Genetic Investigations of Sporadic Inclusion Body Myositis and Myopathies with Structural Abnormalities and Protein Aggregates in Muscle

Qiang Gang

MRC Centre for Neuromuscular Diseases and UCL Institute of Neurology

Supervised by Professor Henry Houlden, Dr Conceição Bettencourt and Professor Michael Hanna

Thesis submitted for the degree of Doctor of Philosophy

University College London

2016

Declaration

I, Qiang Gang, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature……

Date……...

Abstract

The application of whole-exome sequencing (WES) has not only dramatically accelerated the discovery of pathogenic genes of Mendelian diseases, but has also shown promising findings in complex diseases. This thesis focuses on exploring genetic risk factors for a large series of sporadic inclusion body myositis (sIBM) cases, and identifying disease-causing genes for several groups of patients with abnormal structure and/or protein aggregates in muscle. Both conventional and advanced techniques were applied.

Based on the International IBM Genetics Consortium (IIBMGC), the largest sIBM cohort of blood and muscle tissue for DNA analysis was collected as the initial part of this thesis. Candidate gene studies were carried out and revealed a disease modifying effect of an intronic polymorphism in *TOMM40*, enhanced by the *APOE* ε3/ε3 genotype. Rare variants in *SQSTM1* and *VCP* genes were identified in seven of 181 patients, indicating a mutational overlap with neurodegenerative diseases. Subsequently, a first whole-exome association study was performed on 181 sIBM patients and 510 controls. This reported statistical significance of several common variants located on chromosome 6p21, a region encompassing genes related to inflammation/infection.

WES was performed on a group of 35 cases with tubular aggregates/cylindrical spirals, and detected rare variants in known/candidate genes. Disease-causing genes were identified in four families with protein aggregates in muscle also by WES. In one family identified with a novel homozygous deletion in *SBF1* with a rare autosomal-recessive neuromuscular condition, functional analysis was carried out indicating a loss-of-function mechanism underlying the pathogenesis of the disease.

The collection of a large series of sIBM patients through the IIBMGC has been shown here to reveal important genetic findings and will be a valuable resource for the future. WES proved to be important in sIBM and also to be an efficient method to investigate the genetics basis of rare complex muscle disorders.

Acknowledgements

First of all, I would like to thank my supervisors Professor Henry Houlden, Dr Conceição Bettencourt and Professor Michael Hanna for giving me the opportunity to work on this thesis. I express my deepest gratitude for their support, encouragement, scientific guidance and inspiration throughout the last four years.

I specially thank Dr Pedro Machado for his coordination with sIBM study collaborators and his clinical knowledge and guidance when I needed it. I would like to thank Dr Alan Pittman for his guidance in bioinformatics; and also to thank Professor Janice Holton for her help with biopsy images. This PhD would also have not been possible without the support, help and guidance from all the members within the Department of Molecular Neuroscience, MRC Centre for Neuromuscular Diseases, and Neurogenetics Lab, in particular Debbie, Mark, Chris, Hallgeir, Estelle, David L, Lucia, Sarah W, Niccolo, Stephanie, Josh, Steven, Jack, James, Ese, Iwona, Enrico, Zuzanna, Renata, and Karen. Special thanks also go to Alice, Ellen, Monika, Siobhan, and Amelie, for their advice and friendship - sharing an office with them has been a pleasure; my gratefulness in particular to Alice for her time to comment on this thesis.

I am also hugely grateful to my family, in particular to my parents for their never-ending love, support, and belief in me to allow me to pursue my dreams; also to my aunt and uncle, Yi Gang and Zhong-Yi Zhang, for their excellent care and for providing me with a comfortable home since I first arrived in London six years ago. I would also like to thank my cousin, Xiao-Yin, and her husband, Matthew, for their encouragement and also for taking the time to read through this thesis. I also thank my dear friends from UCL, Ya Hu, Bao-Luen Chang, Li-Xun Liu, Jin Si, Xing-Wei Zhao, Xi Ji, and many others with whom I have shared the ups and downs of our PhD life, and made it an unforgettable period of my life.

I would like to extend my thanks to the UCL Impact Studentship and Chinese Scholarship Council, who provided financial support during my PhD. Lastly, thanks

to all the collaborators, patients and their families who have taken part in the studies in this thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

α1ACT	Alpha-1-antichymotyrpsin
Aβ	β-amyloid protein
αB-crystallin	Alpha-crystallin B chain
AβPP	β-amyloid precursor protein
AD	Alzheimer's disease
AH	Ancestral haplotype
ALS	Amyotrophic lateral sclerosis
ANNOVAR	Annotate Variation
ApoE	Apolipoprotein E
ATG5	Autophagy protein 5
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
BACE	Beta-secretase
BAM	Binary Alignment/Map
BD	Becton, Dickinson and Company
bp	Base pair
BSA	Bovine serum albumin
CAD	CRAC-activating domain
CADD	Combined Annotation Dependent Depletion
CAST	Cohort allelic sums test
CBD	Corticobasal degeneration
CC1	Coiled-coil 1
CCL	Chemokine ligand
CCSD	The Consensus CDS project
CDC48	Cell division protein 48
CDG	Congenital disorders of glycosylation
cDNA	Complementary deoxyribonucleic acid
CG69	Complete Genomics database
CGH	Comparative genomic hybridization
CK	Creatine kinase
CMC	Combined and multivariate collapsing

CMS	Congenital myasthenic syndrome
CMT	Charcot-Marie-Tooth disease
cN1A	Cytoplasmic 5'-nucleotidase 1A
CNM	Centronuclear myopathy
COX	Cytochrome *c* oxidase
CRAC	Calcium release-activated channel
CSs	Cylindrical spirals
CT	Computed tomography
Ct	Threshold cycle
CTID	C-terminal inhibitory domain
CXCL	Chemokine ligand
CXCR	Chemokine receptor
dbSNP	Single Nucleotide Polymorphism Database
ddNTP	Dideoxynucleotides
dH_2O	Distilled H_2O
DHPR	Dihydropyridine receptor
DM	Dermatomyositis
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
DP	Depth
DPBS	Dulbecco's phosphate buffered saline
DPR	Dipeptide repeat protein
EC	Exponential combination
ECL	Electrochemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMG	Electromyography
EMMPAT	Evolutionary mixed model for pooled association testing
ENMC	European Neuromuscular Centre
eQTL	Expression quantitative trait loci
ER	Endoplasmic reticulum
EVS	Exome Variants Server
ExAC	Exome Aggregation Consortium

FDR	False discovery rate
fIBM	Familial inclusion body myositis
FTD	Frontotemporal dementia
FTDP-17	Frontotemporal dementia, Parkinsonism linked to chromosome 17
FTLD-U	Tau-negative frontotemporal lobar dementia
FUS	Fused in sarcoma protein
FWER	Family-wise error rate
GATK	Genome Analysis Toolkit
GCTA	Genome-wide Complex Trait Analysis
gDNA	Genomic DNA
GERP++	Genomic Evolutionary Rate Profiling
GO	Gene ontology
GQ	Genotype quality
GRAM	Glucosyltransferase, Rab-like GTPase Activator and Myotubularins
GRCh37	Human genome assembly build 37
GSK3β	Glycogen synthase kinase-3beta
gVCF	Genomic VCF
GWAS	Genome-wide association study
H&E	Hematoxylin & eosin
HEK293	Human embryonic kidney 293
hg	Human genome
HIV	Human immunodeficiency virus
hIBM	hereditary inclusion body myopathy
HLA	Human leukocyte antigen
hnRNP	Heterogeneous nuclear ribonucleoprotein
HPO	Human phenotype ontology
HRP	Horseradish peroxidase
HSPs	Heat shock proteins
HTLV-1	Human T-lymphotropic virus-1
HWE	Hardy-Weinberg Equilibrium
IBD	Identity by descent

IBMPFD	Inclusion body myopathy with early-onset Paget's disease of the bone and frontotemporal dementia
IFN-γ	Interferon gamma
IgG	Immunoglobulin G
IIBMCGS	International IBM Consortium Genetic Study
IIMs	Idiopathic inflammatory myopathies
IL-1	Interleukin-1
Indels	Insertions/deletions
ION	Institute of Neurology
JDM	Juvenile dermatomyositis
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	Long
LASSO	Least absolute shrinkage and selection operator
LC3	Microtubule-associated protein 1A/1B-light chain 3
LD	Linkage disequilibrium
LGC	Laboratory of the Government Chemist
LGMD	Limb girdle muscular dystrophy
LIR	LC3-interaction region
LOAD	Late onset AD
LOD	Logarithm (base 10) of odds
LOF	Loss of function
MAD	Myoadenylate deaminase
MAF	Minor allele frequency
MFM	Myofibrillar myopathy
MFS	Marfan's syndrome
MHC	Major histocompatibility complex
MRC	Medical Research Council
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MSP	Multisystem proteinopathy
mtDNA	Mitochondrial DNA
MTMR	Myotubularin-related
N/A	Not available/ not applicable
NADH-TR	Nicotinamide adenine dinucleotide-tetrazolium reductase

NBR1	Neighbour of BRCA1 gene 1
NCS	Nerve conduction study
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next-generation sequencing
NHNN	National Hospital of Neurology and Neurosurgery
NOS	Nitric oxide synthase
$Np40^{+}$	Nonyl phenoxypolyethoxylethanol
NRES	National Research Ethics Service
OMIM	Online Mendelian inheritance in man
OR	Odds ratio
PAMs	Protein aggregate myopathies
PAS	Periodic acid-Schiff
PB1	Phox and Bem1
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD	Parkinson's disease
PDB	Paget's disease of the bone
PGAM	Phosphoglycerate mutase
PM	Polymyositis
PolyPhen2	Polymorphism Phenotyping v2
PSP	Progressive supranuclear palsy
Ptdlns3*P*	Phosphatidylinositol 3-phosphate
Ptdlns[3,5]P_2	Phosphatidylinositol 3,5-bisphosphate
PTH	Parathyroid hormone
PTP/DSP	tyrosine/dual-specificity phosphatase super-family
QC	Quality control
QQ	Quantile-quantile
RAN	Repeat-associated non-ATG
RBT	Replication-based test
RFLP	Restriction fragment length polymorphism
RID	Rac-Induced recruitment domain
RIN	RNA integrity number
RNA	Ribonucleic acid
RP-PCR	Repeat-primed PCR

RR	Relative risk
RT-qPCR	Real-time quantitative PCR
RVAS	Rare variant association study
RyR	Ryanodine receptor
S	Short
SAM	Sequence Alignment/Map
SD	Standard deviation
SDH	Succinate dehydrogenase
SDS	Sodium dodecyl sulphate
SERCAs	Sarcoplasmic endoplasmic reticulum calcium ATPases
sIBM	Sporadic inclusion body myositis
SID	SET-interacting domain
SIFT	Sorts intolerant from tolerant
SKAT	Sequence-based kernel association test
SKAT-O	Optimal SKAT
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SOAR	STIM1-Orai-activating region
SOCE	Store-operated calcium entry
SR	Sarcoplasmic reticulum
SST	Serum Separation Tubes
SSU	Sum of squared score
TAM	Tubular aggregate myopathy
TAs	Tubular aggregates
TBE	Tris-Borate-EDTA
TBS	Tumour necrosis factor receptor-associated factor 6 (TRAF6)-binding site
TBS (buffer)	Tris-buffered saline
TDP-43	Transactive response (TAR) DNA-binding protein-43
TNF-α	Tumour necrosis factor alpha
TOMM40	Translocase of outer mitochondrial membrane 40
TRAF6	Tumour necrosis factor receptor-associated factor 6
T-tubule	Transverse tubule
UBA	Ubiquitin associated domain

UBB^{+1}	Mutated ubiquitin
UCL	University College London
UCLA	University of California, LA
UCSC	University of California, Santa Cruz
UPS	Ubiquitin-proteasome system
UV	Ultraviolet
VCF	Variant call format
VEP	Variant Effect Predictor
VL	Very long
VQSR	Variant quality score recalibration
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WSC	Weighted-sum collapsing

List of Publications Arising from This Thesis

1. **Rare variants in *SQSTM1* and *VCP* genes and risk of sporadic inclusion body myositis.**
   Gang Q, Bettencourt C, Machado PM, Brady S, Holton JL, Pittman AM, Hughes D, Healy E, Parton M, Hilton-Jones D, Shieh PB, Zanoteli E, Camargo LV, De Paepe B, De Bleecker J, Shaibani A, Ripolone M, Violano R, Moggio M, Barohn RJ, Dimachkie MM, Mora M, Mantegazza R, Zanotti S, Singleton AB, Hanna MG, Houlden H; Muscle Study Group and the International IBM Genetics Consortium. Neurobiol Aging. 2016 Nov;47:218.e1-218.e9.

2. **Genetic advances in sporadic inclusion body myositis**.
   Gang Q, Bettencourt C, Houlden H, Hanna MG, Machado PM. Curr Opin. 2015 Nov;27(6):586-94.

3. **The effects of an intronic polymorphism in *TOMM40* and *APOE* genotypes in sporadic inclusion body myositis.**
   Gang Q, Bettencourt C, Machado PM, Fox Z, Brady S, Healy E, Parton M, Holton JL, Hilton-Jones D, Shieh PB, Zanoteli E, De Paepe B, De Bleecker J, Shaibani A, Ripolone M, Violano R, Moggio M, Barohn RJ, Dimachkie MM, Mora M, Mantegazza R, Zanotti S, Hanna MG, Houlden H; Muscle Study Group and the International IBM Genetics Consortium. Neurobiol Aging. 2015 Apr;36(4):1766.

4. **Sporadic inclusion body myositis: the genetic contributions to the pathogenesis.**
   Gang Q, Bettencourt C, Machado P, Hanna MG, Houlden H. Orphanet J Rare. 2014 Jun 19;9:88.

5. **Ongoing developments in sporadic inclusion body myositis.**
   Machado PM, Ahmed M, Brady S, ***Gang Q***, Healy E, Morrow JM, Wallace AC, Dewar L, Ramdharry G, Parton M, Holton JL, Houlden H, Greensmith L, Hanna MG. Curr Rheumatol Rep. 2014 Dec;16(12):477.

Published Abstracts

1. **Sporadic inclusion body myositis: genetic risk factors and exome sequencing**
   Gang Q, Machado PM, Brady S, Hilton-Jones D, Houlden H, Hanna M; The International IBM genetics Consortium.
   UK Neuromuscular Translational Research Conference, 2013; UCL Neuroscience Symposium, 2013; UCL Queen Square Symposium, 2013

2. **Using exome sequencing to investigate disease-causing mutations of muscle disorders with protein aggregates**
   Gang Q, Bettencourt C, Machado PM, Brady S, Parton M, Holton JL, Hilton-Jones D, Hanna MG, Houlden H; Muscle Study Group and the International IBM Genetics Consortium.
   UK Neuromuscular Translational Research Conference, 2014; UCL Neuroscience Symposium, 2014; UCL Queen Square Symposium, 2014

3. **Using whole-exome sequencing to identify mutations of *SQSTM1* and *VCP* in inclusion body myositis**
   Gang Q, Bettencourt C, Machado PM, Brady S, Holton JL, Healy E, Parton M, Hilton-Jones D, Hanna MG, Houlden H; Muscle Study Group and the International IBM Genetics Consortium.
   The MRC SAB Meeting, 2014; The 116th Meeting of the British Neuropathological Society, 2015; UK Neuromuscular Translational Research Conference, 2015 (Oral presentation); UCL Neuroscience Symposium, 2015; The Muscle Study Group Meeting, 2015; American Society of Human Genetics, 2015

4. **Analysing whole-exome sequencing data in a large cohort of sporadic inclusion body myositis and control individuals**

   Gang Q, Bettencourt C, Machado PM, Brady S, Holton JL, Hughes D, Healy E, Parton M, Hilton-Jones D, Shieh PB, Zanoteli E, Camargo LV, De Paepe B, De Bleecker J, Shaibani A, Ripolone M, Violano R, Moggio M, Barohn RJ, Dimachkie MM, Mora M, Mantegazza R, Zanotti S, Pittman AM, Singleton AB, Hanna MG, Houlden H; Muscle Study Group and the International IBM Genetics Consortium.

   UK Neuromuscular Translational Research Conference, 2016; UCL Neuroscience Symposium, 2016

The published/accepted papers can be found in the additional CD attached at the end of this thesis.

Chapter 1 Introduction

The development of complex diseases is influenced by multiple genetic and environmental factors, and many causes still remain unclear. Investigations into the genetic risk factors or disease-causing genes are important in improving our knowledge of the pathogenesis of these complex diseases. Understanding such disorders at the genetic level has important clinical utility in the areas of diagnosis, counselling and treatment, as well as identifying novel therapeutic targets and agents that add to our current therapeutic armamentarium. This thesis will explore the genetic landscape of patients with sporadic inclusion body myositis (sIBM) and myopathies with abnormal structures and protein aggregates in muscle.

In this chapter, the categories of myopathies and myositis will first be introduced. Next, an overview of clinical and pathological phenotypes, diagnosis criteria, and possible pathogenesis of sIBM, as well as our current knowledge of its genetic basis, will be discussed. This is followed by an introduction of myopathies with structural abnormalities and protein aggregates, and their clinical and genetic heterogeneity. Myopathies with tubular aggregates and cylindrical spirals will be discussed in detail. Lastly, the conventional and advanced genetic techniques used to investigate genetic causes of diseases will be discussed.

1.1 Phenotypic Spectrum of Myopathies and Myositis

1.1.1 Skeletal Muscle

Skeletal muscle is a complex assembly of interacting proteins, which have the capability to generate force and execute movement of the skeletal components by their contractile properties in response to a stimulus from the motor unit activation (Lunn et al., 2009). The muscle contains bundles of muscle fibres which are called fascicles. Muscle fibres are composed of myofibrils arranged in parallel, therefore the myofibril is the basic unit of muscle. Each myofibril is composed of a number of sarcomeres arranged end-to-end, containing many repeating functional multiprotein complexes of actin, myosin and titin. Thus, the sarcomere is known as the smallest

functional unit of muscle to form its contractile apparatus. “Thin” actin filaments and “thick” myosin filaments interdigitate with each other like cross-bridges. When the thin filaments are pulled towards the centre the sarcomere is shortened, the force of contraction is developed (Wilkie, 1968). (Figure 1-1)

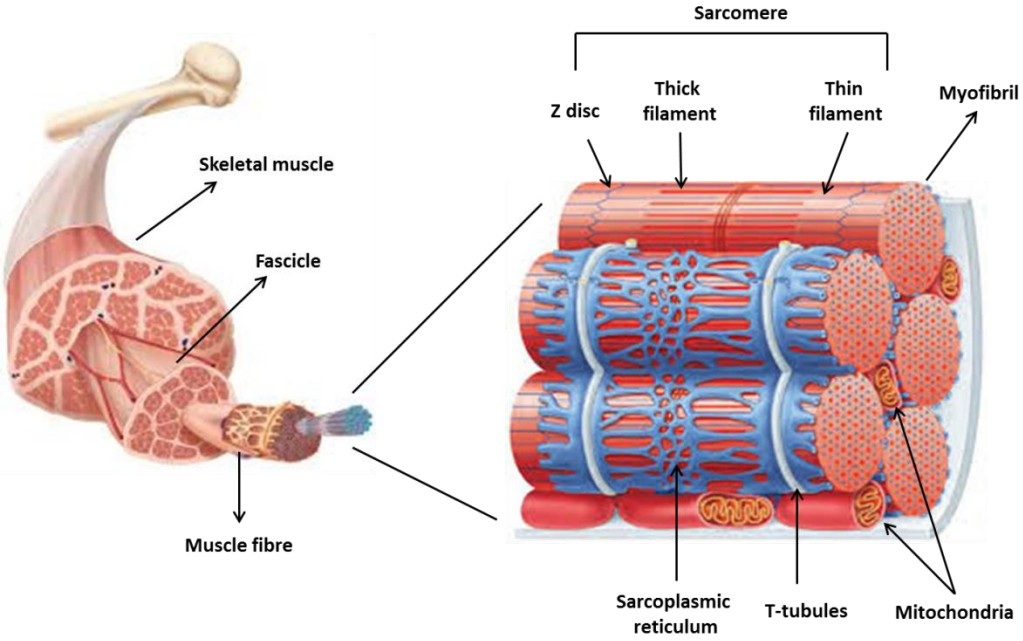


Figure 1-1 Basic structure of skeletal muscle

(Re-produced from www.coursewareobjects.com and www.phschool.com)

When the electrical signal from a motor neuron triggers the muscle fibre to generate force, two networks of tubular structures are involved in the signal transduction: the transverse (T-) tubules and the sarcoplasmic reticulum (SR) (Figure 1-1). The T-tubules allow transmission of the signal into the depths of the muscle fibre to initiate contraction. The SR contains the intracellular storage pool of calcium ions which are released to trigger an energy-dependent interaction between the actin and myosin filaments of the sarcomere to cause the contraction. As the calcium ions are taken back up to the SR and the calcium level falls, the sarcomere extends and then the muscle relaxes. This process is called excitation-contraction coupling. In addition to the actin and myosin, there are also a number of other structural proteins are important in forming a link between myofibrillar components, the plasma membrane and the basal lamina. This is known as the dystrophin-associated protein complex. Various ion channels also span the cell membrane and allow the influx and efflux of sodium, potassium, calcium and chloride, which lead to the generation of resting

potential and of the action potential. The mitochondria are distributed within the muscle fibre to provide adenosine triphosphate (ATP), the energy supply for all energy-dependent processes within the fibre. Muscle is a also highly metabolically active tissue, so it is vulnerable to any abnormalities of energy release. Genetic or acquired defects of any of the proteins or processes above may result in muscle disease (Wilkie, 1968, Lunn et al., 2009).

1.1.2 Classification of Myopathies

Myopathies are a group of muscular disorders in which the most common symptom is muscle weakness due to a primary structural or functional impairment of skeletal muscle fibre, without involving the nervous system (Lunn et al., 2009). Other symptoms can include muscle cramps, stiffness, spasm, fatigue, exercise intolerance and myalgia. Myopathies can be divided into two main categories – inherited myopathies and acquired myopathies. The pattern of muscle weakness, and the absence or presence of a family history of myopathy help to distinguish between the two types. An inherited myopathy often presents an early age of onset with a relatively longer duration of disease compared with an acquired myopathy which is suggested to have a later age of onset with a sudden or subacute presentation. However, some inherited disorders can present in adulthood with variable penetrance within the family (e.g. families with inclusion body myopathy with Paget's disease of the bone and frontotemporal dementia), and some acquired myopathies present in childhood due to genetic susceptibility (e.g. patients with juvenile dermatomyositis). Inherited myopathies can be sub-classified as muscular dystrophies, myotonic syndromes, other channelopathies, myofibrillar myopathies, metobalic myopathies, congenital myopathies, and myosin storage and myosin-related disease. Acquired myopathies can encompass inflammatory myopathies, myopathies with systemic features, and axial myopathy (Visser, 2013). A summary of common inherited and acquired myopathies is provided in Table 1-1.

Table 1-1 Classification of inherited and acquired myopathies (Modified from (Visser, 2013))

Category	Sub-classification	Muscular Disorders
Inherited myopathies	Muscular dystrophies	• Duchenne and Beck muscular dystrophies • Limb-girdle muscular dystrophies including Emery-Dreifuss muscular dystrophies • Facioscapulohumeral dystrophy • Oculopharyngeal muscular dystrophy • Bethlem/Ullrich myopathy • Distal myopathies
	Myotonic syndromes	• Myotonic dystrophies • Myotonia congenita (Thomsen, Becker) • Paramyotonia congenita
	Other ion channel disorders	• Hyperkalemic periodic paralysis • Hypokalemic periodic paralysis
	Myofibrillar myopathies	
	Metabolic myopathies	• Disorders of lipid metabolism • Glycogen storage disorders • Mitochondrial myopathy
	Congenital myopathies	• Central and multicore myopathy • Nemaline myopathies • Centronuclear myopathies
	Myosin storage and myosin-related disease	
Acquired myopathies	Inflammatory myopathies	• Polymyositis • Dermatomyositis • Inclusion body myositis • Nonspecific or overlap myositis • Necrotizing autoimmune myopathy • Macrophagic myofasciitis • Focal myositis • Myositis associated with sarcoidosis • Myotoxic medication
	Myopathies with systemic features	• Myoglobinuria (metabolic, inherited disease, drugs and medication induced) • Amyloid myopathy • Endocrine myopathy • Critical illness polyneuromyopathy
	Axial myopathy	• Dropped head, bent spine syndrome

1.1.3 Idiopathic Inflammatory Myopathies

Idiopathic inflammatory myopathies (IIMs) form a group of rare diseases but represent the most commonly acquired muscle disorders, known as dermatomyositis (DM), polymyositis (PM) and sporadic inclusion body myositis (sIBM) (Holton et al., 2013). The incidence of IIMs estimates varying between studies. In a recent study, an overall annual incidence is suggested to be approximately one in 100,000 (Dimachkie et al., 2014). All three share the main clinical features of proximal lower limb wasting and weakness. However, they may be distinguished by their specific clinical presentations and the findings on investigations. A brief comparison between DM, PM, and sIBM is listed in Table 1-2. IIMs are thought to have an autoimmune aetiology, although they are caused by separate pathological mechanisms. For example, it has been proposed that degeneration may play a key role in sIBM. However, when the typical features and changes are not present, distinguishing one from another condition can be difficult. Treatment of DM and PM mainly relies on immunosuppression, however, sIBM is not responsive to treatment with immune-modulating agents (Holton et al., 2013, Visser and van der Kooi, 2014). In this thesis, sIBM is one of the main focuses and will be discussed in detail in the following sections.

Table 1-2 Comparison of clinical and pathological features of IIMs (Modified from (Holton et al., 2013))

Feature	Dermatomyositis	Polymyositis	Inclusion Body Myositis
Clinical feature			
Age of onset	Peak 30-50 years (juvenile DM peak ≈ 7 years)	> 20 years	> 30 years
Male: female	1:2	1:2	3:1
Skin involvement	Yes	No	No
Pattern of weakness	Proximal, symmetrical	Proximal, symmetrical	Quadriceps, long finger flexors, often asymmetrical
Myalgia	Generalized	Uncommon	Uncommon
Dysphagia	Yes	Yes	Yes
Association with malignancy	Yes (20%)	No	No
Response to immunosuppression	Yes	Yes	No
Creatine kinase	Normal or up to 50x normal	Normal or up to 50x normal	Normal or up to 12x normal
Electromyography	Myopathic	Myopathic	Mixed myopathic and neurogenic
MRI	Muscle edema, predominantly proximal and symmetrical	Muscle edema, predominantly proximal and symmetrical	Fatty infiltration, asymmetrical, quadriceps and distal limbs
Muscle biopsy feature			
Fibre atrophy	Perifascicular	Scattered	Scattered
Fibre necrosis	Predominantly perifascicular	Scattered	Scattered
Fibre regeneration	Predominantly perifascicular	Scattered	Scattered
Rimmed vacuoles	No	No	Yes
Major histocompatibility complex class I up-regulation	Yes	Yes	Yes
T cells	Predominantly $CD4^+$	Predominantly $CD8^+$	Predominantly $CD8^+$
B cells	Yes	Uncommon	Uncommon
Capillary depletion	Predominantly perifascicular	No	No

1.2 Sporadic Inclusion Body Myositis (sIBM)

1.2.1 Prevalence of sIBM

Sporadic inclusion body myositis (sIBM) is the most common form of idiopathic inflammatory myopathy in over 50s (Needham et al., 2008a). A prevalence of 1-71 cases per million has been reported in different populations with a male predominance, increasing to 139 per million among people over 50 years of age (Table 1-3) (Lindberg et al., 1994, Badrising et al., 2000, Kaipiainen-Seppanen and Aho, 1996, Phillips et al., 2000, Felice and North, 2001, Needham et al., 2008a, Wilson et al., 2008, Oflazer et al., 2011, Suzuki et al., 2012, Tan et al., 2013, Molberg and Dobloug, 2016). However, these figures may substantially underestimate the true prevalence of the disease, as it is often misdiagnosed initially (Machado et al., 2013, Cortese et al., 2013, Cox et al., 2011).

Table 1-3 Prevalence of sporadic IBM in different populations (Adapted from (Gang et al., 2014))

Study	Country	Diagnostic criteria	Prevalence (per million population)	Prevalence > 50 years (per million population)
Lingberg et al. 1994	Sweden	Biopsy and clinical data	2.2	N/A
Kaipiainen-Seppanen and Aho 1996	Finland	Not mentioned	0.9	N/A
Philips et al. 2000	Australia	Griggs criteria	9.3	35.3
Badrising et al. 2000	Netherlands	ENMC criteria	4.9	16
Felice and North 2001	Connecticut, USA	Griggs criteria	10.7	28.9 (> 45 yrs)
Needham and Mastaglia 2008	Western Australia	Clinical and biopsy criteria	14.9	51.2
Wilson et al. 2008	Olmsted County, USA	Griggs criteria	71	N/A
Oflazer et al. 2011	Istanbul, Turkey	Own criteria from the study (biopsy + clinical data)	1.0	6.0
Suzuki et al. 2012	Japan	Clinical and biopsy criteria	1.3 (in 1991) 9.8 (in 2003)	N/A N/A
Tan et al. 2013	South Australia	Biopsy and clinical data	50.5	139.3
Molberg et al. 2016	Norway	ENMC criteria	33	N/A

ENMC = European Neuromuscular Centre; N/A = not available.

1.2.2 Clinical and Pathological Phenotypes of sIBM

Clinically, sIBM is characterised by progressive weakness and atrophy of quadriceps femoris and deep finger flexors (Machado et al., 2013, Cortese et al., 2013). Many patients become wheelchair dependent and severely disabled within 10-15 years of symptom onset, although disease progression is heterogeneous (Cortese et al., 2013, Cox et al., 2011). Dysphagia resulting from oesophageal and pharyngeal muscle involvement affects up to 80% of sIBM patients. This can be a significant problem as it predisposes the patient to malnutrition, aspiration and pneumonia (Cortese et al., 2013, Cox et al., 2011, Machado et al., 2013).

Inflammatory and degenerative features are characteristically seen in muscle tissue in sIBM. Inflammatory features include endomysial inflammation, invasion of non-necrotic fibres by inflammatory cells and up-regulation of major histocompatibility complex (MHC) class I molecules (Muntzing et al., 2003). Degenerative features include the formation of rimmed vacuoles, tubulofilaments seen on electron microscopy (EM) and the accumulation of many proteins which may be myotoxic. These include β-amyloid, heat shock proteins (HSPs), p62 and the cytoplasmic mislocalisation of ribonucleic acid (RNA)-binding proteins such as transactive response (TAR) deoxyribonucleic acid (DNA)-binding protein-43 (TDP-43) (Askanas and Engel, 2006, Weihl et al., 2008, Salajegheh et al., 2009, Nogalska et al., 2009), collectively referred to as 'inclusions' (Askanas and Engel, 2001). Ragged-red and cytochrome *c* oxidase (COX) deficient fibres are also seen reflecting mitochondrial changes in sIBM muscle tissue (Oldfors et al., 2006). (Figures 1-2 and 3-4)

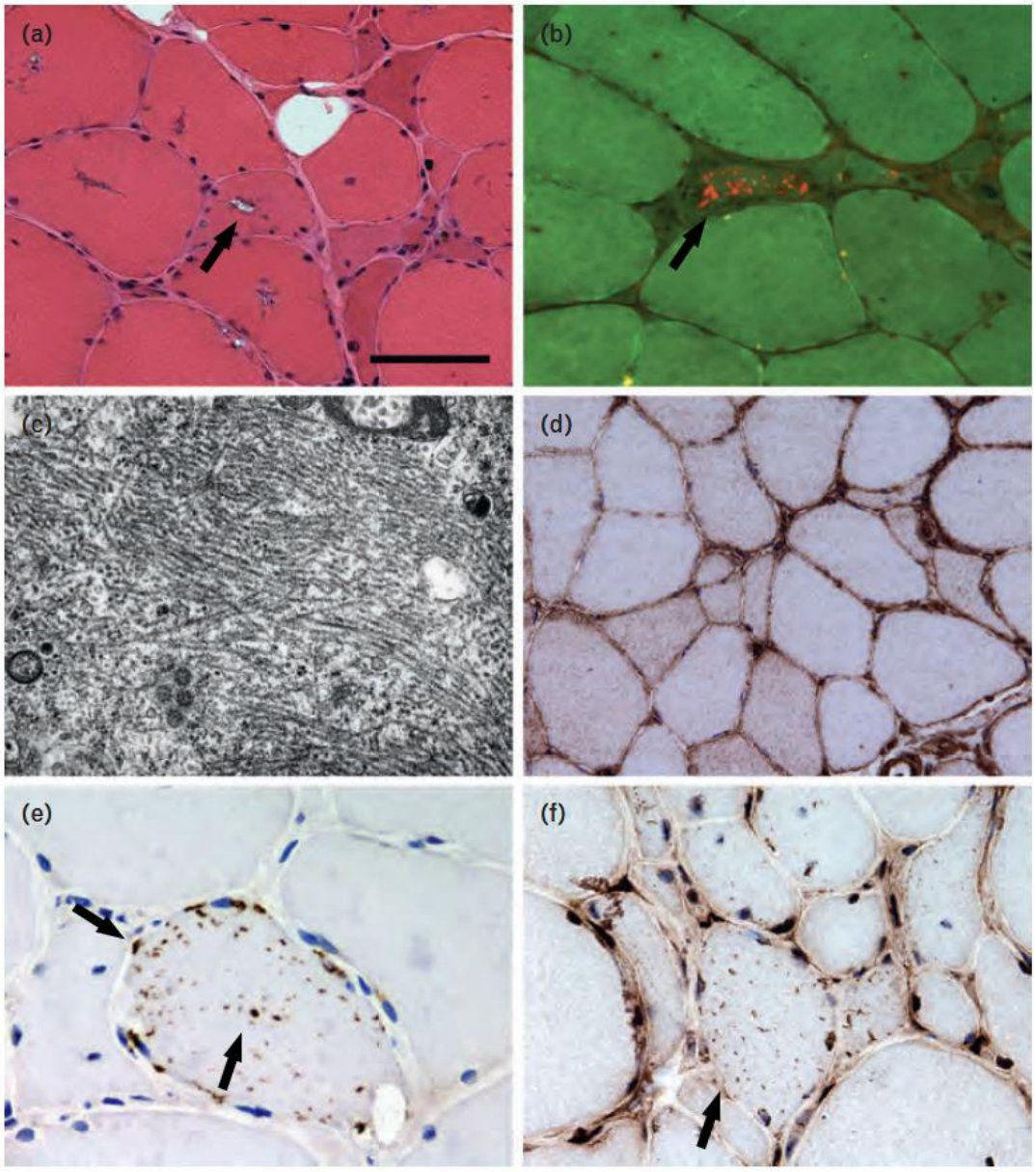


Figure 1-2 Pathological features observed in muscle biopsy of a IBM patient

(a) shows muscle fibres containing rimmed vacuoles (pointed by an arrow); (b) shows amyloid in a tissue section stained by Congo red and visualised under fluorescent light (pointed by an arrow); Tubulofilaments are observed in fibres using electron microscopy (c); Immunohistochemical stainings show increased sarcolemmal and sarcoplasmic MHC Class I expression (d), and fibres containing sarcoplasmic p62 positive aggregates (e), and TDP-43 positive aggregates with loss of normal myonuclear TDP-43 staining (f) (p62 positive and TDP-43 positive aggregates pointed by arrows). Scale bar represents 50 μm in (a), (b), (d), (f); 25 μm in (e); and 0.7 μm in (c). (Adapted from (Machado et al., 2013))

1.2.3 The Diagnosis of sIBM

The clinical diagnosis of sIBM is confirmed by muscle biopsy assisted by clinical presentation, electromyography (EMG), determination of serum muscle-enzyme levels and muscle imaging with magnetic resonance imaging (MRI) or computed tomography (CT) (Dalakas, 2006b). Currently, the Griggs Criteria (Griggs et al., 1995, Tawil and Griggs, 2002), European Neuromuscular Centre (ENMC) Criteria (Verschuuren J J, 1997, Rose, 2013) and Medical Research Council (MRC) Centre for Neuromuscular Diseases Diagnostic Criteria (Benveniste and Hilton-Jones, 2010, Hilton-Jones et al., 2010) are the most widely used diagnosis criteria for IBM. Tables 1-4, 1-5 and 1-6 are the ones used as the inclusion criteria for the IBM genetic studies included in this thesis. Recent studies reported that autoantibodies against cytoplasmic 5'-nucleotidase 1A showed highly specific to sIBM, which could provide for an IBM blood diagnostic test (Larman et al., 2013, Herbert et al., 2015).

Table 1-4 1995 Griggs Criteria for IBM (Adapted from (Griggs et al., 1995))

Criteria Type	Features
A. Clinical features	1. Duration of illness > 6 months 2. Age of onset > 30 years old 3. Muscle weakness - Must affect proximal and distal muscles of arms and legs *and* Patient must exhibit at least one of the following features: a. Finger flexor weakness b. Wrist flexor greater than wrist extensor weakness c. Quadriceps muscle weakness (equal to or less than MRC grade 4)
B. Laboratory features	1. Serum creatine kinase < 12 times normal 2. Muscle biopsy a. Inflammatory myopathy with mononuclear cell invasion of nonnecrotic muscle fibres b. Vacuolated muscle fibres c. Either - Intracellular amyloid deposits - 15 to 18 nm tubulofilaments by electron microscopy 3. Electromyography must be consistent with features of an inflammatory myopathy (long-duration potentials are acceptable)
Definite inclusion body myositis	**Satisfies criteria B2 (muscle biopsy criteria)**
Possible inclusion body myositis	**Satisfies B2a + (A1, 2, 3) + (B1, 3)**

Table 1-5 1997 ENMC Diagnostic Criteria for IBM (Adapted from (Verschuuren J J, 1997))

Criteria Type	Features
Clinical features	1. Presence of muscle weakness 2. Weakness of forearm muscles, particularly finger flexors, or wrist flextors > wrist extensors 3. Slowly progressive course 4. Sporadic disease
Histopathological features	5. Mononuclear inflammatory infiltrate with invasion of non-necrotic muscle fibres 6. Rimmed vacuoles 7. Ultrastructure: tubulofilaments of 16 to 21 nm
Definite inclusion body myositis	**Satisfies 1, 2, 3, 4, 5, 6 or 1, 3, 4, 5, 6, 7**
Probable inclusion body myositis	**Satisfies 1, 2, 3, 4, 5 or 1, 3, 4, 5, 6**

Table 1-6 MRC Centre for Neuromuscular Diseases Diagnostic Criteria for IBM (Adapted from (Benveniste and Hilton-Jones, 2010, Hilton-Jones et al., 2010))

Classification	Features
Pathologically defined inclusion body myositis	Conforming to the Griggs Criteria: Invasion of non-necrotic fibres by mononuclear cells, AND rimmed vacuoles, AND either intracellular amyloid deposits or 15-18 nm filaments
Clinically defined inclusion body myositis	**Clinical features** 1. Duration weakness > 12 months 2. Age > 35 years 3. Weakness of finger flexion > shoulder abduction AND of knee extension > hip flexion **Pathological features** Invasion of non-necrotic fibres by mononuclear cells or rimmed vacuoles or increased MHC-I, but no intracellular amyloid deposit is or 15-18 nm filaments
Possible inclusion body myositis	**Clinical criteria** 1. Duration weakness > 12 months 2. Age > 35 years 3. Weakness of finger flexion > shoulder abduction OR of knee extension > hip flexion **Pathological criteria** Invasion of non-necrotic fibres by mononuclear cells or rimmed vacuoles or increased MHC-I, but no intracellular amyloid deposit is or 15-18 nm filaments

1.2.4 Possible Pathogenesis of sIBM

Despite the first description of sIBM being over 50 years ago, the aetiology of sIBM is still unclear and no effective treatment is available (Greenberg, 2012, Machado et al., 2013). sIBM is a complex disease. The primary pathogenic mechanism is still controversial and it is believed that many factors including genetic and environmental factors contribute to the disease process.

Currently there are two popular theories for the pathogenesis of sIBM – an autoimmune pathway and a degeneration pathway. Ageing is also considered to be an important factor in contributing to both degeneration and mitochondrial

abnormalities seen in sIBM (Figure 1-3). All three are discussed in detail in the following sections.

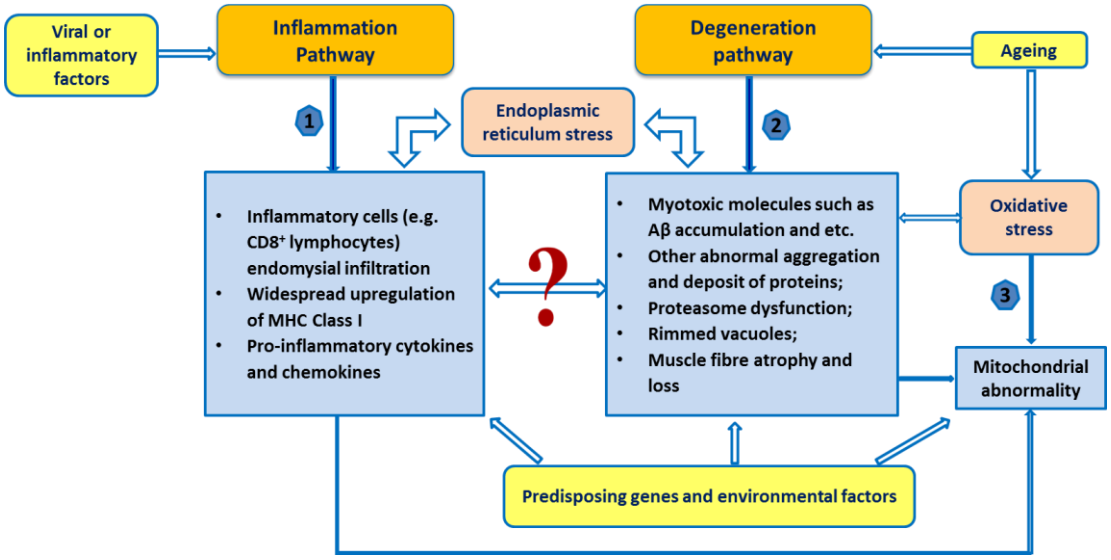


Figure 1-3 Possible pathogenesis of sIBM

This diagraph shows multifactorial events that are proposed to be involved in the pathogenesis of sIBM including genetic and environmental factors. Inflammation and degeneration pathways are two main theories. Ageing is also considered as to be an important factor. (Modified from (Gang et al., 2014))

1.2.4.1 Immune-Related Pathway

The inflammatory pathway theory proposes that viral or inflammatory triggering factors lead to the clonal expansion of $CD8^+$ T cells and T cell-mediated cytotoxicity, which ultimately result in damage or death of muscle fibres (Dalakas, 2008). This theory is supported by the increased prevalence of sIBM in patients with autoimmune disorders and human immunodeficiency virus (HIV) and human T-lymphotropic virus-1 (HTLV-1) infection (Dalakas, 2006a). There is increased messenger RNA (mRNA) expression of a range of cytokines, including interleukin-1 (IL-1), tumour necrosis factor alpha (TNF-α) and interferon gamma (IFN-γ), as well as increased expression of chemokines and chemokine receptors in sIBM muscle, including chemokine ligand 9 (CXCL-9), chemokine ligand 3 (CCL-3), chemokine ligand 4 (CCL-4), chemokine ligand 10 (CXCL-10), chemokine ligand 2 (CCL-2) and chemokine receptor 3 (CXCR-3) (Schmidt et al., 2008). A further study observed that sIBM patients presented a systemic immune activation with Th1 polarization involving IFN-γ pathway and $CD8^+CD28^-$ T cells associated with peripheral regulatory T cells (Allenbach et al., 2014). This further supports the above theory.

These released proinflammatory mediators induce abundant expression of MHC Class I molecules, which exert a stressor effect on endoplasmic reticulum (ER). Activated ER stress triggers an up-regulation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), a transcription factor which promotes further cytokine expression. This leads to a self-sustaining T-cell response (Dalakas, 2006b). In addition, evidence has shown that intramuscular cytokines and chemokines contribute not only to the overexpression of β-amyloid precursor protein (AβPP) but also the overproduction of β-amyloid protein (Aβ) and phosphorylated tau, as well as other cell-stress and degeneration-associated molecules such as ubiquitin and heat-shock protein alpha-crystallin B chain (αB-crystallin) by myofibres (Dalakas, 2008, Kitazawa et al., 2008, Schmidt et al., 2008). In Schmidt et al.'s study there was a linear relationship between the mRNA level of cytokines and chemokines, and that of AβPP expression (Schmidt et al., 2008). An IBM-transgenic mouse model study conducted by Kitazawa et al. supports such interaction between inflammation and sIBM related degenerative proteins. It was found that acute and chronic intramuscular inflammation, induced by the injection of lipopolysaccharide, increased the levels of AβPP, Aβ and phosphorylated tau in muscle mediated by glycogen synthase kinase-3beta (GSK3β) activation (Kitazawa et al., 2008). Therefore, degenerative changes and ER stress are considered as secondary mechanisms induced by increased intracellular cytokines and chemokines (Dalakas, 2008).

The idea of primary inflammatory reaction in sIBM pathogenic process has directed several clinical trials in immunotherapies. However, sIBM is poorly responsive to even vigorous immunosuppression. Even where there is histopathological evidence of reduced inflammation, this does not correlate with clinical improvement (Barohn et al., 1995). In addition, up-regulation of MHC Class I molecules and ER stress of muscle are not specific to sIBM, and can also be seen in other inflammatory myopathies. This suggests that other mechanisms may contribute to the muscle damage in sIBM. Therefore, some investigators are in favour of a degenerative process as the primary pathogenesis of sIBM.

1.2.4.2 Degeneration-Related Pathway

Identification of aberrant protein aggregates in sIBM vacuolated muscle fibres has demonstrated remarkable parallels with those features seen in brain tissues in Alzheimer's disease (AD) and Parkinson's disease (PD) with Lewy bodies. This theory suggests that several complex processes are primarily involved in the pathogenesis of sIBM as discussed below, and inflammation is secondary to the degeneration-associated changes in sIBM muscle fibres (Askanas and Engel, 2005a).

First, aggregates of multiple proteins are present in sIBM muscle fibres, such as plaque-like Aβ and phosphorylated tau (tau is a highly soluble microtubule-associated protein), α-synuclein and cellular prion proteins. Askanas and Engel proposed that increased expression of AβPP, Aβ and possible β-amyloid peptide oligomers may be an early upstream event leading to intracellular abnormal signal-transduction and transcription (Askanas and Engel, 2005a). In addition, a mechanism for abnormal formation of AβPP was demonstrated that results in increased generation of the toxic Aβ42 fragments. For instance, beta-secretase 1 (BACE1) and beta-secretase 2 (BACE2) (transmembrane β-secretases cleaving AβPP at the N-terminal of Aβ), and nicastrin and presenilins (γ-secretases cleaving AβPP at the C-terminal of Aβ), are increased in sIBM muscle fibres and accumulate in aggregates colocalising with Aβ (Nogalska et al., 2010d, Askanas et al., 2009). Amyloid oligomers and overexpressed AβPP have also been found to cause mitochondrial dysfunction *in vitro* (Askanas et al., 1996b), and result in calcium dyshomeostasis *in vivo* (Moussa et al., 2006). These can have detrimental effects on the muscle fibres.

Protein aggregation may be also caused by proteins unfolding or misfolding, thus sIBM is also considered as a 'conformational disorder'. HSPs, endogenous chaperone proteins, prevent aberrant protein-protein interactions and promote correct protein folding (Brown, 2007). HSPs are up-regulated as an endogenous cytoprotective mechanism when stress-induced activation of the heat shock response occurs (Kalmar and Greensmith, 2009). Accumulation of HSPs in sIBM muscle also provides evidence for protein mishandling (Cacciottolo et al., 2013a). The formation of misfolded proteins is suggested through inhibition of the 'ubiquitin-proteasome system' (UPS) (Askanas and Engel, 2005b). The mutated ubiquitin (UBB^{+1}) inhibits the proteasome, and its accumulation was proposed to be a marker for proteasomal

dysfunction in the brain (van Leeuwen et al., 2006). Accumulation of mutated ubiquitin (UBB^{+1}) in muscle fibres of sIBM suggests a similar mechanism in muscle under an ageing cellular environment combined with other factors such as oxidative stress (Fratta et al., 2004).

In addition, the autophagy-lysosome system is another major proteolytic pathway of the cell for removing and eliminating unfolded and toxic proteins as well as abnormal and dysfunctional organelles in skeletal muscle (Sandri, 2010). There is increasing evidence suggesting that impairment of this pathway could be responsible for the formation of rimmed vacuoles associated with the accumulation of misfolded proteins such as Aβ, α-synuclein, and BACE in sIBM patient muscle fibres in a similar way to alterations during neurodegeneration (Wong and Cuervo, 2010). Increased level of expression of autophagic markers p62, microtubule-associated protein 1A/1B-light chain 3 (LC3), beclin 1, autophagy protein 5 (ATG5), neighbour of BRCA1 gene 1 protein (NBR1) and clathrin in sIBM myofibres also supports the above theory (Hiniker et al., 2013, Girolamo et al., 2013, Nogalska et al., 2009, Lunemann et al., 2007, Nogalska et al., 2010c, D'Agostino et al., 2011). Impaired autophagic activity could be due to blockade of downstream autophagosome-lysosome fusion and lysosomal activity (Girolamo et al., 2013). A study indicated endoplasmic reticulum stress in causing impaired lysosomal degradation (Nogalska et al., 2010c), however the mechanism occurring in sIBM muscle fibres is still unknown.

Secondly, abnormal accumulation of lipoprotein receptors and free cholesterol was found in sIBM muscle fibres (Jaworska-Wilczynska et al., 2002). Co-localisation of these increased lipoprotein receptors in sIBM inclusions with abnormally accumulated Aβ, apolipoprotein E (ApoE) and free cholesterol, suggests that cholesterol may increase β-amyloid production and further induce Aβ misfolding and aggregation. The mechanism of this process is not known. One possibility could be that the accumulation of cholesterol might be due to perturbed trafficking of cholesterol in sIBM muscle fibres, and be further involved in direct or indirect inhibition of AβPP α-secretase (Askanas and Engel, 2003).

Thirdly, free-radical toxicity may participate in the pathogenesis of sIBM. Indicators of oxidative stress accumulate in sIBM muscle fibres, such as nitric oxide synthase (NOS) leading to very toxic peroxynitrite (Yang et al., 1996). This could potentiate cascade of abnormal protein accumulation in the muscle.

In addition, endoplasmic reticulum stress induced by oxidative stress and the accumulation of unfolded proteins in the ER might have a detrimental effect on the muscle fibres through increased activation of NF-κB and decreased sirtuin 1 deacetylase activity (Nogalska et al., 2007, Nogalska et al., 2010a). These further increase AβPP transcription. Therefore, Aβ accumulation may create a self-perpetuating pathogenic cycle.

Furthermore, the finding of remnants of the myonuclear membrane molecules and a single-stranded DNA-binding protein in or near the rimmed vacuoles supports the hypothesis that myonuclear disintegration leads to the formation of rimmed vacuoles. This results in a progressive reduction of the number of myonuclei and further progressive muscle atrophy (Karpati and O'Ferrall, 2009).

1.2.4.3 Ageing and Mitochondrial Abnormalities

The milieu of ageing muscle fibres could be another important factor in the development of sIBM because the disease becomes more prevalent in people over 50 years of age. There may be diminished cellular defence mechanisms due to changes associated with ageing in muscle, resulting in muscle fibre vacuolar-degeneration, atrophy and cell death (Askanas et al., 2012).

Mitochondrial abnormalities in sIBM muscle might be related not only to the ageing changes, abnormal AβPP processing and the effects of proinflammatory cytokines, but also to the subsequent oxidative stress event (Askanas et al., 2012). It has been postulated that mutations are likely to occur due to mitochondrial DNA (mtDNA) damage induced by oxidative stress (Oldfors et al., 2006). MtDNA mutations and associated respiratory deficiency may cause the muscle atrophy and weakness seen in sIBM patients. These opinions support the suggestion that mitochondrial abnormalities are secondary to the cellular changes in sIBM development rather than being a primary pathogenic contributor. Interestingly, a recent study found

mitochondrial injury not only in muscle but also in peripheral blood mononuclear cells of sIBM patients. Although the altered mitochondrial phenotype presents a different lesion profile in both tissues, COX enzymatic deficiency is a common pathogenic parallelism (Catalan-Garcia et al., 2016). The authors suggested that this was probably due to different compensatory mechanisms related to the disease. However, they did not found increased levels of lipid peroxidation (an indicator of oxidative damage produced by reactive oxygen species in cellular lipid compounds). The study further supports the involvement of mitochondria in the aetiology of sIBM. Whether oxidative stress play a systemic role in sIBM and whether COX activity could be used as a biomarker of the disease should be assessed in further studies.

1.2.5 Genetic Contributions to sIBM

Despite sIBM not being a Mendelian inherited disease, multiple genetic risk factors are being shown to play important roles in the development and progression of sIBM. Furthermore, the prevalence of sIBM differs between different ethnic groups. This is likely due to differences in genetic makeup and differences in environmental factors of different geographical regions (Needham et al., 2007). Current reported genetic studies on sIBM primarily used candidate gene analysis approach on the basis of clinico-pathological features of sIBM, as well as familial cases or inherited diseases with features resembling sIBM (Gang et al., 2015).

1.2.5.1 Immune-Related Genetic Factors

Garlepp and colleagues first reported the strong association between human leukocyte antigen (*HLA*)-*DR3* with the extended 8.1 ancestral haplotype (AH) and sIBM (Garlepp et al., 1994). This association was subsequently confirmed in a series of further studies in Caucasians with *HLA-A*01, -B*0801, -DRB1*0301, -DRB1*0101, -DRB1*1301, -DQB1*0201,* and *-DQA1*05* (Badrising et al., 2004, Koffman et al., 1998, Lampe et al., 2003, Mastaglia et al., 2009, Needham et al., 2008c, Price et al., 2004, O'Hanlon et al., 2005, Rojana-udomsart et al., 2012). Furthermore, Rojana-udomsart and colleagues reported that the risk factors *HLA-DRB1*0301* and **1301* also influence the age of onset and the severity of muscle weakness (Rojana-udomsart et al., 2012). These results indicate that interactions at the *HLA-DRB1* locus may contribute to both disease susceptibility and clinical

phenotype. The *HLA-DRB1* locus encodes the beta chains of the Class II HLA-DR molecules which present antigenic peptides to $CD4^+$ T cells. This provides further support for the involvement of cellular immune mechanisms in the pathogenesis of sIBM. Other alleles and haplotypes have also been associated with increased risk of sIBM, including the 35.2AH (*HLA-B*3501, -DRB1*0101*) in Caucasians (Price et al., 2004) and the 52.1AH (*HLA-B*5201, -DRB1*1502*) in Japanese populations (Scott et al., 2006). Interestingly, a number of *HLA* alleles and haplotypes, including *HLA-DR53* and *HLA-DRB1*0401, *0701, * 0901, *1101, *1501, HLA-DRB1*0301/*0401* and *HLA-DRB1*0301/*0701* diplotypes, and *HLA-DRB1*1501*, were found to be protective against sIBM (Rojana-udomsart et al., 2012, Badrising et al., 2004, Rojana-udomsart et al., 2013).

A potential region of MHC-associated susceptibility genes was first defined between *HLA-DR* and *C4* (Kok et al., 1999). This region was further suggested in the border of the Class II and III regions, between *HLA-DRB1* and pre-B-cell leukemia homeobox2 (*PBX2*, also known as *HOX12*) loci (Price et al., 2004). A further recombination mapping was applied consisting of three separate cohorts to refine the probably 8.1 AH-derived sIBM susceptibility region from butyrophilin-like MHC class II-associated gene (*BTNL2*) to telomeric gene of *HLA-DRB1*. This includes three protein-coding genes, *BTNL2, HLA-DRA* and *HLA-DRB3* (Scott et al., 2011). Other genes in the above regions have also been considered as candidate genes, such as notch 4 gene (*NOTCH4*; a transmembrane receptor which regulates cell fate decisions), advanced glycosylation end product-specific receptor gene (*AGER*, previously known as *RAGE*), testis-specific basic protein gene (*C6orf10*, also known as *TSBP*), and G-protein signalling modulator 3 gene (*GPSM3*, previously known as *G18*) (Kok et al., 1999, Price et al., 2004).

Two *NOTCH4* gene polymorphisms (rs422951 and rs72555375) have shown a strong association (odds ratio [OR] > 2) with sIBM in two independent Caucasian cohorts (Scott et al., 2012). These variants (or other variants in linkage disequilibrium with them) may play a role in sIBM pathogenesis by generating antigenic peptides and can be regarded as markers for sIBM susceptibility. They may also influence the binding and affinity of the notch 4 receptor, and the presentation of the gene product at the cell surface, respectively (Scott et al., 2012).

The MHC contains important genetic data which have been reported in association with sIBM. Among these 8.1 AH is predominant in Caucasians, supporting an immune-mediated mechanism underlying sIBM. Identification of the actual genes within the MHC will be crucial to understanding the role of the MHC in the pathogenesis of sIBM.

In addition to the genes within the MHC regions, other genes have also been studied. Mutations in the three-prime repair exonuclease 1 (*TREX1*) gene are associated with a spectrum of autoimmune diseases, such as systemic lupus erythematosus and Aicardi-Goutiere (Mezei et al., 1999). Regarding the theory supporting the idea that sIBM is primarily immune-mediated disease, Cox et al. tested 54 patients with sIBM for *TREX1* mutations by direct sequencing, but no pathogenic mutations were found (Cox et al., 2010). However, this cannot rule out a role for rare *TREX1* mutations in sIBM. Further studies should aim to investigate whether *TREX1* accumulation is a secondary effect rather than playing an important pathological role in sIBM (Cox et al., 2010).

1.2.5.2 Degeneration-Related Genetic Factors

Genes encoding proteins deposited in sIBM muscle fibres are another group of important candidate genes. Among them, apolipoprotein E (*APOE*) gene, β-amyloid precursor protein (*APP*) gene, microtubule-associated protein tau (*MAPT*) gene, prion protein (*PRNP*) gene, p62 (*SQSTM1*) gene and the expanded *C9orf72* hexanucleotide repeat were studied in this thesis, and will be discussed in detail in Chapter 3. In addition, other genes that may contribute to the genetic architecture of sIBM are discussed below.

Presenilin (*PSEN*)

Abnormal accumulation of presenilin protein occurs in the brains of both sporadic and familial AD patients (Levey et al., 1997). Autosomal dominant inheritance of mutations in presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*) genes, on chromosome 14 and chromosome 1 respectively, were found to be responsible for early-onset familial AD and also for late-onset AD (LOAD) (Sherrington et al., 1995, Gerrish et al., 2012). The exact roles of presenilin are still not clear, however it has been considered that mutations in presenilin genes cause early onset familial AD by

increasing synthesis of cytotoxic Aβ42 (Scheuner et al., 1996). Abnormal accumulation of presenilin 1 protein was first reported immunohistochemically in muscle fibres in both sIBM and autosomal recessive hereditary IBM (hIBM2) by Askanas et al. (Askanas et al., 1998). However, they did not find increased presenilin 1 mRNA in IBM abnormal muscle fibres. This indicates that there might be some similarities shared between AD and IBM in the pathogenic mechanism of presenilin deposits, however, the mutations in *PSEN1* and *PSEN2* have not been investigated in both sIBM and hIBM cases.

Dysferlin (*DYSF*)

Dysferlin is a modular type II transmembrane protein newly identified as a binding partner for AβPP and co-aggregated with Aβ42 in sIBM muscle fibres (Cacciottolo et al., 2013b). In normal human muscle, dysferlin is immunohistochemically present in the muscle fibre sarcolemma and involved in plasmalemmal repair after injury as well as in T-tubule contraction and calcium homeostasis. Mutations in *DYSF* gene are known to cause a range of autosomal recessive myopathies, called dysferlinopathies including Miyoshi myopathy and limb girdle muscular dystrophy type 2B (LGMD2B) (Barthelemy et al., 2011). Immunohistochemically similar to dysferlinopathies, dysferlin is prominently reduced in the muscle fibre sarcolemma in sIBM muscle biopsies. In addition, there is abnormal distribution of dysferlin into the cytoplasmic protein aggregates which co-localise with Aβ42 aggregates. Accordingly, it has been suggested that dysferlin might link to the pathogenesis of sIBM, although the possible mechanisms of this phenomenon is unknown (Cacciottolo et al., 2013b). Thereby, *DYSF* gene should be investigated as a candidate gene of sIBM in the future.

Alpha-1-Antichymotyrpsin (*SERPINA3*)

Alpha-1-antichymotyrpsin (α1ACT; *SERPINA3*, previously known as *ACT*) is considered as a major acute-phase protein and belongs to the serpin superfamily specifically inhibiting serine proteases (Law et al., 2006). Abnormal accumulation of α1ACT in sIBM muscle was first described by Bilak and colleagues. Similarly to AD brains, α1ACT is located closely with Aβ containing ubiquitinated amyloid plaques (Bilak et al., 1993). The genotype of *SERPINA3/AA* (*ACT* -15A/T polymorphism in signal peptide [rs4934]) is also found to alter the AD risk associated with *APOE*

ε4/ε4 genotype (Kamboh et al., 1995). However in a study (Gossrau et al., 2004) analysing *SERPINA3* and *APOE* gene polymorphisms in 35 sIBM patients, there was no significant correlation between distinct *SERPINA3* and *APOE* genotypes and the risk of developing sIBM. It could also be due to the small size of the patient cohort, of which *SERPINA3/AA* genotype was only present in seven cases and *APOE ε4/ε4* genotype was not present in this cohort. Studies on larger numbers of cases are required to verify whether the observation of α1ACT protein deposits in affected muscles is an epiphenomenon.

TAR DNA-Binding Protein-43 (*TARDBP*)

TAR DNA-binding protein-43 (TDP-43) is found as a major component of ubiquitinated inclusions in amyotrophic lateral sclerosis (ALS) and tau-negative frontotemporal lobar dementia (FTLD-U) (Arai et al., 2006, Neumann et al., 2006). Mutations in *TARDBP* have also been detected in both sporadic and familial forms of ALS (Kabashi et al., 2008, Sreedharan et al., 2008). Several studies (Hernandez Lain et al., 2011, Olive et al., 2009, Salajegheh et al., 2009, Weihl et al., 2008) have examined sarcoplasmic TDP-43 inclusions in sIBM muscle fibres, which contrasts to normal tissue since TDP-43 immunostaining predominates in the nucleus in normal tissue. Similar TDP-43 deposits were also identified in inclusion body myopathy with early-onset Paget's disease of the bone and frontotemporal dementia (IBMPFD), hIBM2 and other myofibrillar myopathies (MFMs) (Olive et al., 2009). These changes are similar to those seen in brain tissue of ALS and FTLD-U, indicating that sIBM may share some pathological pathways with these two diseases. TDP-43 could potentially interfere with the normal functions of regulating gene expression at the transcriptional level, regulating gene splicing and stabilising extranuclear RNAs that maintain myofibre protein production (Cortese et al., 2014). In a study by Salajegheh et al. (Salajegheh et al., 2009), six sIBM patients samples were sequenced for the *TARDBP* gene. However, no exonic mutation was found.

Fused in Sarcoma (*FUS*) Protein

The Fused in Sarcoma (*FUS*) protein shares functional and structural homology with TDP-43 (Lagier-Tourenne and Cleveland, 2009). A number of discovered mutations in *FUS* gene enhanced the prominence of gene expression-mediating proteins in ALS pathogenesis (Lattante et al., 2013). Truncated TDP-43 fragments, decreased normal

forms of FUS protein and increased FUS fragments were demonstrated in sIBM, though FUS protein is suggested not to be a major component of sarcoplasmic sIBM inclusions. (Hernandez Lain et al., 2011) Mutations in *FUS* have not been investigated in sIBM cases. Whether these abnormal FUS fragments may play a pathogenic role in sIBM still needs to be clarified.

Other Heterogeneous Nuclear Ribonucleoproteins (*hnRNPs*)

Abnormal distribution of other heterogeneous nuclear ribonucleoproteins, such as hnRNPA1 and hnRNPA2B1, has been observed as a shared feature between sIBM and other degenerative diseases such as ALS, frontotemporal dementia (FTD), and IBMPFD (Kim et al., 2013, Cortese et al., 2014, Pinkus et al., 2014). Similarly to TDP-43, these proteins were also depleted in myonuclei and formed part of sarcoplasmic aggregates in sIBM muscle (Pinkus et al., 2014). However, the expression levels of these *hnRNPs* were not significantly altered when sIBM muscle was compared with normal muscle (Pinkus et al., 2014). Mutations in both *hnRNPA1* and *hnRNPA2B1* genes have been reported in pedigrees with multisystem proteinopathy (MSP), which is a rare inherited syndrome presenting with features of IBM, FTD, ALS and PDB (Kim et al., 2013). However, a recent study in a Dutch population (including 31 sIBM patients) did not identify any mutation in these two genes (Seelen et al., 2014). The toxicity caused by disruption of RNA homeostasis was suggested to contribute to myofibre injury in sIBM due to the sarcoplasmic aggregation of hnRNPs (Pinkus et al., 2014). However, there is still lack of evidence of the genetic contribution of these hnRNPs to the disease pathogenesis. Genetic screening in a larger cohort of sIBM patients is required.

Cytoplasmic 5'-Nucleotidase 1A (*NT5C1A*)

Larman and colleagues identified cytoplasmic 5'-nucleotidase 1A (cN1A; *NT5C1A*) accumulation by immunohistochemistry in perinuclear regions and rimmed vacuoles in both sIBM and fIBM muscle fibres, indicating a potential contribution of cN1A protein during myonuclear degradation (Larman et al., 2013). In this study they also sequenced the coding regions of both *NT5C1A* and *NT5C1B* genes, but no protein-altering variants were present in their patient group (Larman et al., 2013).

1.2.5.3 Mitochondrial DNA and Nuclear Genes Encoding Mitochondrial Components

There is increasing evidence suggesting a pathogenic role for altered mitochondrial function in sIBM. COX deficient muscle fibres are a common histopathological feature of patients with sIBM (Brady et al., 2014), and they are associated with single/multiple deletions in mtDNA (Oldfors et al., 1995, Santorelli et al., 1996, Kok et al., 2000, Oldfors et al., 2006, Rygiel et al., 2016). A recent study also reported several exonic variants in nuclear genes encoding mitochondrial components in a group of 16 sIBM patients (Rygiel et al., 2015). An analysis of mtDNA deletion(s) in a group of sIBM was also included in this thesis discussed in Chapter 3 in further detail.

1.2.5.4 Genes Associated with sIBM-Like Disorders

Familial cases with IBM (fIBM) are clinically and histologically similar to sIBM, having been reported in two or more affected siblings in the same family (Sivakumar et al., 1997, Tateyama et al., 2003, Ranque-Francois et al., 2005), and also in affected twins (Amato and Shebert, 1998). Among these cases, a strong association with the *HLA* class II *DR3* allele (*DRB*0301/0302*) was reported in four Caucasian families (Sivakumar et al., 1997, Ranque-Francois et al., 2005), and an association with the *DR15(2)/4* allele (*DRB1*1502/0405*) was reported in two Japanese sisters (Tateyama et al., 2003). These associations suggest that *HLA* class II genotypes may be genetic predisposing factors for fIBM and that associations may differ between ethnic groups.

Hereditary inclusion body myopathies (hIBMs) are a group of clinically heterogeneous muscle disorders with autosomal recessive or autosomal dominant inheritance. They share some pathological features with sIBM. The use of the term "myopathy" instead of "myositis" is due to the earlier age of onset, the rarity of lymphocytic inflammation and the absence of MHC class I up-regulation (Broccolini and Mirabella, 2015). There are three main subtypes of hIBMs that are summarised in Table 1-7 with their associated genes.

Table 1-7 Subtypes of hereditary inclusion body myopathies (Adapted from (Gang et al., 2015))

hIBM subtype	Estimated prevalence	Disease OMIM#	Disease inheritance	Gene	Common mutations
GNE myopathy (Distal myopathy with rimmed vacuoles [DMRV]/ Nonaka myopathy) – hIBM2	About 1/1,000,000	600737	Autosomal recessive	*GNE*	Homozygous mutation p.Met712Thr in all Middle Eastern patients Homozygous mutation p.Val572Leu in the majority of Japanese patients
hIBM with Paget's disease of the bone and frontotemporal dementia (IBMPFD)	< 1/1,000,000	167320	Autosomal dominant	*VCP*	Hot spot missense mutations at codon 155
hIBM with congenital joint contractures and external ophthalmoplegia – hIBM3	< 1/1,000,000	605637	Autosomal dominant	*MYH2*/ *MYHC-IIa*	Heterozygous missense mutation p.Glu706Lys in a Swedish family

OMIM = online Mendelian inheritance in man; *GNE* = glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase; *VCP* = valosin containing protein; *MYH2/MYHC-IIa* = myosin heavy chain 2/ myosin heavy chain IIa.

Various other neuromuscular disorders showing rimmed vacuoles in the muscle fibres are important in the differential diagnosis of sIBM. Genes associated with these "rimmed vacuolar myopathies" could also be considered as candidate genes for sIBM genetic studies (Gang et al., 2015) (Table 1-8).

Table 1-8 Rimmed vacuolar myopathies with features resembling sIBM and their associated genes (Adapted from (Gang et al., 2015))

Rimmed vacuolar myopathy	Disease OMIM#	Disease inheritance	Gene
A leukoencephalopathy and a vacuolar myopathy resembling IBM	N/A	Autosomal recessive	Laminin alpha 2 (*LAMA2*) (Di Blasi et al., 2000)
Oculopharyngeal muscular dystrophy	164300	Autosomal dominant	Poly(A)-binding protein-2 (*PABPN1*)
Emery-Dreifuss muscular dystrophy	310300	X-linked recessive	Emerin (*EMD*)
Limb girdle muscular dystrophies	159000 (LGMD1A)	Autosomal dominant	Myotilin (*MYOT*)
	601954 (LGMD2G)	Autosomal recessive	Telethonin (*TCAP*)
	609115 (LGMD1G)	Autosomal dominant	Heterogeneous nuclear ribonucleoprotein D-like (*HNRNPDL*)
Rigid spine syndrome	602771	Autosomal recessive	Selenoprotein N,1 (*SEPN1*)
Myofibrillar myopathies	601419 (MFM1, initially called hIBM1)	Autosomal dominant or recessive	Desmin (*DES*)
	608810 (MFM2)	Autosomal dominant	Alpha-B-crystallin (*CRYAB*)
	609200 (MFM3)	Autosomal dominant	Myotilin (*MYOT*)
	609452 (MFM4)	Autosomal dominant	LIM domain-binding 3 (*LDB*, previously known as *ZASP*)
	609524 (MFM5)	Autosomal dominant	Filamin C (*FLNC*)
	612954 (MFM6)	Autosomal dominant	BCL2-associated athanogene 3 (*BAG3*)

OMIM = online Mendelian inheritance in man; LGMD = limb girdle muscular dystrophy; MFM = myofibrillar myopathy; N/A = not available.

1.2.6 Animal Models of sIBM

Animal models are important tools to help with clarification of the pathogenic mechanisms of human diseases and to explore promising novel therapies. sIBM is

considered as a complex multigenic disorder, so a single transgenic model would not be sufficient to explore the aetiology of the disease. Several animal models of sIBM have previously been reported (Katsumata and Ascherman, 2008). First, transgenic mouse models of sIBM overexpressing AβPP in affected muscle fibres have been reported in several studies (Fukuchi et al., 1998, Sugarman et al., 2002, Sugarman et al., 2006, Kitazawa et al., 2008, Kitazawa et al., 2009). Such models have re-capitulated pathological features seen in sIBM, including vacuolation, increased protein accumulation, centric nuclei, and inflammation in muscle fibres. These transgenic mice have also developed muscle weakness, and deficiency in motor performance. In an alternative transgenic model of IBM, reported by Delaunay et al., the transgenic mice carry the RING finder ubiquitin ligase (*RNF5*) under the control of beta-actin or muscle specific promoters. These mice exhibit an early onset of muscle wasting, degeneration, and extensive fibre regeneration (Delaunay et al., 2008). Rabbits fed with cholesterol-enriched diets were proposed as a new model by Chen and colleagues. In these animals, they identified features of IBM including vacuolation in muscle fibres, increased number of mononuclear inflammatory cells, and increased protein deposition (Chen et al., 2008). It was also recently used to study the pathogenic overlap between LOAD and sIBM (Liu et al., 2016), as many neuropathological changes of LOAD can be reproduced in this model. In addition, a mutant valosin-containing protein (VCP) mouse was also studied recently, a model of IBMPFD. These animals developed progressive muscle weakness, and also classic pathological characteristics of IBM including rimmed vacuoles, inflammation, and TDP-43 pathology (Ahmed et al., 2016, Custer et al., 2010).

1.3 Myopathies with Specific/Rare Structure Abnormalities and with Protein Aggregates

Myopathies can also be grouped by distinct pathological features observed in muscle biopsies. These features in muscle fibres can be specific/rare abnormal structures or abnormal accumulated proteins. A variety of structural defects have been identified in muscle, and some of them are characteristic of particular disorders. The disorders are usually named after such features, such as hyaline body myopathy, reducing body myopathy, spheroid body myopathy, Zebra body myopathy, tubular aggregate myopathy, cylindrical spirals myopathy, crystalloid inclusion myopathy, fingerprint

body myopathy, nemaline myopathies, core/minicore myopathies, and centronuclear myopathies (Ravenscroft et al., 2014, Goebel et al., 2013). These muscle disorders can be congenital, acquired, familial or sporadic, so their clinical presentations are highly heterogeneous, and sometimes lack genetic clarification. The precise mechanisms underlying these conditions also remain uncertain.

Protein aggregate myopathies (PAMs) are a group of neuromuscular conditions with characteristic pathomorphological aggregation of proteins within muscle fibres. They can be sarcoplasmic and/or nuclear protein aggregates (Schroder, 2013). The clinical onset of PAMs varies from childhood to late adulthood, and they usually present a progressive course leading to severe disability and premature death (Schroder, 2013). PAMs can be divided into acquired and hereditary forms with autosomal dominant, autosomal recessive, or X-linked inheritance. The vast majority of PAMs are due to mutations in genes coding for a wide variety of cellular proteins resulting in defects in their structure and function. They comprise cytoskeletal proteins (desmin, plectin, αB-crystallin, α-actin, myosins, titin, filamin C, myotilin, ZASP, BAG-3, FHL1), enzymes (GNE, calpain-3), proteins of autophagy (LAMP2, VCP, TRIM32, VMA21), DNA-binding proteins (PABPN1), and selenoproteins (SEPN1) (Schroder, 2013, Malicdan and Nishino, 2013). There are also a number of sporadic and genetically unsolved cases, which adds to the genetic complexity of PAMs. The precise molecular pathways that lead to abnormal protein aggregation are still unclear. However, recent studies have brought some insight into the molecular pathogenesis of these pathological features and clinical phenotypes. It has been proposed that the formation of protein aggregates may be due to defects in multiple aspects including extralysosomal proteolytic degradation as well as development, integration and maturation of sarcomeric and extrasarcomeric protein components (Goebel and Muller, 2006, Schroder, 2013). This may explain why some lesions within muscle fibres contain multiple protein aggregates. Investigations of muscle diseases with unidentified proteins in terms of clinical, genetic and myopathological phenotypes may further expand the spectrums of PAMs.

This thesis covers four main conditions from those listed above – myopathies with tubular aggregates, myopathies with cylindrical spirals, myopathies with cytoplasmic

bodies, and centronuclear/myotubular myopathies. These will be discussed in the following sections.

1.3.1 Myopathies with Tubular Aggregates and Cylindrical Spirals

Tubular aggregates (TAs), which are unusual membranous structures in skeletal muscle, were first described in 1964 (Engel, 1964). They are found predominantly in type II muscle fibres, although they have been described in both type I and type II fibres (Rohkamm et al., 1983, Pierobon-Bormioli et al., 1985, Cameron et al., 1992). By enzyme histochemistry, TAs appear bright red with modified Gomori trichrome, and stain with periodic acid-Schiff (PAS). They also react for myoadenylate deaminase (MAD), even without substrate, nonspecific esterase, and nicotinamide adenine dinucleotide-tetrazolium reductase (NADH-TR). But they do not react for the mitochondrial oxidative enzymes succinate dehydrogenase (SDH) or COX (Goebel et al., 2013). It is now well established that these structures originate from a complete sarcoplasmic reticulum (SR) containing a number of different SR components, including calsequestrin (CASQ), ryanodine receptor (RyR), triadin, sarcoplasmic endoplasmic reticulum calcium ATPases (SERCAs), sarcalumenin, and dihydropyridine receptor (DHPR) (Chevessier et al., 2004, Boncompagni et al., 2012, Salviati et al., 1985, Chevessier et al., 2005). In addition, they also contain heat-shock protein epitopes (Martin et al., 1991) and dysferlin (Ikezoe et al., 2003). Ultrastructurally, TAs are characterised by the abnormal accumulation of densely packed tubules of variable forms and sizes located in skeletal muscle fibres (Pavlovicova et al., 2003). There are several subtypes of tubules, appearing single-walled, containing amorphous, granular mildly electron-dense material, double-walled or containing several smaller inner tubules, or aggregates of tubule-filamentous structures or filamentous tubules (Pavlovicova et al., 2003). Double-walled or tubules with some amorphous content seem to predominate in different clinical conditions, and yet there is no known correlation between the type of disease and the type of TA (Pavlovicova et al., 2003). Immunohistochemical and EM images of TAs can be found in Chapter 5, Figures 5-5 and 5-6.

Cylindrical spirals (CSs) are a rare structural finding in muscle biopsies characterised by the accumulation of spiral lamellae in skeletal muscle fibres, which were first

described by Carpenter et al. (Carpenter et al., 1979). The abnormal accumulations of CSs are observed mainly or exclusively in the subsarcolemma of type II fibres, and appear bluish with hematoxylin & eosin (H&E) stain, and bright red with Gomori trichrome. CSs are also lightly stained with PAS, strongly or faintly coloured with NADH-TR stain, strongly stained with non-specific esterase and MAD, but not reactive to SDH or COX or myosin adenosine triphosphatase (ATPase) (Goebel et al., 2013). Occasionally, fibres with CSs resemble fibres with TAs or ragged red fibres, particularly when stained with Gomori trichrome and MAD, but their reaction with menadione α-glycerophosphate helps to distinguish them from TAs, as TAs are only lightly stained with menadione α-glycerophosphate reaction (Malfatti et al., 2015). Under EM, clusters of CSs present in variable lengths and measuring 1-5 μm in diameter, and spirals are composed of up to 20 concentric lamellae with a rounded central core continuous at its ends with sarcoplasm containing small vesicles and/or glycogen granules. Each lamella appears trilaminar with no direct continuity with T-tubular, longitudinal sarcoplasmic reticulum or terminal cisternae (Goebel et al., 2013). CSs are thought to originate from a SR component. A recent study found that CSs were highly immunoreactive to the longitudinal SR protein SERCA1 but were not immunoreactive to the junctional SR proteins RYR1 and CASQ, indicating that CSs arise from longitudinal SR (Xu et al., 2015). Therefore, antibodies for RYR1 and CASQ proteins could also help to differentiate between CSs and TAs. Immunohistochemical and EM images of CSs can be found in Chapter 5, Figures 5-17 and 5-18.

1.3.1.1 Clinical and Genetic Heterogeneity of Myopathies with Tubular Aggregates

TAs are found mainly in males, but have been reported in both genders in some familial cases (Ghosh et al., 2010). TAs are present in a wide variety of different disorders, which may occur either sporadically or in a hereditary manner. These include hypokalaemic paralysis, myotonia congenita, malignant hyperthermia, gyrus atrophy, inflammatory myopathy, alcohol/drug-induced myopathies, Whipple's disease, porphyria cutanea tarda, phosphoglycerate mutase (PGAM) deficiency, and recently a case of a mitochondrial encephalopathy (Engel, 1964, Rosenberg et al., 1985, Oh et al., 2006, Morgan-Hughes, 1998, Wedatilake et al., 2015). The major clinical phenotypes of myopathies with TAs include exertional myalgia, muscle

cramps and stiffness with or without weakness, slowly progressive proximal weakness, fatigability and periodic paralysis. TAs are not a consistent finding in many of these disorders and often form a relatively minor feature of the overall pathology in muscle (Morgan-Hughes, 1998). One study suggested that myopathies with TAs could be divided into two groups: i) myopathies with large TAs (TA surface area above 125 μm^2), corresponding to the primary TA myopathies, found in type II fibres and sometimes also in type I fibres, many with a family history, and absent of any other associated disorder; ii) myopathies with small TAs, found exclusively in type II fibres, mostly without a family history, and presence of another associated disease, where TAs are considered to be a feature of a secondary condition (Funk et al., 2013).

Four distinctive clinical syndromes have been described in which TAs are the major abnormality in skeletal muscle from patients: i) autosomal recessive limb girdle myasthenia, ii) progressive limb girdle weakness which may be sporadic or inherited in an autosomal dominant or recessive fashion, iii) exercise related muscle pain, cramps and stiffness, and iv) gyrate atrophy of the choroid and retina (Morgan-Hughes, 1998). In addition to the above disorders, a number of case reports have also described TAs as a feature in skeletal muscle of patients with other clinical conditions. These include cases associated with pupillary abnormalities suggesting smooth muscle dysfunction may co-occur with TAs in some cases (Jacques et al., 2002, Shahrizaila et al., 2004). The presence of TAs was also reported in two alcoholic patients with chronic liver disease but no apparent neuromuscular disorder (del Villar Negro et al., 1982). A case with an isolated late onset distal myopathy, was reported to be associated with multiple deletions of mtDNA and high densities of TAs in muscle fibres (Garrard et al., 2002), which also adds to the growing list of clinical phenotypes associated with TAs.

Mutations in eight genes have so far been implicated in the formation of TAs (Table 1-9). These genes are mainly associated with four clinical phenotypes – dominantly inherited tubular aggregates myopathy, phosphoglycerate mutase deficiency (glycogenosis type X), congenital myasthenic syndromes, and periodic paralysis. These will be discussed in the following sections.

Table 1-9 Genes previously reported in cases with tubular aggregates

Gene ID	Gene OMIM#	Clinical Phenotype (OMIM#)	Age of onset	Inheritance
STIM1	605921	Tubular aggregate myopathy (160565) Stormorken syndrome (185070)	Child - adulthood	Autosomal dominant
ORAI1	610277	Tubular aggregate myopathy (160565) Stormorken syndrome (185070)	Child - adulthood	Autosomal dominant
PGAM2	612931	Glycogen storage disease X/ phosphoglycerate mutase deficiency (261670)	8-52 years	Autosomal recessive
GFPT1	138292	Congenital myasthenia, 12 (610542)	1-40s years	Autosomal recessive
DPAGT1	191350	Congenital myasthenia, 13 (614750)	Infancy/ neonatal – early childhood	Autosomal recessive
ALG2	607905	Congenital myasthenia, 14 (616228)	Infancy – 4 years	Autosomal recessive
SCN4A	603967	Hypokalemic periodic paralysis, type 2 (613345) Paralysis periodica paramyotonia	2 years – childhood	Autosomal dominant
SERAC1	614725	3-methylglutaconic aciduria with deafness, encephalopathy, and Leigh-like syndrome (614739)	Neonatal	Autosomal recessive

OMIM = online Mendelian inheritance in man.

Tubular Aggregate Myopathy with Dominant *STIM1* and *ORAI1* Mutations

Stromal interaction molecule 1 (STIM1) is a calcium sensing protein in the endoplasmic reticulum membrane, which binds to ORAI1, a calcium release-activated channel (CRAC), mediating store-operated calcium entry in the plasma membrane. This calcium influx is responsible not only for maintaining cell signalling, but also many other cellular functions such as exocytosis, gene expression and apoptosis (Lacruz and Feske, 2015). Four *STIM1* mutations (c.216C>G [p.His72Gln], c.251A>G [p.Asp84Gly], c.325C>A [p.His109Asn], and c.326A>G [p.His109Arg]) were reported in four autosomal dominant families affected by nonsyndromic tubular

aggregate myopathy (TAM) (Bohm et al., 2013). *STIM1* was the first gene to be identified in association with dominantly inherited TAM. Later, the mutation of p.His109Asn and a novel mutation of c.343A>T (p.Ile115Phe) in *STIM1* were also described in three unrelated childhood onset TAM patients. This taken together with the first study suggests that p.His109Asn is a hotspot mutation related to early onset muscle weakness (Hedberg et al., 2014). Additionally, five missense mutations (c.239A>C [p.Asp80Thr], c.286C>G [p.Leu96Val], c.322T>A [p.Phe108Ile], c.322T>C [p.Phe108Leu], and c.242G>A [p.Gly81Asp]) in *STIM1* have also been reported in two further studies on TAM patients (Bohm et al., 2014, Walter et al., 2015). All of these mutations lie in the calcium-binding EF-hand domain of *STIM1* and functional analysis demonstrated that these mutations trigger abnormal STIM1 clustering, thereby enhancing the activity of the CRAC channel and resulting in an increased calcium influx in both SR and cytoplasm in TAM cells (Bohm et al., 2013). Interestingly, a recent study reported a novel heterozygous frameshift mutation c.1450_1451insGA (p.lle484ArgfsTer21) that is located in the cytoplasmic C-terminal inhibitory domain (CTID) of STIM1. Functional analysis found that intracellular calcium influx was decreased by this mutation (Okuma et al., 2016).

A further gain-of-function mutation in *STIM1*, c.910C>T (p.Arg304Trp), was also reported in patients from several TAM families but with a rare condition known as Stormorken syndrome (Misceo et al., 2014, Nesin et al., 2014). Stormorken syndrome is an autosomal dominant disorder characterised as congenital miosis, bleeding diathesis, thrombocytopenia and proximal muscle weakness. Furthermore, a gain-of-function mutation in *ORAI1*, c.734C>T (p.Pro245Leu), was for the first time reported as a genetic cause of a patient with a Stormorken-like syndrome of congenital miosis and TAM (Nesin et al., 2014). Subsequently, another two missense mutations, c.292G>A (p.Gly98Ser) and c.412C>T (p.Leu138Phe), in *ORAI1* were reported in three families affected by dominantly inherited TAM with hypocalcaemia (Endo et al., 2015). These further expand the spectrum of phenotypes associated with dominantly inherited TAM. Functional studies showed that these mutations altered calcium homeostasis by different mechanisms. The p.Arg304Trp in *STIM1*, located in the coiled-coil 1 (CC1) domain, appears to play a role in modulating fast calcium-dependent inactivation, which might also have effect on the haemostatic system of platelets (Nesin et al., 2014). The p.Pro245Leu in *ORAI1* incompletely deactivates

the CRAC channel (Nesin et al., 2014), while the p.Gly98Ser and p.Leu138Phe cause constitutive CRAC channel activation in a STIM1-independent manner (Endo et al., 2015). Although loss-of-function mutations have also been identified in *STIM1* and *ORAI1* and are linked to immune deficiency and muscular myopathy, none of the cases present with TAs in muscle. (Feske et al., 1996, Feske et al., 2006) Therefore, the *STIM1* and *ORAI* related dominant TAM is likely to be associated with the dysregulation of calcium homeostasis (Okuma et al., 2016).

Phosphoglycerate Mutase Deficiency

Phosphoglycerate mutase (PGAM) deficiency is a rare autosomal recessive muscle glycogen-storage disease characterised by exercise intolerance, cramps and recurrent myoglobinuria (Tsujino et al., 1995). It is more commonly seen in African-Americans, but has also been reported in Italian, Japanese and Pakistani patients (Naini et al., 2009, Tsujino et al., 1995). Seven mutations in the *PGAM2* gene from 14 patients have been reported to be associated with PGAM deficiency since it was first described in 1981 (Salameh et al., 2013). Interestingly, TAs were described in six of the patients (40%) (Salameh et al., 2013), but have never been associated with other muscle glycolysis disorders. The c.233G>A (p.Trp78Ter) mutation is the most common mutation and has only been seen in African-American patients (Oh et al., 2006, Salameh et al., 2013). The c.268C>T (p.Arg90Trp) mutation was reported in two Italian cases, however, TAs were not observed (Tsujino et al., 1995, Toscano et al., 1996). Homozygous frameshift mutation c.del532_533G (p.Gly178*fs*30Ter) and stop-gain mutation c.538C>T (p.Arg180Ter) were two other mutations reported in an Italian and a Pakistani patients with TAs, respectively (Naini et al., 2009). PGAM is a dimeric glycolytic enzyme, and deficiency of PGAM could result in a terminal block in glycogenolysis (DiMauro et al., 1981). The p.Trp78Ter and p.Arg90Trp mutations are close to an active site and were suggested to reduce enzymatic activity (Tsujino et al., 1993). Interestingly, two patients with TAs showed increased calcium concentration and calcium adenine triphosphate activity in muscle, and treatment with the calcium release inhibitor dantrolene sodium improved their ischemic exercise-induced contractures (Vissing et al., 1999). In spite of this, it is still difficult to establish the specific trigger and relationship between PGAM insult and the formation of TAs.

Congenital Myasthenic Syndromes with Mutations in *GFPT1*, *DPAGT1*, and *ALG2*

Congenital myasthenic syndromes (CMSs) are a heterogeneous group of genetic diseases characterised by defects in neuromuscular transmission. This may be due to multiple mechanisms which in general cause fatigable muscle weakness. The onset of disease is usually during infancy and childhood (Hantai et al., 2013). Whilst mutations in 17 genes have been associated with CMSs, the presence of TAs in CMSs has only been reported with mutations in three genes: *GFPT1* (glutamine-fructose-6-phosphate transaminase 1), *DPAGT1* (dolichyl-phosphate N-acetyl-glucosaminephosphotransferase 1), and *ALG2* (alpha-1,3/1,6-mannosyltransferase). They are all involved in the early steps of N-linked (asparagine-linked) protein glycosylation and mutations in these genes are suggested to lead to reduced levels of acetylcholine receptors at the endplate region due to defects in glycosylation of acetylcholine receptor subunits (Belaya et al., 2012, Huh et al., 2012, Cossins et al., 2013, Hantai et al., 2013, Zoltowska et al., 2013). Patients with mutations in these three genes usually share similar clinical phenotypes of limb-girdle pattern of myasthenia, further indicating similar mechanisms involved with these genes. Since TAs were observed in all but two younger cases with *GFPT1* mutations, and all patients with *DPAGT1* mutations, TAs have become a useful pathological marker for these two types of myasthenia (Hantai et al., 2013). Although the primary mechanisms of CMSs with TAs caused by these genes have not yet been elucidated, it seems that the formation of TAs may also be linked to impairment in the N-linked protein glycosylation pathway.

Myopathy with *SCN4A* Mutations

Mutations in the gene encoding the voltage-gated sodium channel α subunit (*SCN4A*) are responsible for several skeletal muscle disorders including periodic paralysis (hyper, hypo, or normokalaemic), potassium-aggravated myotonia, permanent myotonia, and congenital myasthenic syndrome (Nicole and Fontaine, 2015). TAs have been observed in patients with hyper and hypokalaemic periodic paralysis (Bradley et al., 1990, Gold and Reichmann, 1992). To date, TAs together with mutations in *SCN4A* have been reported in a Caucasian family of hypokalemic periodic paralysis (c.2014C>G, p.Arg672Gly) (Sternberg et al., 2001) and additionally in a Chinese family of paralysis periodica paramyotonica (c.2111C>T,

p.Thr704Met) (Luan et al., 2009). These studies suggested that accumulation of tubules might be a common pathological change for myopathy with *SCN4A* mutations. The formation of TAs may be a result of a persistent sarcolemmal dysfunction under ER stress affected by the genetic mutation (Sternberg et al., 2001, Luan et al., 2009). Genetic screening in additional cases and functional work are required to explore the relationship between *SCN4A* mutations and the presence of TAs.

Mitochondrial Encephalopathy with a *SERAC1* Mutation

Recently, Wedatilake et al. reported a 13-year-old boy with progressive mitochondrial encephalopathy in the Leigh syndrome spectrum. They found very frequent but relatively small subsarcolemmal TAs in his most recent, but not in his neonatal muscle biopsy. Further genetic investigation by whole-exome sequencing revealed a known homozygous pathogenic splice mutation c.1403+1G>C in the mitochondrial gene *SERAC1* (serine active site containing 1) (Wedatilake et al., 2015). *SERAC1* encodes a phospholipase located at the interface between mitochondria and ER in the mitochondria-associated membrane fraction, which is involved in cholesterol trafficking and lipid remodelling (Wortmann et al., 2012). They hypothesised that the delayed formation of TAs was a result of aggregation of ER constituents in the intervening period due to inefficient lipid membrane remodelling caused by the *SERAC1* mutation (Wedatilake et al., 2015).

1.3.1.2 Possible Pathogenesis of Tubular Aggregates

The clinical and genetic heterogeneity of myopathies with TAs presents a formidable challenge to understand the mechanisms underlying the formation of TAs and establish their function in the disease process. However, the remarkable consistency of these specific pathological features suggests that TAs in different disorders may share a common mechanism (Morgan-Hughes, 1998). Currently, the popular hypotheses are that the formation of TAs is an adaptive response to i) abnormalities of intracellular calcium homeostasis, and ii) defects in N-linked protein glycosylation pathway. Although a study which screened 10 candidate genes whose coding proteins were along the calcium regulation pathway in 15 TAM cases failed to identify any disease-causing variants (De Paula et al., 2012), studies in future with more cases with TAs are still required. Candidate genes involved in the two possible

pathogenic pathways should be investigated to compare genetic overlaps between cases with similar phenotypes. There may be other pathogenic pathway in addition to the two described above; therefore, identifying new genes associated with TAs is also important.

1.3.1.3 Clinical Phenotypes Associated with Cylindrical Spirals and Their Possible Pathogenesis

Only 18 cases (males =10) with cylindrical spirals have been reported in 13 studies to date (Rapuzzi et al., Carpenter et al., 1979, Bove et al., 1980, McDougall et al., 1980, Yamamoto et al., 1982, Gibbels et al., 1983, Thomas et al., 1984, Danon et al., 1989, Taratuto et al., 1991, Baker et al., 1997, Wolfe et al., 1997, Malfatti et al., 2015, Xu et al., 2015) (Table 1-10). The age of onset ranged from infant to adulthood. The patients presented with a variety of clinical phenotypes, suggesting CSs are not disease-specific features. Interestingly, a correlation between the activity of NADH and clinical phenotype was reported in a recent study. By reviewing previously reported studies, it showed that patients with low NADH activity presented with muscle weakness while those with high NADH activity presented with cramps or myotonia. This suggests a role for NADH in the pathogenesis of diseases with CSs (Xu et al., 2015). Among the reported patients, six were sporadic, and 12 were familial cases. Although a common pattern of inheritance is difficult to be established, the formation of CSs has been speculated to be an acquired genetic defect in calcium handling (Gibbels et al., 1983, Malfatti et al., 2015). Given the similarities between CSs and TAs, *STIM1*, *STIM2*, and *ORAI1* genes were sequenced in two siblings with CSs and no genetic mutation was found (Xu et al., 2015). To date no other genetic factors have been identified. Further investigations are required to identify possible genetic abnormality in patients with CSs.

Table 1-10 Clinical phenotypes of published cases with cylindrical spirals (Modified from (Malfatti et al., 2015))

Author	Age	Sex	Family History	Reported Diagnosis	Major Symptoms	Other Features in Biopsy	Serum CK
Carpenter et al.	53	M	Unknown	Malignancy	Muscle painful cramps exacerbated by exercise	Tubular aggregates	Unknown
Carpenter et al.	53	M	Gait disorders	Heredofamiliar spinocerebellar degeneration	Progressive gait ataxia	Subsarcolemmal mitochondrial aggregates with paracrystalline inclusions	Unknown
Bove et al.	31	F	Mother of the case below	Percussion myotonia	Exertional muscle cramps, stiffness, progress with age	No	Normal
Bove et al.	10	M	Son of the case above (sister with similar myotonic symptoms but normal muscle biopsy)	Percussion myotonia	Lid lag; no complaints of cramps or stiffness	No	Normal
McDougal et al.	20	M	Unknown	Melorheostosis	Pain and weakness of left thigh	No	Unknown
Yamamoto et al.	27	M	Unknown	Abnormal muscle mitochondria	Pain and cramps of proximal leg muscles	Unknown	Unknown
Gibbels et al.	60	M	Unknown	Alcoholism, diabetes, nicotine abuse, polyneuropathy	Wasting of forearm muscles, aching leg with stiffness	Single degenerating spirals resembling concentric lamellating bodies and myelin-like bodies	Normal
Danon et al.	42	M	Unknown	Dementia	Severe leg pain, left-sided transient episodic weakness	Tubular aggregates	Normal
Thomas et al.	25	F	Familial Behr's syndrome	Behr's syndrome, optic atrophy mental	Pyramidal weakness, dysarthria, ataxia	Mitochondrial accumulation, mitochondrial paracrystalline	Normal

				deterioration, nystagmus		inclusions	
Taratuto et al.	70	F	Muscle weakness, gait disorders, motor impairment, scoliosis (mother of the case below)	Autosomal dominant neuromuscular disorder	Lower limb weakness, myopathic facial features	Dilation of terminal cisternae	High
Taratuto et al.	52	M	Muscle weakness, gait disorders, motor impairment, scoliosis (son of the case above)	Autosomal dominant neuromuscular disorder	Progressive lower weakness, gait instability, myopathic facial features	Inclusions recalled ragged fibres or rimmed vacuoles	High
Rapuzzi et al.	30	M	Unknown	High serum CK	Asymptomatic	No	High
Baker et al.	4m	F	History of a brother dead at birth after normal pregnancy	D-2-hydroxyglutaric aciduria	Hypotonia, seizures, cardiomyopathy	No	Unknown
Wolfe et al.	31	M	Exertional cramps and myoglobinuria	Rhabodomyolysis, schizophrenia	Exertional cramps, myoglobinuria	No	High
Malfatti et al.	12	F	Sister of the case below	Congenital myopathy with epileptic encephalopathy	Muscle hypotonia, epilepsy, psychomotor delay	No	Normal
Malfatti et al.	13	F	Sister of the case above	Congenital myopathy with epileptic encephalopathy	Muscle hypotonia, epilepsy, psychomotor delay	No	Normal
Xu et al.	53	F	Sister of the case below	Exercise intolerance, myalgia	Progressive muscle weakness, exercise intolerance, myalgia	Unknown	Normal
Xu et al.	45	F	Sister of the case above	Exercise intolerance	Abnormal gait, exercise intolerance, lower limb weakness	Lipofuscin-like bodies, enlarged cisternae	Normal

CK = creatine kinase; M = male; F = female; m = months

1.3.2 Myopathies with Cytoplasmic Bodies

Cytoplasmic bodies are a crucial morphological hallmark of PAMs. They are an eosinophilic structure with a circular profile consisting of intermediate filaments (Sewry and Goebel, 2013). Cytoplasmic bodies are nonspecific but can be a common feature of some MFMs, and also of IBM (IBM has been discussed in previous sections) (Sewry and Goebel, 2013). MFMs are the largest group of PAMs. They are most commonly inherited in an autosomal dominant format and occasionally in an autosomal recessive or X-linked manner (Goebel and Muller, 2006). MFMs can be further sub-divided according to the genetic causes. The six major forms of MFMs are summarised in Table 1-8. The mutations in these genes have all been traced to defects in Z-disk associated proteins (Behin et al., 2015). Other subtypes of MFMs include selenoproteinopaty caused by mutations in the selenoprotein gene (*SEPN1*, OMIM#606210) (Ferreiro et al., 2004), laminopathy caused by mutations in the Lamin A/C gene (*LMNA*, OMIM#150330) (D'Amico et al., 2005), and hereditary myopathy with early respiratory failures caused by mutation in the titin gene (*TTN*, OMIM#188840) (Ohlsson et al., 2012, Pfeffer et al., 2012). The majority of patients with MFMs present in adulthood with muscle weakness, more distally than proximally, which is frequently associated with cardiomyopathy. Some patients have an infantile or juvenile disease onset (Behin et al., 2015). Individual MFMs subtype has distinct clinical phenotypes and muscle imaging features. Genetic testing is essential in establishing an accurate diagnosis of MFM. In addition, the use of laser microdissection and mass spectrometry-based proteomics technologies can also improve the diagnosis and enlarge the spectrum of MFMs (Claeys and Fardeau, 2013, Olive et al., 2013).

1.3.3 Centronuclear Myopathies and Myotubular Myopathy

Centronuclear myopathies (CNMs) are a genetically heterogeneous group of congenital myopathies. They are defined by the presence of abundant nuclei placed in rows in the central part of the myofibres as the most prominent histopathological feature (Romero, 2010). The most common genetic causes of classical CNMs are attributed to mutations in the *MTM1* gene encoding for the 3'-phosphoinositides phosphatase myotubularin 1, the *DNM2* gene encoding for dynamin-2, and the *BIN1*

gene encoding for amphiphysin 2 (Romero and Laporte, 2013, Ravenscroft et al., 2014) (Table 1-11). X-linked mode of CNMs presents with characteristic central nuclei resembling myotubes in a high number of small muscle fibres. Therefore, it is often referred to as myotubular myopathy (Romero and Laporte, 2013). Recently, mutations in the *RYR1* gene encoding the ryanodine receptor 1, the *TTN* gene encoding titin, the *MTMR14* gene encoding myotubularin related 14, and the *CCDC78* gene encoding the coiled-coil domain-containing protein 78 have also been reported in patients with CNM-like histology (Jungbluth and Gautel, 2014). Although there are overlaps between CNMs, each of these forms of CNMs has their distinct histopathological features, clinical phenotypes, and mode of inheritance. The presence of necklace fibres which is considered as a peculiar morphological marker has been reported in both type I and type II fibres of the *MTM1*-related CNM (Bevilacqua et al., 2009), and occasionally seen in CNM due to *DNM2* and *RYR1* mutations (Wilmshurst et al., 2010, Ravenscroft et al., 2014). Genetic testing is essential for an accurate diagnosis of CNM. However, many CNM patients are still genetically undefined. The use of new technologies such as next-generation sequencing may allow clinicians and researchers to provide a genetic diagnosis for these patients.

Table 1-11 Genes implicated in centronuclear/myotubular myopathies

Gene (OMIM#)	Protein	Function	Inheritance	Disease (OMIM#)
MTM1 (300415)	Myotubularin	Phosphoinositide phosphatase	X-linked recessive	Myotubular myopathy (310400)
DNM2 (602378)	Dynamin 2	GTPase in membrane trafficking	Autosomal dominant	Centronuclear myopathy 1 (160150)
BIN1 (601248)	Amphiphysin 2	Membrane remodelling	Autosomal recessive	Centronuclear myopathy 2 (255200)
RYR1 (180901)	Ryanodine receptor 1	Calcium channel of the triad	Autosomal recessive	Centronuclear myopathy with external ophthalmoplegia (255320)
TTN (188840)	Titin	Myofibrillar backbone for contractile machinery	Autosomal recessive	*TTN*-associated congential centronuclear myopathy
MTMR14 (611089)	Myotubularin related 14	Phosphoinositide phosphatase	Autosomal dominant	Modifier of centronuclear myopathy (160150)
CCDC78 (614666)	Coiled-coil domain-containing protein 78	Unknown	Autosomal dominant	Centronuclear myopathy 4 (614807)

OMIM = online Mendelian inheritance in man.

