglioside-induced differentiation-associated protein-1 is mutant in Charcot-Marie-Tooth disease type 4A/8q21. *Nat Genet, 30*(1), 21-22.

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., et al. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology, 12*(1), R10-R10.

Berrueta, L., Tirnauer, J. S., Schuyler, S. C., Pellman, D., & Bierer, B. E. (1999). The APC-associated protein EB1 associates with components of the dynactin complex and cytoplasmic dynein intermediate chain. *Curr Biol, 9*(8), 425-428.

Bhatheja, K., & Field, J. (2006). Schwann cells: origins and role in axonal maintenance and regeneration. *Int J Biochem Cell Biol, 38*(12), 1995-1999.

Bielli, A., Thörnqvist, P.-O., Hendrick, A. G., Finn, R., Fitzgerald, K., & McCaffrey, M. W. (2001). The Small GTPase Rab4A Interacts with the Central Region of Cytoplasmic Dynein Light Intermediate Chain-1. *Biochemical and Biophysical Research Communications, 281*(5), 1141-1153.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol, Chapter 19*, Unit 19 10 11-21.

Boerkoel, C. F., Takashima, H., Stankiewicz, P., Garcia, C. A., Leber, S. M., Rhee-Morris, L., et al. (2001). Periaxin Mutations Cause Recessive Dejerine-Sottas Neuropathy. *American journal of human genetics, 68*(2), 325-333.

Bondurand, N., Girard, M., Pingault, V., Lemort, N., Dubourg, O., & Goossens, M. (2001). Human Connexin 32, a gap junction protein altered in the X-linked form of Charcot-Marie-Tooth disease, is directly regulated by the transcription factor SOX10. *Hum Mol Genet, 10*(24), 2783-2795.

Bottillo, I., De Luca, A., Schirinzi, A., Guida, V., Torrente, I., Calvieri, S., et al. (2007). Functional analysis of splicing mutations in exon 7 of NF1 gene. *BMC Med Genet, 8*, 4.

Braathen, G. J., Hoyer, H., Busk, O. L., Tveten, K., Skjelbred, C. F., & Russell, M. B. (2015). Variants in the genes DCTN2, DNAH10, LRIG3, and MYO1A are associated with intermediate Charcot-Marie-Tooth disease in a Norwegian family. *Acta Neurol Scand, 134(1), 67–75.* .

Bradl, M., & Lassmann, H. (2010). Oligodendrocytes: biology and pathology. *Acta Neuropathologica, 119*(1), 37-53.

Brennan, K. M., Bai, Y., Pisciotta, C., Wang, S., Feely, S. M., Hoegger, M., et al. (2015a). Absence of Dystrophin Related Protein-2 disrupts Cajal bands in a patient with Charcot-Marie-Tooth disease. *Neuromuscul Disord, 25*(10), 786-793.

Brennan, K. M., Bai, Y., & Shy, M. E. (2015b). Demyelinating CMT-what's known, what's new and what's in store? *Neurosci Lett, 596*, 14-26.

Brewer, M. H., Chaudhry, R., Qi, J., Kidambi, A., Drew, A. P., Menezes, M. P., et al. (2016). Whole Genome Sequencing Identifies a 78 kb Insertion from Chromosome 8 as the Cause of Charcot-Marie-Tooth Neuropathy CMTX3. *PLoS Genet, 12*(7), e1006177.

Bromberg, Y., & Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research, 35*(11), 3823-3835.

Brunham, L. R., & Hayden, M. R. (2013). Hunting human disease genes: lessons from the past, challenges for the future. *Hum Genet, 132*(6), 603-617.

Bull, P. C., Thomas, G. R., Rommens, J. M., Forbes, J. R., & Cox, D. W. (1993). The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nat Genet, 5*(4), 327-337.

Cantalupo, G., Alifano, P., Roberti, V., Bruni, C. B., & Bucci, C. (2001). Rab-interacting lysosomal protein (RILP): the Rab7 effector required for transport to lysosomes. *EMBO J, 20*(4), 683-693.

Caputi, M., Kendzior, R. J., & Beemon, K. L. (2002). A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. *Genes & Development, 16*(14), 1754-1759.

Caroppo, P., Le Ber, I., Clot, F., Rivaud-Pechoux, S., Camuzat, A., De Septenville, A., et al. (2014). DCTN1 mutation analysis in families with progressive supranuclear palsy-like phenotypes. *JAMA Neurol, 71*(2), 208-215.

Cartegni, L., Chew, S. L., & Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet, 3*(4), 285-298.

Cartegni, L., Hastings, M. L., Calarco, J. A., Stanchina, E. d., & Krainer, A. R. (2006). Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes, SMN1 and SMN2. *American journal of human genetics, 78*(1), 63-77.

Chance, P. F., Alderson, M. K., Leppig, K. A., Lensch, M. W., Matsunami, N., Smith, B., et al. (1993). DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell, 72*(1), 143-151.

Charcot, J.-M., & Marie, P. (1886). Sur une forme particulière d'atrophie musculaire progressive, souvent familiale, debutant par les pieds et les jambes et atteignant plus tard les mains (in French). *Rev. Med., 6*, 97–138.

Chen, T., Wu, D., Moskaluk, C. A., & Fu, Z. (2013). Distinct expression patterns of ICK/MAK/MOK protein kinases in the intestine implicate functional diversity. *PLoS One, 8*(11), e79359.

Chen, X.-J., Levedakou, E. N., Millen, K. J., Wollmann, R. L., Soliven, B., & Popko, B. (2007). Proprioceptive Sensory Neuropathy in Mice with a Mutation in the Cytoplasmic Dynein Heavy Chain 1 Gene. *J. Neurosci., 27*(52), 14515-14524.

Chen, Y. Z., Bennett, C. L., Huynh, H. M., Blair, I. P., Puls, I., Irobi, J., et al. (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet, 74*(6), 1128-1135.

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences, 106*(45), 19096-19101.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One, 7*(10), e46688.

Choi, Y. J., Hyun, Y. S., Nam, S. H., Koo, H., Hong, Y. B., Chung, K. W., et al. (2015). Novel Compound Heterozygous Nonsense PRX Mutations in a Korean Dejerine-Sottas Neuropathy Family. *Journal of Clinical Neurology (Seoul, Korea), 11*(1), 92-96.

Chow, C. Y., Zhang, Y., Dowling, J. J., Jin, N., Adamska, M., Shiga, K., et al. (2007). Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. [10.1038/nature05876]. *Nature, 448*(7149), 68-72.

Chuang, J. Z., Milner, T. A., & Sung, C. H. (2001). Subunit heterogeneity of cytoplasmic dynein: Differential expression of 14 kDa dynein light chains in rat hippocampus. *J Neurosci, 21*(15), 5501-5512.

Cioroiu, C., & Brannagan, T., III. (2014). Peripheral Neuropathy. *Current Geriatrics Reports, 3*(2), 83-90.

Cogli, L., Progida, C., Thomas, C. L., Spencer-Dene, B., Donno, C., Schiavo, G., et al. (2013). Charcot-Marie-Tooth type 2B disease-causing RAB7A mutant proteins show altered interaction with the neuronal intermediate filament peripherin. *Acta Neuropathol, 125*(2), 257-272.

Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C. B., Ficarazzi, F., et al. (2013). Comparative In Vitro and In Silico Analyses of Variants in Splicing Regions of BRCA1 and BRCA2 Genes and Characterization of Novel Pathogenic Mutations. *PLoS One, 8*(2), e57173.

Cooper, G. M., Stone, E. A., Asimenos, G., Branch, N. C. S. P. G. T., Nih Intramural Sequencing Center, N. H. G. R. I. N. I. o. H. B. M. U. S. A., Green, E. D., et al. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research, 15*(7), 901-913.

Cuesta, A., Pedrola, L., Sevilla, T., Garcia-Planells, J., Chumillas, M. J., Mayordomo, F., et al. (2002). The gene encoding ganglioside-induced differentiation-associated protein 1 is mutated in axonal Charcot-Marie-Tooth type 4A disease. *Nat Genet, 30*(1), 22-25.

Cuevas, P., Carceller, F., & Gimenez-Gallego, G. (1995). Acidic fibroblast growth factor prevents post-axotomy neuronal death of the newborn rat facial nerve. *Neurosci Lett, 197*(3), 183-186.

Cullinane, A. R., Vilboux, T., O'Brien, K., Curry, J. A., Maynard, D. M., Carlson-Donohoe, H., et al. (2011). Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia. *J Invest Dermatol, 131*(10), 2017-2025.

Culver-Hanlon, T. L., Lex, S. A., Stephens, A. D., Quintyne, N. J., & King, S. J. (2006). A microtubule-binding domain in dynactin increases dynein processivity by skating along microtubules. *Nat Cell Biol, 8*(3), 264-270.