1.4 Application of Genetic Techniques in Mutation Detection and Gene Discovery

A wide range of genetic techniques have been developed since the 20th century to advance the discovery of disease-causing genes. This section will first discuss three widely used traditional genetic techniques – Sanger sequencing, genetic linkage analysis, and genome-wide association study (GWAS), and move on to Next-Generation Sequencing (NGS) technologies which are the most popular approaches in the last 5-10 years.

1.4.1 Traditional Genetic Techniques

1.4.1.1 Sanger Sequencing

The first method of DNA sequencing was invented by Frederick Sanger in 1977, and is therefore referred to as Sanger sequencing (Sanger et al., 1977). Its principle lies in that DNA polymerase is randomly inhibited by a small amount of modified dideoxynucleotides (ddNTPs) which are mixed with normal deoxyribonucleotide (dNTP). This produces newly synthesized DNA fragments of different lengths, which can be subsequently determined by capillary electrophoresis. Sanger sequencing has been the most widely used sequencing method in the past four decades. It is considered to be the gold standard and the most reliable and accurate sequencing method for mutation detection. Currently, Sanger sequencing is also regularly used to validate the results from NGS to rule out artefacts. However, its main limitation is that samples have to be processed individually, and polymerase chain reaction (PCR) amplification of each coding sequence needs to be performed separately. When many genetic regions of interest require sequencing, this becomes a cumbersome and laborious process.

1.4.1.2 Genetic Linkage Analysis

Genetic linkage analysis is a powerful tool for detecting the chromosomal location of disease associated genes. It relies on the principle that genes which reside closely together on a chromosome have a tendency to be inherited together during meiosis, and are therefore genetically "linked" (Dawn Teare and Barrett, 2005). It is used for genetic mapping of Mendelian and complex traits with familial aggregation. A logarithm (base 10) of odds (LOD) score is used as a statistical test to determine whether the mutation is likely to be linked to a genetic marker within a given pedigree (Morton, 1955). Traditionally, it is followed by positional cloning to narrow the candidate region until the gene and its mutations are found. It has the benefit that the chromosomal location of the disease gene can be identified without knowing the underlying biological defect. It has been very successful to assist the isolation of genes for all types of neurologic diseases (Pulst, 1999). However, genetic linkage analysis usually requires a large pedigree with high penetrance and family members with both affected and unaffected status. The more family members are used, the more powerful the linkage analysis is. However, for complex polygenic disorders,

the analysis becomes complicated, and the results of genetic linkage studies have proved hard to reproduce (Altmüller et al., 2001).

1.4.1.3 Genome-Wide Association Study

GWAS was introduced to population studies to examine the association between genetic markers and complex diseases instead of linkage analysis. In old fashion studies, microsatellites, tracts of repetitive DNA which contain tandemly repeated 1-5 base pair (bp) units of DNA, were widely used as genetic markers (Putman and Carbone, 2014). However, size determining for hundreds of microsatellites is a time-consuming and labour-intensive process. Consequently, a single nucleotide polymorphisms (SNP) array-based genotyping technology has become a more popular method to analyse 100,000+ SNPs across the entire genome in a large cohort of patients with a disease, as well as in a matched group of unaffected individuals (Manolio, 2010). GWAS is also a non-candidate-driven approach. The association is defined when the frequency of a SNP in the affected population is significantly higher than that in the control population. Since the first successful GWAS identified the disease loci in patients with age-related macular degeneration in 2005 (Klein et al., 2005), hundreds of studies have been published, reporting over thousands trait-associated SNPs (Welter et al., 2014). GWAS has also proved successful particularly for detecting genetic variants with population frequencies of 5% or more in common complex disorders, as well as discovering novel biological pathways underlying polygenic diseases (Hirschhorn, 2009). However, there are also some limitations with GWAS. Firstly, GWAS usually requires a large sample size, which can be an issue for relatively rare diseases. In addition, these studies also have potentials to produce false-positive results due to genotyping errors and population stratification. Furthermore, the SNP-arrays for GWAS usually do not detect efficiently genetic variants with population frequencies lower than 5% which can have high risk effects on complex diseases (Pearson and Manolio, 2008, Lee et al., 2014).

1.4.2 Next-Generation Sequencing

Sanger sequencing method, known as “first-generation sequencing”, is a very slow and labour-intensive process when in face of large-scale sequencing projects. This challenge catalysed the development of massively parallel and high-throughput

sequencing techniques, referred as “Next-Generation Sequencing” (NGS) (Jarvie, 2005). NGS has been applied to genome sequencing, RNA-sequencing, ChIP-sequencing (DNA-protein interactions), and characterising epigenome (de Magalhaes et al., 2010). In this thesis, NGS is relevant to three main aspects: whole-exome sequencing (WES), whole-genome sequencing (WGS), and targeted-gene sequencing. Advances in these NGS techniques have made a revolutionary change in the field of gene hunting for Mendelian diseases and even complex disorders due to the enormous reduction in cost in the past 5-10 years.

Despite NGS technologies differing in their chemistries and preparation protocols, they share the same basic workflow. Step one is the preparation of the genomic DNA library. Genomic DNA is shattered intro small DNA fragments, and then specific adaptor sequences are added to each end of the sheared DNA. This allows each DNA fragment to be hybridized to the flow cell and then to be amplified and sequenced in parallel. In WES and targeted-gene sequencing protocols, DNA fragments are hybridized to probes that are complimentary to all the known exons or the targeted regions. This process is therefore called “capture the exon sequences”. The remaining DNA fragments are subsequently washed away. In WGS, the DNA fragments are sequenced without the capture step. The sequences of the fragments are read cyclically in parallel, and in most cases using the form of fluorescent or electrical signals, which are detected by an imaging sensor system. Lastly, the large amount of data is processed in a bioinformatics pipeline to align the sequences to a reference genome, identify the genetic variants and then annotate the function of each variant (Bras et al., 2012). (Figure 1-4)

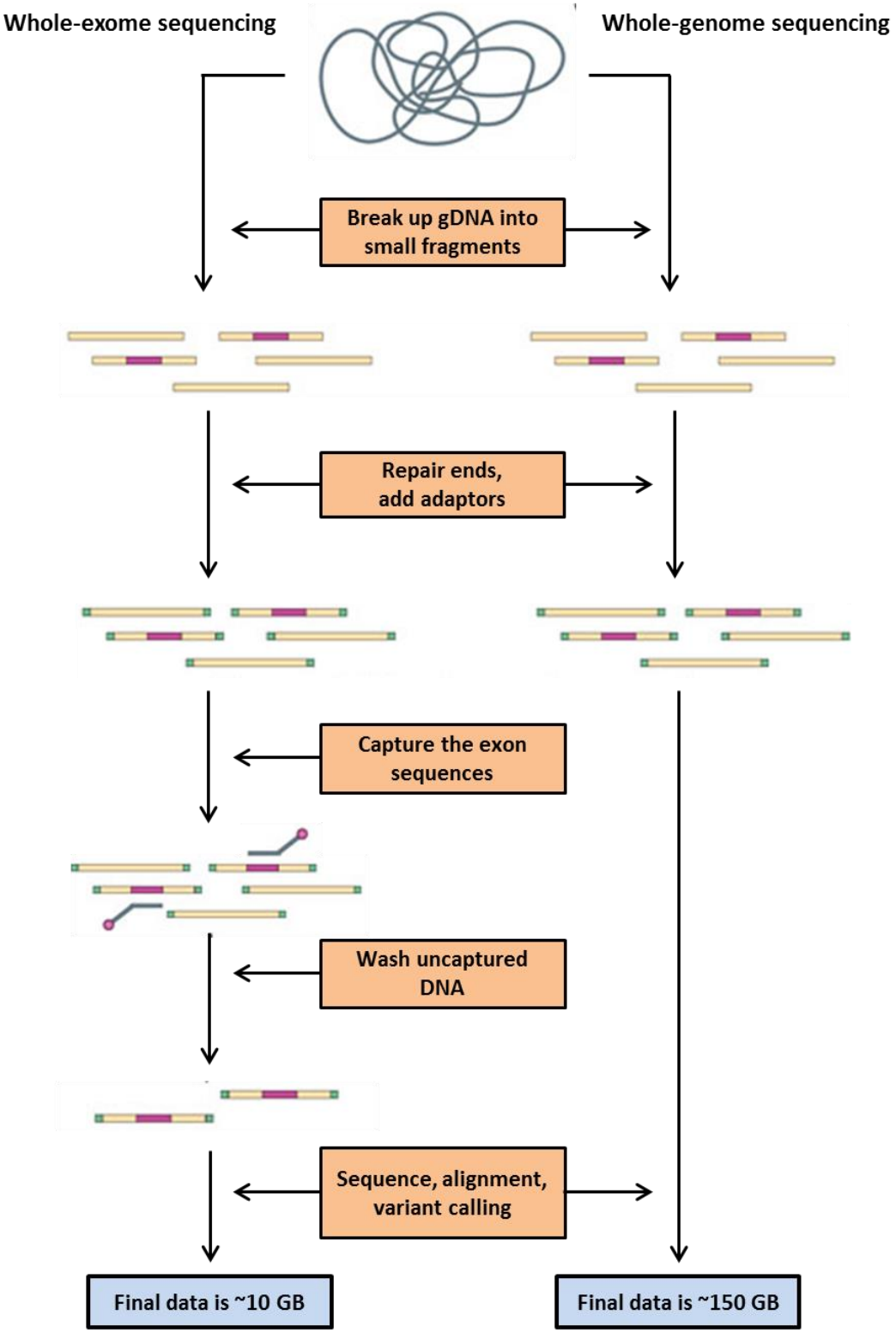


Figure 1-4 Simplified workflows for whole-exome and whole-genome sequencing

(Modified from (Bras et al., 2012))

1.4.2.1 Whole-Exome Sequencing and Whole-Genome Sequencing

Genes in protein-coding regions constitute approximately 1-2% of the human genome (the "exome"). However, they harbour roughly 85% of disease-causing

mutations (Choi et al., 2009). Therefore, WES, a technique to sequence complete coding regions, is considered as a cost-efficient strategy with the potential to contribute to the understanding of common and rare human diseases. WES is also the method of choice for the studies included in this thesis.

An average of 20,000 – 24,000 variants (mostly benign variants) is usually generated by WES for each individual. A major challenge is to assign the causality among these variants, unless a known mutation is found. A series of functional annotation and pathogenic prediction software tools can be used to filter the variants according to their structural effect, frequency in the population, and their conservation during evolution. In addition, combination of inheritance pattern with expression and functional data can enable WES to reach its full potential as a tool for gene discovery.

There are five main applications of WES used in both research and diagnostic settings: i) WES of multiple unrelated patients or families with similar phenotype combined with inheritance pattern to discover the disease-causing genes and mutations; ii) WES of an affected child and two unaffected parents as a trio to discover de novo mutations; iii) WES of distant affected relatives in a pedigree to reduce candidate genes; iv) a rapid screen for candidate disease-causing genes or predisposing genes when there is a previous hypothesised biological pathway; v) Whole-exome wide association analysis to compare the WES data from a large cohort of unrelated cases and controls (Olgiati et al., 2016).

WGS is expected to become the method of choice in the near future, as the cost is continually reducing. WGS provides the most comprehensive genetic information for each individual, which could answer the question for those phenotypes influenced by the non-coding variants. However, aside from the fact that the current cost of WGS remains substantial, the interpretation and storage of the enormous amount of data generated by WGS is extremely challenging (Olgiati et al., 2016). Therefore, WES still currently is the cheaper and more popular choice in both research and clinical diagnostic applications.

1.4.2.2 Targeted-Gene Sequencing

Targeted-gene sequencing (panels) is an alternative approach when the budget is limited for WES, in which only specific genes are sequenced (Gnirke et al., 2009). Both commercially available custom-designed and pre-designed panels have been used in current published studies (Balabanski et al., 2014, Weihl et al., 2015). These genes are often specific to relevant phenotypes or biological pathways. This method enables increased coverage of the targeted regions of interest, and also quickens and simplifies the analysis (Ballester et al., 2016). Therefore, targeted-gene panels, particularly disease-specific sequencing panels are thought to be more useful for genetic diagnosis in known genes for diseases with genetic heterogeneity in comparison to WES.

The comparison between some traditional sequencing approaches and NGS technologies is listed in Table 1-12. Selection of appropriate genetic technique and the experimental design should rely on the type of variants, the frequency of the disease of interest, the mode of inheritance, penetrance and the cohort size, as well as the funding budget (Olgiati et al., 2016).

Table 1-12 Advantages and disadvantages of different genetic approaches (Modified from (Gang et al., 2015))

Genetic approaches		Advantages	Disadvantages
Traditional Sanger sequencing		• Focus on a small number of known or candidate genes • Capability to identify polymorphisms with low allele frequency • Deep sequencing and confirmation of the genes identified in GWAS or NGS	• Low time and cost-efficiency and labour-intensive for large-scale sequencing projects
Linkage analysis		• A highly useful tool for monogenic disorders and Mendelian diseases • Non-candidate-driven approach	• Large pedigrees with many affected and unaffected family members are required • The results are hard to reproduce for complex disorders
Genome-wide association studies (GWAS)		• A complete, unbiased picture of the genome • No prior hypotheses required	• Large numbers of samples are required • Insensitive to both structural and rare variants
Next-generation sequencing (NGS)	Targeted-gene sequencing	• High coverage targeted for gene panels (e.g. candidate pathway genes) • Low complexity • Relatively less costly compared with whole-exome and whole-genome sequencing	• May potentially miss important genes • A *priori* knowledge of candidate gene(s) is required
	Whole-exome sequencing (WES)	• Sequencing covers all the protein coding, some promoter and short flanking intronic regions and delivers panels of genes • Ability to identify rare causal/risk coding variants for complex disorders using case-control association analysis • Less expensive than WGS	• High complexity and bioinformatics demand • Uneven/low coverage of certain parts of the exome
	Whole-genome sequencing (WGS)	• Ability to report protein coding regions and the remainder of genome, including introns, promoters and intergenic regions • More suitable for copy number variants (CNVs) detection • Ability to identify rare causal/risk variants for complex disorders using case-control association analysis	• Huge complexity and bioinformatics demand • Overall cost is high

1.5 Thesis Aims

The overall aim of this thesis is to investigate the genetic architecture of sporadic inclusion body myositis (sIBM), and also to improve the understanding of genetic heterogeneity of patients with specific/rare abnormal structure and protein aggregates in muscle fibres. This thesis is focused on the following four aspects:

1. To screen a series of candidate genes and investigate their associations in a group of patients with sIBM. (Chapter 3)

2. To identify the high risk genetic variants associated with sIBM. Whole-exome wide association analysis is performed in a large cohort of sIBM cases and healthy controls to improve the understanding of the underlying pathways in sIBM. (Chapter 4)

3. To use whole-exome sequencing (WES) to identify the disease-causing genes for a group of patients with tubular aggregates and cylindrical spirals in muscle fibres. This will provide a genetic diagnosis for many of the patients, and also contribute to expanding the clinical and genetic spectrum of these conditions. (Chapter 5)

4. To use WES to uncover the genetic basis of three families with protein aggregates and a family with a rare form of myotubular myopathy. This will establish the benefit of using WES in genetic diagnosis for patients with rare muscular conditions. (Chapter 6)

Chapter 2 Materials and Methods

2.1 Patients Collection and Ethical Approval

2.1.1 IBM Patients

This study is supported by the International IBM Consortium Genetic Study (IIBMCGS). The inclusion criteria and exclusion criteria for IBM patients are listed in Table 2-1. The study proforma is available in Appendix I - Table 1.

Table 2-1 Inclusion and exclusion criteria of IIBMCGS

Inclusion criteria	
(Patient is included as long as fulfils one of the three criteria below.)	
1	Definite or possible IBM according to the Griggs Criteria (Griggs et al., 1995)
2	Definite or probable IBM according to ENMC Criteria (Vershuuren et al., 1997)
3	Pathologically, clinically defined or possible IBM according to MRC Centre for Neuromuscular Diseases Criteria (Hilton-Jones et al., 2010, Benveniste and Hilton-Jones, 2010)
Exclusion criteria	
1	Known co-existent other neuromuscular disease
2	Known co-existent primary genetic disease
3	Known co-existent primary mitochondrial disease
4	Known communicable disease that would make trial tissue a potential infection risk

The sIBM samples were collected from two main resources: i) stored DNA samples or muscle tissue from patients previously diagnosed of sIBM, and ii) blood samples from new patients being seen in the clinic and diagnosed of sIBM. DNA samples from sIBM patients diagnosed at the MRC Centre for Neuromuscular Diseases and National Hospital of Neurology and Neurosurgery (NHNN) were obtained by searching the NHNN DNA database. The available DNA samples were requested by the author of this thesis from the NHNN Neurogenetics Laboratory. The study proforma forms of selected cases were completed by reviewing their clinical notes or online clinical database record. This work was done by the author of this thesis with the help from Dr Pedro Machado and research nurse Ms Iwona Skorupinska from the MRC Centre for Neuromuscular Diseases. Muscle tissues of sIBM patients from

NHNN were obtained by reviewing patients' muscle biopsy reports. The available tissues were requested by the author of this thesis with the help from Professor Janice Holton and her colleagues from the Department of Pathology in the Institute of Neurology (ION), University College London (UCL). Blood samples from new patients were collected by Iwona Skorupinska and study proforma forms were completed by Dr Pedro Machado in the clinic.

The IIBMCGS was coordinated by Dr Pedro Machado. Once the collaboration with other centres/institutes was set up, the blood collection kits with study documents were prepared and shipped to each collaborator for prospective blood collection. Blood collection kits and the documents (described in Section 2.1.3.1) were prepared, and the shipment of delivering kits and collecting samples were organised by the author of this thesis. All the collected samples were stored in -80 °C until use, and recorded in a digital database which was set up and managed by the author of this thesis. To confirm the diagnosis of each patient, all the available study proforma were reviewed and missing information was chased by the author of this thesis with the help from Dr Pedro Machado and Ms Iwona Skorupinska.

2.1.2 Other Patients

Patients with tubular aggregates were identified by reviewing their muscle biopsy reports available through the MRC Centre for Neuromuscular Diseases and NHNN by the author of this thesis with the help from Professor Henry Houlden, Professor Janice Holton and staff from the Department of Pathology, ION. Cases from other centres/institutes were confirmed and kindly provided by the neurologists from Belfast Hospital in Northern Ireland, and Neuromuscular Unit, BioBank of Skeletal Muscle, Nerve Tissue, DNA and Cell Lines and Neuromuscular Diseases and Neuroimmunology Unit Muscle Cell Biology Lab in Italy. In addition, two siblings with cytoplasmic body myopathy, two unrelated patients with cylindrical spirals myopathy, two siblings with hIBM, and a family with a genetically undiagnosed myotubular myopathy from NHNN, and another patient with hIBM from Belgium were also selected for our investigations.

2.1.3 Samples Collection

2.1.3.1 Prospective Blood Collection

From each patient we obtained approximately 25 ml of whole blood on a single visit: a) 2 x 5 ml of blood for DNA extraction collected into ethylenediaminetetraacetic acid (EDTA) tubes (Becton, Dickinson and Company [BD] Vacutainer™, USA) and stored frozen, b) 1 x Tempus™ Blood RNA Tube (Applied Biosystems™, USA) (3 ml of blood) for mRNA analysis, c) 1 x lithium heparin tube (4.5 ml) (BD Vacutainer™, USA) for plasma and blood protein analysis, and d) 1 x SST™ (Serum Separation Tubes) II advance tube (BD Vacutainer™, USA) (5 ml) for serum (SST™ was prepared only for the centres with the facilities of doing serum extraction). All serum samples were extracted on the same day of collection in accordance with the manufacturer's instructions. The other samples were stored at -80 °C until use.

2.1.3.2 Tissue Samples

Stored flash frozen muscle tissues from patients were kindly provided by the Department of Pathology, ION and study collaborators when they were available. Material Transfer Agreement (MTA) documents for requesting DNA and tissue samples from biobanks in Italy were completed by the author of this thesis. Pathologically confirmed normal post-mortem frontal lobe cortex and muscle tissue from a control group were kindly provided by the Queen Square Brain Bank in London (brain tissues) and the MRC Sudden Death Brain and Tissue Bank in Edinburgh (muscle tissues), UK.

2.1.4 Ethical Approval

The studies included in this thesis were approved by the National Research Ethics Service (NRES) Committee London – Queen Square (REC reference: 12/LO/1557), London, UK. Study consent was obtained from all patients and/or family members.

2.2 Genetic Studies

2.2.1 DNA Extraction from Blood and Muscle Tissue

Genomic DNA (gDNA) extraction from blood samples was performed in the Neurogenetics Laboratory by the clinical diagnostic service team of the NHNN. This was performed using 5-10 ml of fresh or frozen whole blood samples using Flexigene kit (Qiagen, Germany) according to manufacturer's instructions.

gDNA extraction from muscle samples was performed by the Laboratory of the Government Chemist (LGC) Genomics (Germany) using the Sbeadex Tissue kit (LGC Genomics, Germany).

2.2.2 DNA Concentration and Purity

The concentration and quality of DNA was measured using a NanoDrop ND-1000 spectrophotometer following the manufacturer's instructions (NanoDrop Technologies, USA). Concentration was assessed at 260 nm. Purity was estimated by the 260/280 and 260/230 absorbance ratios, and the spectra of the ratios between 1.8-2.0 and 1.8-2.2 respectively were considered as DNA samples of the appropriate quality. The concentration of DNA was adjusted to 20 ng/µl or 50 ng/µl with autoclaved distilled H_2O (dH_2O).

2.2.3 Polymerase Chain Reaction (PCR) and Sanger Sequencing

2.2.3.1 Primer Design

The previously published primers of *APP* exons 16 and 17 were used (Levy et al., 1990) and the primers of *SQSTM1* and *VCP* were kindly provided by Rita Louro Guerreiro and Lee Darwent within the Department of Molecular Neuroscience, ION. Primers of other target candidate genes/exons analyses were designed using Primer3 software (Rozen and Skaletsky, 2000). Reference sequences of the human genome assembly build 37 (GRCh37) were obtained from Ensembl Genome Browser (http://www.ensembl.org/index.html). The transcripts of the genes chosen for primer design, PCR and sequencing reactions in this thesis are listed in Table 2-2. The

sequences of all primers that were used in this thesis are available in Appendix I – Table 2. Additional analyses of specific loci investigated by other methods are further detailed below.

Table 2-2 List of the transcripts used for sequencing of genes

Gene	Transcript	Gene	Transcript
STIM1	ENST00000616714	*ATP2A1*	ENST00000357084
ORAI1	ENST00000616379	*DPAGT1*	ENST00000354202
PGAM2	ENST00000297283	*ALG14*	ENST00000370205
SCN4A	ENST00000435607	*MYOT*	ENST00000239926
SBF1	ENST00000380817	*CASQ1*	ENST00000368078
SBF2	ENST00000256190		

2.2.3.2 PCR Reactions

To amplify the target region, a PCR mix was made of 12.5 µl of Roche PCR MasterMix (Roche Applied Science, Germany), 9.5 µl of autoclaved dH_2O, 1 µl of forward primer (10 µM), 1 µl of reverse primer (10 µM) and 1µl of DNA (20 ng/µl) for each reaction. PCR was performed on an Eppendorf Mastercycler thermal cycler at an optimal annealing temperature programme. The PCR conditions of standard 60-50 touchdown programme were as follows:

94 °C for 10 minutes
25 cycles of: 94 °C for 30 seconds
60 °C for 30 seconds (-0.4 °C/cycle)
72 °C for 45 seconds
12 cycles of: 94 °C for 30 seconds
50 °C for 30 seconds
72 °C for 45 seconds
72 °C for 5 minutes
hold at 4 °C.

2.2.3.3 Agarose Gel Electrophoresis for Inspection of PCR Products

Two percent agarose gel was prepared by adding 1.5 g of agarose powder (Roche Applied Science, Germany) to 75 ml of 1x Tris-Borate-EDTA (TBE) buffer (10x TBE buffer: 121.1 g Tris [Sigma-Aldrich, USA], 61.8 g anhydrous boric acid

[Merck-Millipore, USA], 7.4 g EDTA [VWR International, USA], made up with dH2O; pH to 8.3) and heating up the mixture to dissolution, and also adding 2.5 µl of ethidium bromide. After gel polymerisation in an electrophoresis tray, 3 µl of each PCR product mixed with 3 µl orange loading dye (60% glycerol, 40% dH2O, teaspoon of orange G powder [Sigma-Aldrich, USA]) were loaded in each well and run at 60 V for 30 minutes to confirm success of the PCR reaction. A 1 kb DNA ladder (Qiagen, Germany) was used to judge the size of the amplified fragments. Visualization of the bands was done under an Ultraviolet (UV) transilluminator, and digital photographs were taken using the Syngene GeneGenius image acquisition system and GeneSnap software (Synoptics, UK).

2.2.3.4 Purification of PCR Product

Five microliters of each PCR product was purified in 2 µl of a "home-made" enzyme mix "Exosap" (1 ml of "Exosap" was made of 50 µl of Exo I [Thermo Fisher Scientific, USA], 200 µl of Fast-AP [Thermo Fisher Scientific, USA] and 750 µl of autoclaved dH_2O) under the following conditions:

37 °C for 30 minutes
80 °C for 15 minutes
hold at 4 °C.

2.2.3.5 Sequencing Reaction

For each PCR product, a sequencing reaction mix was prepared using 4.5 µl autoclaved dH_2O, 2 µl of BigDye® Terminator v3.1 5x sequencing buffer (Applied Biosystems, USA), 1 µl of forward/reverse primer (10 µM), 0.5 µl of BigDye® Terminator v3.1 (Applied Biosystems, USA), and 2 µl of clean PCR product. The reaction was run under the following conditions:

25 cycles of: 96 °C for 10 seconds
50 °C for 5 seconds
60 °C for 4 minutes
hold at 4 °C.

2.2.3.6 Clean-Up of the Sequencing Reaction

Sequencing products were purified using Corning® FiltrEX™ 96 well filter plates (Sigma) with Sephadex columns. Sephadex was manually prepared per plate as follows: 40 ml of autoclaved dH_2O was mixed thoroughly with 2.9 g of Sephadex G-50 Bioreagent (Sigma-Aldrich, USA) and left to hydrate for at least 30 minutes at room temperature. Subsequently, 350 µl of the mixture was added to each well of the plate. Each plate was spun at 700xg for 3 minutes on top of an empty collection plate. All 10 µl of the sequencing products were added to the centre of the columns and spun at 910xg for 5 minutes on top of a clean, labelled PCR plate for collection of the purified sequencing products.

2.2.3.7 Sequencing Analysis

The purified samples were run on an ABI3730XL Genetic Analyzer (Applied Biosystems, USA) immediately. Sequencing data were manually analysed using SeqScape v2.5 software (Applied Biosystems, USA).

2.2.4 Mutation Nomenclature

Standard nomenclature from the Human Genome Variation Society (http://www.hgvs.org/mutnomen/) was used to name the variants described in this thesis based on published cDNA sequences. Variants described for the first time in the text were written with both the nucleotide and the amino acid change, denoted by c. and p. respectively. Subsequently, only the amino acid change was used.

2.2.5 Restriction Endonuclease Analysis

Restriction fragment length polymorphism (RFLP) was used for *APOE* genotyping. Primers were used as by Ingelsson and colleagues (Ingelsson et al., 2003), and are available in Appendix I – Table 2. The PCR reaction was done in a total volume of 20 µl containing 10 µl of Roche PCR MasterMix (Roche Applied Science, Germany), 4 µl of autoclaved dH_2O, 2 µl of forward primer (10 µM), 2 µl of reverse primer (10 µM), 1 µl of 5% dimethylsulfoxide (DMSO, Sigma-Aldrich, USA), and 1 µl of DNA (50 ng/µl) for each reaction. The PCR conditions used were as follows:

94 °C for 1 minute
15 cycles of: 94 °C for 30 seconds
58 °C for 30 seconds
72 °C for 30 seconds
16 cycles of: 94 °C for 30 seconds
58 °C for 30 seconds (-0.3 °C per cycle)
14 cycles of: 94 °C for 30 seconds
52 °C for 30 seconds
72 °C for 30 seconds
72 °C for 5 minutes
hold at 4 °C.

The PCR products were resolved on a 2% agarose gel as described previously for amplification inspection.

The RFLP reaction was carried out in a total volume of 20 µl containing 5 µl of *APOE* PCR product, 0.5 µl of Hin6I restriction endonuclease (10 U/µl, Thermo Scientific, USA), 2 µl of 10x Buffer Tango (Thermo Scientific, USA), and 12.5 µl of autoclaved dH_2O for each reaction. The RFLP mixture was incubated at 37 °C for 4 hours. The *APOE* RFLP products were then run on a 4% agarose gel (10 µl of RFLP product and 2 µl of 6x orange loading dye) at 125 V for 80 minutes. GelPilot Midrange Ladder (100 bp, Qiagen, Germany) was used to judge the size of the fragments. The *APOE* genotypes were determined by visualizing the band patterns in the gel (size and number of bands) as shown in Figure 2-1 (Ingelsson et al., 2003). (Chapter 3)

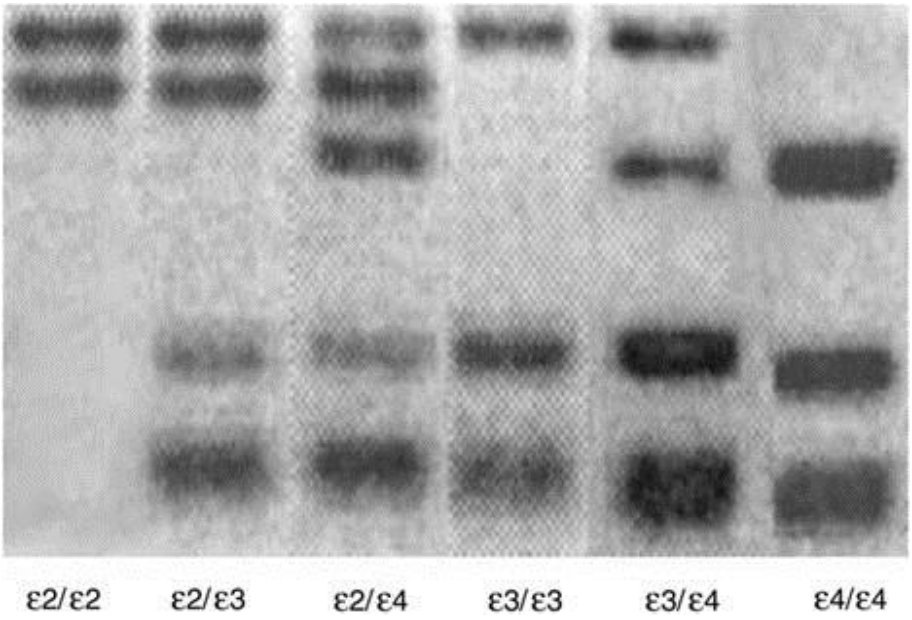


Figure 2-1 *APOE* genotyping by restriction fragment length polymorphism analysis

The band patterns can be transferred into *APOE* genotypes (Adapted from (Ingelsson et al. 2003)).

2.2.6 *MAPT* Genotyping

An analysis of the 238 bp *MAPT* haplotype (H2) deletion in intron 9 upstream of *MAPT* gene exon 10 was performed as described previously (Baker et al., 1999) to determine the *MAPT* haplotypes by visualizing PCR products on an agarose gel. Primer sequences were as reported in the above study, and are available in Appendix I – Table 2. (Chapter 3)

2.2.7 *PRNP* Genotyping at Codon 129

An Allelic Discrimination Assay for the 7500 Fast Real-Time PCR System (Applied Biosystems, USA) was used to determine *PRNP* common methionine/valine (M/V) polymorphism at codon 129. Two TaqMan® MGB probes, fluorescent FAM™ dye and fluorescent VIC® dye (Applied Biosystems, USA), were used to target a SNP site for M or V, respectively. The PCR mix was made to a total volume of 5 µl containing 2.5 µl of TaqMan® SNP Genotyping Assay (Applied Biosystems, USA), 0.05 µl of forward primer (90 µM), 0.05 µl of reverse primer (90 µM), 0.05 µl of probe FAM™ (20 µM), 0.05 µl of probe VIC®, 1.3 µl of autoclaved dH_2O, and 1 µl of DNA (20 ng/µl). The prepared PCR mix was run on the 7500 Fast Real-Time PCR System in UCL MRC Prion Unit. The real-time PCR data was analysed using the

7500 Fast System Software v2.0.6 (Applied Biosystems, USA). The Genotyping Assay and the *PRNP* primers and probes were kindly provided by Gary Adamson from UCL MRC Prion Unit, and are available in Appendix I – Table 2. (Chapter 3)

2.2.8 Fragment Analysis

2.2.8.1 C9orf72 Repeat-Primed PCR (RP-PCR) Analysis

One hundred eighty-seven sIBM patients were screened for the GGGGCC repeat expansion in *C9orf72* gene using a repeat-primed PCR (RP-PCR) as described previously (DeJesus-Hernandez et al., 2011), and are available in Appendix I – Table 2. Each PCR reaction contained 12 µl of Extensor Long PCR Master Mix, Buffer 1 (Thermo Scientific, USA), 10 µl of Betaine (Thermo Scientific, USA), 2 µl of primer mix (0.33 µM of F and R2 primers, and 0.033 µM of R1 primer), and 1 µl of DNA (50 ng/µl). The PCR conditions were as follows:

98 °C for 10 minutes
10 cycles of: 97 °C for 35 seconds
53 °C for 2 minutes
68 °C for 2 minutes
25 cycles of: 97 °C for 35 seconds
53 °C for 2 minutes
68 °C for 2 minutes (-0.3 °C per cycle)
68 °C for 10 minutes
hold at 4 °C.

Following this, a mixture containing 2 µl of the PCR products, 0.3 µl of Liz 500 size standard (Applied Biosystems, USA) and 9.2 µl of HiDi formamide (Applied Biosystems, USA) for each sample was denatured at 95 °C for 3 minutes, and put immediately on ice for 5 minutes. This mixture was then run on an ABI3730XL Genetic Analyzer (Applied Biosystems, USA). The presence/absence of the GGGGCC repeat expansion in *C9orf72* and the approximate repeat size for each sample were determined using GeneMapper v3.7 (Applied Biosystems, USA) software. (Chapter 3)

2.2.8.2 TOMM40 Gene Intronic Polymorphism Analysis

Fluorescence-based fragment size analysis was performed for the polyT repeat in intron 6 of *TOMM40* gene (rs10524523). The primers were as previously described (Cruchaga et al., 2011), and are available in Appendix I – Table 2. The PCR reaction contained 10 µl of Roche PCR MasterMix (Roche Applied Science, Germany), 7 µl of autoclaved dH_2O, 1 µl of forward primer (10 µM), 1 µl of reverse primer (10 µM) and 1 µl of DNA (20 ng/µl) for each sample. The PCR conditions used were as follows:

94 °C for 10 minutes
25 cycles of: 94°C for 30 seconds
67 °C for 30 seconds (-0.2 °C per cycle)
72 °C for 45 seconds
15 cycles of: 94 °C for 30 seconds
62 °C for 30 seconds
72 °C for 45 seconds
72 °C for 10 minutes
hold at 4°C.

The PCR product was mixed with Liz 500 size standard and HiDi formamide, and analysed in the same way as for *C9orf72* repeat previously described. The genotypes of the polyT repeat alleles were classified as per Cruchaga et al.: short [S] (247-266 bp), long [L] (267-276 bp), and very long [VL] (277-281 bp). The base pair numbers do not correspond to the number of (T)n repeats, but refer to the total length of the amplified PCR fragment. The expected length of the PCR product was N+247 base pairs [bp], where "N" represents the number of the polyT residues, and 247 bp was the size of polyT flanking region plus an "A" overhang at the end of the product. (Chapter 3)

2.2.9 Long PCR for mtDNA Deletion(s) Detection

Long PCR was performed on 57 sIBM DNA samples extracted from muscle tissues for large-scale rearrangement detection of mtDNA. The complete mitochondrial genome was amplified in two fragments of 11529 bp (11 kb) and 6706 bp (5 kb). The

primer sequences were kindly provided by Ese Mudanohwo from the NHNN Neurogenetics Laboratory, and are available in Appendix I – Table 2. They were designed to lie close to the origins of replication and so should not be deleted. The PCR reaction consisted of Mix1 containing 2.5 µl of each dNTP (10 mM, Applied Biosystems, USA), 0.4 µl of each primer (100 µM), 1 µl of DNA (50 ng/µl), and 13.2 µl of autoclaved dH_2O; and Mix2 containing 0.75 µl of Expand Polymerase (Roche Expand Template Long PCR kit, Roche Applied Science, Germany), and 5 µl of 10x Buffer3 (the buffer from the kit was defrosted at 37 °C for 1-2 hours and was mixed thoroughly prior to use), and 19.25 µl of autoclaved dH_2O for each sample. A suitable negative control for the 5 kb and a negative and a positive deletion control for the 11 kb must be used in long PCR analysis. The PCR conditions were as follows:

94 °C for 3 minutes
10 cycles of: 94 °C for 10 seconds
60 °C for 30 seconds
68 °C for 10 minutes
25 cycles of: 94 °C for 10 seconds
60 °C for 30 seconds
68 °C for 10 minutes (+20 seconds per cycle)
68 °C for 10 minutes
hold at 4 °C.

Following this, 10 µl of L5647/H607 PCR product and 5 µl of L15788/H5925 PCR product were resolved on a 0.8% agarose gel with a 1 kb ladder (Progema, USA) to identify any deletion in the mtDNA genome. (Chapter 3)

2.3 Whole-Exome Sequencing

2.3.1 Sample Assessment and Preparation

The sample gDNA concentration was measured by using the Qubit Fluorometry system with dsDNA Broad Range (BR) Assay (Invitrogen, USA) and the Qubit 2.0

Fluorometer (Invitrogen, USA). Sample concentration was adjusted to 5 ng/µl and a total amount of 100 ng was used for library preparation.

2.3.2 Library Enrichment

Sequence capture, enrichment, and elution were performed according to the instructions and protocols of Illumina's TruSeq Exome Enrichment (62 Mb) (before 2014) or Nextera Rapid Capture Exome (37 Mb) (since 2014). The samples were run in-house on the Illumina HiSeq2500 (Illumina, USA) for whole-exome sequencing (WES). This was performed by Ms Deborah Hughes at the UCL Institute of Neurology (ION).

2.3.3 Bioinformatics Pipeline for Data Processing

The primary data analysis was performed in-house by Dr Alan Pittman at the UCL ION. Sequence reads from FASTQ files were mapped to the reference human genome (University of California, Santa Cruz [UCSC] hg19) using NovoAlign (Novocraft, Malaysia) and generating Binary Alignment/Map (BAM) files using the Sequence Alignment/Map tools (SAMtools) (Li et al., 2009) (http://samtools.sourceforge.net/). PCR duplicates, as well as reads without a unique mapping location, were removed by Picard tool (http://picard.sourceforge.net). Local realignment of insertions/deletions (indels) and variant calling were performed by Genome Analysis Toolkit (GATK) (DePristo et al., 2011) (https://www.broadinstitute.org/gatk/). Variant quality score recalibration (VQSR) was calculated according to the GATK Best Practices documentations. This is to assign a well-calibrated probability (VQSRLOD: variant quality score log-odds) to each variant call in a call set, and to filter based on the estimation for the accuracy of each call. The GenomeBrowse® software (Golden Helix, USA) was used to inspect the reads depths and the sequencing coverage. The variants were then annotated using the Annotate Variation software (ANNOVAR) (Wang et al., 2010) for gene names, genomic function, the transcript IDs, amino acid changes, segmental duplication, the Single Nucleotide Polymorphism Database (dbSNP) Version 137 (http://www. ncbi.nlm.nih.gov/SNP/) (Sherry et al., 2001), 1000 Genomes (Genomes Project et al., 2015) (http://www.1000genomes.org/), Exome Variants Server (EVS)

(http://evs.gs. washington.edu/EVS/), the Exome Aggregation Consortium (ExAC) Version 0.2 (http://exac. broadinstitute.org) databases, Genomic Evolutionary Rate Profiling (GERP++) (Davydov et al., 2010), Complete Genomics database (CG69) (Drmanac et al., 2010), Online Mendelian Inheritance in Man (OMIM) IDs (Amberger et al., 2015) (http://omim.org/), CliniVar database (Landrum et al., 2016), and Combined Annotation Dependent Depletion (CADD) score (Kircher et al., 2014). The variants were also predicted for the pathogenicity using *in silico* tools including SIFT (Sort Intolerant from Tolerant) (Kumar et al., 2009), PolyPhen2 (Polymorphism Phenotyping v2) (Adzhubei et al., 2010), and MutationTaster (Schwarz et al., 2014).

For the whole-exome association analysis, a series of additional quality control techniques were performed to remove the poor quality samples and variants. The detailed process will be described in Section 4.2.

2.3.4 Variants Selection for Candidate Gene Analysis

Variant filtering was carried out by the author of this thesis. A general filtering strategy for rare variants in candidate genes was suggested to exclude the variants present in public databases (1000 Genomes project, EVS and ExAC) with a minor allele frequency (MAF) ≥ 1% or in a proportion of similarity between the duplication region > 0.97. Synonymous variants and non-coding variants were also excluded. The variants were also removed if they have a MAF ≥ 1% in an ION exome database (it contains all the variants identified in the samples exome sequenced in ION), as they are likely to be artefacts/false positives generated in the sequencing platform. The variant filtering was also based on the mode of inheritance of the phenotype. For an autosomal recessive inheritance, variants were filtered for homozygous or compound heterozygous in candidate genes; while for the autosomal dominant format, variants were filtered for heterozygous in candidate genes. If the phenotype was suggestive of X-linked inheritance, the filtering focused on variants on chromosome X. For sporadic cases, variants were filtered for either recessive mode or *de novo* (variants present in patients but not in the parents) in candidate genes. (Figure 2-2) Pathogenicity scores for the selected variants were predicted by *in silico* tools. The predicted deleterious variants were subsequently analysed for segregation

among other affected/unaffected family members. All the selected variants and segregation required to be confirmed by Sanger sequencing. Specific filtering strategies will be highlighted in each chapter.

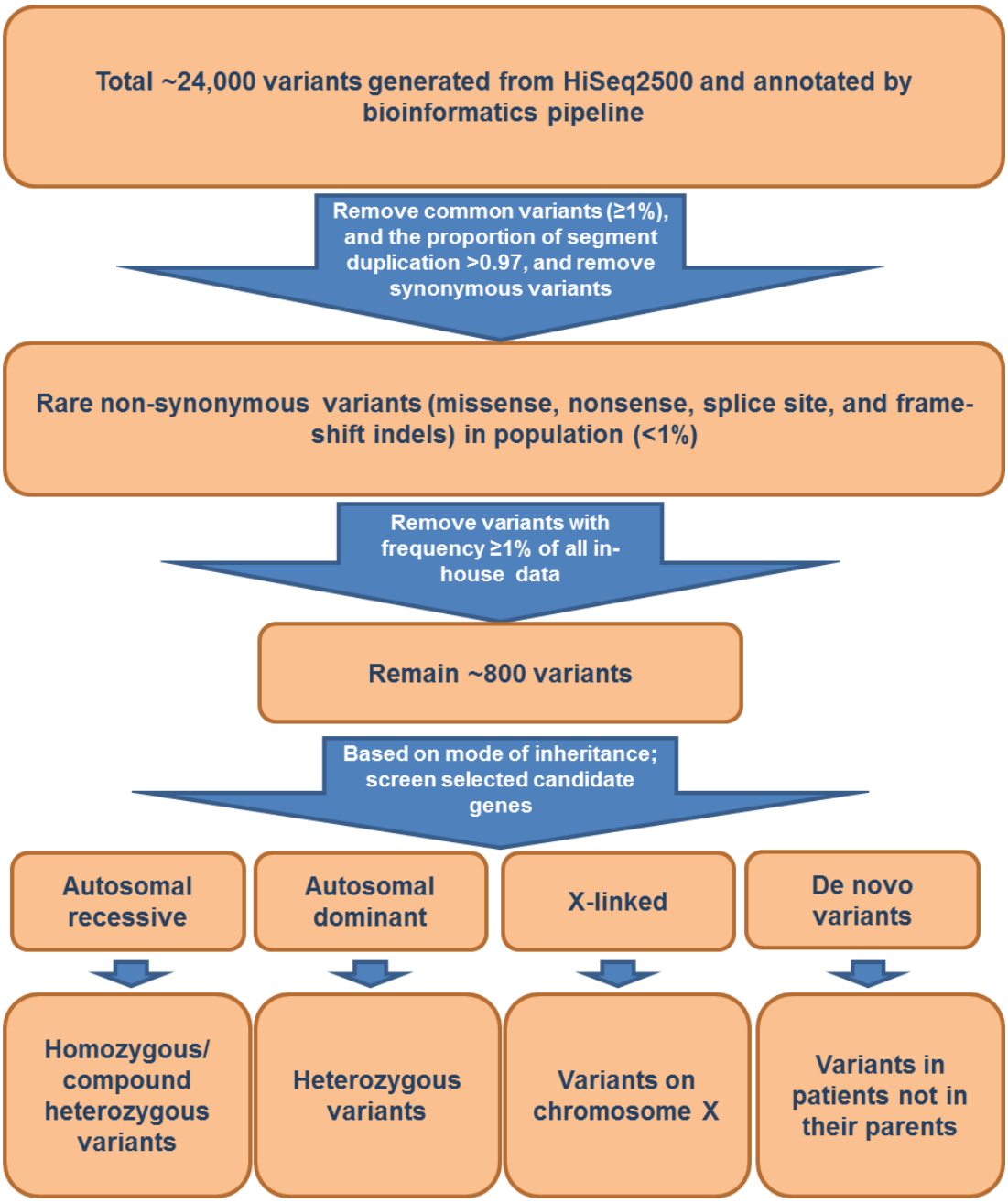


Figure 2-2 A general WES data filtering strategy for rare variants in candidate genes

2.4 Whole-Genome Expression Study

2.4.1 RNA Extraction from Muscle Tissues

Total RNA was isolated from frozen muscle tissue using the miRNeasy kit (Qiagen, Germany). Muscle tissue was cut into a small piece at approximate 10 mg for each sample and placed in a sterile 2 ml tube. After adding 700 µl of QIAzol (Qiagen, Germany), the tissue sample was disrupted and homogenized using TissueRuptor (Qiagen, Germany), and then incubated at room temperature for a minimum of 5 minutes. A volume of 140 µl (20% of QIAzol) of chloroform was added to the homogenized sample and vortexed vigorously for 30 seconds. After incubating at room temperature for 10 minutes and centrifuging at 1200xg at 4 °C for 25 minutes, the aqueous phase (upper layer) of the sample was transferred to a sterile 2 ml RNase free tube. The following steps of RNA extraction were according to the manufacturer's protocol. RNA was eluted and suspended in 30 µl of RNase-free water, and stored at -80 °C until use. Concentration and purity were measured as described in Section 2.2.2 with altered absorbance ratios for RNA. Samples of adequate quality were expected to have a 260/280 ratio of > 2.0, and the 260/230 ratio in the range of 2.0-2.2. The RNA integrity number (RIN) of each RNA sample was assessed using Agilent 2100 Bioanalyzer (Agilent Technologies, USA) with Agilent RNA 6000 Nano kit (Agilent Technologies, USA) according to the manufacture's protocol. The RIN above 6 was considered acceptable for microarray.

2.4.2 Microarray Analysis

Whole-genome expression profiling (total RNA at 600 ng was used) was performed using the Illumina HumanHT-12 v4 Expression BeadChip (Illumina, USA) at UCL Genomics. Raw expression data were log2 transformed and quantile normalized, and differential expression analysis (subgroups of patients vs. controls and all the patients vs controls) was performed using R software the limma Bioconductor package (Ritchie et al., 2015). The scripts were kindly provided by Dr Conceição Bettencourt (Appendix I – Script section). Genes were considered differentially expressed, and used in further analysis, when false discovery rate (FDR) adjusted *p*-value was lower than 0.05, and absolute log2 fold-change was higher than 0.2. Functional enrichment

analysis for Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and Human Phenotype Ontology (HPO) terms was performed using g:Profiler (biit.cs.ut.ee/gprofiler/). The selected up-regulated and down-regulated genes were validated using real-time quantitative PCR (RT-qPCR).

2.4.3 Real-Time Quantitative PCR (RT-qPCR)

Total RNA (600 ng) from each sample was reverse transcribed into complementary deoxyribonucleic acid (cDNA) using random primers from High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, USA). Three replicates per sample were assayed for each target gene in a 384-well format plate using Fast SYBR Green PCR Kit (Applied Biosystems, USA). The primers for *HLA-A* and *HLA-DRA* cDNA levels were used from papers by Ivanidze et al. and by Gersuk et al., respectively (Gersuk and Nepom, 2009, Ivanidze et al., 2011). The primers for *CD74* and *PPIA* (cyclophilin) cDNA levels were designed using Primer3 software and validated by the author of this thesis. The sequences of the primers are available in Appendix I – Table 2. QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, USA) was used and the protocol in the QuantStudio™ 6 Software Guide was followed for running the Standard Curves programme and the Comparative Threshold Cycle (Ct) programme. *PPIA* gene was selected as a reference gene to normalize the expression levels of the genes of interest, as a tendency for down-regulation of a widely used reference gene glyceraldehyde-3-phosphate dehydro-genase (*GAPDH*) was observed in IBM compared with controls by our group. The comparative Ct method, also known as the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008), was used for Ct normalization for each gene and determination of fold-changes in gene expression between patients and controls.

The methods described above in Section 2.4 were used in Chapter 3.

2.5 Protein Biochemistry for Fibroblasts Cell Lines

2.5.1 Cell Harvesting for RNA Extraction

Fibroblasts were generated from skin biopsies taken with informed consent from the affected patient, unaffected parents and unaffected controls. Cells were grown in 10 cm^2 plates and incubated at 37 °C until confluent. Skin biopsies were performed by Professor Henry Houlden. Generation of fibroblasts and all the cell harvesting were performed by Mr Chris Lovejoy at the UCL ION. Two millilitres of 0.05% 1x Trypsin-EDTA (Life Technologies, Thermo Fisher Scientific, USA) were added to a confluent 10 cm^2 plate of fibroblasts. The plate was then placed in the incubator at 37 °C for 5 minutes to ensure sufficient detachment. Subsequently, cells were gently scrapped from the plate with a cell scraper and transferred to a new sterile Eppendorf tube. After spinning down the pellet by centrifugation at 4000 rpm for 1 minute, the cell suspension was transferred to another new sterile Eppendorf tube, and used straightway or kept at -80 °C.

2.5.2 RNA Extraction from Fibroblasts Cell Lines

Total RNA was extracted from fibroblasts using the RNeasy Mini kit (Qiagen, Germany). A volume of 600 µl of RLT lysis buffer from the kit and 600 µl of 70% ethanol was added to the Eppendorf tube of cell suspension from the step above and mixed well by pipetting up and down. Subsequent steps in RNA extraction followed the manufacturer's protocol. RNA was eluted and suspended in 40 µl of RNase-free water, and then stored at -80 °C until use. Concentration, purity and integrity were measured as described in Section 2.4.1.

2.5.3 cDNA Synthesis and Sequencing

cDNA was synthesized using 1 µg total RNA, random primers and MultiScribe reverse transcriptase according to the High-Capacity cDNA Reverse Transcription Kits' protocol (Applied Biosystems, USA). cDNA was stored at -20 °C until used.

Two microliters of the cDNA product was used as a template. PCR and sequencing reactions were carried out as described in Section 2.2.3 with *SBF1* and *SBF2* primers designed to only amplify the specific region in cDNA and not gDNA. The sequences of the primers are available in Appendix I – Table 2.

2.5.4 Cell Harvesting for Western Blotting

Fibroblasts were harvested as described in Section 2.5.1 to confluence. After removing the media, the plate was washed with 1 ml of 1x Dulbecco's phosphate buffered saline (DPBS) buffer (Thermo Fisher Scientific, USA) twice and the buffer was discarded. After this, 300 µl of lysis buffer (50 µM Tris pH8, 150 µM NaCl, 0.5% nonyl phenoxypolyethoxylethanol [$Np40^+$], 1 µM EDTA and 1x complete protease inhibitor cocktail [Roche Applied Science, Germany]) was added to the surface of the plate and the cells were scraped with a cell scraper. Subsequently, cell lysates were frozen at -80 °C for further cell lysis and centrifuged down to remove cell debris at maximum speed for 10 minutes. The supernatant was transferred to a new cold labelled Eppendorf tube stored at -80 °C until use. The pellet was stored at -80 °C.

2.5.5 Protein Estimation

The supernatant obtained in Section 2.5.4 was thawed on ice and centrifuged quickly at 4 °C before use. One microliter of each sample was pipetted in triplicate on a 96-well plate with bovine serum albumin (BSA) (New England BioLabs, UK) standard ranging from 0-5 µg/µl. One microliter of $Np40^+$ lysis buffer was added to each well of BSA standard as background signal. Protein concentration was determined by RC DC^{TM} Protein Assay (Bio-Rad Laboratories, USA) following its protocol.

2.5.6 Western Blot

Samples were adjusted to the same concentration and made up to 50 µl with 12.5 µl 4x sodium dodecyl sulphate (SDS) buffer (Invitrogen, USA), 0.5 µl 1 M DL-Dithiothreitol (Invitrogen, USA) and the necessary volume of $Np40^+$ lysis buffer. All the samples were denatured at 75 °C for 10 minutes. Forty microliters of each sample,

along with 12 µl of a HiMark™ protein standard (Novex) was loaded onto a 10-well 3-8% Tris-Acetate NuPAGE gel (Invitrogen, USA), and run at 50 V for three hours and then at 70 V for one hour in 1x Tri-Acetate running buffer (45 ml 20x NuPAGE Tri-Acetate running buffer [Invitrogen, USA] in 855 ml dH_2O). The samples were then transferred onto a nitrocellulose membrane at 35 V for two hours in 1x transferring buffer (100 ml 10x Tris-Glycine electroblotting buffer [National Diagnostics], 200 ml methanol, and 700 ml dH_2O). The membrane was immersed in Ponceau red (Sigma-Aldrick, USA) to verify protein transfer, and then blocked with 5% milk in 20 ml 1x tris-buffered saline (TBS)-Tween20 buffer (100 ml TBS 10x Solution [Thermo Fisher Scientific, USA] + 900 ml dH_2O + 1 ml TWEEN 20 [Sigma-Aldrich, USA]) for one hour at room temperature. The membrane was then incubated overnight at 4 °C with the myotubularin related 5 (MTMR5) primary antibody (Santa Cruz, USA) at a dilution of 1/250 with 5% of BSA in TBS-Tween20 buffer (5 g BSA [Sigma-Aldrick, USA] + 100 ml 1x TBS-Tween20).

On the following day, the membrane was washed with TBS-Tween20 buffer for 10 minutes three times, and then incubated with the secondary goat anti-mouse immunoglobulin G (IgG)-horseradish peroxidase (HRP) antibody (Santa Cruz, USA) at a dilution of 1/5000 with 1% milk in TBS-Tween20 buffer for one hour at room temperature. After three further washes with TBS-Tween20 buffer, an Electrochemiluminescence (ECL) Chemiluminescence kit (Thermo-Scientific Fisher, USA) with each ECL reagent at a ratio of 1:1 was used as a substrate to detect the protein of interest, developed onto a Super Rx X-ray film (Fujifilm, Japan) in an Amersham autoradiography cassette (GE Healthcare, USA) in a dark room.

The membrane was then washed with TBS-Tween20 buffer three times again and incubated for 30 minutes at room temperature with β-actin primary antibody (Sigma-Aldrich, USA) which was diluted at 1/50,000 with 1% milk in TBS-Tween20 buffer. The remaining process was as described above. Beta-actin was used as a loading control for normalising the protein of interest. The analysis was performed with the ImageJ software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2016).

The methods described in Section 2.5 were used in Chapter 6.

2.6 Muscle Biopsy Pathology

The available slides of muscle biopsies discussed in this thesis were kindly provided from the Department of Pathology, ION and collaborators in Gent University. These slides were further reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

2.7 Statistical Analyses

Hardy-Weinberg Equilibrium (HWE) was tested for the distributions of *APOE* genotypes, *MAPT* genotypes, *PRNP* genotypes in both sIBM patient and control cohorts using Arlequin Ver3.5.1.3. (Excoffier L, 2010). Statistical analyses were performed using online software OpenEpi Version2.3.1 (Dean AG) for comparing the frequencies of *MATP* genotypes, *PRNP* genotypes, and variants in *SQSTM1* and *VCP* genes between sIBM patient and control cohorts in Chapter 3.

Stata 12 (StataCorp LP, USA) was used for the study of *APOE-TOMM40* in sIBM discussed in Chapter 3, Section 3.1. Genotypic frequencies of *APOE* and *TOMM40* polyT between the sIBM patients and control cohorts were compared using the chi-square test or the Fisher's exact test. The associations between *APOE* genotypes, *TOMM40* polyT genotypes, ethnicity, gender and age of onset of sIBM were analysed by linear regression analyses. Gender, tissue and ethnicity were used as covariates for the adjusted analysis in linear regression models.

SPSS Statistics 22 (IBM, USA) was used to perform Mann-Whitney U-Test to analyse the difference in relative gene expression between patient groups and controls from RT-qPCR in Chapter 3, Section 3.2.

PLINK software version 1.9 (https://www.cog-genomics.org/plink2) (Chang et al., 2015) and R 3.1.1 (https://www.r-project.org/) were used for the statistical analyses in whole-exome association study. The methods will be described in detail in Chapter 4.

For all the analyses, p-value < 0.05 was considered statistically significant, unless specified otherwise in each chapter.

Chapter 3 Candidate Gene Screening in sIBM Patients

3.1 Overview of Sample Collections Based on International IBM Genetics Consortium (IIBMGC) and Muscle Study Group

By the 11th March 2016, the IIBMGC and Muscle Study Group had members from 17 specialised centres in seven countries around the world. The detailed sample collections from each centre are listed in Table 3-1. Information of members of IIBMGC and Muscle Study Group are available in Appendix II – Table 1.

Table 3-1 IBM sample collections of IIBMGC and Muscle Study Group (by 11/03/2016)

Institutions/ Research centres (Key participants)	Countries	Status	No. of samples
National Hospital of Neurology and Neurosurgery, Queen Square, London (Q. Gang, C. Bettencourt, P. Machado, J. Holton, M. Parton, H. Houlden, M. Hanna)	UK	Active	80
John Radcliffe Hospital, Oxford (S. Brady, D. Hilton-Jones)	UK	Active	59
MyoNet, Manchester (J. Lilleker, H. Chinoy)	UK	Active	115
University of California, LA (UCLA) (P. Shieh)	USA	Active	25
Nerve and Muscle Center of Texas (A. Shaibani)	USA	Active	15
Kansas University Medical Center (R. Barohn, M. Dimachkie)	USA	Active	13
Penn State Hershey Medical Center (M. Wicklund)	USA	Active	16
The Ohio State University Wexner Medical Center (A. Bartlett, J. Kissel)	USA	Active	9
Brigham & Women's Hospital, Harvard (A. Anthony)	USA	Active	8
Queen Elizabeth II Medical Centre, Western Australia	Australia	Active	25

(M. Needham)			
Royal North Shore Hospital, New South Wales (C. Liang)	Australia	Active	18
Neuromuscular Diseases and Neuroimmunology Unit Muscle Cell Biology Lab (M. Mora)	Italy	Active	19
Neuromuscular Unit, BioBank of Skeletal Muscle, Nerve Tissue, DNA and Cell Lines (M. Ripolone, M. Moggio)	Italy	Active	9
University Hospital Ghent (B. De Paepe and J. De Bleecker)	Belgium	Active	4
Medical School of the University of São Paulo (E. Zanoteli)	Brazil	Active	7
National University of Athens Medical School, Athens (M. Dalakas)	Greece	Active	18
University of California, Irvine (T. Mozaffar)	USA	Active	To be collected
Total			440

There were a total of 171 DNA samples, blood samples (2 x EDTA, 1 x Tempus^{TM} RNA, 1 x plasma; serum samples were available from 36 cases) from 214 cases, and muscle samples from 100 cases. Forty-four cases had both blood and muscle samples available. The sample collection and data management were performed by the author of this thesis with the help from Dr Pedro Machado and Iwona Skorupinska from MRC Centre for Neuromuscular Diseases. The detailed information was provided in Sections 2.1.1 – 2.1.3.

From May 2013 to January 2015, a total of 204 DNA samples were sent for whole-exome sequencing and analysed by a bioinformatics pipeline based in ION (as described in Section 2.3). The available complete study proforma of each patient was reviewed; one case with a known pathogenic homozygous *GNE* mutation identified in whole-exome data and one of a pair of identical twins were excluded. This left a total number of 239 patients who fulfilled the study criteria of sIBM at the time of writing; 181 of which had whole-exome data available. The classifications of fulfilled diagnostic criteria are summarised in Table 3-2.

Table 3-2 Summary of the diagnostic classification of 239 sIBM patients

Diagnostic criteria	Sub-classification	No. of cases
Griggs Criteria	Definite IBM	42
	Possible IBM	140
ENMC Criteria	Definite IBM	143
	Probable IBM	40
MRC Centre for Neuromuscular Diseases	Pathologically defined IBM	34
Diagnostic Criteria	Clinically defined IBM	123
	Possible IBM	60

Candidate gene studies described in this chapter and the whole-exome association study described in Chapter 4 were based on this cohort of 239 sIBM patients. The different sample size in each study was due to the different number of available samples at the time of analysis. Patient demographics of each study cohort is summarised in Table 3-3.

Table 3-3 Patient demographics of each study cohort in Chapters 3 and 4

Study	No. of sIBM (% of the whole cohort)	No. of Caucasians (% of the study cohort)	No. of males (% of the study cohort)	Mean age of onset (years)	No. of blood DNA	No. of muscle DNA
APOE-TOMM40	158 (66.1%)	142 (89.9%)	105 (66.5%)	59.7±9.7	126	32
Whole-exome dataset	181 (75.7%)	150 (82.9%)	119 (65.9%)	59.6±9.6	124	57
MAPT	68 (28.5%)	56 (82.4%)	47 (69.1%)	59.1±10.4	68	-
PRNP	44 (18.4%)	38 (86.4%)	29 (65.9%)	57.5±10.6	44	-
APP	69 (28.9%)	57 (82.6%)	48 (69.6%)	59.5±10.1	69	-
C9orf72	187 (78.2%)	152 (81.3%)	124 (66.3%)	58.5±10.0	164	56
mtDNA	50 (20.9%)	49 (98.0%)	32 (64%)	61.4±10.7	-	50
Total cohort	239 (100%)	195 (81.6%)	159 (66.5%)	59.5±9.9	209	68

3.2 The Effects of an Intronic Polymorphism in *TOMM40* and *APOE* Genotypes in sIBM

3.2.1 Introduction

Due to the many similarities between AD and sIBM, such as the late age of onset, abnormal accumulation of similar proteins and mitochondrial dysfunction, genetic investigations in sIBM have focused on known susceptibility genes for AD (Gang et al., 2014). The Apolipoprotein E (*APOE*, OMIM#107741) ε4 allele has been confirmed as a strong risk factor for developing LOAD (Saunders et al., 1993). However, it has not been shown to be associated with sIBM disease risk (Garlepp et al., 1995, Harrington et al., 1995, Askanas et al., 1996a, Love et al., 1996, Gossrau et al., 2004, Needham et al., 2008b), despite ApoE protein deposits having been identified in both sIBM and hIBM muscle fibres (Askanas et al., 1994, Mirabella et al., 1996).

"Translocase of Outer Mitochondrial Membrane 40" (*TOMM40*, OMIM#608061) is a gene adjacent to and in strong linkage disequilibrium with the *APOE* locus on chromosome 19 (Yu et al., 2007). It encodes the mitochondrial pore protein Tom40 involved in the transport of Aβ and other proteins into mitochondria (Hansson Petersen et al., 2008). An intronic polymorphism of a variable length polyT repeat within the gene (rs10527454, reported as rs10524523 in previous studies) was firstly reported as having a significant influence on the age of onset of LOAD (Roses et al., 2010, Maruszak et al., 2012). This gene was also recently studied in sIBM. A study of 90 Caucasian sIBM patients demonstrated that among carriers of the *APOE* genotypes ε3/ε3 or ε3/ε4, carriage of very long polyT repeats in *TOMM40* was less frequent in sIBM compared to controls and was additionally associated with a later age of onset of symptoms (Mastaglia et al., 2013). These findings suggest that these polymorphisms may be involved in genetic susceptibility to sIBM and may have an effect on the age of onset.

To further investigate the previously reported association between *APOE-TOMM40* and sIBM, *APOE* and *TOMM40* polyT repeat polymorphism were investigated for

their association with disease risk and the age of onset of sIBM in a cohort of 158 sIBM cases and 127 controls.

3.2.2 Methods

Of the 158 sIBM cases (66.1% of the total 239 confirmed sIBM patients), 89 DNA samples were extracted from blood, while 69 DNA samples were extracted from muscle tissue. The control group comprised 127 individuals with no history of neuromuscular disease. DNA samples from these individuals were extracted from pathologically confirmed normal post-mortem frontal lobe cortex or muscle tissue. The examinations of CK, muscle biopsy or neurophysiology of the control individuals are not available.

Restriction fragment length polymorphisms for *APOE* genotyping and fluorescence-based fragment size analysis for *TOMM40* intron 6 polyT repeat genotyping were performed by the author of the thesis as described in Sections 2.2.5 and 2.2.8.2.

3.2.3 Results

3.2.3.1 Patient Demographics

Of the entire cohort of 158 sIBM patients, 142 (89.9%) were Caucasians (66.2% of which were male), and 16 were of other ethnicities (68.8% of which were male), including Oriental Asians, Asian/American Indians and Black Africans. Age of onset ranged from 36 to 85 years (mean 59.7±9.7 years for the entire cohort, mean 60.0±10.0 years for Caucasians, age of onset unknown for one Caucasian patient). The mean age of the control group was 70.5±16.9 years (age unknown for one individual).

3.2.3.2 Association with Risk of sIBM

The *APOE* allelic and genotypic frequencies are shown in Table 3-4, and an example of the *APOE* genotyping on sIBM patients is shown in Figure 3-1. The *APOE* ε2/ε2 genotype was absent in both the sIBM cohort and the control group. No significant differences were found between patients and controls in *APOE* allelic and genotypic frequencies, in both the Caucasian patient subset and the entire cohort.

Table 3-4 Frequencies of *APOE* and *TOMM40* polyT repeat genotypes and alleles in sIBM patients and controls

	sIBM Patients		**Controls**	**Total**
	Caucasians N=142 | Others N= 16		**N=127**	**N=285**
***APOE* genotypes**				
ε2/ε2	0	0	0	0
ε2/ε3	16 (11.3%)	3 (18.8%)	14 (11.0%)	33 (11.6%)
ε2/ε4	6 (4.2%)	0	6 (4.7%)	12 (4.2%)
ε3/ε3	88 (62.0%)	11 (68.8%)	78 (61.4%)	177 (62.1%)
ε3/ε4	29 (20.4%)	2 (12.5%)	26 (20.5%)	57 (20.0%)
ε4/ε4	3 (2.1%)	0	3 (2.4%)	6 (2.1%)
***p*-value = 1.000**				
***APOE* alleles**				
ε2	22 (7.8%)	3 (9.4%)	20 (7.9%)	45 (7.9%)
ε3	221 (77.8%)	27 (84.4%)	196 (77.2%)	444 (77.9%)
ε4	41 (14.4%)	2 (6.3%)	38 (15.0%)	81 (14.2%)
Total	284	32	254	570
***p*-value = 0.983**				
PolyT genotypes				
S-S	32 (22.5%)	4 (25.0%)	22 (17.3%)	58 (20.4%)
S-L	25 (17.6%)	2 (12.5%)	17 (13.4%)	44 (15.4%)
S-VL	53 (37.3%)	8 (50.0%)	42 (33.1%)	103 (36.1%)
L-L	10 (7.0%)	1 (6.3%)	14 (11.0%)	25 (8.8%)
L-VL	5 (3.5%)	0	9 (7.1%)	14 (4.9%)
VL-VL	17 (12.0%)	1 (6.3%)	23 (18.1%)	41 (14.4%)
***p*-value = 0.263**				
PolyT alleles*				
S	142 (50.0%)	18 (56.3%)	103 (40.6%)	245 (46.1%)
L	50 (17.6%)	4 (12.5%)	54 (21.3%)	104 (18.9%)
VL	92 (32.4%)	10 (31.3%)	97 (38.2%)	189 (34.9%)
Total	284	32	254	570
***p*-value = 0.089**				

*S: short polyT repeat (≤ 19 bp); L: long polyT repeat (20-29 bp); VL: very long polyT repeat (≥ 30 bp) in *TOMM40* gene.

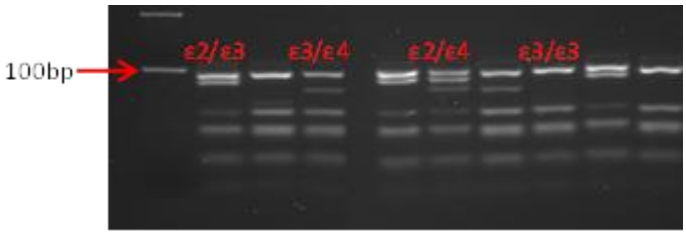


Figure 3-1 *APOE* genotyping using 4% agarose gel electrophoresis

Figure 3-2 shows the polyT repeat genotyping from our patients and Figure 3-3 shows the distribution of *TOMM40* polyT repeat lengths in the sIBM group (range 8-36 repeats) compared to controls (range 11-33 repeats). PolyT with 13 repeats was the most common allele length amongst patients (41.8%) and controls (30.7%). Despite patients had a higher frequency of the S allele (50.6%) compared to controls (40.6%), neither the frequency of *TOMM40* polyT repeat genotypes nor the frequency of polyT alleles differed significantly between groups (Table 3-4).

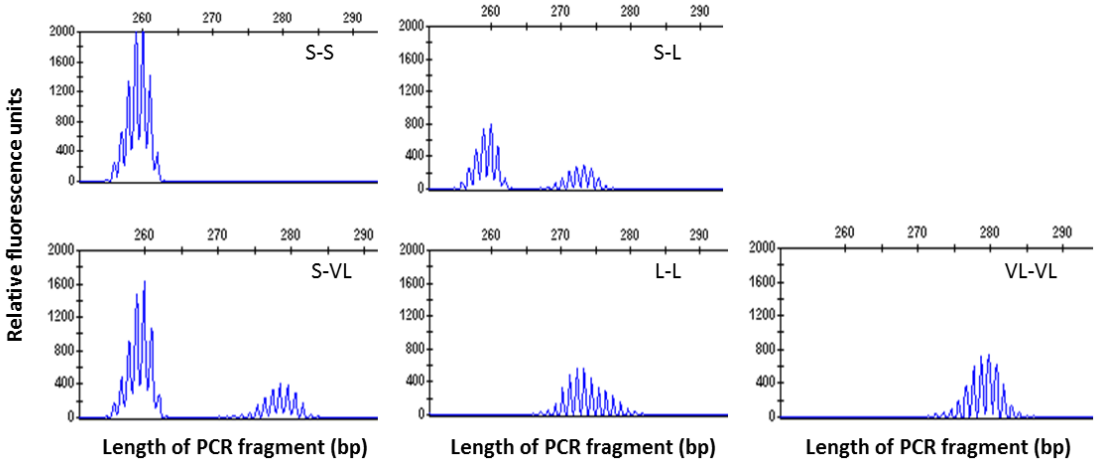


Figure 3-2 Electropherograms for the polyT repeat genotyping (rs10524523)

The figure shows five examples of PCR products of fluorescence-based fragment size analysis separated on an ABI3730XL Genetic Analyzer and visualised by GeneMapper v3.7. The base pair numbers do not correspond to the number of repeat, but refer to the total length of the amplified PCR fragment. The expected length of the PCR product was N+247 base pairs (bp), where "N" represents the number of the polyT residues, and 247 bp was the size of polyT flanking region plus an "A" overhang at the end of the product. In this figure, S (short) = 247-266 bp, L (long) = 267-276 bp, and VL (very long) = 277-281 bp.

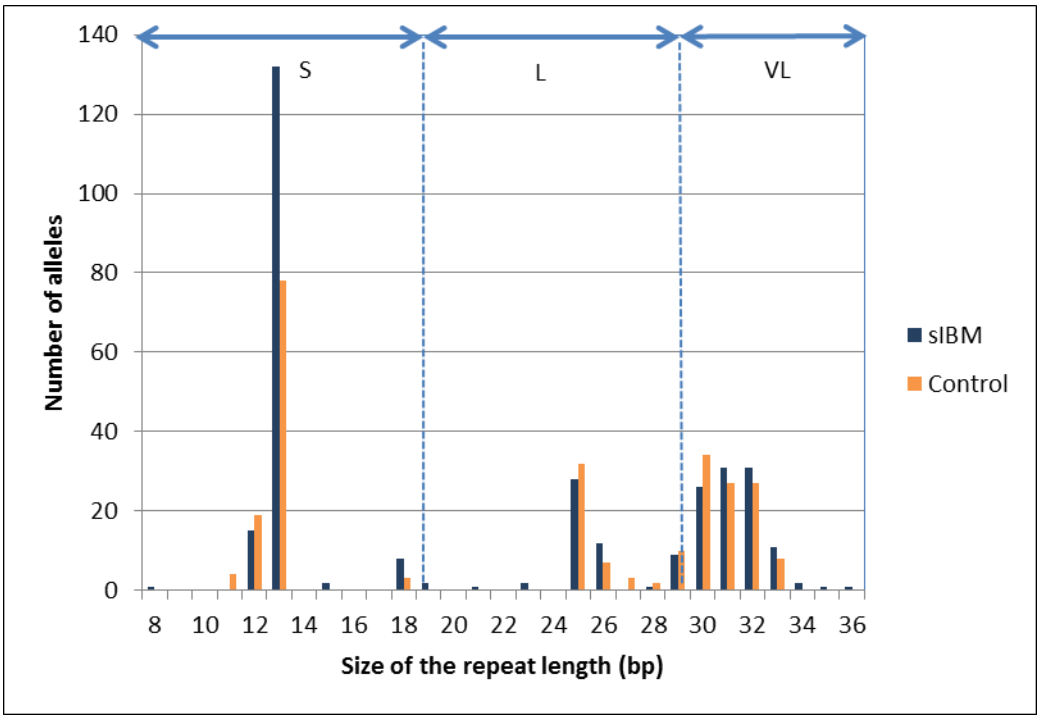


Figure 3-3 Distribution of *TOMM40* polyT repeat lengths in this sIBM patient and control cohort

sIBM patient group N = 158; control group N = 127; S (short) ≤ 19 bp; L (long) = 20-29 bp; VL (very long) ≥ 30 bp.

To clarify the possible interaction between the *APOE* and *TOMM40* genotypes on sIBM risk, we stratified patients and controls by *APOE* genotypes (Table 3-5). Within each *APOE* genotype, no significant differences were found between patients and controls. When the analysis was performed in the combined group of carriers of *APOE* ε3/ε3 and ε3/ε4 genotypes, the difference between those with and without an allele of VL polyT repeats was still not significant. Therefore, *APOE* and *TOMM40* genotypes were not associated with increased risk of sIBM either independently or in combination.

Table 3-5 Frequencies of *TOMM40* polyT repeat genotypes in sIBM patients and controls stratified by *APOE* genotypes

	APOE* ε2/ε3			***APOE* ε2/ε4***			***APOE* ε3/ε3***			***APOE* ε3/ε4***			***APOE* ε4/ε4***		
***TOMM40* polyT**	**sIBM**	**Controls**	**Total**	**sIBM**	**Controls**	**Total**	**sIBM**	**Controls**	**Total**	**sIBM**	**Controls**	**Total**	**sIBM**	**Controls**	**Total**
S-S	5 (26.3%)	1 (7.1%)	6 (18.2%)	0	0	0	27 (27.3%)	19 (24.4%)	46 (26.0%)	3 (9.7%)	2 (7.7%)	5 (8.8%)	1 (33.3%)	0	1 (16.7%)
S-L	1 (5.3%)	2 (14.3%)	3 (9.1%)	3 (50.0%)	3 (50.0%)	6 (50.0%)	10 (10.1%)	4 (5.1%)	14 (7.9%)	13 (41.9%)	8 (30.8%)	21 (36.8%)	0	0	0
S-VL	7 (36.8%)	7 (50.0%)	14 (42.4%)	0	0	0	48 (48.5%)	34 (43.6%)	82 (46.3%)	6 (19.4%)	1 (3.9%)	7 (12.3%)	0	0	0
L-L	0	0	0	1 (16.7%)	2 (33.3%)	3 (25.0%)	1 (1.0%)	3 (3.9%)	4 (2.3%)	7 (22.6%)	6 (23.1%)	13 (22.8%)	2 (66.7%)	3 (100%)	5 (83.3%)
L-VL	0	0	0	2 (33.3%)	1 (16.7%)	3 (25.0%)	1 (1.0%)	0	1 (0.6%)	2 (6.5%)	8 (30.8%)	10 (17.5%)	0	0	0
VL-VL	6 (31.6%)	4 (28.6%)	10 (30.3%)	0	0	0	12 (12.1%)	18 (23.1%)	30 (17.0%)	0	1 (3.9%)	1 (1.8%)	0	0	0
Total	**19**	14	33	**6**	6	12	**99**	78	177	**31**	26	57	**3**	3	6
***p*-value = 0.712**				***p*-value = 1.000**			***p*-value = 0.289**			***p*-value = 0.119**			***p*-value = 1.000**		

S = short; L = long; VL = very long

3.2.3.3 Influence on Age of Onset of sIBM

When analysing the results for possible disease-modifying effects of different *APOE* alleles, six cases with the *APOE* ε2/ε4 genotype were omitted from the analysis due to possible opposing effects of the ε2 and ε4 alleles (by analogy with Alzheimer's disease). This left the following three groups of cases: (a) ε2/ε3, (b) ε3/ε3, and (c) ε3/ε4 and ε4/ε4. Those with the ε2/ε3 genotype had an earlier age of onset (mean age of 56.9) compared with the other two groups (Table 3-6), but this difference did not reach statistical significance. However, being a carrier of the polyT repeat VL allele was significantly associated with a later age of onset by 3.7 years on average (95% CI: 0.4, 6.9; $p = 0.027$, in the adjusted analysis). Since the *APOE* ε3 allele with a VL allele was suggested by Mastaglia et al. to be associated with a later age of onset (Mastaglia et al., 2013), we repeated the analysis for only carriers of the *APOE* ε3/ε3 genotype. We found that *APOE* ε3-*TOMM40* VL allele carriers had an even later age of onset of sIBM by 4.9 years on average (95% CI: 1.1-8.7; $p = 0.013$, in the adjusted analysis). Similar results were found when the analysis was restricted to Caucasians (by 4.0 years for a VL carriage, 95% CI: 0.4-7.6, $p = 0.028$; by 5.4 years for carriage of a VL allele and *APOE* ε3/ε3 genotype, 95% CI: 1.2-9.7, $p = 0.013$). Interestingly, after adjusting for ethnicity, type of tissue and *APOE* and polyT genotypes, there was a trend for a later age of onset in males by 2.7 years when compared to females, although this did not reach statistical significance (Table 3-6).

Table 3-6 Influences of *APOE* alleles, *TOMM40* polyT repeat of VL length, ethnicity and gender on the age of onset of sIBM using standard adjusted linear regression analysis

			Adjusted analysis*	
	Count	Age of onset (mean±SD), years	Regression coefficient (95% CI)	*P*-value
Ethnicity				
Non-Caucasian	16	56.7±5.7	Reference	
Caucasian	141	60.0±10.0	2.8 (-2.3, 7.9)	0.28
Gender				
Female	52	57.9±10.4	Reference	
Male	105	60.6±9.3	2.7 (-0.5, 5.9)	0.095
APOE				
ε2/ε4[a]	6	56.8±5.8		
ε3/ε3	99	60.2±9.8	Reference	
ε2/ε3	19	56.9±10.0	-2.9 (-7.7, 1.9)	0.23
ε3/ε4 and ε4/ε4	33	60.4±9.7	1.6 (-2.4, 5.7)	0.43
***TOMM40* poly-T**				
No VL carriage	74	58.1±9.7	Reference	
VL carriage	83	61.2±9.6	3.7 (0.4, 6.9)	**0.027**
APOE-TOMM40				
ε3/ε3 and polyT non VL carriage	38	57.3±9.9	Reference	
ε3/ε3 and polyT VL carriage	61	62.0±9.4	4.9 (1.1, 8.7)	**0.013**

*Each analysis was adjusted for gender, ethnicity, tissue and genetic factors, except for the variable under study. [a] ε2/ε4 was not included in the regression analysis for *APOE* alleles and the age of onset. SD = standard deviation; CI = confidence interval; VL = very long.

3.2.4 Discussion

This is the largest cohort (158 patients and 127 controls) where the influence of the *APOE* and *TOMM40* genes on sIBM disease risk and features has been investigated to date. Concerning *APOE*, our findings confirmed that the *APOE* ε4 allele is not a susceptibility factor for developing sIBM. *APOE* alleles were also not significantly associated with the age of onset of the disease, but patients with the ε2 allele were likely to present with symptoms slightly earlier. This is in agreement with a previous study (Needham et al., 2008b), where a non-significant trend towards an earlier age of onset in patients with the ε2 allele was also found. An increased risk of an earlier age

of onset in carriers of the *APOE* ε2 allele has also been reported for other diseases (e.g. Machado-Joseph disease/spinocerebellar ataxia type 3, and Parkinson's disease) (Maraganore et al., 2000, Bettencourt et al., 2011). In addition, our findings failed to reproduce the results from a previous *APOE-TOMM40* study by Mastaglia et al. In our study, carriers of the *APOE* ε3/ε3 or ε3/ε4 genotypes with a VL polyT repeat were not under-represented in sIBM compared to controls. The significant association with disease risk observed in the previous study may be a spurious effect observed in a smaller study population, particularly when the analysis was restricted to carriers of the *APOE* ε3/ε3 and ε3/ε4 genotypes. A meta-analysis of combining the previous and current studies could be considered in the future. This can increase the study power. However, carriage of a VL repeat allele was significantly associated with a later age of onset of symptoms in our sIBM cohort. This effect was even more pronounced among those who also had the *APOE* ε3/ε3 genotype. This suggests that the *TOMM40* VL polyT repeat has a disease modifying effect on sIBM by delaying the onset of symptoms, and the *APOE* ε3/ε3 genotype enhances this effect.

It has been hypothesised that the polyT polymorphism may influence susceptibility to LOAD and sIBM by modulating the TOM40 protein or the ApoE protein, thereby altering mitochondrial function (Roses et al., 2010, Mastaglia et al., 2013). Although the association between the polyT repeat locus and LOAD is still debated due to conflicting results (Cruchaga et al., 2011, Jun et al., 2012b), a recent expression study by Linnertz et al. showed that VL polyT is associated with higher expression levels of *APOE* and *TOMM40* mRNA in samples from both normal and LOAD cases (Linnertz et al., 2014). The association between *APOE* and *TOMM40* and sIBM risk was not confirmed in our study, and the exact mechanism of this polyT polymorphism to sIBM remains unknown. An association between the *TOMM40* VL polyT repeat and a later age of onset of sIBM may suggest the mitochondria become less vulnerable in an ageing environment in patients with VL polyT repeat. The hypothesis will need to be justified in further gene expression and functional studies by using human tissues or animal models of sIBM in the future.

This is the first sIBM study where muscle DNA samples from both sIBM patients and pathologically healthy controls were analysed. Thirty-nine cases with DNA from both blood and muscle, and were tested for polyT polymorphism in both DNA

samples. As sIBM is a primary muscle disease, when both blood and muscle tissue were available from the same individual, muscle DNA was used for the analysis, and the regression analyses were adjusted for this potential confounder (tissue differences). Nevertheless, no significant differences in genetic data were seen between the two types of tissue, with the exception of one patient who harboured the *TOMM40* polyT genotype of L-VL in blood, while the S-VL genotype was found in muscle tissue. This may be a somatic change in the polyT length in different tissue environments. This individual finding did not alter our results, as the analyses focused on the influence of VL repeat length.

Additionally, in this study males showed a trend to have a later age of onset than females (Table 3-6). A similar trend was also seen in a previous clinical assessment study of 67 sIBM patients that the age of onset in males (mean age 62, range 55 to 71 years) was slightly later than in females (mean age 61, range 51 to 70 years) (Brady et al., 2013). This may be due to different genetic components between males and females, but the exact mechanism is still unclear. Further studies on a larger sIBM cohort should be performed to clarify the disease-modifying effect of gender on sIBM.

Our study has several limitations. Due to the limited number of non-Caucasian cases, we could not conduct additional analyses stratified by ethnicity. It is therefore important to encourage international collaboration and the recruitment of sIBM patients from different populations, in order to be able to analyse the impact of ethnicity on sIBM features. Furthermore, the data on age of onset (first symptom) was collected retrospectively in both this and previous studies with the potential limitation of self-reported information (recall bias). Roses et al. proposed the necessity for using standardized and informative assessment tools and processes to determine the age of disease development. The age of onset is best measured reliably in prospective studies that individuals are monitored over time for AD studies (Roses et al., 2013). However, this can still be a challenge for IBM studies due to the slow progress of the disease and an often misdiagnosis initially.

In summary, a sIBM disease modifying effect of the *TOMM40* VL polyT repeat was suggested in this study, and this effect was further enhanced by the *APOE* ε3/ε3

genotype. The VL polyT repeat was associated with a later sIBM symptom onset. Although this polymorphism is not associated with increased risk of sIBM, there may be other variants within the *APOE-TOMM40* loci involved sIBM susceptibility. In addition, various other genes, including genes related to immune, degenerative and mitochondrial pathways may contribute to sIBM disease susceptibility (Gang et al., 2014). Identification of new genes and variants is crucial to improving our understanding of this complex disease.

3.3 Identification of Rare Variants in *SQSTM1* and *VCP* Genes in a Large Cohort of sIBM

3.3.1 Introduction

P62, also known as sequestosome 1 (*SQSTM1*), is a shuttle protein that transports polyubiquitinated proteins for their degradation by both ubiquitin-proteasome system and autophagy-lysosome pathway (Seibenhener et al., 2004). It is also recognized as a pathological biomarker with a high sensitivity and specificity in sIBM muscle biopsies (Brady et al., 2014). Genetic variants in *SQSTM1* had not been investigated in sIBM until a recent study using targeted next generation sequencing in a group of 79 patients (Weihl et al., 2015). In that study, only a common *SQSTM1* polymorphism was found, unlikely to contribute to a rare disease. A recent study has reported a splice donor variant in *SQSTM1* in a family with an autosomal dominant distal myopathy and also in an unrelated patient with sporadic distal myopathy (Bucelli et al., 2015). Mutations in valosin-containing protein (*VCP*) gene have been known to cause IBMPFD, and two missense mutations in *VCP* have recently been identified in two unrelated IBM patients, one with sIBM and another with a family history for late onset dementia (Weihl et al., 2015).

These findings, along with denervation in muscle EMG of sIBM patients suggest a possible genetic overlap between sporadic and IBM-like myopathies, and also neurodegenerative diseases. To thoroughly investigate the contribution of *SQSTM1* and *VCP* genes in sIBM, we investigated these two genes using whole-exome sequencing (WES) data from 181 sIBM patients, which was produced as a part of an IIBMGC.

3.3.2 Methods

A total of 181 sIBM exomes were used and 235 neuropathologically healthy controls (aged over 60 years), were selected in this study. The control whole-exome data was kindly provided by Professor Andrew Singleton at the National Institute of Health, USA. We used a candidate gene approach on these WES data to investigate rare

variants in *SQSTM1* and *VCP* genes. The allele frequency of each variant found in sIBM was compared with the ExAC database using Fisher's test.

Whole-genome expression profiling from muscle tissues was performed on three of four patients found with *SQSTM1* variants (Cases 1, 3 and 4), two patients with *VCP* variants (Cases 6 and 7) and five age and gender-matched controls without muscle diseases. The selected genes (*HLA-A*, *CD74*, and *HLA-DRA*) were validated using RT-qPCR. Protocols were as described in Section 2.4.

3.3.3 Results

Of the entire cohort of 181 sIBM patients, 150 (82.9% of this patient cohort) were Caucasians, and 16 (8.8%) were from other ethnicities, including Asian Chinese, Asian/American Indians and Black Africans (ethnicity information was unavailable for 15 patients). The majority of sIBM patients were male (65.9%). Age of onset, which was collected retrospectively, ranged from 31 to 85 years (mean 59.6±9.6 years). The mean age of the 235 healthy ageing controls was 79.1±8.5 years, ranging from 60 to 102 years. Similar to the sIBM cohort, the majority were males (61.7%).

In this sIBM cohort, four rare missense variants in the *SQSTM1* gene (ENST00000389805) were found in four patients (Table 3-7). Two rare missense variants in the *VCP* gene (ENST00000358901) were found in three sIBM patients. Of note, the frequency of the variants *SQSTM1* c.580G>A (p.Gly194Arg) and *VCP* c.475C>T (p.Arg159Cys) was significantly higher in our sIBM cohort when compared with the ExAC database (Fisher's exact test, $p = 0.018$ and $p = 5.346 \times 10^{-5}$, respectively). Four of these variants had previously been reported in patients with ALS (Rubino et al., 2012, Abramzon et al., 2012). Among them, the *SQSTM1* c.1175C>T (p.Pro392Leu) is known to be the most frequent *SQSTM1* mutation in PDB (Laurin et al., 2002); the *VCP* c.79A>G (p.Ile27Val) and p.Arg159Cys have also been found in patients with IBMPFD (Rohrer et al., 2011, Chan et al., 2012), and the *VCP* p.Ile27Val has also recently been reported in one sIBM patient (Weihl et al., 2015). The *SQSTM1* c.350C>T (p.Ala117Val) variant was reported in one early onset AD patient (Cuyvers et al., 2015), while the p.Gly194Arg variant has not been associated with other diseases. Variants found in sIBM patients were absent in our

internal aged controls except for *SQSTM1* variants previously known to be associated with ALS (p.Pro392Leu and c.712A>G [p.Lys238Glu]). With the exception of *SQSTM1* p.Ala117Val, all these variants are located within conserved positions across different species further suggesting that they are functionally relevant.

Table 3-7 *SQSTM1* and *VCP* rare genetic variants in patients with sIBM

Case ID	Gene and region	Variants (Heterozygous)	MAF in 235 neuropathological controls (%)	MAF in 1000 Genomes (%)	MAF in ExAC (%)	MAF in sIBM cohort (%)	GERP++ Score	PolyPhen prediction	Known in other diseases
Case 1 (sIBM)	*SQSTM1* Exon 8	c.1175C>T p.Pro392Leu (rs104893941)	0.213	0.46	0.089	0.275	4.43	Pathogenic	Familial PDB and ALS
Case 2 (sIBM)	*SQSTM1* Exon 3	c.350C>T p.Ala117Val (rs147810437)	0	0.18	0.152	0.275	-5.17	Benign	Early-onset AD
Case 3 (sIBM)	*SQSTM1* Exon 4	c.580G>A p.Gly194Arg	0	-	0.0017	0.275	3.65	Possibly damaging	-
Case 4 (sIBM)	*SQSTM1* Exon 5	c.712A>G p.Lys238Glu (rs11548633)	0.638	0.32	0.242	0.275	3.87	Possibly damaging	ALS
Case 5 (sIBM)	*VCP* Exon 2	c.79A>G p.Ile27Val (rs140913250)	0	0.09	0.054	0.275	5.71	Benign	IBMPFD, ALS and PD
Case 6 (sIBM) and Case7 (sIBM)	*VCP* Exon 5	c.475C>T p.Arg159Cys	0	-	0.00082	0.549	4.62	Possibly damaging	IBMPFD, IBM with PD, and sporadic ALS

SQSTM1 = sequestosome 1 gene; *VCP* = valosin-containing protein gene; sIBM = sporadic inclusion body myositis; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; MAF = minor allele frequency; PDB = Paget's disease of bone; ALS = amyotrophic lateral sclerosis; AD = Alzheimer's disease; IBMPFD = inclusion body myopathy with Paget's disease and frontotemporal dementia; PD = Parkinson's disease.

Tables 3-8 and 3-9 summarise the demographic, clinical and muscle biopsy characteristics of the patients carrying *SQSTM1* and *VCP* variants. The seven sIBM cases fulfilled the MRC 2010 diagnostic category of pathologically defined, clinically defined or possible sIBM. There was also no family history of muscle diseases, and none of the seven patients and their families showed evidence of bone or cognitive problems.

Figure 3-4 illustrates the pathological features of muscle biopsies observed in patients carrying variants in *SQSTM1* only, due to the unavailability of biopsy images of the other patients. P62 positive inclusions were found in three patients with *SQSTM1*. sIBM patients showed a global up-regulation of MHC-I (diffuse pattern, Figure 3-4 G) in comparison with the healthy control (Figure 3-4 H). The biopsies were examined and the images were kindly taken by Professor Janice Holton from the Department of Pathology, ION.

Table 3-8 Demographic and clinical features of sIBM patients carrying variants in the *SQSTM1* and *VCP* genes

Features	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Sex	F	M	M	F	M	F	M
Ethnicity	Caucasian	Caucasian	Indian subcontinent	Caucasian	Caucasian	Caucasian	Caucasian
Age at onset	45	50	71	57	85	74	48
Family history	-	-	-	-	-	-	-
Finger flexor weakness	+	+	+	+	+	+	+
Weakness of KE > HF	-	-	+	+	UNK	-	-
Weakness of FF > SA	+	+	+	+	UNK	+	+
Weakness of WF > WE	+	+	+	-	UNK	-	-
PDB	-	-	-	-	-	-	-
ALS	-	-	-	-	-	-	-
FTD	-	-	-	-	-	-	-
Parkinson's disease	-	-	-	-	-	-	-
Elevated CK (x ULN)	+ (≤15)	+ (≤15)	+ (≤15)	-	-	N/A	+ (≤15)
Neurophysiological investigation	Myopathic	Myopathic	Myopathic	N/A	UNK	N/A	N/A
MRC 2010 sIBM diagnostic category	CLD	CLD	CLD	CLD	PO	PO	PO

F = female; M = male; KE = knee extension; HF = hip flexion; FF = finger flexion; SA = shoulder abduction; WF = wrist flexion; WE = wrist extension; PDB = Paget's disease of bone; ALS = amyotrophic lateral sclerosis; FTD = frontotemporal dementia; CK = creatine kinase; ULN = upper limit of normal; UNK = unknown; N/A = not available; CLD = Clinically defined; PO = Possible.

Table 3-9 Muscle biopsy features of sIBM patients carrying variants in the *SQSTM1* and *VCP* genes

Muscle biopsy features	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Endomysial exudate	+	+	+	+	+	+	+
MHC-I up-regulation	+	+	+	+	+	+	+
Partial invasion	+	+	+	-	+	-	-
Rimmed vacuoles	+	+	+	+	+	-	-
p62 (sarcoplasmic and intranuclear inclusions)	+	+	+	N/A	N/A	N/A	N/A
TDP-43	N/A	+	N/A	N/A	N/A	N/A	N/A
15-18 (or 16-21) nm filaments	N/A	N/A	N/A	-	N/A	N/A	N/A
COX deficient fibres	+	+	-	+	-	+	+

MHC: = major histocompatibility complex; TDP-43 = transactivation response DNA-binding protein-43; COX = cytochrome *c* oxidase; N/A = not available.

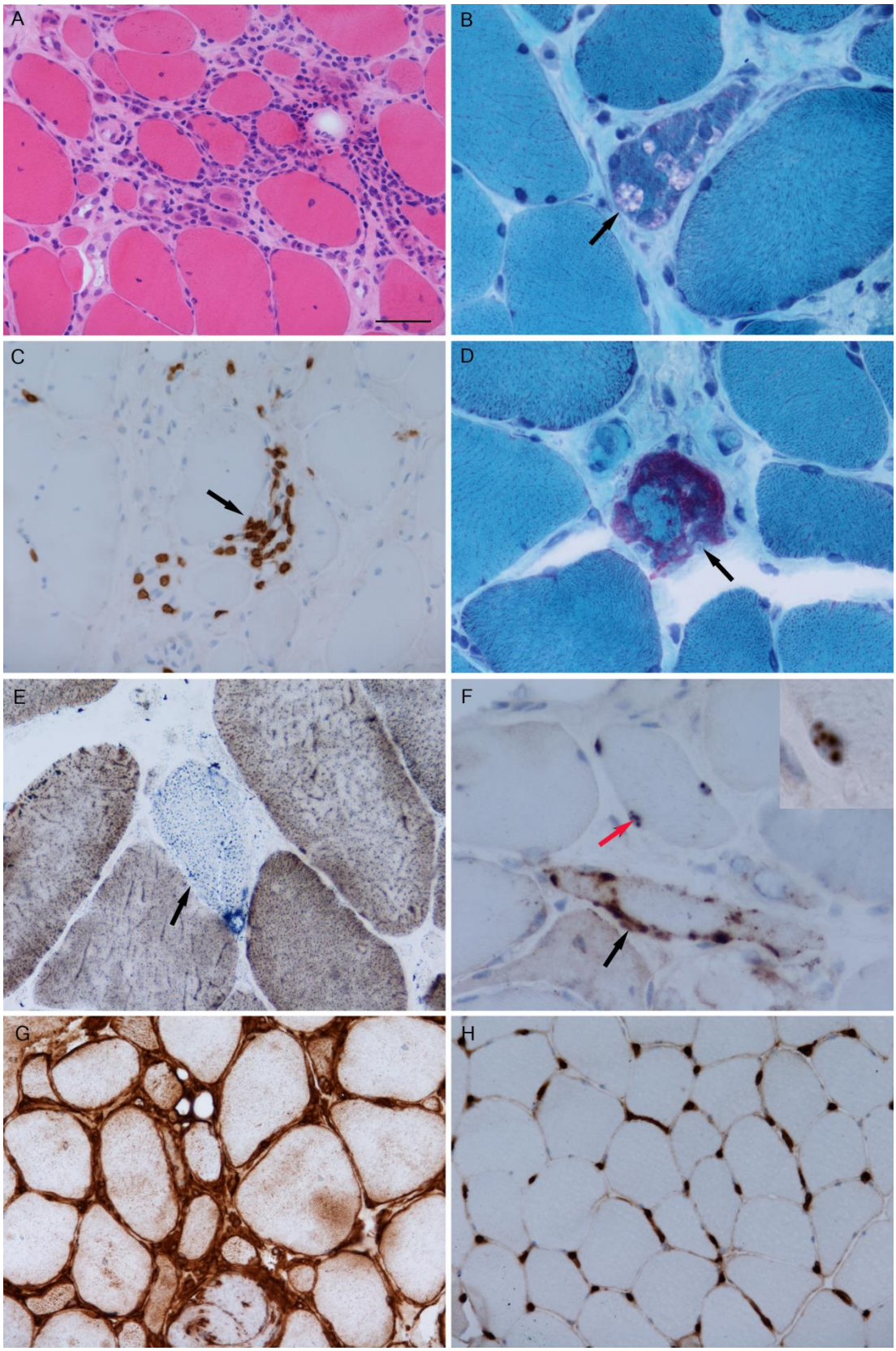


Figure 3-4 Pathological features observed in sIBM patients

There was variation in fibre size with endomysial inflammation, increased internal nucleation and fibre regeneration (A, haematoxylin and eosin). Rimmed vacuoles (black arrow) were found in all cases (B, Gomori trichrome). The inflammatory infiltrate contains T lymphocytes (C, CD3, pointed by

a black arrow). A ragged red fibre (black arrow) was observed in Case 2 (D, Gomori trichrome). Cytochrome *c* oxidase negative fibres (black arrow) were identified (E, cytochrome *c* oxidase/succinic dehydrogenase). P62 immunoreactive sarcoplasmic inclusions (black arrow) were identified (F) in addition to sparse intranuclear inclusions (red arrow and inset). MHC Class I was diffusely increased in patients with *SQSTM1* variants (G), in comparison with a normal control (H). Scale bar represents 50 μm in A, C, and G-H; 25 μm in B and D-F; 10 μm in the inset in F. Panels A and G are from Case 1; panels B-F are from Case 2. Images for stainings on healthy control muscle were not available except MHC Class I marker. The biopsies were examined and images were kindly taken by Professor Janice Holton from the Department of Pathology, ION.

To further understand molecular changes occurring in sIBM, particularly those related to the variants in *SQSTM1* and *VCP*, we performed whole-genome expression analysis on muscle tissues. Most of the differentially expressed genes were found when we compared the *SQSTM1* sIBM patient group with controls (Appendix II - Table 2 and Table 3). Thirty-three genes were up-regulated and seven genes were down-regulated. The small number of available tissue samples (n=2) from patients with *VCP* variants prevented statistical analysis of this sIBM patient group. The expression of *SQSTM1* and *VCP* did not show significant differences between any patient group and controls, but a significant up-regulation of MHC genes (class I - *HLA-A*, and class II – *CD74* and *HLA-DRA*) was seen in the group of patients carrying *SQSTM1* variants compared to controls (Appendix II - Table 3). This finding is consistent with the MHC-I patterns observed in the patients' biopsies (Figure 3-4). RT-qPCR analysis of those MHC genes validated their up-regulation in sIBM; this was particularly evident when comparing the *SQSTM1* group with the controls, with significant up-regulation of all analysed genes (Figure 3-5). Functional enrichment analysis of up-regulated genes in the *SQSTM1* patient group showed a significant over-representation of several GO terms related to immune response, MHC protein complex, and endosome vesicles. KEGG pathways that were significantly over-represented mostly related to inflammatory, autoimmune and infectious diseases (Appendix II - Table 4). The small number of dysregulated genes found in the expression microarray analysis data prevented functional enrichment analysis for other comparison groups.

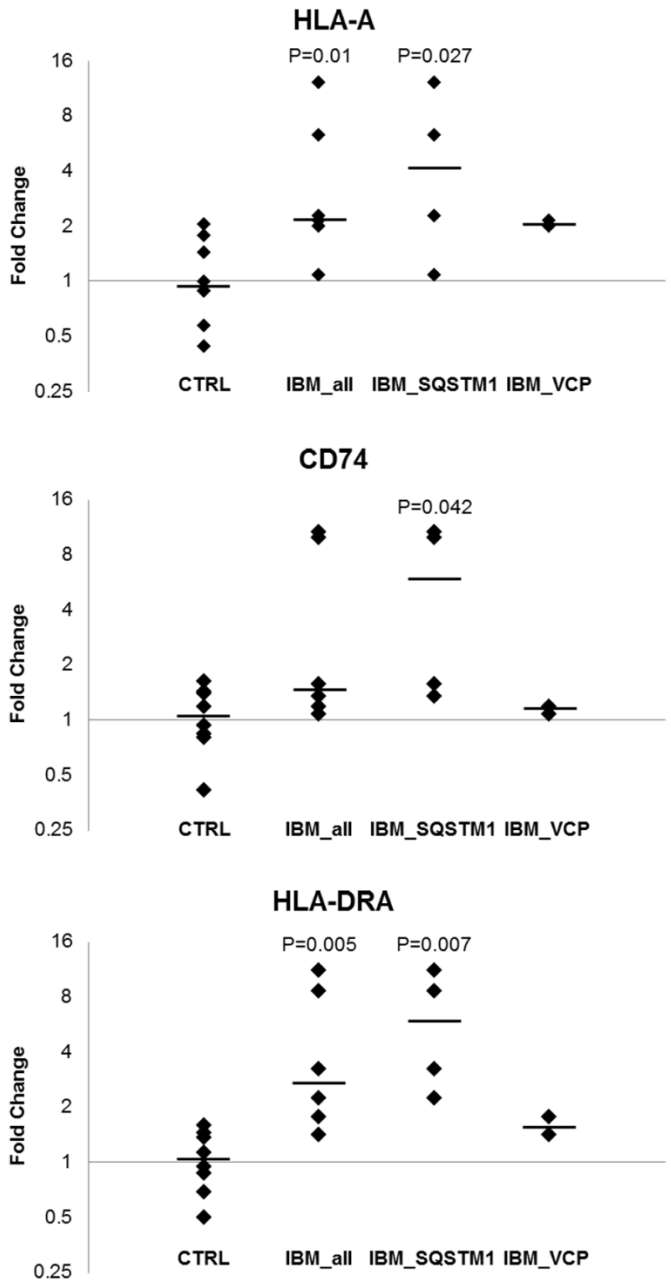


Figure 3-5 Distribution of the fold change for the expression of three MHC genes in sIBM groups compared to controls as determined by RT-qPCR

All expression levels were normalised to the expression of the reference gene, *PPIA*, and relative to the mean normalised expression of all the controls. The solid black lines denote the medians. Only significant Mann-Whitney U-test *p*-values (< 0.05) regarding comparisons of patient groups with controls are shown. Gene abbreviations: *HLA-A* = major histocompatibility complex, class I, A; *CD74* = CD74 molecule, MHC class II invariant chain; *HLA-DRA* = major histocompatibility complex, class II, DR alpha. CTRL = control.

3.3.4 Discussion

Using WES we identified rare missense variants in the *SQSTM1* and *VCP* genes in seven sIBM cases. The frequency of patients with rare *SQSTM1* and *VCP* variants in the sIBM cohort was 3.8%. Two cases had previously been reported with *VCP* variants (Weihl et al., 2015). Our study extended this finding in a larger cohort of sIBM patients. Regarding *SQSTM1*, this is to our knowledge the first report where likely pathogenic variants in this gene were observed in sIBM patients.

The *SQSTM1* gene encodes for sequestosome 1/p62 (referred as p62), which is a scaffold protein consisting of 440-amino acids and several domains or motifs. These domains include an N-terminal Phox and Bem1p (PB1) domain, a tumour necrosis factor receptor-associated factor 6 (TRAF6)-binding site (TBS), a LC3-interaction region (LIR) and a C-terminal ubiquitin associated domain (UBA) (Rea et al., 2014) (Figure 3-6). P62 is a multifunction protein participating in a number of different biological pathways (Komatsu et al., 2012), including the autophagy pathway and various transduction pathways such as NF-κB signaling and apoptosis. Mutations in *SQSTM1* were first identified in PDB (Laurin et al., 2002), a chronic disease of bone that can cause skeletal deformity and fractures. Mutations in this gene have been found in 25–50% of familial and 5–10% of sporadic PDB cases (Ralston and Layfield, 2012). In addition, mutations in *SQSTM1* are also known to contribute to 1–3.5% of patients with ALS/FTD with or without family history (Rubino et al., 2012), a similar frequency to that which was found in our sIBM cohort. Mutations in *SQSTM1* are widespread along the gene. However, the missense mutation p.Pro392Leu located in the C-terminal UBA domain where most mutations lie, is the most frequent *SQSTM1* mutation in all the different clinical phenotypes (Fecto et al., 2011, Laurin et al., 2002). A mouse model with sqstm1 p.Pro394Leu mutation (Daroszewska et al., 2011), equivalent to *SQSTM1* p.Pro392Leu in humans, developed a human PDB-like phenotype, and showed dysregulation of autophagy and enhanced autophagosome formation. The *SQSTM1* p.Lys238Glu variant has also been reported in one sporadic ALS case (Rubino et al., 2012), and lies in the TBS where p62 interacts with TRAF6, a critical component of the NF-κB pathway in response to multiple factors, including proinflammatory cytokines (Fecto et al., 2011). The *SQSTM1* p.Gly194Arg variant has not been observed in other diseases

and was absent in our aged neuropathologically healty control exomes, and it is worth noting that it has been found to be over-represented our sIBM cohort. Although the *SQSTM1* p.Ala117Val variant is predicted as benign, it was absent from our aged controls and recently was reported in a patient with early-onset AD (Cuyvers et al., 2015), and thus cannot be excluded as a risk factor for sIBM.

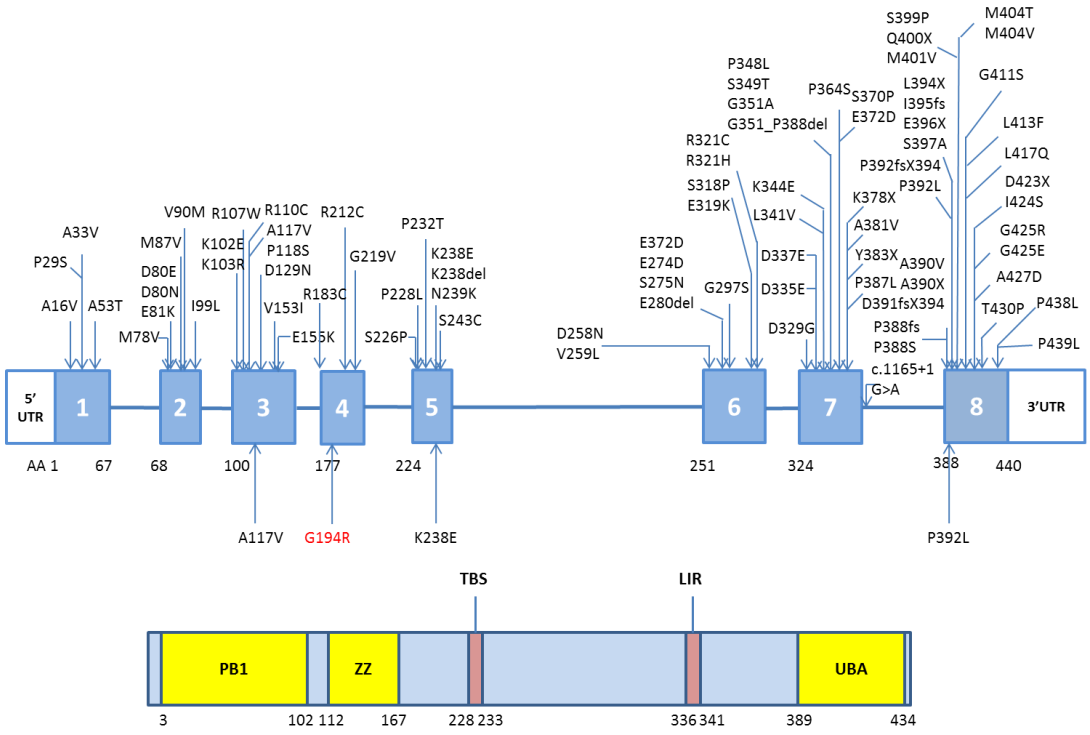


Figure 3-6 *SQSTM1* gene structure and protein domains

The mutations listed above the gene structure were previously reported to be associated with diseases. The ones below were identified in our sIBM cohort, and the one labelled in red was not reported in other diseases before. PB1 = Phox and Bem1p domain; ZZ = zinc finger domain; TBS = tumour necrosis factor receptor-associated factor 6 binding site; LIR = LC3-interaction region; UBA = ubiquitin-associated domain.

The *VCP* gene encodes for the ATPase valosin-containing protein, which consists of two ATPase domains, D1 and D2, cell division protein 48 (CDC48) domain, N-terminal domain and C-terminal tail (Meyer and Weihl, 2014) (Figure 3-7). *VCP* plays a main role in proteasomal degradation of misfolded proteins and is also involved in critical signalling pathways, membrane fusion, cell cycle controls, and more importantly facilitating cargo sorting via endosomal/autophagy pathway (Meyer and Weihl, 2014). Mutations in *VCP* are known to cause IBMPFD (Watts et al., 2004), PD (Majounie et al., 2012b), and are associated with ALS with or without

FTD (Johnson et al., 2010). The *VCP* p.Ile27Val variant has been previously reported as potentially pathogenic (Rohrer et al., 2011, Majounie et al., 2012b), and was recently found in another patient with sIBM (Weihl et al., 2015). Functional analysis of this variant showed an increase in p62 and LC3II protein levels (Weihl et al., 2015), suggesting that it may cause disruption to autophagosome maturation (Ju et al., 2009). The *VCP* p.Arg159Cys variant found to be over-represented in our sIBM cohort was previously reported as pathogenic and associated with IBMPFD (Bersano et al., 2009) and sporadic ALS (Abramzon et al., 2012). Two additional mutations were also found at this amino acid residue in familial ALS (p.Arg159Gly) (Johnson et al., 2010) and IBMPFD (p.Arg159His) (Haubenberger et al., 2005). The *VCP* p.Arg159Cys variant lies within the highly conserved CDC48 domain of the protein, which is involved in ubiquitin binding and protein-protein interaction, and a hotspot for *VCP* mutations (Bersano et al., 2009).

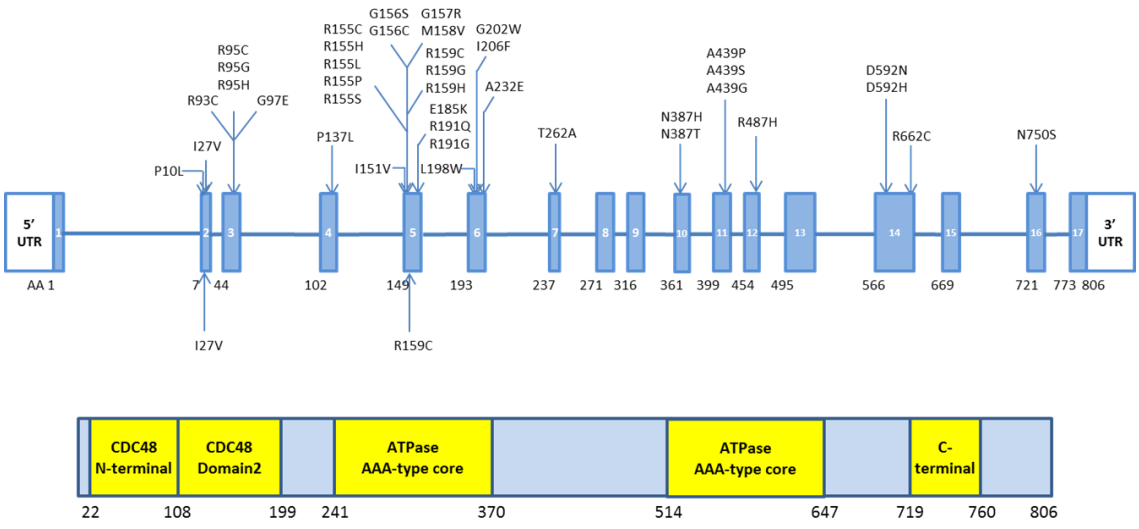


Figure 3-7 *VCP* gene structure and protein domains

The mutations listed above the gene structure were previously reported to be associated with diseases. The ones below were identified in our sIBM cohort.

We reviewed the clinical and pathological details of all sIBM patients with *SQSTM1* and *VCP* variants, and confirmed that none had developed symptoms of PDB, FTD, ALS, or had family history of these diseases or of muscle weakness. Whether these patients would develop PDB or FTD phenotypes is not predicable. It would be interesting to follow-up these patients and also request DNA samples from their parents and offsprings if possible. Based on patients' current phenotypes and family

history, these rare variants in *SQSTM1* and *VCP* are considered as risk factors for sIBM. VCP staining was not available for all the patients, but p62 positive inclusions were seen in all four biopsies where p62 staining was available, including three sIBM cases with *SQSTM1* p.Pro392Leu, p.Ala117Val, and p.Gly194Arg variants. P62 positive inclusions were reported in 26 (14.4%) out of our 181 patients. Among these 26 cases, 23 were from NHNN. P62 staining became a routine staining at NHNN only recent years, however, it is not available in many our collaborators' labs. It would be worthwhile to include p62 as a routine staining in more labs in the future.

DNA samples from both blood and muscle were only available from Cases 1-3. DNA samples from Case 4, Case 6 and Case 7 were from muscle, while DNA from Case 5 was from blood. In Cases 1-3, the *SQSTM1* rare variants were found in their both tissues. This suggests that these rare variants are not somatic variants in muscle. The expression levels of neither *SQSTM1* nor *VCP* mRNA were significantly altered in patients when compared to controls, suggesting that the missense variants found in these patients did not alter the corresponding gene expression at the mRNA level. P62 and VCP aggregates in the muscle could be a result of either increased protein stability, or impairment in the ubiquitin-proteasome system or autophagy-lysosome pathway, or both due to the rare variants in *SQSTM1* and *VCP* genes (Sandri, 2010).

MHC class I (*HLA-A*) and II (*HLA-DRA*) genes were significantly up-regulated in patients compared to healthy muscle controls by RT-qPCR, which is consistent with a previous study (Ivanidze et al., 2011). The other MHC class II gene (*CD74*) was significantly up-regulated only in the *SQSTM1* patient group. Statistical analysis could not be carried out for the *VCP* patient group due to small sample size, but there was also a trend for the up-regulation of these inflammatory markers. The up-regulation observed in the *SQSTM1* sIBM patient group was more pronounced than in the *VCP* patient group. This is the first time that different patterns of inflammatory markers between patients with sIBM are suggested by mRNA analysis. Upregulation of MHC Class I staining was reported in muscle biopsies of all the seven patients with *SQSTM1* and *VCP* genes, but the biopsy slides of *VCP* patient group were not available for a further assessment. We hypothesize that MHC expression could be a potential differentiating factor that directs the clinical phenotype of *SQSTM1* or *VCP* towards sIBM as opposed to the other neurodegenerative conditions. Further cases

should be analysed, especially to compare with sIBM patients without *SQSTM1* and *VCP* genes, and other neurodegenerative disorders with these two genes.

In conclusion, we report for the first time likely pathogenic *SQSTM1* variants and expand the spectrum of *VCP* variants in sIBM. Our findings suggest that variants in these genes constitute genetic susceptibility factors for sIBM, like other multisystem proteinopathy (MSP) phenotypes. The findings from this study also expand the clinico-pathological spectrum of diseases associated with the *SQSTM1* and *VCP* genes, and the overlap between sIBM and IBMPFD/ALS/FTD suggests that muscle and brain diseases share similar pathogenic pathways which may be important for further discovery of biomarkers, genes, and therapeutic targets.

3.4 Analyses of Other Candidate Genes in sIBM

3.4.1 Introduction

Abnormal protein accumulation and abnormal mitochondria are common features observed in the muscle of sIBM patients, which overlaps with those seen in brains from patients with neurodegenerative diseases. Although mutations in genes encoding the accumulated proteins have been found to be associated with different neurodegenerative diseases, the pathogenicity of these genes in sIBM is still uncertain. In this section, a few candidate gene studies were performed based on the sIBM cohort described in Section 3.1. The genes examined in this chapter and the justification for including them are discussed below.

3.4.1.1 MAPT

Phosphorylated-tau has been reported to be a major protein present in sIBM muscle aggregates (Maurage et al., 2004, Askanas and Engel, 2008). This is strikingly similar to those found in taupathies which share hyperphosphorylated deposited tau protein in the brain (Maurage et al., 2004). Different microtubule-associated protein tau (*MAPT*) gene mutations have been identified in many different tauopathies including AD, frontotemporal dementia, Parkinsonism linked to chromosome 17 (FTDP-17), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD). The majority of them affect exons 9-13 (van Slegtenhorst et al., 2000). In addition to *MAPT* mutations, a common extended haplotype (H1) in *MAPT* appears to be a risk factor for sporadic PSP (Baker et al., 1999). A 238 bp deletion in *MAPT* intron 9 is inherited as a part of the less common H2 sub-haplotype (Baker et al., 1999). Interestingly, a recent study reported *MAPT* splicing changes in sIBM, specifically in exon 6 (Cortese et al., 2014). This suggests that *MAPT* gene plays a role in sIBM.

3.4.1.2 PRNP

Prion protein and mRNA have been described as being abnormally accumulated in vacuolated muscle fibres in both sIBM and hIBM (Sarkozi et al., 1994, Askanas et al., 1993b). Lampe et al. reported that homozygosity for methionine at codon 129 (M/M) of the prion protein (*PRNP*) gene was found in 14 out of 22 sIBM patients

(64%). This was more prevalent in sIBM compared to controls (Lampe et al., 1999). However, this finding was not replicated in later studies with a larger cohort of 41 sIBM patients (51.2% M/M) (Orth et al., 2000).

3.4.1.3 APP

It has been widely recognized that there are abnormal accumulations of Aβ, β-amyloid precursor protein (*APP*), and AβPP-mRNA in both sIBM and hIBM muscle fibres (Askanas et al., 1992, Askanas et al., 1993a, Sarkozi et al., 1993). Aβ42 is considered more cytotoxic and more prone to self-association and oligomerization than Aβ40, thereby exerting a detrimental effect in AD brain (Masuda et al., 2009). Similarly, increased Aβ42 oligomers were demonstrated preferentially in sIBM affected fibres (Nogalska et al., 2010b) and also in plasma from patients with sIBM (Abdo et al., 2009). These indicate that the intra-muscle fibre accumulation of Aβ42 oligomers in sIBM contributes to the pathogenic cascade (Nogalska et al., 2010b). The mechanisms of overexpression of *APP* gene in sIBM have not been fully understood. The fact that mutations in exons 16 and 17 of *APP* gene have been found to be pathogenic in AD, resulting in early-onset familial AD, suggests that mutations in this gene might also play a role in sIBM. However, no coding *APP* mutations were found in a previous small study (Sivakumar et al., 1995).

3.4.1.4 C9orf72

An expanded GGGGCC hexanucleotide repeat in the noncoding region between exons 1a and 1b of the *C9orf72* gene, located on chromosome 9p21, was first identified as the most common genetic cause of familial ALS, FTD or a combination of both phenotypes and TDP-43-based pathology in two independent studies (DeJesus-Hernandez et al., 2011, Renton et al., 2011). Further studies also found *C9orf72* repeat expansion in other neurodegenerative diseases including AD (Majounie et al., 2012a), PD (Lesage et al., 2013), PSP (Origone et al., 2013), ataxic syndromes and corticobasal degeneration (Lindquist et al., 2013), sporadic Creutzfeldt-Jakob disease, and Huntington disease-like syndrome (Beck et al., 2013). The normal repeat length was less than 20 units in most healthy controls (Renton et al., 2011), and the maximum repeat length was estimated at 23-30 units depending on studies (DeJesus-Hernandez et al., 2011, Renton et al., 2011, Gijselinck et al., 2012). However, in a UK population, large expansions (> 400 repeats) were found at

a frequency of around 1 in 600 individuals (Beck et al., 2013). The estimated pathogenic expansions associated with disease ranged from over 400 units to 4,400 units (DeJesus-Hernandez et al., 2011, Renton et al., 2011, Beck et al., 2013).

The exact pathogenic mechanisms of the large hexanucleotide repeat expansion are yet unknown; different potential mechanisms have been proposed (Todd and Petrucelli, 2016). Reduced expression of C9orf72 levels were found in the frontal cortex of ALS/FTD brains with large expansion, indicating that loss of C9orf72 function may play a role (Waite et al., 2014). A gain-of-function mechanism was also suggested in which the expanded repeat forms toxic RNA foci co-localised with hnRNPA1 in C9orf72-ALS motor neurons which is believed to sequester RNA-binding proteins and impair their function in RNA processing (Donnelly et al., 2013, Sareen et al., 2013). In addition, the repeat also produces abnormal accumulation of toxic dipeptide repeat proteins (DPRs) in the central nervous system in an unconventional manner termed repeat-associated non-ATG (RAN) translation (Mori et al., 2013, Kanekura et al., 2016). Accumulation of DPRs was found to cause cell death *in vitro* and also in a fly model (Kwon et al., 2014, Mizielinska et al., 2014). Although the normal function of C9orf72 is still unknown, C9or72 was found to co-localise with several endocytic and autophagy-related Rabs (Farg et al., 2014), as well as ubiquitin-binding proteins such as p62 and ubiquilin-2 (Bieniek et al., 2013). This suggests that C9orf72 may be involved in endosomal and autophagy-related trafficking. A recent study reported a C9orf72 knockout mouse model leading to lysosomal accumulation, macrophage and microglial dysfunction and age-related neuroinflammation, suggesting C9orf72 may also play a key role in innate immune cells (O'Rourke et al., 2016).

As there is a genetic overlap between sIBM and ALS/FTD (such as *SQSTM1* and *VCP* genes discussed in Section 3.3), it was questioned whether the *C9orf72* large expansion would be also associated with sIBM. However, no such expansion was detected in a previous study of 79 sIBM patients (Weihl et al., 2015).

3.4.1.5 Mitochondrial DNA Deletion(s)

Mitochondrial abnormalities are another important pathological feature in sIBM muscle biopsies. These consist of ragged-red fibres and mostly show enzyme

histochemical deficiency of COX activity. These changes are more prevalent in sIBM than in polymyositis, dermatomyositis and normal ageing muscle fibres (Rifai et al., 1995). It is therefore of great interest to investigate another group of susceptibility factors – mtDNA. An accumulation of mtDNA molecules with large-scale deletions was found in many COX-deficient ragged-red fibres in sIBM patients (Oldfors et al., 1993). There were multiple different mtDNA deletions in different muscle fibres, but usually one predominant type of deletion was present in each COX-deficient fibre (Moslemi et al., 1997, Santorelli et al., 1996). Thirty-three different deletions were identified by sequencing four patients with sIBM. The majority of mtDNA deletion breakpoints identified in sIBM muscle fibres span the region between nt8029-8032 and nt16066-16078 (Moslemi et al., 1997). A recent study on 26 Caucasian sIBM patients used long-extension PCR and qPCR analysis. They reported a significantly higher relative amount of mtDNA deletions in patients with a high amount of COX-deficient fibres (5.81-23.50%) compared to those with less COX-deficient fibres (0.48-2.65%) ($p < 0.001$) and the age-matched controls ($p < 0.01$) (Lindgren et al., 2015). A similar trend was also observed in another study on 16 sIBM patients conducted by Rygiel et al (Rygiel et al., 2015). The authors hypothesised a link involving cross-talk between inflammation, accumulation of mtDNA rearrangement, and fibre atrophy in sIBM. They showed a strong correlation between the severity of T lymphocyte infiltration and the proportion of COX-deficient fibres. Furthermore, fibres with respiratory dysfunction were more prone to be atrophic compared with fibres with normal respiratory function (Rygiel et al., 2015). Lindgren et al. also evaluated several nuclear DNA genes encoding proteins including *POLG* (OMIM#174763), *C10orf2* (OMIM#606075), *DNA2* (OMIM#60 1810), *MGME1* (OMIM#615076), *POLG2* (OMIM#604983), *OPA1* (OMIM#604 983), *TK2* (OMIM#188250), *TYMP* (OMIM#131222), *RRM2B* (OMIM#604712), and *SLC25A4* (OMIM#103220) (Lindgren et al., 2015). The mutations in these genes can be associated with somatic mtDNA deletions or disruption of mitochondrial function in post-mitotic tissues. Mutations in *POLG* are especially considered as a major cause of disorders with multiple mtDNA deletions. Overall, four *POLG* heterozygous variants including a previously unreported missense variant, were identified in eight sIBM patients in this study (Lindgren et al., 2015). No association between abnormal length of CAG repeats in *POLG* and sIBM was found in this group of 26 patients. Moreover, heterozygous exonic variants were also identified in

other six genes except *SLC25A4*, *TK2* and *OPA1*. Among them, the frequency of a synonymous variant in *RRM2B* (p.V69=) was found to be significantly higher in sIBM patients compared with the control population from the EVS database. A synonymous variant in *C10orf2* (p.H524=) had also not been previously reported. Although synonymous variants are less likely damaging, they may have a negative effect via altering mRNA splicing and stability. The effect of the variants identified in this study remains to be further investigated. Overall, these findings indicate that the long-term inflammation in sIBM muscle fibres could contribute to DNA damage leading to the accumulation of mtDNA deletions (Lindgren et al., 2015). The variants in nuclear DNA could also affect mtDNA replication and maintenance. However, there is also a variation in the degree of mitochondrial involvement between patients (Rygiel et al., 2015), implying differences in the sensitivity of mtDNA to damage between individuals, or different mechanisms contributing to the pathogenic process. Further studies with more cases and longitudinal studies are necessary in the future.

3.4.2 Results

3.4.2.1 MAPT

A 238 bp deletion in intron 9 of the *MAPT* gene was screened in 68 Caucasian sIBM patients and 159 British controls. This was the first time that this *MAPT* haplotype was screened in a sIBM cohort. The PCR product with a size of around 500 bp corresponds to the common extended haplotype (H1) in the *MAPT* gene; while the product with a size of around 250 bp has a deletion in intron 9, and corresponds to the less common H2 sub-haplotype (Figure 3-8). The frequency of each genotype was compared between the sIBM cohort and the control group. However, no statistically significant difference in the frequencies of *MAPT* genotypes was found between the patients and the controls (Table 3-10).

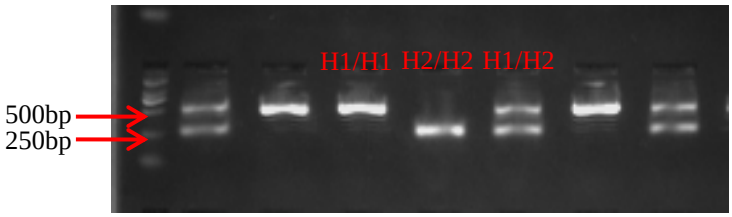


Figure 3-8 *MAPT* genotype PCR products in 2% agarose gel electrophoresis

H1 = H1 haplotypes; H2 = H2 haplotype.

Table 3-10 Frequencies of *MAPT* genotypes in sIBM patient group and a British control group

MAPT genotypes	sIBM patients (n=68)	British controls (n=159)	Mid-*P* value (two-tailed)	Odds ratio (95% CI)
H1/H1	38 (0.559)	90 (0.566)	0.919	0.97 (0.55-1.73)
H1/H2	28 (0.412)	58 (0.365)	0.507	1.22 (0.68-2.18)
H2/H2	2 (0.029)	11 (0.069)	0.254	0.41 (0.06-1.70)

H1 = H1 haplotype; H2 = H2 haplotype; CI = confidence interval.

3.4.2.2 PRNP

The frequencies of *PRNP* genotypes at codon 129 were compared between 44 Caucasian sIBM patients and 398 European controls. Seventeen patients (39%) were homozygous for methionine (M), eight patients (18%) were homozygous for valine (V), and 19 patients (43%) were heterozygous (Figure 3-9). The control data were pooled from five European studies as summarised by Orth et al. (Orth et al., 2000). In our sIBM cohort of 44 patients an even lower frequency of M/M genotype (39%) and an increased V/V genotype (18%) were found compared with the findings from the two previous studies. However, no significant difference was found in any genotype between the patients and the controls (Table 3-11).

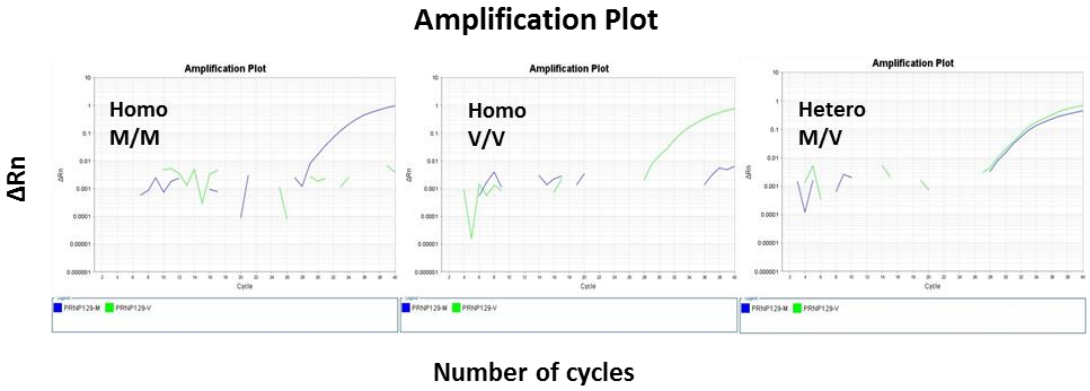


Figure 3-9 Genotyping of *PRNP* at codon 129 by TaqMan® SNP Genotyping Assay

The figure shows examples of amplification plots of *PRNP* common methionine/valine polymorphism at codon 129 determined by TaqMan® SNP Genotyping Assay on 7500 Fast Real-Time PCR System. ΔRn represents the normalised value of fluorescent signal. M = methionine (blue); V = valine (green).

Table 3-11 Frequencies of *PRNP* genotypes at codon 129 between sIBM patients and European controls

***PRNP* Codon 129 genotypes**	**Our sIBM patients (n=44)**	**European controls* (n=398)**	**Mid-*P* value (two-tailed)**	**Odds ratio (95% CI)**	**Lamp et al. 1999 (n=22) [Germany]**	**Orth et al. 2000 (n=41) [UK]**
M/M	17 (0.39)	156 (0.39)	0.951	0.98 (0.51-1.85)	14 (0.64)	21 (0.512)
M/V	19 (0.43)	198 (0.49)	0.415	0.77 (0.40-1.44)	7 (0.32)	19 (0.464)
V/V	8 (0.18)	44 (0.11)	0.185	1.79 (0.73-3.99)	1 (0.05)	1 (0.024)

M = methionine; V = valine; CI = confidence interval. *Pooled from five European studies (Orth et al., 2000).

3.4.2.3 APP

Exons 16 and 17 of the *APP* gene were screened for mutations in 69 sIBM patients by Sanger sequencing. No coding mutations were identified. This finding is consistent with a previous small study in both sIBM and fIBM patients.

3.4.2.4 C9orf72

A total of 187 sIBM samples were screened for GGGGCC repeats in *C9orf72* gene by RP-PCR, as described in Section 2.2.8.1. Thirty-three of them were screened on both blood and muscle DNA samples. None of them were found to have a large expansion of several hundred of repeats, and most of the 33 patients had consistent repeat length between blood and muscle. Of note, there were five samples with repeat length of more than 20 units in one allele. One sample had an allele of 27 repeat units, while another had an allele of 35 repeat units which is outside the common normal range (Figure 3-10). *C9orf72* sizing PCR and southern blot were also performed on these two samples to confirm the length of repeats by Ese Mudanohwo from the NHNN Neurogenetics Laboratory. The results of sizing PCR are shown in Figure 3-10. Southern blot did not detect typical pathogenic large expansion in both samples. The image of the southern blot is available in Appendix II – Figure 1.

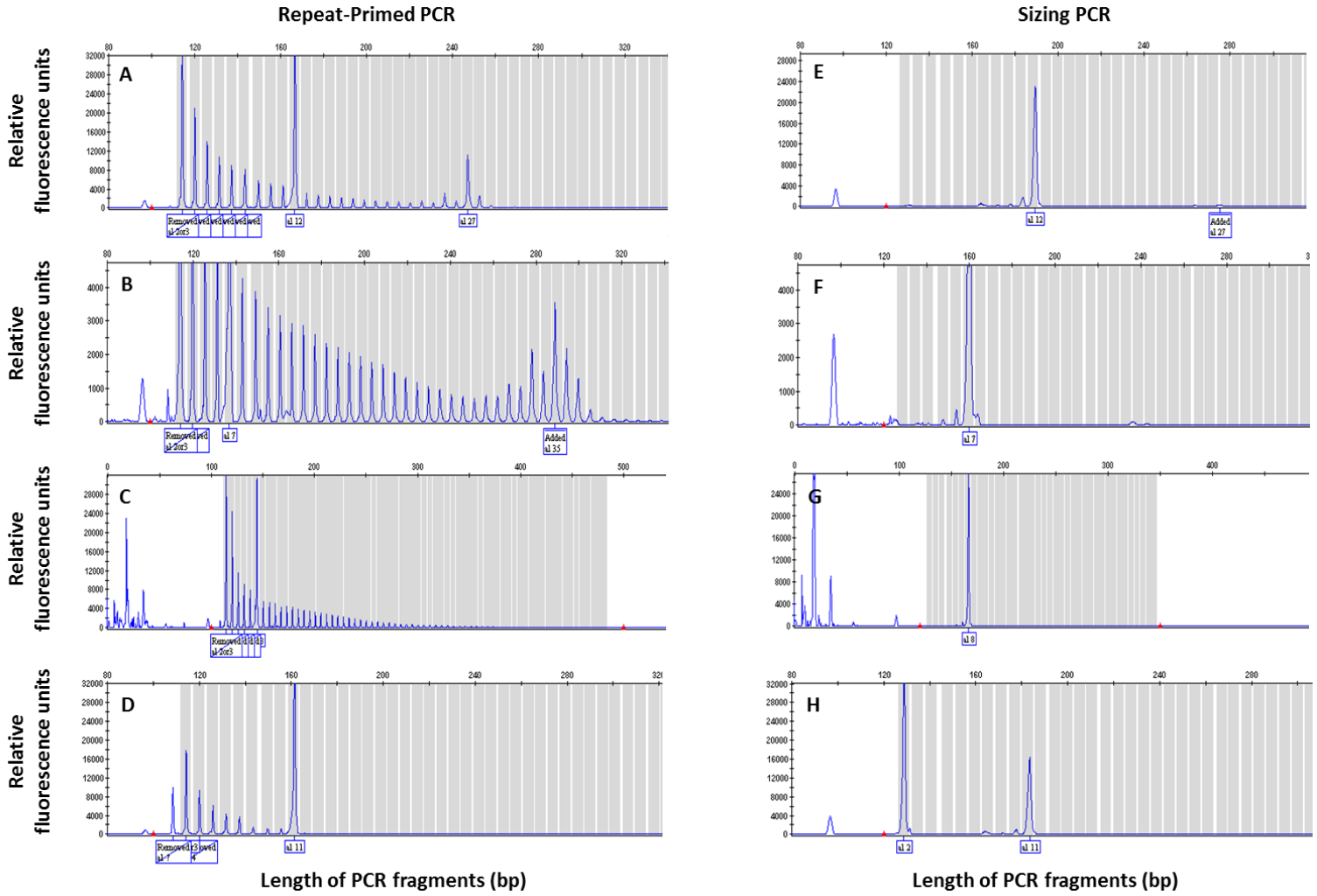


Figure 3-10 Repeat-primed and sizing PCRs for *C9orf72*

PCR products of repeat-primed PCR (left panels) and fluorescent fragment length analyses (right panels) separated on an ABI3730XL Genetic Analyzer and visualised by GeneMapper v3.7 software. Each peak (length > 100 bp) represents a repeat unit of GGGGCC. A sIBM patient carrying a large allele with 27 repeats and a small allele with 12 repeats (A and E); A sIBM patient carrying a large allele with 35 repeats and a small allele with 7 repeats (B and F); A positive control carrying an allele with 8 repeats and the other allele with a large *C9orf72* repeat expansion (C and G); A negative control with alleles of 2 repeats and 11 repeats respectively (D and H). (The data of sizing PCR was performed and kindly provided by Ese Mudanohwo from the NHNN Neurogenetics Laboratory).

3.4.2.5 Mitochondrial DNA

Long PCR was performed on 50 muscle DNA samples from sIBM patients to detect mtDNA deletion(s) as described in Section 2.2.9. No deletions in mtDNA were confirmed in this group of patients. (Appendix II - Figure 2)

3.4.3 Discussion

3.4.3.1 MAPT

Although the more common haplotype (H1) of *MAPT* is significantly over-represented in patients with other taupathies such as PSP (Baker et al., 1999), no such association was found in our cohort of sIBM patients. The presence or absence of this deletion in intron 9 does not contribute to the accumulation of tau protein in sIBM muscle. However, our study does not exclude other variants (including point mutations and splicing variants) beyond this region which may play a role in the pathogenesis of the disease. Screening the whole gene could be considered in future studies.

3.4.3.2 PRNP

Our analysis of the *PRNP* gene in 44 sIBM patients did not demonstrate a significant difference in methionine homozygosity at codon 129 between patients and normal controls. Whilst an increased frequency of valine homozygosity was found in our patient group, this did not reach statistical significance when compared with the controls. This suggests that polymorphisms at *PRNP* codon 129 are not genetic risk factors for sIBM. The association between homozygous methionine at codon 129 and the disease reported in the initial study is likely due to the small sample size.

3.4.3.3 APP

Direct sequencing of *APP* gene exons 16 and 17 in 80 sIBM patients did not identify any coding variants or splicing variants in these regions. This indicates that the accumulation of AβPP protein in muscle may not be caused by mutations in the exons 16 and 17 of *APP*. Instead, mutations in other parts of *APP* or other genes might modulate the *APP* gene transcription, leading to increased levels of AβPP mRNA. Future studies screening the complete *APP* gene could be considered.

3.4.3.4 C9orf72

This study identified two patients with 27 and 35 units of the hexanucleotide GGGGCC repeat amongst 187 sIBM patients. This is the first time that we found a *C9orf72* repeat of more than 25 units in sIBM patients. The typical pathogenic expansions associated with diseases consist of several hundred or thousands of units. However, expansions between 20 and several hundred units have also been reported in both patients with neurodegenerative syndromes and normal individuals (Millecamps et al., 2012, Beck et al., 2013, Gomez-Tortosa et al., 2013). Pathogenicity of these intermediate repeat lengths and the size of the smallest expansion unit that confers disease risk are still unclear. Some studies have suggested that they may be susceptible to unfaithful inheritance between generations or somatic instability (Beck et al., 2013, Waite et al., 2014). One study reported that a ~70 repeat *C9orf72* allele from the unaffected father expanded during parent-offspring transmission and passed a ~1750-unit *C9orf72* repeat expansion on to four children who were the first generation affected by ALS (Xi et al., 2015). This supported that intermediate expansions can potentially be unstable and result in larger expansion in next generations. A recent study reported a 30-unit repeat expansion inducing pathological lesions containing p62-positive inclusions with DPRs and RNA foci in brain tissues of a cognitively normal individual. In this study, the authors suggest that small expansion may be required to initiate the pathological changes, but are not sufficiently long to trigger the entire disease cascade (Gami et al., 2015). Both blood and muscle samples from 33 cases were also screened for *C9orf72* repeats. Six muscle samples failed in the RP-PCR, probably due to the very low concentration of DNA, and there were insufficient samples to repeat the experiment. No difference in the repeat length was found in the remaining cases. The patients with repeats of 27 units and 35 units do not have muscle samples available to investigate somatic instability.

3.4.3.5 Mitochondrial DNA

No confirmed mtDNA deletions were found amongst 57 muscle DNA samples. The main reason for this was that the low concentration of DNA extracted from muscle tissue was not sufficient to give clear PCR products, and the limited amount of samples was not enough to confirm suspected cases on Southern blot. Therefore, the possibility of mtDNA deletions in these sIBM patients cannot be excluded. A repeat

experiment should be considered at later stages of the project once more samples have been acquired.

3.5 Overall Conclusion

A total number of 239 sIBM patients have been fulfilled at least one of the three diagnostic criteria for sIBM – the Griggs criteria, ENMC 1997 criteria and MRC 2010 criteria based on IIBMGC (Table 3-2). Currently, there is not a "gold standard" criteria for diagnosing sIBM. Many diagnostic categories were proposed by different study groups. Twenty-four previously reported diagnostic criteria for sIBM were evaluated in a recent study where concluded that they all performed with high specificity (≥ 97%) but wide-ranging sensitivities (11%-84%) (Lloyd et al., 2014). The true sIBM patients would be potentially missed by low sensitivity diagnostic criteria. This will add difficulty to the patient recruitment especially for studying such a rare disease. However, to increase the sensitivity by reducing the specificity would also increase the chances of "contamination" with other diseases such as polymositis. It is challenging to balance both sensitivity and specificity using single diagnostic criteria. Therefore, despite such a limitation of current available diagnosis criteria, applying the three most widely used criteria above as our study inclusion criteria would allow us to increase the sensitivity and also keep a high specificity of patient recruitment.

In this chapter, I studied several candidate genes in the sIBM patients group, including *APOE-TOMM40* genotypes, *SQSTM1* and *VCP* genes, *MAPT* genotypes, *PRNP* codon 129 genotypes, *APP* gene (exons 16 and 17), hexanucleotide repeat expansion in the *C9orf72* gene, and mtDNA deletion(s). I reported that i) the *TOMM40* VL polyT repeat has a disease modifying effect which is associated with a later onset of sIBM symptoms, and this effect is further enhanced by the *APOE* ε3/ε3 genotype; ii) rare variants in *SQSTM1* and *VCP* genes are present in approximately 4% of our sIBM cohort, which suggests a genetic overlap between sIBM and other neurodegenerative diseases such as ALS and FTD; iii) two sIBM patients carrying *C9orf72* intermediate repeat expansions at 27-unit and 35-unit respectively, but their pathogenicity is still unclear. No association was found in other genes.

The results of some candidate gene studies were negative, such as in analyses for *MAPT* genotypes, *PRNP* genotypes, and *APP* gene. This may be due to small sample sizes. However, an issue of citation bias was reported in a study. The author indicated that high citation rate of some immunohistochemical findings of abnormal protein accumulation (e.g. β amyloid precursor or β amyloid) in sIBM muscle stainings were caused by citation bias (Greenberg, 2009). The negative findings of studying the genes encoding for some accumulated proteins may also be a result of contentious immunohistochemical findings. In this case, validations of many protein deposits are very important by different research groups in further studies on sIBM muscle tissues, in order to avoid the misleading to candidate genetic studies.

In addition to the genes discussed above, other candidate genes could also be studied in the future, such as the ones associated with ALS/FTD. Next-generation sequencing has become a fast and cost-efficient approach to identify new genes and genetic variants; however it is currently limited in exploring large deletions/insertions and repeats in gDNA, as well as defects in mtDNA. Therefore, it is important to combine conventional genetic techniques and next-generation sequencing in studying the genetics of a complex disease like sIBM. A WES association study on a large sIBM cohort will be discussed in the next chapter.

Chapter 4 Whole-Exome Sequencing Association Analyses in a Large Cohort of sIBM and Neuropathologically Healthy Controls

4.1 Introduction

sIBM is a complex disease which could be influenced by polygenic factors. Current genetic studies on sIBM are mainly candidate gene association analyses by Sanger sequencing, genotyping, and targeted next generation sequencing. However, these approaches are not sufficient to identify all the genetic events underlying a complex disease like sIBM. A traditional population-based genome-wide association analysis (GWAS) is a well-established method for gene discovery without prior hypotheses. This has been particularly successful in identifying novel common genetic loci for the common diseases. However, such a study requires several thousands of cases to be able to detect the likely modest underlying genetic relative risks (Zondervan and Cardon, 2007). Such a large sample size would be impractical for rare diseases like sIBM. In addition, the common variants tend to have relatively small effects or low-risk, and explain only a fraction of the overall heritability (Manolio et al., 2009). Evolutionary theory predicts that the deleterious variants should be rare as a result of purifying selection, and the variants with low-frequency likely explain a greater proportion of the heritability (Eyre-Walker, 2010, Gibson, 2011). This suggests that the rare variants which are usually defined as having minor allele frequency (MAF) < 1% or < 0.5% present modest or high-risk for the complex traits especially in rare diseases and rare forms of common diseases (Gibson, 2011). However, the rare variants are not well-detected in GWAS genotyping arrays. Therefore, new genetic approaches are required to study rare variant association with sIBM.

4.1.1 Rare Variant Association Study (RVAS)

With the advent of next-generation sequencing and affordability of whole-exome sequencing (WES), many causal variants have been identified for Mendelian disorders (Ng et al., 2010a, Ng et al., 2010b). Some rare variants have also been identified as genetic susceptibility for complex diseases (Guerreiro et al., 2013, Cruchaga et al., 2014, Lange et al., 2014). The main challenge of traditional single variant association methods is the substantial decrease in power when analysing rare variants. For example, if one assumes 5% disease prevalence and a significance level of 5×10^{-8}, the sample sizes required to achieve 80% power with an odds ratio (OR) = 1.4 are 54,000 and 540,000 for a MAF = 0.01 and 0.001, respectively (Lee et al., 2014). The common approach for increasing power is to aggregate variants in sets based on their functional annotation. The variants can be grouped into analysis models according to different sources, for example, i) all exonic variants located in a given gene, ii) a subset of variants based on their functional impact such as stop-gain, missense, frameshift, splicing and etc, iii) a subset of exonic variants based on the possibility of pathogenicity predicted by SIFT and PolyPhen, iv) a subset of exonic variants with large functional scores from computational tools such as GERP++ or CADD (Nicolae, 2016). As a growing number of studies now are using WES to investigate the rare variant association with the disease phenotypes, numerous methods for rare variant association tests have been developed in addition to single variant association tests. There are three main methods – burden tests, variance component tests, and combination tests (Lee et al., 2014, Nicolae, 2016). These will be discussed in the following paragraphs. Other methods are also available such as adaptive burden tests, exponential combination (EC) test, least absolute shrinkage and selection operator (LASSO), replication-based test (RBT), and Bayesian methods (Lee et al., 2014, Nicolae, 2016).

4.1.1.1 Burden Tests

Burden tests function by collapsing information (e.g. counting the number of minor alleles) for multiple genetic variants within a gene or a defined region of the genome into a single genetic score, and testing for association between the score and a trait (Madsen and Browning, 2009, Morris and Zeggini, 2010). Several burden methods have been proposed, such as cohort allelic sums test (CAST) (Morgenthaler and

Thilly, 2007) and a combination of rare and common variants in the combined and multivariate collapsing (CMC) test (Li and Leal, 2008). The burden score can be modified such as the weighted-sum collapsing (WSC) approach to implement a weight based on functional category and/or allele frequency of variants in the model (Madsen and Browning, 2009). However, burden tests are limited by its implicit assumption that all rare alleles in a set are associated with a trait and act in the same direction (e.g. all risk or all protective) (Morgenthaler and Thilly, 2007, Li and Leal, 2008, Madsen and Browning, 2009, Morris and Zeggini, 2010).

4.1.1.2 Variance Component Tests

Variance component tests include the sequence-based kernel association test (SKAT) (Wu et al., 2011), C-alpha test (Neale et al., 2011), the evolutionary mixed model for pooled association testing (EMMPAT) (King et al., 2010), and the sum of squared score (SSU) test (Pan, 2009). They allow a mixture model of effects across a group of rare variants and effect sizes of different magnitude, and test for association by evaluating the distribution of the aggregated score test statistics of these variants. For example, SKAT is robust to aggregate both variants with positive effects and variants with negative effects (Wu et al., 2011). Therefore, this class of methods are more powerful than burden tests if a region contains many non-causal variants or there is a mixture of both risk and protective variation. However, compared to burden tests, variance component tests lose power if the majority of variants are causal with the same direction of effect.

4.1.1.3 Combination Tests

Combination tests are designed to blend strengthens of both burden tests and variance component tests. An optimal SKAT (SKAT-O) method (Lee et al., 2012b, Lee et al., 2012a) is proposed as a linear combination of SKAT statistic (T_S) and burden test statistic (T_B) with the weight coefficient (α) or data adaptive. Its test statistics are given by $T_O = \alpha T_S + (1 - \alpha)T_B$. When there is no prior information on genetic architecture, combined tests are a more attractive choice to have robust power across a range of disease models. Therefore, SKAT-O was used for all the gene-based models in this study.

4.1.2 Aim of the Study

The aim of this study is to explore sIBM case-control WES associations by single variant and gene-based analyses on both common and rare variants, in order to identify high-risk genetic factors for sIBM in European populations.

4.2 Method

4.2.1 Patient and Control Cohorts

This study is an IIBMGC and Muscle Study Group endorsed project. A total number of 181 sIBM exomes (the same cohort as in Section 3.3) were used for the whole-exome association study. The control cohort consisted of a total number of 510 neuropathologically healthy individuals (five individuals with age information missing; mean age = 55.2 ± 25.7 years) kindly provided by Professor Andrew Singleton at the National Institute of Health, USA.

4.2.2 Whole-Exome Sequencing and Bioinformatics

The WES of the sIBM patient cohort was performed at ION as described in Sections 2.3.1-2.3.2. The whole-exome data of controls were generated by using Illumina TruSeq Exome Enrichment (62 Mb) Kit (Illumina) on the platform HiSeq2500 (Illumina) at the National Institute of Health, USA. The alignment of sequence reads from FASTQ files to generate BAM files, the removal of PCR duplicates, and realignment of indels were performed on both patients and controls using the ION bioformatics pipeline as described in Section 2.3.3. After that, each BAM file was run through a program called HaplotypeCaller which is capable of calling SNPs and indels simultaneously via local re-assembly of haplotypes. This is more accurate when calling regions that are traditionally difficult to call, such as when they contain different types of variants close together. The output is so-called genomic VCF (variant call format) (gVCF) which can be used for joint genotyping of multiple samples. This is an efficient way to process samples in a large cohort size. A number of 50-100 gVCFs were merged as a batch via CombineGVCFs and then passed to GenotypeGVCFs to generate a combined and genotyped VCF file. The programs

HaplotypeCaller, CombineGVCFs and GenotypeGVCFs are based on GATK the Best Practice protocol. Subsequently, the VCF file was filtered for poor genotype quality, which will be described in the next section. The above workflow is shown as Figure 4-1. The script files, such as Haplotype_Caller.sh, Genotype_gVCF.sh, and Filter_VCF.sh, are available in Appendix III - 1. These steps of data processing were performed by Dr Alan Pittman using the ION bioinformatics pipeline.

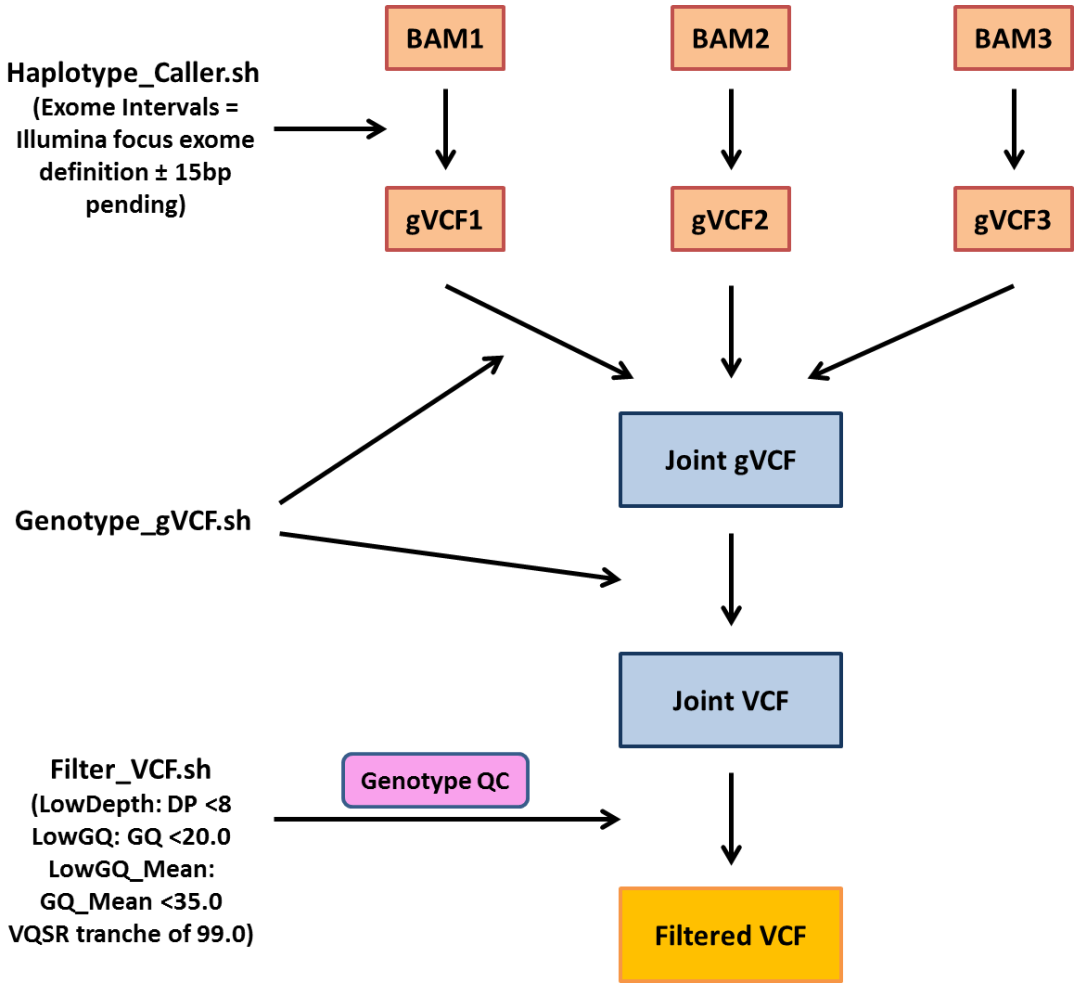


Figure 4-1 Workflow of the first stage data processing to generate a joint filtered VCF file

This figure shows an example of generating a joint VCF file using three samples. It is filtered for the low depth and low genotype quality (GQ), which is called genotype quality control (QC), and then it is also filtered for variants outside the range of VQSR tranche of 99.0. In this study, the sIBM exomes and controls exomes were processed separately, and each batch of the joint gVCF file contains 50-100 samples. Eventually, all the sIBM batches and control batches were combined together for the downstream QC and analyses. This step was performed by Dr Alan Pittman on the ION bioinformatics pipeline.

4.2.3 Quality Controls (QC)

Despite the advantages of WES, it can potentially produce genotyping errors throughout the sequencing process, especially at sites with low coverage or variants with low minor allele frequency (MAF), which are called as systematic biases. Issues with sampling or potentially unknown ethnic background could also increase the likelihood of producing biases. Introduction of these biases into genetic case-control association analyses will lead to an increased number of false-positive and false-negative associations. Therefore, a thorough quality control (QC) of the WES data is critical to identify the problematic markers and samples, which should be removed before further analyses to reduce the number of false-positive and false-negative associations. The QC protocol used in this study was modified from the recommended protocol for GWAS by Anderson et al. (Anderson et al., 2010). Three main steps of QC were included – genotype QC, variant QC, and sample QC. The genotype QC was performed on the ION bioinformatics pipeline as shown in Figure 4-1. The variant QC and sample QC were mainly performed using an open-source C/C++ tool, PLINK software version 1.9 (https://www.cog-genomics.org/plink2) (Chang et al., 2015) by the author of the thesis, and a workflow was shown as Figure 4-2. All the scripts for PLINK are available in Appendix III - 2.

4.2.3.1 Genotype QC

Genotype QC is an additional step which is not included in the Anderson's protocol, but is recommended by Carson et al. specifically for population-based WES studies (Carson et al., 2014). First, this uses genotype-level quality metrics including depth of data (DP), genotype quality (GQ), and an average of GQ (GQ_Mean), which are generated via GATK variant calling program. DP value represents the number of reads passing QC used to calculate the genotype at a specific site in a specific sample. GQ is a Phred-scaled value representing the confidence that the called genotype is the true genotype. GQ_Mean represents the average value of GQ at a specific site across all the samples in one batch. Higher values for DP, GQ and GQ_Mean generally reflect more accurate genotype calls. The recommended filtering thresholds of these metrics are a minimum of eight reads for DP, a minimum of 20 for GQ, and a minimum of 35 for GQ_Mean (Carson et al., 2014). The genotypes which have

either DP < 8, GQ < 20, or GQ_Mean < 35, were labelled as "NoCall" via GATK VariantFiltration and SelectVariants functions.

A VQSR (variant quality score recalibration) score was assigned to each SNP and indel at a call site. A threshold of VQSR at 99.0 is recommended by GATK maintaining 99% sensitivity for the "true" variants.

4.2.3.2 Variant QC

Variant QC consists of two steps: i) identification of SNPs showing a significant deviation from Hardy-Weinberg equilibrium (HWE), and ii) identification of SNPs with an excessive missing genotype call rate. The criteria to filter out the poor-quality variants have varied between studies. Since the sIBM and control samples were sequenced in two different centres and using different reagents as mentioned previously, a stringent QC strategy was applied in this study.

Extensive deviation from HWE can be indicative of a genotyping or genotype-calling error; however, deviations from HWE may also indicate selection. For example, a case sample showing deviations from HWE may indicate that the loci is in association with the disease, which should not be removed from further investigation. Therefore, only control samples were used when testing for deviations for HWE. The *p*-value threshold for declaring SNPs to be in HWE was recommended at 0.001 by PLINK. This means that SNPs with a HWE *p*-value less than 0.001 were removed.

The missingness per variant was calculated across the whole dataset. A filtering threshold of the missingness per variant at 0.015 was used. This means that the variants with call rate below 98.5% were removed.

4.2.3.3 Sample QC

Sample QC consists of six steps: i) identification of individuals with potential sample contamination, ii) identification of individuals with excessive missing genotype call rate, iii) identification of individuals with discordant sex information, iv) identification of individuals with outlying proportion of heterozygote genotypes, v) identification of duplicated or related individuals, and vi) identification of individuals of divergent ancestry.

Contamination Detection

The software VerifyBamID (http://genome.sph.umich.edu/wiki/VerifyBamID) was used to check whether the reads were contaminated as a mixture of two samples (Jun et al., 2012a). It requires a VCF file containing external genotypes or allele frequency information and the BAM file of each sample as input files. In this study, 1000 Genomes project phase1 data with allele frequency information was formatted into the required input VCF file, and predicted genotypes are based on provided allele frequencies. Freemix Fraction is widely used to estimate the contamination per sample in next-generation sequencing. The program treats reads as coming from two samples, and estimates fraction of reads that come from a second sample. A FREEMIX value ≥ 0.03 and a large log likelihood ratio value (FREELK1-FREELK0) suggest a potential contamination as proposed by the software.

Call Rate per Sample

The genotype failure rate is an important measure of DNA sample quality. After calculating the distribution of missing genotype call rates across the whole dataset, the threshold of 0.2 was selected. The individuals with more than 20% missing genotypes were excluded.

Gender Mismatch

To detect discordancy between genotype information and the recorded sex for each individual is to calculate the homozygosity rate (estimated by coefficient of inbreeding [F]) across all X-chromosome SNPs and compare these with the expected value. Typically, males are expected to have a homozygosity rate of 1, and females to have a homozygosity rate smaller than 0.2 (Anderson et al., 2010). When a sample with discordant sex information is detected, the sample should be excluded from further analysis, unless it can be confirmed that the sex was recorded incorrectly. In this study, when the homozygosity rate (F value) was more than 0.2 but less than 0.6, the genotype data was considered to have obscure sex information, which caused the individual to be removed.

Heterozygosity Rate per Sample

Heterozygosity rate per individual is another measure of DNA sample quality. The distribution of mean heterozygosity is calculated across the whole dataset, and an

excessive or reduced proportion of heterozygote genotypes indicate a potential DNA sample contamination or inbreeding. A heterozygosity rate ± 3X standard deviation (SD) from the mean is recommended (Anderson et al., 2010).

Duplicates and Relatedness

To identify duplicate and related individuals, the degree of shared ancestry for a pair of individuals (identity by descent, IBD) is calculated. This analysis works best when only independent SNPs are included in the analysis. A high-LD-regions.txt containing regions of extended linkage disequilibrium (LD) is available online from the paper by Anderson et al., and this was used to remove the high LD regions from our cohort. The expectation is that IBD = 1 for duplicates or monozygotic twins, IBD = 0.5 for first-degree relatives, IBD = 0.25 for second-degree relatives and IBD = 0.125 for third-degree relatives (Anderson et al., 2010).

Population Stratification

The frequency of variants within a gene can differ between populations. A mix of different ethnic groups in a case-control study, also called population stratification bias, can result in false positive or false negative associations. Principal component analysis (PCA), a multivariate statistical method, is the most common method for identifying differences in ancestry. The PCA model was applied to all the samples and HapMap genotype data from individuals of European, Asian and African origin (Anderson et al., 2010), using a tool called Genome-wide Complex Trait Analysis (GCTA) (http://cnsgenomics.com/software/gcta/pca.html) (Yang et al., 2011) which is now built into PLINK version 1.9. The PCA was plotted using R 3.1.1 (https://www.r-project.org/) in RStudio. In this study, only European ancestries from both cases and controls were selected for the association analyses. The scripts for R commands are available in Appendix III - 3.

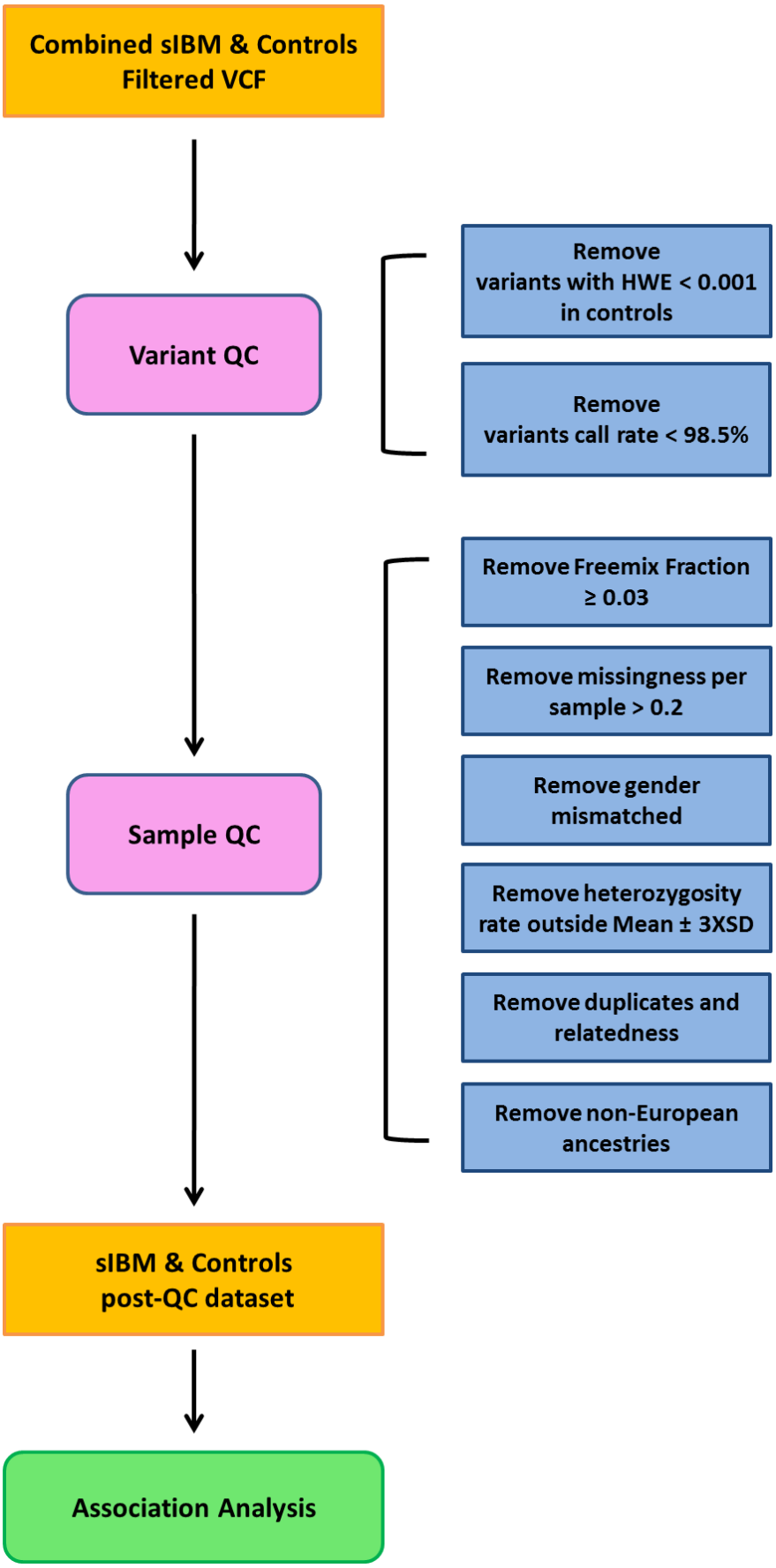


Figure 4-2 Workflow of variant and sample QC for whole-exome association analyses

4.2.4 Association Analyses and Statistics

4.2.4.1 QQ Plot

The log quantile-quantile (QQ) plot of *p*-value distributions was created by using the R software package to check for systematic bias after all the QC steps. QQ plot is to compare the observed distribution of *p*-values for all SNPs on the y-axis with their expected values on the x-axis under the null hypothesis of no association. A deviation of *p*-values from the y = x line corresponds to a deviation from the null hypothesis. A deviation among the bottom 90% of the *p*-values from the y = x line suggests the existence of excessive false positive associations due to systematic bias (such as the possibility of population stratification). A sharp deviation from the y = x line among the top right corner (the smallest *p*-values) implies a potential true association (Balding, 2006). A genomic inflation factor λ is calculated among the bottom 90% of the *p*-values to estimate the influence of population stratification on distribution of χ^2 association test statistics. It is the expectation that, $\lambda = 1$ if the data follows the normal χ^2 distribution. If $\lambda > 1$, this indicates that population stratification may exist and need to be corrected (Anderson et al., 2010). WES produces a large amount of rare variants and deviations could be caused by some extremely rare variants. Therefore, in this analysis only common SNPs (MAF > 0.05) were selected. The scripts for R commands are available in Appendix III - 4.

4.2.4.2 Power Calculation

The study power was calculated retrospectively based on the current sample size after QC using an online software tool GAS Power Calculator (Skol et al., 2006) (http://csg.sph.umich.edu//abecasis/CaTS/gas_power_calculator/index.html).

4.2.4.3 Single Variant Analyses

Association analyses for single variants were performed using specific commands in PLINK version 1.9. All the scripts are available in Appendix III - 5. First, a standard case-control allelic association analysis based on χ^2 test was performed on the dataset containing both common and rare variants. Subsequently, it was tested for association using logistic regression including sex as a covariate. Bonferroni's adjustment, Holm, Sidak and false discovery rate (FDR) methods, and permutation were performed to correct the multiple testing. The family-wise error rate (FWER)

threshold was set at α = 0.05. The adjusted *p*-value after correcting for multiple testing below 0.05 is considered as significant.

Haploview software (Barrett et al., 2005) was used to create a Manhattan plot for the χ^2 allelic test of association. SNPs with significant *p*-values are easy to distinguish, corresponding to those values with large log10 *p*-values.

4.2.4.4 Gene-Based Analyses

SKAT-O method was applied to analyse the multiple variant effects within each gene using R software package "SKAT" (Lee et al., 2012b). Five different models for defining variants were used for analyses: i) all variants regardless of frequency (All Variants Model), ii) all non-synonymous variants regardless of frequency (Coding Model), iii) all non-synonymous variants with MAF < 1% (Coding Model for Rare Variants), iv) all non-synonymous variants with MAF < 1% which were predicted as not benign by SIFT and PolyPhen (Not Benign Model for Rare Variants), and v) only stop-gain, frameshift and canonical splice variants with MAF < 1% (Loss-of-Function [LOF] Model for Rare Variants). The Variant Effect Predictor (VEP) software (McLaren et al., 2010) was used to annotate the consequences of the variants passed the QC and select the variants according to the above five groups. The scripts for R and VEP are available in Appendix III – 6 and 7. Analyses for the above five models were first performed on all genes across the whole dataset. Then, analyses were performed on a group of 127 candidate genes. This contains genes associated with neurodegenerative diseases (such as ALS, FTD, AD, and Prion diseases), genes associated muscle diseases (such hIBM, myopathies with rimmed vacuoles, and other IBM-like myopathies), immune genes previously studied in sIBM, genes encoding RNA binding proteins, gene encoding mitochondrial proteins, and genes suggested by personal communications from collaborators or conferences (Table 4-1). Sex was included as a covariate. Bonferroni's adjustment was used for correcting multiple testing. The FWER threshold was set at α = 0.05. *P*-value for Bonferroni adjustment < α* (α* = α/n [n is the number of gene sets and the number of analysed groups]) is considered as significant.

Table 4-1 List of candidate genes for gene-based analyses (to be extended when more candidate genes are proposed)

Gene Group	Gene ID
Genes associated with ALS and FTD	*TARDBP, OPTN, DAO, PRPH, TBK1, ANG, FUS, PFN1, ALS2, DCTN1, VAPB, SOD1, NEFH, CHMP2B, NEK1, MATR3, SQSTM1, FIG4, PON1, PON2, SETX, C9orf72, VCP, UBQLN2, UBQLN1, ATXN2, HNRNPA1, SPG11, GRN, TUBA4A, SPAST, SS18L1, CHCHD10, EWSR1, VEGFA, ELP3, GLE1, SIGMAR1, HDAC6, CTST*
Gene associated with other neurodegenerative diseases, such as AD, PD, and Prion disease	*APOE, MAPT, PSEN1, PSEN2, SERPINA3, TREM2, PRNP, APP, AKT1, ATN1, CAND1, CRHR1, TYROBP, TTC27, EP300*
Genes related to muscle diseases, such as hIBM, IBM-like and rimmed vacuolar myopathies	*GNE, LAMA2, MYOT, DPM3, LMNA, SEPN1, POMGNT1, BAG3, LDB3, CRYAB, ANO5, BSCL2, SGCG, PABPN1, SYNE1,SYNE2, POMT2, CAPN3, MYH2, TCAP, SGCA, SMCHD1, FKRP, TTN, DES, DYSF, TMEM43, DAG1, GMPPB, TRAPPC11, SGCB, SGCD, FLNC, TNPO3, DNAJB6, ISPD, HNRNPA2B1, FKTN, TRIM32, POMT1, FHL1. EMD, DMD, HNRNPDL, DUX4*
Immune genes studied in IBM previously	*NT5C1A, HLA-DRB1, HLA-DRA, CD74, BTNL2*
Genes encoding mitochondrial proteins	*POLG, TOMM40, C10orf2, TYMP, SLC25A4*
Genes encoding other RNA binding proteins	*DNAJB2, TIA1, TAF15, PTRF, CEP112, IRF3*
Genes suggested by personal communications from collaborators/ conferences (not published)	*UNC45B, KIDINS220, FYCO1, CCR2, CCR3, CCR5, CCRL2, LTF, XCR1, CXCR6*

ALS = amyotrophic lateral sclerosis; FTD = frontotemporal dementia; AD = Alzheimer's disease; PD = Parkinson's disease; hIBM = hereditary inclusion body myopathy.

4.3 Results

4.3.1 Coverage of Whole-Exome Sequencing

All the sIBM samples were sequenced in-house, and the metrics of the exomes were only available for sIBM cohort by the time of writing. Of a total number of 181 sIBM samples, 32 were sequenced by TruSeq Exome Enrichment (62 Mb, Illumina)

protocol and 149 were by Nextera Rapid Capture Exome (37 Mb, Illumina) protocol. There were only eight exomes which had less than 80% of the target covered at 10X reads. The average percentage of the target covered at 10X was 91% for the whole sIBM cohort (Table 4-2).

Table 4-2 Metrics of sIBM whole-exome data

Sequencing Protocol	Average % Target Covered at 10X	Average % Target Covered at 20X
TruSeq = 32	86.6%	75.2%
Nextera = 149	92.0%	80.2%
Total = 181	91.0%	79.3%

4.3.2 Variant QC

There were a total number of 691 samples (sIBM = 181; controls = 510) and 371264 variants in the whole dataset after genotype QC. The variant failure call rate was calculated across the whole dataset before variant QC shown in Figure 4-3. First, variants with an HWE *p*-value less than 0.001 in controls were removed from the whole dataset. This left a total number of 367834 variants. Then, variants that failed the call rate at 98.5% were removed. This left a total number of 105883 variants into further analyses.

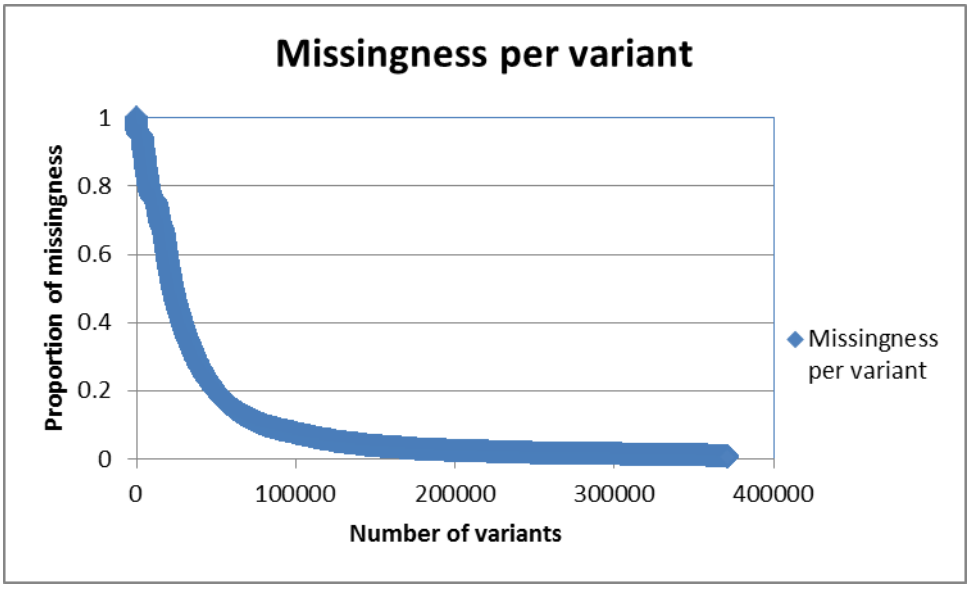


Figure 4-3 Distribution of missingness per variant across the whole dataset before variant QC

Each diamond symbol represents a variant. Variants with missingness > 1.5% were removed.

4.3.3 Sample QC

4.3.3.1 Freemix Fraction

Due to the control BAM files taking up a large data-computing space, freemix fraction was calculated for each patient sample only, but not for controls within the time constraint of this thesis (Figure 4-4). Thirty-one samples had a freemix fraction ≥ 0.03, suggesting potential contamination amongst them. These were removed at end of sample QC.

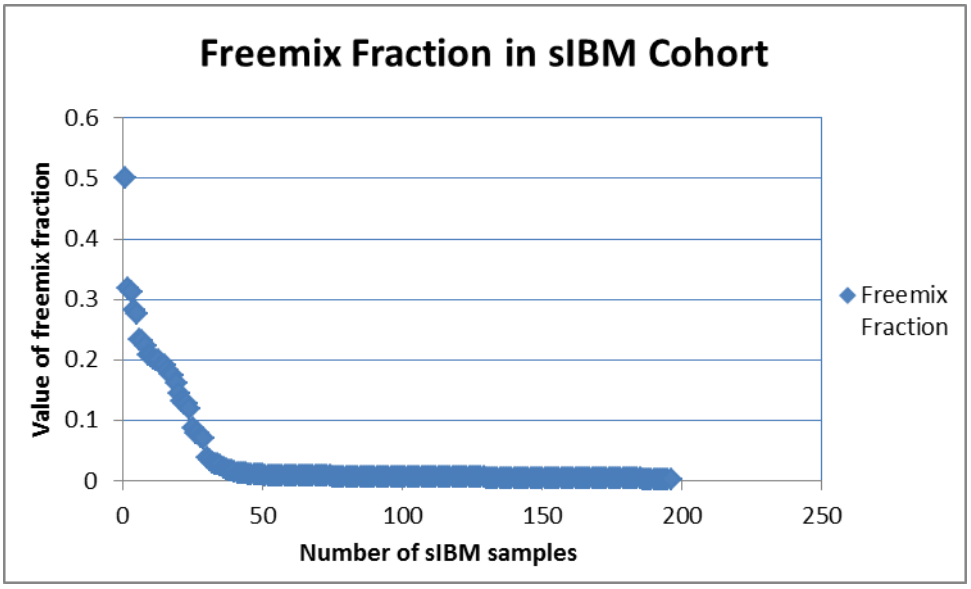


Figure 4-4 Distribution of the freemix fraction in sIBM cohort

Each diamond symbol represents a sample. The samples with freemix fraction ≥ 0.03 were removed.

4.3.3.2 Genotype Call Rate per Sample

The genotype failure call rate per sample was calculated across the whole dataset before and after variant QC, shown in Figures 4-5 and 4-6. The majority of the samples had a failure call rate below 20%. The number of samples with a failure call rate above 20% was reduced after variant QC, suggesting that the quality of genotype call rate per sample was improved after removing the low quality variants. Therefore, the samples which passed variant QC but still had a failure call rate above 20% were removed at the end of sample QC. Eight control samples were identified to have failed this threshold.

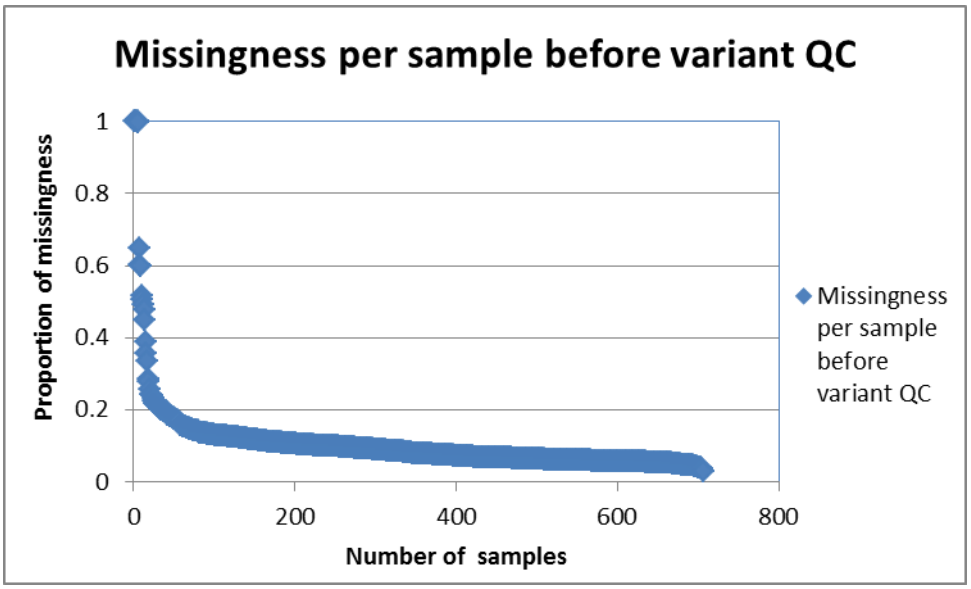


Figure 4-5 Distribution of missingness per sample across the whole dataset before variant QC

Each diamond symbol represents a sample. No sample was removed at this point.

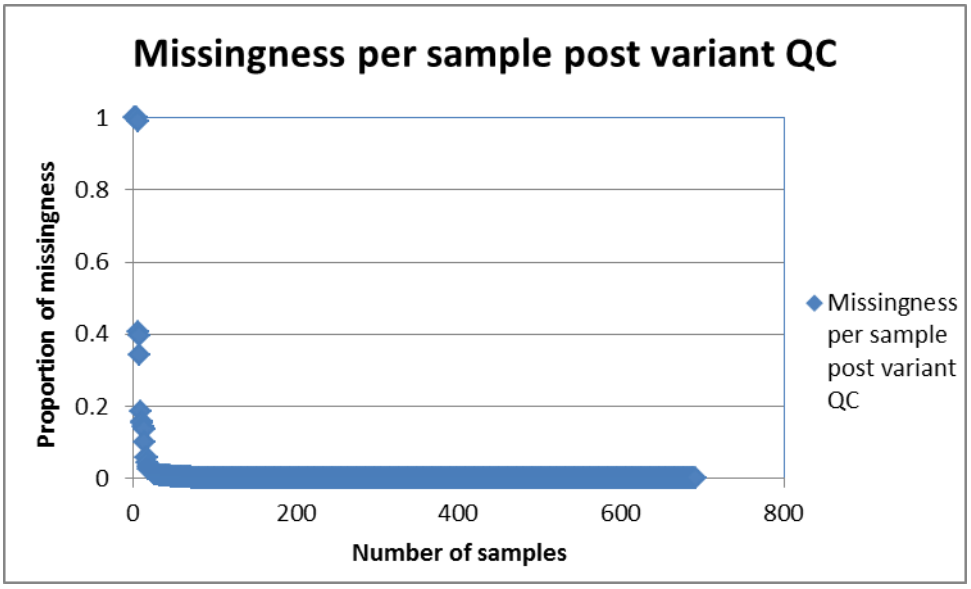


Figure 4-6 Distribution of missingness per sample across the whole dataset post variant QC

Each diamond symbol represents a sample. The samples with missingness > 0.2 were removed.

4.3.3.3 Gender Mismatch

The whole dataset before variant QC was used for calculating the X chromosome homozygosity rate (coefficient of inbreeding [F]) per sample (Figure 4-7), as the variants on X and Y chromosomes would be removed by variant QC due to their low call rate. A total number of 49 samples (sIBM = 18; controls = 31) were identified either with inconclusive sex information (F = 0.2 – 0.6) or with discordant gender information, which were removed at the end of sample QC.

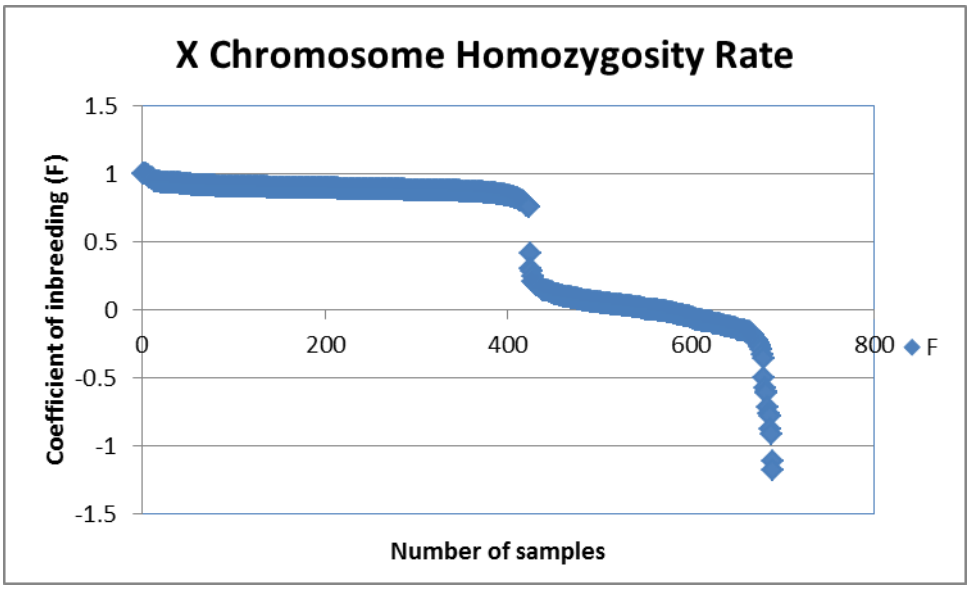


Figure 4-7 Distribution of X chromosome homozygosity rate per sample across the whole dataset before variant QC

Each diamond symbol represents a sample. The samples with F between 0.2 and 0.6 were removed.

4.3.3.4 Heterozygosity Rate

The whole dataset passed variant QC was used for calculating heterozygosity rate per sample across the whole dataset (Figure 4-8). The range of the threshold (Mean ± 3XSD) was 0.019972 - 0.051636. There were 23 samples (sIBM = 16; controls = 7) outside this range, which were removed at the end of sample QC.

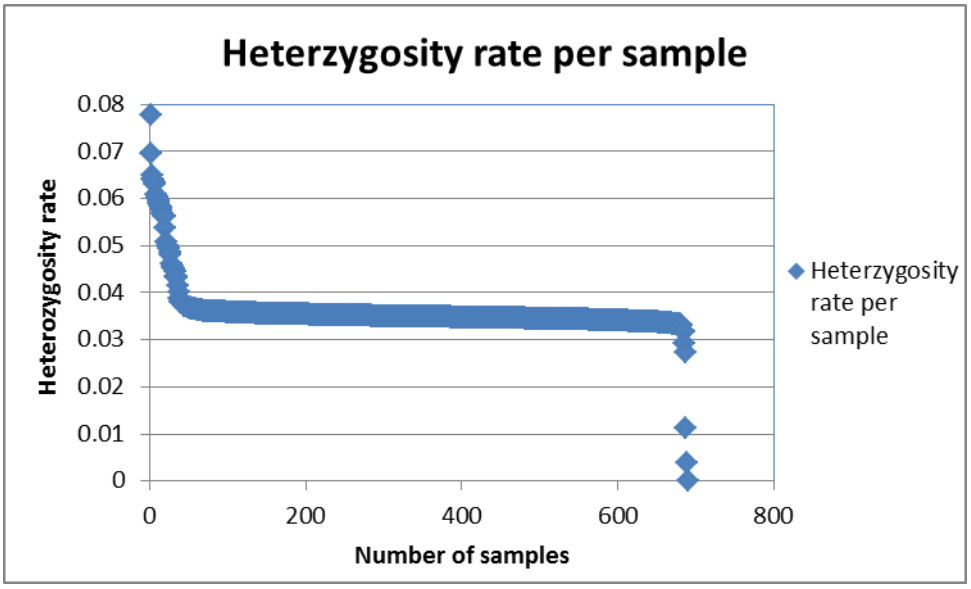


Figure 4-8 Distribution of heterozygosity rate per sample across the whole dataset

Each diamond symbol represents a sample. The samples with heterozygosity rate outside the range between 0.019972 and 0.051636 were removed.

4.3.3.5 Duplicated and Related Samples

Two control samples were identified with IBD = 1, suggesting that these two were duplicates. One of the two controls was removed. Two sIBM samples were identified with IBD = 0.9993 and the other two sIBM samples with IBD = 0.5. After checking the study proforma of each individual, they were four unrelated patients. This indicates there may be contamination during sampling and sample handling. Therefore, all these four sIBM samples were removed at the end of sample QC.

4.3.3.6 Population Stratification

Ancestry clusterings of both sIBM cases and control individuals in combination with HapMap genotype data from individuals with European (CEU), Asian (CHB+JPT) and African origin (YRI) were shown in a PCA plot (Figure 4-9). The majority of cases and controls were overlapped with each other and close to the HapMap European population, suggesting that they had matched European ancestry background. There were 13 sIBM and eight control samples with PC1 > 0.0 or PC2 < 0.0 were identified as outliers, which were removed at the end of sample QC.

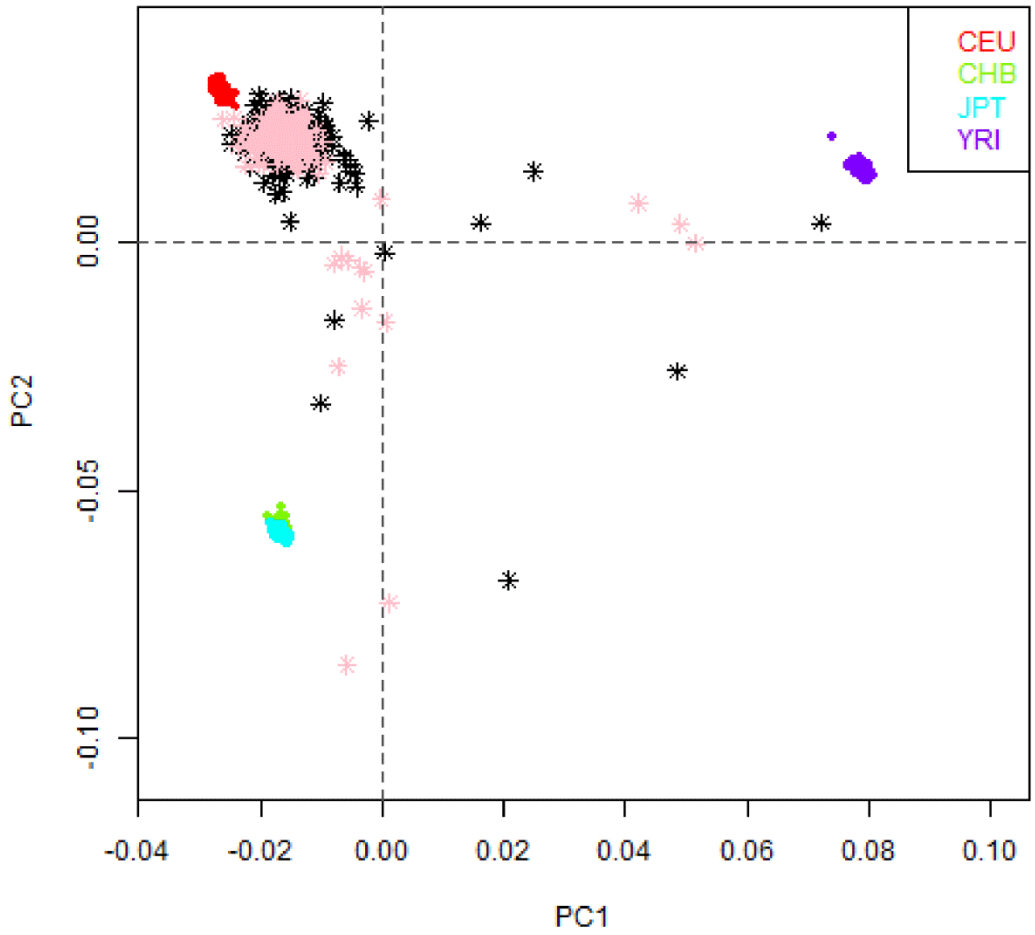


Figure 4-9 A principal component analysis (PCA) plot showing population stratification of the whole dataset

This shows population stratification of both sIBM cases and controls in combination with HapMap genotype data from Europe (CEU), Asia (CHB+JPT) and Africa (YRI). Each asterisk (*) represents one sample. Pink = cases; Black = controls; Red = CEU (Utah residents with ancestry from northern and western Europe); Green = CHB (Han Chinese in Beijing, China); Light blue = JPT (Japanese in Tokyo, Japan); Purple = YRI (Yoruba in Ibadan, Nigeria); PC1 = principal component 1; PC2 = principal component 2.

All the outliers from each sample QC step above were removed at this stage. A total of 601 samples (sIBM = 138; controls = 463) were left for further analyses.

4.3.4 Data Metrics Post QC

4.3.4.1 QQ Plot

The QQ plot was generated for the results from a simple χ^2 allelic test of association on common SNPs only of 138 sIBM cases and 463 controls. The QQ plot shows that the variants from sIBM cases and controls were generally appropriately matched

(Figure 4-10). The *p*-values adhere relatively closely to the y = x line for the bottom 90% of the values. The points deviating from the y = x line on the right corner may indicate real associations. The genomic inflation factor λ is 1.060995, indicating that cases and controls are well matched for ethnicity.

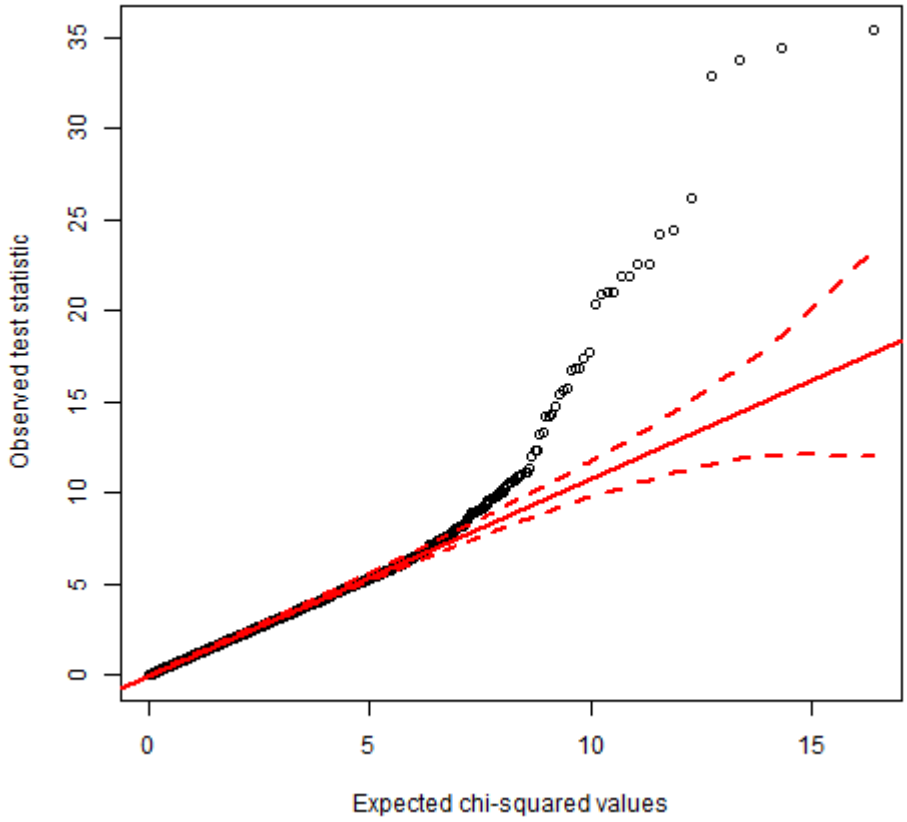


Figure 4-10 Quantile-quantile plot

This shows a quantile-quantile (QQ) plot of the results from a simple χ^2 allelic test of association on common SNPs only of 138 cases and 463 controls which passed the QC. The points deviating from the y = x line on the right corner may indicate real associations. The genomic inflation factor λ is 1.060995.

4.3.4.2 Annotation by VEP

A total number of 105883 variants of all the sIBM and control exomes which passed the QC were annotated by VEP. General statistics of the annotation was shown in Table 4-3. Among them, 27341 variants (25.8%) were novel and the rest 78542 (74.2%) were previously known variants.

Table 4-3 General statistics of the sIBM and control whole-exome dataset by VEP annotation

General Statistics	Count
Lines of input read	105911
Variants processed	105883
Variants remaining after filtering	105883
Lines of output written	105883
Novel/ known variants	27341 (25.8%)/ 78542 (74.2%)
Overlapped genes	22431
Overlapped transcripts	124675
Overlapped regulatory features	8976

VEP = Variant Effect Predictor.

The variants were also annotated for their consequence types by VEP. About half of the total variants were missense variants. A pie chart shows the proportion of each variant consequence type as Figure 4-11.

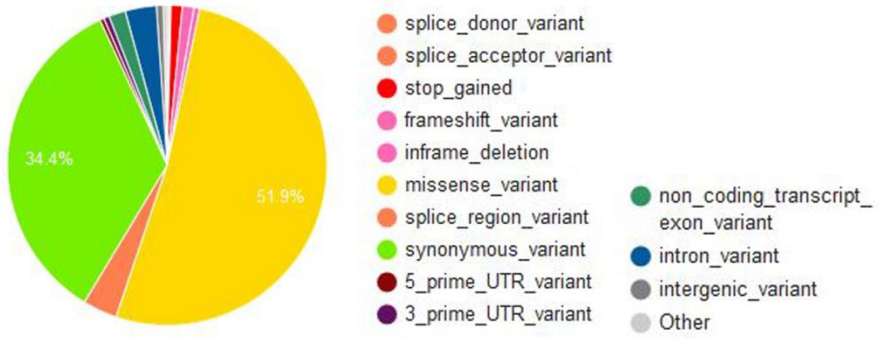


Figure 4-11 Proportion of variant consequences annotated by VEP

VEP also provided the pathogenicity predication of each variant by both SIFT and PolyPhen. More than a third of the variants were predicted as deleterious/probably damaging/possible damaging by the two *in silico* tools (Figure 4-12).

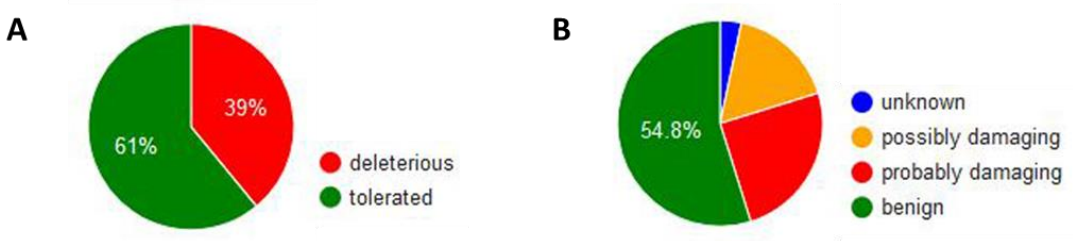


Figure 4-12 Pathogenicity predication of variants by SIFT and PolyPhen

Pie chart (A) shows the proportion of prediction of variants by SIFT; Pie chart (B) shows the proportion of prediction of variants by PolyPhen.

In addition, VEP summarised the number of variants by chromosome and the distribution of variants on each chromosome. The figures are available in Appendix III - 8.

4.3.4.3 Study Power

When the prevalence of sIBM is assumed to be 7.1×10^{-5}, to reach a genome-wide significance level of 5×10^{-7}, the study can achieve only 23.0% power at the current sample size (sIBM = 138; control = 463) with a relative risk (RR) = 2.0 for MAF = 0.2. However, the study can achieve 97.9% power at the current sample size with a RR = 3.0 for MAF = 0.2.

4.3.5 Single Variant Association Analyses

A simple χ^2 allelic test of sIBM case-control whole-exome association showed that strongest association signals were located on chromosomes 6 and 15. These were shown in a Manhattan Plot (Figure 4-13). Table 4-4 shows details of 16 variants in 10 genes with significant *p*-values in an allelic test of association according to at least one adjustment method for multiple testing. With the exception of *SLC24A1*, located on chromosome 15, eight other genes are mapped to chromosome 6p21.3 and one in 6p22.1 which are within or adjacent to MHC class I-III regions. The majority of the variants are common SNPs, and only the variant in *SLC24A1* has a MAF < 5% in the ExAC database.

A logistic regression analysis of sIBM case-control whole-exome association was also performed and adjusted by gender as a covariate. The top 15 SNPs with lowest *p*-values were listed in Table 4-5. All these SNPs are located on chromosome 6 and overlapped with the findings from the above allelic association test. The rare variant in *SLC24A* observed in the allelic association test was absent from the logistic regression analysis. This is probably because the logistic regression model is not well-defined if the SNP is monomorphic in cases or controls. Two SNPs (rs3132554 and rs3130983) in *PSORS1C1* with an odds ratio (OR) at 0.5114 (95% CI: 0.3847 – 0.6799) suggest that they decrease the risk of the disease; while the rest of SNPs with ORs > 1.0 suggest their effects of increasing the risk of the disease.

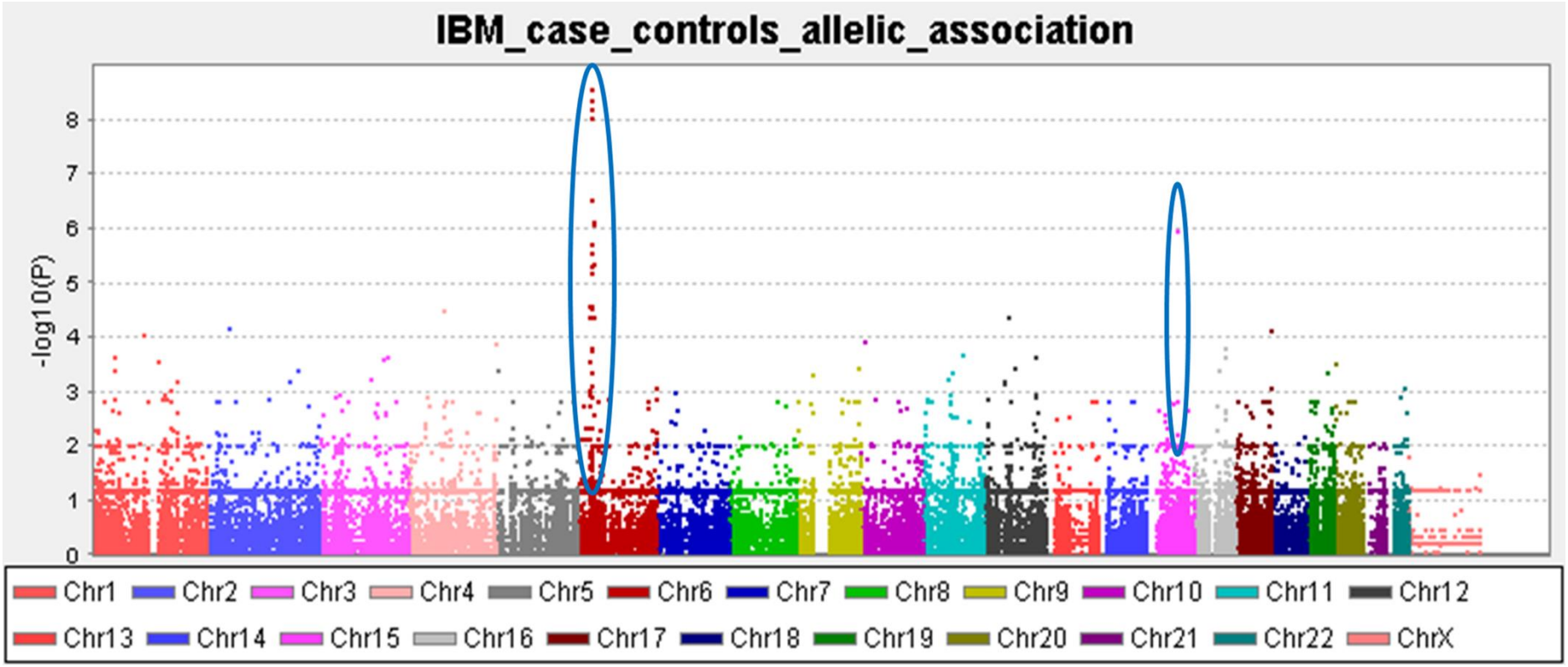


Figure 4-13 Manhattan plot of association *p*-values from a simple χ^2 allelic test on a sIBM and control whole-exome dataset

The plot shows –log10 *p*-value for each variant against chromosome location. Values for each chromosome are shown in different colours for visual effect. Two regions are highlighted where markers have strongest association signals.