Dalal, A., Bhowmik, A. D., Agarwal, D., & Phadke, S. R. (2015). Exome sequencing & homozygosity mapping for identification of genetic aetiology for spastic ataxia in a consanguineous family. *The Indian Journal of Medical Research, 142*(2), 220-224.

Davidson, G., Murphy, S., Polke, J., Laura, M., Salih, M., Muntoni, F., et al. (2012). Frequency of mutations in the genes associated with hereditary sensory and autonomic neuropathy in a UK cohort. *J Neurol, 259*(8), 1673-1685.

Davis, T. A., Loos, B., & Engelbrecht, A. M. (2014). AHNAK: the giant jack of all trades. *Cell Signal, 26*(12), 2683-2693.

de Brito, O. M., & Scorrano, L. (2008). Mitofusin 2 tethers endoplasmic reticulum to mitochondria. [10.1038/nature07534]. *Nature, 456*(7222), 605-610.

De Sandre-Giovannoli, A., Chaouch, M., Kozlov, S., Vallat, J. M., Tazir, M., Kassouri, N., et al. (2002). Homozygous defects in LMNA, encoding lamin A/C nuclear-envelope proteins, cause autosomal recessive axonal neuropathy in human

(Charcot-Marie-Tooth disorder type 2) and mouse. *Am. J. Hum.Genet., 70*(726–736).

Deinhardt, K., Salinas, S., Verastegui, C., Watson, R., Worth, D., Hanrahan, S., et al. (2006). Rab5 and Rab7 control endocytic sorting along the axonal retrograde transport pathway. *Neuron, 52*(2), 293-305.

Delague, V., Bareil, C., Tuffery, S., Bouvagnet, P., Chouery, E., Koussa, S., et al. (2000). Mapping of a new locus for autosomal recessive demyelinating Charcot-Marie-Tooth disease to 19q13.1-13.3 in a large consanguineous Lebanese family: exclusion of MAG as a candidate gene. *Am J Hum Genet, 67*(1), 236-243.

Dierick, I., Baets, J., Irobi, J., Jacobs, A., De Vriendt, E., Deconinck, T., et al. (2008). Relative contribution of mutations in genes for autosomal dominant distal hereditary motor neuropathies: a genotype-phenotype correlation study. *Brain, 131*(Pt 5), 1217-1227.

Dobyns, W. B., Reiner, O., Carrozzo, R., & Ledbetter, D. H. (1993). Lissencephaly. A human brain malformation associated with deletion of the LIS1 gene located at chromosome 17p13. *JAMA, 270*(23), 2838-2842.

Drew, A. P., Zhu, D., Kidambi, A., Ly, C., Tey, S., Brewer, M. H., et al. (2015). Improved inherited peripheral neuropathy genetic diagnosis by whole-exome sequencing. *Molecular Genetics & Genomic Medicine, 3*(2), 143-154.

Dubourg, O., Azzedine, H., Yaou, R. B., Pouget, J., Barois, A., Meininger, V., et al. (2006). The G526R glycyl-tRNA synthetase gene mutation in distal hereditary motor neuropathy type V. *Neurology, 66*(11), 1721-1726.

Duchen, L. W. (1974). A dominant hereditary sensory disorder in the mouse with deficiency of muscle spindles: the mutant Sprawling. *J Physiol, 237*(2), 10P-11P.

Dupuis, L., Fergani, A., Braunstein, K. E., Eschbach, J., Holl, N., Rene, F., et al. (2009). Mice with a mutation in the dynein heavy chain 1 gene display sensory neuropathy but lack motor neuron disease. *Experimental Neurology, 215*(1), 146-152.

Eckley, D. M., Gill, S. R., Melkonian, K. A., Bingham, J. B., Goodson, H. V., Heuser, J. E., et al. (1999). Analysis of dynactin subcomplexes reveals a novel actin-related protein associated with the arp1 minifilament pointed end. *J Cell Biol, 147*(2), 307-320.

Egan, M. J., McClintock, M. A., Hollyer, I. H., Elliott, H. L., & Reck-Peterson, S. L. (2015). Cytoplasmic dynein is required for the spatial organization of protein aggregates in filamentous fungi. *Cell Rep, 11*(2), 201-209.

Einarsdottir, E., Carlsson, A., Minde, J., Toolanen, G., Svensson, O., Solders, G., et al. (2004). A mutation in the nerve growth factor beta gene (NGFB) causes loss of pain perception. *Hum Mol Genet, 13*(8), 799-805.

Emond, M. J., Louie, T., Emerson, J., Zhao, W., Mathias, R. A., Knowles, M. R., et al. (2012). Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat Genet, 44*(8), 886-889.

Engeholm, M., Sekler, J., Schöndorf, D. C., Arora, V., Schittenhelm, J., Biskup, S., et al. (2014). A novel mutation in LRSAM1 causes axonal Charcot-Marie-Tooth disease with dominant inheritance. *BMC Neurology, 14*, 118-118.

Engele, J., & Bohn, M. C. (1992). Effects of acidic and basic fibroblast growth factors (aFGF, bFGF) on glial precursor cell proliferation: age dependency and brain region specificity. *Dev Biol, 152*(2), 363-372.

Eschbach, J., & Dupuis, L. (2011). Cytoplasmic dynein in neurodegeneration. *Pharmacology & Therapeutics, 130*(3), 348-363.

Eura, Y., Ishihara, N., Yokota, S., & Mihara, K. (2003). Two mitofusin proteins, mammalian homologues of FZO, with distinct functions are both required for mitochondrial fusion. *J Biochem, 134*(3), 333-344.

Evgrafov, O. V., Mersiyanova, I., Irobi, J., Van Den Bosch, L., Dierick, I., Leung, C. L., et al. (2004). Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat Genet, 36*(6), 602-606.

Exome Variant Server. NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. *(URL: http://evs.gs.washington.edu/EVS/), [2015 accessed]*.

Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., et al. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med, 14*(7), 723-730.

Fairbrother, W. G., Yeo, G. W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P. A., et al. (2004). RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res, 32*(Web Server issue), W187-190.

Fang, Y., & Fullwood, M. J. (2016). Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics, Proteomics & Bioinformatics, 14*(1), 42-54.

Farrer, M. J., Hulihan, M. M., Kachergus, J. M., Dachsel, J. C., Stoessl, A. J., Grantier, L. L., et al. (2009). DCTN1 mutations in Perry syndrome. *Nat Genet, 41*(2), 163-165.

Feder, J. N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D. A., Basava, A., et al. (1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet, 13*(4), 399-408.

Feng, J., Zhang, Z., Wu, X., Mao, A., Chang, F., Deng, X., et al. (2013). Discovery of Potential New Gene Variants and Inflammatory Cytokine Associations with Fibromyalgia Syndrome by Whole Exome Sequencing. *PLoS One, 8*(6), e65033.

Ferencakovic, M., Solkner, J., & Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol, 45*, 42.

Fiorillo, C., Moro, F., Yi, J., Weil, S., Brisca, G., Astrea, G., et al. (2014). Novel dynein DYNC1H1 neck and motor domain mutations link distal spinal muscular atrophy and abnormal cortical development. *Hum Mutat, 35*(3), 298-302.

Fleckner, J., Rasmussen, H. H., & Justesen, J. (1991). Human interferon gamma potently induces the synthesis of a 55-kDa protein (gamma 2) highly homologous to rabbit peptide chain release factor and bovine tryptophanyl-tRNA synthetase. *Proc Natl Acad Sci U S A, 88*(24), 11520-11524.

Franke, T. F., Kaplan, D. R., Cantley, L. C., & Toker, A. (1997). Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science, 275*(5300), 665-668.

Gentil, B. J., Benaud, C., Delphin, C., Remy, C., Berezowski, V., Cecchelli, R., et al. (2005). Specific AHNAK expression in brain endothelial cells with barrier properties. *J Cell Physiol, 203*(2), 362-371.

Gentil, B. J., & Cooper, L. (2012). Molecular basis of axonal dysfunction and traffic impairments in CMT. *Brain Res Bull, 88*(5), 444-453.

Gillespie, C. S., Sherman, D. L., Blair, G. E., & Brophy, P. J. (1994). Periaxin, a novel protein of myelinating Schwann cells with a possible role in axonal ensheathment. *Neuron, 12*(3), 497-508.

Goecks, J., Nekrutenko, A., Taylor, J., & Team, T. G. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology, 11*(8), R86.

Gonzaga-Jauregui, C., Lupski, J. R., & Gibbs, R. A. (2012). Human genome sequencing in health and disease. *Annu Rev Med, 63*, 35-61.

Goodson, H. V., Skube, S. B., Stalder, R., Valetti, C., Kreis, T. E., Morrison, E. E., et al. (2003). CLIP-170 interacts with dynactin complex and the APC-binding protein EB1 by different mechanisms. *Cell Motil Cytoskeleton, 55*(3), 156-173.

Gorlov, I. P., Gorlova, O. Y., Frazier, M. L., & Amos, C. I. (2003). Missense mutations in hMLH1 and hMSH2 are associated with exonic splicing enhancers. *Am J Hum Genet, 73*(5), 1157-1161.