Table 4-4 SNPs from an allelic test of sIBM case-control whole-exome association study showing the strongest association signals

Chr	Position _A1_A2	SNP	Gene ID	MAF in sIBM	MAF in Controls	MAF in ExAC	CHISQ	Unadjusted *P*-value	Bonferroni	Holm	Sidak single step	FDR BH	*P*-value after 1000000 Permutation
6p21.33	31602967_A_G	rs1046089	*PRRC2A*	0.543	0.344	0.357	35.38	$2.71x10^{-9}$	$2.29x10^{-4}$	$2.29x10^{-4}$	$2.29x10^{-4}$	$1.78x10^{-4}$	$6.47x10^{-4}$
6p21.33	30919391_A_G	rs3094086	*DPCR1*	0.293	0.140	0.068	34.43	$4.42x10^{-9}$	$3.73x10^{-4}$	$3.73x10^{-4}$	$3.73x10^{-4}$	$1.78x10^{-4}$	$1.57x10^{-4}$
6p22.1	30078330_T_C	rs2239529	*TRIM31*	0.275	0.129	0.111	33.74	$6.31x10^{-9}$	$5.33x10^{-4}$	$5.33x10^{-4}$	$5.33x10^{-4}$	$1.78x10^{-4}$	$4.98x10^{-4}$
6p22.1	30078275_T_C	rs2523989	*TRIM31*	0.275	0.130	0.114	32.92	$9.61x10^{-9}$	$8.12x10^{-4}$	$8.12x10^{-4}$	$8.12x10^{-4}$	$2.03x10^{-4}$	$7.33x10^{-4}$
6p21.33	31929014_A_C	rs437179	*SKIV2L*	0.467	0.302	0.230	26.16	$3.14x10^{-7}$	$2.65x10^{-2}$	$2.65x10^{-2}$	$2.62x10^{-2}$	$5.31x10^{-3}$	$1.15x10^{-2}$
6p21.32	32362639_A_G	rs41535850	*BTNL2*	0.272	0.144	0.155	24.39	$7.88x10^{-7}$	$6.66x10^{-2}$	$6.66x10^{-2}$	$6.45x10^{-2}$	$1.04x10^{-2}$	$3.40\ x10^{-2}$
6p21.32	32362669_G_T	rs41449245	*BTNL2*	0.272	0.144	0.155	24.22	$8.58x10^{-7}$	$7.25x10^{-2}$	$7.25x10^{-2}$	$6.99x10^{-2}$	$1.04x10^{-2}$	$3.87x10^{-2}$
15q22.31	65918176_C_CCTG	rs370680044	*SLC24A1*	0.026	0	0.013	23.74	$1.10x10^{-6}$	$9.30x10^{-2}$	$9.30x10^{-2}$	$8.88x10^{-2}$	$1.16x10^{-2}$	$5.81x10^{-2}$
6p21.33	31084163_G_A	rs3132554	*PSORS1C1*	0.341	0.503	0.410	22.61	$1.99x10^{-6}$	$1.68x10^{-1}$	$1.68x10^{-1}$	$1.55x10^{-1}$	$1.68x10^{-2}$	$1.40x10^{-1}$
6p21.33	31084792_T_C	rs3130983	*PSORS1C1*	0.341	0.503	0.424	22.61	$1.99x10^{-6}$	$1.68x10^{-1}$	$1.68x10^{-1}$	$1.55x10^{-1}$	$1.68x10^{-2}$	$1.40x10^{-1}$
6p21.33	31237323_A_G	rs68094471	*HLA-C*	0.463	0.310	0.265	21.87	$2.91x10^{-6}$	$2.46x10^{-1}$	$2.46x10^{-1}$	$2.18x10^{-1}$	$2.05x10^{-2}$	$1.34x10^{-1}$

6p21.33	31237333_T_A	rs66772001	*HLA-C*	0.463	0.310	0.265	21.87	$2.91x10^{-6}$	$2.46x10^{-1}$	$2.46x10^{-1}$	$2.18x10^{-1}$	$2.05x10^{-2}$	$1.34x10^{-1}$
6p21.32	32362702_C_T	rs41342846	*BTNL2*	0.263	0.144	0.155	21.04	$4.49x10^{-6}$	$3.80x10^{-1}$	$3.80x10^{-1}$	$3.16x10^{-1}$	$2.71x10^{-2}$	$1.60x10^{-1}$
6p21.32	32362703_T_G	rs41521946	*BTNL2*	0.263	0.144	0.155	21.04	$4.49x10^{-6}$	$3.80x10^{-1}$	$3.80x10^{-1}$	$3.16x10^{-1}$	$2.71x10^{-2}$	$1.60x10^{-1}$
6p21.32	32097290_T_C	rs28732176	*FKBPL*	0.018	0.107	0.098	20.87	$4.91x10^{-6}$	$4.15x10^{-1}$	$4.15x10^{-1}$	$3.40x10^{-1}$	$2.77x10^{-2}$	$3.39x10^{-1}$
6p21.33	31839309_C_T	rs2242665	*SLC44A4*	0.602	0.447	0.401	20.37	$6.38x10^{-6}$	$5.39x10^{-1}$	$5.39x10^{-1}$	$4.17x10^{-1}$	$3.37x10^{-2}$	$2.95x10^{-1}$

Chr = chromosome; A1 = minor allele code; A2 = major allele code; MAF = minor allele frequency; ExAC = Exome Aggregation Consortium; CHISQ = Chi-squared statistic; FDR = false discovery rate; BH = Benjamini and Hochberg; N/A = not available.

Table 4-5 Top 15 SNPs from a logistic regression analysis of sIBM case-control association study adjusted by gender

Chr	Position _A1_A2	SNP	Gene ID	OR	SE	L95	U95	STAT	P
6p21.33	30919391_A_G	rs3094086	*DPCR1*	2.931	0.1826	2.049	4.193	5.889	3.89×10^{-9}
6p21.33	31602967_A_G	rs1046089	*PRRC2A*	2.276	0.1442	1.715	3.019	5.703	1.18×10^{-8}
6p22.1	30078330_T_C	rs2239529	*TRIM31*	2.768	0.1797	1.946	3.937	5.665	1.47×10^{-8}
6p22.1	30078275_T_C	rs2523989	*TRIM31*	2.738	0.1796	1.926	3.893	5.609	2.04×10^{-8}
6p21.33	31929014_A_C	rs437179	*SKIV2L*	2.172	0.1514	1.614	2.923	5.122	3.02×10^{-7}
6p21.32	32362639_A_G	rs41535850	*BTNL2*	2.344	0.1748	1.664	3.302	4.873	1.10×10^{-6}
6p21.32	32362669_G_T	rs41449245	*BTNL2*	2.338	0.1748	1.66	3.294	4.858	1.18×10^{-6}
6p21.33	31237323_A_G	rs68094471	*HLA-C*	1.97	0.1468	1.477	2.626	4.617	3.90×10^{-6}
6p21.33	31237333_T_A	rs66772001	*HLA-C*	1.97	0.1468	1.477	2.626	4.617	3.90×10^{-6}
6p21.33	31084163_G_A	rs3132554	*PSORS1C1*	0.5114	0.1452	0.3847	0.6799	-4.616	3.90×10^{-6}
6p21.33	31084792_T_C	rs3130983	*PSORS1C1*	0.5114	0.1452	0.3847	0.6799	-4.616	3.90×10^{-6}
6p21.32	32362702_C_T	rs41342846	*BTNL2*	2.238	0.1771	1.582	3.167	4.548	5.42×10^{-6}
6p21.32	32362703_T_G	rs41521946	*BTNL2*	2.238	0.1771	1.582	3.167	4.548	5.42×10^{-6}
6p21.32	32097290_T_C	rs28732176	*FKBPL*	1.89	0.1437	1.426	2.504	4.427	9.54×10^{-6}
6p21.33	31839309_C_T	rs2242665	*SLC44A4*	1.792	0.1419	1.357	2.366	4.11	3.96×10^{-5}

Chr = chromosome; A1 = minor allele code; A2 = major allele code; OR = odds ratio; SE = standard error; L95 = lower bound on 95% confidence interval for odds ratio; U95 = Upper bound on 95% confidence interval for odds ratio; STAT = coefficient t-statistic; P = asymptotic *p*-value for t-statistic.

4.3.6 Gene-Based Association Analyses

Variants were selected by VEP into groups as mentioned in Section 4.2.4.4. No variant was left when filtering for the not benign model for rare variants among candidate genes. Therefore, SKAT-O association tests adjusted by gender were performed in five models for all genes and four models in candidate genes. The Bonferroni correction threshold α* for a genome-wide 5% significance level for each test was suggested in Table 4-6.

Table 4-6 Summary of each model for SKAT-O association tests

Analysis Model	Number of Gene Sets (N)	Number of SNPs	Bonferroni correction $\alpha^* = \alpha/(N + 9)$
All genes – All Variants Model	15321	105883	3.2×10^{-6}
All genes – Coding Model	12923	56918	3.8×10^{-6}
All genes – Coding Model with MAF < 1%	12446	47495	4.0×10^{-6}
All genes – Not Benign Model with MAF < 1%	7108	13491	7.0×10^{-6}
All genes – LOF Model with MAF < 1%	2048	2502	2.43×10^{-5}
Candidate genes – All Variants Model	129	1638	3.62×10^{-4}
Candidate genes – Coding Model	116	920	4.0×10^{-4}
Candidate genes – Coding Model with MAF < 1%	115	784	4.0×10^{-4}
Candidate genes – LOF Model with MAF < 1%	21	26	1.67×10^{-3}

MAF = minor allele frequency; LOF = loss-of-function; α = 0.05.

No genes under any model for all genes passed the Bonferroni correction threshold. Top 20 genes with lowest *p*-values in each model were listed in Table 4-7. *SLC24A1* was the only gene overlapping in the top genes identified in the single variant allelic association analysis. It was also the top gene in "All Variants Model", "Coding Model", and "Coding Model with MAF < 1%". In addition, there were 15 genes which were among the top 20 genes in at least two analysis models. However, the top 20 genes in "LOF with MAF < 1%" model were not overlapped with any other model.

Table 4-7 Top 20 genes in each model of SKAT-O test on all genes

Model	Gene ID	Chromosome	CCDS ID	Markers Tested	*P*-Value
All Variants Model α* = 3.2x10^{-6}					
	SLC24A1	15	CCDS45284.1	11	5.39x10^{-5}
	BEND4	4	CCDS47048.1	3	1.66 x10^{-4}
	TRIM40	6	CCDS69069.1	2	1.77x10^{-4}
	CDKL2	4	CCDS3570.1	3	2.22x10^{-4}
	SLC35B2	6	CCDS34462.1	5	2.61x10^{-4}
	FNDC9	5	CCDS4337.1	7	3.64x10^{-4}
	SEC22B	1	N/A	14	3.64x10^{-4}
	SENP3-EIF4A1	17	N/A	6	4.53x10^{-4}
	MSH2	2	CCDS1834.1	11	4.73x10^{-4}
	PRKRA	2	CCDS2279.1	1	5.39x10^{-4}
	TAF1L	9	CCDS35003.1	25	6.03x10^{-4}
	OBP2A	9	CCDS6992.1	1	6.05x10^{-4}
	ADAMTS20	12	CCDS31778.2	14	6.98x10^{-4}
	ITCH	20	CCDS13234.1	2	7.25x10^{-4}
	ZNF80	3	CCDS2979.1	7	9.14x10^{-4}
	CCDC158	4	CCDS43242.1	13	9.60x10^{-4}
	ENDOV	17	CCDS54172.1	4	9.76x10^{-4}
	MRPL32	7	CCDS5468.1	1	1.00x10^{-3}
	LIG3	17	CCDS11284.2	3	1.03x10^{-3}
	ATP2A2	12	CCDS9143.1	11	1.03x10^{-3}
Coding Model α* = 3.8 x10^{-6}					
	SLC24A1	15	CCDS45284.1	9	9.54x10^{-6}
	INPP5F	10	CCDS7616.1	8	9.96x10^{-5}
	C5orf42	5	CCDS34146.2	8	2.21x10^{-4}
	TRIM40	6	CCDS69069.1	2	2.66x10^{-4}
	KCNMB1	5	CCDS4373.1	4	3.94x10^{-4}
	MSH2	2	CCDS1834.1	7	4.24x10^{-4}
	PRKRA	2	CCDS2279.1	1	4.74x10^{-4}
	SLC35B2	6	CCDS34462.1	4	5.45x10^{-4}
	OBP2A	9	CCDS6992.1	1	6.06x10^{-4}
	ADAMTS20	12	CCDS31778.2	11	6.16x10^{-4}
	ZNF573	19	CCDS12508.1	4	6.73x10^{-4}
	SNX15	11	CCDS8089.1	1	7.39x10^{-4}
	BEND4	4	CCDS47048.1	2	7.68x10^{-4}
	GBP3	1	CCDS717.2	4	7.75x10^{-4}

	TUB	11	CCDS7786.1	5	8.18×10^{-4}
	KIR3DX1	19	N/A	3	9.08×10^{-4}
	CTU2	16	CCDS45545.1	4	9.62×10^{-4}
	LRRC66	4	CCDS43229.1	3	1.05×10^{-3}
	MED12L	3	CCDS33876.1	3	1.09×10^{-3}
	SLC25A45	11	CCDS41670.1	3	1.12×10^{-3}
Coding Model with MAF < 1% $\alpha^* = 4.0 \times 10^{-6}$					
	SLC24A1	15	CCDS45284.1	7	8.33×10^{-6}
	INPP5F	10	CCDS7616.1	8	1.28×10^{-4}
	SLC35B2	6	CCDS34462.1	4	1.65×10^{-4}
	C5orf42	5	CCDS34146.2	7	2.21×10^{-4}
	TUB	11	CCDS7786.1	4	3.04×10^{-4}
	TCTN1	12	CCDS41834.1	8	3.16×10^{-4}
	ADAM33	20	CCDS13058.1	2	3.26×10^{-4}
	OBP2A	9	CCDS6992.1	1	6.70×10^{-4}
	BEND4	4	CCDS47048.1	2	6.87×10^{-4}
	KCNMB1	5	CCDS4373.1	3	7.63×10^{-4}
	GBP3	1	CCDS717.2	3	7.75×10^{-4}
	LIG3	17	CCDS11284.2	3	7.77×10^{-4}
	LRRC66	4	CCDS43229.1	3	7.92×10^{-4}
	FAM221B	9	CCDS43799.2	3	8.37×10^{-4}
	NEDD4L	18	CCDS45872.1	3	8.52×10^{-4}
	WASF3	13	CCDS9318.1	3	8.79×10^{-4}
	SAC3D1	11	CCDS41668.1	2	8.81×10^{-4}
	NOB1	16	CCDS10884.1	5	9.35×10^{-4}
	MED12L	3	CCDS33876.1	3	9.77×10^{-4}
	CES2	16	CCDS10825.1	3	1.05×10^{-3}
Not Benign Model with MAF < 1% $\alpha^* = 7.0 \times 10^{-6}$					
	OBP2A	9	CCDS6992.1	1	6.05×10^{-4}
	RIMS2	8	CCDS55269.1	2	7.40×10^{-4}
	ESYT3	3	CCDS3101.2	3	8.41×10^{-4}
	MCM3AP	21	CCDS13734.1	3	8.78×10^{-4}
	TUBAL3	10	CCDS7066.2	3	9.37×10^{-4}
	ZSWIM5	1	CCDS41319.1	6	9.91×10^{-4}
	SEC24A	5	CCDS43363.1	3	1.08×10^{-3}
	EYA4	6	CCDS5165.1	3	1.08×10^{-3}
	ANO2	12	N/A	7	1.13×10^{-3}
	MAP3K4	6	CCDS34565.1	3	1.14×10^{-3}
	CCDC77	12	CCDS8503.1	3	1.15×10^{-3}
	RNF6	13	CCDS9316.1	3	1.15×10^{-4}

	PARP9	3	CCDS3014.1	3	$1.24x10^{-3}$
	SFT2D2	1	CCDS1271.1	1	$2.32x10^{-3}$
	INCA1	17	CCDS54074.1	1	$2.32x10^{-3}$
	PTPRB	12	CCDS44943.1	4	$2.38x10^{-3}$
	TBX5	12	CCDS9173.1	1	$2.56x10^{-3}$
	PTPRD	9	CCDS43786.1	2	$2.62x10^{-3}$
	SPATA20	17	CCDS11571.1	2	$2.78x10^{-3}$
	MAST2	1	CCDS41326.1	4	$3.04x10^{-3}$
LOF Model with MAF < 1% α* = $2.43x10^{-5}$					
	ADAM33	20	CCDS13058.1	1	$2.28x10^{-4}$
	ADAM20	14	CCDS32111.1	1	$2.25x10^{-3}$
	IPP	1	CCDS30702.1	1	$2.27x10^{-3}$
	OR6B1	7	CCDS43667.1	2	$2.76x10^{-3}$
	CPNE1	20	CCDS46595.1	2	$2.94x10^{-3}$
	IGSF22	11	CCDS41625.2	2	$3.42x10^{-3}$
	IRAK3	12	CCDS8975.1	2	$4.58x10^{-3}$
	ZNF662	3	CCDS46807.1	2	$4.96x10^{-3}$
	ZNF442	19	CCDS12271.1	2	$5.18x10^{-3}$
	ZNF615	19	CCDS59418.1	2	$5.18x10^{-3}$
	LILRA1	19	CCDS12901.1	2	$5.20x10^{-3}$
	CHRNG	2	CCDS33400.1	2	$5.67x10^{-3}$
	BAMBI	10	CCDS7162.1	2	$5.80x10^{-3}$
	GSTK1	7	CCDS47730.1	1	$8.38x10^{-3}$
	GPR151	5	CCDS34266.1	1	$8.66x10^{-3}$
	TAS2R3	7	CCDS5867.1	1	$9.11x10^{-3}$
	ZMYM5	13	CCDS31942.1	1	$1.45x10^{-2}$
	DLEC1	3	CCDS2672.2	1	$1.52x10^{-2}$
	EMR2	19	CCDS32935.1	1	$1.54x10^{-2}$
	SPATA31D1	9	CCDS47986.1	2	$1.62x10^{-2}$

CCDS = the Consensus CDS project; MAF = minor allele frequency; LOF = loss-of-function; N/A = not available.

A group of selected candidate genes may be enriched with lower *p*-values when comparing with all genes based association tests. However, none of the candidate genes in any test model reached statistical significance. The top 20 genes with lowest *p*-values in each model were listed in Table 4-8. Three genes including *SIGMAR1*, *SYNE1*, and *TRAPPC11*, were present in all four models. The *VCP*, *GRN*, *KIDINS220*, *MYH2*, *DYSF*, *LAMA2*, and *DAG1* genes were overlapped in three models. Other 11 genes were present in two models. *BTNL2* with a *p*-value of

$8.47x10^{-2}$ among the top 20 genes in the "All Variants Model" also had common variants that passed the significant threshold in the single variant allelic association test.

Table 4-8 Top 20 genes in each model of SKAT-O test on candidate genes

Model	Gene ID	Chromosome	CCDS ID	Marker Tested	*P*-Value
All Variants Model α* = $3.62x10^{-4}$					
	SIGMAR1	9	CCDS6562.1	3	$3.91x10^{-3}$
	GRN	17	CCDS11483.1	3	$1.75x10^{-2}$
	SYNE1	6	CCDS5236.2	110	$1.94x10^{-2}$
	SERPINA3	14	CCDS32150.1	12	$2.44x10^{-2}$
	TRAPPC11	4	CCDS34112.1	11	$2.48x10^{-2}$
	RNF123	3	CCDS33758.1	1	$3.43x10^{-2}$
	POLG	15	CCDS10350.1	19	$4.53x10^{-2}$
	VCP	9	CCDS6573.1	10	$4.67x10^{-2}$
	LDB3	10	CCDS53550.1	9	$5.81x10^{-2}$
	OXTR	3	CCDS2570.1	2	$7.89x10^{-2}$
	LAMA2	6	CCDS5138.1	35	$8.30x10^{-2}$
	BTNL2	6	N/A	7	$8.47x10^{-2}$
	MYHAS	17	N/A	5	$8.67x10^{-2}$
	DAG1	3	CCDS2799.1	7	$9.97x10^{-2}$
	CCRL2	3	CCDS46814.1	7	$1.01x10^{-1}$
	TMEM43	3	CCDS2618.1	8	$1.03x10^{-1}$
	SRCAP	16	CCDS10689.2	26	$1.06x10^{-1}$
	KIDINS220	2	CCDS42650.1	16	$1.11x10^{-1}$
	LMNA	1	CCDS1129.1	3	$1.28x10^{-1}$
	NT5C1A	1	CCDS440.1	2	$1.35x10^{-1}$
Coding Model α* = $4.0x10^{-4}$					
	VCP	9	CCDS6573.1	4	$1.32\ x10^{-2}$
	SIGMAR1	9	CCDS6562.1	2	$1.58\ x10^{-2}$
	SYNE1	6	CCDS5236.2	63	$2.321x10^{-2}$
	CCRL2	3	CCDS46814.1	1	$2.86x10^{-2}$
	MATR3	5	CCDS4210.1	1	$3.05x10^{-2}$
	NEFH	22	CCDS13858.1	2	$3.11x10^{-2}$
	TUBA4A	2	CCDS2438.1	1	$3.12x10^{-2}$
	GMPPB	3	CCDS2802.1	1	$3.43x10^{-2}$
	GRN	17	CCDS11483.1	1	$3.43x10^{-2}$
	MYH2	17	CCDS11156.1	10	$3.75x10^{-2}$

	POMT1	9	CCDS6943.1	11	1.23×10^{-1}
	KIDINS220	2	CCDS42650.1	9	1.26×10^{-1}
	FLNC	7	CCDS43644.1	11	1.35×10^{-1}
	DYSF	2	CCDS46324.1	23	1.47×10^{-1}
	HLA-DRA	6	CCDS4750.1	1	1.62×10^{-1}
	TRAPPC11	4	CCDS34112.1	8	1.92×10^{-1}
	BSCL2	11	CCDS44627.1	2	2.31×10^{-1}
	DAG1	3	CCDS2799.1	5	2.33×10^{-1}
	FIG4	6	CCDS5078.1	2	2.36×10^{-1}
	PSEN1	14	CCDS9812.1	2	2.46×10^{-1}
Coding Model with MAF < 1% $\alpha^* = 4.0 \times 10^{-4}$					
	TRAPPC11	4	CCDS34112.1	5	1.84×10^{-3}
	VCP	9	CCDS6573.1	4	1.31×10^{-2}
	MATR3	5	CCDS4210.1	1	3.05×10^{-2}
	NEFH	22	CCDS13858.1	1	3.05×10^{-2}
	TUBA4A	2	CCDS2438.1	1	3.12×10^{-2}
	SIGMAR1	9	CCDS6562.1	1	3.13×10^{-2}
	GMPPB	3	CCDS2802.1	1	3.43×10^{-2}
	GRN	17	CCDS11483.1	1	3.43×10^{-2}
	MYH2	17	CCDS11156.1	10	3.80×10^{-2}
	DYSF	2	CCDS46324.1	21	5.20×10^{-2}
	LAMA2	6	CCDS5138.1	16	5.91×10^{-2}
	KIDINS220	2	CCDS42650.1	8	1.28×10^{-1}
	FLNC	7	CCDS43644.1	11	1.35×10^{-1}
	DCTN1	2	CCDS1939.1	10	1.58×10^{-1}
	SYNE1	6	CCDS5236.2	52	1.60×10^{-1}
	DAG1	3	CCDS2799.1	5	2.27×10^{-1}
	BSCL2	11	CCDS44627.1	2	2.31×10^{-1}
	PSEN1	14	CCDS9812.1	2	2.46×10^{-1}
	DES	2	CCDS33383.1	2	2.48×10^{-1}
	UNC45B	17	CCDS11292.1	2	2.50×10^{-1}
LOF Model with MAF < 1% $\alpha^* = 1.67 \times 10^{-3}$					
	CCR2	3	CCDS46814.1	1	3.12×10^{-2}
	SPG11	15	CCDS10112.1	1	3.12×10^{-2}
	SIGMAR1	9	CCDS6562.1	1	3.13×10^{-2}
	TRAPPC11	4	CCDS34112.1	1	3.39×10^{-2}
	GLE1	9	CCDS35154.1	1	3.41×10^{-2}
	CAPN3	15	CCDS45245.1	1	3.45×10^{-2}
	DYSF	2	CCDS46324.1	3	4.76×10^{-2}
	ALS2	2	CCDS42800.1	2	6.08×10^{-2}

	GNG3	11	CCDS8032.1	1	$7.04x10^{-2}$
	LAMA2	6	CCDS5138.1	1	$7.05x10^{-2}$
	ATN1	12	CCDS31734.1	1	$7.08x10^{-1}$
	SYNE2	14	CCDS9761.2	1	$7.08x10^{-1}$
	SS18L1	20	CCDS13491.1	1	$7.08x10^{-1}$
	MAVS	20	CCDS33437.1	1	$7.09x10^{-1}$
	FYCO1	3	CCDS2734.1	1	$7.09x10^{-1}$
	SYNE1	6	CCDS5236.2	0	1
	PON3	7	CCDS5639.1	1	1
	POMT1	9	CCDS6943.1	0	1
	SERPINA3	14	CCDS32150.1	0	1
	MYH2	17	CCDS11156.1	0	1

CCDS = the Consensus CDS project; MAF = minor allele frequency; LOF = loss-of-function; N/A = not available.

4.4 Discussion

This is the first WES case-control association study on a large cohort of 181 sIBM patients and 510 neuropathologically healthy control individuals. Single variant association analyses found significant associations with 15 common SNPs in genes (*PRRC2A*, *DPPCR1*, *TRIM31*, *SIKV2L*, *BTNL2*, *PSORS1C1*, *HLA-C*, *FKBPL* and *SLC44A4*) located close or within MHC class I-III regions on chromosome 6 and one rare variant in *SLC24A1* on chromosome 15 after correcting for multiple testing. Findings of logistic regression analysis adjusted by gender suggest that the SNP rs3094086 in *DPCR1* has the strongest association to increased risk of disease (OR = 2.9, 95% CI: 2.049 – 4.193), while two SNPs (rs3132554 and rs3130983) in *PSORS1C1* show a protective effect (both with an OR = 0.5 and 95% CI: 0.3847 – 0.6799). Interestingly, a strong association between *HLA-DR3* and components of the 8.1 MHC AH with sIBM has been reported in previous studies, and a susceptibility locus was mapped between *HLA-DR* and *C4* on chromosome 6 (Kok et al., 1999, Lampe et al., 2003, Needham et al., 2008c, Scott et al., 2011, Rojana-udomsart et al., 2013, Price et al., 2004). Our findings from the single variant association analyses suggest an overlapped region on chromosome 6 with previous studies. This provides more evidence for the role of MHC in sIBM. *HLA* region is in high degree of linkage disequilibrium (LD) that exists between alleles at neighboring loci, and also contains the extreme degree of polymorphism. Therefore, it is very difficult to decipher the

causal variants from those in LD. Imputation of the variants in *HLA* regions will be important to determine in which *HLA* alleles these common SNPs identified in our study lie. In addition, whole-exome sequencing targets all the protein-coding regions and this will miss the intronic variants that could be markers for specific *HLA* haplotypes. Immunochip in sIBM would be a more specific technique for studying *HLA* alleles. Of note, the 8.1 MHC AH has also shown the strongest association with other idiopathic inflammatory myopathies (IIMs), such as dermatomyositis (DM), juvenile dermatomyositis (JDM), and polymositis (PM). Associations with alleles outside the 8.1 AH reveal differences between DM, JDM, and PM (Rothwell et al., 2016). The exact immue mediated mechanism leading to sIBM or other IIMs is still unclear. This may be due to the impact of different *HLA* loci, or interaction between different other genetic or environmental factors and inflammatory factors.

To boost the power of studying the rare variant association, aggregated multiple variants in a gene were analysed as a unit. By using VEP, variants were also weighted based on their functional annotation, and their frequency such as non-synonymous variants, variants with MAF < 1%, variants predicted as not benign by SIFT and PolyPhen, or LOF variants. To further increase the analysis power, the number of genes was reduced by selecting a group of 127 candidate genes, each belonging to at least one of six groups (Table 4-1). These included i) genes associated with neurodegenerative diseases (such as ALS, FTD, AD, and Prion diseases), ii) genes associated muscle diseases (such hIBM, myopathies with rimmed vacuoles, and other IBM-like myopathies), iii) immune genes previously studied in sIBM, iv) genes encoding for mitochondrial proteins, v) genes encoding for RNA binding proteins, and vi) genes known from private communications or conferences. Overall, five test models across all genes and four test models across the candidate genes were analysed as shown in Table 4-6. However, no significant association was detected either across all genes or only across the candidate genes for any test model. *SLC24A1* was the top gene with lowest *p*-value in three models across all genes analyses, and a rare variant in *SLC24A1* was also found to be associated with sIBM in single variant analysis. *SLC24A1* encoded sodium-calcium exchanger expressed in the retina (Riazuddin et al., 2010). Its mutations have been reported as a cause of autosomal recessive congenital stationary night blindness (OMIM#613830), but has not yet been implicated in neuromuscular diseases. The rare variant found in our

sIBM cohort is an inframe deletion (c.1759_1761delCTG, p.Leu591del) in exon 2 of the *SLC24A1* gene, locating in one of the multitransmembrane helix. The function of this region of the SLC24A1 protein is unclear. Therefore, it is hard to suggest the role of *SLC24A1* in sIBM, although its *p*-values are very close to the significance threshold. Validation of this region on chromosome 15 is required in future replication studies.

The failure to reach significance at the gene level is probably due to our study being underpowered to detect genetic effects. Several aspects could be involved. First, the sample size of our study was not sufficiently large to achieve the statistical power required. The sample size was usually in the thousands in studies that successfully identified significant associations in other diseases. However, reaching such a number would be unfeasible for rare diseases like sIBM which has a prevalence of 1-71 per million of the population, even with a great effort from international collaborations. Secondly, the patients in our cohort were diagnosed as pathologically defined sIBM, clinically defined sIBM, probable sIBM, and possible sIBM according to three diagnosis criteria we used in this study. This could potentially increase the heterogeneity of clinical phenotypes in the patient cohort, which may also decrease the analysis power. However, the sample size would be dramatically reduced if analyses were only based on each diagnostic category. The balance between the purity of the phenotype and the sample size is a significant challenge that researchers commonly have to face, especially when the total number of cases is limited. In addition, genetics of sIBM is likely to be polygenic. This means that the more genes are involved, the weaker the effect each gene accounts, and the less the power of the analysis is to detect the association with the current sample size. Furthermore, gene-based tests sometimes can lead to loss of power when only one or a very few of the variants in a gene are associated with the phenotype, and many variants have no effect, and causal variants are also low frequency variants. Single variant tests showed clear evidence of association but gene-based tests showed weaker signals in our study. This is likely because these genes contain a very small number of common variants that are associated with sIBM, while rare coding variants contribute less to the disease. Furthermore, conflicting prediction results on a same variant can sometimes be generated by *in silico* tools, SIFT and PolyPhen. This will cause a

potential mix of benign variants and true pathogenic variants in the analysis model, which may decrease the power to detect the pathogenic gene.

QC is one of the most difficult challenges and also the most time consuming process in dealing with whole-exome data, and it plays a very important role as a gatekeeper to filter out the low-quality samples and variants to avoid false positive associations. A good QC strategy determines the reliability of the results of association analyses. However, in contrast to the QC for GWAS, there is no standard QC filtering threshold for whole-exome association studies. It largely depends on the quality of the raw data from sequencing. Researchers need to adjust their thresholds according to the outcomes of QC. For example, in a recent whole-exome association analysis on ALS, the cut-offs of VQSR tranche varied from 97% to 99.9% for whole-exomes from different centres (Cirulli et al., 2015). In our study, as the whole-exome data of cases and control were generated by two institutions using different sequence protocols, this increased the heterogeneity of the whole-exome dataset. Therefore, after testing three different QC filtering strategies, we finally applied a stringent QC strategy - a combination of genotype QC, variant QC and sample QC. Variants were labelled as missing ("NoCall") when they failed the genotype QC (read depth < 8, GQ < 20, GQ_mean < 35, or outside the range of a VQSR tranche at 99.0%). Subsequently, any variant with a call rate < 98.5% was removed. This left almost a third of the total original variants for the final association analyses. The QQ plot showed that there was no excess of genetic inflation after QC, suggesting that our stringent QC strategy successfully removed the access of systematic bias. However, a stringent QC strategy could also potentially remove true associations, which causes false negatives. This means if the true association exists among the other two thirds of the variants that have been removed during the QC, the association would have never been detected. Therefore, generating a cohort of good quality whole-exome data is crucial at the first stage for any kind of whole-exome analysis. An appropriate QC strategy should be chosen based on the quality of the data, in order to reduce the chance of both false positives and false negatives.

The design of whole-exome association analysis for sIBM aimed to investigate the associations of rare variants in coding regions with the risk of the disease. However, if rare variants in coding regions contribute little to the genetics of sIBM, whole-

exome studies would be not able to identify the associations. This may be another explanation of the failure of our gene-based analyses. In this case, a whole-genome sequencing study (WGS) should be designed, when rare non-coding regulatory variants account for much of the missing heritability.

Despite of the challenges and limitations of using WES for association analysis, it is still a faster and more cost-efficient method to explore variations in coding regions than Sanger sequencing and WGS, especially when the genetic architecture of the disease is unclear. Our study identified that several common SNPs in MHC regions passed the genome-wide significance threshold by single variant association analyses, which is consistent with some previous studies on the association between MHC genes and the risk of sIBM. This suggests that our study design was reliable, although it failed to detect the association at gene level with current sample size.

In conclusion, the first whole-exome association analyses were perform on 181 sIBM cases and 510 controls by our group. A stringent QC strategy using a series of bioinformatics techniques was applied to generate a cohort of high quality whole-exome dataset for association analyses. Our study suggested a strong association between common SNPs in the MHC regions on chromosome 6 and a rare variant on chromosome 15 and the risk of sIBM, while analyses at the gene levels did not identify any significant association. In the future, more sIBM samples should be sequenced, and sIBM exomes could be grouped based on the diagnostic category for definite sIBM (both pathologically and clinically), probably sIBM, and possible sIBM. Subsequently, the association could be analysed between each sIBM subgroup and controls. When necessary, WGS could be considered to investigate the genetic contribution of non-coding regions to sIBM.

Chapter 5 Whole-Exome Sequencing of a Cohort of Patients with Myopathies with Tubular Aggregates and Cylindrical Spirals

5.1 Introduction

Tubular aggregates (TAs) and cylindrical spirals (CSs) are rare membranous structures found in skeletal muscle fibres. They are both thought to be derived from the abnormal sarcotubular system and have similar histopathological appearances that are bright red with modified Gomori trichrome (Goebel et al., 2013). However, they have distinct ultrastructural appearances which can distinguish one from the other. TAs consist of densely packed groups of tubules mostly 50-70 nm in diameter, while CSs are composed of up to 20 concentric lamellae measuring 1-5 μm in diameter and up to 300 μm in longitudinally oriented fibres clustered in subsarcolemmal areas (Pavlovicova et al., 2003, Goebel et al., 2011, Goebel et al., 2013). Both TAs and CSs are unspecific pathological features. Myopathies with TAs and CSs have highly heterogeneous clinical phenotypes (Goebel et al., 2013). TAs have been observed in four major disease categories including dominant tubular aggregate myopathy (TAM), phosphoglycerate mutase deficiency, congenital myasthenic syndromes, and periodic paralysis. Their details were discussed in Section 1.3.1. CSs were only reported in 18 sporadic and familial cases so far with a variety of clinical phenotypes, summarised in Table 1-10. Eight genes have been reported to be associated with cases with TAs listed in Table 1-9. However, many cases with TAs still lack of genetic clarification, and no gene has been reported to be responsible for myopathies with CSs.

In this chapter, a cohort of patients was selected based on the pathological findings of TAs and/or CSs in muscle biopsies, and then their clinical phenotypes were reviewed. We hypothesized that genetic components could contribute to a portion of the

diseases with TAs and CSs in muscle. Whole-exome sequencing was applied to investigate the genetic basis of TAs and CSs. The analysis aims to identify mutations in known genes and candidate genes, as well as to identify novel genes associated with TAs and CSs in order to improve our knowledge of the pathogenesis of these rare pathological structures.

5.2 Methods

Samples from a total of 35 genetically undiagnosed patients with TAs (four cases were from two families) and two patients with CSs were collected for genetic investigation. Twenty-four cases were identified by searching ION pathology database for records of TAs or CSs in their biopsy reports. Most of the patients were previously seen or are currently being seen by neurologists at NHNN. Four cases from two families with TAs were kindly provided by neurologists in Northern Ireland, four were from the Neuromuscular Unit, BioBank of Skeletal Muscle, Nerve Tissue, DNA and Cell Lines in Italy, and five were from Neuromuscular Diseases and Neuroimmunology Unit Muscle Cell Biology Lab in Italy. Among them, 13 cases had a positive family history, 14 were sporadic, and 10 had unknown family history. Clinical phenotypes and available muscle biopsies were reviewed. The biopsy slides were stained by staff at the Department of Pathology, ION. Biopsies were reviewed and the images were kindly taken by Professor Janice Holton from the Department of Pathology, ION.

WES was performed on 33 unrelated index patients with TAs, and two unrelated cases with CSs. The sequencing was carried out in-house on the Illumina HiSeq2500 platform by Ms Deborah Hughes as described in Sections 2.3.1-2.3.2. With the exception of Case ID 29970, the data for this case were kindly provided by a former colleague at the MRC Centre for Neuromuscular Diseases. The preliminary data were processed using the ION bioinformatics pipeline by Dr Alan Pittman. In this study, the BAM files of all the samples were merged into a master BAM file. Joint variant calling was performed on the master BAM file by GATK. This could potentially increase the accuracy of overall variant calling, and also be able to reveal the genetic overlaps between cases. Then it was annotated using ANNOVAR as described in Section 2.3.3. Lastly, the master file was annotated with a list of the

reported neuromuscular diseases genes, which were downloaded from the online GeneTable of Neuromuscular Disorders (http://www.musclegenetable.fr/) updated by August of 2015. New genes that were reported after this time point were added to the list manually.

Initially, the merged master file of all the exomes was analysed to look for possible mutations in known genes or candidate genes shared between different cases. A candidate gene searching strategy was used for the non-synonymous rare variants (MAF < 1%), filtering as described in Section 2.3.4. The genes include eight known TAs genes, genes which encode proteins related to sarcoplasmic reticulum, and the genes involved in N-linked protein glycosylation associated with congenital myasthenic syndromes (Table 5-1). If known genes or candidate genes were not found in the master exome file, the same candidate gene filtering approach was performed in each individual sample to confirm whether any variant of interest was missing from the master file. If no positive result was identified, samples from other family members were needed to look for other neuromuscular genes or new genes. The workflow of the filtering strategy was shown in Figure 5-1. Sanger sequencing was performed to confirm the variants identified in WES. Only the confirmed variants are presented and discussed in the following sections.

Table 5-1 Major candidate genes for myopathies with TAs and CSs

Candidate genes	Gene ID (OMIM#)
Known TAs genes	*STIM1*, *ORAI1*, *PGAM2*, *GFPT1*, *DPAGT1*, *ALG2*, *SCN4A*, *SERAC1* (Table 1-9)
Genes for SR related protein in muscle	*CASQ1* (114250, AD), *RYR1* (180901, AD/AR), *TRDN* (603283, AR), *ATP2A1* (108730, AR), *DHPR* (N/A), SRL (604992)
Other genes involved in N-linked protein glycosylation associated with CMS	*ALG14* (612866, AR), *GMPPB* (615320, AR)

TAs = tubular aggregates; CSs = cylindrical spirals; SR = sarcoplasmic reticulum; CMS = congenital myasthenic syndrome; AD = autosomal dominant; AR = autosomal recessive; N/A = not available.

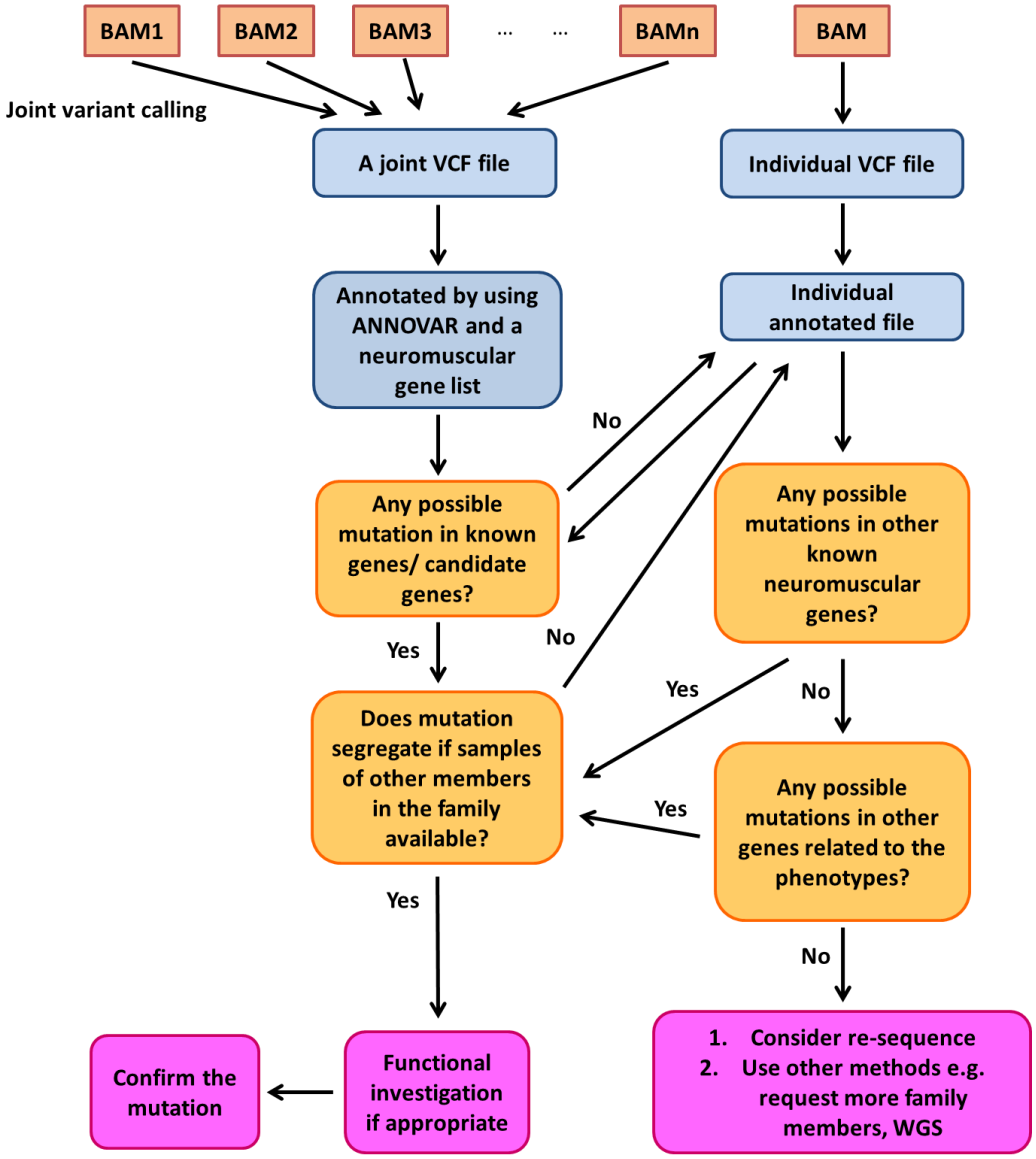


Figure 5-1 Workflow of the variant filtering strategy for this cohort of cases with TAs and CSs

5.3 Results

5.3.1 Coverage and Success Rate of Whole-Exome Sequencing in this Cohort

Two different WES protocols were used due to the development of techniques in the department in 2014. Fourteen patients were sequenced using TruSeq protocol (62 Mb, Illumina) before 2014. Subsequently a further 20 patients were sequenced using Nextera protocol (37 Mb, Illunima). The coverage of WES is variable between samples, mainly because the chemicals of both TruSeq and the first version of Nextera protocols were in relative poor quality, and many of the samples were sequenced during this period of time. The metrics of the whole-exome data is shown

in Table 5-2, with the exception of Case ID 29970. The average coverage at 10x read depths of the samples was 79.1%. The average coverage at 10x read depths of the samples using TruSeq protocol was 76.8%, and that of the samples using Nextera protocol was 80.7%. Although the coverage at 10x read depths was less than 70% in four samples processed by Nextera compared to only three by TruSeq, the average coverage of whole-exome data was improved by switching to the Nextera protocol.

Table 5-2 Metrics of the whole-exomes of cases with TAs and CSs

Case ID/ DNA ID	Sequencing Protocol	Total Reads	Mean Target Coverage	% Target Covered at 2X	% Target Covered at 10X	Total Number of Variants
A-III-5 (78385)	Nextera	91409108	71.3x	99.0%	96.6%	56587
B-III-3 (78386)	Nextera	69971884	51.0x	99.0%	94.8%	54810
Case 1 (MB8561)	Nextera	29969542	25.8x	95.3%	78.3%	20917
Case 2 (19917)	TruSeq	58448580	22.6x	85.3%	53.1%	19267
Case 3 (29308)	TruSeq	117747024	48.0x	92.6%	78.2%	23165
Case 4 (64413)	Nextera	50654784	26.3x	92.1%	69.8%	19932
Case 5 (66205)	TruSeq	106519880	53.5x	97.0%	90.1%	21723
Case 6 (MB8249)	TruSeq	147539478	55.7x	92.1%	73.9%	24879
Case 7 (60595)	TruSeq	108676888	52.2x	97.3%	90.5%	22245
Case 8 (4159)	Nextera	46527292	30.8x	98.6%	87.8%	49532
Case 9 (66204)	TruSeq	67742344	38.1x	92.2%	76.7%	21149
Case 10 (66210)	TruSeq	70758710	38.0x	93.5%	79.6%	21267
Case 11 (9-52)	Nextera	64729386	44.2x	98.3%	92.3%	22966
C-II-1 (60052)	Nextera	109768498	55.2x	95.8%	84.0%	21529
Case 12	Nextera	63291864	36.7x	98.2%	91.4%	27563

(12-1650)						
64412	TruSeq	92188620	42.9x	92.5%	78.0%	21495
40725	Nextera	54414016	30.7x	79.2%	47.1%	17226
66209	TruSeq	95820580	53.9x	94.0%	83.9%	21710
66203	Nextera	57696882	29.1x	90.4%	78.0%	21545
66206	Nextera	67194014	38.4x	89.7%	70.0%	20062
64416	TruSeq	80918072	30.9x	91.8%	77.0%	20953
65142	Nextera	44867036	23.1x	90.6%	65.3%	18147
63568	Nextera	80861034	55.3x	99.1%	93.9%	23732
19817	TruSeq	86441924	33.6x	92.8%	65.2%	19917
64414	TruSeq	99825222	49.2x	96.7%	89.3%	21868
2589	Nextera	105212566	76.9x	99.1%	97.0%	23605
1285	Nextera	66909070	40.6x	98.5%	92.8%	22771
MB7684	Nextera	28838958	24.0x	96.2%	77.2%	20765
MB7052	Nextera	45631002	40.0x	97.0%	87.0%	22314
4580	TruSeq	88728932	37.5x	87.9%	63.4%	22368
4893	TruSeq	156206714	60.8x	92.7%	76.2%	24332
9469	Nextera	59537924	34.4x	86.7%	64.0%	19513
3212	Nextera	56741288	28.2x	92.5%	70.4%	19369
3938	Nextera	51674428	31.7x	92.9%	75.6%	19627

TAs = tubular aggregates; CSs = cylindrical spirals.

Ten index cases with TAs were found to have mutations or probably damaging variants in five known TA genes, including *STIM1*, *ORAI1*, *PGAM2*, *SCN4A*, and *DPAGT1*. Three other cases with TAs were found to have probable damaging variants in two candidate genes, *ALG14*, and *CASQ1*. In addition, two cases with CSs were found to have likely damaging variants in the candidate gene *ATP2A1*. A summary of these variants is shown in Table 5-3, and the detailed results will be discussed in the following sections. Whilst the pathogenicity of some of the identified variants is currently unconfirmed, preliminary results suggest a rate of 15/35 (42.9%) of WES for the detection of known or candidate genes in this cohort of myopathies with TAs and CSs.

Table 5-3 Summary of rare variants found in known/candidate genes in this cohort

Patient	Gene	Phenotype	Mutation
A-II-4	*STIM1*	Limb-girdle muscular dystrophy	p.Asp84Glu (het)
A-III-5	*STIM1*	Exercise-induced myalgia	p.Asp84Glu (het)
B-III-3	*STIM1*	Congenital TAM	p.Glu255Val (het)

B-II-1	*STIM1*	TAM	p.Glu255Val (het)
Case 1	*STIM1*	TAM	p.Leu92Val (het)
Case 2	*ORAI1*	Muscle weakness, fatigue, myalgia	p.Val107Met (het)
Case 3	*ORAI1*	Progressive muscle weakness	p.Gly98Ser (het)
Case 4	*PGAM2*	Exercise-induced myalgia, myoglobinuria	p.Arg10Gln (hom)
Case 5	*PGAM2*	Lambert-Eaton myasthenic syndrome	p.Arg10Trp (het)
Case 6	*PGAM2*	Myalgia, muscle cramps	p.Gly178fs30Ter (hom)
Case 7	*SCN4A*	Myotonic syndrome	p.Val445Leu (het)
Case 8	*DPAGT1*	Exercise-induced muscle spasms and cramps	p.Gly251Glu (het)
Case 9	*ALG14*	Epilepsy, hearing loss, facial weakness	p.Arg104Ter (het) p.Ala11Thr (het)
Case 10	*ALG14*	Possible myasthemia gravis	p.Ala11Thr (het)
Case 11	*CASQ1*	Congenital scapularperoneal	p.Ile385Thr (het)
C-II-1	*ATP2A1*	Exercise-induced myalgia, stiffness, and fatigability	p.Arg637Trp (het) p.Gln920X (het)
Case 12	*ATP2A1*	Autoimmune myasthenia gravis	P.Pro540Leu (hom)

TAM = tubular aggregate myopathy; het = heterozygous; hom = homozygous.

5.3.2 Known Genes in This Cohort

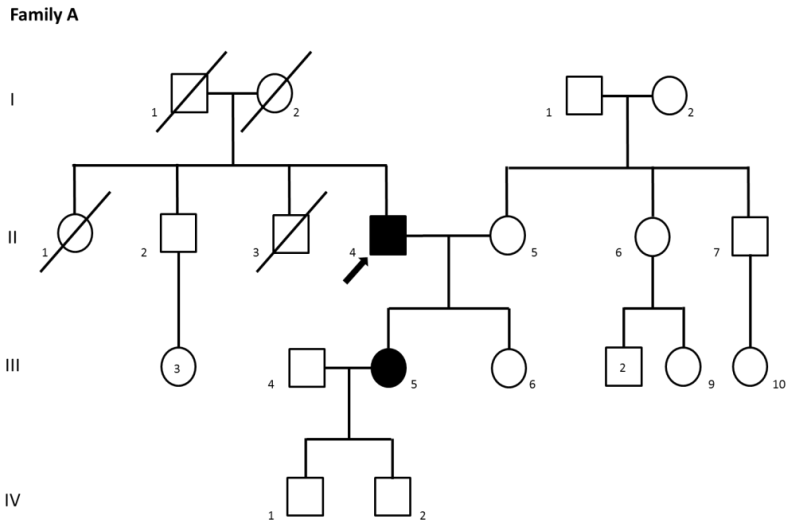
5.3.2.1 STIM1

Two families (A and B) and one single case were found to have rare variants in the *STIM1* gene. Both families were reviewed by Dr John McConville and Professor Parick Marrison in Northern Ireland (Figure 5-2). Family A's phenotypes were previously published by Cameron et al. (Cameron et al., 1992). In Family A, only the index case (Family A-II-4) and one of his daughters (Family A-III-5) were affected. The parents of the index and his three siblings had no weakness or other neuromuscular symptoms. Two children of the affected daughter were also free of these symptoms. Histology of muscle biopsy showed features of TAs in the index patient, while histology of the daughter showed relatively normal fibres with occasional central nuclei. Under EM, there was a similar ultrastructural appearance in both patients. TAs were found in all muscle fibres and were located within the sarcoplasm and adjacent to the sarcolemma. Large TAs were predominant in all fibres in the index patient, while they were much smaller in the daughter. Particulate glycogen was diffusely between the tightly packed tubules (Cameron et al., 1992).

In Family B, the index case (Family B-III-3) was diagnosed with a congenital myopathy with tubular aggregates. The father (B-II-1) of the index case was also affected with tubular aggregate myopathy with the same phenotype. There is no definite history on the other family members. The grandfather (B-I-1) of the index died aged 67 years of heart problems, while the grandmother (B-I-2) died of severe arthritis at the age of 63. She did develop some muscle wasting towards the end of life. None of other siblings of the father and the index patient had primary muscle problems. Detailed information of investigations (e.g. CK level, EMG, and genetic test) and the images of muscle biopsies were not available at the time of writing.

Case 1 was provided by the Neuromuscular Unit, BioBank of Skeletal Muscle, Nerve Tissue, DNA and Cell Lines in Italy. There was limited information available. She had developed muscle weakness in the lower limbs and myalgia from the age of 15. She had an elevated CK level at 2000 IU/L, and a normal EMG. Her family history was unknown. She had a history of thyroidectomy and was on replacement therapy. Muscle biopsy was performed at the age of 51, which showed appearance of TAs in muscle fibres. The images of the biopsy were not available by the time of writing.

The clinical phenotypes of the affected patients from Family A, Family B and Case 1 are summarised in Table 5-4.



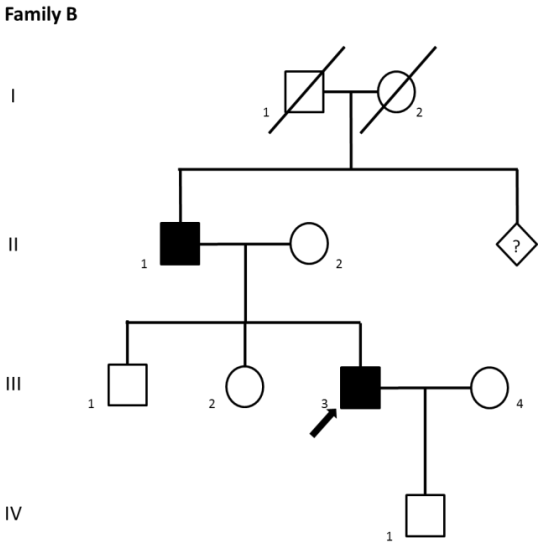


Figure 5-2 Pedigrees of the Family A and Family B

Black symbols represent affected patients; Black arrow points the index patient of the family; Symbols with slash indicate the individual was dead; Numbers inside the symbols indicate multiple individuals with the same gender; Diamond with a question mark indicates that the number of individuals with unknown gender was unclear by the time of writing this thesis.

Table 5-4 Clinical phenotypes of patients in Family A, Family B and Case 1

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other history	CK level	EMG/NCS	Muscle Biopsy
A-II-4 (index)	M	Early childhood	Y	Limb-girdle muscular dystrophy; progressive muscle weakness	Lordosis and walk with a waddling gait; bilateral restriction of eye adduction and upward gaze	Heavy alcohol uptake for several years, but stopped nine months before the consultation	1550 IU/L	EMG: Some increase in polyphasic activity in motor unit potentials; NCS: a possible demyelination change in the median nerve	IHC: features of TAs; EM: confirmed TAs; large TAs predominantly in all fibres types
A-III-5	F	About nine	Y	Progressive muscle weakness, and muscle pain after exercise	Lordosis and walk with a waddling gait; a complex ophthalmoparesis with mostly upgase and some horizontal restrictions	N/A	700 IU/L	EMG: a predominantly myopathic pattern; NCS: within normal limits	IHC: relatively normal fibres with increased EM: confirmed TAs; TAs much smaller compared with above
B-III-3 (index)	M	UK	Y	Congenital myopathy with tubular aggregates	Ophthalmoparesis; mild concentric left ventricular hypertrophy and mild dilated left atrium	N/A	N/A	N/A	EM confirmed TAs
B-II-1	M	UK	Y	TAM	N/A	N/A	N/A	N/A	EM confirmed TAs
Case 1	M	15	UK	TAM; muscle weakness in lower limbs and myalgia	N/A	Thyroidectomy with substitution treatment	2000 IU/L	N/A	EM confirmed TAs

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; IHC = immunohistochemistry; EM = electron microscopy; TAs = tubular aggregates; TAM = tubular aggregate myopathy; Y = yes; N/A = not available; UK = unknown.

WES was performed on the DNA samples from A-III-5, B-III-3 and Case 1. Three missense variants c.252T>A (p.Asp84Glu), c.764A>T (p.Glu255Val), and c.274C>G (p.Leu92Val) in the *STIM1* gene were identified in the three patients, respectively (Table 5-5 and Figure 5-3). The p.Asp84Glu variant was further confirmed to be segregating in A-II-4. The p.Glu255Val variant was segregating in B-II-1. DNA samples of other family members were not available. Amino acids at these three positions were highly preserved across species (Figure 5-3). These missense variants are all absent from the 1000 Genomes project, EVS and ExAC databases, and are predicted as pathogenic by SIFT, PolyPhen2 or MutationTaster.

Table 5-5 Variants found in the *STIM1* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	PolyPhen2/SIFT /MutationTaster
A-II-4 (index) A-III-5	p.Asp84Glu (het)	N/A; N/A; N/A	4.21	N/A	D/T/D
B-III-3 (index) B-II-1	p.Glu255Val (het)	N/A; N/A; N/A	5.38	4.98	D/D/D
Case 1	p.Leu92Val (het)	N/A; N/A; N/A	2.55	N/A	D/D/D

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; N/A = not available; D = damaging; T = tolerated.

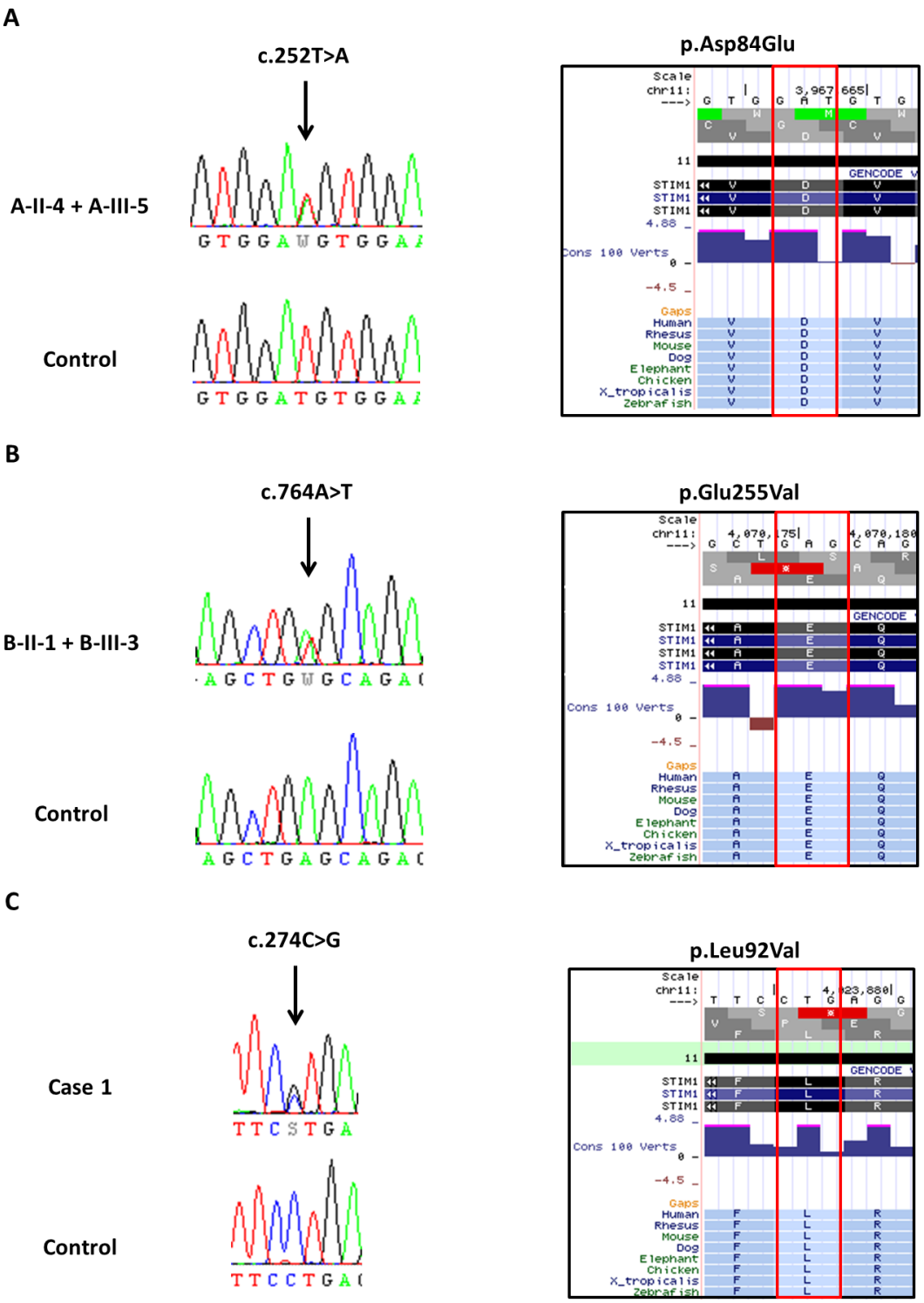


Figure 5-3 Missense variants in the *STIM1* gene

(A) Electropherograms of DNA for the missense variant c.252T>A found in Family A, and conservation of the Aspartic acid at position 84; (B) Electropherograms of DNA for the missense variant c.764A>T found in Family B, and conservation of the Glutamic acid at position 255 across the species; (C) Electropherograms of DNA for the missense variant c.274C>G found in Case 1, and conservation of the Leucine at position 92 across species.

These three missense variants are novel and have not been reported previously in public databases. A mutation at codon 84 was reported in a family with TAM, but for a different amino acid substitution (p.Asp84Gly) (Bohm et al., 2013). Both p.Asp84Glu and p.Leu92Val lie in the EF-hand domain which is known to be a mutation hotspot in the *STIM1* gene and causes dominant TAM by a gain-of-function mechanism (Figure 5-4). The EF-hand is a calcium-binding domain in the ER luminal N-terminus of STIM1, which acts as a calcium sensor. When there is a reduction of calcium level in the ER, calcium is released from the EF-hand, and the EF-hand and SAM domains undergo a further conformational change to dimerise with the neighbouring STIM1 molecules. Mutations in the EF-hand domain are considered to alter calcium binding and/or impair the interaction with the SAM domain (Lacruz and Feske, 2015). A previous study on the mutation p.Asp84Gly in transfected C2C12 myoblasts showed that this mutation triggered STIM1 clustering independent of calcium store depletion. This suggested that the mutant STIM1 was unable to sense calcium in the SR lumen thereby being oligomerised and clustered at the SR plasma-membrane junctions to activate CRACs. This study also demonstrated an increased calcium level in both SR and cytoplasm in cells (Bohm et al., 2013). Therefore, the two mutations p.Asp84Glu and p.Leu92Val are postulated to underlie the similar pathogenic mechanism to induce the phenotypes of the disease. Functional confirmation of this hypothesis would be required in the future.

The p.Glu255Val mutation is located in the cytoplasmic CC1 domain. In the inactive state, CC1 domain binds to CRAC-activating domain (CAD) / STIM1-Orai-activating region (SOAR) forming a folded STIM1 C-terminus. When STIM1 is activated by the depletion of ER calcium, the conformation of CC1-CAD/SOAR is extended. This is a critical process in binding to ORAI1 in the plasma membrane and to activate the CRAC channel (Lacruz and Feske, 2015). Mutations in this region may affect the interaction between CC1 and CAD/SOAR in the inactive state. Only one mutation (p.Arg304Trp) has previously been reported in the CC1 domain in patients with familial Stormorken syndrome (Misceo et al., 2014, Morin et al., 2014, Nesin et al., 2014). Functional analysis confirmed that the p.Arg304Trp substitution acted as an activating mutation to the CRAC channel and caused elevated calcium level in cells (Misceo et al., 2014, Nesin et al., 2014). Further analysis would be required to understand the exact consequence of the p.Arg255Glu.

On reviewing previous papers on cases with STIM1-related TAs, when the mutations were located in the EF-hand domain, the phenotypes of the patients were predominantly in the skeletal muscle and were called TAM. Mutations in CC1 domain also affected platelets, and were associated with Stormorken syndrome. Interestingly, although Family B had a mutation in the CC1 domain of STIM1, all of our affected patients presented mainly with symptoms of TAM (with elevated CK level when measured) and there was no evidence suggestive of Stormorken syndrome. Both Family A and Family B presented with ophthalmoparesis, and ophthalmoparesis has been reported in eight other patients (Lacruz and Feske, 2015). The two affected patients in Family A also had severe lordosis, which for the first time is in association with TAs. Therefore, our findings expand the clinical phenotypes of TAM related to the *STIM1* mutations.

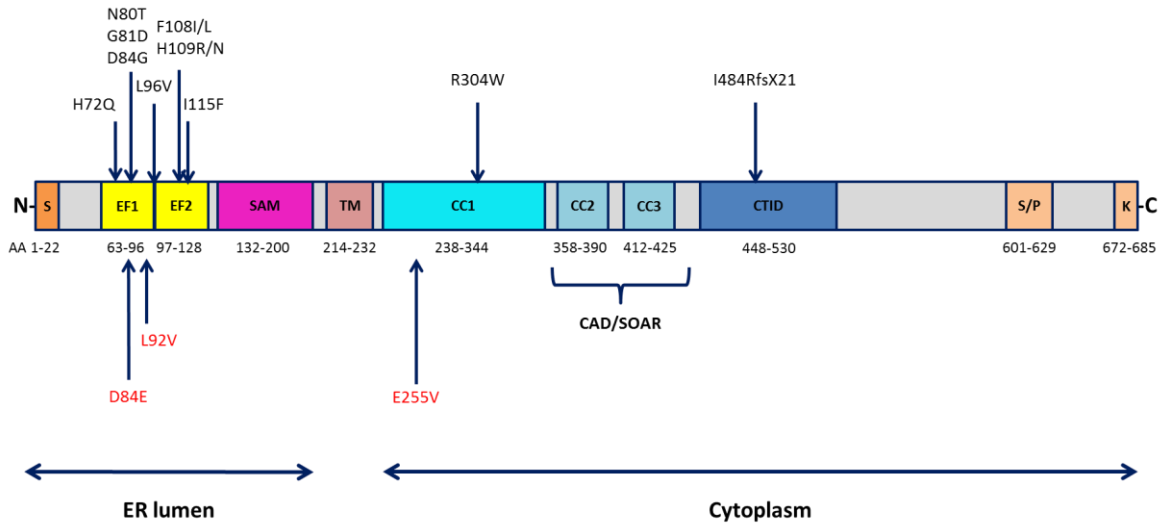


Figure 5-4 Protein domains of STIM1

The missense and frameshift mutations on the top were previously reported in cases with TAM, the missense variants found in our cases are labelled in red at the bottom. N = N-terminal; S = serine-rich domain; EF1 = canonical EF-hand domain; EF2 = noncanonical EF-hand domain; SAM = sterile alpha motif; TM = transmembrane domain; CC1-CC3 = coiled-coil domains; CAD/SOAR = CRAC activation domain/ STIM1-ORAI1 activation region; CTID = C-terminal inhibitory domain; S/P = serine- and proline-rich domain; K = polylysine; C = C-terminal; ER = endoplasmic reticulum.

5.3.2.2 ORAI1

Rare variants in the *ORAI1* gene were identified in two cases of this cohort. Case 2 and Case 3 were two unrelated patients both previously seen by neurologists at NHNN. Case 2 was previously published as a case report in 2002 (Jacques et al.,

2002). This patient started experiencing symptoms of upper limb weakness and fatigue, and myalgia in the lower limbs from the age of 47. Case 3 first experienced symptoms aged 17. She developed progressive proximal weakness that affected the lower limbs more than the upper limbs, until she became wheelchair-dependent outdoors. There was no family history of muscle disease in either case. A muscle biopsy from Case 2 was performed at the age of 50 and it showed features of severe accumulations of TAs in both type I and type II fibres (Figure 5-5). EM confirmed the presence of accumulated double-walled TAs (Figure 5-5 H). Case 3 had a muscle biopsy at the age of 40. Many of the fibres had subsarcolemmal and internal appearance of TAs and these were also confirmed on EM (Figure 5-6). Both muscle biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION. A summary of the clinical phenotypes of both patients is provided in Table 5-6.

Table 5-6 Clinical phenotypes of Case 2 and Case 3

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK Level	EMG/NCS	Muscle Biopsy
Case 2	M	47	N	Muscle weakness and fatigue in the upper limbs and myalgia in the lower limbs	A mild saddle nose deformity and pectus excavatum. Both pupils were markedly constricted to 1.5 mm in the dark. They responded briskly to light and accommodation, and were unresponsive to mydriatic agents	Less than six units of alcohol per week; no exposure to myotoxic drugs	376-3854 IU/L	Mild chronic myopathic but not neuropathic changes	Figure 5-5
Case 3	F	17	N	Slowly progressive muscle weakness in the proximal limbs (legs > arms)	Hypothyroidism and hypocalcaemia secondary to thyroidectomy	A history of thyrotoxicosis and underwent thyroidectomy	274 IU/L	EMG: mild limb girdle myopathy; NCS: no evidence of peripheral neuropathy	Figure 5-6

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; N = no.

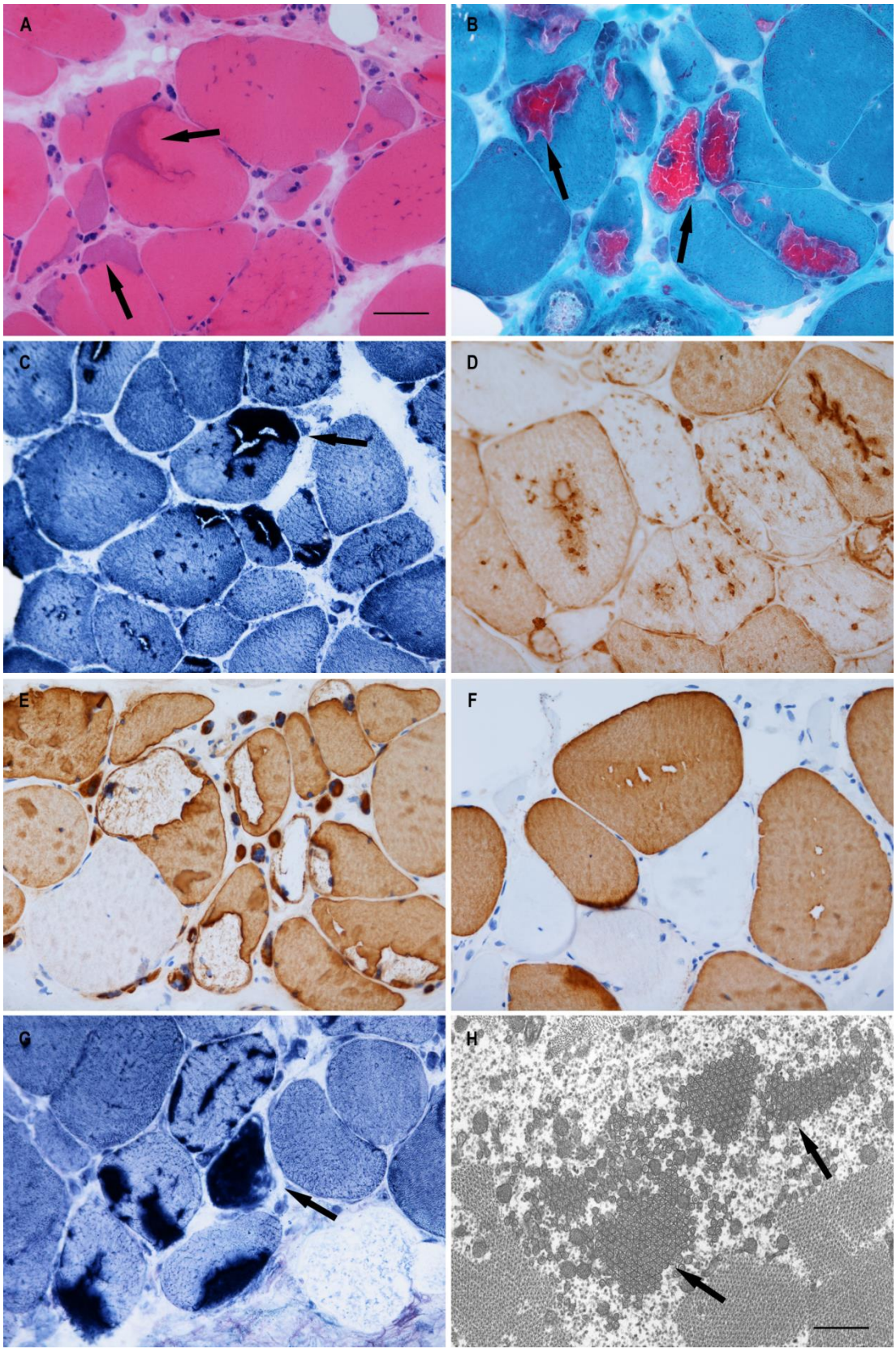


Figure 5-5 Muscle biopsy features observed in Case 2

In H&E staining, there was variation in fibre size with frequent atrophic fibres and increased connective tissue, and marked basophilic inclusions in many fibres (A); Gomori trichrome staining showed inclusions in bright red and severely affected fibres (B); The inclusions were stained dark in

NADH-TR (C) and adenylate deaminase (G); The inclusions were also positive for SERCA II (D); Fast myosin (E) and slow myosin (F) stainings suggested both type I and type II fibres were affected; Accumulations of double-walled tubular aggregates were confirmed in electron microscopy (EM) (H). Inclusions of tubular aggregates were pointed by arrows in images A-C, G and H. Scale bar represents 50 μm in A-G for histology images; Scale bar in H represents 500 nm in EM. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. Biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

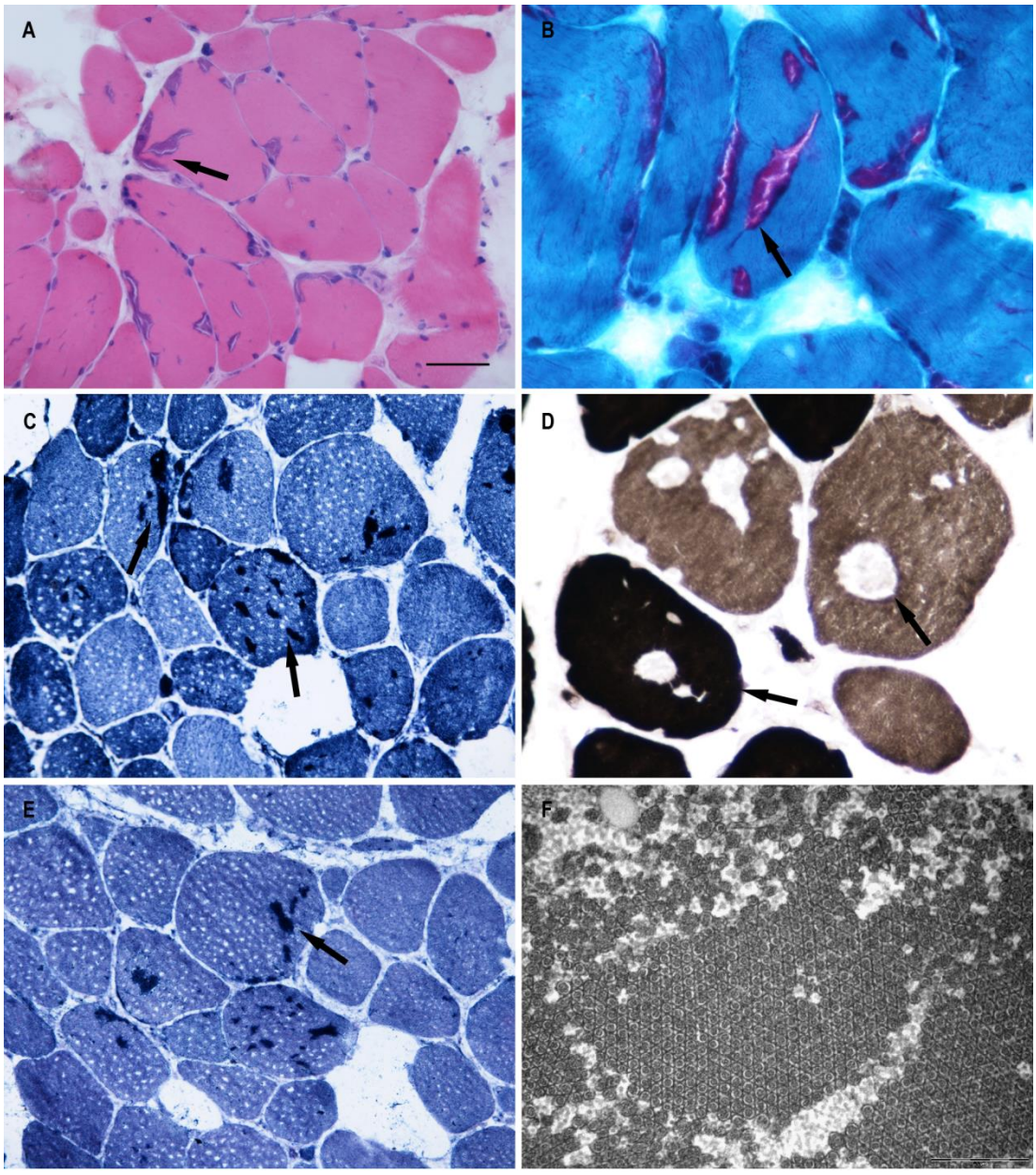


Figure 5-6 Muscle biopsy features observed in Case 3

There was increased variation in fibre size with subsarcolemma and internal basophilic inclusions in many of the fibres in H&E (A); Gomori trichrome showed the inclusions stained in red (B); The inclusions were darkly stained in NADH-TR (C) and adenylate deaminase (E); The type II fibres were lightly stained compared with the type I fibres in NADH-TR(C), and the type II fibres were darker

stained in ATPase pH9.4 (D), and both markers confirmed that inclusions were present in both type I and type II fibres; Accumulations of tubular aggregates were confirmed in EM (F). Inclusions of tubular aggregates were pointed by arrows in images A-E. Scale bar in histology images represents 50 μm in A, C and E; 25 μm in B and D; Scale bar in EM represents 500 nm. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. Biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

Two missense variants c.319G>A (p.Val107Met) and c.292G>A (p.Gly98Ser) in the *ORAI1* gene were identified in Case 2 and Case 3, respectively by WES, and confirmed by Sanger sequencing (Table 5-7 and Figure 5-7). Amino acid residues at these two positions are highly conserved across species. The p.Val107Met variant found in Case 2 is a novel variant, and is predicted as probably deleterious. The p.Gly98Ser variant has been reported in two families with dominant TAM and hypocalcemia (Endo et al., 2015). Both variants are absent from the 1000 Genomes project, EVS and ExAC databases.

Table 5-7 Variants found in the *ORAI1* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	Polyphen2/SIFT/ MutationTaster
Case 2	p.Val107Met (het)	N/A; N/A; N/A	5.36	4.32	D/D/D
Case 3	p.Gly98Ser (het)	N/A; N/A; N/A	3.94	5.39	D/D/D

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; N/A = not available; D = damaging

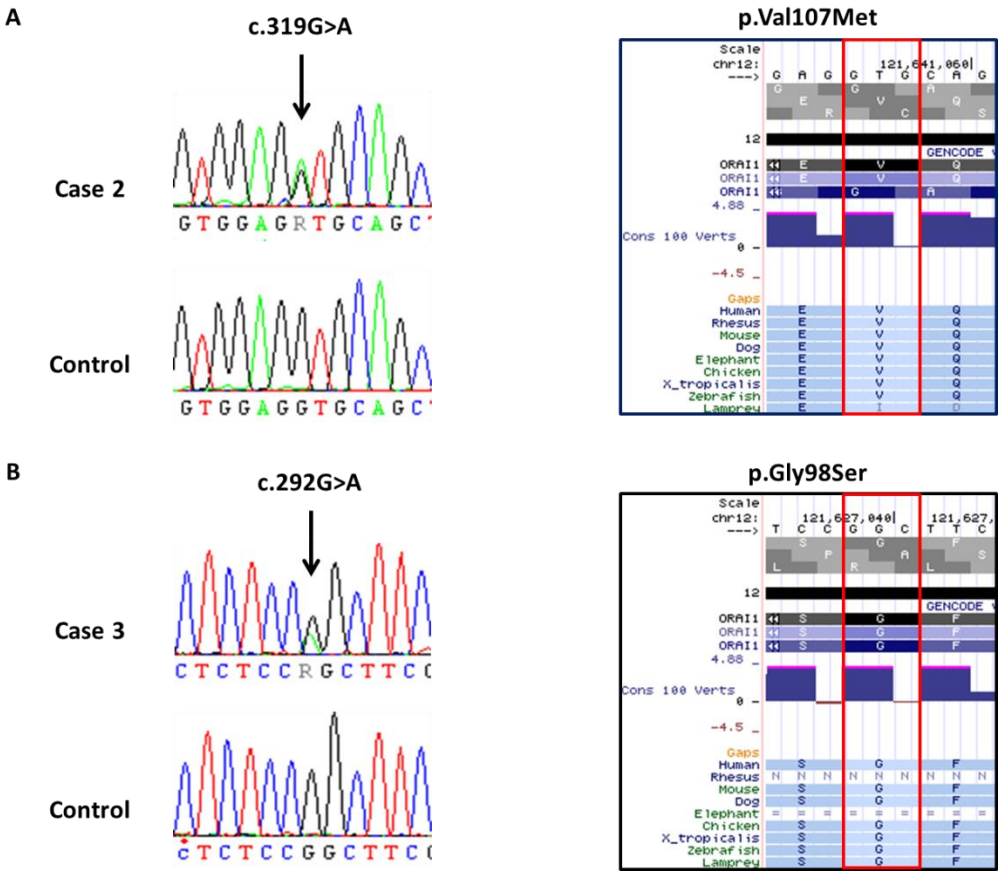


Figure 5-7 Missense variants in the *ORAI1* gene

(A) Electropherograms of DNA for the heterozygous missense variant c.319G>A found in Case 2, and conservation of the Valine at position 107; (B) Electropherograms of DNA for the heterozygous missense variant c.292G>A found in Case 3, and conservation of the Glycine at position 98.

ORAI1 is a tetra-spanning transmembrane protein in the plasma membrane, and acts as an ion-conducting pore subunit of the CRAC channel. ORAI1 consists of four alpha-helical transmembrane domains (M1-M4), two extracellular loops, and intracellular N- and C-terminus. The functional CRAC channel is a hexamer of ORAI1 proteins, of which the central ion pore is composed of a ring of six M1 domains. The p.Gly98Ser mutation is located in the transmembrane M1 domain directly beside the pore. A previous study showed that this mutation caused constitutive opening of the channel, leading to increased calcium influx into the cytoplasm independent of either calcium stores in SR or STIM1 activation (Endo et al., 2015). Amino acid residues in the M1 domain are thought to have functions in both channel activity and the ion selectivity of store-operated calcium entry (SOCE) channels. Gly98 has been suggested to be involved in the channel activity *in vitro* (Zhang et al., 2011). The mutation replaces Gly98 with the hydrophilic amino acid

serine, which may alter the constitutive channel activity (Endo et al., 2015). Val107 is the first residue located in the extracellular loop and adjacent to the M1 domain. Glu106 is located towards the outer end of the pore and is responsible for calcium binding and the high calcium selectivity of the CRAC channel. Although the exact function of Val107 residue is still unknown, replacing valine at position 107 with methionine may also confer the protein with a hydrophilic property. This may affect the adjacent Glu106 residue and cause constitutive channel activation. Functional confirmation of pathophysiological function of Val107 residue would be required in the future.

Both of the patients had no family history of their similar phenotypes, suggesting these are probably *de novo* mutations. Information on their off-springs and DNA samples from family members were both unavailable. Case 2 harboured the novel mutation and had TAM as well as miosis. Similar phenotypes had previously been seen in two other familial TAM cases with a mutation in *ORAI1* (c.734C>T, p.Pro245Leu) (Nesin et al., 2014). Because these patients lacked other symptoms present in Stormorken patients, such as thrombocytopenia, bleeding diathesis and asplenia, they were also described as Stormorken-like syndrome (Nesin et al., 2014). Case 3 harboured a known mutation p.Gly98Ser in *ORAI1* and presented with progressive muscle weakness. Previously reported patients with the same mutation had decreased serum calcium levels and rather low parathyroid hormone (PTH) levels, which were considered as PTH deficient hypoparathyroidism. Case 3 had hypocalcaemia, but it may be a result of thyroidectomy in her later teens which damaged her parathyroid glands. Therefore, whether there is a correlation between the mutation and decreased PTH secretion and hypocalcemia is unknown.

5.3.2.3 PGAM2

Three patients from the cohort were found to harbour interesting variants in the *PGAM2* gene. Case 4 and Case 5 were two unrelated patients both previously seen by neurologists at NHNN. Case 4 is a 41-year old Egyptian patient from a consanguineous family. He initially presented symptoms at around 26 years of age after he underwent a period of vigorous exercise and within hours developed loin pain and anuria. He had no muscle weakness in the limbs. Case 5 was diagnosed with Lambert Eaton myasthenic syndrome. He had proximal muscle weakness, myalgia

and muscle stiffness particularly affecting the facial muscles. Muscle biopsy from Case 4 showed numerous muscle fibres with TAs which were often subsarcolemmal but also internal in type II fibres (Figure 5-8 A-D). Both the phosphorylase and adenylate deaminase activities were increased in TAs, which also had increased glycogen content. Muscle biopsy from Case 5 showed atrophy in type II fibres, occasional muscle fibres that had been reduced to clumps of pyknotic nuclei, and occasional TAs in fibres (Figure 5-8 E-H). The lipid and glycogen content of the fibres were normal. Muscle biopsies of both patients were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

Case 6 was provided by the Neuromuscular Unit, BioBank of Skeletal Muscle, Nerve Tissue, DNA and Cell Lines in Italy. Case 6 is a 47-years old man from a consanguineous family where the parents are first cousins. He started to have symptoms at a few months of age and developed myalgia and muscle cramps. He had an elevated CK level at 606-76000 IU/l. EMG showed a myopathic pattern. Muscle biopsy from the biceps was performed at the age of 38, and showed the presence of subsarcolemmal TAs in over 50% of fibres with PAS-positive vacuoles. The biopsy images were not available by the time of writing.

The clinical phenotypes of the three patients has been summarised in Table 5-8.

Table 5-8 Clinical phenotypes of Cases 4-6

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK Level	EMG/NCS	Muscle Biopsy
Case 4	M	26	Consanguineous parents	Myalgia and anuria after intense exercise; myoglobinuria	Rhabdomyolysis	Haemodialysis due to acute renal failure; a history of probable viral induced thrombocytopenia at the age of 7	> 2000 IU/L	N/A	IHC: Figure 5-7 A-D
Case 5	M	31	N	Lambert-Eaton myasthenic syndrome (LEMS) with positive anti-VGCC antibodies (diagnosed at the age of 39); aching and stiffness in the calves and thighs with progressive weakness	Bilateral cataracts; poor vision in his right eye following a work related accident; mildly poor eye closure on the left side; a left sided lower motor neuron facial palsy; weight loss; a tight compressive central chest pain with no radiation; excessive sleepness during the day.	Treatment with nocturnal CPAP for a couple of years due to respiratory muscle weakness; previous sleep studies revealed frequent episodes of desaturation overnight, and was on nocturnal NIPPV treatment for 2 years. Previously heavy smoker	Normal	EMG: a disorder of NMJ trans-mission; the specific changes associated with LEMS were not seen; no neurogenic or myopathic features in the respiratory muscles	IHC: Figure 5-7 E-H
Case 6	M	A few months	Consanguineous parents	Myalgia and muscle cramps	N/A	N/A	606-76000 IU/L	EMG: myopathic pattern	IHC: presence of subsarcolemmal TAs in over 50% of fibres with PAS-positive vacuoles

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; IHC = immunohistochemistry; F = female; M = male; N = no; N/A = not available; NMJ = neuromuscular junction; VGCC = voltage-gated calcium channels; CPAP = continuous positive airway pressure; NIPPV = nasal intermittent positive pressure ventilation; TAs = tubular aggregates; PAS = periodic acid-Schiff.

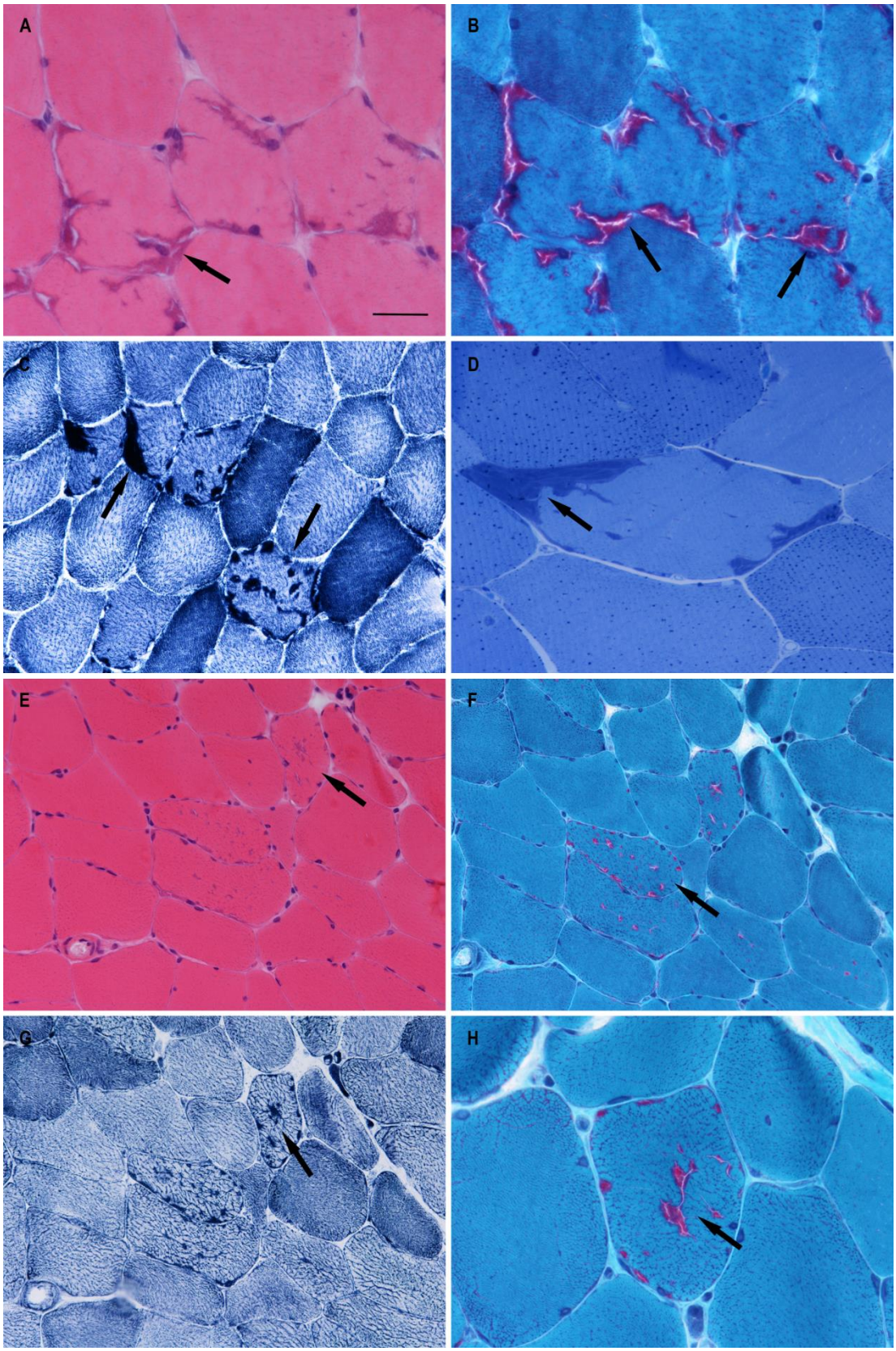


Figure 5-8 Muscle biopsy features observed in Case 4 and Case 5

Case 4 (A-D): There were numerous subsarcolemma and internal basophilic inclusions suggesting TAs in muscle fibres stained in H&E (A), while these inclusions were stained as red in Gomori trichrome (B), and were stained as dark in NADH-TR and mainly in type II fibres (C); In Toluidine

blue, the inclusions were in dark blue, with small dark dots in fibres representing lipid droplets (D); Case 5 (E-H): There were small basophilic inclusions suggesting TAs in muscle fibres in H&E (E), which were stained in red in Gomori trichrome (F and H); The small dark stained inclusions mostly in type II fibres shown in NADH-TR (G); Appearance of red stained TAs in one fibre (H). Inclusions of TAs were pointed by arrows in all the images. Scale bar represents 25 μm in A, B, D, and H; represents 50 μm in C, E-G. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. Biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

WES was performed on the three cases, and found variants in the *PGAM2* gene. This gene encodes muscle phosphoglycerate mutase-2 which is a glycolytic enzyme that catalyses the interconversion of 2-phosphoglycerate and 3-phosphoglycerate using 2,3-bisphosphoglycerate as a cofactor (Grisolia and Carreras, 1975). A homozygous missense variant c.29G>A (p.Arg10Gln) was identified in Case 4 (Figure 5-9 A), a heterozygous missense variant c.28C>T (p.Arg10Trp) in Case 5 (Figure 5-9 A), and a homozygous frameshift deletion c.532delG (p.Gly178fs30Ter) in Case 6 (Figure 5-9 B). All three variants are absent from the 1000 Genomes project database, and were found with frequency less than 0.02% in the ExAC database. Amino acid residues at these two positions are highly conserved across species (Figure 5-9). Both p.Arg10Gln and p.Arg10Trp mutations are predicted to be possibly deleterious by PolyPhen2. The p.Gly178fs30Ter deletion causes a frameshift in exon 2 resulting in premature termination codon (Table 5-9). DNA samples from their parents were not available for segregation analysis.

Table 5-9 Variants and a frameshift deletion found in the *PGAM2* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	Polyphen2/SIFT/ MutationTaster
Case 4	p.Arg10Gln (hom)	N/A; N/A; 0.00004066	5.38	4.78	D/D/D
Case 5	p.Arg10Trp (het)	N/A; N/A; 0.0001057	5.38	4.55	D/D/D
Case 6	p.Gly178fs30Ter (hom)	N/A; 0.00008; 0.00005694	N/A	N/A	N/A

MAF = minor allele frequency; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; hom = homozygous; het = heterozygous; N/A = not available; D = damaging

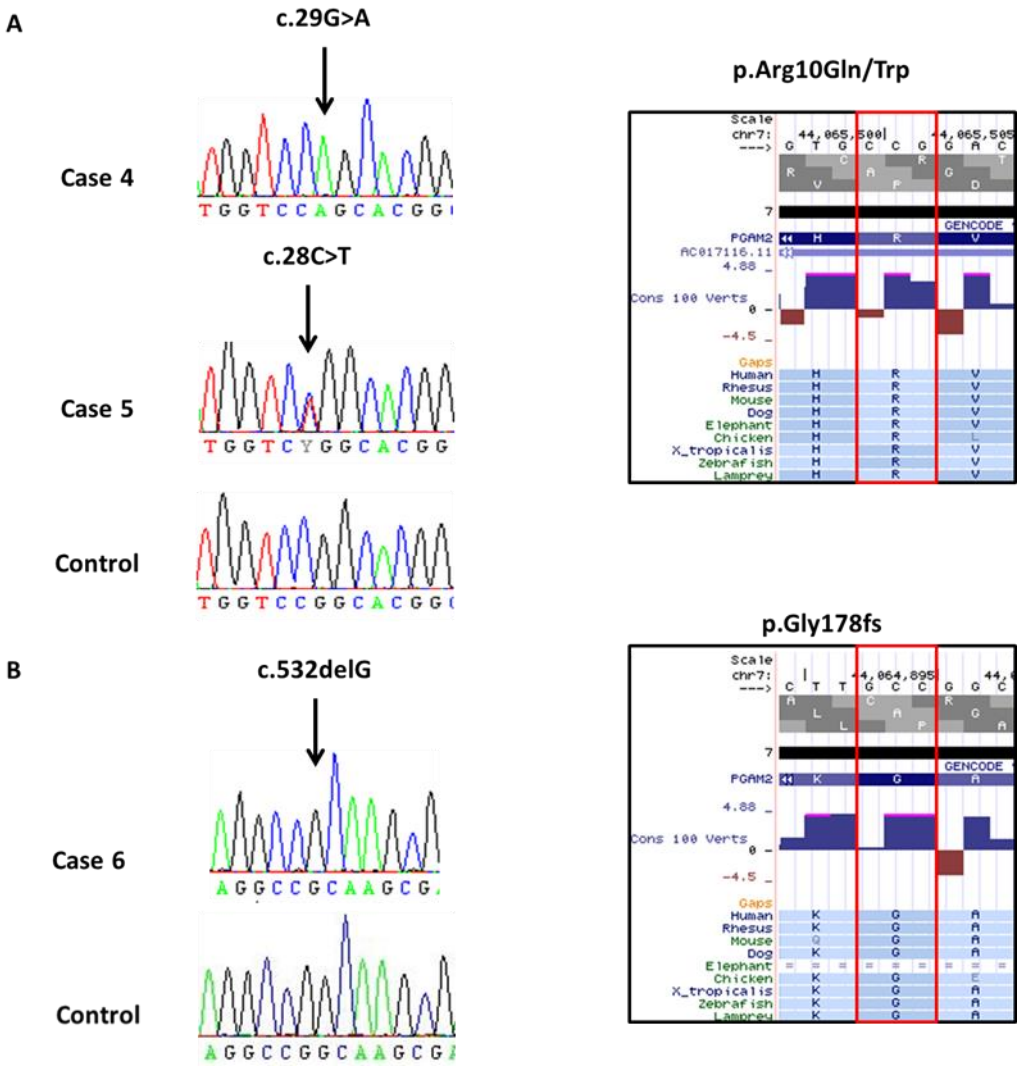


Figure 5-9 Missense variants and a frameshift deletion in the *PGAM2* gene

(A) Electropherograms of DNA for the homozygous c.29G>A and heterozygous c.28C>T missense variants found in Case 4 and Case 5, and conservation of the Arginine at position 10; (B) Electropherograms of DNA for the heterozygous c.532delG frameshift deletion found in Case 6, and conservation of the Glycine at position 178.

Case 4 with homozygous p.Arg10Gln substitution and Case 6 with homozygous p.Gly178fs30Ter mutation had the typical clinical presentation of PGAM deficiency since they both had myalgia and muscle cramps, and elevated serum CK levels. Case 4 additionally had episodes of myoglobinuria. They were both from consanguineous families and their genetic findings fulfilled the autosomal recessive inherited phenotypes. The compound heterozygous variant p.Arg10Gln and deletion p.Gly178fs30Ter were previously reported in an Italian patient in his 40s with persistently elevated CK but without exercise intolerance, myoglobinuria, nor TAs in his biopsy (Tonin et al., 2009). Homozygous p.Gly178fs30Ter was reported in

another Italian patient in his 60s with exercise intolerance and myalgia, and TAs in the biopsy (Naini et al., 2009). PGAM activities were found reduced to 3% and 5% of normal respectively in these two previous cases (Naini et al., 2009, Tonin et al., 2009), although the PGAM activity was not investigated in our patients. Case 6 had symptom onset from infancy, which indicated that the frameshift deletion may cause a more severe phenotype compared with the missense mutations. The previous case with homozygous p.Gly178fs30Ter had a late onset of symptoms, and was additionally diagnosed with a possible late-onset statin myopathy. The authors suggested that clinically silent metabolic myopathies might be unmasked due to the exposure to statins (Naini et al., 2009).

Case 5 had a diagnosis of Lambert-Eaton myasthenic syndrome with aching and stiffness in the calves. The heterozygous missense variant p.Arg10Trp found in Case 5 caused a different amino acid change at the same codon as Case 4. This variant has not been previously reported in patients with PGAM deficiency, and its function was unclear. It is next to codon 11 (His11) that is considered to be an active site of the enzyme. Substitution of the charged arginine with the uncharged tryptophan may have deleterious consequences particularly due to its close proximity to the active site (Tsujino et al., 1993). Thus, this suggests that p.Arg10Trp may be pathogenic. Heterozygote for a mutation (Gly97Asp) was reported in a Japanese family with PGAM deficiency (Hadjigeorgiou et al., 1999). No additional variant was found in other protein coding regions of *PGAM2* by Sanger sequencing, but there could be other types of variants such as copy number variants or large indels in introns which were missed by Sanger sequencing. However, our patient did not present with exercise intolerance, myoglobinuria or elevation of serum CK level, which makes this missense variant unlikely the culprit to the disease. Biochemical analysis of the patient's PGAM activity would be needed to confirm the effect of this variation. Variants in other muscle genes should also be investigated as a next step.

5.3.2.4 SCN4A

A variant of interest in *SCN4A* was found in one patient of the cohort. This patient (Case 7) is currently being seen by neurologists at NHNN. He is a 65-year old who was normal healthy baby with normal motor milestones. He first developed symptoms around the age of 12. Whilst playing rugby, he suddenly "could not move".