Gu, W., Xu, Y., Xie, X., Wang, T., Ko, J.-H., & Zhou, T. (2014). The role of RNA structure at 5′ untranslated region in microRNA-mediated gene regulation. *RNA, 20*(9), 1369-1375.

Guilbot, A., Williams, A., Ravise, N., Verny, C., Brice, A., Sherman, D. L., et al. (2001). A mutation in periaxin is responsible for CMT4F, an autosomal recessive form of Charcot-Marie-Tooth disease. *Hum Mol Genet, 10*(4), 415-421.

Gundry, C. N., Vandersteen, J. G., Reed, G. H., Pryor, R. J., Chen, J., & Wittwer, C. T. (2003). Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clin Chem, 49*(3), 396-406.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., et al. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature, 306*(5940), 234-238.

Habura, A., Tikhonenko, I., Chisholm, R. L., & Koonce, M. P. (1999). Interaction mapping of a dynein heavy chain. Identification of dimerization and intermediate-chain binding domains. *J Biol Chem, 274*(22), 15447-15453.

Hafezparast, M., Klocke, R., Ruhrberg, C., Marquardt, A., Ahmad-Annuar, A., Bowen, S., et al. (2003). Mutations in Dynein Link Motor Neuron Degeneration to Defects in Retrograde Transport. *Science, 300*(5620), 808-812.

Haghighi, A., Tiwari, A., Piri, N., Nürnberg, G., Saleh-Gohari, N., Haghighi, A., et al. (2014). Homozygosity Mapping and Whole Exome Sequencing Reveal a Novel Homozygous <italic>COL18A1</italic> Mutation Causing Knobloch Syndrome. *PLoS One, 9*(11), e112747.

Han, H., & Kursula, P. (2014). Periaxin and AHNAK nucleoprotein 2 form intertwined homodimers through domain swapping. *J Biol Chem, 289*(20), 14121-14131.

Hantke, J., Chandler, D., King, R., Wanders, R. J., Angelicheva, D., Tournev, I., et al. (2009). A mutation in an alternative untranslated exon of hexokinase 1 associated with hereditary motor and sensory neuropathy -- Russe (HMSNR). *Eur J Hum Genet, 17*(12), 1606-1614.

Harada, A., Takei, Y., Kanai, Y., Tanaka, Y., Nonaka, S., & Hirokawa, N. (1998). Golgi vesiculation and lysosome dispersion in cells lacking cytoplasmic dynein. *J Cell Biol, 141*(1), 51-59.

Harding, A. E. (1993). Inherited neuronal atrophy and degeneration predominantly of lower motor neurons. In P. A. P. .In:Dyck P. J.; Thomas P. K.; Griffin J. W.; Low, J. F. (Eds.), Peripheral Neuropathy (Vol. Pp. 1051-1064,

Harding, A. E., & Thomas, P. K. (1980). The clinical features of hereditary motor and sensory neuropathy types I and II. *Brain, 103*(2), 259-280.

Harms, M. B., Ori-McKenney, K. M., Scoto, M., Tuck, E. P., Bell, S., Ma, D., et al. (2012). Mutations in the tail domain of DYNC1H1 cause dominant spinal muscular atrophy. *Neurology, 78*(22), 1714-1720.

Hartline, D. K., & Colman, D. R. (2007). Rapid conduction and the evolution of giant axons and myelinated fibers. *Curr Biol, 17*(1), R29-35.

Hashimoto, M., Sagara, Y., Langford, D., Everall, I. P., Mallory, M., Everson, A., et al. (2002). Fibroblast Growth Factor 1 Regulates Signaling via the Glycogen Synthase Kinase-3β Pathway: IMPLICATIONS FOR NEUROPROTECTION. *Journal of Biological Chemistry, 277*(36), 32985-32991.

Hayasaka, K., Takada, G., & Ionasescu, V. V. (1993). Mutation of the myelin P0 gene in Charcot-Marie-Tooth neuropathy type 1B. *Hum Mol Genet, 2*(9), 1369-1372.

Higuchi, Y., Hashiguchi, A., Yuan, J., Yoshimura, A., Mitsui, J., Ishiura, H., et al. (2016). Mutations in MME cause an autosomal-recessive Charcot–Marie–Tooth disease type 2. *Annals of Neurology, 79*(4), 659-672.

Hildebrand, M. S., Tankard, R., Gazina, E. V., Damiano, J. A., Lawrence, K. M., Dahl, H.-H. M., et al. (2015). PRIMA1 mutation: a new cause of nocturnal frontal lobe epilepsy. *Annals of Clinical and Translational Neurology, 2*(8), 821-830.

Hirokawa, N. (1998). Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science, 279*(5350), 519-526.

Hirokawa, N., Noda, Y., Tanaka, Y., & Niwa, S. (2009). Kinesin superfamily motor proteins and intracellular transport. [10.1038/nrm2774]. *Nat Rev Mol Cell Biol, 10*(10), 682-696.

Hohaus, A., Person, V., Behlke, J., Schaper, J., Morano, I., & Haase, H. (2002). The carboxyl-terminal region of ahnak provides a link between cardiac L-type Ca2+ channels and the actin-based cytoskeleton. *FASEB J, 16*(10), 1205-1216.

Holleran, E. A., Tokito, M. K., Karki, S., & Holzbaur, E. L. (1996). Centractin (ARP1) associates with spectrin revealing a potential mechanism to link dynactin to intracellular organelles. *J Cell Biol, 135*(6 Pt 2), 1815-1829.

Hoogenraad, C. C., Akhmanova, A., Howell, S. A., Dortland, B. R., De Zeeuw, C. I., Willemsen, R., et al. (2001). Mammalian Golgi-associated Bicaudal-D2 functions in the dynein–dynactin pathway by interacting with these complexes. *The EMBO Journal, 20*(15), 4041-4054.

Houlden, H., Blake, J., & Reilly, M. M. (2004). Hereditary sensory neuropathies. *Curr Opin Neurol, 17*(5), 569-577.

Houlden, H., Laura, M., Wavrant-De Vrieze, F., Blake, J., Wood, N., & Reilly, M. M. (2008). Mutations in the HSP27 (HSPB1) gene cause dominant, recessive, and sporadic distal HMN/CMT type 2. *Neurology, 71*(21), 1660-1668.

Howrigan, D., Simonson, M., & Keller, M. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics, 12*(1), 460.

Huang, J., Roberts, Anthony J., Leschziner, Andres E., & Reck-Peterson, Samara L. (2012). Lis1 Acts as a Clutch between the ATPase and Microtubule-Binding Domains of the Dynein Motor. *Cell, 150*(5), 975-986.

Huang, Y., Laval, S. H., van Remoortere, A., Baudier, J., Benaud, C., Anderson, L. V. B., et al. (2007). AHNAK, a novel component of the dysferlin protein complex, redistributes to the cytoplasm with dysferlin during skeletal muscle regeneration. *The FASEB Journal, 21*(3), 732-742.

Hung, A. Y., & Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. *J Biol Chem, 277*(8), 5699-5702.

Hunter, M., Bernard, R., Freitas, E., Boyer, A., Morar, B., Martins, I. J., et al. (2003). Mutation screening of the N-myc downstream-regulated gene 1 (NDRG1) in patients with Charcot-Marie-Tooth Disease. *Hum Mutat, 22*(2), 129-135.

International HapMap Project. (2003). The International HapMap Project. *Nature, 426*(6968), 789-796.

Irobi, J., De Jonghe, P., & Timmerman, V. (2004a). Molecular genetics of distal hereditary motor neuropathies. *Hum Mol Genet, 13 Spec No 2*, R195-202.

Irobi, J., Van Impe, K., Seeman, P., Jordanova, A., Dierick, I., Verpoorten, N., et al. (2004b). Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nat Genet, 36*(6), 597-601.

Jacques, T. S., Skepper, J. N., & Navaratnam, V. (1999). Fibroblast growth factor-1 improves the survival and regeneration of rat vagal preganglionic neurones following axon injury. *Neurosci Lett, 276*(3), 197-200.

Johansson, M., Rocha, N., Zwart, W., Jordens, I., Janssen, L., Kuijl, C., et al. (2007). Activation of endosomal dynein motors by stepwise assembly of Rab7-RILP-p150Glued, ORP1L, and the receptor betalll spectrin. *J Cell Biol, 176*(4), 459-471.

Jordanova, A., De Jonghe, P., Boerkoel, C. F., Takashima, H., De Vriendt, E., Ceuterick, C., et al. (2003). Mutations in the neurofilament light chain gene (NEFL) cause early onset severe Charcot-Marie-Tooth disease. *Brain, 126*(Pt 3), 590-597.

Jordanova, A., Irobi, J., Thomas, F. P., Van Dijck, P., Meerschaert, K., Dewil, M., et al. (2006). Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot-Marie-Tooth neuropathy. *Nat Genet, 38*(2), 197-202.

Jordens, I., Fernandez-Borja, M., Marsman, M., Dusseljee, S., Janssen, L., Calafat, J., et al. (2001). The Rab7 effector protein RILP controls lysosomal transport by inducing the recruitment of dynein-dynactin motors. *Current Biology, 11*(21), 1680-1685.

Kabzinska, D., Kochanski, A., Drac, H., Ryniewicz, B., Rowinska-Marcinska, K., & Hausmanowa-Petrusewicz, I. (2005). Autosomal recessive axonal form of Charcot-Marie-Tooth Disease caused by compound heterozygous 3'-splice site and Ser130Cys mutation in the GDAP1 gene. *Neuropediatrics, 36*(3), 206-209.

Kage, M., Yang, Q., Sato, H., Matsumoto, S., Kaji, R., Akiguchi, I., et al. (2001). Acidic fibroblast growth factor (FGF-1) in the anterior horn cells of ALS and control cases. *Neuroreport, 12*(17), 3799-3803.

Kalaydjieva, L., Gresham, D., Gooding, R., Heather, L., Baas, F., de Jonge, R., et al. (2000). N-myc downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. *Am J Hum Genet, 67*(1), 47-58.

Kennerson, M. L., Nicholson, G. A., Kaler, S. G., Kowalski, B., Mercer, J. F. B., Tang, J., et al. (2010). Missense Mutations in the Copper Transporter Gene ATP7A Cause X-Linked Distal Hereditary Motor Neuropathy. *American journal of human genetics, 86*(3), 343-352.

Kennerson, M. L., Yiu, E. M., Chuang, D. T., Kidambi, A., Tso, S. C., Ly, C., et al. (2013). A new locus for X-linked dominant Charcot-Marie-Tooth disease (CMTX6) is caused by mutations in the pyruvate dehydrogenase kinase isoenzyme 3 (PDK3) gene. *Hum Mol Genet, 22*(7), 1404-1416.

Kijima, K., Numakura, C., Izumino, H., Umetsu, K., Nezu, A., Shiiki, T., et al. (2005). Mitochondrial GTPase mitofusin 2 mutation in Charcot-Marie-Tooth neuropathy type 2A. *Hum Genet, 116*(1-2), 23-27.

Kikkawa, M. (2013). Big steps toward understanding dynein. *J Cell Biol, 202*(1), 15-23.

Kim, H. J., Hong, S. H., Ki, C. S., Kim, B. J., Shim, J. S., Cho, S. H., et al. (2005). A novel locus for X-linked recessive CMT with deafness and optic neuropathy maps to Xq21.32-q24. *Neurology, 64*(11), 1964-1967.

Kim, H. J., Sohn, K. M., Shy, M. E., Krajewski, K. M., Hwang, M., Park, J. H., et al. (2007). Mutations in PRPS1, which encodes the phosphoribosyl pyrophosphate synthetase enzyme critical for nucleotide biosynthesis, cause hereditary peripheral neuropathy with hearing loss and optic neuropathy (cmtx5). *Am J Hum Genet, 81*(3), 552-558.

King, S. J., Brown, C. L., Maier, K. C., Quintyne, N. J., & Schroer, T. A. (2003). Analysis of the Dynein-Dynactin Interaction In Vitro and In Vivo. *Molecular Biology of the Cell, 14*(12), 5089-5097.

King, S. J., & Schroer, T. A. (2000). Dynactin increases the processivity of the cytoplasmic dynein motor. *Nat Cell Biol, 2*(1), 20-24.

Kirchmair, R., Tietz, A. B., Panagiotou, E., Walter, D. H., Silver, M., Yoon, Y. S., et al. (2007). Therapeutic angiogenesis inhibits or rescues chemotherapy-induced peripheral neuropathy: taxol- and thalidomide-induced injury of vasa nervorum is ameliorated by VEGF. *Mol Ther, 15*(1), 69-75.

Kirov, A., Kacer, D., Conley, B. A., Vary, C. P., & Prudovsky, I. (2015). AHNAK2 Participates in the Stress-Induced Nonclassical FGF1 Secretion Pathway. *J Cell Biochem, 116*(8), 1522-1531.

Kolb, S. J., Snyder, P. J., Poi, E. J., Renard, E. A., Bartlett, A., Gu, S., et al. (2010). Mutant small heat shock protein B3 causes motor neuropathy: utility of a candidate gene approach. *Neurology, 74*(6), 502-506.

Komuro, A., Masuda, Y., Kobayashi, K., Babbitt, R., Gunel, M., Flavell, R. A., et al. (2004). The AHNAKs are a class of giant propeller-like proteins that associate

with calcium channel proteins of cardiomyocytes and other cells. *Proc Natl Acad Sci U S A, 101*(12), 4053-4058.

Kosambi, D. D. (1943). THE ESTIMATION OF MAP DISTANCES FROM RECOMBINATION VALUES. *Annals of Eugenics, 12*(1), 172-175.

Kotruchow, K., KabziŃSka, D., Hausmanowa-Petrusewicz, I., & KochaŃSki, A. (2013). A late-onset and mild form of Charcot-Marie-Tooth disease type 2 caused by a novel splice-site mutation within the Mitofusin-2 gene. *Acta Myologica, 32*(3), 166-169.

Krawczak, M., Reiss, J., & Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet, 90*(1-2), 41-54.

Ku, C. S., Naidoo, N., & Pawitan, Y. (2011). Revisiting Mendelian disorders through exome sequencing. *Hum Genet, 129*(4), 351-370.

Kuffler, D. P. (2013). Platelet-rich plasma and the elimination of neuropathic pain. *Mol Neurobiol, 48*(2), 315-332.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc, 4*(7), 1073-1081.

Laird, J. M., Mason, G. S., Thomas, K. A., Hargreaves, R. J., & Hill, R. G. (1995). Acidic fibroblast growth factor stimulates motor and sensory axon regeneration after sciatic nerve crush in the rat. *Neuroscience, 65*(1), 209-216.

LaMonte, B. H., Wallace, K. E., Holloway, B. A., Shelly, S. S., Ascaño, J., Tokito, M., et al. (2002). Disruption of Dynein/Dynactin Inhibits Axonal Transport in Motor Neurons Causing Late-Onset Progressive Degeneration. *Neuron, 34*(5), 715-727.

Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. [10.1038/ng1195-241]. *Nat Genet, 11*(3), 241-247.

Lander, E. S., & Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science, 236*(4808), 1567-1570.

Lathrop, G. M., Lalouel, J. M., Julier, C., & Ott, J. (1984). Strategies for multilocus linkage analysis in humans. *Proceedings of the National Academy of Sciences of the United States of America, 81*(11), 3443-3446.

Latour, P., Thauvin-Robinet, C., Baudelet-Mery, C., Soichot, P., Cusin, V., Faivre, L., et al. (2010). A major determinant for binding and aminoacylation of tRNA(Ala) in cytoplasmic Alanyl-tRNA synthetase is mutated in dominant axonal Charcot-Marie-Tooth disease. *Am J Hum Genet, 86*(1), 77-82.

Lee, D. Y., Moon, J., Lee, S. T., Jung, K. H., Park, D. K., Yoo, J. S., et al. (2015a). Dysregulation of long non-coding RNAs in mouse models of localization-related epilepsy. *Biochem Biophys Res Commun, 462*(4), 433-440.

Lee, D. Y., Moon, J., Lee, S. T., Jung, K. H., Park, D. K., Yoo, J. S., et al. (2015b). Distinct Expression of Long Non-Coding RNAs in an Alzheimer's Disease Model. *J Alzheimers Dis, 45*(3), 837-849.

LeGuern, E., Guilbot, A., Kessali, M., Ravisé, N., Tassin, J., Maisonobe, T., et al. (1996). Homozygosity mapping of an autosomal recessive form of demyelinating Charcot-Marie-Tooth disease to chromosome 5q23-q33. *Hum. Mol. Genet, 5*(1685–1688).

Leipold, E., Liebmann, L., Korenke, G. C., Heinrich, T., Giesselmann, S., Baets, J., et al. (2013). A de novo gain-of-function mutation in SCN11A causes loss of pain perception. *Nat Genet, 45*(11), 1399-1404.

Lejeune, F., & Maquat, L. E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol, 17*(3), 309-315.

Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., et al. (2015). Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv*.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, X., Hu, Z., Liu, L., Xie, Y., Zhan, Y., Zi, X., et al. (2015). A SIGMAR1 splice-site mutation causes distal hereditary motor neuropathy. *Neurology, 84*(24), 2430-2437.

Ligon, L. A., Shelly, S. S., Tokito, M., & Holzbaur, E. L. (2003). The microtubule plus-end proteins EB1 and dynactin have differential effects on microtubule polymerization. *Mol Biol Cell, 14*(4), 1405-1417.