This subsequently became a persistent and frequent feature. He also found that his muscle became very stiff and painful, so that he was unable to walk after rugby trainings. Mostly symptoms affected his legs but occasionally he also had stiffness in the hands, abdominal wall muscles and the eyelids. He did not report symptoms of jaw stiffness and dysphagia. He became less active and developed significant and widespread osteoarthritis. He also had a history of obstructive sleep apnoea with morning headache and daytime fatigue. He denied sensory loss but occasional tingling in his hands. He had some lower back pain due to multiple previous operations for likely a lumbar vertebral fracture following trauma. He had two children and they both probably had similar phenotypes of muscle stiffness according to the clinical letter, but the detailed phenotypes were unknown. His mother was deceased and his father was in his 90s with some possible mild symptoms of muscle stiffness. He was a non-smoker and had a modest alcohol intake. On examinations, there was hypertrophy in his calves. He had eyelid myotonia which was probably paradoxical. Cranial nerve was normal. Tone was normal in all four limbs with preserved power, but areflexic throughout. He had distal sensory loss to pinprick to the wrists and ankles bilaterally with reduced proprioception to the ankles. CK level was mildly elevated around 204-216 IU/L. EMG showed significant length dependent sensory motor polyneuropathy with both axonal features and evidence of motor axonal loss. Muscle biopsy at age 59 revealed features suggestive of acute denervation, and a small number of fibres with TAs (Figure 5-10). The biopsy was reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

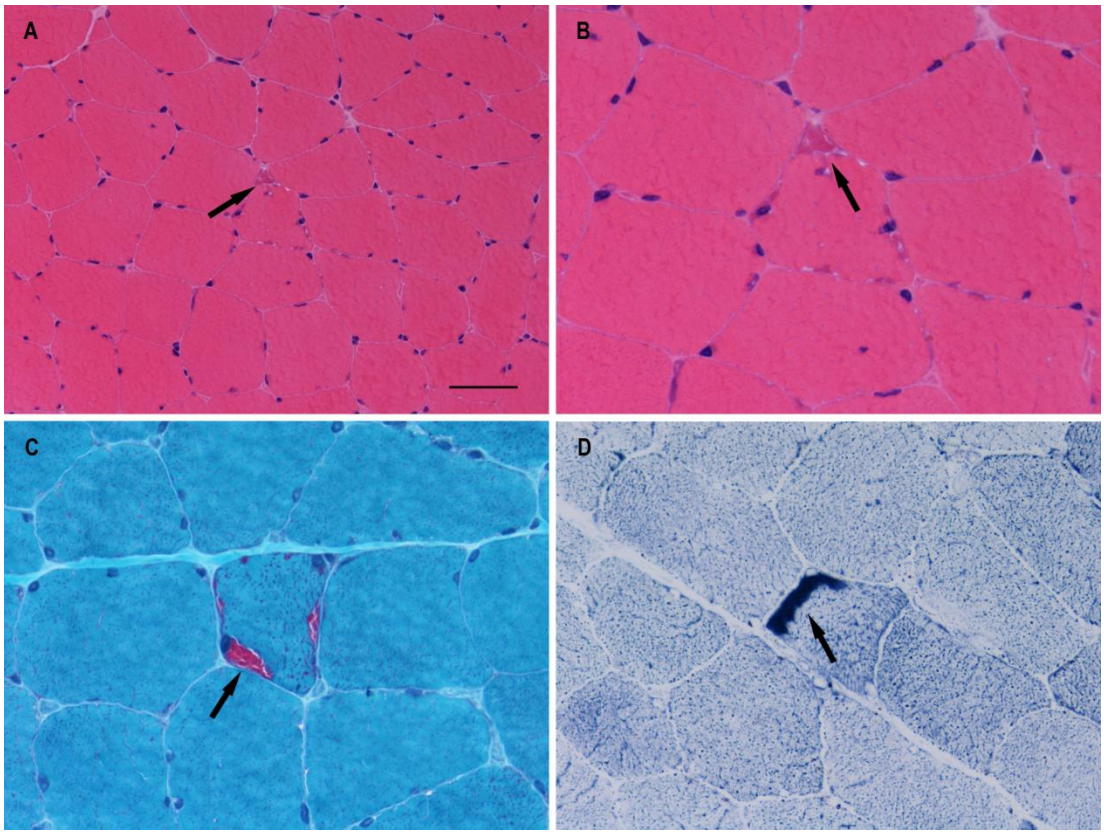


Figure 5-10 Muscle biopsy features in Case 7

There were increased variation in fibre size and scattered atrophic fibres which were mostly angular or polygonal in outline. Occasional fibres contain basophilic inclusions suggestive of TAs in H&E (A); TAs in one fibre stained in H&E was shown in (B); TAs in one fibre stained as red in Gomori trichrome (C), and darkly stained in NADH-TR (D). Inclusions of TAs were pointed by arrows in all the images. Scale bar represents 50 μm in A, and 25 μm in B-D. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. Biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

WES was performed on DNA sample from Case 7, and identified a heterozygous novel missense variant in exon 9 c.1333G>C (p.Val445Leu) of *SCN4A* (Figure 5-11). *SCN4A* encodes the alpha subunit of the predominant voltage-gated sodium channel found in skeletal muscle. This variant was absent from the 1000 Genomes project, EVS and ExAC databases, and was predicted as probably deleterious (Table 5-10). The samples from his two affected children were not available for segregation analysis at the time of writing.

Table 5-10 Variant found in the *SCN4A* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	Polyphen2/SIFT /MutationTaster
Case 7	p.Val445Leu (het)	N/A; N/A; N/A	N/A	4.49	D/D/D

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; N/A = not available; D = damaging.

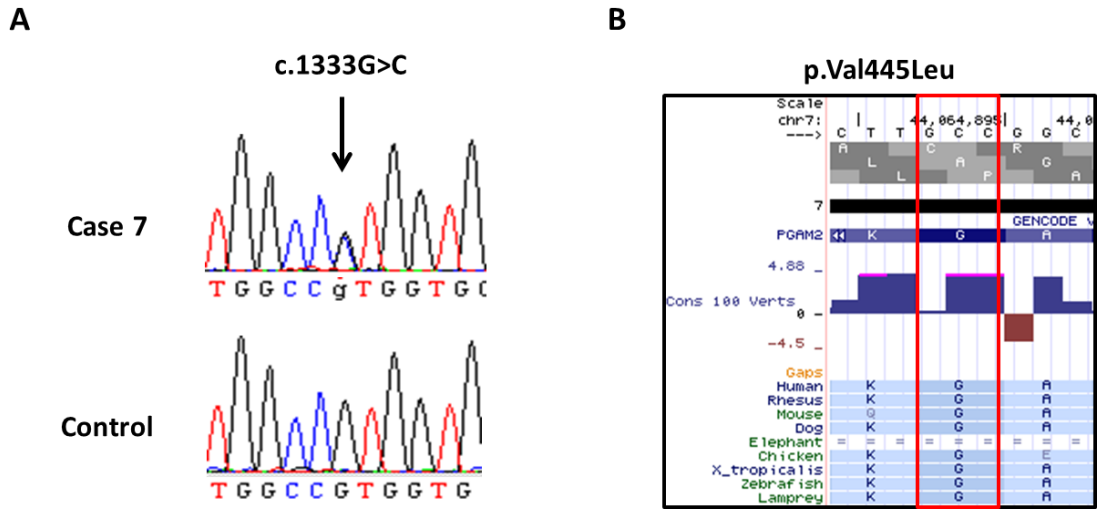


Figure 5-11 Missense variant in the *SCN4A* gene

(A) Electropherograms of DNA for the heterozygous c.1333G>C found in Case 7; (B) Conservation of the Valine at position 445.

The p.Val445Leu variant is located in the sixth transmembrane segment of domain 1 of the skeletal muscle sodium channel (*SCN4A*). A valine to methionine substitution at the same position 445 (p.Val445Met) was reported previously in a family with an unusual autosomal dominant painful congenital myotonia (Rosenfeld et al., 1997), and also in French Canadian patients with painful generalised myotonia and severe muscle hypertrophy (Dupré et al., 2009). A study revealed that the mutation p.Val445Met caused the disease by subtle defects in the inactivation properties of the sodium channel. The mutant sodium channel exhibited a small non-inactivating current during short test depolarizations, and a shift in the voltage-dependence of channel activation to more negative potentials. This resulted in slowing the recovery time from inactivation (Wang et al., 1999). Our patient presented muscle stiffness and hypertrophy in his calves but did not report myalgia. Interestingly, his EMG showed evidence of axonal neuropathy, which was not seen in previous cases with p.Val445Met mutation. The differences in the phenotypes may be due to the different

amino acid substitution. Although the segregation in other members of the family could not be performed, taken the likely autosomal dominant inheritance of this family, and the clinical phenotypes of the patient, we consider that p.Val445Leu variant is likely to be the culprit to cause the patient's myotonic phenotype. Whether the neuropathic feature is also related to this variant is unclear, which cannot exclude the possibility of other genetic contributions. The functional analysis of this novel variant is currently ongoing by Dr Roope Mannikko within the Department of Molecular Neuroscience.

5.3.2.5 DPAGT1

Variants in the *DPAGT1* gene were found in one patient of this cohort. This patient (Case 8) was previously seen by neurologists at NHNN. He had exercise-induced muscle spasms since childhood. TAs were observed in his muscle biopsy (images were not available at the time of writing). His mother also had similar muscle cramps. His two siblings were well. Details of phenotypic features are provided in Table 5-11.

WES identified a novel heterozygous missense variant c.752G>A (p.Gly251Glu) in *DPAGT1*. This variant was predicted as deleterious and was confirmed by Sanger sequencing. As *DPAGT1* is an autosomal recessive gene, Sanger sequencing was performed to check if another variant in the other protein coding region was missed by WES. This only found a common missense variant c.1177A>G (p.Ile393Val) in exon 9 of *DPAGT1*, and it was thought to be a benign variant. (Table 5-12 and Figure 5-12)

Table 5-11 Clinical phenotypes of Case 8

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK level	EMG/NCS	Muscle Biopsy
Case 8	M	Childhood	Y	Muscle spasms in toes and arches of both feet; symptoms triggered by exercise; symptoms worse in cold or wet weather; jaw cramp when yawning	Had a history of parasethesia in both hands on waking; non-Hodgkin's lymphoma	On treatment with Quinidine; smoker (10 cigarettes/day), moderate alcohol intake	527 IU/L	EMG: limited information as patient fainted; NCS: normal	IHC: features of TAs confined to type II fibres; EM: confirmed TAs

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; N = no; N/A = not avialable; IHC = immunohistochemistry; EM = electron microscopy; TAs = tubular aggregates.

Table 5-12 Variants found in the *DPAGT1* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	Polyphen2/SIFT /MutationTaster
Case 8	p.Gly251Glu (het)	N/A; N/A; N/A	5.55	5.23	D/D/D
	p.Ile393Val (het)	0.43;0.39;0.42	N/A	N/A	B/N/A/B

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; N/A = not available; D = damaging; B = benign.

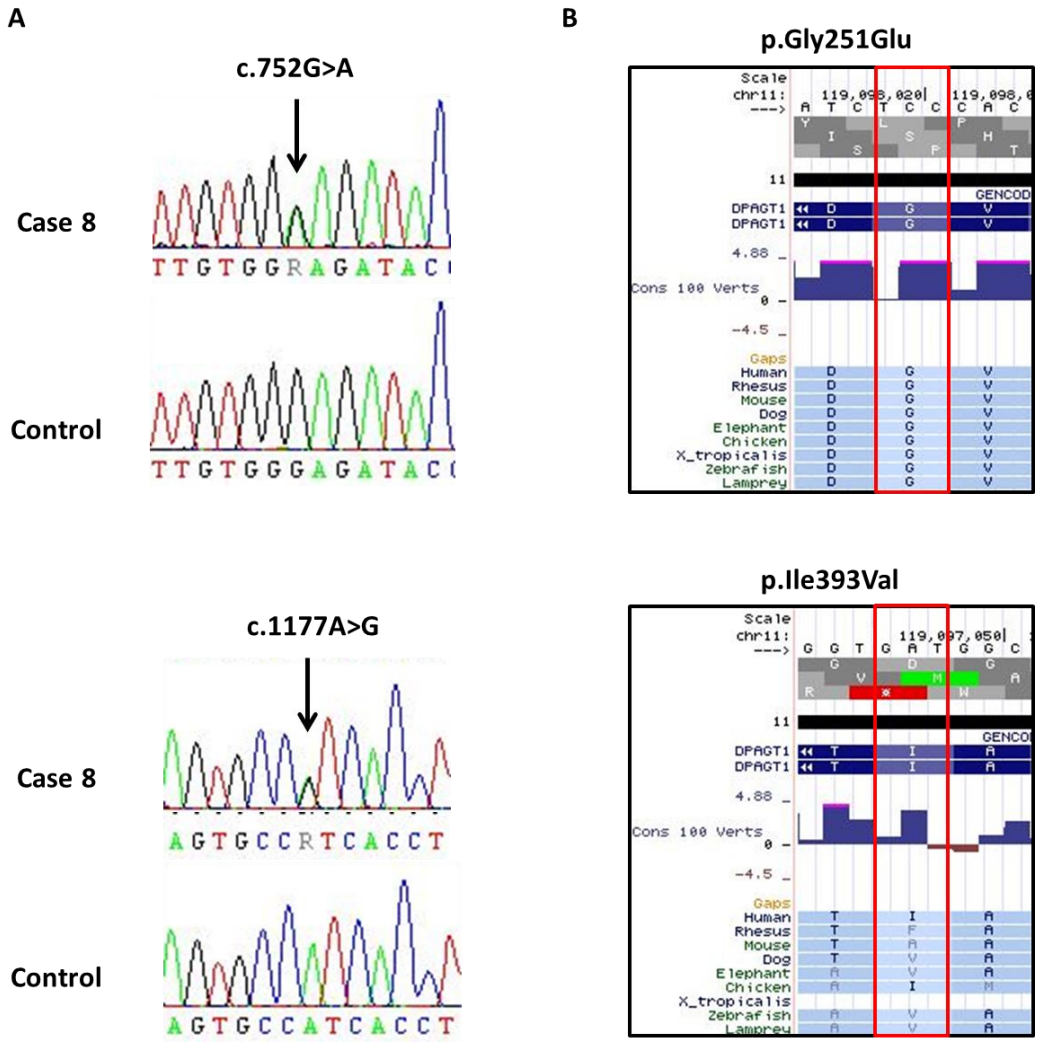


Figure 5-12 Missense variants in the *DPAGT1* gene

(A) Electropherograms of DNA for the heterozygous c.752G>A and c.1177A>G found in Case 8; (B) Conservation of the Glycine at position 251 and Isoleucine at position 393.

DPAGT1 encoded dolichyl-phosphate N-acetylglucosaminephosphotransferase 1 is an essential enzyme for N-linked protein glycosylation. Mutations in *DPAGT1* are

associated with an autosomal recessive form of CMS (Finlayson et al., 2013). Although compound heterozygous missense variants were found in Case 8, the p.Ile393Val is likely to be a polymorphism, which suggests that the patient is a carrier of a novel, likely pathogenic variant in *DPAGT1*. In addition, the mother of the patient had similar cramps suggesting an autosomal dominant form of inheritance; therefore, the variants found in *DPAGT1* are unlikely to contribute to the disease. Other muscle genes should be considered in future investigation, and the DNA sample of the mother should also be requested if possible.

5.3.3 Candidate Genes in This Cohort

5.3.3.1 ALG14

Variants in a candidate gene *ALG14* were found in two cases of the cohort. Case 9 and Case 10 are two unrelated cases who were seen by neurologists at NHNN. Case 9 first developed symptoms in his early 20s. The main manifestations were generalised tonic-clonic seizures, impaired hearing and bilateral facial weakness. He had a positive family history. His mother also had a history of epilepsy since sustaining a head injury at the age of five. She also had a history of classical migraine associated with vertigo. In addition, two other family members also had hearing difficulties. Case 10 first presented at the age of 19. He was diagnosed with possible myasthenia gravis. He did not have family members reporting similar problems. Features of TAs were seen in both their muscle biopsies. The biopsy images of both cases were not available by the time of writing. The summary of their clinical phenotypes is provided in Table 5-13.

Table 5-13 Clinical phenotypes of Case 9 and Case 10

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK Level	EMG/NCS	Muscle Biopsy
Case 9	M	Early 20s	Y	Grand mal seizures, triggered by tilting his head back to the right or turning to the right; these occurred once every one or three weeks; probably inherited cochlear hearing loss	Bilateral facial weakness; posterior fossa atrophy; slow learner; slightly dysarthic; arachnodactyly and Marfanoid appearance; mentally dull; mild gait ataxia; mild weakness of orbicularis oculi	Fall with a head injury aged 35 with skull fracture and transient left facial palsy; longstanding recurrent left shoulder dislocation	Normal	Normal study without evidence of a large fibre neuropathy or myopathy	IHC: suggestive of TAs but no other abnormalities; EM: about 5% of fibres with TAs; they were associated with areas of glycogen rich sarcoplasm; no evidence of mitochondrial abnormalities
Case 10	M	19	N	Possible myasthenia gravis; progressive muscle weakness in the legs, impaired gait with foot-drop and trunkal weakness; dysphagia	Intermittent and variable diplopia; mild neuropathy with severe involvement of lateral popliteal nerve; gross ataxia with possible functional component	Severe injury to right scapula aged 14. At 16 yrs he was loss of conscious for approximate half an hour, after which he developed imbalance, difficulty in focussing, headache and diplopia	Normal	EMG: myopathic changes in several affected muscles; NCS: normal	IHC: scattered large and small muscle fibres, internal nuclei, and presence of TAs

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; N = no; Y = yes; IHC = immunohistochemistry; EM = electron microscopy; TAs = tubular aggregates.

WES was performed on both cases. Compound heterozygous stop-gain mutation c.310C>T (p.Arg104Ter) and missense variant c.31G>A (p.Ala11Thr) in *ALG14* were identified in Case 9, and a same heterozygous missense variant p.Ala11Thr was also found in Case 10 (Figure 5-13). Both variants have frequency below 1% in the 1000 Genomes project, EVS and ExAC databases. There were no other samples available from the family of Case 9 to confirm the segregation. (Table 5-14)

Table 5-14 Variants found in the *ALG14* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	Polyphen2/SIFT /MutationTaster
Case 9	p.Arg104Ter (het)	N/A; 0.000077; 0.00005774	5.39	7.34	N/A/N/A/D
	p.Ala11Thr (het)	0.0018; 0.008227; 0.006945	-0.309	N/A	B/T/B
Case 10	p.Ala11Thr (het)	0.0018; 0.008227; 0.006945	-0.309	N/A	B/T/B

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; N/A = not available; D = damaging; B = benign; T = tolerated.

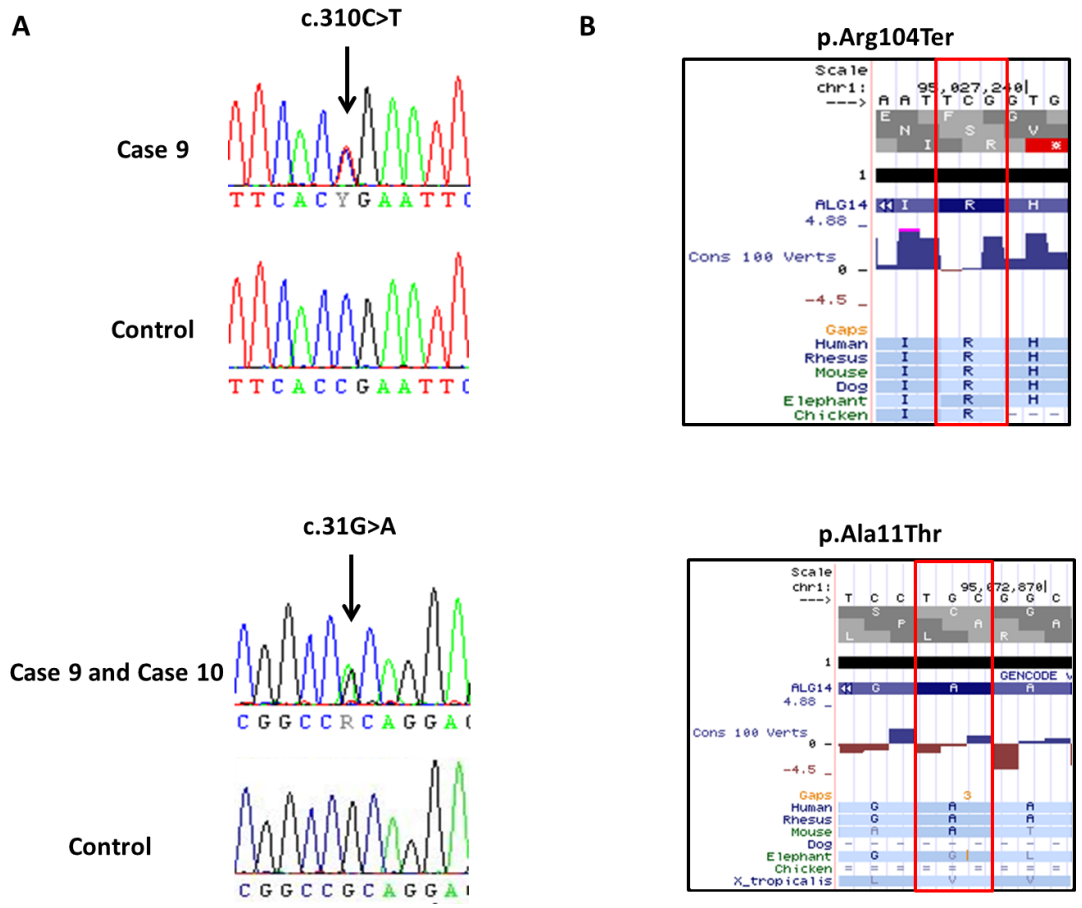


Figure 5-13 Variants in the *ALG14* gene

(A) Electropherograms of DNA for the heterozygous stop-gain mutation c.310G>C and missense variant c.31G>A found in Case 9 and Case 10; (B) Conservation of the Arginine at position 104 and Alanine at position 11.

The heterozygous variant c.310C>T was previously reported to be pathogenic in a family of two siblings with a diagnosis of probable congenital myasthenic syndrome and without TAs in muscle biopsies (Cossins et al., 2013). The variant introduces a premature stop codon after amino acid residue Arg104, resulting in a truncated protein by 112 amino acid residues. The second variant c.31G>A causes a non-synonymous substitution Ala11Thr, and this is predicted as benign by both SIFT and PolyPhen2. To assess whether this substitution affects protein expression, the wild-type and the mutant p.Ala11Thr were transfected into human embryonic kidney 293 (HEK293) cells, and this followed by western blot analysis to visualize the expression of protein (Figure 5-14 A). The p.Ala11Thr variant severely reduced the expression of ALG14, suggesting that it is likely to be pathogenic (Figure 5-14 B). The functional work was carried out in collaboration with Professor David Beeson in Oxford.

ALG14, similarly to *ALG2* and *DPAGT1*, encodes a protein involved in the early steps of N-linked glycosylation. Studies in yeast have shown that ALG14 is a membrane protein and plays an essential role in recruiting ALG13 to the endoplasmic reticulum and form a heterodimer. Together with ALG13, ALG14 also interacts with ALG7 (the human orthologue is DPAGT1) to form a multiglycosyltransferase. It catalyses the first two steps of the biosynthesis of lipid-linked oligosaccharide (LLO) precursor for N-glycan assembly. A severe truncation of ALG14 caused by p.Arg104Ter would result in failure in the recruitment of ALG13 to the endoplasmic reticulum. The other mutation p.Ala11Thr resulted in markedly reduced expression of ALG14 in HEK293 cells. It was not possible to analyse the number of endplate acetylcholine receptors in this study. However, knockdown of ALG14 cell culture models have showed a reduced expression of cell-surface acetylcholine receptor (Cossins et al., 2013). This results in a reduced synaptic response to acetylcholine at the neuromuscular junctions, which causes the phenotypes of CMS. In addition to the symptoms in muscle, defects in N-linked glycosylation pathway may affect many different organs including the nervous system and the intestines (Freeze, 2006). Although the molecular mechanism underlying this has yet to be determined, mutations in *DPAGT1*, *ALG2*, or *GMPPB* have been reported in patients affected with severe multisystem disorders (Thiel et al., 2003, Carss et al., 2013, Wu et al., 2003), also called congenital disorders of glycosylation (CDG). Case 9 presented with generalised seizures, hearing loss and bilateral facial weakness, which is likely due to the combination of the two severely deleterious mutations in *ALG14* leading to a complex multisystem disorder. The other patient, Case 10, with a heterozygous p.Ala11Thr had a more myasthenic presentation. Due to the recessive inheritance of glycosylation related CMS, it is likely that there is another variant on the second allele which has not been picked up by exome sequencing. No variant was found in other protein coding regions and splicing regions of *ALG14* by Sanger sequencing. However, a variant in the promotor or 3'UTR regions, or a large-scale deletion in non-coding regions still cannot be excluded.

Three mutations (p.Arg104Ter, p.Pro65Leu, and p.Ala11Thr) in *ALG14* have been found in two studies so far including our study. In the first study, the patients were two siblings with a clinical diagnosis of probable CMS without TAs in muscle

biopsies (Cossins et al., 2013). Our study was the first time that *ALG14* mutations were identified in patients with TAs, and the patient Case 9 presented with symptoms of a complex multisystem disorder. Therefore, our findings expand the genetic and phenotypic spectrum of *ALG14* mutations.

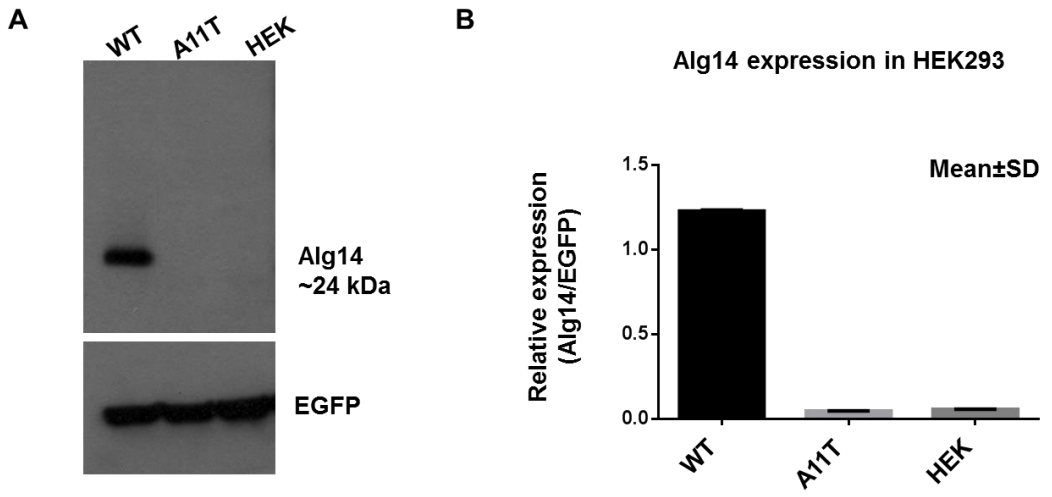


Figure 5-14 Functional analysis of the variant p.Ala11Thr in the *ALG14* gene

HEK293 cells were seeded at $3x10^5$ per well in 6-well plates and transfected with 3 µg pDNA3.1/hygro(+) HA-tagged wild type and mutant (Ala11Thr, A11T) Alg14 constructs using polyethyleneimine. 48 hours after transfection, cells were harvested and lysed. Whole cell lysates were subjected to western blotting. (A) The expression of HA-tagged wild type and mutant Alg14 were detected using mouse monoclonal anti-HA antibody (ab18181, Abcam), HRP-conjugated anti-mouse secondary antibody (Dako) and ECL (GM Healthcare). Transfection efficiency was verified by co-transfection of EGFP. This represents an example of three experiments carried out. Densitometry of bands was analysed using ImageJ software and the protein expression indicated as Alg14:EGFP was then quantitated (B). The level of Alg14 expression was the mean value of three experiments, and the error bar represents the standard deviation. EGFP = enhanced green fluorescent protein; HA = human influenza hemagglutinin; HRP = horseradish peroxidase; WT = wild type. (This work was carried out in collaboration with Professor David Beeson's lab in Oxford, and the experiment and the data analysis were kindly performed by Wei-Wei Liu.)

5.3.3.2 CASQ1

A variant in the *CASQ1* gene was found in one patient of the cohort. The patient (Case 11) was previously seen by neurologists at NHNN. He had a diagnosis of congenital scapuloperoneal syndrome. Appearance of TAs was reported in his muscle biopsies. The biopsy images were not available by the time of writing. Details of the phenotypic features are provided in Table 5-15.

Table 5-15 Clinical phenotypes of Cases 11

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK level	EMG/NCS	Muscle Biopsy
Case 11	M	Infancy (floppy baby at birth)	N	Congenital scapular peroneal syndrome; marked scapular weakness, more prominent weakness and wasting of his anterior and posterior calf muscles, marked weakness in knee flexion and extension and preservation of deltoid; no fatigability	Cervical radiculopathy	Right cervical radiculopathy, decompression in 2007, and in 2012; Left cervical radiculopathy, decompression in 2009. Genetic test on FSHD gene negative	621 IU/L after 5 minutes exercise	EMG: neurogenic changes in C5/C7 and left C5/C6 nerve root lesion (2009); no SFEMG abnormality in right EDL and right orbicularis occuli (2010); active denervation in C5/C6 on a background of chronic partial denervation (2012)	IHC: suspicious of mitochondrial proliferation; considerable number of TAs; dystrophy panel normal

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; N = no; Y = yes; IHC = immunohistochemistry; N/A = not available; FSHD = facioscapulohumeral muscular dystrophy; SFEMG = single-fibre EMG; EDL = extensor digitorum longus; TAs = tubular aggregates.

WES identified a rare heterozygous missense variant c.1154T>C (p.Ile385Thr) in *CASQ1* in Case 11. This was predicted as deleterious by both PolyPhen and MutationTaster (Table 5-16 and Figure 5-15). However, no DNA samples from his family members were available for the segregation analysis.

Table 5-16 Variant found in the *CASQ1* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	Polyphen2/SIFT /MutationTaster
Case 11	p.Ile385Thr (het)	N/A; 0.000154; 0.0001193	4.05	N/A	D/T/D

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; N/A = not available; D = damaging; T = tolerated.

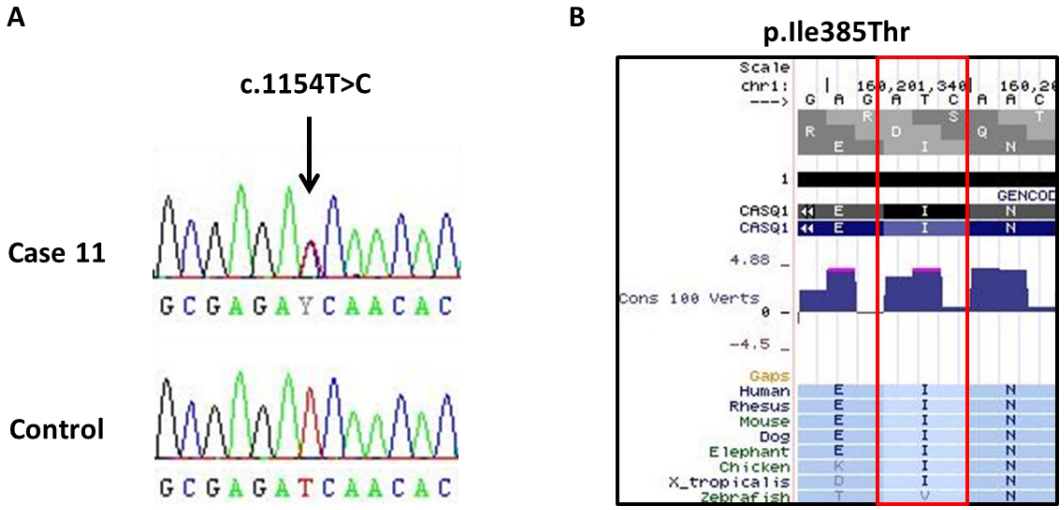


Figure 5-15 Missense variant in the *CASQ1* gene

(A) Electropherograms of DNA for the heterozygous missense variant c.1154T>C found in Case 11; (B) Conservation of the Isoleucine at position 385.

CASQ1 encoding calsequestrin 1 is a high-capacity and moderate-affinity calcium-binding protein predominantly located at terminal cisternae and expressed in fast skeletal muscle fibres. It acts as the main calcium buffer of the sarcoplasmic reticulum (Fujii et al., 1990). A heterozygous missense mutation p.Asp244Gly, the first mutation in *CASQ1*, was identified as the cause of an autosomal dominant vacuolar myopathy with CASQ1 aggregates in four Italian families and one unrelated sporadic case (Rossi et al., 2014). The patients presented with one or more symptoms

including episodes of muscle cramping, reduced muscle strength, fatigue, and elevated plasma CK levels. The stainings of CASQ1, RYR1, and SERCA were positive in their muscle biopsies (Rossi et al., 2014), however, the appearance of TAs was not observed. In this same study, altered calcium release kinetics were observed in muscle fibres from two patients harbouring the mutation, suggesting the mutation may reduce the ability of CASQ1 to store calcium efficiently (Rossi et al., 2014). The same mutation was reported in a further two studies, but again no TAs were found (Di Blasi et al., 2015, D'Adamo et al., 2016).

The missense variant p.Ile385Thr in *CASQ1* found heterozygously in Case 11 is very rare with a MAF of 0.01% in the ExAC and the EVS databases, and is predicted as damaging by PolyPhen2 and MutationTaster. Case 11 had symptoms of congenital scapular peroneal syndrome with marked muscle weakness but no fatigue, and an elevated CK level after a short time of exercise. He was a sporadic case as no other family members reported similar phenotypes. This suggests that either there could be another recessive gene involved in his family or this may be a de novo mutation in the patient himself. The heterozygous mutation p.Asp244Gly in *CASQ1* was found in one sporadic case in the study by Rossi et al. (Rossi et al., 2014). However, unlike the mutation p.Asp244Gly, the p.Ile385Thr in *CASQ1* has not been found in other patients with similar phenotypes, and no DNA samples from his parents were available for sequencing. Therefore, it is uncertain whether the variant p.Ile385Thr could cause the disease. DNA samples from his parents should be acquired for further investigation if possible. A known gene *TRPV4* associated with congenital scapular peroneal syndrome was negative in the patient's exome. Contribution of other genes cannot be excluded.

5.3.3.3 ATP2A1

Notable variations in the *ATP2A1* gene were seen in two members of the cohort. The index patient (C-II-1, Figure 5-16) from Family C was seen by neurologists at NHNN. He has longstanding symptoms of exercise-induced myalgia since childhood. This included pain in his legs with walking, in his arms and hands when lifting heavy objects and lower back pain. There was no definite history of a second wind and taking glucose prior to exercise did not relieve the symptoms. He also had significant post-exercise fatigue and muscle pain and stiffness which lasted for several days.

There is no history of myoglobinuria. Both of his parents and the sister were not reported to have neuromuscular symptoms. One of his daughters (III-1) seemed to suffer a severe chronic regional pain, however, her detailed clinical phenotypes were unknown.

Case 12 was originally seen in Warwick Hospital but her care was subsequently transferred to NHNN and was reviewed by Professor Michael Hanna. She had a normal teenage and early adult life, but started to notice a weakness in her legs at the age of 44. She started experiencing difficulty in going upstairs, particularly tall steps. In addition, she found when she did repetitive activities with her hands, she developed weakness in griping. She also felt fatigue more easily. She did not report swallowing problems but she noticed that after talking for a long time her mouth felt heavy. She had no pain and no sensory symptoms. There was no history of diplopia, ptosis or skin rash. There was limited information available on her investigation results by the time of writing. She was diagnosed of autoimmune myasthenia gravis with positive antibodies. There was no family history of neuromuscular disease.

Clinical phenotypes of both patients were summarised in Table 5-17. Interestingly, numerous collections of CSs were noted in both patients' muscle biopsies (Figures 5-17 and 5-18). The images of the biopsies were kindly provided by Dr Estelle Healy previously from UCL MRC Centre for Neuromuscular Disease.

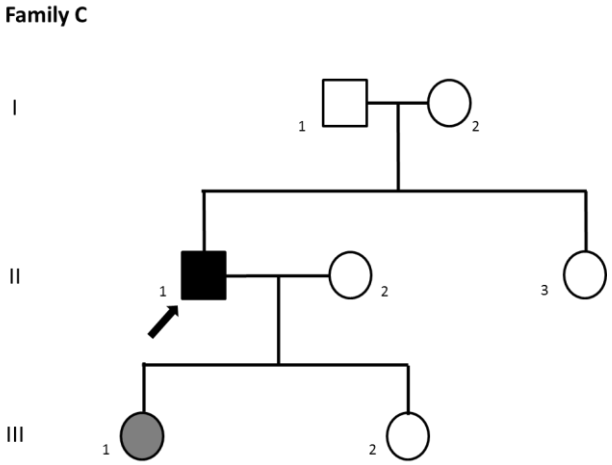


Figure 5-16 Pedigree of Family C

Black symbol represents affected patients, grey symbol represents individual affected with uncertain clinical status; Black arrow points the index patient of the family.

Table 5-17 Clinical phenotypes of patient from Family C and Case 13

Case ID	Sex	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK level	EMG/NCS	Muscle Biopsy
C-II-1	M	Child hood	Y	Exercise intolerance; post-exercise myalgia, stiffness and fatigability, no myoglobinuria	Left-sided elbow contracture and quite a tight iliotibial band; possible sleep apnoea	A history of urinary difficulties; developed back/neck pain after walking; rarely had alcohol; ex-smoker	374 IU/L	EMG and NCS normal	EM: presence of cylindrical spirals, mostly in type II fibres; excess COX deficient fibres (Figure 5-17)
Case 12	F	44	N	Autoimmune myasthenia gravis; muscle weakness particularly in proximal thighs when going upstairs, and also grip weakness on repetitive activities with her hands; fatigability	Her mouth feels heavy after prolonged speech; no diplopia, ptosis or skin rash	Deep vein thrombosis	263 IU/L	EMG revealed decrement; details not available	EM: presence of cylindrical spirals (Figure 5-18)

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; N = no; Y = yes; EM = electron microscopy; COX = cytochrome *c* oxidase.

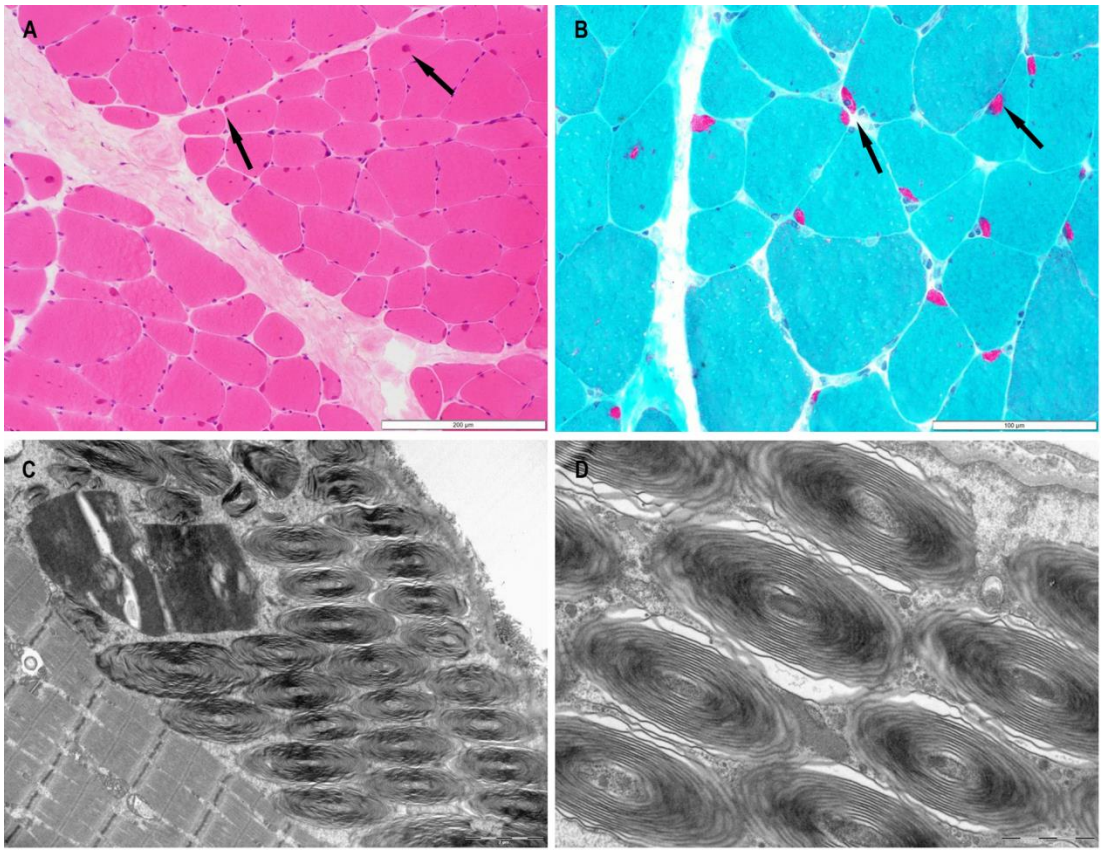


Figure 5-17 Muscle biopsy features in C-II-1

There were increased variation in fibre size and scattered atrophic fibres with internal nuclei. Some of fibres contain basophilic inclusions suggestive of CSs in H&E (A); CSs in fibres stained as red in Gomori trichrome (B); Accumulations of CSs observed in EM (C and D). Inclusions of CSs were pointed by arrows in images A and B. Scale bar represents 200 μm in A, 100 μm in B, 2 μm in C and 1 μm in D. Stainings of muscle biopsies were performed by staff at Department of Pathology, ION. The images were kindly provided by Dr Estelle Healy previously from the MRC Centre for Neuromuscular Disease, UCL.

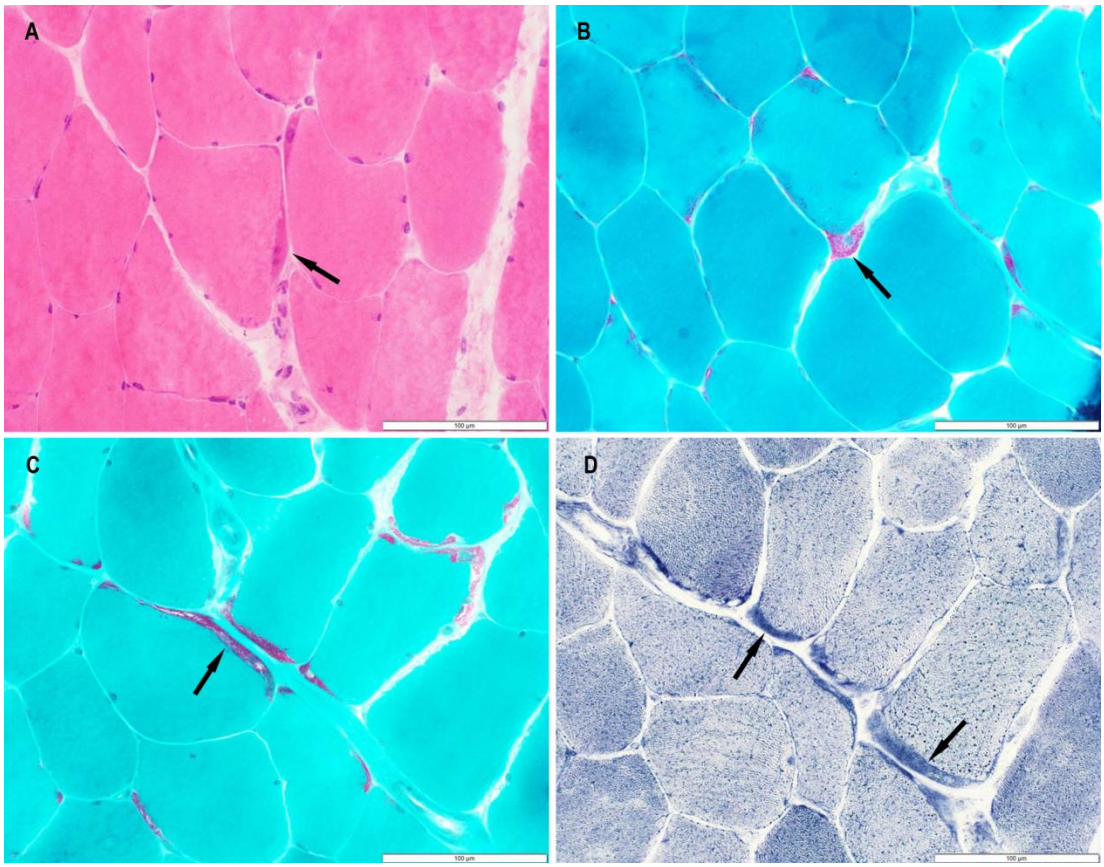


Figure 5-18 Muscle biopsy features in Case 12

Two fibres contain basophilic inclusions suggestive of CSs in H&E (A); CSs in fibres stained as red in Gomori trichrome (B and C); CSs in fibres darkly stained in myoadenylate deaminase (D). Inclusions of CSs were pointed by arrows in all the images. Scale bar represents 100 μm in A-D. Stainings of muscle biopsies were performed by staff at Department of Pathology, ION. The images were kindly provided by Dr Estelle Healy previously from the MRC Centre for Neuromuscular Disease, UCL.

DNA samples from the index patient (C-II-1) from Family C and Case 12 were sent for WES. Searching for rare variants in an autosomal recessive pattern, compound heterozygous missense and stop-gain variants c.1909C>T (p.Arg637Trp) and c.2758C>T (p.Gln920Ter) in *ATP2A1* were identified in patient C-II-1, and a homozygous missense variant c.1619C>T (p.Pro540Leu) in the same gene was found in Case 12 (Table 5-18 and Figure 5-19). Segregation analysis showed that the unaffected mother (C-I-2) of patient C-II-1 harboured a heterozygous variant p.Arg637Trp, while the unaffected sister (C-II-3) carried neither of the variants. The DNA sample of the daughter (C-III-1) was not available for the analysis. The variant p.Arg637Trp is very rare with a frequency of 8.25×10^{-6} in the ExAC database, and the Arginine residue at codon 637 is highly conserved. The p.Gln920Ter causes a

premature stop codon at position 920, resulting in a truncation of 81 amino acid residues of the protein, and this variant is novel and absent from the public databases. Both are predicted to be damaging. The p.Pro540Leu is also a rare non-synonymous variant with a frequency of 0.22% in the ExAC database, however, it is predicted to be a benign variant. (Table 5-18).

Table 5-18 Variants found in the *ATP2A1* gene

Case ID	Variant	MAF in 1000G/EVS/ExAC	GERP++	CADD	PolyPhen2/SIFT/ MutationTaster
C-II-1	p.Arg637Trp (het)	N/A; 0.000077; 0.000008132	3.59	4.45	D/D/D
	p.Gln920Ter (het)	N/A; N/A; N/A	4.26	9.10	N/A/D/D
Case 12	p.Pro540Leu (hom)	0.0046; 0.00962; 0.002684	2.53	N/A	B/T/B

MAF = minor allele frequency; 1000G = 1000 Genomes project; EVS = Exome Variants Server; ExAC = Exome Aggregation Consortium; GERP = Genomic Evolutionary Rate Profiling; CADD = Combined Annotation Dependent Depletion; het = heterozygous; hom = homozygous; N/A = not available; D = damaging; B = benign; T = tolerated.

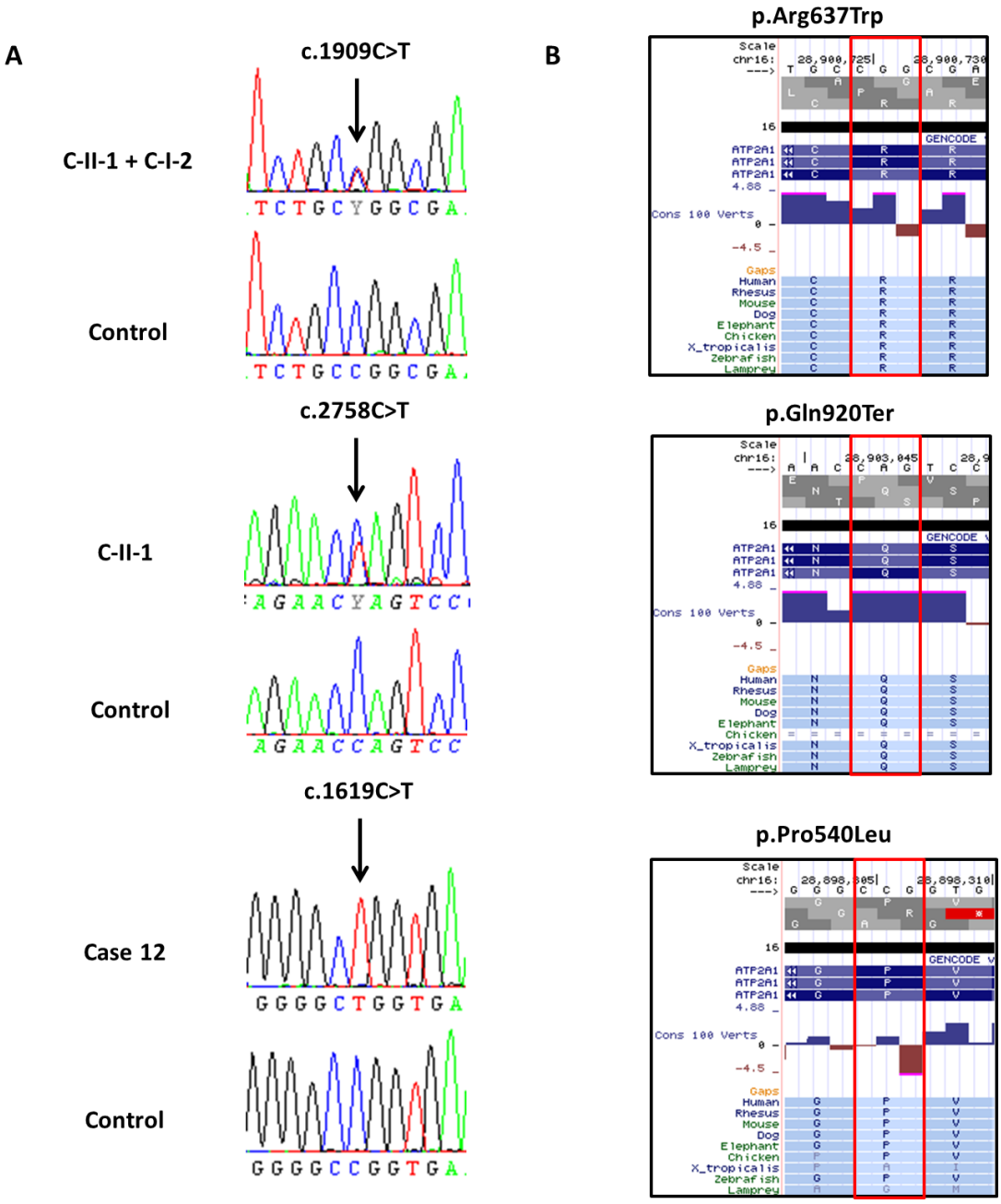


Figure 5-19 Variants in the *ATP2A1* gene

(A) Electropherograms of DNA for the compound heterozygous missense variant c.1909C>T and stop-gain variant c.2758C>T in Family C-II-1, and heterozygous missense variant c.1909C>T in Family C-I-2, and homozygous missense variant c.1619C>T in Case 12; (B) Conservation of the Arginine at position 637, Glutamine at position 920, and Proline at position 540.

ATP2A1 encodes ATPase sarcoplasmic reticulum Ca^{2+} transporting 1 (SERCA1) which is an intracellular pump located in the SR of the muscle fibres, transporting calcium ions from the cytoplasm into the SR (Periasamy and Kalyanasundaram, 2007). Mutations in this gene have been known to cause a rare inherited autosomal recessive form of Brody disease due to reduced activity of SERCA (Odermatt et al.,

1996). Brody disease is characterised by increasing impairment of muscle relaxation during exercise and silent cramps, and so far the disease has mostly been diagnosed by measurement of SERCA1 activity (Benders et al., 1994). The variants identified in our patients have not been reported in patients with Brody disease before. The index patient (C-II-1) from Family C had compound heterozygous variants in *ATP2A1* and his unaffected mother was a carrier of one heterozygous variant, which shows a recessive inherited pattern of the gene. He presented with exercise intolerance and post-exercise myalgia which is compatible with Brody disease. The measurement of SERCA activity on the muscle biopsy from the patient C-II-1 (harbouring variants p.Arg637Trp and p.Gln920Ter) was performed by our collaborators in Radboud University Nijmegen Medical Centre, the Netherlands. The compound SERCA activity was measured in whole muscle homogenates obtained from muscle biopsy as previously reported (Benders et al., 1994). It showed that the compound SERCA activity of the patient C-II-1 was at 80.2 mU/mg protein, which was slightly higher than the upper range of the control subjects at 60.0 mU/mg protein reported by Voermans et al. (Voermans et al., 2012). The western blot on SERCA protein was not performed due to insufficient muscle biopsy tissue of the patient. This very likely ruled out the diagnosis of Brody disease of this patient. The Case 12 was a sporadic case with a diagnosis of autoimmune myasthenia gravis. The patient had a homozygous missense variant p.Pro540Leu in *ATP2A1*, which also fulfils an inheritance of recessive form, and variants found in other candidate genes were all heterozygous. However, the p.Pro540Leu is predicted as a benign polymorphism, and the muscle tissue of Case 12 was not available for the functional analyses. Therefore, the relevance between the variants in *ATP2A1* and these two cases is still uncertain.

In a recent study, CSs were found highly immunoreactive for the longitudinal SR protein SERCA1 but not junctional SR proteins (Xu et al., 2015). This is the first time that rare variants in *ATP2A1* were identified in patients with CSs. Although the exact mechanisms by which these rare variants in *ATP2A1* cause disease remain unknown, they may have some contributions to the formation of this rare abnormal structure of CSs in muscle fibres instead of causing the clinical phenotypes. Of course, the possibility that other genes are responsible for the phenotypes cannot be excluded. Investigations on more cases would be necessary in the future.

5.3.4 Others Cases

There are five cases whose clinical phenotypes were not available by the time of writing this thesis, and the analysis for these cases had to be interrupted after filtering for the muscle genes. The preliminary analysis on the rest of the 15 cases did not identify a variant(s) in known or selected candidate genes for myopathies with TAs. The clinical phenotypes of these cases were summarised in Table 5-19. Four of them had positive family history and 10 were sporadic cases, however, no samples were available from their affected/unaffected family members. This brings challenges of filtering variants and identifying the possible disease-causing genes. According to the strategy suggested in Figure 5-1, the variants were filtered for MAF < 1%, non-synonymous and read depth ≥10 in the genes related to neuromuscular diseases. The filtered variants in the exomes of these 15 cases were also listed in the Appendix IV – Table 1. Acquiring detailed information of the patients and any other affected family members as well as their DNA samples will be necessary for further investigations.

Table 5-19 Summary of clinical phenotypes of other cases

Case ID	Sex	Age	AAO	FxH	Main Symptoms	Other Symptoms	Other History	CK Level	EMG/NCS	Muscle Biopsy
64412	M	91	16	Y	Exercise induced pain syndrome; muscle weakness, myalgia, and stiffness	Marfan's syndrome; glaucoma at age 45	Classical migraine with positive visual symptoms	N/A	N/A	IHC: features of TAs
40725	M	68	Early 30s	Y (sister and youngest daughter affected)	TAM with myalgia, cramps and stiffness; restless leg syndrome; myoclonic jerks	Myoadenylate deaminase deficiency; early cataracts.	N/A	N/A	EMG: increased nerved hyperexcitability repeat	IHC: features of TAs
66209	M	63	17	N	Progressive oculo-motor disorder with muscle weakness; impaired vertical eye movements and ptosis	Chorea and grimacing movement of face; psychosis; depression	N/A	N/A	N/A	EM: confirmed TAs
66203	M	61	N/A	N	TAM and hypoPP	UK	Had salt restricted diet	N/A	N/A	IHC: features of TAs
66206	M		63	N	Late onset dystonia and cervical root damage; weakness in upper limbs	Torticollis	N/A	N/A	N/A	IHC: features of TAs
64416	M	72	N/A	N	Exercise induced pain syndrome and fatigability	Depression	N/A	N/A	N/A	IHC: features of TAs in type II fibres
29970	M	69	Around 30	Y (older daughter affected)	TAM and periodic paralysis; myalgia, fatigability and muscle cramps	Statin/exercise induced acidosis	Episodic muscle weakness since hitting head on a steel girder; MI aged 45; diabetes at 50 yrs	N/A	N/A	IHC: features of TAs in type II fibres

65142	M	66	58	N	TAM and axial myopathy post statin; rapidly progressive muscle weakness, particular involvement of the paraspinal and gluteal muscles; fatigability and cramps	Occasional some blurring of vision, no diploplia	Atrial fibrillation with pacemaker; a history of statin treatment	N/A	N/A	IHC: features of TAs
63568	M (Hin-du)	66	29	N	Exercise induced pain syndrome; muscle weakness and myalgia	Chronic renal failure; hypertension; diabetes	Previous nutritional deficiencies with osteomalacia; had uveitis and retinitis thought to be Harada's syndrome, and treated with steroids at 34yrs; bilateral osteonecrosis of the femoral head and had right hip replacement aged 35	N/A	N/A	IHC: features of TAs
19817	M	65	48	N	MSA type C/ cerebellar ataxia; impaired balance, slurred speech and worsening in the quality of his handwriting	Suffered from urinary symptoms	N/A	N/A	N/A	IHC: features of TAs
64414	M	76	N/A	N/A	TAM and exercise induced pain syndrome	N/A	N/A	N/A	N/A	IHC: features of TAs
2589	M	93	12	Y	Essential tremor; mild axonal neuropathy; muscle pain syndrome	N/A	Hiatus hernia aged 44yrs; thyrotoxicosis treated with Neomercazole, complicated by thrombocytopenia aged 50; angina aged 61	N/A	N/A	IHC: features of TAs
1285	M	88	48	N	Muscle weakness and unexplained muscle pain;	Occasional tendency to choke	N/A	N/A	N/A	IHC: features of TAs

					fatigability but unlike those with myasthenia					
MB7684	M	44	One month	N	TAM; muscle weakness and myalgia	N/A	N/A	N/A	N/A	IHC: features of TAs
MB7052	M	N/A	N/A	N	TAM; difficulty in standing and walking	N/A	N/A	N/A	N/A	IHC: features of TAs
4580	M	N/A		N/A	N/A	N/A	N/A	N/A	N/A	IHC: features of TAs
4893	M	N/A		N/A	N/A	N/A	N/A	N/A	N/A	IHC: features of TAs
9469	M	58		N/A	N/A	N/A	N/A	N/A	N/A	IHC: features of TAs
3212	M	48		N/A	N/A	N/A	N/A	N/A	N/A	IHC: features of TAs
3938	M	28		N/A	N/A	N/A	N/A	N/A	N/A	IHC: features of TAs

AAO = age at onset; FxH = family history; CK = creatine kinase; EMG = electromyograms; NCS = nerve conduction studies; F = female; M = male; N = no; Y = yes; N/A = not available; UK = unknown; IHC = immunohistochemistry; MSA = multiple system atrophy; TAM = tubular aggregate myopathy; TAs = tubular aggregates; hypoPP = hypokalaemic periodic paralysis.

5.4 Discussion

Myopathies with TAs and CSs are a group of highly clinically heterogeneous diseases. So far mutations in eight genes have been reported to be responsible for cases with TAs in different inheritance patterns, while no gene has been identified in cases with CSs. This chapter investigated the genetic basis of a group of 35 index patients with TAs and CSs using WES technique with a focus on previously reported known genes and possible candidate genes.

Despite an average coverage of whole-exome data at 10x read depths of 79.1% in all samples, rare/novel variants in four known genes were successfully identified in eight unrelated patients with TAs and were suggested as pathogenic to the disease (Table 5-3). This constituted 22.9% of the cohort. These include three novel heterozygous missense variants (p.Asp84Glu, p.Leu92Val, and p.Glu255Val) in *STIM1* identified in two families (Families A and B) and one unrelated case (Case 1), two heterozygous missense variants (p.Gly98Ser is known, and p.Val107Mat is novel) in *ORAI1* identified in two unrelated sporadic cases (Cases 2 and 3), one known homozygous missense variant (p.Arg10Glu) and one known homozygous frameshift deletion (p.Gly178fs30Ter) in *PGAM2* identified in two unrelated patients (Cases 4 and 6), and a novel heterozygous missense variant (p.Val445Thr) in *SCN4A* identified in one patient (Case 7). Two patients (Cases 5 and 8) were also found with rare/novel heterozygous missense variants in known gene *PGAM2* (p.Arg10Trp) and *DPAGT1* (p.Gly251Glu), respectively. Both are autosomal recessive genes, and Sanger sequencing the whole protein coding region did not find another rare variant. Therefore, they may be just carriers for these genes. In addition to the known genes, rare variants in three candidate genes were also found in our patients. Compound heterozygous variants (p.Arg104Ter is known, and p.Ala11Thr is novel) in an N-linked glycosylation pathway related gene *ALG14* were found in one patient (Case 9). Functional analysis on p.Ala11Thr suggested it was pathogenic. The heterozygous p.Ala11Thr was also seen in a different patient (Case 10). A possible damaging heterozygous missense variant (p.Ile385Thr) in SR-related gene *CASQ1* was found in one sporadic case (Case 11). Moreover, compound heterozygous variants (p.Arg637Trp and p.Gln920Ter) and a homozygous missense variant (Pro540Leu) in another SR-related gene *ATP2A1* were found in two unrelated patients with CSs.

This study therefore expanded the knowledge surrounding the potential genetic causes of myopathies with TAs and CSs (e.g. novel variants in *STIM1* and *ORAI1* genes) as well as expanding the phenotypic spectrum associated with some of the genes discussed (e.g. *ALG* gene). Further functional analysis is important to determine the causality of these newly identified variants.

Although the exact mechanisms of how these variants cause disease and the formation of TAs/CSs remains unknown and would require further investigations, we identified pathogenic/likely pathogenic rare variants in both calcium related genes and N-linked glycosylation related genes. These genetic findings may provide further evidence for the previous hypothesis that disordered calcium homeostasis and impaired N-linked glycosylation pathway are involved in the formation of the TAs. The interaction of ORAI1 and STIM1 proteins activates the opening of CRAC and SOCE channels, which plays an essential role in calcium signalling pathway resulting in calcium influx into the cytosol (Lacruz and Feske, 2015). There have been 16 mutations and probably pathogenic variants in *STIM1* and four mutations and probably pathogenic variants in *ORAI1* found in patients with TAs including the ones from our cohort. These mutations disrupt the calcium-dependent activation/inactivation of CRAC/SOCE in gain-of-function or loss-of-function form, resulting in aberrant calcium homoeostasis (Lacruz and Feske, 2015, Okuma et al., 2016). TAs originate from the SR. It is suggested that dysregulated intraluminal or cytosolic calcium level could potentially cause swelling of SR cisternae and their extension into longitudinally oriented tubules, subsequently resulting in the formation of TAs (Bohm et al., 2013). A similar process was also proposed for the sequential formation TAs in ageing skeletal muscle fibres (Boncompagni et al., 2012). SR-related proteins involved in the uptake and storage of calcium, such as CASQ1, RYR1 and SERCA, have been observed to be components of TAs, so it is assumed that the mutations/variants in these genes may also lead to a similar pathogenic pathway to form TAs.

Another major group of genetic factors related to TAs are recessive mutations in genes encoding proteins involved in N-linked glycosylation pathway, including previously reported *GFPT1*, *DPAGT1* and *ALG2*, and *ALG14* found in our cohort. Protein N-linked glycosylation plays a critical role in a wide variety of biological

processes, such as protein folding, cellular targeting and motility, and immune response (Larkin and Imperiali, 2011). Both STIM1 and ORAI1 are glycosylated proteins. The importance of N-glycosyl modification for the function of STIM1 was reported in a recent study where a mutation at N-glycosylation site in STIM1 resulted in a strong gain-of-function effect by increasing the number of active CRAC channels (Kilch et al., 2013). Thus, it is possible that mutations in genes involved in the N-linked glycosylation pathway may cause inappropriate glycosylation of STIM1 and other proteins. This leads to abnormal protein folding and impaired calcium homeostasis, and eventually triggers the proliferation of SR tubules and formation of TAs. Further analyses of CRAC channel activation in skeletal muscles of patients with mutations in this group of genes would be required in the future.

The pathogenesis of *PGAM2*- and *SCN4A*-related TAs is still a mystery. However, interestingly, two patients with PGAM deficiency and TAs were observed with markedly increased calcium concentration and calcium adenine triphosphate activity in muscles (Naini et al., 2009). It is speculated that abnormal calcium homeostasis is likely to be a common final process leading to the formation of TAs, which is induced by diverse conditions. These may include defects in calcium signalling pathway and N-linked glycosylation pathway, PGAM deficiency, probably also exposure to drugs, toxins, and hypoxia. Meanwhile, there may be other mechanisms underlying the pathogenesis of TAs given that in more than half of the cases in our cohort we did not identify rare genetic variants in the pathways discussed above. However, all these data further confirm that TAs are a gradually changing SR architecture and also an unspecific pathological change as an adaptive response of the SR to various insults to the muscle fibres.

CSs also originate from the SR, and their formation has been considered underlying a similar hypothesis to the formation of TAs. However, pathological studies suggested that CSs originate from the longitudinal SR, which differs from TAs as their derivation is from the whole SR (Xu et al., 2015). In addition, genetic findings of recessive variants in *ATP2A1* encoding the longitudinal SR protein SERCA1 in our two cases with CSs may also indicate that CSs are likely a distinct pathological change to the specific defect in the longitudinal SR protein. This is the first attempt at identifying potential genetic factors in cases with CSs, thus more cases are

required for the pathological, genetic, and functional investigations to prove the hypothesis and reveal the mystery of the formation of CSs.

The findings of this study showed that WES is a good tool for genetic diagnosis of patients with TAs and CSs. As the cost of WES keeps reducing, it is a cheaper and faster method, and allows for a small amount of DNA for analysis and potential novel findings when studying a cohort of patients with genetic and clinical heterogeneity compared with Sanger sequencing. In addition, it allows for patients to be investigated in parallel to look for shared mechanism. However, despite our genetic findings of known genes and candidate genes in this cohort of cases with TAs and CSs, there are some additional difficulties in determining the disease-causing genes for the patients. The first one is the technical problems induced by either quality of DNA samples or quality of whole-exome sequencing chemicals. In this study, many samples were sequenced during a change from Illumina TruSeq kit to Nextera kit in our department, and unfortunately the chemicals in both versions of kits were in relatively poor quality. Many other colleagues suffered with the same issues of low coverage of their samples during that period. If a causative gene locates in the region which is poorly covered, or even uncovered, this will be impossible to provide a genetic diagnosis for the patient. Re-sequencing can be considered if the coverage is very poor and no potential causative gene is found. The second one is the lack of clinical information, and the unavailability of samples of other family members. There are no thorough clinical records of some of the patients, especially those provided by the two biobanks in Italy. Some patients could be traced back to many decades ago, and the follow-up of the disease development had been lost. Some family members of the patients were reluctant to partake in further investigations. Without detailed clinical information and access to samples of other family members, it is difficult to interpret the variants found in the patients. In this case, it could not be determined whether the genetic findings are correlated with the phenotypes, and whether they are segregating in the family. Thirdly, there may be other phenotypes correlated with the genes, which is out of our knowledge. It would not be easy to interpret the pathogenicity of the gene if the patients with mutations have novel clinical features. In addition, the high number of rare variants in each patient is another challenge for data interpretation. Each patient usually has about 4-22 rare variants left after the hard filtering for neuromuscular related genes.

Therefore, looking for variants in known genes and candidate genes was our priority at the first stage of the analysis. Furthermore, there would also be a risk of false positive caused by misleading of similar features with other diseases. Therefore, re-analysis would be required after more new genes or phenotypes are reported. Lastly, some patients' phenotypes may not be caused by genetic defects, which would lead an unfruitful search in their exome data.

The preliminary analysis focused on eight known genes and eight selected candidate genes. This left a large amount of data provided by WES unexplored. Initial analyses were also mainly looking for the pathogenic mutations in a specific gene in each patient, while rare variants in genes within the same pathway could have had a cumulative effect in causing the disease. In addition, some of the patients had complex phenotypes which may attribute multiple genetic factors. Therefore, it is possible to analyse other neuromuscular related genes or even outside this range to potentially reveal novel genes. However, such analyses were beyond the time constraints of this thesis.

Overall, in this chapter, WES was successfully applied in a cohort of patients with myopathies with TAs and CSs, and the results from this study expanded the phenotypic and genotypic spectrums of myopathies with TAs and CSs. Despite the issues and challenges raised in the data interpretation, WES is still the most efficient approach to help provide genetic diagnosis for patients with TAs and CSs, as this group is both clinically and genetically heterogeneous. In addition to the significance of the thorough clinical information of the patients and the availability of samples from other family members, it is also important to accompany these with pathological findings and further functional work to confirm the genetic findings.

Chapter 6 Whole-Exome Sequencing to Identify Disease-Causing Genes in Four Families with Other Rare Myopathies

In this Chapter, the aim is to investigate the genetic basis in four families with rare myopathies and protein aggregates on muscle biopsies using whole-exome sequencing (WES). The clinical phenotypes and the available muscle biopsies of the patients were reviewed. Most muscle biopsy slides were stained and obtained from the Department of Pathology, ION. The muscle biopsies slides of Family F were stained by Dr Boel De Paepe from Gent Hospital, Belgium, and were kindly provided to us for images. All the images were kindly taken by Professor Janice Holton from the Department of Pathology, ION.

6.1 Family D with Myofibrillar Myopathy

6.1.1 Clinical and Pathological Details of the Two Affected Siblings

Two siblings from a consanguineous Italian family (Family D) were seen previously by neurologists at NHNN. The index patient (D-II-3, Figure 6-1) presented with progressive muscle weakness in his late 40s, while the other affected sibling (D-II-1) gradually developed muscle weakness and difficulty in standing or ascending stairs from 68 years old of age. The index patient was also diagnosed with dilated cardiomyopathy and atrio-ventricular dissociation. Both siblings had cytoplasmic bodies but no inflammatory infiltrates in their muscle biopsies. Rimmed vacuoles were seen in fibres from D-II-3 but not from D-II-1. Myotilin, desmin and ubiquitin positive inclusions were observed in D-II-1, but these stains were not available in D-II-3. These were suggestive of a type of myofibrillar myopathy (MFM) (Figure 6-2) There was no family history of muscular problems. Details of other siblings in this

family were unknown. Of note, DNA samples of both patients were extracted for storage, and to our knowledge no genetic tests had been carried out before.

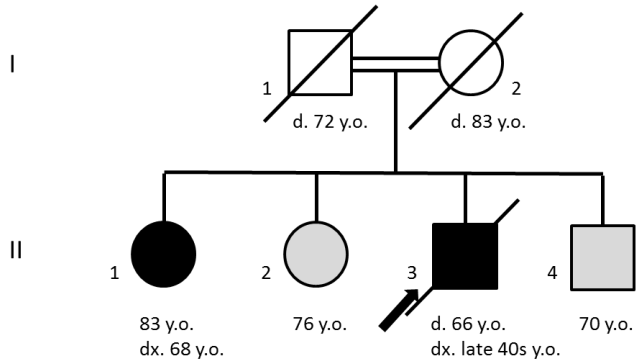


Figure 6-1 Pedigree of the Family D with two siblings with cytoplasmic body myopathy

Black symbols represent affected patients, light grey symbols represent individuals with unknown clinical status; Black arrow points the index patient of the family; Symbols with a slash indicate a deceased individual.

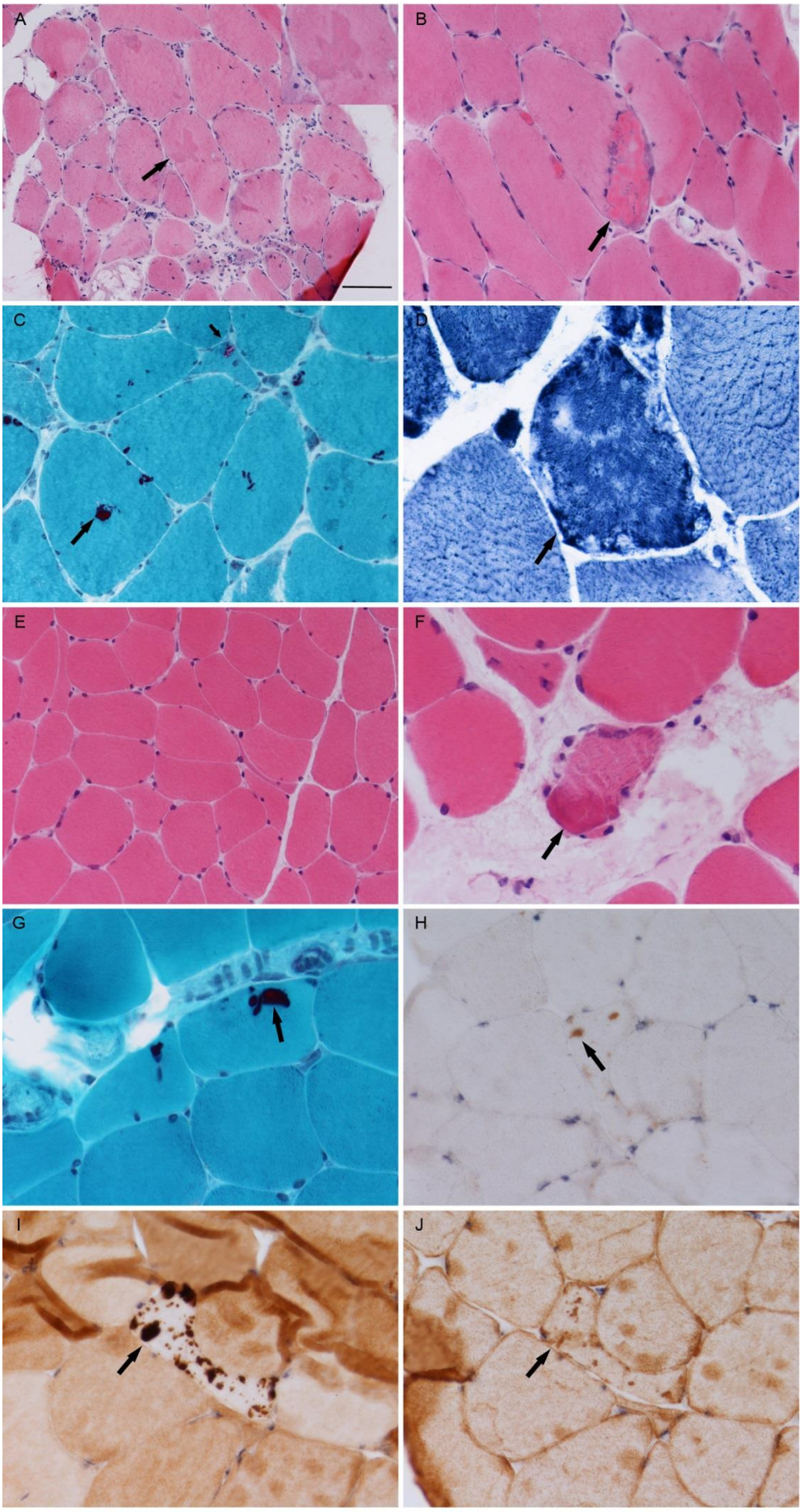

A
B
C
D
E
F
G
H
I
J

Figure 6-2 Biopsy features of two affected siblings from Family D

A-D were from the index patient D-II-3: in H&E staining, there was variation in fibre size with increased internal nuclei and abnormal internal structure (arrow and top right corner inset in A) and abnormal protein aggregates stained eosinophilic (B); Gomori trichrome staining (C) showed cytoplasmic bodies in dark (big arrow) and a rimmed vacuole (small arrow); nicotinamide adenine dinucleotide dehydrogenase-tetrazolium reductase (NADH-TR) stain showed a fibre with disturbed internal architecture with a moth-eaten appearance (D). E-J were from the sister (D-II-1) of the index patient: in H&E there was mild variation in fibre size (E) and one fibre with eosinophilic abnormal internal structure (F); there were two cytoplasmic bodies in fibres in Gomori trichrome staining (G); H-J showed ubiquitin-positive (H), myotilin-positive (I) and desmin-positive (J) inclusions in the same fibre; these were more predominant in myotilin staining. Abnormal protein aggregates were pointed by arrows. Scale bar represents 100 μm in A; 50 μm in the inset in A and B, C and E; 25 μm in D and F-J. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. The biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

6.1.2 Genetic Investigations on Family D

WES was performed on the DNA sample of the index patient (D-II-3) using Nextera Rapid Capture Exome protocol (Illumina). The metrics of whole-exome data is shown in Table 6-1. The filtering strategy for non-synonymous variants with a MAF < 1% was used as described in Section 2.3.4. As the patients were from a consanguineous family, it fulfilled a pattern of autosomal recessive inheritance. The variant search focused on the genes with compound heterozygous or homozygous variants. In addition, the muscle biopsy was compatible with a kind of MFM. Therefore, genes associated with MFM, such as *CRYAB*, *DES*, *LDB3*, *FLNC*, *BAG3*, and *MYOT*, were considered as candidate genes. This left a homozygous missense variant c.179C>T (p.Ser60Phe) in the exon 2 of the myotilin (*MYOT*) gene (OMIM#604103), and no variant in other genes was found. This variant is absent in the 1000 Genomes project database, and with a MAF of 0.008% in the EVS and of 0.004% in the ExAC database. It was subsequently confirmed in the patient, and its segregation was also confirmed in the other affected sibling (D-II-1) by Sanger sequencing. DNA samples from parents and other family members were not available for screening.

Table 6-1 Whole-exome sequencing metrics for patient D-II-3

Total Reads	Mean Target Coverage	% Target Covered at 2x	% Target Covered at 10x	% Target Covered at 20x	Total Number of Variants
84691420	49.3x	97.6%	88.6%	76.3%	21980

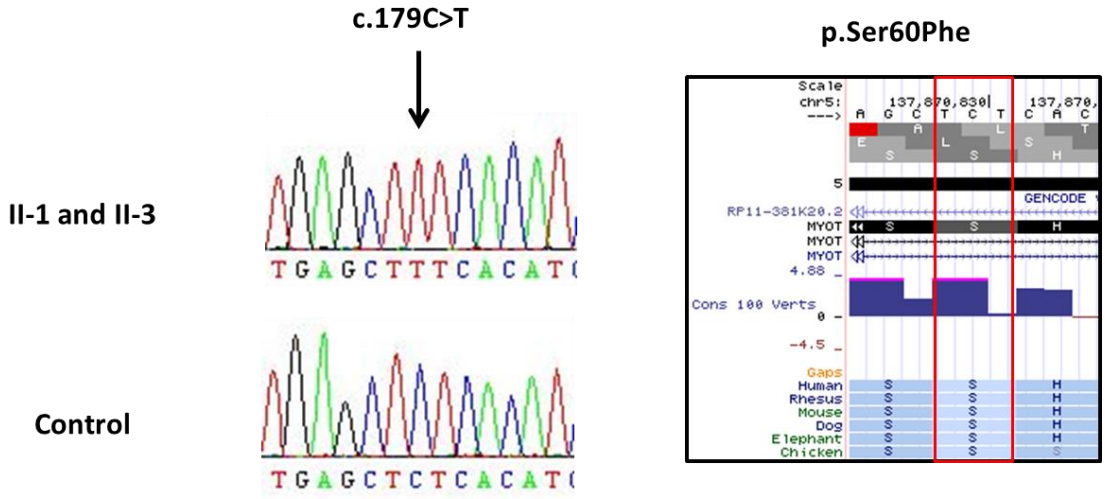


Figure 6-3 Mutation in the *MYOT* gene

This figure shows electropherograms of DNA for the homozygous missense mutation c.179C>T, and conservation of Serine at position 60.

6.1.3 Discussion

This variant is a known mutation previously reported as heterozygous in a patient with MFM. The patient also had a cardiac involvement, but had no family history (Selcen and Engel, 2004). The *MYOT* gene encodes a key Z-disk component, myotilin protein, which is expressed strongly in skeletal muscle. Myotilin binds to the main component of Z filaments, α-actinin, and cross-links actin filaments, which plays a significant role in the stability of thin filaments during muscle contraction. Myotilin consists of a unique serine-rich domain at the N-terminus and two immunoglobulin-like (Ig-like) domains at the C-terminus (Salmikangas et al., 1999, Bang et al., 2001, Salmikangas et al., 2003). Mutations in *MYOT* have been associated with autosomal dominant form of limb-girdle muscular dystrophy (OMIM#159000) (Hauser et al., 2000), MFM3 (OMIM#609200) (Selcen and Engel, 2004, Selcen, 2008), spheroid body myopathy (OMIM#182920) (Foroud et al., 2005) and distal myopathy (Berciano et al., 2008). Apart from one mutation which was found in exon 9 in a Japanese patient (Shalaby et al., 2009), the other previous

reported *MYOT* mutations were located in the serine-rich exon 2 suggesting this is a hotspot of mutations in this gene. The p.Ser60Phe mutation may affect the hydrophobic stretch of myotilin and thus alter its membrane association. The study had failed to detect substantive localization of myotilin to the sarcolemma, and the effect of this mutation on the Z-disk is obscure (Selcen and Engel, 2004). A study expressed mutant p.Ser60Cys, a different residue substitution in codon 60, in mice and found the mutant myotilins were prone to aggregate in skeletal muscles. Other Z-disk associated proteins such as BAG3 actin, desmin and filamin C were also observed, which is similar to the observations in patients. In addition, myofibril disorganization with a disrupted Z-disk was revealed in the mutant expressing muscles by EM (Keduka et al., 2012). Although the exact role of the p.Ser60Phe to the disease still remains under investigation, it may have a similar mechanism as p.Ser60Cys.

In muscle biopsies, cytoplasmic bodies were particular features seen in both patients. The presence of cytoplasmic bodies and myotilin and desmin positive inclusions in fibres are compatible with those observed in MFM muscle. Therefore, their pathological features together with our genetic findings allow us to believe that these two siblings have a type of MFM3 caused by the mutation in *MYOT*. The ages of onset in our two patients (II-3 at late 40s and II-1 at 68 years old) were earlier than the previous patient (in his 80s). This is probably due to the influence of homozygous mutations on the development of the symptoms. Their parents did not develop muscle problems as carriers with a heterozygous mutation, which might be because they died of other problems (father died at 72 years old and mother died at 83 years old) before their muscular symptoms could develop.

6.2 Two Families with Hereditary Inclusion Body Myopathy

6.2.1 Clinical and Pathological Details of the Affected Members in Two Families

Family E

Two siblings from Family E (E-II-2 and E-II-4, Figure 6-4) were seen by neurologists at NHNN. The index patient (E-II-2) had progressive predominantly

distal wasting and weakness, mainly affecting lower limbs, since his late 40s. The muscle biopsy showed myopathic muscle with type I fibre predominance, small areas of fibre necrosis and regeneration with rimmed vacuoles, and no significant inflammation or ragged red fibres, but with p62 positive inclusions in fibres (Figure 6-5). The patient was from a non-consanguineous family of nine siblings. The elder brother (E-II-1) and a younger sister (E-II-4) both developed similar phenotypes in their 50s. E-II-4 had progressive distal weakness in both upper and lower limbs and also had facial muscle weakness. Her muscle biopsy was reported to suggest changes of neurogenic atrophy (biopsy not available for review). Both E-II-2 and E-II-4 had a diagnosis of probable hereditary inclusion body myopathy (hIBM) as shown in their clinical notes. Genetic investigations had been previously performed on both E-II-2 and E-II-4 for CAG repeat expansion in the *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, and *ATXN7* genes, and also for the *DRPLA*, *CX32*, *TTR* genes and common mtDNA mutations A3243G, A8344G, and T8993C/G. They were all negative for both patients. The three affected siblings were all deceased by the time of this study. The daughter of the index patient was known to also be affected and developed difficulty in walking at 23 years old, but refused further investigation. Other members of the family had not been reported to have muscular problems and DNA samples were not available from them.

Family F

Family F was seen by Professor Jan De Bleecker and his colleagues in Gent Hospital, Belgium. DNA sample from the index patient (F-II-1, Figure 6-4) was kindly provided for genetic investigation. The index patient is currently 64 years old and started to have symptoms from 57 years old. He had progressive muscle weakness, particularly in finger flexor and knee extensor, and weakness also in wrist extension and more pronounced in wrist flexion. Muscle biopsy demonstrated moderate inflammatory infiltrates with MHC Class I positive fibres. There were multiple vacuoles but they were not rimmed. P62 positive inclusions were also seen in the fibres (Figure 6-6). His EMG showed a myopathic change. The patient fulfilled the diagnostic criteria of IBM, and was initially included in our sIBM cohort. After seeing the patient in a recent clinic, a positive family history was noted that both his father and sister were affected (Figure 6-4). This suggests a dominant form of inheritance, therefore, the patient was excluded from the sIBM cohort.

The father (F-I-1) had slowly progressive (for at least seven years) motor neuron disease (peripheral more than central). He developed frontal release signs but no signs of dementia. Muscle biopsy at the age of 75 showed chronic neurogenic atrophy with prominent fibre type grouping and no inflammation. He was deceased by the time of reviewing his clinical information. The sister (F-II-2) had slightly asymmetric proximal and distal weakness in her legs, and milder proximal weakness in the arms. Tendon reflexes were normal. She had no cognitive or behavioural problems. Muscle biopsy showed groups of atrophic fibres without fibre type grouping, but with increased internal nuclei and some split fibres. There was endomysial inflammation possibly invading a few non-necrotic muscle fibres, as well as many rimmed vacuoles. Areas of myofibrillar degeneration were irregular and devoid of oxidative activity. The biopsies of the father and the sister were reviewed by Professor Jan De Bleecker in Gent Hospital, and images of their biopsies were not available by the time of writing. Whether any genetic test was done for the father and the sister was unknown. Professor Jan De Bleecker will see the sister himself in her next appointment for detailed information and also obtain blood samples from her for segregation analysis.

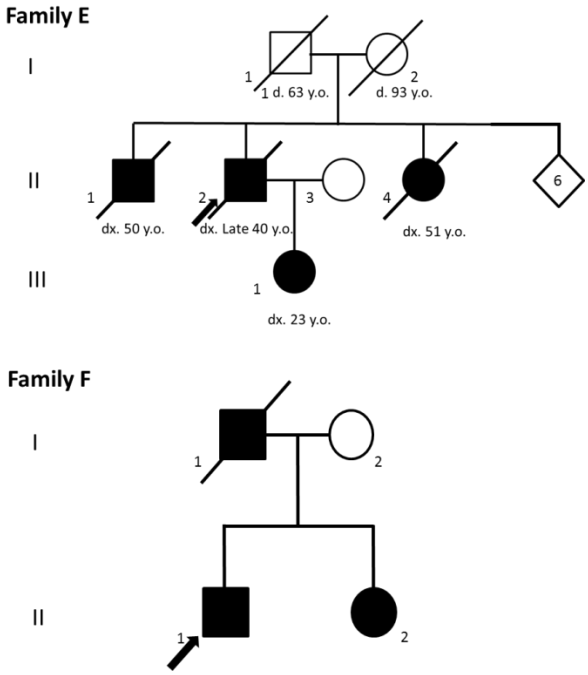


Figure 6-4 Pedigrees of Family E and Family F with hereditary inclusion body myopathy

Black symbols represent affected patients; Black arrow points the index patient of the family; Symbols with slash indicate the individuals were dead; Diamond with a number inside indicates the number of individuals with unknown gender.

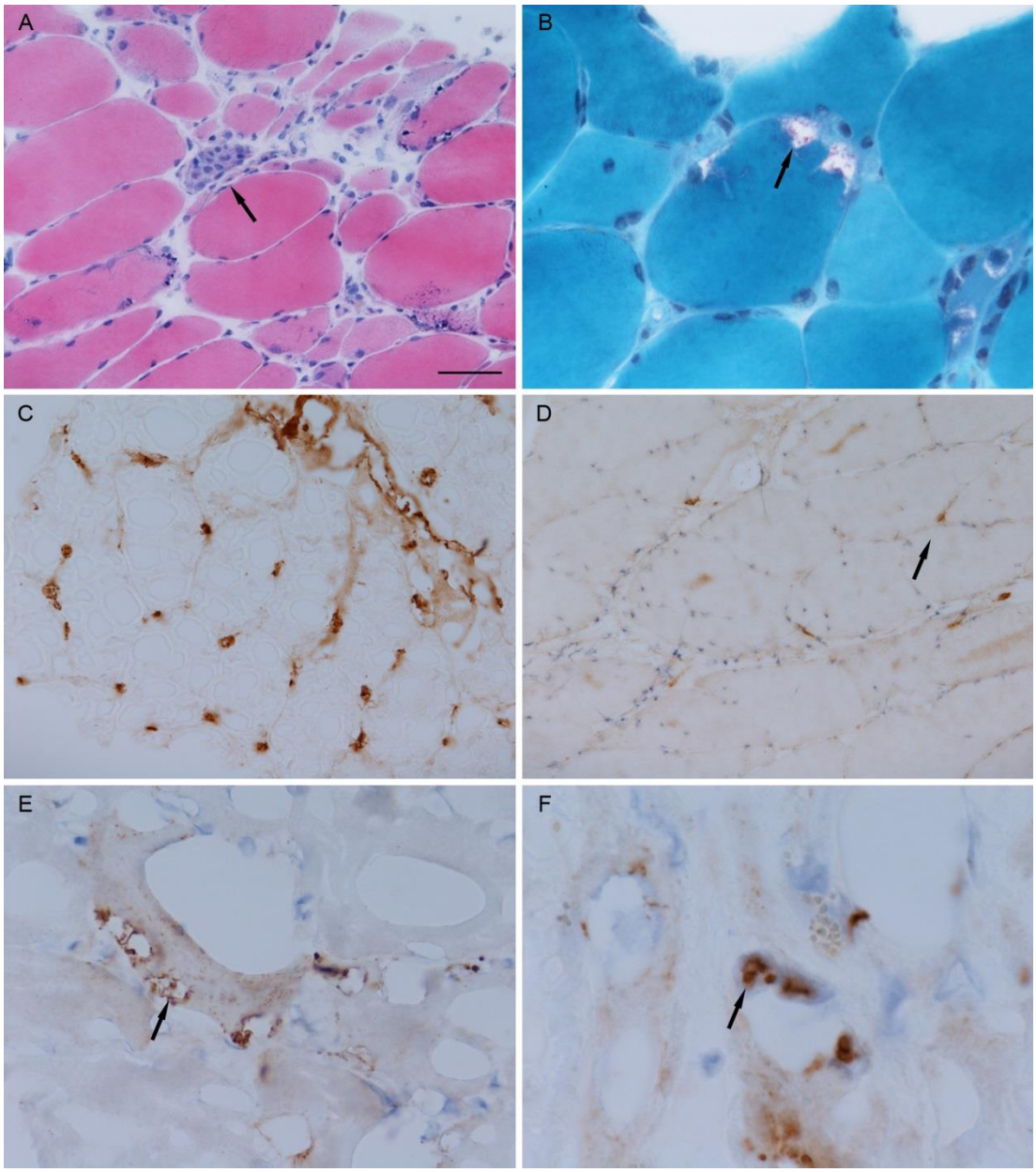


Figure 6-5 Biopsy features from the index patient II-2 from Family E

There was a moderate variation in fibre size with mild endomysial inflammation and fibre necrosis (A, H&E), and rimmed vacuoles (B, Gomori trichrome); MHC Class I stain was normal (C). The mild perymysial inflammatory infiltration contained sparse T lymphocytes (D, CD8); P62 immunoreactive sarcoplasmic (E) and intranuclear (F) inclusions were seen in fibres. The large vacuoles/spaces seen in above slides were due to the processing artefacts and do not represent pathogenic features. The key pathological features were pointed by arrows. Scale bar represents 50 μm in A, C and D; 25 μm in B and E; 10 μm in F. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. The biopsy was reviewed and the images were taken by Professor Janice Holton from the Department of Pathology, ION.

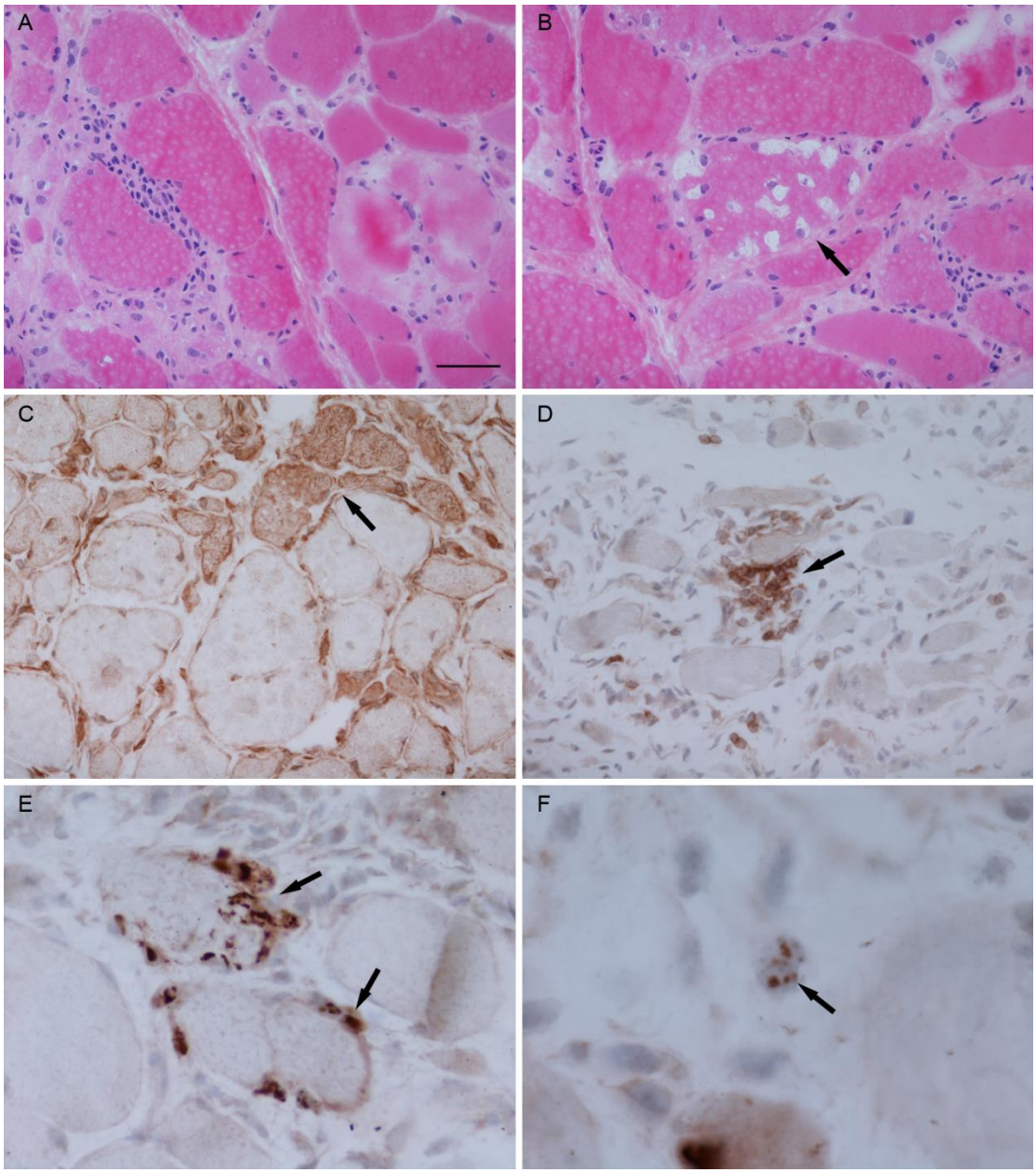


Figure 6-6 Biopsy features from the index patient II-1 from Family F

There was marked variation in fibre size with increased internal nuclei, moderate endomysial inflammatory invasion of intact fibres and several necrotic fibres (A, H&E); Vacuoles were seen in fibres but not rimmed (B, H&E); MHC Class I staining was increased at the sarcolemma and sarcoplasm, mostly affecting atrophic fibres (C); T lymphocytes were present in the endomyosium with infiltration of fibres (D, CD3); P62 immunoreactive sarcoplasmic (E) and intranuclear (F) inclusions were seen in fibres. The key pathological features were pointed by arrows. Scale bar represents 50 μm in A-D; 25 μm in E; 10 μm in F. The biopsy slides of this patient were stained and kindly provided by Dr Boel De Paepe from Gent Hospital, Belgium. The biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

6.2.2 Genetic Investigations on Two Families

WES was performed on the index patients E-II-2 and F-II-1 using TruSeq Exome Enrichment protocol (Illumina). The metrics for the whole-exome data is shown in Table 6-2. For E-II-2, the quality of the exome data was not very good as it has only 44.4% of the whole-exome with minimum of 10x reads. Therefore, searching for candidate genes would be a priority. The filtering strategy for non-synonymous variants with a MAF < 1% was used as described in Section 2.3.4. As the family was likely to demonstrate autosomal dominant inheritance, the variant selection was focused on the genes with a heterozygous variant. This left a number of 506 variants after the above filtering. To further narrow down the number of variants, genes associated with dominant hereditary IBM and vacuolar myopathies were considered as candidate genes including *VCP*, *MYH2*, *PABPN1*, *MYOT*, *HNRNPDL*, *SQSTM1*, *LD3*, and *CRYAB* genes. A heterozygous missense variant c.277C>T (p.Arg93Cys) in *VCP* gene was found in E-II-2 (Figure 6-7 A), and no variant was found in other candidate genes. This variant was absent in the 1000 Genomes project, EVS and ExAC databases, and was previously reported as pathogenic in a family with IBMPFD (Guyant-Marechal et al., 2006). The p.Arg93Cys in E-II-2 was subsequently confirmed by Sanger sequencing. The segregation of this variant was also confirmed in the affected sister (E-II-4). The DNA samples from other affected/unaffected members from the family were not available.

F-II-1 was initially included in the sIBM whole-exome project for analysing rare variants in candidate genes *SQSTM1* and *VCP*. A heterozygous missense variant c.374G>A (p.Gly125Asp) in *VCP* was found (Figure 6-7 B). After discovering his positive family history, the exome was re-analysed for patient F-II-2, and no variant in other candidate genes was found. The DNA samples from the affected father and the sister in Family F were not available to be analysed the segregation by the time of writing. The p.Gly125Asp is a novel variant which is absent in the public databases, and the Glycine at position 125 is highly conserved across different species (Figure 6-7 B).

Table 6-2 Whole-exome sequencing metrics for patients E-II-2 and F-II-1

Case ID	Total Reads	Mean Target Coverage	% Target Covered at 2x	% Target Covered at 10x	% Target Covered at 20x	Total Number of Variants
E-II-2	37966124	17.5x	59.9%	44.4%	33.7%	18823
F-II-1	161343738	87.8x	98.0%	94.7%	88.9%	18777

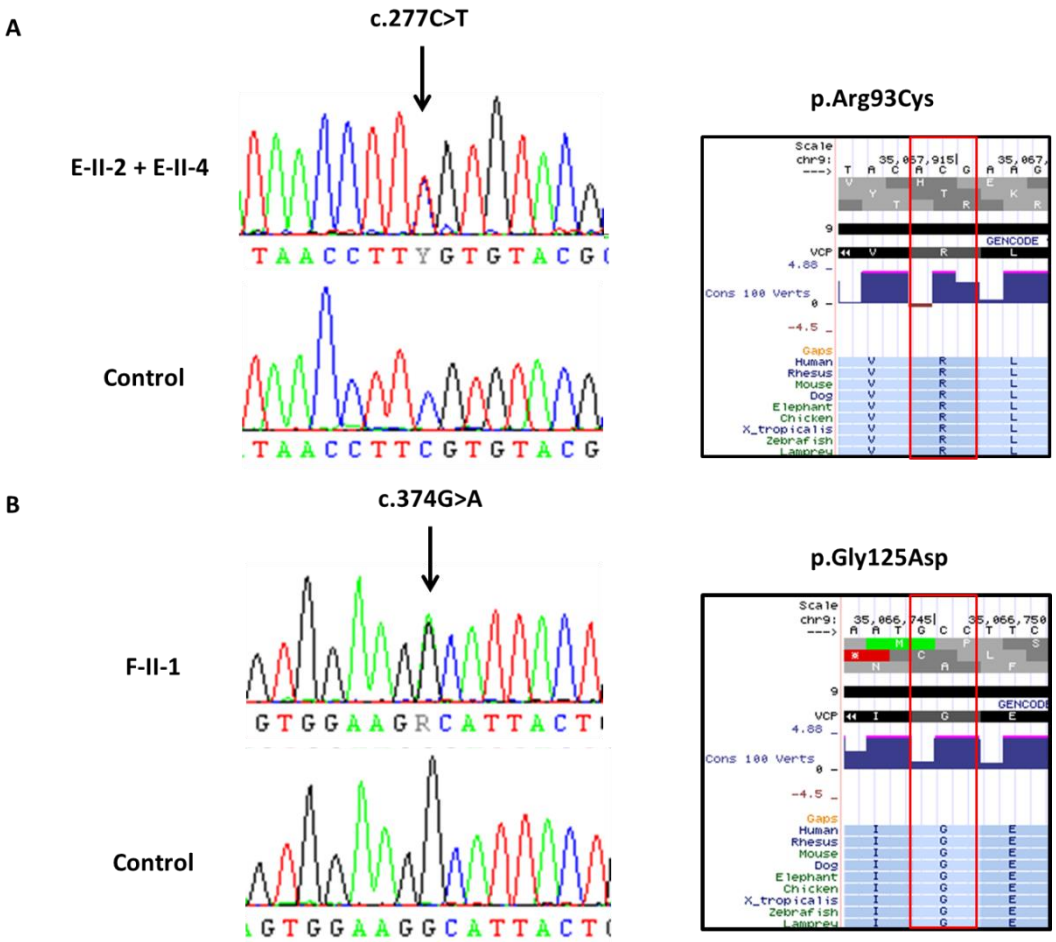


Figure 6-7 Variants in the *VCP* gene

(A) Electropherograms of DNA for the heterozygous missense mutation c.277C>T in *VCP* gene in Family E and conservation of Arginine at position 93; (B) Electropherograms of DNA for the heterozygous missense variant c.374G>A in *VCP* gene in Family F and conservation of Glycine at position 125.

6.2.3 Discussion

Mutations in the *VCP* gene are known as the cause of an autosomal dominant form of IBMPFD (OMIM#167320). IBMPFD is a rare multisystem disorder with a variable

penetrance of the three main features (Tucker et al., 1982, Watts et al., 2004). The rare variants in *VCP* gene have also been found in our sIBM cohort as discussed in Section 3.3.

To date, more than 30 missense mutations in the *VCP* gene have been reported across North American, European, Korean and Chinese familial cases (Haubenberger et al., 2005, Watts et al., 2007, Gidaro et al., 2008, Kimonis et al., 2008, Stojkovic et al., 2009, Kim et al., 2011, Gu et al., 2013). VCP is a member of the 'ATPases associated with a variety of activities (AAA-ATPase)' superfamily, and it is associated with a variety of cellular activities. These include cell cycle control, membrane fusion and the ubiquitin-proteasome degradation pathway (Watts et al., 2004). A number of independent studies have supported that disruption of a specific function of VCP leads to inclusion body formation (Watts et al., 2004). A study also proposed a mechanism that VCP deficiency causes profound mitochondrial uncoupling and decreased ATP levels, ultimately leading to cell death (Bartolome et al., 2013).

In Family E, the *VCP* p.Arg93Cys found in the two affected siblings has been previously reported in a family with IBMPFD (Guyant-Marechal et al., 2006). It was suggested to be pathogenic by blocking the normal protein degradation and triggering the accumulation of ubiquitinated and aggregated proteins *in vitro* (Poksay et al., 2011). Both patients had not developed cognitive or bone problem prior to their death. Muscle biopsy of the patient E-II-2 was compatible with the pathological features of inclusion body myopathy where it usually shows less inflammatory infiltration and negative MHC Class I but with rimmed vacuoles. Therefore, this mutation is very likely also the culprit to the phenotypes of the family.

The *VCP* p.Gly125Asp found in F-II-1 is novel, being absent in 1000 Genomes project, EVS and ExAC databases. The variation lies within the highly conserved CDC48 domain of the VCP protein involved in ubiquitin binding and protein-protein interaction, and is a hotspot for *VCP* mutations (Bersano et al., 2009). The variant was also predicted as damaging by PolyPhen2. Both the father and the sister were affected showing a dominant form of inheritance, which is compatible with the *VCP*-related hIBM. None of the affected members reported with bone problems, however,

the father developed frontal release signs. Interestingly, the muscle biopsy features seen in the patient F-II-1 were more compatible with inclusion body myositis, which showed moderate inflammatory infiltration with invasion of non-necrotic, increased MHC Class I fibres and also p62 positive inclusions. This indicates the challenge to differentiate between sIBM and hIBM, and also highlights the importance of collecting detailed clinical phenotypes as well as family history to provide an accurate diagnosis for patients. P62 accumulations in muscle also supports that the mutations in the *VCP* gene may lead to abnormal protein homeostasis. Further functional work on the p.Gly125Asp is required in the future.