Lim, C. M., Cater, M. A., Mercer, J. F., & La Fontaine, S. (2006). Copper-dependent interaction of dynactin subunit p62 with the N terminus of ATP7B but not ATP7A. *J Biol Chem, 281*(20), 14006-14014.

Linhares, N. D., Freire, M. C. M., Cardenas, R. G. C. C. L., Pena, H. B., Bahia, M., & Pena, S. D. J. (2014). Exome sequencing identifies a novel homozygous variant in NDRG4 in a family with infantile myofibromatosis. *European Journal of Medical Genetics, 57*(11–12), 643-648.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408.

Lloyd, T. E. (2012). Axonal transport disruption in peripheral nerve disease: From Jack's discoveries as a resident to recent contributions. *J Peripher Nerv Syst, 17 Suppl 3*, 46-51.

Lloyd, T. E., Machamer, J., O'Hara, K., Kim, J. H., Collins, S. E., Wong, M. Y., et al. (2012). The p150(Glued) CAP-Gly Domain Regulates Initiation of Retrograde Transport at Synaptic Termini. *Neuron, 74*(2), 344-360.

Lo, K. W.-H., Kogoy, J. M., Rasoul, B. A., King, S. M., & Pfister, K. K. (2007). Interaction of the DYNLT (TCTEX1/RP3) Light Chains and the Intermediate Chains Reveals Novel Intersubunit Regulation during Assembly of the Dynein Complex. *Journal of Biological Chemistry, 282*(51), 36871-36878.

Lo Nigro, C., Chong, C. S., Smith, A. C., Dobyns, W. B., Carrozzo, R., & Ledbetter, D. H. (1997). Point mutations and an intragenic deletion in LIS1, the lissencephaly causative gene in isolated lissencephaly sequence and Miller-Dieker syndrome. *Hum Mol Genet, 6*(2), 157-164.

Makokha, M., Hare, M., Li, M., Hays, T., & Barbar, E. (2002). Interactions of cytoplasmic dynein light chains Tctex-1 and LC8 with the intermediate chain IC74. *Biochemistry, 41*(13), 4302-4311.

Marchesi, C., Milani, M., Morbin, M., Cesani, M., Lauria, G., Scaioli, V., et al. (2010). Four novel cases of periaxin-related neuropathy and review of the literature. *Neurology, 75*(20), 1830-1838.

Marg, A., Haase, H., Neumann, T., Kouno, M., & Morano, I. (2010). AHNAK1 and AHNAK2 are costameric proteins: AHNAK1 affects transverse skeletal muscle fiber stiffness. *Biochem Biophys Res Commun, 401*(1), 143-148.

Matise, T. C., Chen, F., Chen, W., De La Vega, F. M., Hansen, M., He, C., et al. (2007). A second-generation combined linkage physical map of the human genome. *Genome Res, 17*(12), 1783-1786.

Matsunami, N., Smith, B., Ballard, L., Lensch, M. W., Robertson, M., Albertsen, H., et al. (1992). Peripheral myelin protein-22 gene maps in the duplication in chromosome 17p11.2 associated with Charcot-Marie-Tooth 1A. *Nat Genet, 1*(3), 176-179.

McDonald, K. K., Stajich, J., Blach, C., Ashley-Koch, A. E., & Hauser, M. A. (2012). Exome analysis of two limb-girdle muscular dystrophy families: mutations identified and challenges encountered. *PLoS One, 7*(11), e48864.

McLaughlin, H. M., Sakaguchi, R., Giblin, W., Wilson, T. E., Biesecker, L., Lupski, J. R., et al. (2012). A recurrent loss-of-function alanyl-tRNA synthetase (AARS) mutation in patients with Charcot-Marie-Tooth disease type 2N (CMT2N). *Hum Mutat, 33*(1), 244-253.

McLaughlin, H. M., Sakaguchi, R., Liu, C., Igarashi, T., Pehlivan, D., Chu, K., et al. (2010). Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. *Am J Hum Genet, 87*(4), 560-566.

Melkonian, K. A., Maier, K. C., Godfrey, J. E., Rodgers, M., & Schroer, T. A. (2007). Mechanism of Dynamitin-mediated Disruption of Dynactin. *Journal of Biological Chemistry, 282*(27), 19355-19364.

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet, 10*(3), 155-159.

Mersiyanova, I. V., Perepelov, A. V., Polyakov, A. V., Sitnikov, V. F., Dadali, E. L., Oparin, R. B., et al. (2000). A new variant of Charcot-Marie-Tooth disease type 2 is probably the result of a mutation in the neurofilament-light gene. *Am J Hum Genet, 67*(1), 37-46.

Misko, A., Jiang, S., Wegorzewska, I., Milbrandt, J., & Baloh, R. H. (2010). Mitofusin 2 is necessary for transport of axonal mitochondria and interacts with the Miro/Milton complex. *J Neurosci, 30*(12), 4232-4240.

Mitchell, D. J., Blasier, K. R., Jeffery, E. D., Ross, M. W., Pullikuth, A. K., Suo, D., et al. (2012). Trk Activation of the ERK1/2 Kinase Pathway Stimulates Intermediate Chain Phosphorylation and Recruits Cytoplasmic Dynein to Signaling Endosomes for Retrograde Axonal Transport. *The Journal of Neuroscience, 32*(44), 15495-15510.

Mohan, H., Friese, A., Albrecht, S., Krumbholz, M., Elliott, C. L., Arthur, A., et al. (2014). Transcript profiling of different types of multiple sclerosis lesions yields FGF1 as a promoter of remyelination. *Acta Neuropathol Commun, 2*, 168.

Morgan, T. H. (1915). *The Mechanism of Mendelian Heredity*: H. Holt.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American journal of human genetics, 7*(3), 277-318.

Morton, N. E. (1984). Linkage and association. *Prog Clin Biol Res, 147*, 245-265.

Moughamian, A. J., & Holzbaur, E. L. (2012). Dynactin is required for transport initiation from the distal axon. *Neuron, 74*(2), 331-343.

Mroczek, M., Kabzinska, D., & Kochanski, A. (2015). Molecular pathogenesis, experimental therapy and genetic counseling in hereditary sensory neuropathies. *Acta Neurobiol Exp (Wars), 75*(2), 126-143.

Munch, C., Meyer, R., Linke, P., Meyer, T., Ludolph, A. C., Haas, J., et al. (2007). The p150 subunit of dynactin (DCTN1) gene in multiple sclerosis. *Acta Neurol Scand, 116*(4), 231-234.

Munch, C., Rosenbohm, A., Sperfeld, A. D., Uttner, I., Reske, S., Krause, B. J., et al. (2005). Heterozygous R1101K mutation of the DCTN1 gene in a family with ALS and FTD. *Ann Neurol, 58*(5), 777-780.

Munch, C., Sedlmeier, R., Meyer, T., Homberg, V., Sperfeld, A. D., Kurt, A., et al. (2004). Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. *Neurology, 63*(4), 724-726.

Narasimhan, V. M., Hunt, K. A., Mason, D., Baker, C. L., Karczewski, K. J., Barnes, M. R., et al. (2016). Health and population effects of rare gene knockouts in adult humans with related parents. *Science, 352*(6284), 474-477.

Neely, M. D., Erickson, H. P., & Boekelheide, K. (1990). HMW-2, the Sertoli cell cytoplasmic dynein from rat testis, is a dimer composed of nearly identical subunits. *J Biol Chem, 265*(15), 8691-8698.

Nelis, E., Erdem, S., Van Den Bergh, P. Y., Belpaire-Dethiou, M. C., Ceuterick, C., Van Gerwen, V., et al. (2002). Mutations in GDAP1: autosomal recessive CMT with demyelination and axonopathy. *Neurology, 59*(12), 1865-1872.

Neveling, K., Martinez-Carrera, Lilian A., Hölker, I., Heister, A., Verrips, A., Hosseini-Barkooie, Seyyed M., et al. (2013). Mutations in BICD2, which Encodes a Golgin and Important Motor Adaptor, Cause Congenital Autosomal-Dominant Spinal Muscular Atrophy. *The American Journal of Human Genetics, 92*(6), 946-954.

Newsway, V., Fish, M., Rohrer, J. D., Majounie, E., Williams, N., Hack, M., et al. (2010). Perry syndrome due to the DCTN1 G71R mutation: A distinctive levodopa responsive disorder with behavioral syndrome, vertical gaze palsy, and respiratory failure. *Movement Disorders, 25*(6), 767-770.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature, 461*(7261), 272-276.

Nicholson, G., & Myers, S. (2006). Intermediate forms of Charcot-Marie-Tooth neuropathy: a review. *Neuromolecular Med, 8*(1-2), 123-130.

Nicholson, G. A., Valentijn, L. J., Cherryson, A. K., Kennerson, M. L., Bragg, T. L., DeKroon, R. M., et al. (1994). A frame shift mutation in the PMP22 gene in hereditary neuropathy with liability to pressure palsies. *Nat Genet, 6*(3), 263-266.

Niemann, A., Ruegg, M., La Padula, V., Schenone, A., & Suter, U. (2005). Ganglioside-induced differentiation associated protein 1 is a regulator of the mitochondrial network: new implications for Charcot-Marie-Tooth disease. *J Cell Biol, 170*(7), 1067-1078.

Nixon, R. A. (2006). Autophagy in neurodegenerative disease: friend, foe or turncoat? *Trends Neurosci, 29*(9), 528-535.

Oates, Emily C., Rossor, Alexander M., Hafezparast, M., Gonzalez, M., Speziani, F., MacArthur, Daniel G., et al. (2013). Mutations in BICD2 Cause Dominant Congenital Spinal Muscular Atrophy and Hereditary Spastic Paraplegia. *The American Journal of Human Genetics, 92*(6), 965-973.

Ott, J., Wang, J., & Leal, S. M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet, 16*(5), 275-284.

Pagani, F., & Baralle, F. E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. [10.1038/nrg1327]. *Nat Rev Genet, 5*(5), 389-396.

Palais, R. A., Liew, M. A., & Wittwer, C. T. (2005). Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. *Anal Biochem, 346*(1), 167-175.

Pankiv, S., Clausen, T. H., Lamark, T., Brech, A., Bruun, J. A., Outzen, H., et al. (2007). p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J Biol Chem, 282*(33), 24131-24145.

Park, C., Chen, X., Yang, J. R., & Zhang, J. (2013). Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A, 110*(8), E678-686.

Patel, P. I., Roa, B. B., Welcher, A. A., Schoener-Scott, R., Trask, B. J., Pentao, L., et al. (1992). The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. *Nat Genet, 1*(3), 159-165.

Peeters, K., Bervoets, S., Chamova, T., Litvinenko, I., De Vriendt, E., Bichev, S., et al. (2015). Novel mutations in the DYNC1H1 tail domain refine the genetic and clinical spectrum of dyneinopathies. *Hum Mutat, 36*(3), 287-291.

Penno, A., Reilly, M. M., Houlden, H., Laura, M., Rentsch, K., Niederkofler, V., et al. (2010). Hereditary sensory neuropathy type 1 is caused by the accumulation of two neurotoxic sphingolipids. *J Biol Chem, 285*(15), 11178-11187.

Pericak-Vance, M. A. (2001). Analysis of genetic linkage data for Mendelian traits. *Curr Protoc Hum Genet, Chapter 1*, Unit 1 4.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res, 29*(9), e45.

Pfister, K. K., Shah, P. R., Hummerich, H., Russ, A., Cotton, J., Annuar, A. A., et al. (2006). Genetic analysis of the cytoplasmic dynein subunit families. *PLoS Genet, 2*(1), e1.

Phillis, R., Statton, D., Caruccio, P., & Murphey, R. K. (1996). Mutations in the 8 kDa dynein light chain gene disrupt sensory axon projections in the Drosophila imaginal CNS. *Development, 122*(10), 2955-2963.

Poirier, K., Lebrun, N., Broix, L., Tian, G., Saillour, Y., Boscheron, C., et al. (2013). Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat Genet 45: 639-647.*

Prior, T. W., Krainer, A. R., Hua, Y., Swoboda, K. J., Snyder, P. C., Bridgeman, S. J., et al. (2009). A positive modifier of spinal muscular atrophy in the SMN2 gene. *Am J Hum Genet, 85*(3), 408-413.

Puliti, A., Caridi, G., Ravazzolo, R., & Ghiggeri, G. M. (2007). Teaching molecular genetics: chapter 4—positional cloning of genetic disorders. *Pediatric Nephrology (Berlin, Germany), 22*(12), 2023-2029.

Puls, I., Jonnakuty, C., LaMonte, B. H., Holzbaur, E. L. F., Tokito, M., Mann, E., et al. (2003). Mutant dynactin in motor neuron disease. [10.1038/ng1123]. *Nat Genet, 33*(4), 455-456.

Puls, I., Oh, S. J., Sumner, C. J., Wallace, K. E., Floeter, M. K., Mann, E. A., et al. (2005). Distal spinal and bulbar muscular atrophy caused by dynactin mutation. *Annals of Neurology, 57*(5), 687-694.

Punetha, J., Monges, S., Franchi, M. E., Hoffman, E. P., Cirak, S., & Tesi-Rocha, C. (2015). Exome Sequencing Identifies DYNC1H1 Variant Associated With Vertebral Abnormality and Spinal Muscular Atrophy With Lower Extremity Predominance. *Pediatr Neurol, 52*(2), 239-244.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Raju, R., Palapetta, S. M., Sandhya, V. K., Sahu, A., Alipoor, A., Balakrishnan, L., et al. (2014). A Network Map of FGF-1/FGFR Signaling System. *J Signal Transduct, 2014*, 962962.

Rasmussen, R. (2001). Quantification on the LightCycler *Rapid cycle real-time PCR* (pp. 21-34): Springer.

Reed, G. H., Kent, J. O., & Wittwer, C. T. (2007). High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics, 8*(6), 597-608.

Reilly, M. M., & Shy, M. E. (2009). Diagnosis and new treatments in genetic neuropathies. *Journal of Neurology, Neurosurgery & Psychiatry, 80*(12), 1304-1314.

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res, 39*(17), e118.

Rinaldi, C., Grunseich, C., Sevrioukova, I. F., Schindler, A., Horkayne-Szakaly, I., Lamperti, C., et al. (2012). Cowchock syndrome is associated with a mutation in apoptosis-inducing factor. *Am J Hum Genet, 91*(6), 1095-1102.

Romashkova, J. A., & Makarov, S. S. (1999). NF-kappaB is a target of AKT in anti-apoptotic PDGF signalling. *Nature, 401*(6748), 86-90.

Ross, J. (1995). mRNA stability in mammalian cells. *Microbiol Rev, 59*(3), 423-450.

Rossor, A. M., Evans, M. R., & Reilly, M. M. (2015). A practical approach to the genetic neuropathies. *Pract Neurol, 15*(3), 187-198.

Royer-Pokora, B., Kunkel, L. M., Monaco, A. P., Goff, S. C., Newburger, P. E., Baehner, R. L., et al. (1986). Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. *Nature, 322*(6074), 32-38.

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol, 132*, 365-386.

Sabet, A., Li, J., Ghandour, K., Pu, Q., Wu, X., Kamholz, J., et al. (2006). Skin biopsies demonstrate MPZ splicing abnormalities in Charcot-Marie-Tooth neuropathy 1B. *Neurology, 67*(7), 1141-1146.

Saillour, Y., Carion, N., Quelin, C., Leger, P. L., Boddaert, N., Elie, C., et al. (2009). LIS1-related isolated lissencephaly: spectrum of mutations and relationships with malformation severity. *Arch Neurol, 66*(8), 1007-1015.

Salim, C., Boxberg, Y. V., Alterio, J., Fereol, S., & Nothias, F. (2009). The giant protein AHNAK involved in morphogenesis and laminin substrate adhesion of myelinating Schwann cells. *Glia, 57*(5), 535-549.

Saporta, A. S., Sottile, S. L., Miller, L. J., Feely, S. M., Siskind, C. E., & Shy, M. E. (2011). Charcot-Marie-Tooth disease subtypes and genetic testing strategies. *Ann Neurol, 69*(1), 22-33.

Saporta, M. A., & Shy, M. E. (2013). Inherited peripheral neuropathies. *Neurol Clin, 31*(2), 597-619.

Sapunar, D., Kostic, S., Banozic, A., & Puljak, L. (2012). Dorsal root ganglion - a potential new therapeutic target for neuropathic pain. *J Pain Res, 5*, 31-38.

Saxena, S., Bucci, C., Weis, J., & Kruttgen, A. (2005). The small GTPase Rab7 controls the endosomal trafficking and neuritogenic signaling of the nerve growth factor receptor TrkA. *J Neurosci, 25*(47), 10930-10940.

Saxton, W. M., & Hollenbeck, P. J. (2012). The axonal transport of mitochondria. *Journal of Cell Science, 125*(9), 2095-2104.

Schafer, D. A., Gill, S. R., Cooper, J. A., Heuser, J. E., & Schroer, T. A. (1994). Ultrastructural analysis of the dynactin complex: an actin-related protein is a component of a filament that resembles F-actin. *J Cell Biol, 126*(2), 403-412.

Scherer, S. S., Xu, Y. T., Bannerman, P. G., Sherman, D. L., & Brophy, P. J. (1995). Periaxin expression in myelinating Schwann cells: modulation by axon-glial interactions and polarized localization during development. *Development, 121*(12), 4265-4273.

Schiavo, G., Greensmith, L., Hafezparast, M., & Fisher, E. M. (2013). Cytoplasmic dynein heavy chain: the servant of many masters. *Trends Neurosci, 36*(11), 641-651.

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., & Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A, 109*(36), 14508-14513.