6.3 Family G with Rare Autosomal-Recessive Myotubular Myopathy

6.3.1 Clinical and Pathological Details of the Two Affected Siblings

The index case (G-II-2, Figure 6-8) was seen by Professor Henry Houlden at NHNN and consented for obtaining his pictures for research and publication purpose. He is a 28 year old patient, born following a normal pregnancy and delivery, but at birth had syndactyly of two fingers of both hands and two toes of both feet. He also had pes cavus of both feet (clinical features are shown in Figure 6-9 D-F and K-M). He started walking by age of $1^1/_2$, and was seen by a local paediatrician at the age of three because of his clumsiness when walking. By the age of four, he also had difficulty pronouncing consonants. He was noted to display poor pencil grasp at the age of eight, and received a diagnosis of dyspraxia with hypotonia. In addition, he had visual abnormalities with bilaterally dilated pupils (unreactive to light) at the age of two, and had a history of photophobia dating back to the age of three. Over the years, the patient had symptoms of slowness of swallowing, choking on food, blurred and double vision, imbalance, dragging his legs, easy fatigue and weakness. These symptoms, particularly his gait, speech, and eyelid drop, were worse in the evening. The patient currently wears a fibrotic foot matrix daily to support his walking. Upon examination, he had obvious strabismus, gaze evoked nystagmus to the left as well as on up-gaze, and he had broken pursuit movements. His pupils were not reactive to light or accommodation. His fundi were normal. The rest of cranial nerve examination was normal. Power in lower limbs was generally at grade 4-/5

throughout. Distal weakness in wrist extension was at 4-/5, and finger extension and flexion were 3/5. He displayed symmetrical hypo-reflexia with bilateral extensor plantars, however his coordination was normal. Sensory examination showed that he had a patchy pin-prick loss in a glove and stocking distribution. Blood tests were negative for acetylcholine receptor antibodies, anti-MUSK antibody and anti-voltage gated calcium channel (VGCC) antibody. The CK level was normal. Metabolic testing was normal. EMG/NCS showed moderately severe axonal sensory-motor neuropathy (motor >> sensory) and polyneuropathy. There was no evidence of defect at the neuromuscular junction. There were mild chronic neurogenic changes on sampling the left orbicularis oculi, probably due to a lesion at the level of the facial nucleus. MRI of head and whole spine showed mature damage in periaqueductal grey matter of upper tegmentum, and the prominence of the central canal at T7/8 and the cord was mildly thinned throughout but was otherwise normal. The muscle biopsy showed marked myopathic features with prominent internal nuclei and appearances of necklace fibres, which suggest a type of myotubular myopathy (Figure 6-10). The biopsy was reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION. Previous genetic analyses were negative for three common mutations in mtDNA (A3243G, A8344G, and T8993C/G), CAG repeat expansion in the *AR* gene, and other genes including *MTM1*, *DNM2, AMPH2*/*BIN1*, *CFL2*, *MT-ATP6, MT-ATP8*, *POLG*, *GAN* and *SACS*.

The patient had a positive family history. His elder sister (G-II-1) also had similar pupils and disc appearance to the index patient, but with less pronounced problems with walking. On examination, she had similar sensation loss in a glove and stock distribution, and also had abnormal colouration of both feet suggesting a sensory neuropathy (Figure 6-9 H-J). The family is consanguineous, as the parents are second cousins. Both of their parents and the younger brother were not affected. Overall, the phenotype in this family is consistent with an autosomal recessive form of inheritance.

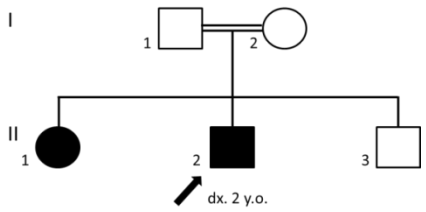


Figure 6-8 Pedigree of the Family G with two affected siblings

Black symbols represent affected patients; Black arrow points the index patient of the family.

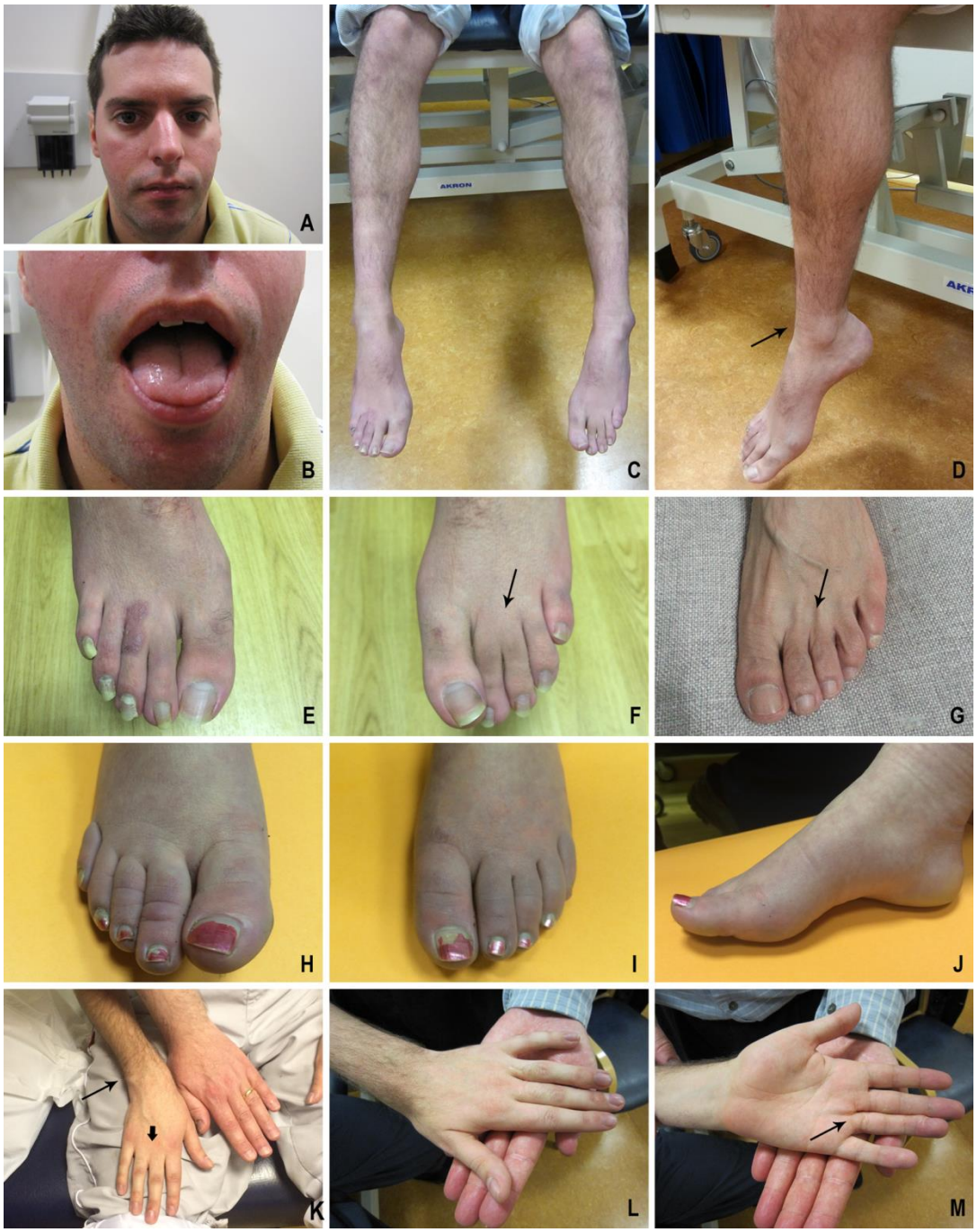


Figure 6-9 Clinical features of the affected patients in Family G

(A-F) and (K-M) were from the index patient II-2, and (H-J) were from the affected sister II-1. The tongue of the index patient was atrophic (B); Muscle wasting in both calves (C); Pes cavus and slimed ankle as pointed by arrow in (D); Syndactyly of two toes of both feet (E-F) as pointed by arrow, compared to a normal foot (G); Abnormal colour of both feet of the affected sister (H-J); Syndactyly of two fingers of both hands (K-M) pointed by a small arrow in compared to a normal hand on the right (K), a slimed wrist pointed by an arrow in (K), and a surgery scar on left hand pointed by arrow in (M). (Both patients consented for obtaining their pictures for research and publication purpose.)

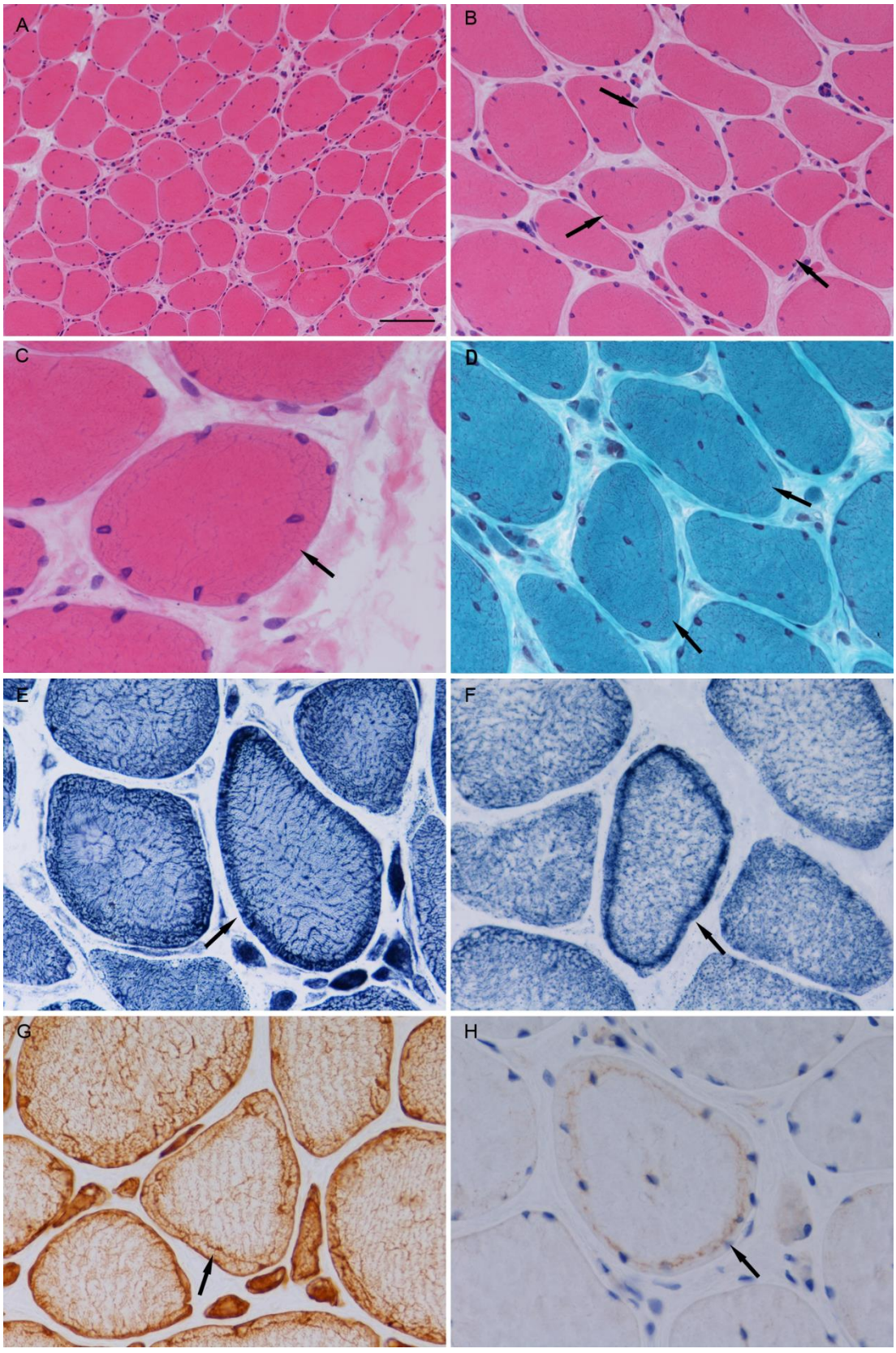


Figure 6-10 Biopsy features of the index patient II-2 in Family G

H&E staining showed an increased variation in fibre size and frequent internal nuclei (A); Internal nuclei were often distributed in a linear fashion displaced within the fibre and laying along a basophilic line resembling a necklace fibre (arrows in B, and C); Gomori trichrome staining showed

some increase in the density of mitochondrial staining in a peripheral band in occasional fibres (arrow in D), and there were no ragged red fibres; NADH-TR (E) and SDH (F) stainings confirmed the impression of necklace fibres with an increase in the intensity of peripheral staining with a band-like pattern; (G) showed some desmin positive staining in the peripheral band of many fibres and (H) showed fine granular p62 positive staining along the ring visible in a necklace fibre. Necklace fibres were pointed by arrows. Scale bar represents 100 μm in A; 50 μm in B; 25 μm in C-D and E-H. Stainings of muscle biopsy were performed by staff at Department of Pathology, ION. Biopsies were reviewed and images were taken by Professor Janice Holton from the Department of Pathology, ION.

6.3.2 Genetic Investigations on Family G

WES was performed on the index patient G-II-2 using Nextera Rapid Capture Exome protocol (Illumina). The metrics of the whole-exome data are shown in Table 6-3. The filtering strategy for non-synonymous variants with a MAF < 1% was used as described in Section 2.3.4. As the patient has unaffected consanguineous parents, the variants were further filtered for genes with homozygous variants or with compound heterozygous variants. This left a total of 118 genes. Then, these genes were annotated with a list of reported neuromuscular genes downloaded from the online GeneTable of Neuromuscular Disorders (http://www.musclegenetable.fr/) updated by August of 2015. New genes that were reported after this time point were added to the list manually. This left five genes including *SLC52A3*, *SBF1*, *TTN*, *ATXN1* and *ATXN3*. The depths of variants in *SLC52A3* and *TTN* were below 11 reads, which is so low that is unlikely to be real. Both *ATXN1* and *ATXN3* are autosomal dominant genes, which are not compatible with the recessive inheritance of the family. This only left a novel homozygous frameshift deletion (c.5477-5478del; p.1826-1826del) in exon 40 of the *SBF1* (SET-binding factor 1, also called myotubularin-related 5 [MTMR5]) gene. This gene is known to cause an autosomal recessive subtype of Charcot-Marie-Tooth disease (CMT). The variant was subsequently confirmed by Sanger sequencing. Segregation analysis on both parents and two siblings found that the affected sister (G-II-1) also harboured the same homozygous frameshift deletion in *SBF1*, while both parents as well as the unaffected sibling were heterozygous carriers of the deletion in *SBF1* (Figure 6-11). To determine if this deletion is transcribed into mRNA, cDNA sequencing was performed using mRNA extracted from fibroblasts of the index patient, his parents and a normal control (Figure 6-12). The process was as described in Sections 2.5.1-

2.5.3. The deletion occurs in the last exon of *SBF1* gene and causes a frameshift of the amino acid codon. This leads to a premature stop codon at 1850 amino acid position and produces a truncated protein by 43 amino acid residues. This deletion is novel and is absent in the public databases.

In addition, as the muscle biopsy of the patient suggested a type of myotubular myopathy, genes associated with autosomal recessive myotubular myopathies and other recessive myotubularin-related genes (*MTMR2*, *MTMR3*, *MTMR4*, *MTMR7-9*, *MTMR12*, *SBF2*, and *RYR1*) were also searched in the exome. Then, compound heterozygous missense variants c.3292C>G (p.Leu1098Val) and c.3433A>G (p.Arg1145Gly) in exon 26 of the *SBF2* (SET-binding factor 2, also called myotubularin-related 13 [MTMR13]) gene were found in the index patient G-II-2. This is also a CMT gene. The p.Leu1098Val is with a MAF = 2.1% in the ExAC database, and the p.Arg1145Gly is a very rare variant with a MAF = 0.02%. Both amino acid residues Leucine and Arginine at codons 1098 and 1145 in *SBF2* are highly conserved across different species (Figure 6-11) and were predicted to be damaging by PolyPhen. The variants were subsequently confirmed by Sanger sequencing (Figure 6-11). Both parents and his brother were carriers of a heterozygous missense variant in *SBF2* respectively, however, the affected sister G-II-1 was absent of the variants in *SBF2*.

Table 6-3 Whole-exome sequencing metrics for patient G-II-2

Total Reads	Mean Target Coverage	% Target Covered at 2X	% Target Covered at 10X	% Target Covered at 20X	Total Number of Variants
44467160	33.6x	97.5%	84.8%	66.2%	21636

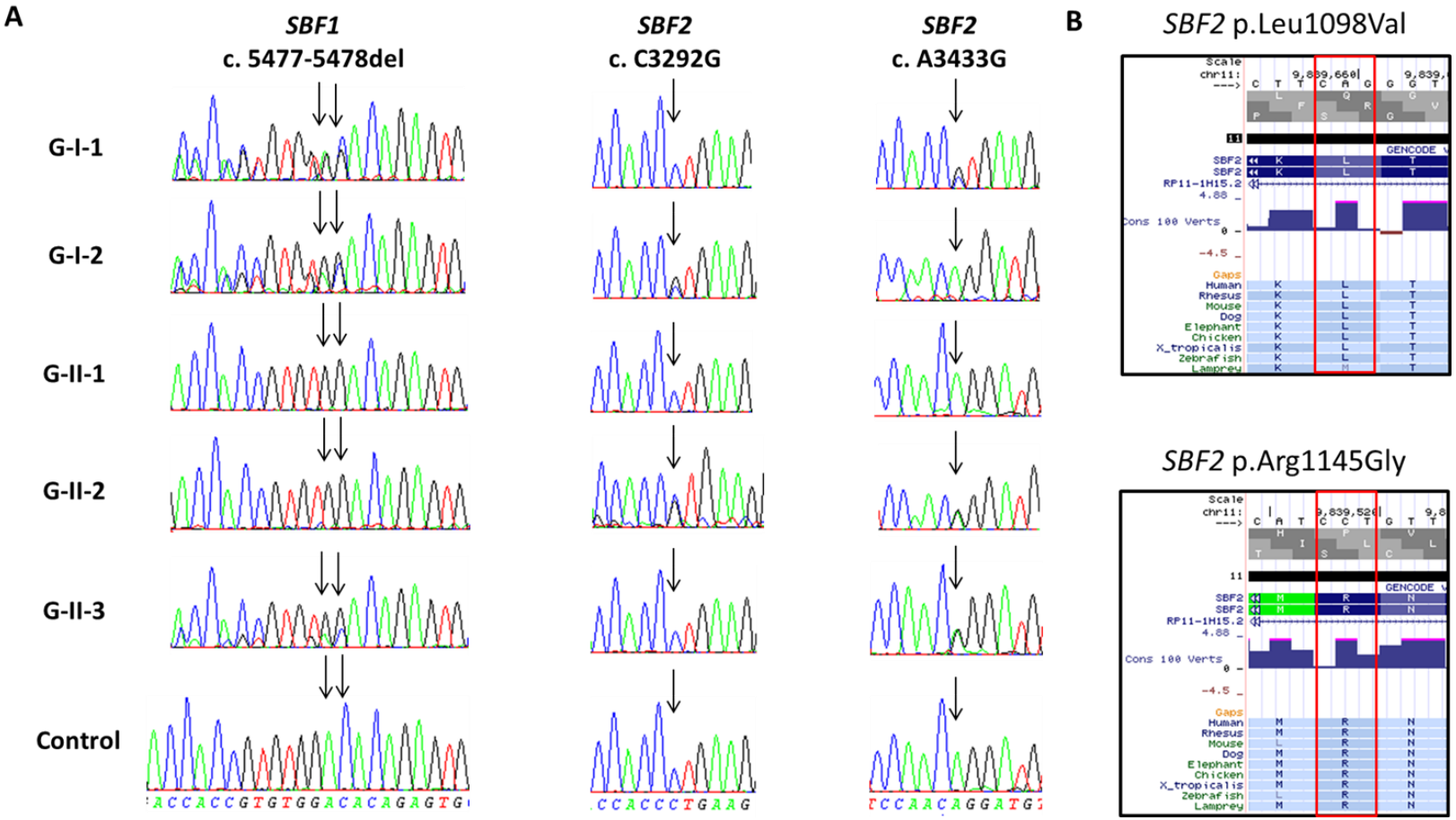


Figure 6-11 Deletion and variants identified in the *SBF1* and *SBF2* genes

(A) Electropherograms of DNA for a frameshift deletion in *SBF1* and compound heterozygous missense variants in *SBF2*; (B) conservation of the amino acids at position 1098 and 1145 in the *SBF2* gene across species.

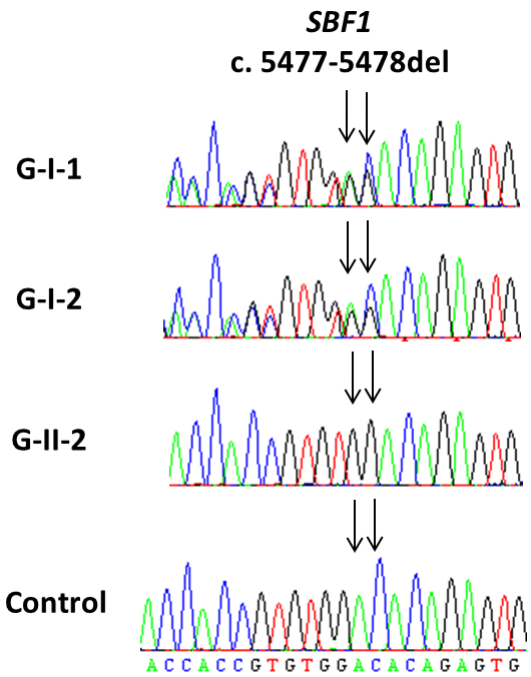


Figure 6-12 Electropherograms of cDNA for the frameshift deletion in the *SBF1* gene

G-I-1 and G-I-2 are parents; G-II-2 is the index patient; control is a normal healthy control.

6.3.3 Protein Levels of the MTMR5 (*SBF1*) and MTMR13 (*SBF2*) in Patient Fibroblasts

Although the p.Leu1098Val in *SBF2* is with a MAF > 1% in the population, the possibility of affecting the function of the protein still cannot be ruled out when it is combined with a very rare variant. Therefore, western blot analysis was performed to investigate the impact of variants on the expression of protein levels of *SBF1* and *SBF2* using fibroblasts from the index patient (G-II-2), both the parents (G-I-1 and G-I-2), and three normal healthy controls. The fibroblasts were not available from the affected sister (G-II-1) for the analysis. The protocol of western blot was as described in Sections 2.5.4-2.5.6.

There was a markedly reduced level of MTMR5 protein in the patient with homozygous frameshift deletion compared with both heterozygous parents and controls. The levels of MTMR5 protein in unaffected parents were similar. (Figure 6-13 A and B) The sample size was not sufficient to perform statistical analysis, but there was a trend that the level of MTMR5 protein was lower in carriers than controls (Figure 6-13 C). This suggests that the homozygous frameshift deletion p.1826-1826del in *SBF1* causes a marked decay of the protein level. Western blot for

MTMR13 antibody showed many unspecific bands which were not feasible to quantify the protein level. However, the bands close to the molecular weight of MTMR13 protein (208kDa) showed no obvious difference between the patient, unaffected parents and controls. It seems that the compound heterozygous missense variants in *SBF2* do not affect the expression of the protein (Figure 6-14).

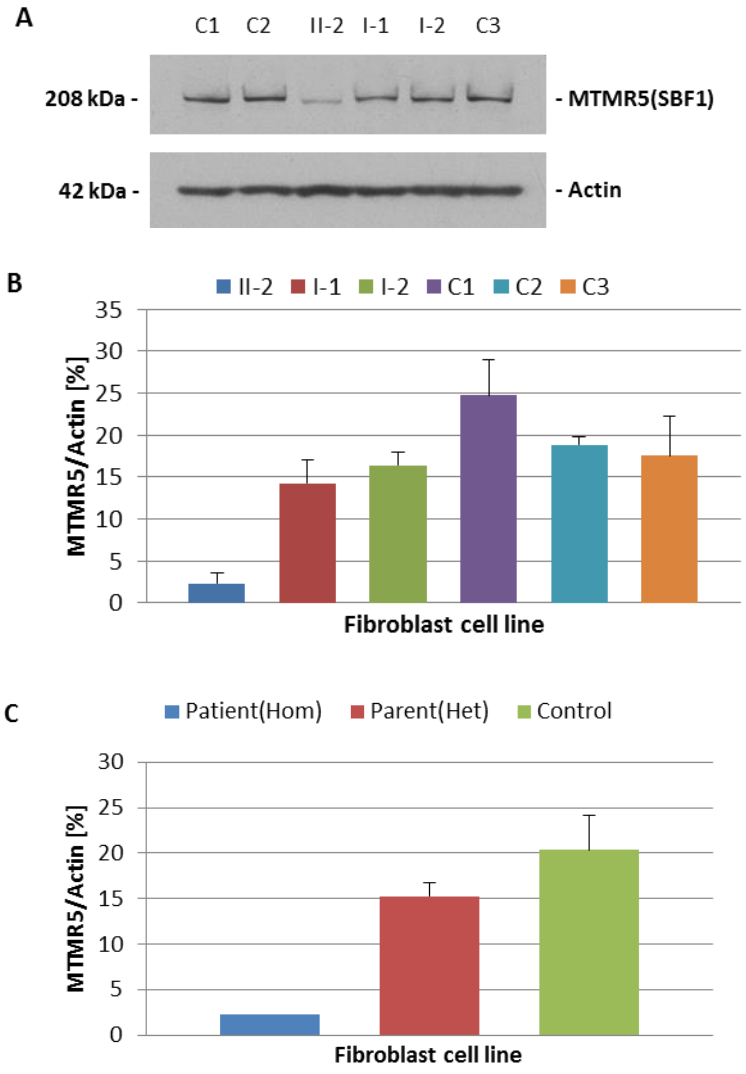


Figure 6-13 Analysis of MTMR5 protein level in fibroblasts

(A) Western blot analysis for MTMR5 and actin proteins in fibroblasts (This represents an example of three experiments carried out); (B) Protein levels of MTMR5 in all six fibroblast cell lines (Each protein level is a mean value of three experiements); (C) MTMR5 levels compared between the patient, carriers (mean of I-1 and I-2) and controls (mean of C1-C3). Error bar represents the standard deviation. (II-2 is the index patient, I-1 and I-2 are parents [carriers] from Family G; C1-C3 are three normal controls. Hom = homozygous; Het = heterozygous.)

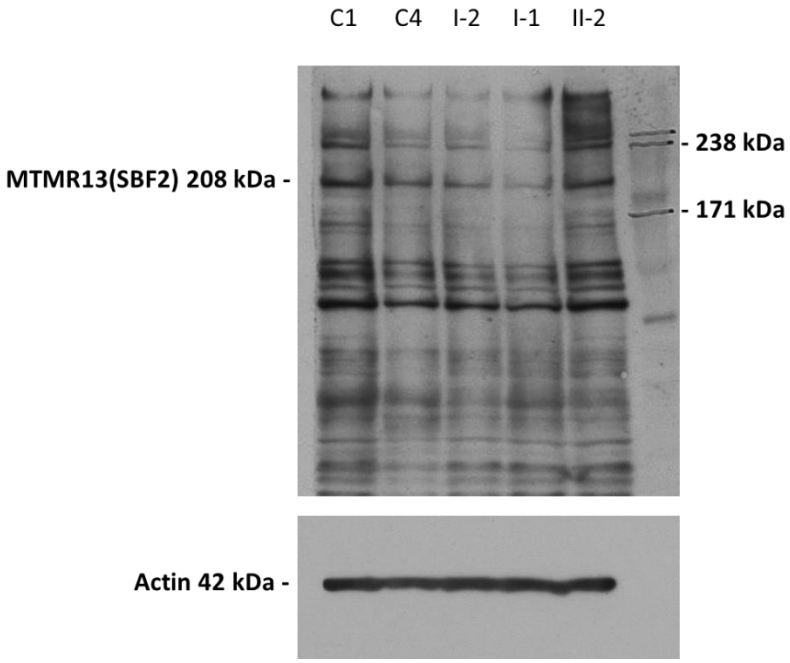


Figure 6-14 Western blot of MTMR13 protein level in fibroblasts

The membrane was incubated overnight at 4 °C with the MTMR13 primary antibody (Santa Cruz, USA) at a dilution of 1/500 with 5% BSA in TBS-Tween buffer. On the following day, after being washed with TBS-Tween buffer for 10 minutes three times, the membrane was incubated with the secondary goat anti-mouse IgG-HRP antibody (Santa Cruz, USA) at a dilution of 1/5000 with 1% milk in TBS-Tween buffer for one hour at room temperature. MTMR13 antibody in fibroblasts showed many unspecific bands which were not feasible to quantify the protein level. However, the bands close to the molecular weight of MTMR13 protein (208kDa) showed no obvious difference between the patient, unaffected parents and controls. The protein level of Lane I-1 was slightly less, probably due to less ECL enhancer distributed in that area. (II-2 is the index patient; I-1 and I-2 are parents [carriers]; C1 and C4 are two normal controls).

6.3.4 Discussion

We identified a novel homozygous frameshift deletion in *SBF1* and compound heterozygous missense variants in *SBF2* in an autosomal recessive family. The two siblings were affected with unusual neuropathic and myopathic phenotypes in addition to necklace fibres in muscle biopsy of the index patient (G-II-2). There was a markedly reduced level of MTMR5 protein in the index patient, and the MTMR5 level in the affected sister (G-II-1) with a same homozygous deletion in *SBF1* should be similar as G-II-2. This suggests the causality of the homozygous deletion in a loss-of-function fashion.

The *SBF1* and *SBF2* genes encode proteins MTMR5 and MTMR13, members of the myotubularin-related (MTMR) family. MTMR proteins are a large family of phosphoinositide (PI) 3-phosphatases which all share the same protein domains core including GRAM domain (Glucosyltransferase, Rab-like GTPase Activator and Myotubularins), RID (Rac-Induced recruitment Domain), PTP/DSP homology (tyrosine/dual-specificity phosphatase super-family), and SID (SET-interacting domain) (Laporte et al., 2003). Some members of the family contain a catalytically active PTP/DSP domain, such as MTMR2, dephosphorylating phosphatidylinositol 3-phosphate (Ptdlns3*P*) and phosphatidylinositol 3,5-bisphosphate (Ptdlns[3,5]P_2). They are two lipids which regulate endo-lysosomal membrane traffic (Laporte et al., 2003). Some members contain catalytically inactive substitutions in the phosphatase domain, thus they belong to a subgroup called "pseudophosphatases" (Laporte et al., 2003). MTMR5 and MTM13 are homologues which share 59% overall sequence identity, and both are catalytically inactive phosphatases (Azzedine et al., 2003). Studies suggested that MTMR5 and MTMR13 interact with MTMR2 directly via their coiled-coil domains increasing the phosphatase activity of MTMR2 as well as commanding its subcellular localization (Kim et al., 2003, Robinson and Dixon, 2005). Mutations in MTMR genes are known as the cause of an autosomal recessive demyelinating Charcot-Marie-Tooth disease type 4B (CMT4B) family. CMT4B is characterised by severe axonal loss and focal myelin outfoldings of the nerves from the patients (Previtali et al., 2007). Subtypes CMT4B1, CMT4B2, and CMT4B3 are implicated in mutations in *MTMR2* (OMIM#603557) (Bolino et al., 2000), *SBF2* (OMIM#607697) (Senderek et al., 2003), and *SBF1* (OMIM#603560) (Nakhro et al., 2013), respectively.

It has been suggested that loss of function of MTMR13 leads to alterations in MTMR2 function and subsequent dysfunction of membrane trafficking, endocytic/exocytic processes and neuron-Schwann cell interactions. This effect is proved both *in vivo* and *in vitro* via imbalanced 3-phosphoinositide signalling (Berger et al., 2006), as well as via altered Rab21 GTPase activity (Jean et al., 2012). MTMR13 -/- mice showed myelin outfoldings and infoldings in motor and sensory peripheral nerves as pathological hallmarks of CMT4B2 (Tersar et al., 2007, Robinson et al., 2008). The level of MTMR2 was decreased by ~50% in MTMR13-/- sciatic nerves, suggesting that MTMR13 influences MTMR2 stability (Robinson et

al., 2008). MTMR2 also plays a crucial role in axon development and maintenance by interacting with the neurofilament light chain (Goryunov et al., 2008). This could be affected due to altered function of MTMR13.

Loss of function of MTMR5 was suggested to cause a defect in late stages of spermatogenesis in mice (Firestein et al., 2002) and studies on humans showed that rare variants in *SBF1* are linked to male infertility (Kuzmin et al., 2009). The pathophysiological pathways of MTMR5-induced neuropathy have not been fully elucidated. As MTMR5 is the closest homologue of MTMR13, it is possible that the loss of function in MTMR5 could also induce dysfunction in neuronal or Schwann cells via a similar mechanism with the MTMR2/MTMR13 complex. This may explain the strikingly similar phenotypes of patients with recessive mutations in the *MTMR2*, *SBF2* and *SBF1* genes. In this study, the protein level of MTMR5 was markedly reduced due to the homozygous frameshift deletion in the patient, indicating a loss-of-function mechanism underlying the pathogenesis of the disease. Whether this causes the patient to be infertile is as yet unknown. The protein level of MTMR13 seemed not to be affected by the compound heterozygous missense variants found in *SBF2*. Both variants lying in/near the PTP/DSP domain may act as contributing factors which affect the cellular function of MTMR13. This would explain why the index patient with defects in both *SBF1* and *SBF2* genes has more severe phenotypes compared to his sister who only has homozygous deletion in *SBF1*.

Clinical phenotypes of our patients with *SBF1* and *SBF2* mutations were similar to a previous report of a Saudi family with CMT4B3 and a homozygous missense variant in *SBF1*. The affected family members also had strabismus and syndactyly in addition to progressive weakness of the legs and neuropathy (Alazami et al., 2014, Bohlega et al., 2011). Our cases are the second family identifying *SBF1* gene with syndactyly, which supports a hypothesis of a non-neuro-specific developmental role of this gene.

Necklace fibres seen in muscle biopsy are another intriguing feature observed in the index patient. These fibres were usually described to be small and obliquely oriented with an increased density of mitochondria and sarcoplasmic reticulum profiles, and

internal nuclei usually aligned like the necklace (Bevilacqua et al., 2009). Necklace fibres with nuclear internalisation have been reported as a histological hallmark of X-linked myotubular myopathy related to mutations in the phosphoinositide phosphatase myotubularin 1 (*MTM1*) gene in several studies (Romero, 2010, Hedberg et al., 2012, Bevilacqua et al., 2009). A report of a patient with severe X-linked *MTM1*-related myotubular myopathy suggested an increase in the number of necklace fibres during the disease course (Gurgel-Giannetti et al., 2012). Recently, necklace fibres were also reported in a late-onset dynamin-2 (*DNM2*) gene related centronuclear myopathy (Casar-Borota et al., 2015). Previously, the presence of necklace fibres in muscle biopsy indicated a diagnosis of myotubular myopathy. Our study is the first to report necklace fibres in *SBF1*-related phenotypes, which expands the clinical and molecular genetic spectrum of this specific pathological feature. This may also indicate a common pathogenic mechanism between *MTM1*, *DNM2* and *SBF1* associated myopathies. In addition, since necklace fibres are not a specific feature in muscle biopsy, it is important to combine with patients' clinical phenotypes and inheritance pattern to identify the potential causative gene.

6.4 Overall Conclusion

In this chapter, four families with rare myopathies were investigated using WES. A homozygous missense known mutation (p.Ser60Phe) in *MYOT* was identified in the two affected siblings in Family D with myofibrillar myopathy and cytoplasmic bodies in muscle fibres. A heterozygous known missense mutation (p.Arg93Cys) and a novel missense variant (p.Gly125Asp) in *VCP* were identified in Family E and Family F, respectively. Both families were considered to have hIBM and p62 positive inclusions were observed in the affected muscle fibres. A novel homozygous deletion (p.1826-1826del) in *SBF1* and compound heterozygous missense variants (p.Leu1098Val and p.Arg1145Gly) in *SBF2* were found in Family G with an unusual subtype of CMT4B with syndactyly and necklace fibres in muscle biopsy of the index patient. Western blot on the fibroblasts from Family G showed a markedly reduced level of MTMR5 protein suggesting a loss-of-function mechanism for the pathogenesis of the disease.

This chapter has shown that WES can be used to identify a disease-causing gene that was missed by Sanger sequencing, and further help to provide genetic diagnosis of rare myopathies for previously undiagnosed families. Whilst the mutations may have been missed if they locate in a splice site or outside the exons, WES can still provide more thorough screening than Sanger sequencing. However, there are still some issues with WES which mean that the process of analysis is not always straightforward. Firstly, sometimes the quality of WES may not be decent probably due to the quality of samples, the reagents of the exome capture protocol, or misalignment in the pipeline. These can cause a failure of the sequencing or generate poor WES data with low coverage. For example, the WES of E-II-2 has only 44.4% of all the exons with a coverage of more than 10 reads. This means that the data from more than half of the protein-coding regions was potentially not very reliable because of a low read depth. The mutation would have been missed if it is located in the poor regions of the exome. Therefore, a good quality of WES is an essential first step of mutation discovery. Secondly, it is common to have several hundred variants left even after a hard filtering. It was lucky to identify variants in known genes which were selected according to the clinical phenotypes and genetic inheritance, despite the small pedigree of all four families. Small families would have challenges to find the culprit of the disease, especially there is a novel disease-causing gene. A hard filtering may also miss synonymous variants with functions in splicing or gene regulation. In addition, unavailable samples or any missing information from the key family members (first degree relationship of the index patient) would bring more difficulties to analysis as discussed in the previous chapter. Therefore, this once again highlights the importance of obtaining the samples of affected/unaffected families as well as detailed pedigree information of clinical and pathological phenotypes when it comes to a genetic investigation of a family with heterogeneous diseases. Re-analysis of unsolved exome data would be also necessary when the knowledge of the know genes is improved and new genes are discovered in similar phenotypes.

Chapter 7 General Conclusions

The aim of this thesis was to investigate the genetic basis of two main disease groups: i) sporadic inclusion body myositis (sIBM), and ii) myopathies with rare/specific structural abnormalities in muscle to improve our understanding of the pathogenesis of these diseases. Both traditional sequencing and whole-exome sequencing (WES) techniques were applied to achieve this.

7.1 Genetics of sIBM

Since sIBM was first described in 1971 (Yunis and Samaha, 1971), our knowledge of this disease has greatly improved. Studies have proposed that the pathogenesis of sIBM is likely to be influenced by multifactorial events, including inflammation, degeneration, mitochondrial dysfunction, and ageing, as well as genetic and environmental factors (Gang et al., 2014, Askanas et al., 2015, Needham and Mastaglia, 2016, Rygiel et al., 2015). However, the primary cause remains controversial after decades of research.

Based on the International IBM Genetics Consortium and Muscle Study Group, we started to collected sIBM samples by searching in-house DNA and pathology databases, and set up collaborations from September 2012. By the 11th March 2016, DNA, blood, and muscle tissue from a total of 440 IBM patients had been collected from 17 specialised centres in seven countries around the world (Table 3-1). Two-hundred and thirty-nine patients had been confirmed to fulfil the study criteria for sIBM, and 181 of the DNA samples had completed WES. This is the largest sIBM patient cohort and also the largest sIBM whole-exome cohort for genetic investigations to date. Studies in both Chapters 3 and 4 were based on this cohort of sIBM patients.

In Chapter 3, several genes were investigated using diverse genetic techniques by a candidate gene screening approach.

Firstly, we studied the association between the risk of sIBM and features of the *APOE* genotypes and an intronic polymorphism (polyT repeat) in the *TOMM40* gene in a large group of 158 sIBM patients and 127 age-matched controls (Section 3.2). This study confirmed that the *APOE* ε4 allele was not a susceptibility factor for developing sIBM. *APOE* alleles were also not significantly associated with the age of onset of the disease, but patients with the ε2 allele were more likely to present with symptoms slightly earlier. These findings are consistent with a previous study (Needham et al., 2008b). However, our study failed to reproduce the results from a previous *APOE-TOMM40* study on a group of 90 Caucasian sIBM patients (Mastaglia et al., 2013). In our study, carriers of the *APOE* ε3/ε3 or ε3/ε4 genotypes with an intronic very long (VL) polyT repeat (≥ 30 bp) were not under-represented in sIBM compared to controls. This polymorphism was also not associated with the risk of sIBM in our study. However, the carriers of a VL repeat allele had a significantly later age of onset (3.7 years later) of sIBM symptoms compared to other patients. Among those who also had the *APOE* ε3/ε3 genotype, the age of onset was delayed by 4.9 years. These findings suggest that the *TOMM40* VL polyT repeat has a disease modifying effect on sIBM by delaying the onset of symptoms, and the *APOE* ε3/ε3 genotype enhances this effect.

Secondly, two genes *SQSTM1* and *VCP* were screened in the 181 sIBM whole-exome data (Section 3.3). In this study, six rare missense variants (MAF < 1%) in the *SQSTM1* and *VCP* genes were found in seven patients, with a frequency of 3.8% in this sIBM cohort (Table 3-6). Five of these variants have been reported in patients with ALS, IBMPFD, familial PDB, early-onset AD, and PD. Of note, the frequency of the variants *SQSTM1* p.Gly194Arg and *VCP* p.Arg159Cys was significantly higher in our sIBM cohort compared with the ExAC database. Subsequently, whole-genome expression analysis found that MHC class I (*HLA-A*) and II (*HLA-DRA*) genes were significantly up-regulated in these patients compared to controls. The up-regulation of these inflammatory markers was more pronounced in the *SQSTM1* patient group than in the *VCP* patient group. Our study is the first report on likely pathogenic *SQSTM1* variants and also expands the spectrum of *VCP* variants in sIBM. We also suggest different degree of up-regulation of inflammatory markers between sIBM patients with different genetic variants. Our findings suggest that rare variants in these genes constitute genetic susceptibility factors for sIBM. In addition,

this study expands the clinico-pathological spectrum of diseases associated with the *SQSTM1* and *VCP* genes. The overlap between sIBM and neurodegenerative diseases suggests that muscle and brain diseases share similar pathogenic pathways. This may be important for further discovery of biomarkers, genes, and therapeutic targets.

Lastly, we also reported the screening of appropriate cohorts of sIBM for a *MAPT* haplotype with a deletion in intron 9, *PRNP* codon 129 genotypes, coding variants in *APP* exons 16 and 17, *C9orf72* repeat expansion, and mtDNA deletion(s) (Section 3.3). No association was found between *MAPT* genotypes, *PRNP* codon 129 genotypes and sIBM patients. No coding variant was found in *APP* exons 16 and 17, and no defined deletion was found in mtDNA (on muscle tissues) in the tested sIBM patients. However, as these genes were not sequenced completely, this does not exclude the potential presence of pathogenic variants in other regions of these genes. Unfortunately the DNA samples extracted from patient muscle tissues were not of sufficient quality for detecting mtDNA deletion(s). Future studies screening mtDNA will be required. Interestingly, two patients were identified with an allele of 27 and 35 units of the hexanucleotide GGGGCC repeat respectively amongst 187 sIBM patients. This is the first report of a *C9orf72* repeat of more than 25 units (repeat < 25 units is considered to be within common normal range) in sIBM patients. However, the pathogenicity of these intermediate repeats and the size of the smallest expansion unit that confers disease risk are still unclear.

Chapter 4 demonstrated the first whole-exome association study that performed on a cohort of 181 sIBM patients and 510 neuropathologically healthy individuals. This is the largest WES study in sIBM patients to date. All the sIBM samples were sequenced in-house, while the control samples were sequenced at NIH. All the sequencing raw data went through an in-house bioinformatics pipeline. A total of 371264 variants were generated in 691 samples. A stringent quality control (QC) strategy was employed for genotype QC, variant QC, and sample QC using a series of bioinformatics tools (Figure 4-2). Population stratification was also analysed to exclude non-European samples. This left a total of 105883 variants in 601 samples (sIBM = 138; controls = 463). Single variant association analyses found significant associations with 15 common SNPs located on chromosome 6 and one rare variant on chromosome 15 after correcting for multiple testing (Figure 4-13 Manhattan plot,

and Table 4-4). The genes with these SNPs on chromosome 6 include *PRRC2A*, *DPCR1*, *TRIM31*, *SIKV2L*, *BTNL2*, *PSORS1C1*, *HLA-C*, *FKBPL* and *SLC44A4*. These are all located close or within MHC class I-III regions. The rare variant on chromosome 15 belongs to the *SLC24A1* gene. Logistic regression analysis adjusted by gender showed that the SNP rs3094086 in *DPCR1* has the strongest association with increased risk of disease (OR = 2.9, 95% CI: 2.049 – 4.193), while two SNPs (rs3132554 and rs3130983) in *PSORS1C1* have a protective effect (both with an OR = 0.5 and 95% CI: 0.3847 – 0.6799) (Table 4-5). Subsequently, gene-based analyses were performed to aggregate multiple variants in a gene unit for all genes and also for a group of selected candidate genes. Variants were weighted based on their functional annotation and their frequency such as non-synonymous variants, variants with MAF < 1%, variants predicted to be not benign by SIFT and PolyPhen, or LOF variants (Table 4-6). However, no gene passed the Bonferroni-correction threshold for significance either among all genes or only among the candidate genes for any test model. It is likely that our study is under powered to detect the genetic effects at gene level. In addition, if variants outside rather than within the coding regions contribute much of the risk of sIBM, these would be undetectable by WES.

WES has shown the power to uncover all of the coding regions, especially rare variants that are suggested to be "high impact" and usually evade detection by GWAS. This makes WES a more powerful approach than the conventional GWAS in studying rare complex diseases such as sIBM. However, data clean-up (QC) is one of the biggest challenges and the most time-consuming part of the population-based association analysis, as WES produces many systematic bias/artefacts which could result in false-positive findings. In spite of the challenges and difficulties, we managed to establish and optimise a first bioinformatics QC pipeline for such analysis in our department. Significant findings from single variant association analyses proved that the design and data process of WES applied in this study were reliable. From this study, it addresses the importance of generating homogeneous high quality WES data, which can provide reliable findings and also can make life easier for the researchers. However, as a large number of samples were required for population-based studies, they usually rely on multi-centre collaborations, which make it unavoidable that data come from heterogeneous sources. Therefore, a careful

study design and choice of the optimal QC protocol are essential for the downstream association analyses.

Our findings from these two chapters suggest that both degenerative and inflammatory factors contribute to the pathogenesis of sIBM. Candidate gene study reporting rare variants in *SQSTM1* and *VCP* as susceptibility factors for sIBM patients showed an overlap between this muscle disease and other neurodegenerative disorders. This indicates that the accumulation of proteins in muscle may follow similar biological pathways to those in the brain, eventually causing death of muscle cells and muscle atrophy. On the other hand, the whole-exome wide association study reported significant association between common SNPs in MHC regions and the risk of sIBM. This finding is in agreement with previous studies that reported strong associations with MHC haplotypes in smaller cohorts of sIBM patients by HLA typing. Our study provides further evidence of inflammatory factors to sIBM using a different technique without prior hypotheses. Whilst the primary cause of sIBM remains a puzzle, our studies confirmed some of the previously published findings as well as providing new insights into the genetic landscape of sIBM. In addition, the findings from our studies also support the theory that sIBM is a polygenic muscle disease. There are still many unknown areas of sIBM genetics. More samples will be needed to increase the study power at gene level.

7.2 Genetics of Myopathies with Tubular Aggregates and Cylindrical Spirals

Myopathies with tubular aggregates (TAs) and cylindrical spirals (CSs) are a group of rare muscular disorders with high clinical and genetic heterogeneity. Mutations in eight genes have so far been reported in cases and families with myopathies with TAs. No gene has been identified for the current reported 18 cases with CSs. Chapter 5 reported a WES project on the largest cohort of patients with myopathies with TAs and CSs to date with the aim of identifying their disease-causing genes and further improving the knowledge of the possible pathogenesis of these disorders. Both TAs and CSs are considered to originate from components of SR. Studies have also proposed that abnormalities of intracellular calcium homeostasis, and defects in N-linked protein glycosylation pathway may play an important role in the formation of

TAs. Therefore, this project was mainly focused on the known genes and the candidate genes proposed to be related to these two pathways.

A total of 33 genetically undiagnosed unrelated patients with TAs and two unrelated patients with CSs were selected for WES after carefully reviewing their muscle biopsy reports or available biopsy slides. A candidate gene searching strategy (Figure 5-1) was used to filter for the non-synonymous rare variants (MAF < 1%) in genes listed in Table 5-1. Rare variants in known or candidate genes were found in 42.9% of cases in this cohort. Among these, three known mutations and five novel variants in four known genes (*STIM1*, *ORAI1*, *PGAM2*, and *SCN4A*) were suggested to be pathogenic to the phenotypes in patients from two families with TAM and six unrelated patients with TAs. This constituted 22.9% of our cohort. In addition, rare variants in three candidate genes (*ALG14*, *CASQ1*, and *ATP2A1*) were also found in this cohort. Compound heterozygous variants (p.Arg104Ter is known, and p.Ala11Thr is novel) in *ALG14* were found in one patient with complex phenotypes and with TAs in muscle. Whilst the p.Ala11Thr variant was predicted to be benign, it was suggested to be pathogenic by a subsequent functional analysis. This is the first time that pathogenic variants in *ALG14* were reported in patients with TAs and also were suggested to be the potential culprit in the disease. One possible damaging heterozygous missense variant in *CASQ1* was found in one patient. However, its pathogenicity to the phenotype still requires further investigation. Compound heterozygous variants (p.Arg637Trp and p.Gln920Ter) and a homozygous missense variant (Pro540Leu) in *ATP2A1* were found in two unrelated patients with CSs. In contrast to TAs, CSs were found to be highly immunoreactive for the longitudinal SR protein SERCA1 (*ATP2A1*) but not for junctional SR proteins (Xu et al., 2015). This is the first time that rare variants in *ATP2A1* were identified in patients with CSs, though the exact mechanisms by which these rare variants in *ATP2A1* cause disease remain unknown.

This is the largest ever cohort of myopathies with TAs and CSs that have been collected and studied to date. Our findings not only expand the knowledge surrounding the underlying genetic basis of myopathies with TAs and CSs, but also expand the clinical phenotypic spectrum associated with some of the genes discussed. Our study reported pathogenic or likely pathogenic rare variants in both calcium

related genes and N-linked glycosylation related genes. These findings provide further evidence for the previous hypothesis that the formation of TAs and CSs are likely secondary to disordered biological mechanisms underlying these two pathways. However, the exact mechanisms of how these variants cause disease and lead to the formation of TAs/CSs remains unknown and requires further investigation. In addition, there are still 20 cases in which variants in either known genes or candidate genes were not found. There may be other genes and different mechanisms causing the phenotypes in these patients. However, most of them lacked information and samples from their affected/unaffected family members, or had incomplete clinical information about the patients themselves. This also raises a major challenge in interpreting the variants found in patients and determining whether the genetic findings were responsible for the phenotypes when using WES in small families or sporadic cases with complex phenotypes. It also becomes more difficult when multiple genes are involved in these disorders.

Despite the challenges and limitations in interpreting the data in this project, WES has shown to be a promising tool in studying the genetic basis for patients with myopathies with TAs and CSs. This is a more efficient method than the traditional techniques, especially in combination with a candidate gene screening approach.

7.3 Genetic Investigations on Four Families with Myopathies with Protein Aggregates

Chapter 6 continued to use WES to study four genetically undiagnosed families with myopathies with protein aggregates or myotubular myopathy. The aim was to identify the genetic culprit and provide a genetic diagnosis for these families, as well as to show the benefit of applying WES in studying the rare muscular conditions.

Family D had two siblings who presented with late-onset (around 50s) progressive muscle weakness who were both found to harbour cytoplasmic bodies in their muscle biopsies. The two siblings were suggested to be affected by a subtype of myofibrillar myopathy (MFM). After WES of the index patient, a candidate gene search strategy was used to filter rare variants in six established genes associated with MFM (Table 1-8). A homozygous missense variant p.Ser60Phe in exon 2 of the *MYOT* gene was

found and also segregated in the other sibling. This is a previously known mutation that was reported as heterozygosity in a patient in his 80s. The mutation, which was located in the hotspot of mutations in *MYOT*, was suggested to affect the hydrophobic stretch of myotilin, thus altering its function. Functional analysis on the mechanism of this mutation was not performed in this study. However, in combination with the clinical phenotypes and pathological features, as well as the genetic findings of these two siblings, the mutation p.Ser60Phe in *MYOT* is very likely to be the culprit of the disease. Their earlier age of onset is probably due to the impact of the homozygosity.

Family E had three siblings affected with progressive muscle weakness in distal limbs. Index patient (E-II-2) and his affected sister (E-II-4) were seen previously by neurologists at NHNN. Muscle biopsy of the index patient showed myopathic changes with rimmed vacuoles, no inflammation, and p62 positive inclusions in fibres. The sister's biopsy showed some neurogenic changes. Previous genetic tests on both patients were negative for spinocerebellar ataxia genes and common mtDNA mutations. The index patient in Family F presented with sIBM-like symptoms of progressive muscle weakness particularly in finger flexors and knee extensors. His muscle biopsy was also suggestive of IBM and also contained p62 positive inclusions. Both his father and sister were affected, but only the DNA sample of the index patient was available. Patients from both of the two families had a diagnosis suggestive of a form of hIBM. Therefore, a candidate gene search strategy was also used for investigating genes associated with hIBM and IBM-like vacuolar myopathies. A heterozygous missense variant p.Arg93Cys in *VCP* was found in both siblings in Family E, and a different heterozygous missense variant p.Gly125Asp in *VCP* was found in the index patient of Family F. The p.Arg93Cys variant is a known pathogenic mutation previously reported in a family with IBMPFD, while the p.Gly125Asp variant is novel and predicted as deleterious. Samples from the affected father and sister in Family F are required for segregation analysis, and further functional tests on the pathogenicity of the p.Gly125Asp variant should be performed in the future.

Family G had two siblings affected with a rare form of autosomal recessive myotubular myopathy. The index patient presented with unusually neuropathic and

myopathic phenotypes with syndactyly at birth and abnormal pupils. In addition, appearances of necklace fibres were observed in his muscle biopsy, combined with other features such as myopathic changes and prominent internal nuclei. Previous genetic investigations were all negative. His sister had similar pupils to the index patient. No genetic investigation was previously performed on her. After WES on the index patient, as the family demonstrated an autosomal recessive inheritance, the filtering strategy for non-synonymous rare variants (MAF < 1%) was applied for genes with homozygous variants or with compound heterozygous variants. In addition, a candidate gene search strategy for myotubularin-related genes was also used. These approaches found a novel homozygous frameshift deletion p.1826-1826del in exon 40 of the *SBF1* gene, and compound heterozygous missense variants p.Leu1098Val and p.Arg1145Gly in exon 26 of the *SBF2* gene. Segregation analysis showed that the affected sister also had a homozygous p.1826-1826del in *SBF1*, but no variants in *SBF2*. Their unaffected parents and the other healthy sibling were carriers of a heterozygous variant in two genes. Subsequently, a markedly reduced level of MTM5 protein (*SBF1*) was found in the patient with homozygous deletion compared with both heterozygous parents and normal controls by western blot analysis. This suggests that the homozygous deletion in *SBF1* causes the disease by a loss-of-function mechanism. This is the first report of necklace fibres in *SBF1*-related phenotypes, and also the second family whose affected family member with a defect in the *SBF1* gene had syndactyly. The mechanism of the compound heterozygous variants in *SBF2* is unclear. They may act as contributing factors to affect the function of the protein, which may explain why the index patient had more severe symptoms than his sister.

In this chapter, WES successfully identified the pathogenic genes for the four families. This aids genetic counselling to help patients coping with the disease. WES in combination with a candidate gene search strategy proved that it is a more time and cost efficient method than Sanger sequencing in identifying the genetic culprit for those genetically undiagnosed cases with rare muscular conditions. However, analysing WES data in a small family can sometimes be extremely challenging as also discussed in Chapters 5. It can be an unfruitful task when facing a large number of rare variants from WES without adequate numbers of family members. For this reason, many cases that underwent WES remained without a genetic diagnosis. We

were lucky to identify variants in known genes in all these four families using a candidate gene search approach, and had samples from more than one affected/unaffected family members with the exception of Family F. If there is a novel disease-causing gene for a family, sequencing more family members would be necessary, and a linkage analysis may be required to narrow down the number of candidate genes if it is a large pedigree. Therefore, a successful discovery of a culprit gene usually relies on the availability of samples from multiple members of the family, accurate family pedigree with detailed clinical information, and sufficient funding support. The genetic variants can be predicted for pathogenicity by *in silico* tools such as SIFT and PolyPhen. In addition, functional investigations are necessary to prove the genetic findings, as the functional results can sometimes contradict the prediction.

7.4 Future Work

The sample collection for the IBM genetic consortium study is still ongoing. An additional 200 IBM DNA samples is under WES at the moment. These will add to the current cohort to increase the study power. Both the single variant and gene-based association analyses will be repeated in this even larger cohort. In addition, mitochondrial sequencing can be considered in parallel with WES, as defects in mtDNA have been reported in several studies (Rygiel et al., 2016, Catalan-Garcia et al., 2016). It is also worthwhile to combine this with other techniques, such as genotyping, haplotyping for MHC genes, fragment analysis for repeat variants, and comparative genomic hybridization (CGH) analysis for copy number variants which cannot be detected by WES. Furthermore, RNA sequencing in muscle tissue and lymphoblast lines can be applied to identify transcriptomic profiles and expression quantitative trait loci (eQTL) in sIBM and other inflammatory myopathies. These will allow us to study and dissect splicing abnormalities which may not be detected by exon arrays. Investigating the transcriptomic and genomic factors will be important in revealing dysregulated molecular pathways, identifying a diagnostic signature and discovering biomarkers for future therapeutic strategies. A searchable database for the research community is also planned. This would integrate WES, RNA sequencing and eQTL data, further aligned with the biobank database with disease progression, response to treatment, and muscle imaging.

For the project on myopathies with TAs and CSs, it is important to perform further functional work to confirm the genetic findings from our cohort. It would be worthwhile to continue to acquire more information on the patients and their families for those unsolved cases if possible, and investigation in other muscle genes is needed. Meanwhile, more cases should be studied to confirm our findings or to provide new insights into our knowledge of these rare muscle conditions.

Despite the success in applying WES, a thorough understanding of the genetic background of a disease can be a result of a combination of both traditional techniques and the next-generation sequencing technologies. This is because WES only covers 1-2% of the entire human genome, and so will not be able to detect non-coding regions, large indels, duplications, translocations, and also most of the mtDNA sequences. With the continuing reduction in the cost of whole-genome sequencing (WGS), there will be a shift from WES to WGS in the near future. Whilst the challenges of storing and interpreting data will increase, the aim is to decode the genetic basis of all Mendelian traits as well as complex diseases. This knowledge will aid clinicians and scientists in understanding the molecular process underlying these diseases, and bring insights into disease prevention, novel diagnostic methods, and discovery of novel therapeutic targets, and eventually improve the life of human beings.

Appendices

Appendix I for Chapter 2

Table 1. International IBM Genetic Study – Study Profoma (Designed by Dr Pedro Machado)

1. Study ID number: _____
2. Hospital: _____
3. Patient name: _____
4. Patient Hospital No: _____ 5. Sex (M = male / F = female): _____
6. Patient DOB (dd/mm/yyyy): ____ / ____ / _____ 7. Collection date (dd/mm/yyyy): ____ / ____ / _____
8. Race (WHH = White Hispanic, WHO = Other White, BLA = Black African, BLC = Black Caribbean, ACH = Asian Chinese, AOT = Other Asian, IND = Indian Subcontinent, OTH = Other, please specify): _____

9. Approximate date of onset of IBM symptoms (mm / yyyy): ____ / _____
10. Approximate date of definite diagnosis of IBM (mm / yyyy): ____ / _____
11. Family history of neuromuscular disease (Y = yes, N = no): _____
    a) if yes, please *specify* which relative and which disease: _____

12. Initial symptoms/signs (Y = yes, N = no):
    a) Swallowing problems: ____
    b) Proximal upper extremity involvement: ____
    c) Distal upper extremity involvement: ____
    d) Proximal lower extremity involvement: ____
    e) Distal lower extremity involvement: ____

13. Clinical features (Y = yes, N = no):
    a) Slowly progressive course: ____
    b) Sporadic disease: ____
    c) Weakness of proximal and distal muscles of arms and legs: ____
    d) Finger flexor weakness: ____
    e) Quadriceps muscle weakness (MRC ≤ 4): ____
    f) Weakness of knee extension > hip flexion: ____
    g) Weakness of finger flexion > shoulder abduction: ____
    h) Weakness of wrist flexors > than wrist extensors: ____

14. Biopsy features (Y = yes, N = no, NA = not assessed/not reported):
    a) Endomysial exudate: ____
    b) MHC-I upregulation: ____
    c) Partial invasion (infiltration of non-necrotic fibres by mononuclear cells): ____
    d) Rimmed vacuoles: ____
    e) Intracellular amyloid deposit (Congo-red or crystal violet): ____
    f) Hyperphosphorylated tau (SMI-31 immunoreactivity) : ____
    g) p62 immunoreactivity: ____
    h) TDP-43 immunoreactivity: ____
    i) 15-18 (or 16-21) nm filaments (EM): ____
    j) Cytochrome oxidase deficient fibres (abnormal for age): ____

15. Diagnostic criteria used to classify the patient (Y = yes, N = no):
    a) Griggs Definite: ____ Griggs Possible: ____
    b) European Neuromuscular Centre (ENMC) Definite: ____ ENMC Probable: ____
    c) MRC Pathologically Defined: ____ MRC Clinically Defined: ____ MRC Possible: ____

16. Creatine kinase (NR = normal, AB = abnormal, UK = unknown):
    a) Creatine kinase: _____, highest value if abnormal (IU/L): _____

Table 2. Sequences of the primers involved in this thesis

Gene and Exon	Forward Primer	Reverse Primer
APP Exon 16	TTGGAACAAAGCCCCAAAGTAG	GGCAAGACAAACAGTAGTGGAAA
APP Exon 17	GACCAACCAGTTGGGCAGAG	CATGGAAGCACACTGATTCG
SQSTM1 Exon 1	GCAATGAAGAGAGGGGTCAG	GTCGCCGAGAAGGGAGG
SQSTM1 Exon 2	TGTCCCTTTCATACTTGCCTC	AATTGCTGACCCCTTCATTG
SQSTM1 Exon 3-4	GTGGATTCCATGCTGGAGAG	ATCCTGAATTCTTGCCTTGC
SQSTM1 Exon 5	CAAGATGTCCGGGTTAAAGG	ATCCTCACCAGTGTCAAGGG
SQSTM1 Exon 6	TAGCTGCTTGTGGGGACTG	AGCTCCCTCGGGTTTGTAAG
SQSTM1 Exon 7	GTGCCATTAAAGTCACGCTG	ATTCCTGGCTCTGGAGTTTG
SQSTM1 Exon 8	CAGGGTATGTGTTTCGGGTC	ACCTGCAATTCTACGCAAGC
VCP Exon 1	CAGGAGCCAGCGTTGTTC	TCCTGGTCTCCACCTCTCTG
VCP Exon 2-3	CAAGAACTTGGTCCTGCCTG	GCTTTCTGGTCTAGGGACAGC
VCP Exon 4	TCACTTCCACCCTGTTCTCC	CCCAGCTGACCCCATCTC
VCP Exon 5	TTAAGACAGGTGGGGTGGAG	CCCATCTCAGTCTCCCAAAG
VCP Exon 6	CCTGGACTCAAGCATTCCTC	ATTGCCCCTCTAATCCAAGG
VCP Exon 7	TTACTGCGGAGGGCTTATTC	CCACAATCTTTCTCAAAGGGG
VCP Exon 8-9	CTGTCCTGGGATAATGGCTG	GTCTCAGGAAAACCAGGGTG
VCP Exon 10	AACCACCCTGGTTTTCCTG	AAATGCCAACTCCCATTTCC
VCP Exon 11-12	CCCAGCCCATAAAGGTATTG	CCAATGGTATCAGATCCAGG
VCP Exon 13	GGTTTGAGGCACTAAGGAGTC	GAGCCAGACCCTGTCTCAAG
VCP Exon 14	ATGCTGGTTTCGGATTTCTG	CTCCCAACTACAGTTTGCCC
VCP Exon 15	GGGTTGGTCTAAAGGGAAGG	TTGAATGGATTCACCTCAGC
VCP Exon 16	CCTTTTCCAGAGTGCATTGAC	AAAGAGGGTTAGGACAGGGC
VCP Exon 17-1	CAGGTAAGTTGGTTGGGAGC	TGCAGATGCTTTACTGTGGC
STIM1 Exon 1	CTCCTGGCTTTGCCTCTG	TTTCAACTTGCCCACTTCG
STIM1 Exon 2	TGCTAGAAGCTAAGGATGCTG	GGGAAGAGGCTGTCTAAGTAGC
STIM1 Exon 3	ATGTTGGGTAGATGGAATGTG	TTCTTCTAAGGCCAAGTTGC
STIM1 Exon 4	TCAGCATGACAACAAATGAAAG	TAGACGCTGCCCTGAAAAAG
STIM1 Exon 5	GGGGAGCAATCACCAAGAG	AGGAGACCCAGAGCCTTAGC
STIM1 Exon 6	TGGGAGAGTGTGTCTGTTATGG	GCTGCCATGCTGACAAGG
STIM1 Exon 7	GGAGCTGTCATTTTCCTCTTTG	TGCCTGGCCCATAGTGAC
STIM1 Exon 8	CCTGGGAGAGTTGTAAAGCAG	AATGGGAGGAACAGGGAGAG
STIM1 Exon 9	CAGTTGTGCTAGGAGGAGTGG	GGAAATGGAATTGGTGTTGG
STIM1 Exon 10	CACATATTCTCAAAACTTGTTCCTC	AGGCCACAGACTTCTCAACC
STIM1 Exon 11	TCTCCAGATTGGCATTAGAGG	TGGTCCCGTAGCCTTCAG
STIM1 Exon 12-1	CTCGTGTTGTCCCTCTCTCC	AACTGGAGATGGTGTGTCTGG
STIM1 Exon 12-2	GGCATTACTGGCGCTGAAC	AATGAAGGACCCAAGGATGAG
ORAI1 Exon 1	CAGCAGTCCCGAGCTTCC	CAGCTCTTCCCCTCACTCTG

ORAI1 Exon 2	AGGGTCCCAGAGCCTGAC	GCCTGGCTGCTTCTTGAG
ALG14 Exon 1	CCAGCTCAGGAAAGACCAG	CGGCAAGGCAGAGAAACTAC
ALG14 Exon 2	TGTCCATTGTGGGTTTCTTAATC	GAGACAATTAAACTCAATGTGCA
ALG14 Exon 3	AAGCATGAGATAGGCATTACAACTC	CCTGGCACACTAATAAATGGTAGC
ALG14 Exon 4	ATGCAAAATGCCTCCCTTAG	TCAGACGCCTTTACAAGAAAC
ATP2A1 Exon 14	GAAGGAAAGTGGTGGTCTCTG	CACCAGCTCCACGACAGTG
ATP2A1 Exon 15	CTGACCTTTCACCCCATCC	CTCTCCCCAGCACAGTCATC
ATP2A1 Exon 20	CACCTCCTTCCTCCTCACTG	GGAGCCAGAGACCAGCAC
PGAM2 Exon 1	AATGCTGATTGGCAGTTGG	AGCACTGACGTGTAGCAGATG
PGAM2 Exon 1_2	GCCATCAAGGATGCCAAG	GCTGCTTTCCCTCCCAAG
PGAM2 Exon 2	GCTGCCTTTGCTGGGAAG	ACGCCTAGAAAGCCTGGTC
PGAM2 Exon 3	AAACAGGCTGGGGACAGAG	GGGGGAGGTGCCTTTATTG
MYOT Exon 2	TCCGTAATTCCAGGCTTGTG	TACCCCCAAAACTCCTACCC
DPAGT1 Exon 1	GCCACTCCATACTCTGAGG	CTCCTTCCTTAGCCCTTGC
DPAGT1 Exon 2-3	ACGCCCCCTTTCCTCTTC	CCTGAAGTAGGTGCCATAGG
DPAGT1 Exon 4	AGGTGGCACTACTTCTTTTCC	CAGGCCACACCTTTAACACTC
DPAGT1 Exon 5	AAGGTTTGCAGAGGAAGTGG	ATGATAAGGGCCATCTTTGC
DPAGT1 Exon 6	GGATTGGTAATGCCTCTTGC	TGGAAATAGCCCTTCTTTGG
DPAGT1 Exon 7-9	CAGGGAGGAGGGTTCAAATAG	TGGGACCAGAGAGAGAGGTC
SCN4A Exon 9	ATGTATGGAAAGGGGCACTG	GGCTGAGGGCAGGTAGAAC
CASQ1 Exon 11	CAGAGTCCAGTGAGTGGGAAG	AGACCAGTGGGGCTGAAAG
SBF1 Exon 40	GGACAAGACCAAGCACCAG	GAAGGGCGGGTTACTGACTC
SBF2 Exon 26	ACCCAGCCCAAAAGGAAG	CAGGGCTATGGATGCAAATG
APOE Genotyping	TAAGCTTGGCAGGGCTGTCCAAGGA	ACAGAATTCGCCCCGGCCTGGTACAC
MAPT Genotyping	GGAAGACGTTCTCACTGATCTG	AGGAGTCTGGCTTCAGTCTCTC
PRNP Genotyping	TCAGTGGAACAAGCCGAGTAAG	CATAGTCACTGCCGAAATGTATGAT
PRNP Probes	Methionine: CGGCTA CATGCTGG	Valine: CGGCTACGTGCTGG
C9orf72 Exon 1a	6FAM-CAAGGAGGGAAACAACCGCAGCC	R1: CAGGAAACAGCTATGACCGG GCCCGCCCCGACCACGCCCCGGCC CCGGCCCCGG R2: CAGGAAACAG CTATGACC
TOMM40 Intron 6	FAM-TGCTGACCTCAAGCTGTCCTCTTGC	GAATTGGGGAGGAAGAGAGTGGGAAA
mtDNA 11kb fragment	L5647: TAAGCCCTTACTAGACCAATGGGAC	H607: GTATTGCTTTGAGG AGGTAAG
mtDNA 5kb fragment	L15788: ACCATCATTGGAC AAGTAGC	H5925: ATAGTCAACGGTCGGCGA ACATCAG

HLA-A (cDNA)	CTGAGATGGGAGCTGTCTTC	CTATCTGAGCTCTTCCTCCT
HLA-DRA (cDNA)	AGGCCGAGTTCTATCTGAATCCT	CGCCAGACCGTCTCCTTCT
CD74 (cDNA)	CAGGAAGAGGTCAGCCACATC	GGGAAGACACACCAGCAGTAG
PPIA (cDNA)	CCCACCGTGTTCTTCGACAT	CCAGTGCTCAGAGCACGAAA
SBF1 (cDNA)	CCAGACAGCAGAGAGTGAGAA	CGACAGGCAGCTCTGGAT
SBF2 (cDNA)	AAATCGTCCTGGATGGAATG	CAGGATAGCTCCGGCAGA

Script section – R scripts for limma Bioconductor package (kindly provided by Dr Conceição Bettencourt)

```
#Open the library
library(limma)

#First read the targets file describing the target RNA samples.
targets <- readTargets()

#We read in the expression profiles for both regular and control probes.
x <-
read.ilmn(files="BettencourtSampleProbesIBM_260115.txt",ctrlfiles="BettencourtC
ontrolProbesIBM_260115.txt", probeid="Probe", annotation=c("PROBE_ID",
"SYMBOL", "DEFINITION"), expr="AVG_Signal",other.columns="Detection")

# This reads a EListRaw object.
table(x$genes$Status)

#The component E contains the expression value for each probe.
options(digits=3)
head(x$E)
head(x$genes)
head(x$other)

#The intensities on the log2 scale:
boxplot(log2(x$E),col="red", range=0,ylab="log2 intensity",
xlab="SampleID",cex.axis=0.6)

#The detection values contain p-values for testing whether each probe is more
intense than the negative control probes. Small values are evidence that the probe
corresponds to a truly expressed gene:
head(x$other$Detection)

#Estimate the overall proportion of the regular probes that correspond to expressed
transcript.
pe <- propexpr(x)
mean(pe)
range(pe)
write.csv(pe, file= "prop_expressed_probes.txt")

#The neqc functions performs normexp background correction using negative
controls, then quantile normalizes and finally log2 transforms.
y <- neqc(x)
dim(y)
write.csv(y, file= "nomalizedallprobes.txt")
```

```
#The normalized intensities on the log2 scale.
boxplot(y$E,col="blue",range=0,ylab="log2 intensity",
xlab="SampleID",cex.axis=0.6)

#Filter out probes that are not expressed. We keep probes that are expressed in at least three arrays according to a detection p-values of 5%.
expressed <- rowSums(y$other$Detection < 0.05) >= 3
z <- y[expressed,]
dim(z)
write.csv(z, file= "nomalizedexpressedprobes.txt")

#The intensities on the log2 scale.
boxplot(z$E,range=0,ylab="log2 intensity", xlab="SampleID",cex.axis=0.6)

#A multi-dimensional scaling plot.
plotMDS(z,labels=targets$Disease, cex=0.6)
plotMDS(z,labels=targets$SampleID, cex=0.5)
plotMDS(z,labels=targets$Disease1, cex=0.5)

#Now we look for diferentially expressed genes; to collect the disease/mutant gene combinations into one vector as follows.
DT <- paste(targets$Disease, targets$Disease1, sep = ".")
DT <- factor(DT, levels = c("IBM.SQSTM1","IBM.VCP","CTRL.CTRL"))
design <- model.matrix(~0 + DT)
colnames(design) <- levels(DT)
fit <- lmFit(z, design)

# Comparisons
cont.matrix <- makeContrasts(IBM_SQSTM1vsCTRL = IBM.SQSTM1 -
CTRL.CTRL, IBM_VCPvsCTRL = IBM.VCP - CTRL.CTRL, levels = design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)
topTable(fit2, coef=1, adjust="fdr", sort.by="B")
topTable(fit2, coef=2, adjust="fdr", sort.by="B")
write.table(topTable(fit2, coef=1, adjust="fdr", sort.by="B", number=50000),
file="limma_CTRLvsIBM_SQSTM1.xls", row.names=F, sep="\t")
write.table(topTable(fit2, coef=2, adjust="fdr", sort.by="B", number=50000),
file="limma_CTRLvsIBM_VCP.xls", row.names=F, sep="\t")
write.fit(fit2, file="limma_complete.xls", digits=3, adjust="fdr", method="separate",
sep="\t")
results <- decideTests(fit2)
vennDiagram(results, cex=0.8)
write.csv(results, file = "results.txt")

#Paiwise expression correlation plots.
pairs(z$E, z)

#Complete list of genes with p-values and fold change.
```

```
#Coef=1, so we are just looking at SQSTM1 cases
gene_list_SQSTM1 <- topTable(fit2, coef=1, adjust="fdr", number=1000000,
sort.by="B")

##Coef=2, so we are just looking at VCP cases
gene_list_VCP <- topTable(fit2, coef=2, adjust="fdr", number=1000000,
sort.by="B")
head(gene_list_SQSTM1$logFC)
head(gene_list_SQSTM1$adj.P.Val)
head(gene_list_VCP$logFC)
head(gene_list_VCP$adj.P.Val)
```

#To determine the number of genes that have an absolute log fold change greater than 0.5 and a p-value less than 0.05 can be found using the command.

```
sum(abs(gene_list_SQSTM1$logFC) > 0.2 & gene_list_SQSTM1$adj.P.Val < 0.05)
sum(abs(gene_list_VCP$logFC) > 0.2 & gene_list_VCP$adj.P.Val < 0.05)
```

#To look for differentially expressed genes IBMallvsCTRL; to collect the Disease combinations into one vector as follows.

```
DT <- paste(targets$Disease)
DT <- factor(DT, levels = c("IBM","CTRL"))
design <- model.matrix(~0 + DT)
colnames(design) <- levels(DT)
fit <- lmFit(z, design)

#Comparisons
cont.matrix <- makeContrasts(IBMvsCTRL = IBM - CTRL, levels = design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)
topTable(fit2, coef=1, adjust="fdr", sort.by="B")
write.table(topTable(fit2, coef=1, adjust="fdr", sort.by="B", number=50000),
file="limma_IBMallvsCTRL.xls", row.names=F, sep="\t")
write.fit(fit2, file="limma_complete_IBMall.xls", digits=3, adjust="fdr",
method="separate", sep="\t")
results <- decideTests(fit2)
vennDiagram(results, cex=0.8)
```

Appendix II for Chapter 3

Table 1. All the participants of the IIBMGC and the Muscle Study Group

Institutes/ Research centres	Participants
Department of Molecular Neuroscience, and MRC Centre for Neuromuscular Diseases, Institute of Neurology, University College London, Queen Square, London, UK	Michael G. Hanna, Henry Houlden, Pedro M. Machado, Qiang Gang, Conceicao Bettencourt, Estelle Healy, Matt Parton, and Janice L. Holton, Iwona Skorupinska
Nuffield Department of Clinical Neurosciences, University of Oxford, UK	Stefen Brady and David Hilton-Jones
Neuromuscular Division, Department of Neurology, UCLA Medical Centre, USA	Perry B. Shieh
Department of Neurology, Medical School of the University of São Paulo (FMUSP), São Paulo, Brazil	Edmar Zanoteli and Leonardo Valente de Camargo
Ghent University Hospital, Department of Neurology and Neuromuscular Reference Centre, Ghent, Belgium	Boel De Paepe and Jan De Bleecker
Nerve and Muscle Centre of Texas, Houston, Texas, USA	Aziz Shaibani, Laura Shawver
Neuromuscular Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Dino Ferrari Centre, University of Milan, Milan, Italy	Michela Ripolone, Raffaella Violano, and Maurizio Moggio
The University of Kansas Medical Centre, Kansas City, USA	Richard J. Barohn, Mazen M. Dimachkie, April L. McVey, Mamatha Pasnoor, Melanie Glenn, Omar Jawdat, Jeffrey Statland, and Gabrielle Rico
Neuromuscular Diseases and Neuroimmunology Unit, Fondazione IRCCS Isitituto Neurologico C. Besta, Milano, Italy	Marina Mora, Renato Mantegazza, and Simona Zanotti
Western Australian Neurosciences Research Institute (WANRI), University of Western Australia and Murdoch University, Fiona Stanley Hospital, Perth, Australia	Merrilee Needham and Frank Mastaglia
Royal North Shore Hospital, New South Wales, Australia	Christina Liang
Neuroimmunology Unit , Department of Pathophysiology, National University of Athens Medical School, Athens, Greece	Marinos C. Dalakas and Angie Biba
Brigham & Women's Hospital, Harvard, USA	Amato Anthony
Centre for Musculoskeletal Research, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK	Hector Chinoy and James B. Lilleker
Centre for Integrated Genomic Medical Research, Institute of Population Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK	Janine Lamb and Hazel Platt
MRC/ARUK Institute for Ageing and Chronic Disease, University of Liverpool, Liverpool, UK	Robert G. Cooper
Department of Neurology, Newcastle Upon Tyne Hospitals NHS Trust, UK	James A.L. Miller, Alison Sutherland, and Lilian Fairbairn-smith
Salford Royal NHS Foundation Trust, UK	Mark Roberts
The Ohio State University Wexner Medical Center, USA	Amy Bartlett and John T. Kissel
Penn State Milton S. Hershey Medical Center, USA	Heidi Runk and Matthew Wicklund
University of California, Irvine, USA	Tahseen Mozaffar, Marie Wencel
Derriford Hospital, Plymouth, UK	Elizabeth Househam, David Hilton, and Aditya Shivane

Table 2. Up/down-regulated genes in the microarray analysis of all the sIBM vs controls

PROBE_ID	SYMBOL	DEFINITION	logFC	P.Value	adj.P.Val
ILMN_1700316	*LOC440055*	PREDICTED: Homo sapiens similar to ribosomal protein S12 (LOC440055), mRNA.	0.265176	1.20E-05	0.023722
ILMN_1704315	*LOC389435*	Homo sapiens hCG21078 (LOC389435), mRNA.	0.570352	1.77E-05	0.023722
ILMN_2116366	*RPL12*	Homo sapiens ribosomal protein L12 (RPL12), mRNA.	0.396815	3.66E-05	0.036006
ILMN_1685722	*EIF4A2*	Homo sapiens eukaryotic translation initiation factor 4A, isoform 2 (EIF4A2), mRNA.	-0.42772	9.79E-05	0.047704

Table 3. Up/down-regulated genes in the microarray analysis of the *SQSTM1* sIBM patient group vs controls

PROBE_ID	SYMBOL	DEFINITION	logFC	P.Value	adj.P.Val
ILMN_2203950	*HLA-A*	Homo sapiens major histocompatibility complex, class I, A (HLA-A), mRNA.	1.697545	8.21E-09	5.93E-05
ILMN_1800354	*CST3*	Homo sapiens cystatin C (CST3), mRNA.	1.345043	5.28E-06	0.003326
ILMN_1736567	*CD74*	Homo sapiens CD74 molecule, major histocompatibility complex, class II invariant chain (CD74), transcript variant 1, mRNA.	1.33614	4.33E-06	0.002934
ILMN_1725427	*B2M*	Homo sapiens beta-2-microglobulin (B2M), mRNA.	1.070304	1.40E-08	7.60E-05
ILMN_2152131	*ACTB*	Homo sapiens actin, beta (ACTB), mRNA.	0.999226	0.000573	0.028821
ILMN_2038777	*ACTB*	Homo sapiens actin, beta (ACTB), mRNA.	0.982321	0.000867	0.035755
ILMN_1343291	*EEF1A1*	Homo sapiens eukaryotic translation elongation factor 1 alpha 1 (EEF1A1), mRNA.	0.934575	0.00017	0.015975
ILMN_2157441	*HLA-DRA*	Homo sapiens major histocompatibility complex, class II, DR alpha (HLA-DRA), mRNA.	0.918773	0.0002	0.016682
ILMN_1812392	*TMSB10*	Homo sapiens thymosin beta 10 (TMSB10), mRNA.	0.897223	0.001149	0.041167
ILMN_1683271	*TMSB4X*	Homo sapiens thymosin, beta 4, X-linked (TMSB4X), mRNA.	0.860586	0.000178	0.016206
ILMN_2130441	*HLA-H*	Homo sapiens major histocompatibility complex, class I, H (pseudogene) (HLA-H), non-coding RNA.	0.730665	7.95E-06	0.003728
ILMN_2379644	*CD74*	Homo sapiens CD74 molecule, major histocompatibility complex, class II invariant chain (CD74), transcript variant 2, mRNA.	0.704286	0.000201	0.016701
ILMN_3256742	*LOC100129902*	PREDICTED: Homo sapiens similar to mCG7602 (LOC100129902), mRNA.	0.676803	0.001328	0.04352
ILMN_1778401	*HLA-B*	Homo sapiens major histocompatibility complex, class I, B (HLA-B), mRNA.	0.664872	3.43E-05	0.007274
ILMN_1704315	*LOC389435*	Homo sapiens hCG21078 (LOC389435), mRNA.	0.613575	5.44E-05	0.009308
ILMN_1689655	*HLA-DRA*	Homo sapiens major histocompatibility complex, class II, DR alpha (HLA-DRA), mRNA.	0.611838	0.00037	0.022843

ILMN_2148459	*B2M*	Homo sapiens beta-2-microglobulin (B2M), mRNA.	0.567881	0.000111	0.0124
ILMN_1713086	*RPL27A*	Homo sapiens ribosomal protein L27a (RPL27A), mRNA.	0.565704	0.000731	0.032647
ILMN_2046730	*S100A10*	Homo sapiens S100 calcium binding protein A10 (S100A10), mRNA.	0.531182	0.001396	0.044466
ILMN_1765258	*HLA-E*	Homo sapiens major histocompatibility complex, class I, E (HLA-E), mRNA.	0.463181	0.001422	0.044605
ILMN_1772218	*HLA-DPA1*	Homo sapiens major histocompatibility complex, class II, DP alpha 1 (HLA-DPA1), mRNA.	0.452916	0.000121	0.012932
ILMN_2393765	*IGLL1*	Homo sapiens immunoglobulin lambda-like polypeptide 1 (IGLL1), transcript variant 1, mRNA.	0.433705	9.56E-05	0.011618
ILMN_1752592	*HLA-DRB4*	Homo sapiens major histocompatibility complex, class II, DR beta 4 (HLA-DRB4), mRNA.	0.416486	9.82E-05	0.011618
ILMN_2116366	*RPL12*	Homo sapiens ribosomal protein L12 (RPL12), mRNA.	0.416108	0.000174	0.016128
ILMN_3228688	*LOC730415*	PREDICTED: Homo sapiens hypothetical LOC730415, transcript variant 2 (LOC730415), mRNA.	0.403135	5.44E-05	0.009308
ILMN_1684306	*S100A4*	Homo sapiens S100 calcium binding protein A4 (S100A4), transcript variant 2, mRNA.	0.394127	0.000976	0.03811
ILMN_2058782	*IFI27*	Homo sapiens interferon, alpha-inducible protein 27 (IFI27), transcript variant 2, mRNA.	0.381146	3.56E-06	0.002658
ILMN_2409167	*ANXA2*	Homo sapiens annexin A2 (ANXA2), transcript variant 1, mRNA.	0.333895	0.001462	0.045365
ILMN_2053178	*ACTG1*	Homo sapiens actin, gamma 1 (ACTG1), mRNA.	0.333156	0.00131	0.043378
ILMN_2148819	*TUBA1A*	Homo sapiens tubulin, alpha 1a (TUBA1A), mRNA.	0.30505	4.54E-05	0.00845
ILMN_1796712	*S100A10*	Homo sapiens S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11)) (S100A10), mRNA.	0.277185	0.001248	0.042185
ILMN_1806017	*PSME1*	Homo sapiens proteasome (prosome, macropain) activator subunit 1 (PA28 alpha) (PSME1), transcript variant 1, mRNA.	0.269909	0.000101	0.011694
ILMN_1700316	*LOC440055*	PREDICTED: Homo sapiens similar to ribosomal protein S12 (LOC440055), mRNA.	0.255721	0.000123	0.012932
ILMN_2189936	*RPL36AL*	Homo sapiens ribosomal protein L36a-like (RPL36AL), mRNA.	0.242778	0.000293	0.020117
ILMN_1716678	*NPC2*	Homo sapiens Niemann-Pick disease, type C2 (NPC2), mRNA.	0.216139	4.41E-05	0.008414
ILMN_1810577	*RPS4X*	Homo sapiens ribosomal protein S4, X-linked (RPS4X), mRNA.	0.212198	0.000265	0.01925
ILMN_2038776	*TXN*	Homo sapiens thioredoxin (TXN), mRNA.	0.209643	1.27E-06	0.001451
ILMN_3241758	*POTEF*	Homo sapiens POTE ankyrin domain family, member F (POTEF), mRNA.	0.207325	0.001137	0.040977
ILMN_1794595	*GAMT*	Homo sapiens guanidinoacetate N-methyltransferase (GAMT), transcript variant 1, mRNA.	-0.28579	0.001056	0.039317
ILMN_2371825	*AGL*	Homo sapiens amylo-1, 6-glucosidase, 4-alpha-glucanotransferase (AGL), transcript variant 5, mRNA.	-0.29323	0.001208	0.041646
ILMN_1738383	*EEF2*	Homo sapiens eukaryotic translation elongation factor 2 (EEF2), mRNA.	-0.31977	0.001135	0.040962
ILMN_1687971	*CAPN3*	Homo sapiens calpain 3, (p94) (CAPN3), transcript variant 2, mRNA.	-0.45792	0.001588	0.046861

ILMN_1685722	*EIF4A2*	Homo sapiens eukaryotic translation initiation factor 4A, isoform 2 (EIF4A2), mRNA.	-0.45877	0.000314	0.021052
ILMN_1756469	*GAMT*	Homo sapiens guanidinoacetate N-methyltransferase (GAMT), transcript variant 1, mRNA.	-0.49518	0.000475	0.026259
ILMN_1769191	*GNAS*	Homo sapiens GNAS complex locus (GNAS), transcript variant 4, mRNA.	-0.56318	0.000495	0.026739
ILMN_1789950	*C20orf166*	Homo sapiens chromosome 20 open reading frame 166 (C20orf166), mRNA.	-0.56343	0.000102	0.01172

Table 4. Significantly enriched terms obtained by functional enrichment analysis of dysregulated genes in the *SQSTM1* sIBM patient group vs controls comparisons[a]

TERM ID	TERM TYPE	TERM GROUP	TERM NAME	GENE LIST	P-VALUE
Up-regulated genes selected for functional enrichment analysis					
GO:0001895	BP	13	retina homeostasis	*ACTB,B2M,ACTG1,POTEF*	0.00965
GO:0001906	BP	11	cell killing	*B2M,HLA-E,HLA-A,HLA-B*	0.0419
GO:0002376	BP	11	immune system process	*CD74,ACTB,PSME1,IGLL1,TXN,IFI27,B2M,ANXA2,ACTG1,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	0.00094
GO:0019882	BP	11	antigen processing and presentation	*CD74,PSME1,B2M,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	8.78E-08
GO:0048002	BP	11	antigen processing and presentation of peptide antigen	*CD74,PSME1,B2M,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	1.28E-08
GO:0002428	BP	11	antigen processing and presentation of peptide antigen via MHC class Ib	*B2M,HLA-E*	0.0498
GO:0002474	BP	11	antigen processing and presentation of peptide antigen via MHC class I	*PSME1,B2M,HLA-E,HLA-H,HLA-A,HLA-B*	2.01E-05
GO:0019883	BP	11	antigen processing and presentation of endogenous antigen	*CD74,B2M,HLA-E,HLA-A,HLA-B*	3.38E-08
GO:0002483	BP	11	antigen processing and presentation of endogenous peptide antigen	*B2M,HLA-E,HLA-A,HLA-B*	6.38E-06
GO:0019885	BP	11	antigen processing and presentation of endogenous peptide antigen via MHC class I	*B2M,HLA-A,HLA-B*	0.00149
GO:0019884	BP	11	antigen processing and presentation of exogenous antigen	*CD74,PSME1,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	3.36E-07
GO:0002478	BP	11	antigen processing and presentation of exogenous	*CD74,PSME1,B2M,HLA-DRA,HLA-E,HLA-*	2.44E-07

			peptide antigen	*A,HLA-DPA1,HLA-B*	
GO:0042590	BP	11	antigen processing and presentation of exogenous peptide antigen via MHC class I	*PSME1,B2M,HLA-E,HLA-A,HLA-B*	0.00027
GO:0002479	BP	11	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	*PSME1,B2M,HLA-E,HLA-A,HLA-B*	0.000209
GO:0002480	BP	11	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent	*B2M,HLA-E,HLA-A,HLA-B*	1.13E-06
GO:0002449	BP	11	lymphocyte mediated immunity	*CD74,B2M,HLA-E,HLA-A,HLA-B*	0.0384
GO:0031341	BP	11	regulation of cell killing	*B2M,HLA-E,HLA-A,HLA-B*	0.00314
GO:0031342	BP	11	negative regulation of cell killing	*HLA-E,HLA-A,HLA-B*	0.00307
GO:0002700	BP	11	regulation of production of molecular mediator of immune response	*CD74,B2M,HLA-E,HLA-A*	0.0296
GO:0002706	BP	11	regulation of lymphocyte mediated immunity	*B2M,HLA-E,HLA-A,HLA-B*	0.0403
GO:0002707	BP	11	negative regulation of lymphocyte mediated immunity	*HLA-E,HLA-A,HLA-B*	0.0362
GO:0006955	BP	11	immune response	*CD74,ACTB,PSME1,IGLL1,TXN,IFI27,B2M,ACTG1,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	2.13E-05
GO:0006952	BP	11	defense response	*CD74,ACTB,PSME1,CST3,TXN,IFI27,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000647
GO:0045087	BP	11	innate immune response	*ACTB,PSME1,TXN,IFI27,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000238
GO:0048584	BP	11	positive regulation of response to stimulus	*CD74,ACTB,PSME1,TXN,B2M,ACTG1,S100A4,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0285
GO:0034097	BP	11	response to cytokine	*CD74,IFI27,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0141
GO:0071345	BP	11	cellular response to cytokine stimulus	*CD74,IFI27,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0046
GO:0034341	BP	11	response to interferon-gamma	*B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0002
GO:0071346	BP	11	cellular response to interferon-gamma	*B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	9.11E-05
GO:0034340	BP	11	response to type I interferon	*IFI27,HLA-E,HLA-A,HLA-B*	0.0146
GO:0071357	BP	11	cellular response to type I interferon	*IFI27,HLA-E,HLA-A,HLA-B*	0.0139
GO:0002682	BP	11	regulation of immune system process	*CD74,ACTB,PSME1,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.043
GO:0002684	BP	11	positive regulation of immune system process	*CD74,ACTB,PSME1,B2M,ACTG1,HLA-DRA,HLA-*	0.000873

				E,HLA-A,HLA-DPA1,HLA-B	
GO:0002699	BP	11	positive regulation of immune effector process	*CD74,B2M,HLA-E,HLA-A,HLA-B*	0.00582
GO:0002705	BP	11	positive regulation of leukocyte mediated immunity	*B2M,HLA-E,HLA-A,HLA-B*	0.0168
GO:0002708	BP	11	positive regulation of lymphocyte mediated immunity	*B2M,HLA-E,HLA-A,HLA-B*	0.00729
GO:0002702	BP	11	positive regulation of production of molecular mediator of immune response	*CD74,B2M,HLA-E,HLA-A*	0.00505
GO:0050776	BP	11	regulation of immune response	*CD74,ACTB,PSME1,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000749
GO:0050778	BP	11	positive regulation of immune response	*CD74,ACTB,PSME1,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	3.97E-05
GO:0002821	BP	11	positive regulation of adaptive immune response	*B2M,HLA-E,HLA-A,HLA-B*	0.00818
GO:0002460	BP	11	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	*CD74,B2M,HLA-E,HLA-A,HLA-B*	0.036
GO:0002822	BP	11	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	*B2M,HLA-E,HLA-A,HLA-B*	0.0389
GO:0002824	BP	11	positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	*B2M,HLA-E,HLA-A,HLA-B*	0.00687
GO:0002456	BP	11	T cell mediated immunity	*B2M,HLA-E,HLA-A,HLA-B*	0.00818
GO:0002709	BP	11	regulation of T cell mediated immunity	*B2M,HLA-E,HLA-A,HLA-B*	0.00168
GO:0002711	BP	11	positive regulation of T cell mediated immunity	*B2M,HLA-E,HLA-A,HLA-B*	0.000315
GO:0001909	BP	11	leukocyte mediated cytotoxicity	*B2M,HLA-E,HLA-A,HLA-B*	0.0146
GO:0001910	BP	11	regulation of leukocyte mediated cytotoxicity	*B2M,HLA-E,HLA-A,HLA-B*	0.00182
GO:0001911	BP	11	negative regulation of leukocyte mediated cytotoxicity	*HLA-E,HLA-A,HLA-B*	0.00194
GO:0001913	BP	11	T cell mediated cytotoxicity	*B2M,HLA-E,HLA-A,HLA-B*	0.000209
GO:0001914	BP	11	regulation of T cell mediated cytotoxicity	*B2M,HLA-E,HLA-A,HLA-B*	0.000053
GO:0031343	BP	11	positive regulation of cell killing	*B2M,HLA-E,HLA-A,HLA-B*	0.000878
GO:0001912	BP	11	positive regulation of leukocyte mediated cytotoxicity	*B2M,HLA-E,HLA-A,HLA-B*	0.000514
GO:0001916	BP	11	positive regulation of T cell mediated cytotoxicity	*B2M,HLA-E,HLA-A,HLA-B*	2.12E-05
GO:0019221	BP	11	cytokine-mediated signaling pathway	*CD74,IFI27,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000668
GO:0060337	BP	11	type I interferon signaling pathway	*IFI27,HLA-E,HLA-A,HLA-B*	0.0139

GO:0060333	BP	11	interferon-gamma-mediated signaling pathway	*B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	5.58E-06
GO:0002715	BP	11	regulation of natural killer cell mediated immunity	*HLA-E,HLA-A,HLA-B*	0.0362
GO:0002716	BP	11	negative regulation of natural killer cell mediated immunity	*HLA-E,HLA-A,HLA-B*	0.000815
GO:0002720	BP	11	positive regulation of cytokine production involved in immune response	*CD74,B2M,HLA-A*	0.033
GO:0042269	BP	11	regulation of natural killer cell mediated cytotoxicity	*HLA-E,HLA-A,HLA-B*	0.033
GO:0045953	BP	11	negative regulation of natural killer cell mediated cytotoxicity	*HLA-E,HLA-A,HLA-B*	0.000815
GO:0042270	BP	11	protection from natural killer cell mediated cytotoxicity	*HLA-E,HLA-A,HLA-B*	0.000136
GO:0001775	BP	5	cell activation	*CD74,ACTB,CST3,B2M,ACTG1,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-DPA1*	0.00101
GO:0035740	BP	37	CD8-positive, alpha-beta T cell proliferation	*HLA-E,HLA-A*	0.0299
GO:2001187	BP	37	positive regulation of CD8-positive, alpha-beta T cell activation	*HLA-E,HLA-A*	0.0299
GO:2000564	BP	37	regulation of CD8-positive, alpha-beta T cell proliferation	*HLA-E,HLA-A*	0.0299
GO:2000566	BP	37	positive regulation of CD8-positive, alpha-beta T cell proliferation	*HLA-E,HLA-A*	0.015
GO:0098602	BP	19	single organism cell adhesion	*CD74,ACTB,B2M,ACTG1,S100A10,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.00187
GO:0016337	BP	19	single organismal cell-cell adhesion	*CD74,ACTB,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.0126
GO:0034109	BP	19	homotypic cell-cell adhesion	*CD74,ACTB,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.001
GO:0045785	BP	19	positive regulation of cell adhesion	*CD74,S100A10,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.0299
GO:0022409	BP	19	positive regulation of cell-cell adhesion	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.0475
GO:0034112	BP	19	positive regulation of homotypic cell-cell adhesion	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.022
GO:1903039	BP	19	positive regulation of leukocyte cell-cell adhesion	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.0226
GO:0050870	BP	19	positive regulation of T cell activation	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1*	0.0199
GO:0051099	BP	9	positive regulation of binding	*TXN,B2M,ANXA2,S100A10*	0.0484
GO:0002486	BP	7	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent	*HLA-A,HLA-B*	0.015
GO:0044419	BP	3	interspecies interaction between organisms	*PSME1,RPL27A,B2M,RPL12,RPS4X,HLA-*	0.0481

				DRA,HLA-A,HLA-B	
GO:0044403	BP	3	symbiosis, encompassing mutualism through parasitism	*PSME1,RPL27A,B2M,RPL12,RPS4X,HLA-DRA,HLA-A,HLA-B*	0.0481
GO:0044764	BP	3	multi-organism cellular process	*PSME1,RPL27A,B2M,RPL12,RPS4X,HLA-DRA,HLA-A,HLA-B*	0.0292
GO:0016032	BP	3	viral process	*PSME1,RPL27A,B2M,RPL12,RPS4X,HLA-DRA,HLA-A,HLA-B*	0.0261
GO:0006457	BP	36	protein folding	*CD74,ACTB,TXN,B2M,TUBA1A*	0.0475
GO:0032991	CC	29	macromolecular complex	*CD74,ACTB,PSME1,EEF1A1,RPL36AL,RPL27A,B2M,TUBA1A,ANXA2,ACTG1,RPL12,RPS4X,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	0.0254
GO:0009986	CC	28	cell surface	*CD74,B2M,ANXA2,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	0.00127
GO:0098797	CC	15	plasma membrane protein complex	*B2M,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	0.0109
GO:0042611	CC	15	MHC protein complex	*B2M,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	7.55E-12
GO:0042612	CC	15	MHC class I protein complex	*B2M,HLA-E,HLA-H,HLA-A,HLA-B*	2.85E-09
GO:0022626	CC	24	cytosolic ribosome	*RPL36AL,RPL27A,RPL12,RPS4X*	0.0389
GO:0005576	CC	6	extracellular region	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-B*	2.83E-06
GO:0044421	CC	6	extracellular region part	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,HLA-A,HLA-B*	9.25E-07
GO:0043230	CC	6	extracellular organelle	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,HLA-A,HLA-B*	2.31E-09
GO:0065010	CC	6	extracellular membrane-bounded organelle	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,HLA-A,HLA-B*	2.1E-09

GO:0044444	CC	6	cytoplasmic part	*CD74,ACTB,PSME1,CST3,NPC2,TXN,EEF1A1,RPL36AL,IFI27,RPL27A,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,RPL12,RPS4X,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-DPA1,HLA-B*	0.00323
GO:0031982	CC	6	vesicle	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-DPA1,HLA-B*	1.85E-09
GO:1903561	CC	6	extracellular vesicle	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,HLA-A,HLA-B*	2.31E-09
GO:0044433	CC	6	cytoplasmic vesicle part	*CD74,B2M,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-DPA1,HLA-B*	0.00172
GO:0031988	CC	6	membrane-bounded vesicle	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-DPA1,HLA-B*	8.82E-10
GO:0012506	CC	6	vesicle membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0048
GO:0070062	CC	6	extracellular exosome	*CD74,ACTB,PSME1,CST3,NPC2,IGLL1,TXN,EEF1A1,B2M,TUBA1A,ANXA2,ACTG1,S100A4,POTEF,S100A10,RPL12,RPS4X,HLA-DRA,HLA-E,HLA-A,HLA-B*	2.1E-09
GO:0005768	CC	6	endosome	*CD74,CST3,B2M,TUBA1A,ANXA2,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000162
GO:0044440	CC	6	endosomal part	*B2M,ANXA2,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00365
GO:0016023	CC	6	cytoplasmic membrane-bounded vesicle	*CD74,B2M,ANXA2,HLA-DRA,HLA-E,TMSB4X,HLA-A,HLA-DPA1,HLA-B*	0.0293
GO:0030139	CC	6	endocytic vesicle	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	7.25E-05
GO:0045335	CC	6	phagocytic vesicle	*B2M,HLA-E,HLA-A,HLA-B*	0.0139
GO:0030135	CC	6	coated vesicle	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	2.12E-05

GO:0030659	CC	6	cytoplasmic vesicle membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00365
GO:0030662	CC	6	coated vesicle membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	8.03E-07
GO:0030133	CC	6	transport vesicle	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	1.02E-05
GO:0030134	CC	6	ER to Golgi transport vesicle	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	1.32E-09
GO:0030658	CC	6	transport vesicle membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	5.65E-08
GO:0012507	CC	6	ER to Golgi transport vesicle membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	2.89E-10
GO:0010008	CC	6	endosome membrane	*B2M,ANXA2,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00243
GO:0031901	CC	6	early endosome membrane	*B2M,HLA-E,HLA-A,HLA-B*	0.0434
GO:0030666	CC	6	endocytic vesicle membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	1.86E-06
GO:0030670	CC	6	phagocytic vesicle membrane	*B2M,HLA-E,HLA-A,HLA-B*	0.00271
GO:0098552	CC	1	side of membrane	*CD74,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00423
GO:0098576	CC	1	lumenal side of membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	4E-09
GO:0031300	CC	1	intrinsic component of organelle membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00517
GO:0031227	CC	1	intrinsic component of endoplasmic reticulum membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	6.03E-05
GO:0098553	CC	1	lumenal side of endoplasmic reticulum membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	3.11E-09
GO:0031301	CC	1	integral component of organelle membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00396
GO:0030176	CC	1	integral component of endoplasmic reticulum membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	4.73E-05
GO:0071556	CC	1	integral component of lumenal side of endoplasmic reticulum membrane	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	3.11E-09

GO:0043209	CC	27	myelin sheath	*ACTB,EEF1A1,TUBA1A,ANXA2,ACTG1*	0.0104
GO:0030881	MF	8	beta-2-microglobulin binding	*HLA-E,HLA-H,HLA-A*	0.000815
GO:0003823	MF	10	antigen binding	*CD74,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	1.29E-07
GO:0033218	MF	10	amide binding	*CD74,CST3,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	3.57E-06
GO:0042277	MF	10	peptide binding	*CD74,CST3,HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	1.56E-06
GO:0042605	MF	10	peptide antigen binding	*HLA-DRA,HLA-E,HLA-H,HLA-A,HLA-DPA1,HLA-B*	6.4E-09
CORUM:3055	co	26	Nop56p-associated pre-rRNA complex	*EEF1A1,RPL27A,TUBA1A,RPL12*	0.0485
HP:0001339	hp	21	Lissencephaly	*ACTB,TUBA1A,ACTG1*	0.0469
KEGG:05166	ke	17	HTLV-I infection	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0374
KEGG:05130	ke	4	Pathogenic Escherichia coli infection	*ACTB,TUBA1A,ACTG1*	0.0268
KEGG:04514	ke	38	Cell adhesion molecules (CAMs)	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00219
KEGG:05169	ke	12	Epstein-Barr virus infection	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.0113
KEGG:05330	ke	18	Allograft rejection	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000002
KEGG:03010	ke	2	Ribosome	*RPL36AL,RPL27A,RPL12,RPS4X*	0.0265
KEGG:05332	ke	23	Graft-versus-host disease	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	2.67E-06
KEGG:04940	ke	35	Type I diabetes mellitus	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	4.56E-06
KEGG:04145	ke	14	Phagosome	*ACTB,TUBA1A,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	2.31E-07
KEGG:05168	ke	22	Herpes simplex infection	*CD74,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.000484
KEGG:05320	ke	33	Autoimmune thyroid disease	*HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	1.27E-05
KEGG:04612	ke	31	Antigen processing and presentation	*CD74,PSME1,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	4.18E-10
KEGG:05416	ke	20	Viral myocarditis	*ACTB,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	6.2E-09
MI:hsa-miR-33b	mi	34	MI:hsa-miR-33b	*CD74,TXN,EEF1A1,TUBA1A,RPS4X,TMSB4X*	0.0137
REAC:168256	re	25	Immune System	*CD74,ACTB,PSME1,TXN,IFI27,B2M,ACTG1,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00873
REAC:1280215	re	25	Cytokine Signaling in Immune system	*PSME1,IFI27,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	0.00789

REAC:913531	re	25	Interferon Signaling	*IFI27,B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	7.68E-05
REAC:909733	re	25	Interferon alpha/beta signaling	*IFI27,HLA-E,HLA-A,HLA-B*	0.0054
REAC:877300	re	25	Interferon gamma signaling	*B2M,HLA-DRA,HLA-E,HLA-A,HLA-DPA1,HLA-B*	1.68E-05
REAC:983170	re	30	Antigen Presentation: Folding, assembly and peptide loading of class I MHC	*B2M,HLA-E,HLA-A,HLA-B*	7.59E-05
REAC:156842	re	39	Eukaryotic Translation Elongation	*EEF1A1,RPL27A,RPL12,RPS4X*	0.0163
REAC:156902	re	39	Peptide chain elongation	*EEF1A1,RPL27A,RPL12,RPS4X*	0.013
REAC:1236975	re	16	Antigen processing-Cross presentation	*PSME1,B2M,HLA-E,HLA-A,HLA-B*	0.000398
REAC:1236974	re	16	ER-Phagosome pathway	*PSME1,B2M,HLA-E,HLA-A,HLA-B*	0.000126
REAC:1236977	re	16	Endosomal/Vacuolar pathway	*B2M,HLA-E,HLA-A,HLA-B*	2.41E-06

Abbreviations: GO = Gene Ontology; BP = Biological Process; CC = Cellular Component; MF = Molecular Function; CORUM (CO) = Comprehensive Resource of Mammalian protein complexes; KEGG (KE) = Kyoto Encyclopedia of Genes and Genomes; HP = Human Phenotype Ontology; REAC (RE) = Reactome Pathway Database;

[a]The number of dysregulated genes found in the expression microarray data prevented functional enrichment analysis for other comparison groups.