Schroer, T. A. (2004). Dynactin. *Annu Rev Cell Dev Biol, 20*, 759-779.

Schwarz, J. M., Rodelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. [10.1038/nmeth0810-575]. *Nat Meth, 7*(8), 575-576.

Scoto, M., Rossor, A. M., Harms, M. B., Cirak, S., Calissano, M., Robb, S., et al. (2015). Novel mutations expand the clinical spectrum of DYNC1H1-associated spinal muscular atrophy. *Neurology, 84*(7), 668-679.

Shahrizaila, N., Samulong, S., Tey, S., Suan, L. C., Meng, L. K., Goh, K. J., et al. (2014). X-linked charcot-marie-tooth disease predominates in a cohort of multi-ethnic Malaysian patients. *Muscle & Nerve, 49*(2), 198-201.

Sherman, D. L., Fabrizi, C., Gillespie, C. S., & Brophy, P. J. (2001). Specific Disruption of a Schwann Cell Dystrophin-Related Protein Complex in a Demyelinating Neuropathy. *Neuron, 30*(3), 677-687.

Sherman, D. L., Wu, L. M., Grove, M., Gillespie, C. S., & Brophy, P. J. (2012). Drp2 and periaxin form Cajal bands with dystroglycan but have distinct roles in Schwann cell growth. *J Neurosci, 32*(27), 9419-9428.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research, 29*(1), 308-311.

Shiga, N., Takeshima, Y., Sakamoto, H., Inoue, K., Yokota, Y., Yokoyama, M., et al. (1997). Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J Clin Invest, 100*(9), 2204-2210.

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., et al. (2013). Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Human Mutation, 34*(1), 57-65.

Shtivelman, E., Cohen, F. E., & Bishop, J. M. (1992). A human gene (AHNAK) encoding an unusually large protein with a 1.2-microns polyionic rod structure. *Proceedings of the National Academy of Sciences of the United States of America, 89*(12), 5472-5476.

Siepel, A., & Haussler, D. (2004). Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol, 21*(3), 468-488.

Skre, H. (1974). Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clin Genet, 6*(2), 98-118.

Smith, K., Bromhead, C., Hildebrand, M., Shearer, A. E., Lockhart, P., Najmabadi, H., et al. (2011). Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. *Genome Biology, 12*(9), R85.

Spinosa, M. R., Progida, C., De Luca, A., Colucci, A. M., Alifano, P., & Bucci, C. (2008). Functional characterization of Rab7 mutant proteins associated with Charcot-Marie-Tooth type 2B disease. *J Neurosci, 28*(7), 1640-1648.

Stella, A., Wagner, A., Shito, K., Lipkin, S. M., Watson, P., Guanti, G., et al. (2001). A nonsense mutation in MLH1 causes exon skipping in three unrelated HNPCC families. *Cancer Res, 61*(19), 7020-7024.

Stenson, P. D., Ball, E. V., Howells, K., Phillips, A. D., Mort, M., & Cooper, D. N. (2009). The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics, 4*(2), 69-72.

Strickland, A. V., Schabhuttl, M., Offenbacher, H., Synofzik, M., Hauser, N. S., Brunner-Krainz, M., et al. (2015). Mutation screen reveals novel variants and expands the phenotypes associated with DYNC1H1. *J Neurol, 262*(9), 2124-2134.

Stum, M., McLaughlin, H. M., Kleinbrink, E. L., Miers, K. E., Ackerman, S. L., Seburn, K. L., et al. (2011). An assessment of mechanisms underlying peripheral axonal degeneration caused by aminoacyl-tRNA synthetase mutations. *Mol Cell Neurosci, 46*(2), 432-443.

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., & Lehner, B. (2014). Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. *Cell, 156*(6), 1324-1335.

Susalka, S. J., Nikulina, K., Salata, M. W., Vaughan, P. S., King, S. M., Vaughan, K. T., et al. (2002). The roadblock light chain binds a novel region of the cytoplasmic Dynein intermediate chain. *J Biol Chem, 277*(36), 32939-32946.

Suzuki, Y., Ishikawa, N., Tanihara, M., & Saito, S. (2016). Nontubulation Repair of Peripheral Nerve Gap Using Heparin/Alginate Gel Combined with b-FGF. *Plast Reconstr Surg Glob Open, 4*(1), e600.

Takashima, H., Boerkoel, C. F., De Jonghe, P., Ceuterick, C., Martin, J. J., Voit, T., et al. (2002). Periaxin mutations cause a broad spectrum of demyelinating neuropathies. *Ann Neurol, 51*(6), 709-715.

Takata, R. I., Speck Martins, C. E., Passosbueno, M. R., Abe, K. T., Nishimura, A. L., Dorvalina Da Silva, M., et al. (2004). A new locus for recessive distal spinal muscular atrophy at Xq13.1–q21. *Journal of Medical Genetics, 41*(3), 224-229.

Tang, B. S., Zhao, G. H., Luo, W., Xia, K., Cai, F., Pan, Q., et al. (2005). Small heat-shock protein 22 mutated in autosomal dominant Charcot-Marie-Tooth disease type 2L. *Hum Genet, 116*(3), 222-224.

Tanzi, R. E., Petrukhin, K., Chernov, I., Pellequer, J. L., Wasco, W., Ross, B., et al. (1993). The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. *Nat Genet, 5*(4), 344-350.

Tayebi, N., Jamsheer, A., Flottmann, R., Sowinska-Seidler, A., Doelken, S. C., Oehl-Jaschkowitz, B., et al. (2014). Deletions of exons with regulatory activity at the DYNC1I1 locus are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. *Orphanet J Rare Dis, 9*, 108.

Teer, J. K., & Mullikin, J. C. (2010). Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet, 19*(R2), R145-151.

Tey, S., Ahmad-Annuar, A., Drew, A. P., Shahrizaila, N., Nicholson, G. A., & Kennerson, M. L. (2014). Analysis of dynein intermediate chains, light intermediate chains and light chains in a cohort of hereditary peripheral neuropathies. *Neurogenetics, 15*(4), 229-235.

Tey, S., Ahmad-Annuar, A., Drew, A. P., Shahrizaila, N., Nicholson, G. A., & Kennerson, M. L. (2016). Mutation analysis of genes within the dynactin complex in a cohort of hereditary peripheral neuropathies. *Clin Genet, 90*(2), 127-133.

Timmerman, V., Nelis, E., Van Hul, W., Nieuwenhuijsen, B. W., Chen, K. L., Wang, S., et al. (1992). The peripheral myelin protein gene PMP-22 is contained within the Charcot-Marie-Tooth disease type 1A duplication. *Nat Genet, 1*(3), 171-175.

Timmerman, V., Strickland, A., & Züchner, S. (2014). Genetics of Charcot-Marie-Tooth (CMT) Disease within the Frame of the Human Genome Project Success. *Genes, 5*(1), 13-32.

Tooth, H. (1886). The peroneal type of progressive muscular atrophy. [Ref Type: thesis/dissertation]. *H.K. Lewis*.

Trotta, E. (2016). Selective forces and mutational biases drive stop codon usage in the human genome: a comparison with sense codon usage. *BMC Genomics, 17*, 366.

Tsurusaki, Y., Saitoh, S., Tomizawa, K., Sudo, A., Asahina, N., Shiraishi, H., et al. (2012). A DYNC1H1 mutation causes a dominant spinal muscular atrophy with lower extremity predominance. *Neurogenetics, 13*(4), 327-332.

Tynan, S. H., Purohit, A., Doxsey, S. J., & Vallee, R. B. (2000). Light Intermediate Chain 1 Defines a Functional Subfraction of Cytoplasmic Dynein Which Binds to Pericentrin. *Journal of Biological Chemistry, 275*(42), 32763-32768.

Valentijn, L. J., Bolhuis, P. A., Zorn, I., Hoogendijk, J. E., van den Bosch, N., Hensels, G. W., et al. (1992). The peripheral myelin gene PMP-22/GAS-3 is duplicated in Charcot-Marie-Tooth disease type 1A. *Nat Genet, 1*(3), 166-170.

Vaughan, K. T., & Vallee, R. B. (1995). Cytoplasmic dynein binds dynactin through a direct interaction between the intermediate chains and p150Glued. *J Cell Biol, 131*(6 Pt 1), 1507-1516.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science, 291*(5507), 1304-1351.

Verhoeven, K., De Jonghe, P., Coen, K., Verpoorten, N., Auer-Grumbach, M., Kwon, J. M., et al. (2003). Mutations in the Small GTP-ase Late Endosomal Protein RAB7 Cause Charcot-Marie-Tooth Type 2B Neuropathy. *The American Journal of Human Genetics, 72*(3), 722-727.

Verhoeven, K., De Jonghe, P., Coen, K., Verpoorten, N., Auer-Grumbach, M., Kwon, J. M., et al. (2003). Mutations in the small GTP-ase late endosomal protein RAB7 cause Charcot-Marie-Tooth type 2B neuropathy. *Am J Hum Genet, 72*(3), 722-727.

Vester, A., Velez-Ruiz, G., McLaughlin, H. M., Lupski, J. R., Talbot, K., Vance, J. M., et al. (2013). A loss-of-function variant in the human histidyl-tRNA synthetase (HARS) gene is neurotoxic in vivo. *Hum Mutat, 34*(1), 191-199.

Vilarino-Guell, C., Wider, C., Soto-Ortolaza, A. I., Cobb, S. A., Kachergus, J. M., Keeling, B. H., et al. (2009). Characterization of DCTN1 genetic variability in neurodegeneration. *Neurology, 72*(23), 2024-2028.

Viollet, L., Zarhrate, M., Maystadt, I., Estournet-Mathiaut, B., Barois, A., Desguerre, I., et al. (2004). Refined genetic mapping of autosomal recessive chronic distal spinal muscular atrophy to chromosome 11q13.3 and evidence of linkage disequilibrium in European families. *Eur J Hum Genet, 12*(6), 483-488.

Vissers, L. E. L. M., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., et al. (2010). A de novo paradigm for mental retardation. [10.1038/ng.712]. *Nat Genet, 42*(12), 1109-1112.

von Boxberg, Y., Salim, C., Soares, S., Baloui, H., Alterio, J., Ravaille-Veron, M., et al. (2006). Spinal cord injury-induced up-regulation of AHNAK, expressed in cells delineating cystic cavities, and associated with neoangiogenesis. *Eur J Neurosci, 24*(4), 1031-1041.

von Boxberg, Y., Soares, S., Féréol, S., Fodil, R., Bartolami, S., Taxi, J., et al. (2014). Giant scaffolding protein AHNAK1 interacts with β-dystroglycan and controls motility and mechanical properties of schwann cells. *Glia, 62*(9), 1392-1406.

Vulpe, C., Levinson, B., Whitney, S., Packman, S., & Gitschier, J. (1993). Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. *Nat Genet, 3*(1), 7-13.

Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M. K., Thornton, A. M., Roeb, W., et al. (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet, 87*(1), 90-94.

Wang, J. T., Medress, Z. A., & Barres, B. A. (2012). Axon degeneration: Molecular mechanisms of a self-destruction pathway. *The Journal of Cell Biology, 196*(1), 7-18.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res, 38*(16), e164.

Warner, L. E., Mancias, P., Butler, I. J., McDonald, C. M., Keppen, L., Koob, K. G., et al. (1998). Mutations in the early growth response 2 (EGR2) gene are associated with hereditary myelinopathies. *Nat Genet, 18*(4), 382-384.

Weedon, Michael N., Hastings, R., Caswell, R., Xie, W., Paszkiewicz, K., Antoniadi, T., et al. (2011). Exome Sequencing Identifies a DYNC1H1 Mutation in a Large Pedigree with Dominant Axonal Charcot-Marie-Tooth Disease. *American journal of human genetics, 89*(2), 308-312.

Weisbrich, A., Honnappa, S., Jaussi, R., Okhrimenko, O., Frey, D., Jelesarov, I., et al. (2007). Structure-function relationship of CAP-Gly domains. [10.1038/nsmb1291]. *Nat Struct Mol Biol, 14*(10), 959-967.

Willemsen, M. H., Vissers, L. E. L., Willemsen, M. A. A. P., van Bon, B. W. M., Kroes, T., de Ligt, J., et al. (2012). Mutations in DYNC1H1 cause severe intellectual disability with neuronal migration defects. *Journal of Medical Genetics, 49*(3), 179-183.

Windpassinger, C., Auer-Grumbach, M., Irobi, J., Patel, H., Petek, E., Horl, G., et al. (2004). Heterozygous missense mutations in BSCL2 are associated with distal hereditary motor neuropathy and Silver syndrome. *Nat Genet, 36*(3), 271-276.

Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., & Pryor, R. J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem, 49*(6 Pt 1), 853-860.

Woolfe, A., Mullikin, J. C., & Elnitski, L. (2010). Genomic features defining exonic variants that modulate splicing. *Genome Biology, 11*(2), R20-R20.

Yao, I., Iida, J., Nishimura, W., & Hata, Y. (2002). Synaptic and nuclear localization of brain-enriched guanylate kinase-associated protein. *J Neurosci, 22*(13), 5354-5364.

Ylikallio, E., Poyhonen, R., Zimon, M., De Vriendt, E., Hilander, T., Paetau, A., et al. (2013). Deficiency of the E3 ubiquitin ligase TRIM2 in early-onset axonal neuropathy. *Hum Mol Genet, 22*(15), 2975-2983.

Zemmouri, R., Azzedine, H., Assami, S., Kitouni, N., Vallat, J. M., Maisonobe, T., et al. (2000). Charcot-Marie-Tooth 2-like presentation of an Algerian family with giant axonal neuropathy. *Neuromuscul Disord, 10*(8), 592-598.

Zhu, M., & Zhao, S. (2007). Candidate Gene Identification Approach: Progress and Challenges. *International Journal of Biological Sciences, 3*(7), 420-427.

Zimon, M., Baets, J., Almeida-Souza, L., De Vriendt, E., Nikodinovic, J., Parman, Y., et al. (2012). Loss-of-function mutations in HINT1 cause axonal neuropathy with neuromyotonia. *Nat Genet, 44*(10), 1080-1083.

Zimon, M., Battaloglu, E., Parman, Y., Erdem, S., Baets, J., De Vriendt, E., et al. (2015). Unraveling the genetic landscape of autosomal recessive Charcot-Marie-Tooth neuropathies using a homozygosity mapping approach. *Neurogenetics, 16*(1), 33-42.

Zuchner, S., Mersiyanova, I. V., Muglia, M., Bissar-Tadmouri, N., Rochelle, J., Dadali, E. L., et al. (2004). Mutations in the mitochondrial GTPase mitofusin 2 cause Charcot-Marie-Tooth neuropathy type 2A. *Nat Genet, 36*(5), 449-451.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406-3415.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Publications of research articles

Tey, S., Ahmad-Annuar, A., Drew, A. P., Shahrizaila, N., Nicholson, G. A., Kennerson, M. L. Mutation analysis of genes within the dynactin complex in a cohort of hereditary peripheral neuropathies. Clinical Genetics, 90(2), 127-133 (2016).

Drew, A. P., Zhu, D., Kidambi, A., Ly, C., **Tey, S**., Brewer, M. H. et al. Improved inherited peripheral neuropathy genetic diagnosis by whole-exome sequencing. Molecular Genetics & Genomic Medicine 3, 143-154 (2015).

Tey, S., Ahmad-Annuar, A., Drew, A. P., Shahrizaila, N., Nicholson, G. A., Kennerson, M. L. Analysis of dynein intermediate chains, light intermediate chains and light chains in a cohort of hereditary peripheral neuropathies. Neurogenetics 15, 229-35 (2014).

Shahrizaila, N., Samulong, S., **Tey, S**., Suan, L. C., Meng, L. K., Goh, K. J. et al**.** X-linked charcot-marie-tooth disease predominates in a cohort of multi-ethnic Malaysian patients. Muscle & Nerve, 49(2), 198-201.

Conference presentations

Oral presentations

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Kennerson, M. L., Ahmad-Annuar, A. Gene mapping for a consanguineous family with Charcot-Marie-Tooth (CMT) disease. GeneMappers, Adelaide, Australia (2014).

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Kennerson, M. L., Ahmad-Annuar, A. Exome sequencing of a Malaysian family with Charcot-Marie-Tooth (CMT) disease. 4th International NeuroMalaysia Symposium, Kuala Lumpur (2013).

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Kennerson, M. L., Ahmad-Annuar, A. Investigation of the genetic causes in a Malaysian family with CMT. 5th International CMT Consortium Meeting, Antwerp (2013).

Poster presentations

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Nicholson, G. A., Kennerson, M. L., Ahmad-Annuar, A., Finding the causative gene in a consanguineous family with Charcot-Marie-Tooth (CMT) disease. Human Genome Meeting (HGM2015), Kuala Lumpur (2015).

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Kennerson, M. L., Ahmad-Annuar, A. Whole exome sequencing a consanguineous family in search for a novel genetic cause in Charcot-Marie-Tooth (CMT) disease. American Society of Human Genetics (ASHG), San Diego, CA (2014).

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Kennerson, M. L., Ahmad-Annuar, A. Analysis of Dynein-Dynactin genes in Inherited Peripheral Neuropathies. 3rd International NeuroMalaysia Symposium, Kuala Lumpur (2012).

Tey, S., Shahrizaila, N., Goh, K. J., A., Drew, A. P., Kennerson, M. L., Ahmad-Annuar, A. Analysis of Dynein-Dynactin genes in Inherited Peripheral Neuropathies. 9th GeneMappers Conference, Port Douglas, Cairns, Australia (2012).