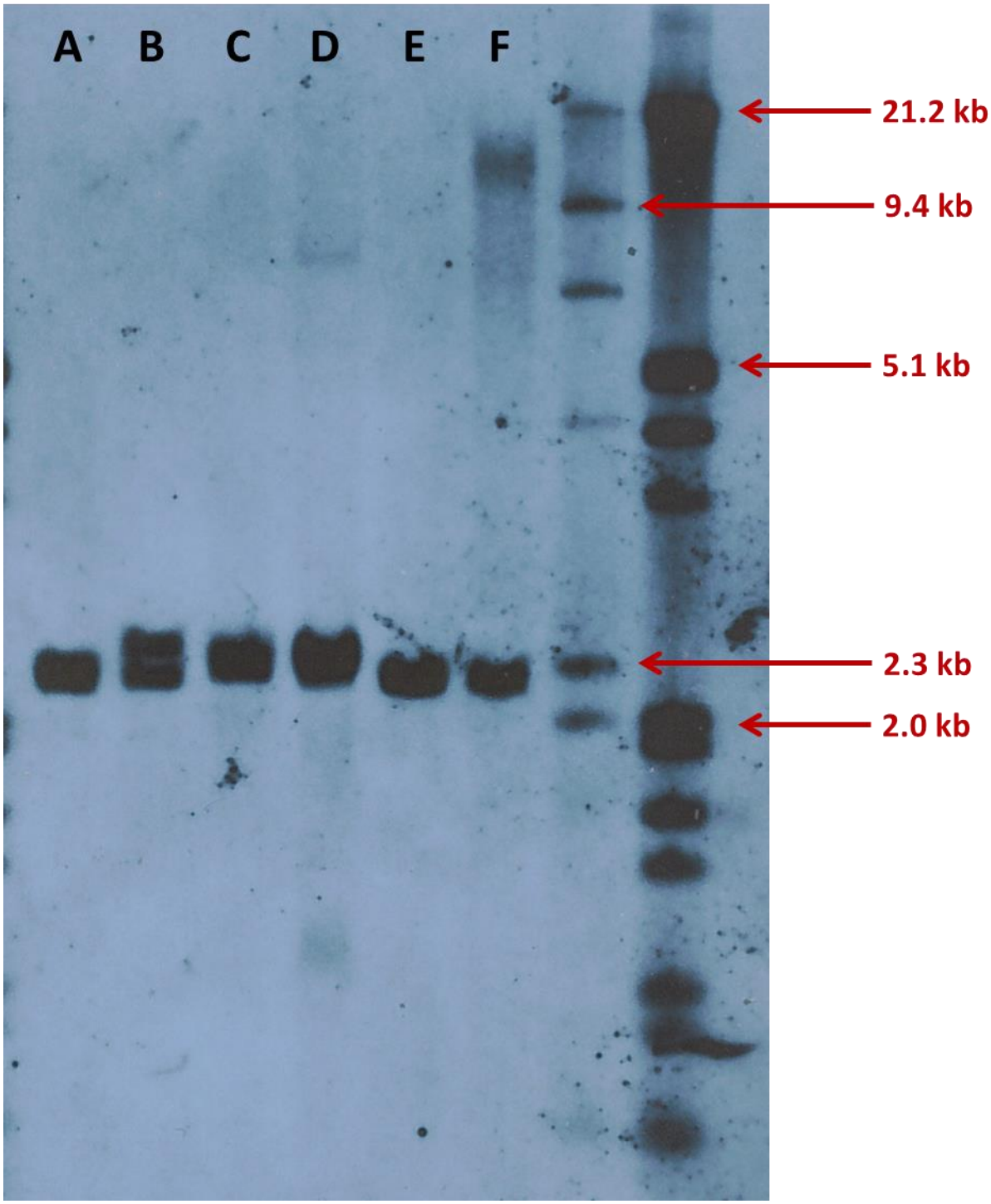


Figure 1. Southern blot of *C9orf72* repeat in different samples. Lanes A, D and E are samples without expansion. The blurred band in lane D between 5.1 kb – 21.2 kb is probably due to the overloading of the sample and incomplete digestion by the enzyme. Lanes B and C are sIBM samples which are confirmed that there are no large expansions. Lane B is the sample with an allele of 35 repeats and the other allele of 7 repeats (There is gap of 168 bp between two bands). Lane C is the sample with an allele of 27 repeats and the other allele with 12 repeats (The two bands are too close to be well separated.). Lane F is a sample with a large expansion at approximately 14 kb. The DIG-labeled DNA molecular-weight markers II (left) and III (right) were used. The experiment was kindly performed by Ese Mudanohwo from NHNN Neurogenetics Laboratory.

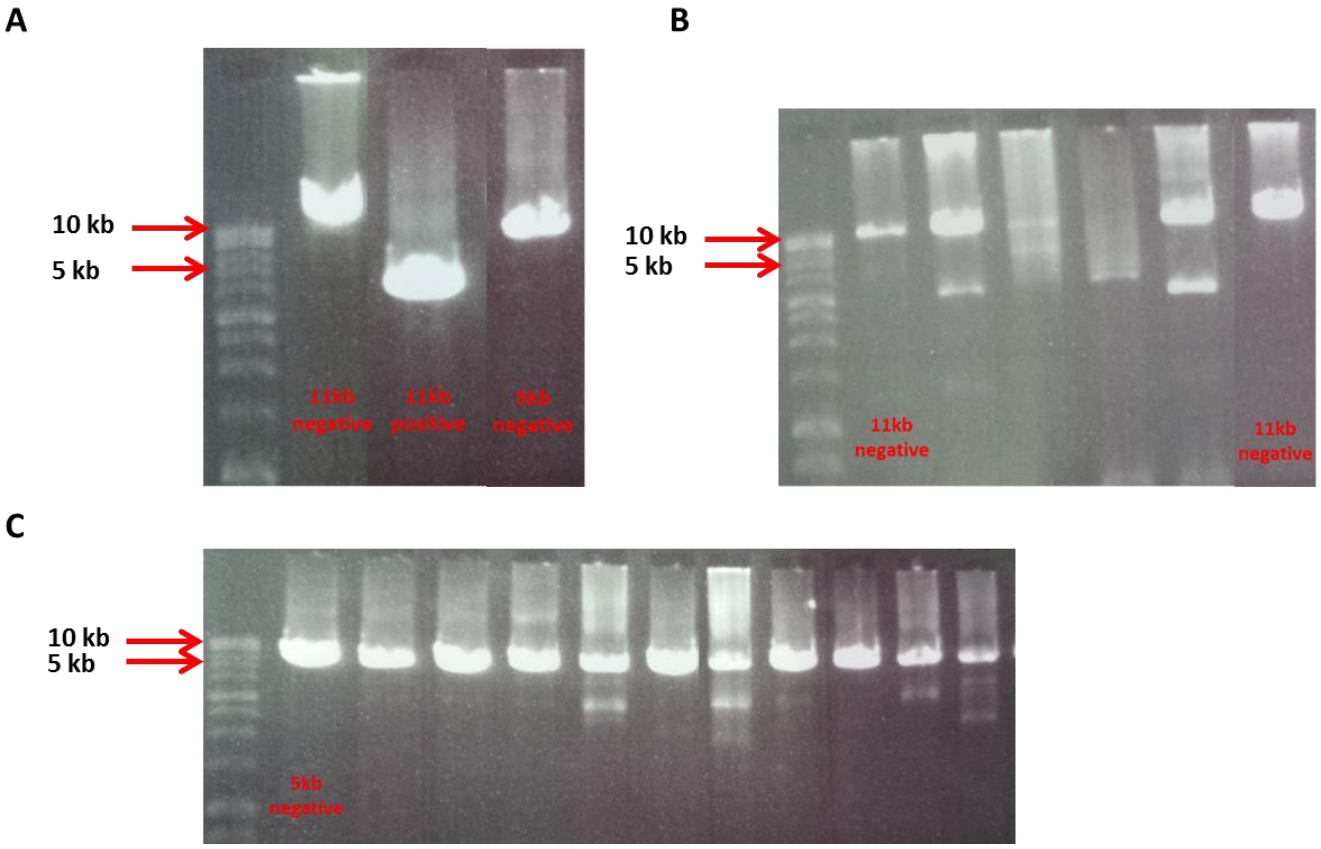


Figure 2. mtDNA long PCR fragments in 0.8% agarose gel electrophoresis run for 60 mins at 45v. (A) 11kb mtDNA long PCR fragments of a negative and a positive control, and a 5kb PCR fragment of a negative control; (B) 11kb mtDNA long PCR fragments of sIBM muscle DNA samples; The labelled ones are negative of a deletion; The others might be non-specific PCR fragments due to the low DNA concentration; (C) 5kb mtDNA long PCR fragments of sIBM muscle DNA samples; The PCR fragments with a same size as the labelled one are negative of a deletion; The others with multiple bands might be non-specific PCR fragments due to the low DNA concentration. A 1 kb ladder was used.

Appendix III for Chapter 4

Command lines involved in the bioinformatics pipeline, quality control, association analyses, and relevant statistics are provided below.

1. **The Workflow of the First Stage Data Processing from BAM Files to Generate a Joint Filtered VCF File.** (kindly provided by Dr Alan Pittman)
   1) Command Lines to Generate BAM files:

```
export
PATH=${PATH}:/data/kronos/NGS_Software_heavy/annovar_Feb2013:/dat
a/kronos/NGS_Software/tabix-0.2.3
export PERL5LIB=/data/kronos/NGS_Software/vcftools_0.1.8/lib:
export JAVA_HOME=/illumina/jre1.7.0_67/bin/java

/data/kronos/NGS_Software/novocraft_v3/novoalign -c 8 -o SAM
$'@RG\tID:apittman-ext-67752\tSM:apittman-ext-67752\tLB:apittman-ext-
67752\tPL:ILLUMINA' --rOQ --hdrhd 3 -H -k -a
CTGTCTCTTATACACATCT -o Soft -t 320 -F ILM1.8 -f
/array/Incoming_HiSeq3000/Unaligned/apittman/apittman-ext-
67752/apittman-ext-67752_S33_L003_R1_001.fastq.gz
/array/Incoming_HiSeq3000/Unaligned/apittman/apittman-ext-
67752/apittman-ext-67752_S33_L003_R2_001.fastq.gz -d
/data/kronos/NGS_Reference/novoalign/human_g1k_v37.fasta.k15.s2.novoin
dex > /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_1.sam

/data/kronos/NGS_Software/samtools-0.1.18/samtools view -bS -t
/data/kronos/NGS_Reference/fasta/human_g1k_v37.fastai -o
/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_1.bam
/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_1.sam  ## make BAM file

/data/kronos/NGS_Software/samtools-0.1.18/samtools sort -m 5000000000
/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_1.bam
/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_sorted_1  ## sort

/data/kronos/NGS_Software/novocraft_v3/novoalign -c 8 -o SAM
$'@RG\tID:apittman-ext-67752\tSM:apittman-ext-67752\tLB:apittman-ext-
67752\tPL:ILLUMINA' --rOQ --hdrhd 3 -H -k -a
CTGTCTCTTATACACATCT -o Soft -t 320 -F ILM1.8 -f
/array/Incoming_HiSeq3000/Unaligned/apittman/apittman-ext-
67752/apittman-ext-67752_S33_L004_R1_001.fastq.gz
/array/Incoming_HiSeq3000/Unaligned/apittman/apittman-ext-
67752/apittman-ext-67752_S33_L004_R2_001.fastq.gz -d
```

/data/kronos/NGS_Reference/novoalign/human_g1k_v37.fasta.k15.s2.novoindex > /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_2.sam

/data/kronos/NGS_Software/samtools-0.1.18/samtools view -bS -t /data/kronos/NGS_Reference/fasta/human_g1k_v37.fastai -o /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_2.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_2.sam ## make BAM file

/data/kronos/NGS_Software/samtools-0.1.18/samtools sort -m 5000000000 /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_2.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_2 ## sort

/data/kronos/NGS_Software/samtools-0.1.18/samtools merge -f /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_1.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_2.bam

rm /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_1.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_2.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_1.sam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_2.sam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_1.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_2.bam

/data/kronos/NGS_Software/samtools-0.1.18/samtools index /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted.bam ##build index

Now remove PCR duplicates using PICARD
java -Xmx10g -jar /data/kronos/NGS_Software/picard-tools-1.75/MarkDuplicates.jar TMP_DIR=/data/kronos/temp ASSUME_SORTED=true REMOVE_DUPLICATES=FALSE INPUT=/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted.bam

OUTPUT=/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_unique.bam METRICS_FILE=/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_picard_metrics.out

/data/kronos/NGS_Software/samtools-0.1.18/samtools index /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_unique.bam ##build index

rm /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted.bam.bai

rm /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752.bam /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752.sam

java -Xmx10g -jar /data/kronos/NGS_Software/picard-tools-1.75/CalculateHsMetrics.jar BAIT_INTERVALS=/data/kronos/NGS_Reference/query_novopile/Nextera_focussed_ExomeTarget_hg19_0bp.tab.intList TARGET_INTERVALS=/data/kronos/NGS_Reference/query_novopile/Nextera_focussed_ExomeTarget_hg19_0bp.tab.intList INPUT=/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_unique.bam OUTPUT=/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752.hybridMetrics

EXTRA STEPS TO PERFORM IndelREALIGNMENT AND VARIANT DETECTION IN GATK
/data/kronos/General_Software/jre1.7.0_67/bin/java -jar /data/kronos/NGS_Software/GATK_v3_3/GenomeAnalysisTK.jar -T RealignerTargetCreator -L /data/kronos/NGS_Reference/GATK_refFiles/INTERVALS.bed -nt 5 -R /data/kronos/NGS_Reference/fasta/human_g1k_v37.fasta -o /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_unique.bam.list -I /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-67752/apittman-ext-67752_sorted_unique.bam --known /data/kronos/NGS_Reference/GATK_refFiles/Mills_and_1000G_gold_standard.indels.hg19_modified.vcf

/data/kronos/General_Software/jre1.7.0_67/bin/java -jar /data/kronos/NGS_Software/GATK_v3_3/GenomeAnalysisTK.jar -T IndelRealigner -targetIntervals /array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-

67752/apittman-ext-67752_sorted_unique.bam.list -I
/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_sorted_unique.bam -R
/data/kronos/NGS_Reference/fasta/human_g1k_v37.fasta -o
/array/Incoming_HiSeq3000/a_aligned_v4/apittman/apittman-ext-
67752/apittman-ext-67752_sorted_unique_realigned.bam --knownAlleles
/data/kronos/NGS_Reference/GATK_refFiles/Mills_and_1000G_gold_stand
ard.indels.hg19_modified.vcf

2) Command Lines in Haplotype_Caller.sh:

```
myIDs= "" #a list of IBM sample IDs
for nID in $myIDs; do
echo " $JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T
HaplotypeCaller –sample_name ${nID} --emitRefConfidence GVCF --
variant_index_type LINEAR --variant_index_parameter 128000 -I
$iDirectory/${nID}/${nID}_sorted_unique_realigned.bam -o
$iDirectory/${nID}/${nID}.raw.snps.indels.g.vcf" >
$oDirectoryGVCF/J_${nID}_H.sh
done
```

3) Command Lines in Genotype_gVCF.sh:

```
mygVCFs="" #a list of gVCF Samples
#Give the analysis batch a name:
BATCH="IBM"
#Remove any old input lists:
rm $temp/$(Foroud et al.)_gVCF_inputs.txt
#First lets make a list of our .gVCF file inputs and paths:
for nID in $mygVCFs; do
echo "--variant $iDirectory/${nID}/${nID}.raw.snps.indels.g.vcf " >>
$temp/$(Foroud et al.)_gVCF_inputs.txt
done

#CombineGVCFs [CombineGVCFS] #
$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T
CombineGVCFs  cat $temp/$(Foroud et al.)_gVCF_inputs.txt` -o
$temp/$BATCH.g.vcf

#Joint Genotyping [GenotypeGVCFS] #
for chr in $chIDs; do
$JAVA -jar $GATK -R $GENOMEREF -L ${chr} -T GenotypeGVCFs -nt
10 --variant $temp/$BATCH.g.vcf --includeNonVariantSites -o
$temp/$(Foroud et al.)_${chr}.vcf
done

#We now have to join all of our split .vcf files back together.
$JAVA -cp $GATK org.broadinstitute.gatk.tools.CatVariants -R
/data/kronos/NGS_Reference/fasta/human_g1k_v37.fasta -V $temp/$(Foroud
et al.)_1.vcf -V $temp/$(Foroud et al.)_2.vcf -V $temp/$(Foroud et al.)_3.vcf
-V $temp/$(Foroud et al.)_4.vcf -V $temp/$(Foroud et al.)_5.vcf -V
```

$temp/$(Foroud et al.)_6.vcf -V $temp/$(Foroud et al.)_7.vcf -V
$temp/$(Foroud et al.)_8.vcf -V $temp/$(Foroud et al.)_9.vcf -V
$temp/$(Foroud et al.)_10.vcf -V $temp/$(Foroud et al.)_11.vcf -V
$temp/$(Foroud et al.)_12.vcf -V $temp/$(Foroud et al.)_13.vcf -V
$temp/$(Foroud et al.)_14.vcf -V $temp/$(Foroud et al.)_15.vcf -V
$temp/$(Foroud et al.)_16.vcf -V $temp/$(Foroud et al.)_17.vcf -V
$temp/$(Foroud et al.)_18.vcf -V $temp/$(Foroud et al.)_19.vcf -V
$temp/$(Foroud et al.)_20.vcf -V $temp/$(Foroud et al.)_21.vcf -V
$temp/$(Foroud et al.)_22.vcf -V $temp/$(Foroud et al.)_X.vcf -V
$temp/$(Foroud et al.)_Y.vcf -out $temp/$(Foroud et al.).vcf

4) Command Lines in Filter_VCF.sh:

```
##SET YOUR FILTER LEVEL HERE##
VQSRFILTERLEVEL="99.0"

#Name_of_VCF_file_to filter_
BATCH="Brain_Controls_Batch_1"

#HardFilter DP and GQ
$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T
VariantFiltration --genotypeFilterExpression "DP < 8" --genotypeFilterName
"LowDepth" -V $temp/$(Foroud et al.).vcf  -o $OUT/$(Foroud et
al.)_HF1.vcf

$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T
VariantFiltration --genotypeFilterExpression "GQ < 20.0 && GQ > 0.0" --
genotypeFilterName "LowGQ"  -V $OUT/$(Foroud et al.)_HF1.vcf   -o
$OUT/$(Foroud et al.)_HF2.vcf

$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T
VariantFiltration --filterExpression "GQ_MEAN < 35.0 && GQ_MEAN >
0.0" --filterName "LowGQmean" -V $OUT/$(Foroud et al.)_HF2.vcf -o
$OUT/$(Foroud et al.)_HF3.vcf

$JAVA -jar $GATK5 -R $GENOMEREF -L $INTERVALS -T
SelectVariants --variant $OUT/$(Foroud et al.)_HF3.vcf --excludeFiltered --
setFilteredGtToNocall -o $OUT/$(Foroud et al.)_HF4.vcf

#VARIANT QUALITY SCORE RECALIBRATION

#VQSR Step 1 [variantRecalibrator] - SNPs
$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T
VariantRecalibrator -input $OUT/$(Foroud et al.)_HF4.vcf  -tranche 100.0 -
tranche 99.9 -tranche 99.0 -tranche 90.0 -mode SNP -recalFile
$OUT/$(Foroud et al.)_SNP.recal -tranchesFile $OUT/$(Foroud et
al.)_SNP.tranches -rscriptFile $OUT/$(Foroud et al.).plots.R -an QD -an MQ
-an MQRankSum -an ReadPosRankSum -an FS -an SOR -an
InbreedingCoeff --maxGaussians 5 -
```

resource:hapmap,known=false,training=true,truth=true,prior=15.0 /data/kronos/smorgan/public_data/hapmap_3.3.hg19.sites.vcf -resource:omni,known=false,training=true,truth=true,prior=12.0 /data/kronos/smorgan/public_data/1000G_omni2.5.hg19.sites.vcf -resource:1000G,known=false,training=true,truth=false,prior=10.0 /data/kronos/smorgan/public_data/1000G_phase1.snps.high_confidence.hg19.sites.vcf -resource:dbsnp,known=true,training=false,truth=false,prior=2.0 /data/kronos/smorgan/public_data/dbsnp_138.hg19.vcf

#VQSR Step 2 [ApplyRecalibration] - SNPs
$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T ApplyRecalibration -mode SNP --ts_filter_level $VQSRFILTERLEVEL -input $OUT/$(Foroud et al.)_HF4.vcf -tranchesFile $OUT/$(Foroud et al.)_SNP.tranches -recalFile $OUT/$(Foroud et al.)_SNP.recal -o $OUT/$(Foroud et al.)_HF4_SNP.recal.snps.vcf

##Now re-do again but for INDELS innit:
#VQSR Step 1 [variantRecalibrator] - INDEL
$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T VariantRecalibrator -input $OUT/$(Foroud et al.)_HF4.vcf -tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0 -mode INDEL -recalFile $OUT/$(Foroud et al.)_INDEL.recal -tranchesFile $OUT/$(Foroud et al.)_INDEL.tranches -rscriptFile $OUT/$(Foroud et al.)_INDEL.plots.R --maxGaussians 4 -resource:mills,known=false,training=true,truth=true,prior=12.0 /data/kronos/smorgan/public_data/nMills_and_1000G_gold_standard.indels.hg19.sites.vcf -resource:dbsnp,known=true,training=false,truth=false,prior=2.0 /data/kronos/smorgan/public_data/dbsnp_138.hg19.vcf -an QD -an DP -an FS -an SOR -an ReadPosRankSum -an MQRankSum -an InbreedingCoeff

#VQSR Step 2 [ApplyRecalibration] - INDEL
$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T ApplyRecalibration -mode INDEL --ts_filter_level $VQSRFILTERLEVEL -input $OUT/$(Foroud et al.)_HF4_SNP.recal.snps.vcf -tranchesFile $OUT/$(Foroud et al.)_INDEL.tranches -recalFile $OUT/$(Foroud et al.)_INDEL.recal -o $OUT/$(Foroud et al.)_HF4_SNP.recal.snps.indel.vcf

vcftools --vcf $OUT/$(Foroud et al.)_HF4_SNP.recal.snps.indel.vcf --max-alleles 4 --recode --recode-INFO-all --out $OUT/$(Foroud et al.)_HF4.recal.snps.indel_alleleReduction.vcf

$JAVA -jar $GATK -R $GENOMEREF -L $INTERVALS -T SelectVariants -V $OUT/$(Foroud et al.)_HF4.recal.snps.indel_alleleReduction.vcf.recode.vcf --excludeFiltered -o $OUT/$(Foroud et al.)_HF4.recal.snps.indel_alleleReduction_VQSRfiltered.vcf

2. **Command Lines in Plink for QC** (performed by the author of the thesis)

#To annotate with dbSNP and then convert a VCF file into a set of binary Plink files (.bed, .bim, .fam)

/data/kronos/General_Software/jre1.7.0_67/bin/java -jar /data/kronos/NGS_Software/GATK_v3_5/GenomeAnalysisTK.jar -R /data/kronos/NGS_Reference/fasta/human_g1k_v37.fasta -T VariantAnnotator -V IBM_combined_case_control_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_1_March_2016.vcf --dbsnp /data/kronos/NGS_Reference/GATK_refFiles/common_all.vcf -o IBM_combined_case_control_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_1_March_2016_dbsnpannotated.vcf

plink --vcf IBM_combined_case_control_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_1_March_2016_dbsnpannotated.vcf --double-id --make-bed --out IBM_combined_case_control_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_1_March_2016_dbsnpannotated

#To select a list of samples for the future analysis
plink --bfile IBM_combined_case_control_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_1_March_2016_dbsnpannotated --double-id --keep IBM_Case_Control_final_list.txt --make-bed --out IBM_combined_case_control_March2016_final_selected

1) Variant QC:
#To check the call rate of all the variants and all the samples across the whole dataset
plink --bfile IBM_combined_case_control_March2016_final_selected --missing --out IBM_combined_case_control_March2016_selected

#To remove the variants with HWE p-value < 0.001 calculated in control samples only
plink --bfile Brain_Controls_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_dbsnpannotated --hwe 0.001 --make-bed --out Brain_Controls_HF4.recal.snps.indel_alleleReduction_VQSRfiltered_99_0_dbsnpannotated_pass_HWE

plink --bfile IBM_combined_case_control_March2016_final_selected --exclude fail_HWE.txt --make-bed --out IBM_combined_case_control_March2016_final_selected_pass_HWE

#To remove the variants with missingness > 0.015
plink --bfile IBM_combined_case_control_March2016_final_selected_pass_HWE --geno

0.015 --make-bed --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015

2) Sample QC:

i. Contamination Detection

```
myIDs="" #a list of IBM sample IDs
for id in $myIDs; do
/data/kronos/Genetics_Software/verifyBamID_1.1.2/verifyBamID/bin/verify
BamID --vcf
/data/kronos/NGS_Reference/1000genomes_phase1/1000genomes_Var_targe
t.snps.recode.vcf --bam
/data/kronos/qgang/Alan/BAM_files/${id}/${id}_sorted_unique_realigned.b
am --out
/data/kronos/qgang/IBM_cases_control/VerifyBamID_test/IBM/${id} --
verbose --ignoreRG --self
done
```

ii. Call Rate per Sample

```
#To remove the samples with missingness > 0.2
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015 --mind 0.2 --make-bed --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_missing_sample
```

iii. Gender Mismatch

```
plink --bfile IBM_combined_case_control_March2016_final_selected --split-
x hg19 --make-bed --out
IBM_combined_case_control_March2016_final_selected_splitX

plink --bfile IBM_combined_case_control_March2016_final_selected_splitX
--check-sex ycount --update-sex IBM_control_genders_final.txt --out
IBM_combined_case_control_March2016_final_selected_splitX
```

iv. Heterozygosity Rate per Sample

```
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015 --het --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015
```

v. Duplicates and Relatedness

```
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers --maf 0.05 --make-bed --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp
```

plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp --exclude high-LD-regions.txt
--range --indep-pairwise 50 5 0.2 --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp

plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp --extract
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp.prune.in --make-bed --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp_pruned

plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp_pruned --genome --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_common_snp_pruned

#To remove the sample outliers and duplicates/relatedness calculated from i, iii, iv, and v steps
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_missing_sample --remove sample_outliers.txt --
remove duplicates_relatedness_list.txt --make-bed --out
IBM_combined_case_control_March2016_selected_less_missing_snp_less_o
utliers_less_related

vi. Population Stratification

#To create a file selected for common SNPs
plink --bfile
IBM_combined_case_control_March2016_selected_less_missing_snp_less_o
utliers_less_related --maf 0.05 --make-bed --out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp

#To remove the high LD regions
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_less_missing
_snp0.015_less_outliers_less_related_common_snp --exclude high-LD-
regions.txt --range --indep-pairwise 50 5 0.2 --out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp

plink --bfile
IBM_combined_case_control_March2016_final_selected_less_missing_snp0

.015_less_outliers_less_related_common_snp --extract
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp.prune.in --make-bed --out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned

#To select HapMap SNPs and merge it with the HapMap data (downloaded from paper by Anderson et al. 2010)
plink --bfile
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned --extract
hapmap3r2_CEU.CHB.JPT.YRI.no-at-cg-snps.txt --make-bed --out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps

plink --bfile
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps --bmerge
hapmap3r2_CEU.CHB.JPT.YRI.founders.no-at-cg-snps.bed
hapmap3r2_CEU.CHB.JPT.YRI.founders.no-at-cg-snps.bim
hapmap3r2_CEU.CHB.JPT.YRI.founders.no-at-cg-snps.fam --make-bed --
out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps.hapmap-
snps_hapmap3r2

plink --bfile
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps --flip
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps.hapmap-
snps_hapmap3r2-merge.missnp --make-bed --out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps_filp

#To remove the multiallelic variants listed in ".missnp" file and repeat the above step
plink --bfile
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-snps_filp --
bmerge hapmap3r2_CEU.CHB.JPT.YRI.founders.no-at-cg-snps.bed
hapmap3r2_CEU.CHB.JPT.YRI.founders.no-at-cg-snps.bim
hapmap3r2_CEU.CHB.JPT.YRI.founders.no-at-cg-snps.fam --make-bed --
out
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-
snps_hapmap3r2

#To calculate the PCA by using GCTA

```
gcta --bfile
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-
snps_hapmap3r2 --make-grm --autosome --out
IBM_Control_final_hapmap3r2
gcta --grm IBM_Control_final_hapmap3r2 --pca 4 --out
IBM_Control_final_hapmap3r2

#To select population information from ".pedind" file, and combine it with
"IBM_Control_hapmap3r2.eigenvec"
awk '{print $6}'
IBM_combined_case_control_March2016_final_selected_less_missing_snp0
.015_less_outliers_less_related_common_snp_pruned.hapmap-
snps_hapmap3r2.pedind > all_population_list.txt
paste IBM_Control_final_hapmap3r2.eigenvec all_population_list.txt >
IBM_Control_final_hapmap3r2.pca.evec
sed -r 's!\t! !g' IBM_Control_final_hapmap3r2.pca.evec >
IBM_Control_final_hapmap3r2.pca_new.evec

#To remove the ethnicity outliers from the whole dataset
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related --remove
ethnicity_outliers_final.txt --make-bed --out
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU
```

3. **R Scripts for Plotting PCA** (performed by the author of the thesis)**:**

```
#Read the file into R.
evec<-read.table("IBM_Control_final_hapmap3r2.pca_new.evec", header=F,
as.is=T)

#Check the dimensions of the file; i.e. list number of rows and columns
dim(evec)

#Check the first few rows and columns to make sure the file read into R
correctly
evec[1:5,]

#Calculate the numbers of subjects within each sample group.
table(evec$V7)

#Define HapMap ethnic groups
evec$V7[evec$V7=="1"]<-"CTRL"
evec$V7[evec$V7=="2"]<-"CASE"
evec$V7[evec$V7=="3"]<-"CEU"
evec$V7[evec$V7=="4"]<-"CHB"
evec$V7[evec$V7=="5"]<-"JPT"
evec$V7[evec$V7=="6"]<-"YRI"
```

```
genomes<-c("CEU","CHB","JPT","YRI")

####  PC Plot 1  ####
xmax<-max(evec$V3)*1.25
xmin<-min(evec$V3)*1.25
ymax<-max(evec$V4)*1.25
ymin<-min(evec$V4)*1.25
c1<-rainbow(4)
png("HapMap_PCs.png")
plot(x=evec[grepl(genomes[1], evec$V7), "V3"], y=evec[grepl(genomes[1],
evec$V7), "V4"], col=c1[1], type="p", ylab="PC2", xlab="PC1", pch=20,
ylim=c(ymin, ymax), xlim=c(xmin, xmax))
points(evec[grepl(genomes[2], evec$V7), "V3"], evec[grepl(genomes[2],
evec$V7), "V4"], col=c1[2], pch=20)
points(evec[grepl(genomes[3], evec$V7), "V3"], evec[grepl(genomes[3],
evec$V7), "V4"], col=c1[3], pch=20)
points(evec[grepl(genomes[4], evec$V7), "V3"], evec[grepl(genomes[4],
evec$V7), "V4"], col=c1[4], pch=20)
points(evec[(evec$V7=="CTRL"), "V3"], evec[(evec$V7=="CTRL"), "V4"],
col="black", pch=8)
points(evec[(evec$V7=="CASE"), "V3"], evec[(evec$V7=="CASE"), "V4"],
col="pink", pch=8)
legend("topright", legend=genomes, text.col=c1)
abline(h=0.0,v=0.0,col="gray32",lty=2)
dev.off()
evec$outliers<-(evec$V3>0.0 | evec$V4<0.0)
list<-evec[evec$outliers,]
list$case_control<-(list$V7=="CASE" | list$V7=="CTRL")
write.table(list[list$case_control, ], "not_european.txt",col.names=F,
row.names=F, quote=F)
```

4. **R Scripts for Plotting QQ-Plot** (performed by the author of the thesis)**:**

```
#Create a qq plot ".chisq.qq.plot.pdf" with a 95% confidence interval
data<-
read.table("IBM_combined_case_control_March2016_selected_final_pass_Q
C0.015_common_SNPs_test1.model", header=TRUE)
png("IBM_Cases_Controls.chisq.qq.plot_common_150316.png")
library(car)
obs<-data[data$TEST=="ALLELIC",]$CHISQ
ALLELIC<-data[data$TEST=="ALLELIC",]

#sorting by CHISQ at ascending order
attach(ALLELIC)
ALLELIC_sort<-ALLELIC[order(CHISQ),]
ALLELIC_sort<-na.omit(ALLELIC_sort)
detach(ALLELIC)

#select the bottom 90% of the CHISQ from ALLELIC_sort
newdata<-ALLELIC_sort[1:((nrow(ALLELIC_sort))*0.9),]
```

```
#calculate genomic inflation factor (lambda)
y<-qchisq(1-ALLELIC_sort$P,1)
lambda<-median(y)/0.456
newobs<-newdata$CHISQ
qqPlot(obs, distribution="chisq", df=1, xlab="Expected chi-squared values",
ylab="Obserbed test statistic", grid=FALSE)
dev.off()
```

5. **Command Lines in Plink for Association Analysis** (performed by the author of the thesis)

```
# Basic descriptive summary
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU --assoc --adjust --out
IBM_Cases_Controls_Association_pass_QC

plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU --assoc --mperm 1000000
--out IBM_Cases_Controls_Association_pass_QC

#Single SNP tests of association
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU --adjust --model-trend --
allow-no-sex --out IBM_Cases_Controls_Association_pass_QC

plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU --model --mperm
1000000 --model-trend --allow-no-sex --out
IBM_Cases_Controls_Association_pass_QC

#Test of association using logistic regression
plink --bfile
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU --logistic --sex --out
IBM_Cases_Controls_Association
```

6. **R Scripts for SKAT-O Analysis** (performed by the author of the thesis)**:**

```
#An example of analysing all genes and all variants
library(SKAT)
#File names
All.Bed<-"IBM_combined_case_control.bed"
All.Bim<-"IBM_combined_case_control.bim"
All.Fam<-"IBM_combined_case_control.fam"
All.SetID<-"AllGene_SetID.txt"
All.Cov<-read.table("Covariate_noheader.txt")
```

```
X<-data.frame(All.Cov)
View(All.Cov)
View(X$V5)
Generate_SSD_SetID(All.Bed, All.Bim, All.Fam, All.SetID, "All.SSD",
"All.Info")

FAM1<-Read_Plink_FAM(All.Fam, Is.binary=TRUE, flag1=1)
View(FAM1)
y<-FAM1$Phenotype
View(y)
X1<-X$V5

SSD.INFO<-Open_SSD("All.SSD", "All.Info")
SSD.INFO$nSample
SSD.INFO$nSets

#To calculate SKATO with covariates
Obj1<-SKAT_Null_Model(y~X1, out_type="D")
Out1_cov.skato.adj<-SKAT.SSD.All(SSD.INFO, obj1,
method="optimal.adj")
All_genes_output_cov.skato.adj.df=out1_cov.skato.adj$results
write.table(All_genes_output_cov.skato.adj.df,
file="All_genes_cov_SKATO.adj_adjusted.txt", col.names=TRUE,
row.names=FALSE)
```

7. **Command Lines for VEP** (performed by the author of the thesis)**:**

```
#Annotation with VEP
perl /data/kronos/NGS_Software/VEP/ensembl-tools-release-
77/scripts/variant_effect_predictor/variant_effect_predictor.pl -i
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU.vcf -offline -cache -dir
~/.vep/ --cache_version 77 --everything --fork 4 -o
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_out.vcf


#Filter for only the Consequences of variants (all non-synonymous variants),
regardless of frequency
perl filter_vep.pl -i
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP.vcf --filter
"Consequence is splice_acceptor_variant or Consequence is
splice_donor_variant or Consequence is stop_gained or Consequence is
frameshift_variant or Consequence is stop_lost or Consequence is start_lost
or Consequence is transcript_amplification or Consequence is
inframe_deletion or Consequence is inframe_insertion or Consequence is
missense_variant or Consequence is protein_altering_variant" --filter
"CANONICAL is YES" --format vcf -o
IBM_combined_case_control_March2016_final_filtered_consequences_only.
vcf
```

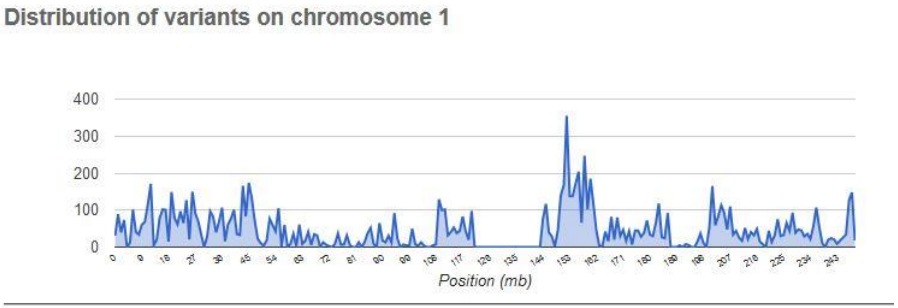
#Filter for the variants MAF<=0.01
perl filter_vep.pl -i
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP.vcf --filter
"EA_MAF < 0.01 or EA_MAF = " --filter "GMAF < 0.01 or GMAF = " --
filter "CANONICAL is YES" --format vcf -o
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP_filtered_frequency0
.01.vcf

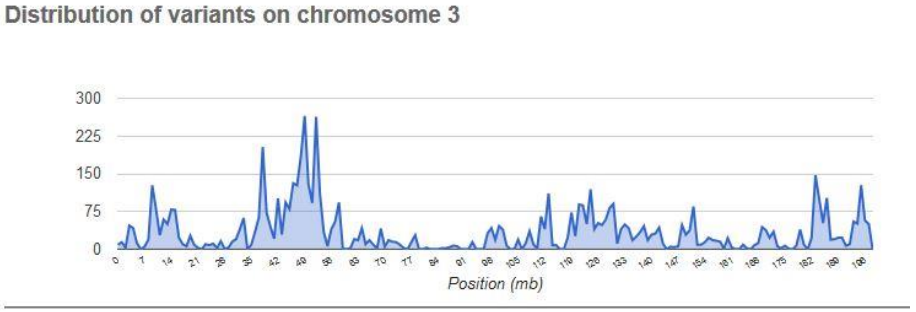
#Filter for all non-synonymous variants except those predicted by SIFT =
tolerated and PolyPhen-2 = benign (all not benign variants) and MAF<=0.01
perl filter_vep.pl -i
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP_filtered_frequency0
.01.vcf --filter "Consequence is splice_acceptor_variant or Consequence is
splice_donor_variant or Consequence is stop_gained or Consequence is
frameshift_variant or Consequence is stop_lost or Consequence is start_lost
or Consequence is transcript_amplification or Consequence is
inframe_deletion or Consequence is inframe_insertion or Consequence is
missense_variant or Consequence is protein_altering_variant" --filter "SIFT
!= tolerated" --filter "PolyPhen != benign" --filter "CANONICAL is YES" --
format vcf -o
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP_filtered_frequency0
.01_filtered_not_benign.vcf

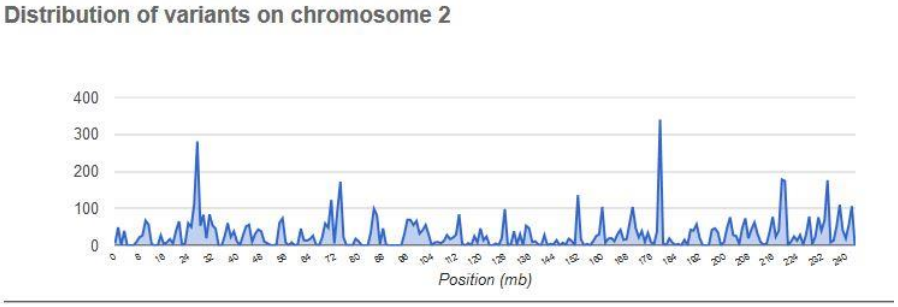
#Filter for LOF variants only and MAF<=0.01 (stop gain, frameshift,
canonical splice variants)
perl filter_vep.pl -i
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP_filtered_frequency0
.01.vcf --filter "Consequence is splice_acceptor_variant or Consequence is
splice_donor_variant or Consequence is stop_gained or Consequence is
frameshift_variant or Consequence is stop_lost or Consequence is start_lost"
--filter "CANONICAL is YES" --format vcf -o
IBM_combined_case_control_March2016_final_selected_pass_HWE_less_
missing_snp0.015_less_outliers_less_related_CEU_VEP_filtered_frequency0
.01_filtered_LOF.vcf

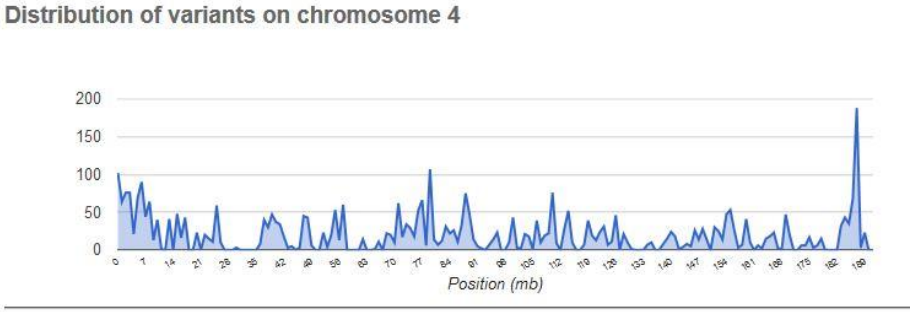
8. Distributions of Variants on Each Chromosome Post QC Annotated by VEP

(X-axis represents the position on each chromosome; Y-axis represents the number of variants.)

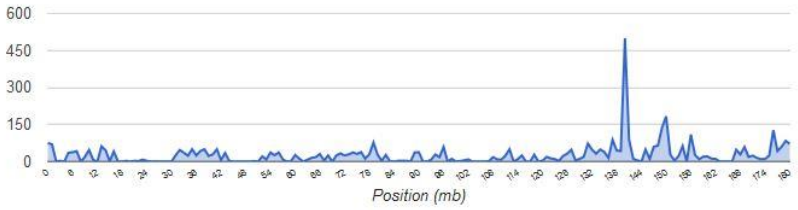




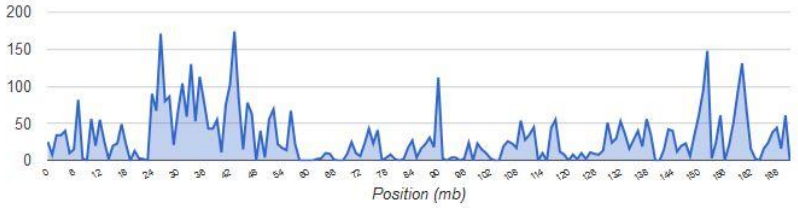




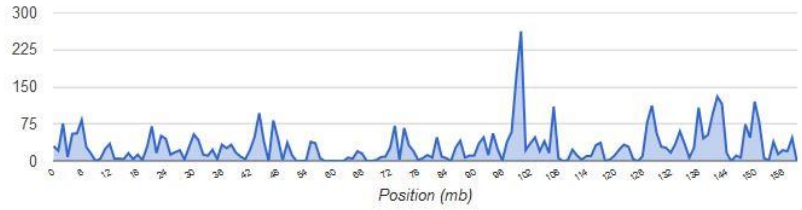
Distribution of variants on chromosome 5



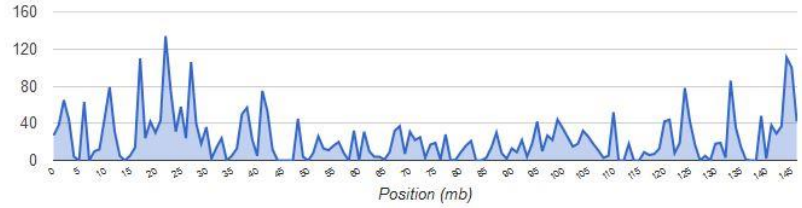
Distribution of variants on chromosome 6



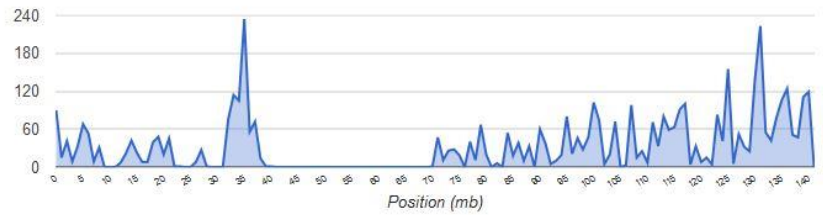
Distribution of variants on chromosome 7



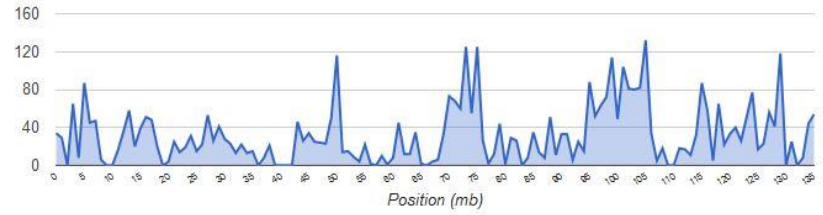
Distribution of variants on chromosome 8



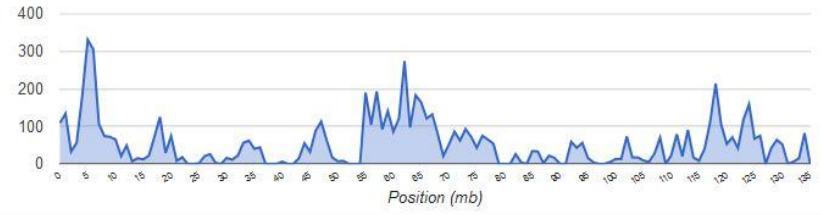
Distribution of variants on chromosome 9



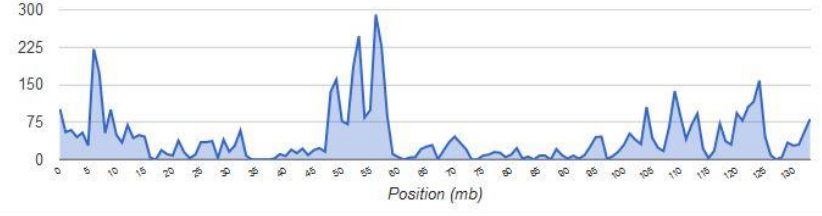
Distribution of variants on chromosome 10



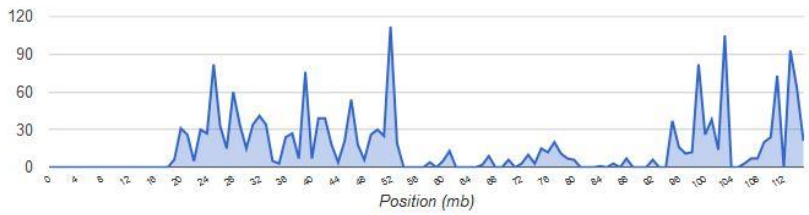
Distribution of variants on chromosome 11



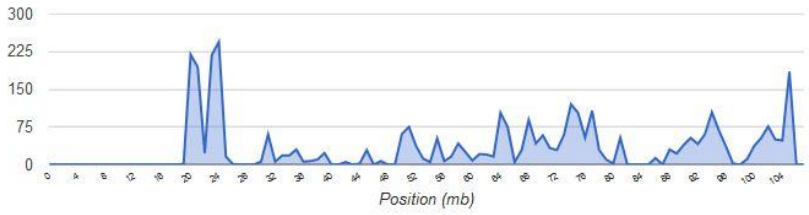
Distribution of variants on chromosome 12



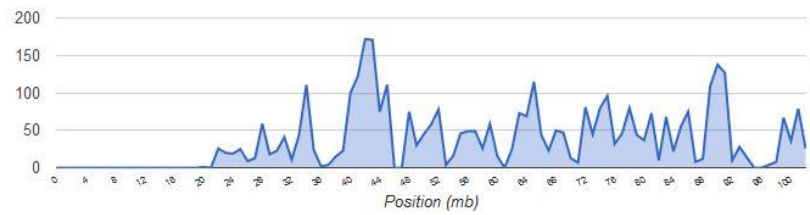
Distribution of variants on chromosome 13



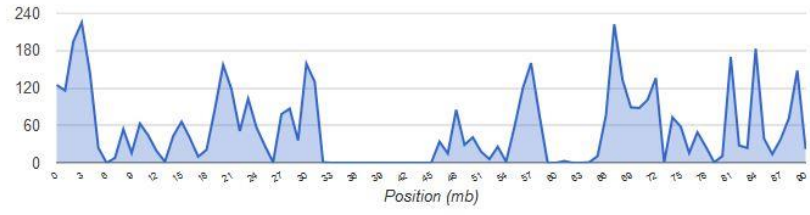
Distribution of variants on chromosome 14



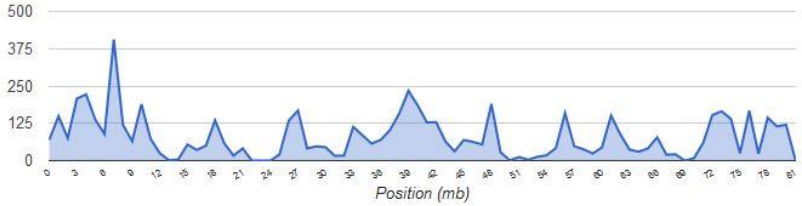
Distribution of variants on chromosome 15



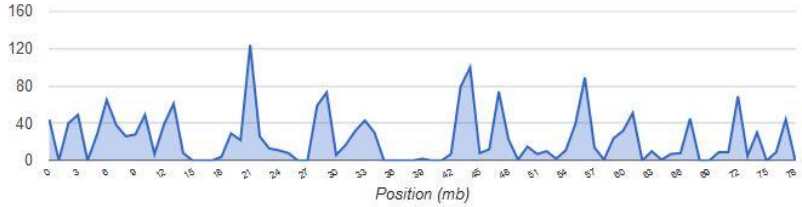
Distribution of variants on chromosome 16



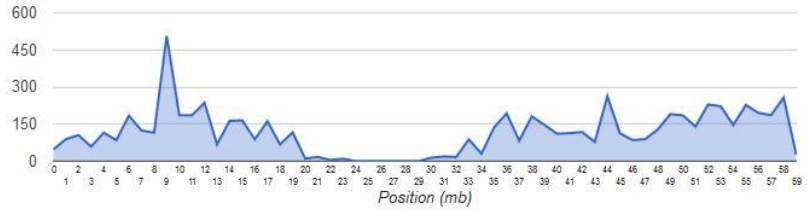
Distribution of variants on chromosome 17



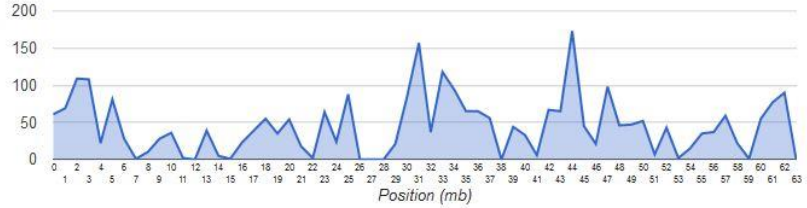
Distribution of variants on chromosome 18



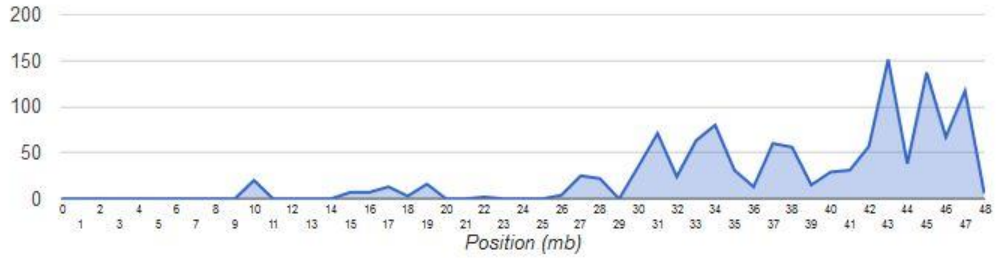
Distribution of variants on chromosome 19



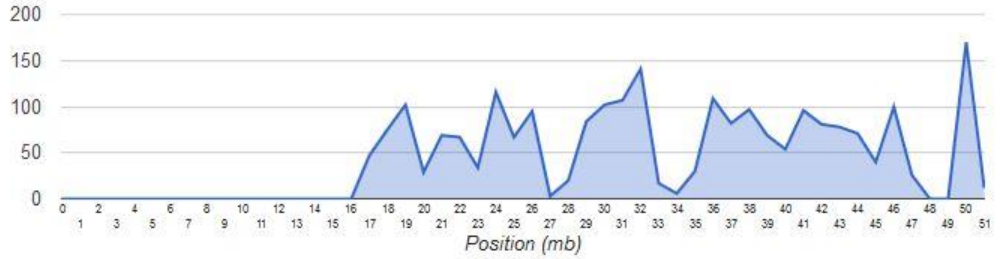
Distribution of variants on chromosome 20



Distribution of variants on chromosome 21



Distribution of variants on chromosome 22



Appendix IV for Chapter 5

Table 1. Variants in Other Muscle Genes in TAM Cases with Uncertain Genetic Results.

Case ID	Gene	Variants	Het/ Hom	RS ID	MAF in 1000G/EVS
64412	*AGL*	p.L409R	Het	rs200459772	N/A; 0.000231
	GPD1L	p.E290D	Het	N/A	N/A; N/A
	ANK2	p.S2785L	Het	rs145895389	0.0009; 0.001768
	ANK2	p.Y3936C	Het	N/A	N/A; 0.000077
	ATXN1	p.Q207H	Het	rs201030692	N/A; N/A
	TBP	p.Q64delinsHQQQ	Het	N/A	N/A; N/A
	TBP	p.Q94delinsQQ	Hom	N/A	N/A; N/A
	AP5Z1	p.R365W	Het	N/A	N/A; N/A
	SETX	p.A289V	Het	N/A	N/A; N/A
	ENTPD1	p.V21M	Het	rs150772804	N/A; 0.000538
	RBM20	p.S455L	Het	rs189569984	0.0014; 0.007227
	MYH8	p.R193H	Het	rs145863180	N/A; 0.000769
40725	*CACNA1S*	p.S1209R	Het	N/A	N/A; N/A
	HSPG2	p.L1731F	Het	rs201491948	N/A; 0.000077
	DYSF	p.V235M	Het	N/A	N/A; N/A
	SCN3B	p.N205S	Het	N/A	N/A; N/A
66209	*SYNE2*	p.Q3771H	Het	rs144596211	0.0014; 0.001845
	SYNE2	p.S6472L	Het	rs150955173	0.0005; 0.001615
	GBE1	p.P40fs	Hom	N/A	N/A; N/A
	ATXN3	p.Q283delinsQQQQQQQQ	Hom	N/A	N/A; N/A
	TTN	p.I10201T	Het	N/A	N/A; N/A
	ABHD5	p.R114L	Het	rs148743497	0.0032; 0.009303
	SYNE1	p.S3346Y	Het	rs150170988	0.0027; 0.004613
	AKAP9	p.Q3265E	Het	N/A	N/A; N/A
	ANO5	p.S796L	Het	rs61910685	0.0032; 0.009227
	AARS	p.G931S	Het	rs149377346	0.0032; 0.007618
	PDYN	p.E192V	Het	rs45469293	0.0041; 0.007535
	JPH2	p.E402K	Het	rs147407445	N/A; 0.000769
66203	*DES*	p.A135V	Het	N/A	N/A; N/A;
	OPA1	p.320_321del	Het	N/A	N/A; N/A;
	TTN	p.L13040S	Het	N/A	N/A; N/A
	ANK2	p.A940G	Het	N/A	N/A; N/A
	ATM	p.I2628fs	Het	N/A	N/A; N/A
66206	*MYH6*	p.R1398Q	Het	rs150815925	N/A; 0.000077
	ATM	p.A2274T	Het	N/A	N/A; N/A
	GNB4	p.318_318del	Het	N/A	N/A; N/A
64416	*CCDC88C*	p.R1537Q	Het	rs200878782	N/A; 0.000492
	GBE1	p.P40fs	Hom	N/A	N/A; N/A
	DDHD2	p.T625P	Het	N/A	N/A; 0.000154
	GNE	p.T95fs	Het	N/A	N/A; N/A
	IKBKAP	p.E1071D	Het	rs140024352	N/A; 0.000615
	DSC2	p.A895fs	Het	N/A	N/A; N/A
	CTDP1	p.T223M	Het	rs149796311	0.0005; 0.000538
29970	*AGRN*	p.A1514T	Het	N/A	0.007
	MFN2	p.G298R	Het	rs41278630	0.001;0.0036
	RYR2	p.R1013Q	Het	N/A	0.001
	RYR2	p.S1400G	Het	rs56229512	0.005
	TTN	p.V8996F	Het	rs111671438	0.003
	TTN	p.T3713S	Het	rs72648925	0.001

	TTN	p.P4650S	Het	N/A	0.003
	COL6A3	p.L205V	Het	rs113716915	0.0018; 0.0047;
	GPD1L	p.E290D	Het	N/A	N/A
	NEFL	p.L472fs	Hom	rs11340767	N/A
	ATM	p.S707P	Het	rs4986761	0.005
	CRYAB	p.G154S	Het	N/A	N/A
	PFKM	p.R696H	Hom	rs41291971	N/A
	ATXN2	p.N491S	Het	rs117851901	0.001
	SYNE2	p.Y1638H	Het	N/A	0.001
	SPG11	p.N1962S	Het	N/A	0.004
	POLG	p.G517V	Het	rs61752783	0.0038; 0.0060
	CACNA1A	p.E735A	Het	rs16019	0.0040; 0.0097
	PRX	p.D1079G	Het	N/A	N/A; N/A
	ETFB	p.P93fs	Het	N/A	N/A
65142	*ACTN2*	p.P9fs	Hom	N/A	N/A; N/A
	NEB	p.M1852R	Het	rs144180493	0.0037; 0.0035
	GNE	c.274-1->ACACACACACACACACAC	Het	N/A	N/A; N/A
	WNK1	p.L740fs	Het	rs34967262	N/A; N/A
	TMPO	p.L653F	Het	rs202035749	N/A; N/A
	SEPT9	p.V508I	Het	N/A	N/A; N/A
63568	*ACTN2*	p.R22G	Het	N/A	N/A; N/A
	RYR2	p.H671R	Het	N/A	N/A; N/A
	TTN	p.A30197G	Het	rs72648273	0.0005; 0.002847
	SPEG	p.R2646Q	Het	rs201143360	0.0009; N/A
	KIAA0196	p.T125I	Het	N/A	N/A; N/A
	TNNT3	p.P103S	Het	rs200540491	N/A; N/A
	WNK1	p.L740fs	Het	rs34967262	N/A; N/A
	TDP1	p.564_564del	Het	rs34694353	N/A; N/A
	PIP5K1C	p.228_229del	Het	N/A	N/A; N/A
	RTN2	p.G390S	Het	rs143937661	0.0022; 0.0025
19817	*SYNE1*	p.E705fs	Het	N/A	N/A; N/A
	CACNB2	p.K170N	Het	rs199539261	N/A; 0.000538
	ILK	p.249_253del	Het	N/A	N/A: N/A
	LITAF	p.T49M	Het	rs141862602	0.0009; 0.001077
	AFG3L2	p.E91G	Het	N/A	N/A; 0.000077
	AFG3L2	p.L772F	Het	rs117182113	0.0018; 0.001461
	KCNC3	p.D477N	Het	rs148033381	N/A; 0.000154
	VAPB	p.R184Q	Het	rs145483046	N/A; 0.000308
64414	*AGRN*	p.R1734H	Het	rs145444272	0.0027; 0.00324
	DCTN1	p.T1249I	Het	rs72466496	0.0005; 0.003306
	TTN	p.W13903C	Het	rs202094100	N/A; 0.00057
	DES	p.R212Q	Het	rs144261171	0.0005; 0.000154
	COL6A3	p.M2001V	Het	rs201539582	N/A; N/A
	AKAP9	p.S1376T	Het	rs144372406	0; 0.000923
	GNE	p.T95fs	Het	N/A	N/A; N/A
	CACNA1C	p.P817S	Het	rs112532048	0.0037; 0.003387
	MYH6	p.D1937N	Het	N/A	N/A; N/A
	TCAP	p.A118V	Het	rs143233087	0.0009; 0.000769
2589	*HSPG2*	p.A2826G	Het	N/A	N/A; N/A
	HSPG2	p.R413Q	Het	rs140621959	0.0010; 0.0021
	LPIN1	p.S203C	Het	N/A	N/A; N/A
	SPAST	p.P206T	Het	rs140094231	N/A; 0.000077
	SLC22A5	p.R254Q	Het	rs200699819	N/A; 0.000077
	CHAT	p.R628Q	Het	rs114545628	0.0009; 0.000384
	WNK1	p.L740fs	Het	rs34967262	N/A; N/A

	SYNE2	p.H625Y	Het	rs192061494	
	CACNA1A	p.E1014K	Het	rs16024	0.0012; 0.0037
	PRX	p.F936V	Het	N/A	N/A; N/A
1285	*ITPR1*	p.V1628A	Het	N/A	N/A; N/A
	SH3TC2	p.D1229V	Het	rs146920285	0.0032; 0.003306
	SYNE1	p.L7750F	Het	N/A	N/A; N/A
	CACNB2	p.A127V	Het	rs200367454	N/A; 0.000154
	RBM20	p.S455L	Het	rs189569984	0.0014; 0.007227
	WNK1	p.L740fs	Hom	rs34967262	N/A; N/A
	POLG	p.L392V	Het	rs145289229	0.0009; 0.001154
	DSG2	p.V56M	Het	rs121913013	0.0005; 0.002588
	PNPLA6	p.V797M	Het	rs145988230	N/A; 0.002076
MB7684	*ACTN2*	p.P9fs	Hom	N/A	N/A; N/A
	FIG4	p.I591V	Het	N/A	N/A; N/A
	TBP	p.Q64delinsQX	Het	N/A	N/A; N/A
	TPM2	p.P303fs	Het	N/A	N/A; N/A
	SPTBN2	p.A1266T	Het	N/A	N/A; 0.000077
	FGF14	p.L56F	Het	N/A	N/A; N/A
	TDP1	p.R562Q	Het	N/A	N/A; N/A
	ATXN3	p.S4fs	Het	N/A	N/A; N/A
	SBF1	p.G427D	Het	N/A	N/A; N/A
MB7052	*ACTN2*	p.P9fs	Hom	N/A	N/A; N/A
	TTN	p.A28373T	Het	N/A	N/A; N/A
	TTN	p.F22511C	Het	N/A	N/A; N/A
	SPEG	p.R146H	Het	N/A	N/A; N/A
	GBE1	p.P40fs	Hom	N/A	N/A; N/A
	TFG	p.P354A	Het	rs111356679	0.0023; 0.001461
	SYNE1	p.S6062L	Het	rs139790539	0.0009; 0.001845
	AP4M1	p.R283W	Het	N/A	N/A; N/A
	INF2	p.R922H	Het	N/A	N/A; N/A
	SPG11	p.L2357F	Het	rs139334167	0.0027; 0.001693
	SPG11	p.R2211H	Het	rs1441650	N/A; 0.001155
	AARS	p.P234S	Het	rs141840552	0.0023; 0.001231
	MYH2	p.S1261R	Het	rs34610503	0.0005; 0.002384
4580	*AGRN*	p.A897V	Het	rs116836855	0.0018; 0.001615
	AGRN	p.D1785N	Het	rs144245019	0.0041; 0.008473
	PGM1	p.S32R	Het	N/A	N/A; N/A
	B3GALNT2	p.S137L	Het	rs140393851	0.0005; 0.000231
	TTN	p.I31631V	Het	rs56347248	0.0014; 0.001475
	TTN	p.S29198F	Het	rs191484894	0.0014; 0.001459
	TTN	p.C27708G	Het	rs150430592	0.0014; 0.001443
	TTN	p.V19029M	Het	rs150661999	0.0014; 0.001464
	TTN	p.A12491T	Het	rs144668626	0.0014; 0.001502
	TTN	p.R7253H	Het	rs149855485	0.0014; 0.001494
	ITPR1	p.V479I	Het	rs41289628	0.0014; 0.005058
	SCN5A	p.T1620M	Het	rs199473282	N/A; N/A
	CYP2U1	p.N331S	Het	rs148983337	0.0014; 0.000923
	DSP	p.V30M	Het	rs121912998	0.0018; 0.001243
	RNF216	p.P233A	Het	N/A	N/A; N/A
	CLCN1	p.H29P	Het	rs146160029	0.0005; 0.002845
	ANO5	p.G231V	Het	rs137854523	0.0005; 0.000846
	PRPH	p.A10D	Het	N/A	N/A; N/A
	ZFYVE26	p.T2352I	Het	rs151166497	0.0037; 0.002614
	TECPR2	p.675_676del	Het	N/A	N/A; N/A
	MYH2	p.R1272C	Het	rs189738574	0.0005; N/A
	KCNJ2	p.T75M	Het	rs104894585	N/A; N/A
4893	*AGRN*	p.A375S	Het	rs138031468	0.0014; 0.005002

	TTN	p.Q25740P	Het	rs201674674	N/A; 0.000166
	TTN	p.R17799Q	Het	rs141973925	0.0014; 0.003334
	COL6A3	p.D563G	Het	rs112913396	0.0009; 0.001692
	GBE1	p.P40fs	Hom	N/A	N/A; N/A
	DNAJB6	p.R237H	Het	rs199612810	N/A; 0.000384
	SETX	p.A1478E	Het	rs143661911	0.0005; 0.000692
	PKP2	p.T338A	Het	rs139851304	0.0009; 0.001461
	PKP2	p.D26N	Het	rs143004808	0.0027; 0.004861
	INF2	p.A1155V	Het	rs199673276	N/A; N/A
9469	*HSPG2*	p.P181T	Het	N/A	N/A; N/A
	SEPN1	p.A105T	Het	rs201692549	N/A; 0.000489
	TTN	p.A30197G	Het	rs72648273	0.0005; 0.002847
	LAMB2	p.F1110S	Het	N/A	N/A; N/A
	GNE	p.92_92del	Het	N/A	N/A; N/A
	TMEM5	p.Y339C	Het	rs150736997	N/A; 0.000077
	MYH6	p.Q277H	Het	rs140660481	N/A; 0.000308
	SPG11	p.G2317D	Het	rs79186522	0.0005; N/A
	JUP	p.I208V	Het	N/A	N/A; N/A
3212	*ACTN2*	p.P9fs	Hom	N/A	N/A; N/A
	TTN	p.I13750T	Het	rs72677243	0.0014; 0.001236
	ITPR1	p.V1477I	Het	rs200646875	N/A; 0.000247
	MTO1	p.D292N	Het	N/A	N/A; N/A
	PRKAG2	p.A368S	Het	N/A	N/A; N/A
	GNE	c.274-1->ACACACACACACACAC	Hom	N/A	N/A; N/A
	SPTBN2	p.A2362T	Het	N/A	N/A; N/A
	WNK1	p.G612S	Het	rs146450828	0.0005; 0.000154
	ZFYVE26	p.R385W	Het	N/A	N/A; N/A
	POMT2	p.W262X	Het	rs200542239	N/A; N/A
	POLG	p.H932Y	Het	rs121918048	N/A; N/A
	DNMT1	p.V1050M	Het	rs187394074	0.0009; N/A
	CACNA1A	p.G637V	Het	N/A	N/A; N/A
	RYR1	p.P1293T	Het	rs146407179	0.0005; N/A
3938	*TARDBP*	p.L5P	Het	rs61730366	0.0014; 0.004383
	ACTN2	p.P9fs	Hom	N/A	N/A; N/A
	SDHA	p.L93Q	Het	N/A	N/A; N/A
	VCL	p.P990L	Het	N/A	N/A; N/A
	DSG2	p.V920G	Het	rs142841727	0.0037; 0.003863
	UBA1	p.A572V	Hom	N/A	N/A; N/A
	FLNA	p.Q1484R	Hom	rs200130356	0.0006; 0.001084
	FLNA	p.G905V	Het	N/A	N/A; N/A

References

ABDO, W. F., VAN MIERLO, T., HENGSTMAN, G. J., SCHELHAAS, H. J., VAN ENGELEN, B. G. & VERBEEK, M. M. 2009. Increased plasma amyloid-beta42 protein in sporadic inclusion body myositis. *Acta Neuropathol,* 118**,** 429-31.

ABRAMZON, Y., JOHNSON, J. O., SCHOLZ, S. W., TAYLOR, J. P., BRUNETTI, M., CALVO, A., MANDRIOLI, J., BENATAR, M., MORA, G., RESTAGNO, G., CHIO, A. & TRAYNOR, B. J. 2012. Valosin-containing protein (VCP) mutations in sporadic amyotrophic lateral sclerosis. *Neurobiol Aging,* 33**,** 2231 e1-2231 e6.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

AHMED, M., MACHADO, P. M., MILLER, A., SPICER, C., HERBELIN, L., HE, J., NOEL, J., WANG, Y., MCVEY, A. L., PASNOOR, M., GALLAGHER, P., STATLAND, J., LU, C. H., KALMAR, B., BRADY, S., SETHI, H., SAMANDOURAS, G., PARTON, M., HOLTON, J. L., WESTON, A., COLLINSON, L., TAYLOR, J. P., SCHIAVO, G., HANNA, M. G., BAROHN, R. J., DIMACHKIE, M. M. & GREENSMITH, L. 2016. Targeting protein homeostasis in sporadic inclusion body myositis. *Sci Transl Med,* 8**,** 331ra41.

ALAZAMI, A. M., ALZAHRANI, F., BOHLEGA, S. & ALKURAYA, F. S. 2014. SET binding factor 1 (SBF1) mutation causes Charcot-Marie-tooth disease type 4B3. *Neurology,* 82**,** 1665-6.

ALLENBACH, Y., CHAARA, W., ROSENZWAJG, M., SIX, A., PREVEL, N., MINGOZZI, F., WANSCHITZ, J., MUSSET, L., CHARUEL, J. L., EYMARD, B., SALOMON, B., DUYCKAERTS, C., MAISONOBE, T., DUBOURG, O., HERSON, S., KLATZMANN, D. & BENVENISTE, O. 2014. Th1 response and systemic treg deficiency in inclusion body myositis. *PLoS One,* 9**,** e88788.

ALTMÜLLER, J., PALMER, L. J., FISCHER, G., SCHERB, H. & WJST, M. 2001. Genomewide Scans of Complex Human Diseases: True Linkage Is Hard to Find. *Am J Hum Genet,* 69**,** 936-950.

AMATO, A. A. & SHEBERT, R. T. 1998. Inclusion body myositis in twins. *Neurology,* 51**,** 598-600.

AMBERGER, J. S., BOCCHINI, C. A., SCHIETTECATTE, F., SCOTT, A. F. & HAMOSH, A. 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res,* 43**,** D789-98.

ANDERSON, C. A., PETTERSSON, F. H., CLARKE, G. M., CARDON, L. R., MORRIS, A. P. & ZONDERVAN, K. T. 2010. Data quality control in genetic case-control association studies. *Nat. Protocols,* 5**,** 1564-1573.

ARAI, T., HASEGAWA, M., AKIYAMA, H., IKEDA, K., NONAKA, T., MORI, H., MANN, D., TSUCHIYA, K., YOSHIDA, M., HASHIZUME, Y. & ODA, T. 2006. TDP-43 is a component of ubiquitin-positive tau-negative inclusions

in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun,* 351**,** 602-11.

ASKANAS, V., ALVAREZ, R. B. & ENGEL, W. K. 1993a. beta-Amyloid precursor epitopes in muscle fibers of inclusion body myositis. *Ann Neurol,* 34**,** 551-60.

ASKANAS, V., BILAK, M., ENGEL, W. K., ALVAREZ, R. B., TOME, F. & LECLERC, A. 1993b. Prion protein is abnormally accumulated in inclusion-body myositis. *Neuroreport,* 5**,** 25-8.

ASKANAS, V. & ENGEL, W. K. 2001. Inclusion-body myositis: newest concepts of pathogenesis and relation to aging and Alzheimer disease. *J Neuropathol Exp Neurol,* 60**,** 1-14.

ASKANAS, V. & ENGEL, W. K. 2003. Unfolding story of inclusion-body myositis and myopathies: role of misfolded proteins, amyloid-beta, cholesterol, and aging. *J Child Neurol,* 18**,** 185-90.

ASKANAS, V. & ENGEL, W. K. 2005a. Molecular pathology and pathogenesis of inclusion-body myositis. *Microsc Res Tech,* 67**,** 114-20.

ASKANAS, V. & ENGEL, W. K. 2005b. Sporadic inclusion-body myositis: a proposed key pathogenetic role of the abnormalities of the ubiquitin-proteasome system, and protein misfolding and aggregation. *Acta Myol,* 24**,** 17-24.

ASKANAS, V. & ENGEL, W. K. 2006. Inclusion-body myositis: a myodegenerative conformational disorder associated with Abeta, protein misfolding, and proteasome inhibition. *Neurology,* 66**,** S39-48.

ASKANAS, V. & ENGEL, W. K. 2008. Inclusion-body myositis: muscle-fiber molecular pathology and possible pathogenic significance of its similarity to Alzheimer's and Parkinson's disease brains. *Acta Neuropathol,* 116**,** 583-95.

ASKANAS, V., ENGEL, W. K., ALVAREZ, R. B. & GLENNER, G. G. 1992. beta-Amyloid protein immunoreactivity in muscle of patients with inclusion-body myositis. *Lancet,* 339**,** 560-1.

ASKANAS, V., ENGEL, W. K., MIRABELLA, M., WEISGRABER, K. H., SAUNDERS, A. M., ROSES, A. D. & MCFERRIN, J. 1996a. Apolipoprotein E alleles in sporadic inclusion-body myositis and hereditary inclusion-body myopathy. *Ann Neurol,* 40**,** 264-5.

ASKANAS, V., ENGEL, W. K. & NOGALSKA, A. 2009. Inclusion body myositis: a degenerative muscle disease associated with intra-muscle fiber multi-protein aggregates, proteasome inhibition, endoplasmic reticulum stress and decreased lysosomal degradation. *Brain Pathol,* 19**,** 493-506.

ASKANAS, V., ENGEL, W. K. & NOGALSKA, A. 2012. Pathogenic considerations in sporadic inclusion-body myositis, a degenerative muscle disease associated with aging and abnormalities of myoproteostasis. *J Neuropathol Exp Neurol,* 71**,** 680-93.

ASKANAS, V., ENGEL, W. K. & NOGALSKA, A. 2015. Sporadic inclusion-body myositis: A degenerative muscle disease associated with aging, impaired muscle protein homeostasis and abnormal mitophagy. *Biochim Biophys Acta,* 1852**,** 633-43.

ASKANAS, V., ENGEL, W. K., YANG, C. C., ALVAREZ, R. B., LEE, V. M. & WISNIEWSKI, T. 1998. Light and electron microscopic immunolocalization of presenilin 1 in abnormal muscle fibers of patients with sporadic inclusion-body myositis and autosomal-recessive inclusion-body myopathy. *Am J Pathol,* 152**,** 889-95.

ASKANAS, V., MCFERRIN, J., BAQUE, S., ALVAREZ, R. B., SARKOZI, E. & ENGEL, W. K. 1996b. Transfer of beta-amyloid precursor protein gene using adenovirus vector causes mitochondrial abnormalities in cultured normal human muscle. *Proc Natl Acad Sci U S A,* 93**,** 1314-9.

ASKANAS, V., MIRABELLA, M., ENGEL, W. K., ALVAREZ, R. B. & WEISGRABER, K. H. 1994. Apolipoprotein E immunoreactive deposits in inclusion-body muscle diseases. *Lancet,* 343**,** 364-5.

AZZEDINE, H., BOLINO, A., TAIEB, T., BIROUK, N., DI DUCA, M., BOUHOUCHE, A., BENAMOU, S., MRABET, A., HAMMADOUCHE, T., CHKILI, T., GOUIDER, R., RAVAZZOLO, R., BRICE, A., LAPORTE, J. & LEGUERN, E. 2003. Mutations in MTMR13, a new pseudophosphatase homologue of MTMR2 and Sbf1, in two families with an autosomal recessive demyelinating form of Charcot-Marie-Tooth disease associated with early-onset glaucoma. *Am J Hum Genet,* 72**,** 1141-53.

BADRISING, U. A., MAAT-SCHIEMAN, M., VAN DUINEN, S. G., BREEDVELD, F., VAN DOORN, P., VAN ENGELEN, B., VAN DEN HOOGEN, F., HOOGENDIJK, J., HOWELER, C., DE JAGER, A., JENNEKENS, F., KOEHLER, P., VAN DER LEEUW, H., DE VISSER, M., VERSCHUUREN, J. J. & WINTZEN, A. R. 2000. Epidemiology of inclusion body myositis in the Netherlands: a nationwide study. *Neurology,* 55**,** 1385-7.

BADRISING, U. A., SCHREUDER, G. M., GIPHART, M. J., GELEIJNS, K., VERSCHUUREN, J. J., WINTZEN, A. R., MAAT-SCHIEMAN, M. L., VAN DOORN, P., VAN ENGELEN, B. G., FABER, C. G., HOOGENDIJK, J. E., DE JAGER, A. E., KOEHLER, P. J., DE VISSER, M. & VAN DUINEN, S. G. 2004. Associations with autoimmune disorders and HLA class I and II antigens in inclusion body myositis. *Neurology,* 63**,** 2396-8.

BAKER, M., LITVAN, I., HOULDEN, H., ADAMSON, J., DICKSON, D., PEREZ-TUR, J., HARDY, J., LYNCH, T., BIGIO, E. & HUTTON, M. 1999. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet,* 8**,** 711-5.

BAKER, N. S., SARNAT, H. B., JACK, R. M., PATTERSON, K., SHAW, D. W. & HERNDON, S. P. 1997. D-2-Hydroxyglutaric Aciduria: Hypotonia, Cortical Blindness, Seizures, Cardiomyopathy, and Cylindrical Spirals in Skeletal Muscle. *J Child Neurol,* 12**,** 31-36.

BALABANSKI, L., ANTOV, G., DIMOVA, I., IVANOV, S., NACHEVA, M., GAVRILOV, I., NESHEVA, D., RUKOVA, B., HADJIDEKOVA, S., MALINOV, M. & TONCHEVA, D. 2014. Next-generation sequencing of BRCA1 and BRCA2 in breast cancer patients and control subjects. *Mol Clin Oncol,* 2**,** 435-439.

BALDING, D. J. 2006. A tutorial on statistical methods for population association studies. *Nat Rev Genet,* 7**,** 781-791.

BALLESTER, L. Y., LUTHRA, R., KANAGAL-SHAMANNA, R. & SINGH, R. R. 2016. Advances in clinical next-generation sequencing: target enrichment and sequencing technologies. *Expert Rev Mol Diagn,* 16**,** 357-72.

BANG, M.-L., MUDRY, R. E., MCELHINNY, A. S., TROMBITÁS, K., GEACH, A. J., YAMASAKI, R., SORIMACHI, H., GRANZIER, H., GREGORIO, C. C. & LABEIT, S. 2001. Myopalladin, a Novel 145-Kilodalton Sarcomeric Protein with Multiple Roles in Z-Disc and I-Band Protein Assemblies. *J Cell Biol,* 153**,** 413-428.

BAROHN, R. J., AMATO, A. A., SAHENK, Z., KISSEL, J. T. & MENDELL, J. R. 1995. Inclusion body myositis: explanation for poor response to immunosuppressive therapy. *Neurology,* 45**,** 1302-4.

BARRETT, J. C., FRY, B., MALLER, J. & DALY, M. J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* 21**,** 263-265.

BARTHELEMY, F., WEIN, N., KRAHN, M., LEVY, N. & BARTOLI, M. 2011. Translational research and therapeutic perspectives in dysferlinopathies. *Mol Med,* 17**,** 875-82.

BARTOLOME, F., WU, H. C., BURCHELL, V. S., PREZA, E., WRAY, S., MAHONEY, C. J., FOX, N. C., CALVO, A., CANOSA, A., MOGLIA, C., MANDRIOLI, J., CHIO, A., ORRELL, R. W., HOULDEN, H., HARDY, J., ABRAMOV, A. Y. & PLUN-FAVREAU, H. 2013. Pathogenic VCP mutations induce mitochondrial uncoupling and reduced ATP levels. *Neuron,* 78**,** 57-64.

BECK, J., POULTER, M., HENSMAN, D., ROHRER, J. D., MAHONEY, C. J., ADAMSON, G., CAMPBELL, T., UPHILL, J., BORG, A., FRATTA, P., ORRELL, R. W., MALASPINA, A., ROWE, J., BROWN, J., HODGES, J., SIDLE, K., POLKE, J. M., HOULDEN, H., SCHOTT, J. M., FOX, N. C., ROSSOR, M. N., TABRIZI, S. J., ISAACS, A. M., HARDY, J., WARREN, J. D., COLLINGE, J. & MEAD, S. 2013. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet,* 92**,** 345-53.

BEHIN, A., SALORT-CAMPANA, E., WAHBI, K., RICHARD, P., CARLIER, R. Y., CARLIER, P., LAFORET, P., STOJKOVIC, T., MAISONOBE, T., VERSCHUEREN, A., FRANQUES, J., ATTARIAN, S., MAUES DE PAULA, A., FIGARELLA-BRANGER, D., BECANE, H. M., NELSON, I., DUBOC, D., BONNE, G., VICART, P., UDD, B., ROMERO, N., POUGET, J. & EYMARD, B. 2015. Myofibrillar myopathies: State of the art, present and future challenges. *Rev Neurol (Paris),* 171**,** 715-29.

BELAYA, K., FINLAYSON, S., SLATER, C. R., COSSINS, J., LIU, W. W., MAXWELL, S., MCGOWAN, S. J., MASLAU, S., TWIGG, S. R., WALLS, T. J., PASCUAL PASCUAL, S. I., PALACE, J. & BEESON, D. 2012. Mutations in DPAGT1 cause a limb-girdle congenital myasthenic syndrome with tubular aggregates. *Am J Hum Genet,* 91**,** 193-201.

BENDERS, A. A., VEERKAMP, J. H., OOSTERHOF, A., JONGEN, P. J., BINDELS, R. J., SMIT, L. M., BUSCH, H. F. & WEVERS, R. A. 1994. Ca2+ homeostasis in Brody's disease. A study in skeletal muscle and cultured muscle cells and the effects of dantrolene an verapamil. *J Clin Invest,* 94**,** 741-8.

BENVENISTE, O. & HILTON-JONES, D. 2010. International Workshop on Inclusion Body Myositis held at the Institute of Myology, Paris, on 29 May 2009. *Neuromuscul Disord,* 20**,** 414-21.

BERCIANO, J., GALLARDO, E., DOMÍNGUEZ-PERLES, R., GARCÍA, A., GARCÍA-BARREDO, R., COMBARROS, O., INFANTE, J. & ILLA, I. 2008. Autosomal-dominant distal myopathy with a myotilin S55F mutation: sorting out the phenotype. *Journal of Neurology, Neurosurgery & Psychiatry,* 79**,** 205-208.

BERGER, P., BERGER, I., SCHAFFITZEL, C., TERSAR, K., VOLKMER, B. & SUTER, U. 2006. Multi-level regulation of myotubularin-related protein-2

phosphatase activity by myotubularin-related protein-13/set-binding factor-2. *Hum Mol Genet,* 15**,** 569-79.

BERSANO, A., DEL BO, R., LAMPERTI, C., GHEZZI, S., FAGIOLARI, G., FORTUNATO, F., BALLABIO, E., MOGGIO, M., CANDELISE, L., GALIMBERTI, D., VIRGILIO, R., LANFRANCONI, S., TORRENTE, Y., CARPO, M., BRESOLIN, N., COMI, G. P. & CORTI, S. 2009. Inclusion body myopathy and frontotemporal dementia caused by a novel VCP mutation. *Neurobiol Aging,* 30**,** 752-8.

BETTENCOURT, C., RAPOSO, M., KAZACHKOVA, N., CYMBRON, T., SANTOS, C., KAY, T., VASCONCELOS, J., MACIEL, P., DONIS, K. C., SARAIVA-PEREIRA, M. L., JARDIM, L. B., SEQUEIROS, J. & LIMA, M. 2011. The APOE epsilon2 allele increases the risk of earlier age at onset in Machado-Joseph disease. *Arch Neurol,* 68**,** 1580-3.

BEVILACQUA, J. A., BITOUN, M., BIANCALANA, V., OLDFORS, A., STOLTENBURG, G., CLAEYS, K. G., LACENE, E., BROCHIER, G., MANERE, L., LAFORET, P., EYMARD, B., GUICHENEY, P., FARDEAU, M. & ROMERO, N. B. 2009. "Necklace" fibers, a new histological marker of late-onset MTM1-related centronuclear myopathy. *Acta Neuropathol,* 117**,** 283-91.

BIENIEK, K. F., MURRAY, M. E., RUTHERFORD, N. J., CASTANEDES-CASEY, M., DEJESUS-HERNANDEZ, M., LIESINGER, A. M., BAKER, M. C., BOYLAN, K. B., RADEMAKERS, R. & DICKSON, D. W. 2013. Tau pathology in frontotemporal lobar degeneration with C9ORF72 hexanucleotide repeat expansion. *Acta Neuropathol,* 125**,** 289-302.

BILAK, M., ASKANAS, V. & ENGEL, W. K. 1993. Strong immunoreactivity of alpha 1-antichymotrypsin co-localizes with beta-amyloid protein and ubiquitin in vacuolated muscle fibers of inclusion-body myositis. *Acta Neuropathol,* 85**,** 378-82.

BOHLEGA, S., ALAZAMI, A. M., CUPLER, E., AL-HINDI, H., IBRAHIM, E. & ALKURAYA, F. S. 2011. A novel syndromic form of sensory-motor polyneuropathy is linked to chromosome 22q13.31-q13.33. *Clin Genet,* 79**,** 193-5.

BOHM, J., CHEVESSIER, F., KOCH, C., PECHE, G. A., MORA, M., MORANDI, L., PASANISI, B., MORONI, I., TASCA, G., FATTORI, F., RICCI, E., PENISSON-BESNIER, I., NADAJ-PAKLEZA, A., FARDEAU, M., JOSHI, P. R., DESCHAUER, M., ROMERO, N. B., EYMARD, B. & LAPORTE, J. 2014. Clinical, histological and genetic characterisation of patients with tubular aggregate myopathy caused by mutations in STIM1. *J Med Genet,* 51**,** 824-33.

BOHM, J., CHEVESSIER, F., MAUES DE PAULA, A., KOCH, C., ATTARIAN, S., FEGER, C., HANTAI, D., LAFORET, P., GHORAB, K., VALLAT, J. M., FARDEAU, M., FIGARELLA-BRANGER, D., POUGET, J., ROMERO, N. B., KOCH, M., EBEL, C., LEVY, N., KRAHN, M., EYMARD, B., BARTOLI, M. & LAPORTE, J. 2013. Constitutive activation of the calcium sensor STIM1 causes tubular-aggregate myopathy. *Am J Hum Genet,* 92**,** 271-8.

BOLINO, A., MUGLIA, M., CONFORTI, F. L., LEGUERN, E., SALIH, M. A. M., GEORGIOU, D.-M., CHRISTODOULOU, K., HAUSMANOWA-PETRUSEWICZ, I., MANDICH, P., SCHENONE, A., GAMBARDELLA, A., BONO, F., QUATTRONE, A., DEVOTO, M. & MONACO, A. P. 2000.

Charcot-Marie-Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2. *Nat Genet,* 25**,** 17-19.

BONCOMPAGNI, S., PROTASI, F. & FRANZINI-ARMSTRONG, C. 2012. Sequential stages in the age-dependent gradual formation and accumulation of tubular aggregates in fast twitch muscle fibers: SERCA and calsequestrin involvement. *Age (Dordr),* 34**,** 27-41.

BOVE, K. E., IANNACCONE, S. T., HILTON, P. K. & SAMAHA, F. 1980. Cylindrical spirals in a familial neuromuscular disorder. *Ann Neurol,* 7**,** 550-6.

BRADLEY, W. G., TAYLOR, R., RICE, D. R. & ET AL. 1990. PRogressive myopathy in hyperkalemic periodic paralysis. *Arch Neurol,* 47**,** 1013-1017.

BRADY, S., SQUIER, W. & HILTON-JONES, D. 2013. Clinical assessment determines the diagnosis of inclusion body myositis independently of pathological features. *J Neurol Neurosurg Psychiatry,* 84**,** 1240-6.

BRADY, S., SQUIER, W., SEWRY, C., HANNA, M., HILTON-JONES, D. & HOLTON, J. L. 2014. A retrospective cohort study identifying the principal pathological features useful in the diagnosis of inclusion body myositis. *BMJ Open,* 4**,** e004552.

BRAS, J., GUERREIRO, R. & HARDY, J. 2012. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci,* 13**,** 453-64.

BROCCOLINI, A. & MIRABELLA, M. 2015. Hereditary inclusion-body myopathies. *Biochim Biophys Acta,* 1852**,** 644-650.

BROWN, I. R. 2007. Heat shock proteins and protection of the nervous system. *Ann N Y Acad Sci,* 1113**,** 147-58.

BUCELLI, R. C., ARHZAOUY, K., PESTRONK, A., PITTMAN, S. K., ROJAS, L., SUE, C. M., EVILA, A., HACKMAN, P., UDD, B., HARMS, M. B. & WEIHL, C. C. 2015. SQSTM1 splice site mutation in distal myopathy with rimmed vacuoles. *Neurology,* 85**,** 665-74.

CACCIOTTOLO, M., NOGALSKA, A., D'AGOSTINO, C., ENGEL, W. K. & ASKANAS, V. 2013a. Chaperone-mediated autophagy components are upregulated in sporadic inclusion-body myositis muscle fibres. *Neuropathol Appl Neurobiol*.

CACCIOTTOLO, M., NOGALSKA, A., D'AGOSTINO, C., ENGEL, W. K. & ASKANAS, V. 2013b. Dysferlin is a newly identified binding partner of AbetaPP and it co-aggregates with amyloid-beta42 within sporadic inclusion-body myositis (s-IBM) muscle fibers. *Acta Neuropathol,* 126**,** 781-3.

CAMERON, C. H., ALLEN, I. V., PATTERSON, V. & AVARIA, M. A. 1992. Dominantly inherited tubular aggregate myopathy. *J Pathol,* 168**,** 397-403.

CARPENTER, S., KARPATI, G., ROBITAILLE, Y. & MELMED, C. 1979. Cylindrical spirals in human skeletal muscle. *Muscle Nerve,* 2**,** 282-287.

CARSON, A. R., SMITH, E. N., MATSUI, H., BRAEKKAN, S. K., JEPSEN, K., HANSEN, J. B. & FRAZER, K. A. 2014. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics,* 15**,** 125.

CARSS, KEREN J., STEVENS, E., FOLEY, A. R., CIRAK, S., RIEMERSMA, M., TORELLI, S., HOISCHEN, A., WILLER, T., VAN SCHERPENZEEL, M., MOORE, STEVEN A., MESSINA, S., BERTINI, E., BÖNNEMANN, CARSTEN G., ABDENUR, JOSE E., GROSMANN, CARLA M., KESARI, A., PUNETHA, J., QUINLIVAN, R., WADDELL, LEIGH B., YOUNG, HELEN K., WRAIGE, E., YAU, S., BRODD, L., FENG, L., SEWRY, C.,

MACARTHUR, DANIEL G., NORTH, KATHRYN N., HOFFMAN, E., STEMPLE, DEREK L., HURLES, MATTHEW E., VAN BOKHOVEN, H., CAMPBELL, KEVIN P., LEFEBER, DIRK J., LIN, Y.-Y. & MUNTONI, F. 2013. Mutations in GDP-Mannose Pyrophosphorylase B Cause Congenital and Limb-Girdle Muscular Dystrophies Associated with Hypoglycosylation of α-Dystroglycan. *The American Journal of Human Genetics,* 93**,** 29-41.

CASAR-BOROTA, O., JACOBSSON, J., LIBELIUS, R., OLDFORS, C. H., MALFATTI, E., ROMERO, N. B. & OLDFORS, A. 2015. A novel dynamin-2 gene mutation associated with a late-onset centronuclear myopathy with necklace fibres. *Neuromuscul Disord,* 25**,** 345-8.

CATALAN-GARCIA, M., GARRABOU, G., MOREN, C., GUITART-MAMPEL, M., HERNANDO, A., DIAZ-RAMOS, A., GONZALEZ-CASACUBERTA, I., JUAREZ-FLORES, D. L., BANO, M., ENRICH-BENGOA, J., EMPERADOR, S., MILISENDA, J., MORENO, P., TOBIAS, E., ZORZANO, A., MONTOYA, J., CARDELLACH, F. & GRAU, J. M. 2016. Mitochondrial-DNA disturbances and deregulated expression of OXPHOS and mitochondrial fusion proteins in sporadic inclusion body myositis. *Clin Sci (Lond).*

CHAN, N., LE, C., SHIEH, P., MOZAFFAR, T., KHARE, M., BRONSTEIN, J. & KIMONIS, V. 2012. Valosin-containing protein mutation and Parkinson's disease. *Parkinsonism Relat Disord,* 18**,** 107-9.

CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience,* 4**,** 1-16.

CHEN, X., GHRIBI, O. & GEIGER, J. D. 2008. Rabbits fed cholesterol-enriched diets exhibit pathological features of inclusion body myositis. *Am J Physiol Regul Integr Comp Physiol,* 294**,** R829-35.

CHEVESSIER, F., BAUCHE-GODARD, S., LEROY, J. P., KOENIG, J., PATURNEAU-JOUAS, M., EYMARD, B., HANTAI, D. & VERDIERE-SAHUQUE, M. 2005. The origin of tubular aggregates in human myopathies. *J Pathol,* 207**,** 313-23.

CHEVESSIER, F., MARTY, I., PATURNEAU-JOUAS, M., HANTAI, D. & VERDIERE-SAHUQUE, M. 2004. Tubular aggregates are from whole sarcoplasmic reticulum origin: alterations in calcium binding protein expression in mouse skeletal muscle during aging. *Neuromuscul Disord,* 14**,** 208-16.

CHOI, M., SCHOLL, U. I., JI, W., LIU, T., TIKHONOVA, I. R., ZUMBO, P., NAYIR, A., BAKKALOGLU, A., OZEN, S., SANJAD, S., NELSON-WILLIAMS, C., FARHI, A., MANE, S. & LIFTON, R. P. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A,* 106**,** 19096-101.

CIRULLI, E. T., LASSEIGNE, B. N., PETROVSKI, S., SAPP, P. C., DION, P. A., LEBLOND, C. S., COUTHOUIS, J., LU, Y. F., WANG, Q., KRUEGER, B. J., REN, Z., KEEBLER, J., HAN, Y., LEVY, S. E., BOONE, B. E., WIMBISH, J. R., WAITE, L. L., JONES, A. L., CARULLI, J. P., DAY-WILLIAMS, A. G., STAROPOLI, J. F., XIN, W. W., CHESI, A., RAPHAEL, A. R., MCKENNA-YASEK, D., CADY, J., VIANNEY DE JONG, J. M., KENNA, K. P., SMITH, B. N., TOPP, S., MILLER, J., GKAZI, A., AL-CHALABI, A., VAN DEN BERG, L. H., VELDINK, J., SILANI, V., TICOZZI, N., SHAW, C. E., BALOH, R. H., APPEL, S., SIMPSON, E.,

LAGIER-TOURENNE, C., PULST, S. M., GIBSON, S., TROJANOWSKI, J. Q., ELMAN, L., MCCLUSKEY, L., GROSSMAN, M., SHNEIDER, N. A., CHUNG, W. K., RAVITS, J. M., GLASS, J. D., SIMS, K. B., VAN DEERLIN, V. M., MANIATIS, T., HAYES, S. D., ORDUREAU, A., SWARUP, S., LANDERS, J., BAAS, F., ALLEN, A. S., BEDLACK, R. S., HARPER, J. W., GITLER, A. D., ROULEAU, G. A., BROWN, R., HARMS, M. B., COOPER, G. M., HARRIS, T., MYERS, R. M. & GOLDSTEIN, D. B. 2015. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science,* 347**,** 1436-41.

CLAEYS, K. G. & FARDEAU, M. 2013. Myofibrillar myopathies. *Handb Clin Neurol,* 113**,** 1337-42.

CORTESE, A., MACHADO, P., MORROW, J., DEWAR, L., HISCOCK, A., MILLER, A., BRADY, S., HILTON-JONES, D., PARTON, M. & HANNA, M. G. 2013. Longitudinal observational study of sporadic inclusion body myositis: implications for clinical trials. *Neuromuscul Disord,* 23**,** 404-12.

CORTESE, A., PLAGNOL, V., BRADY, S., SIMONE, R., LASHLEY, T., ACEVEDO-AROZENA, A., DE SILVA, R., GREENSMITH, L., HOLTON, J., HANNA, M. G., FISHER, E. M. & FRATTA, P. 2014. Widespread RNA metabolism impairment in sporadic inclusion body myositis TDP43-proteinopathy. *Neurobiol Aging,* 35**,** 1491-8.

COSSINS, J., BELAYA, K., HICKS, D., SALIH, M. A., FINLAYSON, S., CARBONI, N., LIU, W. W., MAXWELL, S., ZOLTOWSKA, K., FARSANI, G. T., LAVAL, S., SEIDHAMED, M. Z., DONNELLY, P., BENTLEY, D., MCGOWAN, S. J., MULLER, J., PALACE, J., LOCHMULLER, H. & BEESON, D. 2013. Congenital myasthenic syndromes due to mutations in ALG2 and ALG14. *Brain,* 136**,** 944-56.

COX, F. M., BOON, E. M., VAN DER LANS, C. A., BAKKER, E., VERSCHUUREN, J. J. & BADRISING, U. A. 2010. TREX1 mutations are not associated with sporadic inclusion body myositis. *Eur J Neurol,* 17**,** 1108-9.

COX, F. M., TITULAER, M. J., SONT, J. K., WINTZEN, A. R., VERSCHUUREN, J. J. & BADRISING, U. A. 2011. A 12-year follow-up in sporadic inclusion body myositis: an end stage with major disabilities. *Brain,* 134**,** 3167-75.

CRUCHAGA, C., KARCH, C. M., JIN, S. C., BENITEZ, B. A., CAI, Y., GUERREIRO, R., HARARI, O., NORTON, J., BUDDE, J., BERTELSEN, S., JENG, A. T., COOPER, B., SKORUPA, T., CARRELL, D., LEVITCH, D., HSU, S., CHOI, J., RYTEN, M., HARDY, J., TRABZUNI, D., WEALE, M. E., RAMASAMY, A., SMITH, C., SASSI, C., BRAS, J., GIBBS, J. R., HERNANDEZ, D. G., LUPTON, M. K., POWELL, J., FORABOSCO, P., RIDGE, P. G., CORCORAN, C. D., TSCHANZ, J. T., NORTON, M. C., MUNGER, R. G., SCHMUTZ, C., LEARY, M., DEMIRCI, F. Y., BAMNE, M. N., WANG, X., LOPEZ, O. L., GANGULI, M., MEDWAY, C., TURTON, J., LORD, J., BRAAE, A., BARBER, I., BROWN, K., PASSMORE, P., CRAIG, D., JOHNSTON, J., MCGUINNESS, B., TODD, S., HEUN, R., KOLSCH, H., KEHOE, P. G., HOOPER, N. M., VARDY, E. R., MANN, D. M., PICKERING-BROWN, S., KALSHEKER, N., LOWE, J., MORGAN, K., DAVID SMITH, A., WILCOCK, G., WARDEN, D., HOLMES, C., PASTOR, P., LORENZO-BETANCOR, O., BRKANAC, Z., SCOTT, E., TOPOL, E., ROGAEVA, E., SINGLETON, A. B., KAMBOH, M. I., ST GEORGE-HYSLOP, P., CAIRNS, N., MORRIS, J. C., KAUWE, J.

S. & GOATE, A. M. 2014. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature,* 505**,** 550-4.

CRUCHAGA, C., NOWOTNY, P., KAUWE, J. S., RIDGE, P. G., MAYO, K., BERTELSEN, S., HINRICHS, A., FAGAN, A. M., HOLTZMAN, D. M., MORRIS, J. C. & GOATE, A. M. 2011. Association and expression analyses with single-nucleotide polymorphisms in TOMM40 in Alzheimer disease. *Arch Neurol,* 68**,** 1013-9.

CUSTER, S. K., NEUMANN, M., LU, H., WRIGHT, A. C. & TAYLOR, J. P. 2010. Transgenic mice expressing mutant forms VCP/p97 recapitulate the full spectrum of IBMPFD including degeneration in muscle, brain and bone. *Hum Mol Genet,* 19**,** 1741-55.

CUYVERS, E., VAN DER ZEE, J., BETTENS, K., ENGELBORGHS, S., VANDENBULCKE, M., ROBBERECHT, C., DILLEN, L., MERLIN, C., GEERTS, N., GRAFF, C., THONBERG, H., CHIANG, H. H., PASTOR, P., ORTEGA-CUBERO, S., PASTOR, M. A., DIEHL-SCHMID, J., ALEXOPOULOS, P., BENUSSI, L., GHIDONI, R., BINETTI, G., NACMIAS, B., SORBI, S., SANCHEZ-VALLE, R., LLADO, A., GELPI, E., ALMEIDA, M. R., SANTANA, I., CLARIMON, J., LLEO, A., FORTEA, J., DE MENDONCA, A., MARTINS, M., BORRONI, B., PADOVANI, A., MATEJ, R., ROHAN, Z., RUIZ, A., FRISONI, G. B., FABRIZI, G. M., VANDENBERGHE, R., DE DEYN, P. P., VAN BROECKHOVEN, C. & SLEEGERS, K. 2015. Genetic variability in SQSTM1 and risk of early-onset Alzheimer dementia: a European early-onset dementia consortium study. *Neurobiol Aging,* 36**,** 2005 e15-22.

D'ADAMO, M. C., SFORNA, L., VISENTIN, S., GROTTESI, A., SERVETTINI, L., GUGLIELMI, L., MACCHIONI, L., SAREDI, S., CURCIO, M., DE NUCCIO, C., HASAN, S., CORAZZI, L., FRANCIOLINI, F., MORA, M., CATACUZZENO, L. & PESSIA, M. 2016. A Calsequestrin-1 Mutation Associated with a Skeletal Muscle Disease Alters Sarcoplasmic Ca2+ Release. *PLoS One,* 11**,** e0155516.

D'AGOSTINO, C., NOGALSKA, A., CACCIOTTOLO, M., ENGEL, W. K. & ASKANAS, V. 2011. Abnormalities of NBR1, a novel autophagy-associated protein, in muscle fibers of sporadic inclusion-body myositis. *Acta Neuropathol,* 122**,** 627-36.

D'AMICO, A., BENEDETTI, S., PETRINI, S., SAMBUUGHIN, N., BOLDRINI, R., MENDITTO, I., FERRARI, M., VERARDO, M., GOLDFARB, L. & BERTINI, E. 2005. Major myofibrillar changes in early onset myopathy due to de novo heterozygous missense mutation in lamin A/C gene. *Neuromuscul Disord,* 15**,** 847-50.

DALAKAS, M. C. 2006a. Inflammatory, immune, and viral aspects of inclusion-body myositis. *Neurology,* 66**,** S33-8.

DALAKAS, M. C. 2006b. Sporadic inclusion body myositis--diagnosis, pathogenesis and therapeutic strategies. *Nat Clin Pract Neurol,* 2**,** 437-47.

DALAKAS, M. C. 2008. Interplay between inflammation and degeneration: using inclusion body myositis to study "neuroinflammation". *Ann Neurol,* 64**,** 1-3.

DANON, M. J., CARPENTER, S. & HARATI, Y. 1989. Muscle pain associated with tubular aggregates and structures resembling cylindrical spirals. *Muscle Nerve,* 12**,** 265-272.

DAROSZEWSKA, A., VAN 'T HOF, R. J., ROJAS, J. A., LAYFIELD, R., LANDAO-BASONGA, E., ROSE, L., ROSE, K. & RALSTON, S. H. 2011.

A point mutation in the ubiquitin-associated domain of SQSMT1 is sufficient to cause a Paget's disease-like disorder in mice. *Hum Mol Genet,* 20**,** 2734-44.
DAVYDOV, E. V., GOODE, D. L., SIROTA, M., COOPER, G. M., SIDOW, A. & BATZOGLOU, S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol,* 6**,** e1001025.
DAWN TEARE, M. & BARRETT, J. H. 2005. Genetic linkage studies. *Lancet,* 366**,** 1036-44.
DE MAGALHAES, J. P., FINCH, C. E. & JANSSENS, G. 2010. Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Res Rev,* 9**,** 315-23.
DE PAULA, A. M., BARTOLI, M., COURRIER, S., POUGET, J., LEVY, N., PELLISSIER, J. F., FIGARELLA-BRANGER, D., KRAHN, M. & ATTARIAN, S. 2012. Further heterogeneity in myopathy with tubular aggregates? *Muscle Nerve,* 46**,** 984-5.
DEAN AG, S. K., SOE MM OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.3.1. www.OpenEpi.com. 06/04/2013 ed.
DEJESUS-HERNANDEZ, M., MACKENZIE, I. R., BOEVE, B. F., BOXER, A. L., BAKER, M., RUTHERFORD, N. J., NICHOLSON, A. M., FINCH, N. A., FLYNN, H., ADAMSON, J., KOURI, N., WOJTAS, A., SENGDY, P., HSIUNG, G. Y., KARYDAS, A., SEELEY, W. W., JOSEPHS, K. A., COPPOLA, G., GESCHWIND, D. H., WSZOLEK, Z. K., FELDMAN, H., KNOPMAN, D. S., PETERSEN, R. C., MILLER, B. L., DICKSON, D. W., BOYLAN, K. B., GRAFF-RADFORD, N. R. & RADEMAKERS, R. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron,* 72**,** 245-56.
DEL VILLAR NEGRO, A., MERINO ANGULO, J., RIVERA POMAR, J. M. & AGUIRRE ERRASTI, C. 1982. Tubular aggregates in skeletal muscle of chronic alcoholic patients. *Acta Neuropathol,* 56**,** 250-4.
DELAUNAY, A., BROMBERG, K. D., HAYASHI, Y., MIRABELLA, M., BURCH, D., KIRKWOOD, B., SERRA, C., MALICDAN, M. C., MIZISIN, A. P., MOROSETTI, R., BROCCOLINI, A., GUO, L. T., JONES, S. N., LIRA, S. A., PURI, P. L., SHELTON, G. D. & RONAI, Z. 2008. The ER-bound RING finger protein 5 (RNF5/RMA1) causes degenerative myopathy in transgenic mice and is deregulated in inclusion body myositis. *PLoS One,* 3**,** e1609.
DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNELL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet,* 43**,** 491-8.
DI BLASI, C., MORA, M., PAREYSON, D., FARINA, L., SGHIRLANZONI, A., VIGNIER, N., BLASEVICH, F., CORNELIO, F., GUICHENEY, P. & MORANDI, L. 2000. Partial laminin alpha2 chain deficiency in a patient with myopathy resembling inclusion body myositis. *Ann Neurol,* 47**,** 811-6.
DI BLASI, C., SANSANELLI, S., RUGGIERI, A., MORIGGI, M., VASSO, M., D'ADAMO, A. P., BLASEVICH, F., ZANOTTI, S., PAOLINI, C., PROTASI, F., TEZZON, F., GELFI, C., MORANDI, L., PESSIA, M. & MORA, M. 2015. A CASQ1 founder mutation in three Italian families with protein aggregate myopathy and hyperCKaemia. *J Med Genet,* 52**,** 617-26.

DIMACHKIE, M. M., BAROHN, R. J. & AMATO, A. A. 2014. Idiopathic inflammatory myopathies. *Neurol Clin,* 32**,** 595-628, vii.

DIMAURO, S., MIRANDA, A., KHAN, S., GITLIN, K. & FRIEDMAN, R. 1981. Human muscle phosphoglycerate mutase deficiency: newly discovered metabolic myopathy. *Science,* 212**,** 1277-1279.

DONNELLY, CHRISTOPHER J., ZHANG, P.-W., PHAM, JACQUELINE T., HAEUSLER, AARON R., MISTRY, NIPUN A., VIDENSKY, S., DALEY, ELIZABETH L., POTH, ERIN M., HOOVER, B., FINES, DANIEL M., MARAGAKIS, N., TIENARI, PENTTI J., PETRUCELLI, L., TRAYNOR, BRYAN J., WANG, J., RIGO, F., BENNETT, C. F., BLACKSHAW, S., SATTLER, R. & ROTHSTEIN, JEFFREY D. 2013. RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. *Neuron,* 80**,** 415-428.

DRMANAC, R., SPARKS, A. B., CALLOW, M. J., HALPERN, A. L., BURNS, N. L., KERMANI, B. G., CARNEVALI, P., NAZARENKO, I., NILSEN, G. B., YEUNG, G., DAHL, F., FERNANDEZ, A., STAKER, B., PANT, K. P., BACCASH, J., BORCHERDING, A. P., BROWNLEY, A., CEDENO, R., CHEN, L., CHERNIKOFF, D., CHEUNG, A., CHIRITA, R., CURSON, B., EBERT, J. C., HACKER, C. R., HARTLAGE, R., HAUSER, B., HUANG, S., JIANG, Y., KARPINCHYK, V., KOENIG, M., KONG, C., LANDERS, T., LE, C., LIU, J., MCBRIDE, C. E., MORENZONI, M., MOREY, R. E., MUTCH, K., PERAZICH, H., PERRY, K., PETERS, B. A., PETERSON, J., PETHIYAGODA, C. L., POTHURAJU, K., RICHTER, C., ROSENBAUM, A. M., ROY, S., SHAFTO, J., SHARANHOVICH, U., SHANNON, K. W., SHEPPY, C. G., SUN, M., THAKURIA, J. V., TRAN, A., VU, D., ZARANEK, A. W., WU, X., DRMANAC, S., OLIPHANT, A. R., BANYAI, W. C., MARTIN, B., BALLINGER, D. G., CHURCH, G. M. & REID, C. A. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science,* 327**,** 78-81.

DUPRÉ, N., CHRESTIAN, N., BOUCHARD, J.-P., ROSSIGNOL, E., BRUNET, D., STERNBERG, D., BRAIS, B., MATHIEU, J. & PUYMIRAT, J. 2009. Clinical, electrophysiologic, and genetic study of non-dystrophic myotonia in French-Canadians. *Neuromuscular Disorders,* 19**,** 330-334.

ENDO, Y., NOGUCHI, S., HARA, Y., HAYASHI, Y. K., MOTOMURA, K., MIYATAKE, S., MURAKAMI, N., TANAKA, S., YAMASHITA, S., KIZU, R., BAMBA, M., GOTO, Y., MATSUMOTO, N., NONAKA, I. & NISHINO, I. 2015. Dominant mutations in ORAI1 cause tubular aggregate myopathy with hypocalcemia via constitutive activation of store-operated Ca(2)(+) channels. *Hum Mol Genet,* 24**,** 637-48.

ENGEL, W. K. 1964. MITOCHONDRIAL AGGREGATES IN MUSCLE DISEASE. *J Histochem Cytochem,* 12**,** 46-8.

EXCOFFIER L, L. H. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*.

EYRE-WALKER, A. 2010. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci U S A,* 107 Suppl 1**,** 1752-6.

FARG, M. A., SUNDARAMOORTHY, V., SULTANA, J. M., YANG, S., ATKINSON, R. A. K., LEVINA, V., HALLORAN, M. A., GLEESON, P. A.,

BLAIR, I. P., SOO, K. Y., KING, A. E. & ATKIN, J. D. 2014. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet,* 23**,** 3579-3595.

FECTO, F., YAN, J., VEMULA, S. P., LIU, E., YANG, Y., CHEN, W., ZHENG, J. G., SHI, Y., SIDDIQUE, N., ARRAT, H., DONKERVOORT, S., AJROUD-DRISS, S., SUFIT, R. L., HELLER, S. L., DENG, H. X. & SIDDIQUE, T. 2011. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol,* 68**,** 1440-6.

FELICE, K. J. & NORTH, W. A. 2001. Inclusion body myositis in Connecticut: observations in 35 patients during an 8-year period. *Medicine (Baltimore),* 80**,** 320-7.

FERREIRO, A., CEUTERICK-DE GROOTE, C., MARKS, J. J., GOEMANS, N., SCHREIBER, G., HANEFELD, F., FARDEAU, M., MARTIN, J. J., GOEBEL, H. H., RICHARD, P., GUICHENEY, P. & BONNEMANN, C. G. 2004. Desmin-related myopathy with Mallory body-like inclusions is caused by mutations of the selenoprotein N gene. *Ann Neurol,* 55**,** 676-86.

FESKE, S., GWACK, Y., PRAKRIYA, M., SRIKANTH, S., PUPPEL, S. H., TANASA, B., HOGAN, P. G., LEWIS, R. S., DALY, M. & RAO, A. 2006. A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature,* 441**,** 179-85.

FESKE, S., MULLER, J. M., GRAF, D., KROCZEK, R. A., DRAGER, R., NIEMEYER, C., BAEUERLE, P. A., PETER, H. H. & SCHLESIER, M. 1996. Severe combined immunodeficiency due to defective binding of the nuclear factor of activated T cells in T lymphocytes of two male siblings. *Eur J Immunol,* 26**,** 2119-26.

FINLAYSON, S., PALACE, J., BELAYA, K., WALLS, T. J., NORWOOD, F., BURKE, G., HOLTON, J. L., PASCUAL-PASCUAL, S. I., COSSINS, J. & BEESON, D. 2013. Clinical features of congenital myasthenic syndrome due to mutations in DPAGT1. *J Neurol Neurosurg Psychiatry,* 84**,** 1119-25.

FIRESTEIN, R., NAGY, P. L., DALY, M., HUIE, P., CONTI, M. & CLEARY, M. L. 2002. Male infertility, impaired spermatogenesis, and azoospermia in mice deficient for the pseudophosphatase Sbf1. *J Clin Invest,* 109**,** 1165-72.

FOROUD, T., PANKRATZ, N., BATCHMAN, A., PAUCIULO, M., VIDAL, R., MIRAVALLE, L., GOEBEL, H., CUSHMAN, L., AZZARELLI, B. & HORAK, H. 2005. A mutation in myotilin causes spheroid body myopathy. *Neurology,* 65**,** 1936-1940.

FRATTA, P., ENGEL, W. K., VAN LEEUWEN, F. W., HOL, E. M., VATTEMI, G. & ASKANAS, V. 2004. Mutant ubiquitin UBB+1 is accumulated in sporadic inclusion-body myositis muscle fibers. *Neurology,* 63**,** 1114-7.

FREEZE, H. H. 2006. Genetic defects in the human glycome. *Nat Rev Genet,* 7**,** 537-551.

FUJII, J., WILLARD, H. F. & MACLENNAN, D. H. 1990. Characterization and localization to human chromosome 1 of human fast-twitch skeletal muscle calsequestrin gene. *Somat Cell Mol Genet,* 16**,** 185-9.

FUKUCHI, K., PHAM, D., HART, M., LI, L. & LINDSEY, J. R. 1998. Amyloid-beta deposition in skeletal muscle of transgenic mice: possible model of inclusion body myopathy. *Am J Pathol,* 153**,** 1687-93.

FUNK, F., CEUTERICK-DE GROOTE, C., MARTIN, J. J., MEINHARDT, A., TARATUTO, A. L., DE BLEECKER, J., VAN COSTER, R., DE PAEPE, B., SCHARA, U., VORGERD, M., HAUSLER, M., KOPPI, S., MASCHKE, M.,

DE JONGHE, P., VAN MALDERGEM, L., NOEL, S., ZIMMERMANN, C. W., WIRTH, S., ISENMANN, S., STADLER, R., SCHRODER, J. M., SCHULZ, J. B., WEIS, J. & CLAEYS, K. G. 2013. Morphological spectrum and clinical features of myopathies with tubular aggregates. *Histol Histopathol,* 28**,** 1041-54.

GAMI, P., MURRAY, C., SCHOTTLAENDER, L., BETTENCOURT, C., DE PABLO FERNANDEZ, E., MUDANOHWO, E., MIZIELINSKA, S., POLKE, J. M., HOLTON, J. L., ISAACS, A. M., HOULDEN, H., REVESZ, T. & LASHLEY, T. 2015. A 30-unit hexanucleotide repeat expansion in C9orf72 induces pathological lesions with dipeptide-repeat proteins and RNA foci, but not TDP-43 inclusions and clinical disease. *Acta Neuropathol,* 130**,** 599-601.

GANG, Q., BETTENCOURT, C., HOULDEN, H., HANNA, M. G. & MACHADO, P. M. 2015. Genetic advances in sporadic inclusion body myositis. *Curr Opin Rheumatol,* 27**,** 586-594.

GANG, Q., BETTENCOURT, C., MACHADO, P., HANNA, M. G. & HOULDEN, H. 2014. Sporadic inclusion body myositis: the genetic contributions to the pathogenesis. *Orphanet J Rare Dis,* 9**,** 88.

GARLEPP, M. J., LAING, B., ZILKO, P. J., OLLIER, W. & MASTAGLIA, F. L. 1994. HLA associations with inclusion body myositis. *Clin Exp Immunol,* 98**,** 40-5.

GARLEPP, M. J., TABARIAS, H., VAN BOCKXMEER, F. M., ZILKO, P. J., LAING, B. & MASTAGLIA, F. L. 1995. Apolipoprotein E epsilon 4 in inclusion body myositis. *Ann Neurol,* 38**,** 957-9.

GARRARD, P., BLAKE, J., STINTON, V., HANNA, M. G., REILLY, M. M., HOLTON, J. L., LANDON, D. N. & HONAN, W. P. 2002. Distal myopathy with tubular aggregates: a new phenotype associated with multiple deletions in mitochondrial DNA? *J Neurol Neurosurg Psychiatry,* 73**,** 207-8.

GENOMES PROJECT, C., AUTON, A., BROOKS, L. D., DURBIN, R. M., GARRISON, E. P., KANG, H. M., KORBEL, J. O., MARCHINI, J. L., MCCARTHY, S., MCVEAN, G. A. & ABECASIS, G. R. 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.

GERRISH, A., RUSSO, G., RICHARDS, A., MOSKVINA, V., IVANOV, D., HAROLD, D., SIMS, R., ABRAHAM, R., HOLLINGWORTH, P., CHAPMAN, J., HAMSHERE, M., PAHWA, J. S., DOWZELL, K., WILLIAMS, A., JONES, N., THOMAS, C., STRETTON, A., MORGAN, A. R., LOVESTONE, S., POWELL, J., PROITSI, P., LUPTON, M. K., BRAYNE, C., RUBINSZTEIN, D. C., GILL, M., LAWLOR, B., LYNCH, A., MORGAN, K., BROWN, K. S., PASSMORE, P. A., CRAIG, D., MCGUINNESS, B., TODD, S., JOHNSTON, J. A., HOLMES, C., MANN, D., SMITH, A. D., LOVE, S., KEHOE, P. G., HARDY, J., MEAD, S., FOX, N., ROSSOR, M., COLLINGE, J., MAIER, W., JESSEN, F., KOLSCH, H., HEUN, R., SCHURMANN, B., VAN DEN BUSSCHE, H., HEUSER, I., KORNHUBER, J., WILTFANG, J., DICHGANS, M., FROLICH, L., HAMPEL, H., HULL, M., RUJESCU, D., GOATE, A. M., KAUWE, J. S., CRUCHAGA, C., NOWOTNY, P., MORRIS, J. C., MAYO, K., LIVINGSTON, G., BASS, N. J., GURLING, H., MCQUILLIN, A., GWILLIAM, R., DELOUKAS, P., DAVIES, G., HARRIS, S. E., STARR, J. M., DEARY, I. J., AL-CHALABI, A., SHAW, C. E., TSOLAKI, M., SINGLETON, A. B., GUERREIRO, R., MUHLEISEN, T. W., NOTHEN, M.

M., MOEBUS, S., JOCKEL, K. H., KLOPP, N., WICHMANN, H. E., CARRASQUILLO, M. M., PANKRATZ, V. S., YOUNKIN, S. G., JONES, L., HOLMANS, P. A., O'DONOVAN, M. C., OWEN, M. J. & WILLIAMS, J. 2012. The role of variation at AbetaPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease. *J Alzheimers Dis,* 28**,** 377-87.

GERSUK, V. H. & NEPOM, G. T. 2009. A real-time polymerase chain reaction assay for the rapid identification of the autoimmune disease-associated allele HLA-DQB1*0602. *Tissue Antigens,* 73**,** 335-40.

GHOSH, A., NARAYANAPPA, G., TALY, A. B., CHICKBASAVAIYA, Y. T., MAHADEVAN, A., VANI, S., ATCHAYARAM, N., MOHAPATRA, I. & SUSARALA, K. S. 2010. Tubular aggregate myopathy: a phenotypic spectrum and morphological study. *Neurol India,* 58**,** 747-51.

GIBBELS, E., HENKE, U., SCHÄDLICH, H.-J., HAUPT, W. F. & FIEHN, W. 1983. Cylindrical spirals in skeletal muscle: A further observation with clinical, morphological, and biochemical analysis. *Muscle Nerve,* 6**,** 646-655.

GIBSON, G. 2011. Rare and common variants: twenty arguments. *Nat Rev Genet,* 13**,** 135-45.

GIDARO, T., MODONI, A., SABATELLI, M., TASCA, G., BROCCOLINI, A. & MIRABELLA, M. 2008. An Italian family with inclusion-body myopathy and frontotemporal dementia due to mutation in the VCP gene. *Muscle Nerve,* 37**,** 111-4.

GIJSELINCK, I., VAN LANGENHOVE, T., VAN DER ZEE, J., SLEEGERS, K., PHILTJENS, S., KLEINBERGER, G., JANSSENS, J., BETTENS, K., VAN CAUWENBERGHE, C., PERESON, S., ENGELBORGHS, S., SIEBEN, A., DE JONGHE, P., VANDENBERGHE, R., SANTENS, P., DE BLEECKER, J., MAES, G., BAUMER, V., DILLEN, L., JORIS, G., CUIJT, I., CORSMIT, E., ELINCK, E., VAN DONGEN, J., VERMEULEN, S., VAN DEN BROECK, M., VAERENBERG, C., MATTHEIJSSENS, M., PEETERS, K., ROBBERECHT, W., CRAS, P., MARTIN, J. J., DE DEYN, P. P., CRUTS, M. & VAN BROECKHOVEN, C. 2012. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol,* 11**,** 54-65.

GIROLAMO, F., LIA, A., AMATI, A., STRIPPOLI, M., COPPOLA, C., VIRGINTINO, D., RONCALI, L., TOSCANO, A., SERLENGA, L. & TROJANO, M. 2013. Overexpression of autophagic proteins in the skeletal muscle of sporadic inclusion body myositis. *Neuropathol Appl Neurobiol*.

GNIRKE, A., MELNIKOV, A., MAGUIRE, J., ROGOV, P., LEPROUST, E. M., BROCKMAN, W., FENNELL, T., GIANNOUKOS, G., FISHER, S., RUSS, C., GABRIEL, S., JAFFE, D. B., LANDER, E. S. & NUSBAUM, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol,* 27**,** 182-9.

GOEBEL, H. H., BONNEMANN, C. G. & RARE STRUCTURAL MYOPATHY, C. 2011. 169th ENMC International Workshop Rare Structural Congenital Myopathies 6-8 November 2009, Naarden, The Netherlands. *Neuromuscul Disord,* 21**,** 363-74.

GOEBEL, H. H. & MULLER, H. D. 2006. Protein aggregate myopathies. *Semin Pediatr Neurol,* 13**,** 96-103.

GOEBEL, H. H., SHARMA, M. C., TARATUTO, A. L. & CLAEYS, K. G. 2013. Disorders of Muscle with Rare Structural Abnormalities. *In:* GOEBEL, H. H.,

SEWRY, C. & WELLER, R. O. (eds.) *Muscle Disease: Pathology and Genetics.* SECOND ed. UK: Wiley-Blackwell.

GOLD, R. & REICHMANN, H. 1992. Muscle pathology correlates with permanent weakness in hypokalemic periodic paralysis: a case report. *Acta Neuropathol,* 84**,** 202-206.

GOMEZ-TORTOSA, E., GALLEGO, J., GUERRERO-LOPEZ, R., MARCOS, A., GIL-NECIGA, E., SAINZ, M. J., DIAZ, A., FRANCO-MACIAS, E., TRUJILLO-TIEBAS, M. J., AYUSO, C. & PEREZ-PEREZ, J. 2013. C9ORF72 hexanucleotide expansions of 20-22 repeats are associated with frontotemporal deterioration. *Neurology,* 80**,** 366-70.

GORYUNOV, D., NIGHTINGALE, A., BORNFLETH, L., LEUNG, C. & LIEM, R. K. H. 2008. Multiple disease-linked myotubularin mutations cause NFL assembly defects in cultured cells and disrupt myotubularin dimerization. *J Neurochem,* 104**,** 1536-1552.

GOSSRAU, G., GESTRICH, B., KOCH, R., WUNDERLICH, C., SCHRODER, J. M., SCHROEDER, S., REICHMANN, H. & LAMPE, J. B. 2004. Apolipoprotein E and alpha-1-antichymotrypsin polymorphisms in sporadic inclusion body myositis. *Eur Neurol,* 51**,** 215-20.

GREENBERG, S. A. 2009. How citation distortions create unfounded authority: analysis of a citation network. *BMJ,* 339**,** b2680.

GREENBERG, S. A. 2012. Pathogenesis and therapy of inclusion body myositis. *Curr Opin Neurol,* 25**,** 630-9.

GRIGGS, R. C., ASKANAS, V., DIMAURO, S., ENGEL, A., KARPATI, G., MENDELL, J. R. & ROWLAND, L. P. 1995. Inclusion body myositis and myopathies. *Ann Neurol,* 38**,** 705-13.

GRISOLIA, S. & CARRERAS, J. 1975. Phosphoglycerate mutase from yeast, chicken breast muscle, and kidney (2, 3-PGA-dependent). *Methods Enzymol,* 42**,** 435-50.

GU, J. M., KE, Y. H., YUE, H., LIU, Y. J., ZHANG, Z., ZHANG, H., HU, W. W., WANG, C., HE, J. W., HU, Y. Q., LI, M., FU, W. Z. & ZHANG, Z. L. 2013. A novel VCP mutation as the cause of atypical IBMPFD in a Chinese family. *Bone,* 52**,** 9-16.

GUERREIRO, R., WOJTAS, A., BRAS, J., CARRASQUILLO, M., ROGAEVA, E., MAJOUNIE, E., CRUCHAGA, C., SASSI, C., KAUWE, J. S., YOUNKIN, S., HAZRATI, L., COLLINGE, J., POCOCK, J., LASHLEY, T., WILLIAMS, J., LAMBERT, J. C., AMOUYEL, P., GOATE, A., RADEMAKERS, R., MORGAN, K., POWELL, J., ST GEORGE-HYSLOP, P., SINGLETON, A. & HARDY, J. 2013. TREM2 variants in Alzheimer's disease. *N Engl J Med,* 368**,** 117-27.

GURGEL-GIANNETTI, J., ZANOTELI, E., DE CASTRO CONCENTINO, E. L., ABATH NETO, O., PESQUERO, J. B., REED, U. C. & VAINZOF, M. 2012. Necklace fibers as histopathological marker in a patient with severe form of X-linked myotubular myopathy. *Neuromuscul Disord,* 22**,** 541-5.

GUYANT-MARECHAL, L., LAQUERRIERE, A., DUYCKAERTS, C., DUMANCHIN, C., BOU, J., DUGNY, F., LE BER, I., FREBOURG, T., HANNEQUIN, D. & CAMPION, D. 2006. Valosin-containing protein gene mutations: clinical and neuropathologic features. *Neurology,* 67**,** 644-51.

HADJIGEORGIOU, G. M., KAWASHIMA, N., BRUNO, C., ANDREU, A. L., SUE, C. M., RIGDEN, D. J., KAWASHIMA, A., SHANSKE, S. & DIMAURO, S. 1999. Manifesting heterozygotes in a Japanese family with a

novel mutation in the muscle-specific phosphoglycerate mutase (PGAM-M) gene. *Neuromuscul Disord,* 9**,** 399-402.

HANSSON PETERSEN, C. A., ALIKHANI, N., BEHBAHANI, H., WIEHAGER, B., PAVLOV, P. F., ALAFUZOFF, I., LEINONEN, V., ITO, A., WINBLAD, B., GLASER, E. & ANKARCRONA, M. 2008. The amyloid beta-peptide is imported into mitochondria via the TOM import machinery and localized to mitochondrial cristae. *Proc Natl Acad Sci U S A,* 105**,** 13145-50.

HANTAI, D., NICOLE, S. & EYMARD, B. 2013. Congenital myasthenic syndromes: an update. *Curr Opin Neurol,* 26**,** 561-8.

HARRINGTON, C. R., ANDERSON, J. R. & CHAN, K. K. 1995. Apolipoprotein E type epsilon 4 allele frequency is not increased in patients with sporadic inclusion-body myositis. *Neurosci Lett,* 183**,** 35-8.

HAUBENBERGER, D., BITTNER, R. E., RAUCH-SHORNY, S., ZIMPRICH, F., MANNHALTER, C., WAGNER, L., MINEVA, I., VASS, K., AUFF, E. & ZIMPRICH, A. 2005. Inclusion body myopathy and Paget disease is linked to a novel mutation in the VCP gene. *Neurology,* 65**,** 1304-5.

HAUSER, M. A., HORRIGAN, S. K., SALMIKANGAS, P., TORIAN, U. M., VILES, KRISTI D., DANCEL, R., TIM, R. W., TAIVAINEN, A., BARTOLONI, L., GILCHRIST, J. M., STAJICH, J. M., GASKELL, P. C., GILBERT, J. R., VANCE, J. M., PERICAK-VANCE, M. A., CARPEN, O., WESTBROOK, C. A. & SPEER, M. C. 2000. Myotilin is mutated in limb girdle muscular dystrophy 1A. *Hum Mol Genet,* 9**,** 2141-2147.

HEDBERG, C., LINDBERG, C., MATHE, G., MOSLEMI, A. R. & OLDFORS, A. 2012. Myopathy in a woman and her daughter associated with a novel splice site MTM1 mutation. *Neuromuscul Disord,* 22**,** 244-51.

HEDBERG, C., NICETA, M., FATTORI, F., LINDVALL, B., CIOLFI, A., D'AMICO, A., TASCA, G., PETRINI, S., TULINIUS, M., TARTAGLIA, M., OLDFORS, A. & BERTINI, E. 2014. Childhood onset tubular aggregate myopathy associated with de novo STIM1 mutations. *J Neurol*.

HERBERT, M. K., STAMMEN-VOGELZANGS, J., VERBEEK, M. M., RIETVELD, A., LUNDBERG, I. E., CHINOY, H., LAMB, J. A., COOPER, R. G., ROBERTS, M., BADRISING, U. A., DE BLEECKER, J. L., MACHADO, P. M., HANNA, M. G., PLESTILOVA, L., VENCOVSKY, J., VAN ENGELEN, B. G. & PRUIJN, G. J. 2015. Disease specificity of autoantibodies to cytosolic 5'-nucleotidase 1A in sporadic inclusion body myositis versus known autoimmune diseases. *Ann Rheum Dis*.

HERNANDEZ LAIN, A., MILLECAMPS, S., DUBOURG, O., SALACHAS, F., BRUNETEAU, G., LACOMBLEZ, L., LEGUERN, E., SEILHEAN, D., DUYCKAERTS, C., MEININGER, V., MALLET, J. & PRADAT, P. F. 2011. Abnormal TDP-43 and FUS proteins in muscles of sporadic IBM: similarities in a TARDBP-linked ALS patient. *J Neurol Neurosurg Psychiatry,* 82**,** 1414-6.

HILTON-JONES, D., MILLER, A., PARTON, M., HOLTON, J., SEWRY, C. & HANNA, M. G. 2010. Inclusion body myositis: MRC Centre for Neuromuscular Diseases, IBM workshop, London, 13 June 2008. *Neuromuscul Disord,* 20**,** 142-7.

HINIKER, A., DANIELS, B. H., LEE, H. S. & MARGETA, M. 2013. Comparative utility of LC3, p62 and TDP-43 immunohistochemistry in differentiation of inclusion body myositis from polymyositis and related inflammatory myopathies. *Acta Neuropathol Commun,* 1**,** 29.

HIRSCHHORN, J. N. 2009. Genomewide association studies--illuminating biologic pathways. *N Engl J Med,* 360**,** 1699-701.

HOLTON, J., WEDDERBURN, L. R. & HANNA, M. 2013. Polymyositis, Dermatomyositis, and Inclusion Body Myositis. *In:* GOEBEL, H. H., SEWRY, C. & WELLER, R. O. (eds.) *Muscle Disease: Pathology and Genetics.* SECOND ed. UK: Wiley-Blackwell.

HUH, S. Y., KIM, H. S., JANG, H. J., PARK, Y. E. & KIM, D. S. 2012. Limb-girdle myasthenia with tubular aggregates associated with novel GFPT1 mutations. *Muscle Nerve,* 46**,** 600-4.

IKEZOE, K., FURUYA, H., OHYAGI, Y., OSOEGAWA, M., NISHINO, I., NONAKA, I. & KIRA, J. 2003. Dysferlin expression in tubular aggregates: their possible relationship to endoplasmic reticulum stress. *Acta Neuropathol,* 105**,** 603-9.

INGELSSON, M., SHIN, Y., IRIZARRY, M. C., HYMAN, B. T., LILIUS, L., FORSELL, C. & GRAFF, C. 2003. Genotyping of apolipoprotein E: comparative evaluation of different protocols. *Curr Protoc Hum Genet,* Chapter 9**,** Unit9 14.

IVANIDZE, J., HOFFMANN, R., LOCHMULLER, H., ENGEL, A. G., HOHLFELD, R. & DORNMAIR, K. 2011. Inclusion body myositis: laser microdissection reveals differential up-regulation of IFN-gamma signaling cascade in attacked versus nonattacked myofibers. *Am J Pathol,* 179**,** 1347-59.

JACQUES, T. S., HOLTON, J., WATTS, P. M., WILLS, A. J., SMITH, S. E. & HANNA, M. G. 2002. Tubular aggregate myopathy with abnormal pupils and skeletal deformities. *J Neurol Neurosurg Psychiatry,* 73**,** 324-6.

JARVIE, T. 2005. Next generation sequencing technologies. *Drug Discov Today Technol,* 2**,** 255-60.

JAWORSKA-WILCZYNSKA, M., WILCZYNSKI, G. M., ENGEL, W. K., STRICKLAND, D. K., WEISGRABER, K. H. & ASKANAS, V. 2002. Three lipoprotein receptors and cholesterol in inclusion-body myositis muscle. *Neurology,* 58**,** 438-45.

JEAN, S., COX, S., SCHMIDT, E. J., ROBINSON, F. L. & KIGER, A. 2012. Sbf/MTMR13 coordinates PI(3)P and Rab21 regulation in endocytic control of cellular remodeling. *Mol Biol Cell,* 23**,** 2723-40.

JOHNSON, J. O., MANDRIOLI, J., BENATAR, M., ABRAMZON, Y., VAN DEERLIN, V. M., TROJANOWSKI, J. Q., GIBBS, J. R., BRUNETTI, M., GRONKA, S., WUU, J., DING, J., MCCLUSKEY, L., MARTINEZ-LAGE, M., FALCONE, D., HERNANDEZ, D. G., AREPALLI, S., CHONG, S., SCHYMICK, J. C., ROTHSTEIN, J., LANDI, F., WANG, Y. D., CALVO, A., MORA, G., SABATELLI, M., MONSURRO, M. R., BATTISTINI, S., SALVI, F., SPATARO, R., SOLA, P., BORGHERO, G., GALASSI, G., SCHOLZ, S. W., TAYLOR, J. P., RESTAGNO, G., CHIO, A. & TRAYNOR, B. J. 2010. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron,* 68**,** 857-64.

JU, J. S., FUENTEALBA, R. A., MILLER, S. E., JACKSON, E., PIWNICA-WORMS, D., BALOH, R. H. & WEIHL, C. C. 2009. Valosin-containing protein (VCP) is required for autophagy and is disrupted in VCP disease. *J Cell Biol,* 187**,** 875-88.

JUN, G., FLICKINGER, M., HETRICK, K. N., ROMM, J. M., DOHENY, K. F., ABECASIS, G. R., BOEHNKE, M. & KANG, H. M. 2012a. Detecting and

estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet,* 91**,** 839-48.

JUN, G., VARDARAJAN, B. N., BUROS, J., YU, C. E., HAWK, M. V., DOMBROSKI, B. A., CRANE, P. K., LARSON, E. B., MAYEUX, R., HAINES, J. L., LUNETTA, K. L., PERICAK-VANCE, M. A., SCHELLENBERG, G. D. & FARRER, L. A. 2012b. Comprehensive search for Alzheimer disease susceptibility loci in the APOE region. *Arch Neurol,* 69**,** 1270-9.

JUNGBLUTH, H. & GAUTEL, M. 2014. Pathogenic mechanisms in centronuclear myopathies. *Front Aging Neurosci,* 6**,** 339.

KABASHI, E., VALDMANIS, P. N., DION, P., SPIEGELMAN, D., MCCONKEY, B. J., VANDE VELDE, C., BOUCHARD, J. P., LACOMBLEZ, L., POCHIGAEVA, K., SALACHAS, F., PRADAT, P. F., CAMU, W., MEININGER, V., DUPRE, N. & ROULEAU, G. A. 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet,* 40**,** 572-4.

KAIPIAINEN-SEPPANEN, O. & AHO, K. 1996. Incidence of rare systemic rheumatic and connective tissue diseases in Finland. *J Intern Med,* 240**,** 81-4.

KALMAR, B. & GREENSMITH, L. 2009. Activation of the heat shock response in a primary cellular model of motoneuron neurodegeneration-evidence for neuroprotective and neurotoxic effects. *Cell Mol Biol Lett,* 14**,** 319-35.

KAMBOH, M. I., SANGHERA, D. K., FERRELL, R. E. & DEKOSKY, S. T. 1995. APOE*4-associated Alzheimer's disease risk is modified by alpha 1-antichymotrypsin polymorphism. *Nat Genet,* 10**,** 486-8.

KANEKURA, K., YAGI, T., CAMMACK, A. J., MAHADEVAN, J., KURODA, M., HARMS, M. B., MILLER, T. M. & URANO, F. 2016. Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum Mol Genet*.

KARPATI, G. & O'FERRALL, E. K. 2009. Sporadic inclusion body myositis: pathogenic considerations. *Ann Neurol,* 65**,** 7-11.

KATSUMATA, Y. & ASCHERMAN, D. P. 2008. Animal models in myositis. *Curr Opin Rheumatol,* 20**,** 681-5.

KEDUKA, E., HAYASHI, Y. K., SHALABY, S., MITSUHASHI, H., NOGUCHI, S., NONAKA, I. & NISHINO, I. 2012. In Vivo Characterization of Mutant Myotilins. *Am J Pathol,* 180**,** 1570-1580.

KILCH, T., ALANSARY, D., PEGLOW, M., DÖRR, K., RYCHKOV, G., RIEGER, H., PEINELT, C. & NIEMEYER, B. A. 2013. Mutations of the Ca2+-sensing Stromal Interaction Molecule STIM1 Regulate Ca2+ Influx by Altered Oligomerization of STIM1 and by Destabilization of the Ca2+ Channel Orai1. *Journal of Biological Chemistry,* 288**,** 1653-1664.

KIM, E. J., PARK, Y. E., KIM, D. S., AHN, B. Y., KIM, H. S., CHANG, Y. H., KIM, S. J., KIM, H. J., LEE, H. W., SEELEY, W. W. & KIM, S. 2011. Inclusion body myopathy with Paget disease of bone and frontotemporal dementia linked to VCP p.Arg155Cys in a Korean family. *Arch Neurol,* 68**,** 787-96.

KIM, H. J., KIM, N. C., WANG, Y. D., SCARBOROUGH, E. A., MOORE, J., DIAZ, Z., MACLEA, K. S., FREIBAUM, B., LI, S., MOLLIEX, A., KANAGARAJ, A. P., CARTER, R., BOYLAN, K. B., WOJTAS, A. M., RADEMAKERS, R., PINKUS, J. L., GREENBERG, S. A., TROJANOWSKI, J. Q., TRAYNOR, B. J., SMITH, B. N., TOPP, S., GKAZI,

A. S., MILLER, J., SHAW, C. E., KOTTLORS, M., KIRSCHNER, J., PESTRONK, A., LI, Y. R., FORD, A. F., GITLER, A. D., BENATAR, M., KING, O. D., KIMONIS, V. E., ROSS, E. D., WEIHL, C. C., SHORTER, J. & TAYLOR, J. P. 2013. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature,* 495**,** 467-73.

KIM, S. A., VACRATSIS, P. O., FIRESTEIN, R., CLEARY, M. L. & DIXON, J. E. 2003. Regulation of myotubularin-related (MTMR)2 phosphatidylinositol phosphatase by MTMR5, a catalytically inactive phosphatase. *Proc Natl Acad Sci U S A,* 100**,** 4492-7.

KIMONIS, V. E., MEHTA, S. G., FULCHIERO, E. C., THOMASOVA, D., PASQUALI, M., BOYCOTT, K., NEILAN, E. G., KARTASHOV, A., FORMAN, M. S., TUCKER, S., KIMONIS, K., MUMM, S., WHYTE, M. P., SMITH, C. D. & WATTS, G. D. 2008. Clinical studies in familial VCP myopathy associated with Paget disease of bone and frontotemporal dementia. *Am J Med Genet A,* 146A**,** 745-57.

KING, C. R., RATHOUZ, P. J. & NICOLAE, D. L. 2010. An evolutionary framework for association testing in resequencing studies. *PLoS Genet,* 6**,** e1001202.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet,* 46**,** 310-5.

KITAZAWA, M., TRINH, D. N. & LAFERLA, F. M. 2008. Inflammation induces tau pathology in inclusion body myositis model via glycogen synthase kinase-3beta. *Ann Neurol,* 64**,** 15-24.

KITAZAWA, M., VASILEVKO, V., CRIBBS, D. H. & LAFERLA, F. M. 2009. Immunization with amyloid-beta attenuates inclusion body myositis-like myopathology and motor impairment in a transgenic mouse model. *J Neurosci,* 29**,** 6132-41.

KLEIN, R. J., ZEISS, C., CHEW, E. Y., TSAI, J.-Y., SACKLER, R. S., HAYNES, C., HENNING, A. K., SANGIOVANNI, J. P., MANE, S. M., MAYNE, S. T., BRACKEN, M. B., FERRIS, F. L., OTT, J., BARNSTABLE, C. & HOH, J. 2005. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science,* 308**,** 385-389.

KOFFMAN, B. M., SIVAKUMAR, K., SIMONIS, T., STRONCEK, D. & DALAKAS, M. C. 1998. HLA allele distribution distinguishes sporadic inclusion body myositis from hereditary inclusion body myopathies. *J Neuroimmunol,* 84**,** 139-42.

KOK, C. C., BOYT, A., GAUDIERI, S., MARTINS, R., ASKANAS, V., DALAKAS, M., KIERS, L., MASTAGLIA, F. & GARLEPP, M. 2000. Mitochondrial DNA variants in inclusion body myositis. *Neuromuscul Disord,* 10**,** 604-11.

KOK, C. C., CROAGER, E. J., WITT, C. S., KIERS, L., MASTAGLIA, F. L., ABRAHAM, L. J. & GARLEPP, M. J. 1999. Mapping of a candidate region for susceptibility to inclusion body myositis in the human major histocompatibility complex. *Immunogenetics,* 49**,** 508-16.

KOMATSU, M., KAGEYAMA, S. & ICHIMURA, Y. 2012. p62/SQSTM1/A170: physiology and pathology. *Pharmacol Res,* 66**,** 457-62.

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**,** 1073-81.

KUZMIN, A., JARVI, K., LO, K., SPENCER, L., CHOW, G. Y., MACLEOD, G., WANG, Q. & VARMUZA, S. 2009. Identification of potentially damaging amino acid substitutions leading to human male infertility. *Biol Reprod,* 81**,** 319-26.

KWON, I., XIANG, S., KATO, M., WU, L., THEODOROPOULOS, P., WANG, T., KIM, J., YUN, J., XIE, Y. & MCKNIGHT, S. L. 2014. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science,* 345**,** 1139-1145.

LACRUZ, R. S. & FESKE, S. 2015. Diseases caused by mutations in ORAI1 and STIM1. *Ann N Y Acad Sci,* 1356**,** 45-79.

LAGIER-TOURENNE, C. & CLEVELAND, D. W. 2009. Rethinking ALS: the FUS about TDP-43. *Cell,* 136**,** 1001-4.

LAMPE, J., KITZLER, H., WALTER, M. C., LOCHMULLER, H. & REICHMANN, H. 1999. Methionine homozygosity at prion gene codon 129 may predispose to sporadic inclusion-body myositis. *Lancet,* 353**,** 465-6.

LAMPE, J. B., GOSSRAU, G., KEMPE, A., FUSSEL, M., SCHWURACK, K., SCHRODER, R., KRAUSE, S., KOHNEN, R., WALTER, M. C., REICHMANN, H. & LOCHMULLER, H. 2003. Analysis of HLA class I and II alleles in sporadic inclusion-body myositis. *J Neurol,* 250**,** 1313-7.

LANDRUM, M. J., LEE, J. M., BENSON, M., BROWN, G., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., HOOVER, J., JANG, W., KATZ, K., OVETSKY, M., RILEY, G., SETHI, A., TULLY, R., VILLAMARIN-SALOMON, R., RUBINSTEIN, W. & MAGLOTT, D. R. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res,* 44**,** D862-8.

LANGE, L. A., HU, Y., ZHANG, H., XUE, C., SCHMIDT, E. M., TANG, Z. Z., BIZON, C., LANGE, E. M., SMITH, J. D., TURNER, E. H., JUN, G., KANG, H. M., PELOSO, G., AUER, P., LI, K. P., FLANNICK, J., ZHANG, J., FUCHSBERGER, C., GAULTON, K., LINDGREN, C., LOCKE, A., MANNING, A., SIM, X., RIVAS, M. A., HOLMEN, O. L., GOTTESMAN, O., LU, Y., RUDERFER, D., STAHL, E. A., DUAN, Q., LI, Y., DURDA, P., JIAO, S., ISAACS, A., HOFMAN, A., BIS, J. C., CORREA, A., GRISWOLD, M. E., JAKOBSDOTTIR, J., SMITH, A. V., SCHREINER, P. J., FEITOSA, M. F., ZHANG, Q., HUFFMAN, J. E., CROSBY, J., WASSEL, C. L., DO, R., FRANCESCHINI, N., MARTIN, L. W., ROBINSON, J. G., ASSIMES, T. L., CROSSLIN, D. R., ROSENTHAL, E. A., TSAI, M., RIEDER, M. J., FARLOW, D. N., FOLSOM, A. R., LUMLEY, T., FOX, E. R., CARLSON, C. S., PETERS, U., JACKSON, R. D., VAN DUIJN, C. M., UITTERLINDEN, A. G., LEVY, D., ROTTER, J. I., TAYLOR, H. A., GUDNASON, V., JR., SISCOVICK, D. S., FORNAGE, M., BORECKI, I. B., HAYWARD, C., RUDAN, I., CHEN, Y. E., BOTTINGER, E. P., LOOS, R. J., SAETROM, P., HVEEM, K., BOEHNKE, M., GROOP, L., MCCARTHY, M., MEITINGER, T., BALLANTYNE, C. M., GABRIEL, S. B., O'DONNELL, C. J., POST, W. S., NORTH, K. E., REINER, A. P., BOERWINKLE, E., PSATY, B. M., ALTSHULER, D., KATHIRESAN, S., LIN, D. Y., JARVIK, G. P., CUPPLES, L. A., KOOPERBERG, C., WILSON, J. G., NICKERSON, D. A., ABECASIS, G. R., RICH, S. S., et al.

2014. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet,* 94**,** 233-45.

LAPORTE, J., BEDEZ, F., BOLINO, A. & MANDEL, J.-L. 2003. Myotubularins, a large disease-associated family of cooperating catalytically active and inactive phosphoinositides phosphatases. *Hum Mol Genet,* 12**,** R285-R292.

LARKIN, A. & IMPERIALI, B. 2011. The Expanding Horizons of Asparagine-Linked Glycosylation. *Biochemistry,* 50**,** 4411-4426.

LARMAN, H. B., SALAJEGHEH, M., NAZARENO, R., LAM, T., SAULD, J., STEEN, H., KONG, S. W., PINKUS, J. L., AMATO, A. A., ELLEDGE, S. J. & GREENBERG, S. A. 2013. Cytosolic 5'-nucleotidase 1A autoimmunity in sporadic inclusion body myositis. *Ann Neurol,* 73**,** 408-18.

LATTANTE, S., ROULEAU, G. A. & KABASHI, E. 2013. TARDBP and FUS mutations associated with amyotrophic lateral sclerosis: summary and update. *Hum Mutat,* 34**,** 812-26.

LAURIN, N., BROWN, J. P., MORISSETTE, J. & RAYMOND, V. 2002. Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget disease of bone. *Am J Hum Genet,* 70**,** 1582-8.

LAW, R. H. P., ZHANG, Q., MCGOWAN, S., BUCKLE, A. M., SILVERMAN, G. A., WONG, W., ROSADO, C. J., LANGENDORF, C. G., PIKE, R. N., BIRD, P. I. & WHISSTOCK, J. C. 2006. An overview of the serpin superfamily. *Genome Biol,* 7**,** 216.

LEE, S., ABECASIS, G. R., BOEHNKE, M. & LIN, X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet,* 95**,** 5-23.

LEE, S., EMOND, M. J., BAMSHAD, M. J., BARNES, K. C., RIEDER, M. J., NICKERSON, D. A., CHRISTIANI, D. C., WURFEL, M. M. & LIN, X. 2012a. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet,* 91**,** 224-37.

LEE, S., WU, M. C. & LIN, X. 2012b. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics,* 13**,** 762-775.

LESAGE, S., LE BER, I., CONDROYER, C., BROUSSOLLE, E., GABELLE, A., THOBOIS, S., PASQUIER, F., MONDON, K., DION, P. A., ROCHEFORT, D., ROULEAU, G. A., DÜRR, A. & BRICE, A. 2013. C9orf72 repeat expansions are a rare genetic cause of parkinsonism. *Brain,* 136**,** 385-391.

LEVEY, A. I., HEILMAN, C. J., LAH, J. J., NASH, N. R., REES, H. D., WAKAI, M., MIRRA, S. S., RYE, D. B., NOCHLIN, D., BIRD, T. D. & MUFSON, E. J. 1997. Presenilin-1 protein expression in familial and sporadic Alzheimer's disease. *Ann Neurol,* 41**,** 742-53.

LEVY, E., CARMAN, M. D., FERNANDEZ-MADRID, I. J., POWER, M. D., LIEBERBURG, I., VAN DUINEN, S. G., BOTS, G. T., LUYENDIJK, W. & FRANGIONE, B. 1990. Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. *Science,* 248**,** 1124-6.

LI, B. & LEAL, S. M. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet,* 83**,** 311-21.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LINDBERG, C., PERSSON, L. I., BJORKANDER, J. & OLDFORS, A. 1994. Inclusion body myositis: clinical, morphological, physiological and laboratory findings in 18 cases. *Acta Neurol Scand,* 89**,** 123-31.

LINDGREN, U., ROOS, S., HEDBERG OLDFORS, C., MOSLEMI, A. R., LINDBERG, C. & OLDFORS, A. 2015. Mitochondrial pathology in inclusion body myositis. *Neuromuscul Disord,* 25**,** 281-8.

LINDQUIST, S. G., DUNO, M., BATBAYLI, M., PUSCHMANN, A., BRAENDGAARD, H., MARDOSIENE, S., SVENSTRUP, K., PINBORG, L. H., VESTERGAARD, K., HJERMIND, L. E., STOKHOLM, J., ANDERSEN, B. B., JOHANNSEN, P. & NIELSEN, J. E. 2013. Corticobasal and ataxia syndromes widen the spectrum of C9ORF72 hexanucleotide expansion disease. *Clin Genet,* 83**,** 279-283.

LINNERTZ, C., ANDERSON, L., GOTTSCHALK, W., CRENSHAW, D., LUTZ, M. W., ALLEN, J., SAITH, S., MIHOVILOVIC, M., BURKE, J. R., WELSH-BOHMER, K. A., ROSES, A. D. & CHIBA-FALEK, O. 2014. The cis-regulatory effect of an Alzheimer's disease-associated poly-T locus on expression of TOMM40 and apolipoprotein E genes. *Alzheimers Dement*.

LIU, Q. Y., KOUKIEKOLO, R., ZHANG, D. L., SMITH, B., LY, D., LEI, J. X. & GHRIBI, O. 2016. Molecular events linking cholesterol to Alzheimer's disease and inclusion body myositis in a rabbit model. *Am J Neurodegener Dis,* 5**,** 74-84.

LLOYD, T. E., MAMMEN, A. L., AMATO, A. A., WEISS, M. D., NEEDHAM, M. & GREENBERG, S. A. 2014. Evaluation and construction of diagnostic criteria for inclusion body myositis. *Neurology,* 83**,** 426-33.

LOVE, S., NICOLL, J. A., LOWE, J. & SHERRIFF, F. 1996. Apolipoprotein E allele frequencies in sporadic inclusion body myositis. *Muscle Nerve,* 19**,** 1605-7.

LUAN, X., CHEN, B., LIU, Y., ZHENG, R., ZHANG, W. & YUAN, Y. 2009. Tubular aggregates in paralysis periodica paramyotonica with T704M mutation of SCN4A. *Neuropathology,* 29**,** 579-84.

LUNEMANN, J. D., SCHMIDT, J., DALAKAS, M. C. & MUNZ, C. 2007. Macroautophagy as a pathomechanism in sporadic inclusion body myositis. *Autophagy,* 3**,** 384-6.

LUNN, M., HANNA, M., HOWARD, R., PARTON, M. & REILLY, M. M. 2009. Nerve and Muscle Disease. *In:* CLARKE, C., HOWARD, R., ROSSOR, M. & SHORVON, S. (eds.) *Neurology: a Queen Square Textbook.* First ed. UK: Wiley-Blackwell.

MACHADO, P., BRADY, S. & HANNA, M. G. 2013. Update in inclusion body myositis. *Curr Opin Rheumatol,* 25**,** 763-71.

MADSEN, B. E. & BROWNING, S. R. 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet,* 5**,** e1000384.

MAJOUNIE, E., ABRAMZON, Y., RENTON, A. E., PERRY, R., BASSETT, S. S., PLETNIKOVA, O., TRONCOSO, J. C., HARDY, J., SINGLETON, A. B. & TRAYNOR, B. J. 2012a. Repeat expansion in C9ORF72 in Alzheimer's disease. *N Engl J Med,* 366**,** 283-4.

MAJOUNIE, E., TRAYNOR, B. J., CHIO, A., RESTAGNO, G., MANDRIOLI, J., BENATAR, M., TAYLOR, J. P. & SINGLETON, A. B. 2012b. Mutational analysis of the VCP gene in Parkinson's disease. *Neurobiol Aging,* 33**,** 209 e1-2.

MALFATTI, E., CHAVES, M., BELLANCE, R., VIOU, M. T., SARRAZIN, E., FARDEAU, M. & ROMERO, N. B. 2015. Cylindrical spirals associated with severe congenital muscle weakness and epileptic encephalopathy. *Muscle Nerve,* 52**,** 895-899.

MALICDAN, M. C. & NISHINO, I. 2013. Proteins of Autophagy: LAMP-2, VMA21, VCP, and TRIM32. *In:* GOEBEL, H. H., SEWRY, C. & WELLER, R. O. (eds.) *Muscle Disease: Pathology and Genetics.* SECOND ed. UK: Wiley-Blackwell.

MANOLIO, T. A. 2010. Genomewide association studies and assessment of the risk of disease. *N Engl J Med,* 363**,** 166-76.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R., CHAKRAVARTI, A., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G., HAINES, J. L., MACKAY, T. F., MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MARAGANORE, D. M., FARRER, M. J., HARDY, J. A., MCDONNELL, S. K., SCHAID, D. J. & ROCCA, W. A. 2000. Case-control study of debrisoquine 4-hydroxylase, N-acetyltransferase 2, and apolipoprotein E gene polymorphisms in Parkinson's disease. *Mov Disord,* 15**,** 714-9.

MARTIN, J. E., MATHER, K., SWASH, M. & GRAY, A. B. 1991. Expression of heat shock protein epitopes in tubular aggregates. *Muscle Nerve,* 14**,** 219-25.

MARUSZAK, A., PEPLONSKA, B., SAFRANOW, K., CHODAKOWSKA-ZEBROWSKA, M., BARCIKOWSKA, M. & ZEKANOWSKI, C. 2012. TOMM40 rs10524523 polymorphism's role in late-onset Alzheimer's disease and in longevity. *J Alzheimers Dis,* 28**,** 309-22.

MASTAGLIA, F. L., NEEDHAM, M., SCOTT, A., JAMES, I., ZILKO, P., DAY, T., KIERS, L., CORBETT, A., WITT, C. S., ALLCOCK, R., LAING, N., GARLEPP, M. & CHRISTIANSEN, F. T. 2009. Sporadic inclusion body myositis: HLA-DRB1 allele interactions influence disease risk and clinical phenotype. *Neuromuscul Disord,* 19**,** 763-5.

MASTAGLIA, F. L., ROJANA-UDOMSART, A., JAMES, I., NEEDHAM, M., DAY, T. J., KIERS, L., CORBETT, J. A., SAUNDERS, A. M., LUTZ, M. W. & ROSES, A. D. 2013. Polymorphism in the TOMM40 gene modifies the risk of developing sporadic inclusion body myositis and the age of onset of symptoms. *Neuromuscul Disord.*

MASUDA, Y., UEMURA, S., OHASHI, R., NAKANISHI, A., TAKEGOSHI, K., SHIMIZU, T., SHIRASAWA, T. & IRIE, K. 2009. Identification of physiological and toxic conformations in Abeta42 aggregates. *Chembiochem,* 10**,** 287-95.

MAURAGE, C. A., BUSSIERE, T., SERGEANT, N., GHESTEEM, A., FIGARELLA-BRANGER, D., RUCHOUX, M. M., PELLISSIER, J. F. & DELACOURTE, A. 2004. Tau aggregates are abnormally phosphorylated in inclusion body myositis and have an immunoelectrophoretic profile distinct from other tauopathies. *Neuropathol Appl Neurobiol,* 30**,** 624-34.

MCDOUGALL, J., WILES, C. M. & EDWARDS, R. H. T. 1980. SPIRAL MEMBRANE CYLINDERS IN THE SKELETAL MUSCLE OF A

PATIENT WITH MELORHEOSTOSIS. *Neuropathol Appl Neurobiol,* 6**,** 69-74.

MCLAREN, W., PRITCHARD, B., RIOS, D., CHEN, Y., FLICEK, P. & CUNNINGHAM, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics,* 26**,** 2069-2070.

MEYER, H. & WEIHL, C. C. 2014. The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. *J Cell Sci,* 127**,** 3877-83.

MEZEI, M. M., MANKODI, A., BRAIS, B., MARINEAU, C., THORNTON, C. A., ROULEAU, G. A. & KARPATI, G. 1999. Minimal expansion of the GCG repeat in the PABP2 gene does not predispose to sporadic inclusion body myositis. *Neurology,* 52**,** 669-70.

MILLECAMPS, S., BOILLÉE, S., LE BER, I., SEILHEAN, D., TEYSSOU, E., GIRAUDEAU, M., MOIGNEU, C., VANDENBERGHE, N., DANEL-BRUNAUD, V., CORCIA, P., PRADAT, P.-F., LE FORESTIER, N., LACOMBLEZ, L., BRUNETEAU, G., CAMU, W., BRICE, A., CAZENEUVE, C., LEGUERN, E., MEININGER, V. & SALACHAS, F. 2012. Phenotype difference between ALS patients with expanded repeats in C9ORF72 and patients with mutations in other ALS-related genes. *J Med Genet,* 49**,** 258-263.

MIRABELLA, M., ALVAREZ, R. B., ENGEL, W. K., WEISGRABER, K. H. & ASKANAS, V. 1996. Apolipoprotein E and apolipoprotein E messenger RNA in muscle of inclusion body myositis and myopathies. *Ann Neurol,* 40**,** 864-72.

MISCEO, D., HOLMGREN, A., LOUCH, W. E., HOLME, P. A., MIZOBUCHI, M., MORALES, R. J., DE PAULA, A. M., STRAY-PEDERSEN, A., LYLE, R., DALHUS, B., CHRISTENSEN, G., STORMORKEN, H., TJONNFJORD, G. E. & FRENGEN, E. 2014. A Dominant STIM1 Mutation Causes Stormorken Syndrome. *Hum Mutat*.

MIZIELINSKA, S., GRONKE, S., NICCOLI, T., RIDLER, C. E., CLAYTON, E. L., DEVOY, A., MOENS, T., NORONA, F. E., WOOLLACOTT, I. O., PIETRZYK, J., CLEVERLEY, K., NICOLL, A. J., PICKERING-BROWN, S., DOLS, J., CABECINHA, M., HENDRICH, O., FRATTA, P., FISHER, E. M., PARTRIDGE, L. & ISAACS, A. M. 2014. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science,* 345**,** 1192-4.

MOLBERG, O. & DOBLOUG, C. 2016. Epidemiology of sporadic inclusion body myositis. *Curr Opin Rheumatol*.

MORGAN-HUGHES, J. A. 1998. Tubular aggregates in skeletal muscle: their functional significance and mechanisms of pathogenesis. *Curr Opin Neurol,* 11**,** 439-42.

MORGENTHALER, S. & THILLY, W. G. 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res,* 615**,** 28-56.

MORI, K., WENG, S.-M., ARZBERGER, T., MAY, S., RENTZSCH, K., KREMMER, E., SCHMID, B., KRETZSCHMAR, H. A., CRUTS, M., VAN BROECKHOVEN, C., HAASS, C. & EDBAUER, D. 2013. The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science,* 339**,** 1335-1338.

MORIN, G., BRUECHLE, N. O., SINGH, A. R., KNOPP, C., JEDRASZAK, G., ELBRACHT, M., BREMOND-GIGNAC, D., HARTMANN, K., SEVESTRE, H., DEUTZ, P., HERENT, D., NURNBERG, P., ROMEO, B., KONRAD, K., MATHIEU-DRAMARD, M., OLDENBURG, J., BOURGES-PETIT, E., SHEN, Y., ZERRES, K., OUADID-AHIDOUCH, H. & ROCHETTE, J. 2014. Gain-of-Function Mutation in STIM1 (P.R304W) Is Associated with Stormorken Syndrome. *Hum Mutat,* 35**,** 1221-32.

MORRIS, A. P. & ZEGGINI, E. 2010. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol,* 34**,** 188-93.

MORTON, N. E. 1955. Sequential tests for the detection of linkage. *Am J Hum Genet,* 7**,** 277-318.

MOSLEMI, A. R., LINDBERG, C. & OLDFORS, A. 1997. Analysis of multiple mitochondrial DNA deletions in inclusion body myositis. *Hum Mutat,* 10**,** 381-6.

MOUSSA, C. E., FU, Q., KUMAR, P., SHTIFMAN, A., LOPEZ, J. R., ALLEN, P. D., LAFERLA, F., WEINBERG, D., MAGRANE, J., APRAHAMIAN, T., WALSH, K., ROSEN, K. M. & QUERFURTH, H. W. 2006. Transgenic expression of beta-APP in fast-twitch skeletal muscle leads to calcium dyshomeostasis and IBM-like pathology. *FASEB J,* 20**,** 2165-7.

MUNTZING, K., LINDBERG, C., MOSLEMI, A. R. & OLDFORS, A. 2003. Inclusion body myositis: clonal expansions of muscle-infiltrating T cells persist over time. *Scand J Immunol,* 58**,** 195-200.

NAINI, A., TOSCANO, A., MUSUMECI, O., VISSING, J., AKMAN, H. O. & DIMAURO, S. 2009. Muscle phosphoglycerate mutase deficiency revisited. *Arch Neurol,* 66**,** 394-8.

NAKHRO, K., PARK, J. M., HONG, Y. B., PARK, J. H., NAM, S. H., YOON, B. R., YOO, J. H., KOO, H., JUNG, S. C., KIM, H. L., KIM, J. Y., CHOI, K. G., CHOI, B. O. & CHUNG, K. W. 2013. SET binding factor 1 (SBF1) mutation causes Charcot-Marie-Tooth disease type 4B3. *Neurology,* 81**,** 165-73.

NEALE, B. M., RIVAS, M. A., VOIGHT, B. F., ALTSHULER, D., DEVLIN, B., ORHO-MELANDER, M., KATHIRESAN, S., PURCELL, S. M., ROEDER, K. & DALY, M. J. 2011. Testing for an unusual distribution of rare variants. *PLoS Genet,* 7**,** e1001322.

NEEDHAM, M., CORBETT, A., DAY, T., CHRISTIANSEN, F., FABIAN, V. & MASTAGLIA, F. L. 2008a. Prevalence of sporadic inclusion body myositis and factors contributing to delayed diagnosis. *J Clin Neurosci,* 15**,** 1350-3.

NEEDHAM, M., HOOPER, A., JAMES, I., VAN BOCKXMEER, F., CORBETT, A., DAY, T., GARLEPP, M. J. & MASTAGLIA, F. L. 2008b. Apolipoprotein epsilon alleles in sporadic inclusion body myositis: a reappraisal. *Neuromuscul Disord,* 18**,** 150-2.

NEEDHAM, M., JAMES, I., CORBETT, A., DAY, T., CHRISTIANSEN, F., PHILLIPS, B. & MASTAGLIA, F. L. 2008c. Sporadic inclusion body myositis: phenotypic variability and influence of HLA-DR3 in a cohort of 57 Australian cases. *J Neurol Neurosurg Psychiatry,* 79**,** 1056-60.

NEEDHAM, M. & MASTAGLIA, F. L. 2016. Sporadic inclusion body myositis: A review of recent clinical advances and current approaches to diagnosis and treatment. *Clin Neurophysiol,* 127**,** 1764-73.

NEEDHAM, M., MASTAGLIA, F. L. & GARLEPP, M. J. 2007. Genetics of inclusion-body myositis. *Muscle Nerve,* 35**,** 549-61.

NESIN, V., WILEY, G., KOUSI, M., ONG, E. C., LEHMANN, T., NICHOLL, D. J., SURI, M., SHAHRIZAILA, N., KATSANIS, N., GAFFNEY, P. M., WIERENGA, K. J. & TSIOKAS, L. 2014. Activating mutations in STIM1 and ORAI1 cause overlapping syndromes of tubular myopathy and congenital miosis. *Proc Natl Acad Sci U S A*.

NEUMANN, M., SAMPATHU, D. M., KWONG, L. K., TRUAX, A. C., MICSENYI, M. C., CHOU, T. T., BRUCE, J., SCHUCK, T., GROSSMAN, M., CLARK, C. M., MCCLUSKEY, L. F., MILLER, B. L., MASLIAH, E., MACKENZIE, I. R., FELDMAN, H., FEIDEN, W., KRETZSCHMAR, H. A., TROJANOWSKI, J. Q. & LEE, V. M. 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science,* 314**,** 130-3.

NG, S. B., BIGHAM, A. W., BUCKINGHAM, K. J., HANNIBAL, M. C., MCMILLIN, M. J., GILDERSLEEVE, H. I., BECK, A. E., TABOR, H. K., COOPER, G. M., MEFFORD, H. C., LEE, C., TURNER, E. H., SMITH, J. D., RIEDER, M. J., YOSHIURA, K., MATSUMOTO, N., OHTA, T., NIIKAWA, N., NICKERSON, D. A., BAMSHAD, M. J. & SHENDURE, J. 2010a. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet,* 42**,** 790-3.

NG, S. B., BUCKINGHAM, K. J., LEE, C., BIGHAM, A. W., TABOR, H. K., DENT, K. M., HUFF, C. D., SHANNON, P. T., JABS, E. W., NICKERSON, D. A., SHENDURE, J. & BAMSHAD, M. J. 2010b. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet,* 42**,** 30-5.

NICOLAE, D. L. 2016. Association Tests for Rare Variants. *Annu Rev Genomics Hum Genet*.

NICOLE, S. & FONTAINE, B. 2015. Skeletal muscle sodium channelopathies. *Curr Opin Neurol,* 28**,** 508-514xs.

NOGALSKA, A., D'AGOSTINO, C., ENGEL, W. K., DAVIES, K. J. & ASKANAS, V. 2010a. Decreased SIRT1 deacetylase activity in sporadic inclusion-body myositis muscle fibers. *Neurobiol Aging,* 31**,** 1637-48.

NOGALSKA, A., D'AGOSTINO, C., ENGEL, W. K., KLEIN, W. L. & ASKANAS, V. 2010b. Novel demonstration of amyloid-beta oligomers in sporadic inclusion-body myositis muscle fibers. *Acta Neuropathol,* 120**,** 661-6.

NOGALSKA, A., D'AGOSTINO, C., TERRACCIANO, C., ENGEL, W. K. & ASKANAS, V. 2010c. Impaired autophagy in sporadic inclusion-body myositis and in endoplasmic reticulum stress-provoked cultured human muscle fibers. *Am J Pathol,* 177**,** 1377-87.

NOGALSKA, A., ENGEL, W. K. & ASKANAS, V. 2010d. Increased BACE1 mRNA and noncoding BACE1-antisense transcript in sporadic inclusion-body myositis muscle fibers--possibly caused by endoplasmic reticulum stress. *Neurosci Lett,* 474**,** 140-3.

NOGALSKA, A., TERRACCIANO, C., D'AGOSTINO, C., KING ENGEL, W. & ASKANAS, V. 2009. p62/SQSTM1 is overexpressed and prominently accumulated in inclusions of sporadic inclusion-body myositis muscle fibers, and can help differentiating it from polymyositis and dermatomyositis. *Acta Neuropathol,* 118**,** 407-13.

NOGALSKA, A., WOJCIK, S., ENGEL, W. K., MCFERRIN, J. & ASKANAS, V. 2007. Endoplasmic reticulum stress induces myostatin precursor protein and NF-kappaB in cultured human muscle fibers: relevance to inclusion body myositis. *Exp Neurol,* 204**,** 610-8.

O'HANLON, T. P., CARRICK, D. M., ARNETT, F. C., REVEILLE, J. D., CARRINGTON, M., GAO, X., ODDIS, C. V., MOREL, P. A., MALLEY, J. D., MALLEY, K., DREYFUSS, J., SHAMIM, E. A., RIDER, L. G., CHANOCK, S. J., FOSTER, C. B., BUNCH, T., PLOTZ, P. H., LOVE, L. A. & MILLER, F. W. 2005. Immunogenetic risk and protective factors for the idiopathic inflammatory myopathies: distinct HLA-A, -B, -Cw, -DRB1 and -DQA1 allelic profiles and motifs define clinicopathologic groups in caucasians. *Medicine (Baltimore),* 84**,** 338-49.

O'ROURKE, J. G., BOGDANIK, L., YÁÑEZ, A., LALL, D., WOLF, A. J., MUHAMMAD, A. K. M. G., HO, R., CARMONA, S., VIT, J. P., ZARROW, J., KIM, K. J., BELL, S., HARMS, M. B., MILLER, T. M., DANGLER, C. A., UNDERHILL, D. M., GOODRIDGE, H. S., LUTZ, C. M. & BALOH, R. H. 2016. C9orf72 is required for proper macrophage and microglial function in mice. *Science,* 351**,** 1324-1329.

ODERMATT, A., TASCHNER, P. E., KHANNA, V. K., BUSCH, H. F., KARPATI, G., JABLECKI, C. K., BREUNING, M. H. & MACLENNAN, D. H. 1996. Mutations in the gene-encoding SERCA1, the fast-twitch skeletal muscle sarcoplasmic reticulum Ca2+ ATPase, are associated with Brody disease. *Nat Genet,* 14**,** 191-4.

OFLAZER, P. S., DEYMEER, F. & PARMAN, Y. 2011. Sporadic-inclusion body myositis (s-IBM) is not so prevalent in Istanbul/Turkey: a muscle biopsy based survey. *Acta Myol,* 30**,** 34-6.

OH, S. J., PARK, K. S., RYAN, H. F., JR., DANON, M. J., LU, J., NAINI, A. B. & DIMAURO, S. 2006. Exercise-induced cramp, myoglobinuria, and tubular aggregates in phosphoglycerate mutase deficiency. *Muscle Nerve,* 34**,** 572-6.

OHLSSON, M., HEDBERG, C., BRADVIK, B., LINDBERG, C., TAJSHARGHI, H., DANIELSSON, O., MELBERG, A., UDD, B., MARTINSSON, T. & OLDFORS, A. 2012. Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin. *Brain,* 135**,** 1682-94.

OKUMA, H., SAITO, F., MITSUI, J., HARA, Y., HATANAKA, Y., IKEDA, M., SHIMIZU, T., MATSUMURA, K., SHIMIZU, J., TSUJI, S. & SONOO, M. 2016. Tubular aggregate myopathy caused by a novel mutation in the cytoplasmic domain of STIM1. *Neurology: Genetics,* 2**,** e50.

OLDFORS, A., LARSSON, N. G., LINDBERG, C. & HOLME, E. 1993. Mitochondrial DNA deletions in inclusion body myositis. *Brain,* 116 (Pt 2)**,** 325-36.

OLDFORS, A., MOSLEMI, A. R., FYHR, I. M., HOLME, E., LARSSON, N. G. & LINDBERG, C. 1995. Mitochondrial DNA deletions in muscle fibers in inclusion body myositis. *J Neuropathol Exp Neurol,* 54**,** 581-7.

OLDFORS, A., MOSLEMI, A. R., JONASSON, L., OHLSSON, M., KOLLBERG, G. & LINDBERG, C. 2006. Mitochondrial abnormalities in inclusion-body myositis. *Neurology,* 66**,** S49-55.

OLGIATI, S., QUADRI, M. & BONIFATI, V. 2016. Genetics of movement disorders in the next-generation sequencing era. *Mov Disord,* 31**,** 458-70.

OLIVE, M., JANUE, A., MORENO, D., GAMEZ, J., TORREJON-ESCRIBANO, B. & FERRER, I. 2009. TAR DNA-Binding protein 43 accumulation in protein aggregate myopathies. *J Neuropathol Exp Neurol,* 68**,** 262-73.

OLIVE, M., KLEY, R. A. & GOLDFARB, L. G. 2013. Myofibrillar myopathies: new developments. *Curr Opin Neurol,* 26**,** 527-35.

ORIGONE, P., VERDIANI, S., CIOTTI, P., GULLI, R., BELLONE, E., MARCHESE, R., ABBRUZZESE, G. & MANDICH, P. 2013. Enlarging the clinical spectrum associated with C9orf 72 repeat expansions: Findings in an Italian cohort of patients with Parkinsonian syndromes and relevance for genetic counselling. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration,* 14**,** 479-480.

ORTH, M., TABRIZI, S. J. & SCHAPIRA, A. H. 2000. Sporadic inclusion body myositis not linked to prion protein codon 129 methionine homozygosity. *Neurology,* 55**,** 1235.

PAN, W. 2009. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol,* 33**,** 497-507.

PAVLOVICOVA, M., NOVOTOVA, M. & ZAHRADNIK, I. 2003. Structure and composition of tubular aggregates of skeletal muscle fibres. *Gen Physiol Biophys,* 22**,** 425-40.

PEARSON, T. A. & MANOLIO, T. A. 2008. How to interpret a genome-wide association study. *JAMA,* 299**,** 1335-44.

PERIASAMY, M. & KALYANASUNDARAM, A. 2007. SERCA pump isoforms: their role in calcium transport and disease. *Muscle Nerve,* 35**,** 430-42.

PFEFFER, G., ELLIOTT, H. R., GRIFFIN, H., BARRESI, R., MILLER, J., MARSH, J., EVILA, A., VIHOLA, A., HACKMAN, P., STRAUB, V., DICK, D. J., HORVATH, R., SANTIBANEZ-KOREF, M., UDD, B. & CHINNERY, P. F. 2012. Titin mutation segregates with hereditary myopathy with early respiratory failure. *Brain,* 135**,** 1695-713.

PHILLIPS, B. A., ZILKO, P. J. & MASTAGLIA, F. L. 2000. Prevalence of sporadic inclusion body myositis in Western Australia. *Muscle Nerve,* 23**,** 970-2.

PIEROBON-BORMIOLI, S., ARMANI, M., RINGEL, S. P., ANGELINI, C., VERGANI, L., BETTO, R. & SALVIATI, G. 1985. Familial neuromuscular disease with tubular aggregates. *Muscle Nerve,* 8**,** 291-8.

PINKUS, J. L., AMATO, A. A., TAYLOR, J. P. & GREENBERG, S. A. 2014. Abnormal distribution of heterogeneous nuclear ribonucleoproteins in sporadic inclusion body myositis. *Neuromuscul Disord*.

POKSAY, K. S., MADDEN, D. T., PETER, A. K., NIAZI, K., BANWAIT, S., CRIPPEN, D., BREDESEN, D. E. & RAO, R. V. 2011. Valosin-containing protein gene mutations: cellular phenotypes relevant to neurodegeneration. *J Mol Neurosci,* 44**,** 91-102.

PREVITALI, S. C., QUATTRINI, A. & BOLINO, A. 2007. Charcot-Marie-Tooth type 4B demyelinating neuropathy: deciphering the role of MTMR phosphatases. *Expert Rev Mol Med,* 9**,** 1-16.

PRICE, P., SANTOSO, L., MASTAGLIA, F., GARLEPP, M., KOK, C. C., ALLCOCK, R. & LAING, N. 2004. Two major histocompatibility complex haplotypes influence susceptibility to sporadic inclusion body myositis: critical evaluation of an association with HLA-DR3. *Tissue Antigens,* 64**,** 575-80.

PULST, S. M. 1999. Genetic linkage analysis. *Arch Neurol,* 56**,** 667-72.

PUTMAN, A. I. & CARBONE, I. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol,* 4**,** 4399-428.

RALSTON, S. H. & LAYFIELD, R. 2012. Pathogenesis of Paget disease of bone. *Calcif Tissue Int,* 91**,** 97-113.

RANQUE-FRANCOIS, B., MAISONOBE, T., DION, E., PIETTE, J. C., CHAUVEHEID, M. P., AMOURA, Z. & PAPO, T. 2005. Familial inflammatory inclusion body myositis. *Ann Rheum Dis,* 64**,** 634-7.

RAPUZZI, S., PRELLE, A., MOGGIO, M., RIGOLETTO, C., CISCATO, P., COMI, G., FRANCESCA, F. & SCARLATO, G. *High serum creatine kinase levels associated with cylindrical spirals at muscle biopsy*.

RAVENSCROFT, G., CLARKE, N. F. & LAING, N. G. 2014. Congenital/ultrastructural myopathies. *In:* HILTON-JONES, D. & TURNER, M. R. (eds.) *Oxford Textbook of Neuromuscular Disorders.* First ed. UK: Oxford University Press.

REA, S. L., MAJCHER, V., SEARLE, M. S. & LAYFIELD, R. 2014. SQSTM1 mutations--bridging Paget disease of bone and ALS/FTLD. *Exp Cell Res,* 325**,** 27-37.

RENTON, A. E., MAJOUNIE, E., WAITE, A., SIMON-SANCHEZ, J., ROLLINSON, S., GIBBS, J. R., SCHYMICK, J. C., LAAKSOVIRTA, H., VAN SWIETEN, J. C., MYLLYKANGAS, L., KALIMO, H., PAETAU, A., ABRAMZON, Y., REMES, A. M., KAGANOVICH, A., SCHOLZ, S. W., DUCKWORTH, J., DING, J., HARMER, D. W., HERNANDEZ, D. G., JOHNSON, J. O., MOK, K., RYTEN, M., TRABZUNI, D., GUERREIRO, R. J., ORRELL, R. W., NEAL, J., MURRAY, A., PEARSON, J., JANSEN, I. E., SONDERVAN, D., SEELAAR, H., BLAKE, D., YOUNG, K., HALLIWELL, N., CALLISTER, J. B., TOULSON, G., RICHARDSON, A., GERHARD, A., SNOWDEN, J., MANN, D., NEARY, D., NALLS, M. A., PEURALINNA, T., JANSSON, L., ISOVIITA, V. M., KAIVORINNE, A. L., HOLTTA-VUORI, M., IKONEN, E., SULKAVA, R., BENATAR, M., WUU, J., CHIO, A., RESTAGNO, G., BORGHERO, G., SABATELLI, M., HECKERMAN, D., ROGAEVA, E., ZINMAN, L., ROTHSTEIN, J. D., SENDTNER, M., DREPPER, C., EICHLER, E. E., ALKAN, C., ABDULLAEV, Z., PACK, S. D., DUTRA, A., PAK, E., HARDY, J., SINGLETON, A., WILLIAMS, N. M., HEUTINK, P., PICKERING-BROWN, S., MORRIS, H. R., TIENARI, P. J. & TRAYNOR, B. J. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron,* 72**,** 257-68.

RIAZUDDIN, S. A., SHAHZADI, A., ZEITZ, C., AHMED, Z. M., AYYAGARI, R., CHAVALI, V. R., PONFERRADA, V. G., AUDO, I., MICHIELS, C., LANCELOT, M. E., NASIR, I. A., ZAFAR, A. U., KHAN, S. N., HUSNAIN, T., JIAO, X., MACDONALD, I. M., RIAZUDDIN, S., SIEVING, P. A., KATSANIS, N. & HEJTMANCIK, J. F. 2010. A mutation in SLC24A1 implicated in autosomal-recessive congenital stationary night blindness. *Am J Hum Genet,* 87**,** 523-31.

RIFAI, Z., WELLE, S., KAMP, C. & THORNTON, C. A. 1995. Ragged red fibers in normal aging and inflammatory myopathy. *Ann Neurol,* 37**,** 24-9.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**,** e47.

ROBINSON, F. L. & DIXON, J. E. 2005. The phosphoinositide-3-phosphatase MTMR2 associates with MTMR13, a membrane-associated pseudophosphatase also mutated in type 4B Charcot-Marie-Tooth disease. *J Biol Chem,* 280**,** 31699-707.

ROBINSON, F. L., NIESMAN, I. R., BEISWENGER, K. K. & DIXON, J. E. 2008. Loss of the inactive myotubularin-related phosphatase Mtmr13 leads to a Charcot-Marie-Tooth 4B2-like peripheral neuropathy in mice. *Proc Natl Acad Sci U S A,* 105**,** 4916-21.

ROHKAMM, R., BOXLER, K., RICKER, K. & JERUSALEM, F. 1983. A dominantly inherited myopathy with excessive tubular aggregates. *Neurology,* 33**,** 331-6.

ROHRER, J. D., WARREN, J. D., REIMAN, D., UPHILL, J., BECK, J., COLLINGE, J., ROSSOR, M. N., ISAACS, A. M. & MEAD, S. 2011. A novel exon 2 I27V VCP variant is associated with dissimilar clinical syndromes. *J Neurol,* 258**,** 1494-6.

ROJANA-UDOMSART, A., JAMES, I., CASTLEY, A., NEEDHAM, M., SCOTT, A., DAY, T., KIERS, L., CORBETT, A., SUE, C., WITT, C., MARTINEZ, P., CHRISTIANSEN, F. & MASTAGLIA, F. 2012. High-resolution HLA-DRB1 genotyping in an Australian inclusion body myositis (s-IBM) cohort: an analysis of disease-associated alleles and diplotypes. *J Neuroimmunol,* 250**,** 77-82.

ROJANA-UDOMSART, A., MITRPANT, C., JAMES, I., WITT, C., NEEDHAM, M., DAY, T., KIERS, L., CORBETT, A., MARTINEZ, P., WILTON, S. D. & MASTAGLIA, F. L. 2013. Analysis of HLA-DRB3 alleles and supertypical genotypes in the MHC Class II region in sporadic inclusion body myositis. *J Neuroimmunol,* 254**,** 174-7.

ROMERO, N. B. 2010. Centronuclear myopathies: a widening concept. *Neuromuscul Disord,* 20**,** 223-8.

ROMERO, N. B. & LAPORTE, J. 2013. Centronuclear Myopathies. *In:* GOEBEL, H. H., SEWRY, C. & WELLER, R. O. (eds.) *Muscle Disease: Pathology and Genetics.* SECOND ed. UK: Wiley-Blackwell.

ROSE, M. R. 2013. 188th ENMC International Workshop: Inclusion Body Myositis, 2-4 December 2011, Naarden, The Netherlands. *Neuromuscul Disord,* 23**,** 1044-55.

ROSENBERG, N. L., NEVILLE, H. E. & RINGEL, S. P. 1985. Tubular aggregates. Their association with neuromuscular diseases, including the syndrome of myalgias/cramps. *Arch Neurol,* 42**,** 973-6.

ROSENFELD, J., SLOAN-BROWN, K. & GEORGE, A. L. 1997. A novel muscle sodium channel mutation causes painful congenital myotonia. *Ann Neurol,* 42**,** 811-814.

ROSES, A. D., LUTZ, M. W., AMRINE-MADSEN, H., SAUNDERS, A. M., CRENSHAW, D. G., SUNDSETH, S. S., HUENTELMAN, M. J., WELSH-BOHMER, K. A. & REIMAN, E. M. 2010. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J,* 10**,** 375-84.

ROSES, A. D., LUTZ, M. W., CRENSHAW, D. G., GROSSMAN, I., SAUNDERS, A. M. & GOTTSCHALK, W. K. 2013. TOMM40 and APOE: Requirements for replication studies of association with age of disease onset and enrichment of a clinical trial. *Alzheimers Dement,* 9**,** 132-6.

ROSSI, D., VEZZANI, B., GALLI, L., PAOLINI, C., TONIOLO, L., PIERANTOZZI, E., SPINOZZI, S., BARONE, V., PEGORARO, E., BELLO, L., CENACCHI, G., VATTEMI, G., TOMELLERI, G., RICCI, G., SICILIANO, G., PROTASI, F., REGGIANI, C. & SORRENTINO, V. 2014. A mutation in the CASQ1 gene causes a vacuolar myopathy with

accumulation of sarcoplasmic reticulum protein aggregates. *Hum Mutat,* 35**,** 1163-70.

ROTHWELL, S., COOPER, R. G., LUNDBERG, I. E., MILLER, F. W., GREGERSEN, P. K., BOWES, J., VENCOVSKY, J., DANKO, K., LIMAYE, V., SELVA-O'CALLAGHAN, A., HANNA, M. G., MACHADO, P. M., PACHMAN, L. M., REED, A. M., RIDER, L. G., COBB, J., PLATT, H., MOLBERG, O., BENVENISTE, O., MATHIESEN, P., RADSTAKE, T., DORIA, A., DE BLEECKER, J., DE PAEPE, B., MAURER, B., OLLIER, W. E., PADYUKOV, L., O'HANLON, T. P., LEE, A., AMOS, C. I., GIEGER, C., MEITINGER, T., WINKELMANN, J., WEDDERBURN, L. R., CHINOY, H., LAMB, J. A. & MYOSITIS GENETICS, C. 2016. Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. *Ann Rheum Dis,* 75**,** 1558-66.

ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* 132**,** 365-86.

RUBINO, E., RAINERO, I., CHIO, A., ROGAEVA, E., GALIMBERTI, D., FENOGLIO, P., GRINBERG, Y., ISAIA, G., CALVO, A., GENTILE, S., BRUNI, A. C., ST GEORGE-HYSLOP, P. H., SCARPINI, E., GALLONE, S. & PINESSI, L. 2012. SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Neurology,* 79**,** 1556-62.

RYGIEL, K. A., MILLER, J., GRADY, J. P., ROCHA, M. C., TAYLOR, R. W. & TURNBULL, D. M. 2015. Mitochondrial and inflammatory changes in sporadic inclusion body myositis. *Neuropathol Appl Neurobiol,* 41**,** 288-303.

RYGIEL, K. A., TUPPEN, H. A., GRADY, J. P., VINCENT, A., BLAKELY, E. L., REEVE, A. K., TAYLOR, R. W., PICARD, M., MILLER, J. & TURNBULL, D. M. 2016. Complex mitochondrial DNA rearrangements in individual cells from patients with sporadic inclusion body myositis. *Nucleic Acids Res,* 44**,** 5313-29.

SALAJEGHEH, M., PINKUS, J. L., TAYLOR, J. P., AMATO, A. A., NAZARENO, R., BALOH, R. H. & GREENBERG, S. A. 2009. Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis. *Muscle Nerve,* 40**,** 19-31.

SALAMEH, J., GOYAL, N., CHOUDRY, R., CAMELO-PIRAGUA, S. & CHONG, P. S. 2013. Phosphoglycerate mutase deficiency with tubular aggregates in a patient from Panama. *Muscle Nerve,* 47**,** 138-40.

SALMIKANGAS, P., MYKKÄNEN, O.-M., GRÖNHOLM, M., HEISKA, L., KERE, J. & CARPÉN, O. 1999. Myotilin, a Novel Sarcomeric Protein with Two Ig-like Domains, is Encoded by a Candidate Gene for Limb-Girdle Muscular Dystrophy. *Hum Mol Genet,* 8**,** 1329-1336.

SALMIKANGAS, P., VAN DER VEN, P. F. M., LALOWSKI, M., TAIVAINEN, A., ZHAO, F., SUILA, H., SCHRÖDER, R., LAPPALAINEN, P., FÜRST, D. O. & CARPÉN, O. 2003. Myotilin, the limb-girdle muscular dystrophy 1A (LGMD1A) protein, cross-links actin filaments and controls sarcomere assembly. *Hum Mol Genet,* 12**,** 189-203.

SALVIATI, G., PIEROBON-BORMIOLI, S., BETTO, R., DAMIANI, E., ANGELINI, C., RINGEL, S. P., SALVATORI, S. & MARGRETH, A. 1985. Tubular aggregates: sarcoplasmic reticulum origin, calcium storage ability, and functional implications. *Muscle Nerve,* 8**,** 299-306.

SANDRI, M. 2010. Autophagy in skeletal muscle. *FEBS Lett,* 584**,** 1411-6.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SANTORELLI, F. M., SCIACCO, M., TANJI, K., SHANSKE, S., VU, T. H., GOLZI, V., GRIGGS, R. C., MENDELL, J. R., HAYS, A. P., BERTORINI, T. E., PESTRONK, A., BONILLA, E. & DIMAURO, S. 1996. Multiple mitochondrial DNA deletions in sporadic inclusion body myositis: a study of 56 patients. *Ann Neurol,* 39**,** 789-95.

SAREEN, D., O'ROURKE, J. G., MEERA, P., MUHAMMAD, A. K. M. G., GRANT, S., SIMPKINSON, M., BELL, S., CARMONA, S., ORNELAS, L., SAHABIAN, A., GENDRON, T., PETRUCELLI, L., BAUGHN, M., RAVITS, J., HARMS, M. B., RIGO, F., BENNETT, C. F., OTIS, T. S., SVENDSEN, C. N. & BALOH, R. H. 2013. Targeting RNA Foci in iPSC-Derived Motor Neurons from ALS Patients with a C9ORF72 Repeat Expansion. *Science Translational Medicine,* 5**,** 208ra149-208ra149.

SARKOZI, E., ASKANAS, V. & ENGEL, W. K. 1994. Abnormal accumulation of prion protein mRNA in muscle fibers of patients with sporadic inclusion-body myositis and hereditary inclusion-body myopathy. *Am J Pathol,* 145**,** 1280-4.

SARKOZI, E., ASKANAS, V., JOHNSON, S. A., ENGEL, W. K. & ALVAREZ, R. B. 1993. beta-Amyloid precursor protein mRNA is increased in inclusion-body myositis muscle. *Neuroreport,* 4**,** 815-8.

SAUNDERS, A. M., STRITTMATTER, W. J., SCHMECHEL, D., GEORGE-HYSLOP, P. H., PERICAK-VANCE, M. A., JOO, S. H., ROSI, B. L., GUSELLA, J. F., CRAPPER-MACLACHLAN, D. R., ALBERTS, M. J. & ET AL. 1993. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology,* 43**,** 1467-72.

SCHEUNER, D., ECKMAN, C., JENSEN, M., SONG, X., CITRON, M., SUZUKI, N., BIRD, T. D., HARDY, J., HUTTON, M., KUKULL, W., LARSON, E., LEVY-LAHAD, E., VIITANEN, M., PESKIND, E., POORKAJ, P., SCHELLENBERG, G., TANZI, R., WASCO, W., LANNFELT, L., SELKOE, D. & YOUNKIN, S. 1996. Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med,* 2**,** 864-70.

SCHMIDT, J., BARTHEL, K., WREDE, A., SALAJEGHEH, M., BAHR, M. & DALAKAS, M. C. 2008. Interrelation of inflammation and APP in sIBM: IL-1 beta induces accumulation of beta-amyloid in skeletal muscle. *Brain,* 131**,** 1228-40.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc,* 3**,** 1101-8.

SCHRODER, R. 2013. Protein aggregate myopathies: the many faces of an expanding disease group. *Acta Neuropathol,* 125**,** 1-2.

SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods,* 11**,** 361-2.

SCOTT, A. P., ALLCOCK, R. J., MASTAGLIA, F., NISHINO, I., NONAKA, I. & LAING, N. 2006. Sporadic inclusion body myositis in Japanese is associated with the MHC ancestral haplotype 52.1. *Neuromuscul Disord,* 16**,** 311-5.

SCOTT, A. P., LAING, N. G., MASTAGLIA, F., DALAKAS, M., NEEDHAM, M. & ALLCOCK, R. J. 2012. Investigation of NOTCH4 coding region

polymorphisms in sporadic inclusion body myositis. *J Neuroimmunol,* 250**,** 66-70.

SCOTT, A. P., LAING, N. G., MASTAGLIA, F., NEEDHAM, M., WALTER, M. C., DALAKAS, M. C. & ALLCOCK, R. J. 2011. Recombination mapping of the susceptibility region for sporadic inclusion body myositis within the major histocompatibility complex. *J Neuroimmunol,* 235**,** 77-83.

SEELEN, M., VISSER, A. E., OVERSTE, D. J., KIM, H. J., PALUD, A., WONG, T. H., VAN SWIETEN, J. C., SCHELTENS, P., VOERMANS, N. C., BAAS, F., DE JONG, J. M., VAN DER KOOI, A. J., DE VISSER, M., VELDINK, J. H., TAYLOR, J. P., VAN ES, M. A. & VAN DEN BERG, L. H. 2014. No mutations in hnRNPA1 and hnRNPA2B1 in Dutch patients with amyotrophic lateral sclerosis, frontotemporal dementia, and inclusion body myopathy. *Neurobiol Aging,* 35**,** 1956 e9-1956 e11.

SEIBENHENER, M. L., BABU, J. R., GEETHA, T., WONG, H. C., KRISHNA, N. R. & WOOTEN, M. W. 2004. Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. *Mol Cell Biol,* 24**,** 8055-68.

SELCEN, D. 2008. Myofibrillar myopathies. *Curr Opin Neurol,* 21**,** 585-589.

SELCEN, D. & ENGEL, A. G. 2004. Mutations in myotilin cause myofibrillar myopathy. *Neurology,* 62**,** 1363-71.

SENDEREK, J., BERGMANN, C., WEBER, S., KETELSEN, U.-P., SCHORLE, H., RUDNIK-SCHÖNEBORN, S., BÜTTNER, R., BUCHHEIM, E. & ZERRES, K. 2003. Mutation of the SBF2 gene, encoding a novel member of the myotubularin family, in Charcot–Marie–Tooth neuropathy type 4B2/11p15. *Hum Mol Genet,* 12**,** 349-356.

SEWRY, C. & GOEBEL, H. H. 2013. General Pathology of Muscle Disease. *In:* GOEBEL, H. H., SEWRY, C. & WELLER, R. O. (eds.) *Muscle Disease: Pathology and Genetics.* SECOND ed. UK: Wiley-Blackwell.

SHAHRIZAILA, N., LOWE, J. & WILLS, A. 2004. Familial myopathy with tubular aggregates associated with abnormal pupils. *Neurology,* 63**,** 1111-3.

SHALABY, S., MITSUHASHI, H., MATSUDA, C., MINAMI, N., NOGUCHI, S., NONAKA, I., NISHINO, I. & HAYASHI, Y. K. 2009. Defective Myotilin Homodimerization Caused by a Novel Mutation in MYOT Exon 9 in the First Japanese Limb Girdle Muscular Dystrophy 1A Patient. *Journal of Neuropathology & Experimental Neurology,* 68**,** 701-707.

SHERRINGTON, R., ROGAEV, E. I., LIANG, Y., ROGAEVA, E. A., LEVESQUE, G., IKEDA, M., CHI, H., LIN, C., LI, G., HOLMAN, K., TSUDA, T., MAR, L., FONCIN, J. F., BRUNI, A. C., MONTESI, M. P., SORBI, S., RAINERO, I., PINESSI, L., NEE, L., CHUMAKOV, I., POLLEN, D., BROOKES, A., SANSEAU, P., POLINSKY, R. J., WASCO, W., DA SILVA, H. A., HAINES, J. L., PERKICAK-VANCE, M. A., TANZI, R. E., ROSES, A. D., FRASER, P. E., ROMMENS, J. M. & ST GEORGE-HYSLOP, P. H. 1995. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature,* 375**,** 754-60.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res,* 29**,** 308-11.

SIVAKUMAR, K., CERVENAKOVA, L., DALAKAS, M. C., LEON-MONZON, M., ISAACSON, S. H., NAGLE, J. W., VASCONCELOS, O. &

GOLDFARB, L. G. 1995. Exons 16 and 17 of the amyloid precursor protein gene in familial inclusion body myopathy. *Ann Neurol,* 38**,** 267-9.

SIVAKUMAR, K., SEMINO-MORA, C. & DALAKAS, M. C. 1997. An inflammatory, familial, inclusion body myositis with autoimmune features and a phenotype identical to sporadic inclusion body myositis. Studies in three families. *Brain,* 120 (Pt 4)**,** 653-61.

SKOL, A. D., SCOTT, L. J., ABECASIS, G. R. & BOEHNKE, M. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet,* 38**,** 209-13.

SREEDHARAN, J., BLAIR, I. P., TRIPATHI, V. B., HU, X., VANCE, C., ROGELJ, B., ACKERLEY, S., DURNALL, J. C., WILLIAMS, K. L., BURATTI, E., BARALLE, F., DE BELLEROCHE, J., MITCHELL, J. D., LEIGH, P. N., AL-CHALABI, A., MILLER, C. C., NICHOLSON, G. & SHAW, C. E. 2008. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science,* 319**,** 1668-72.

STERNBERG, D., MAISONOBE, T., JURKAT-ROTT, K., NICOLE, S., LAUNAY, E., CHAUVEAU, D., TABTI, N., LEHMANN-HORN, F., HAINQUE, B. & FONTAINE, B. 2001. Hypokalaemic periodic paralysis type 2 caused by mutations at codon 672 in the muscle sodium channel gene SCN4A. *Brain,* 124**,** 1091-9.

STOJKOVIC, T., HAMMOUDA EL, H., RICHARD, P., LOPEZ DE MUNAIN, A., RUIZ-MARTINEZ, J., CAMANO, P., LAFORET, P., PENISSON-BESNIER, I., FERRER, X., LACOUR, A., LACOMBLEZ, L., CLAEYS, K. G., MAURAGE, C. A., FARDEAU, M. & EYMARD, B. 2009. Clinical outcome in 19 French and Spanish patients with valosin-containing protein myopathy associated with Paget's disease of bone and frontotemporal dementia. *Neuromuscul Disord,* 19**,** 316-23.

SUGARMAN, M. C., KITAZAWA, M., BAKER, M., CAIOZZO, V. J., QUERFURTH, H. W. & LAFERLA, F. M. 2006. Pathogenic accumulation of APP in fast twitch muscle of IBM patients and a transgenic model. *Neurobiol Aging,* 27**,** 423-32.

SUGARMAN, M. C., YAMASAKI, T. R., ODDO, S., ECHEGOYEN, J. C., MURPHY, M. P., GOLDE, T. E., JANNATIPOUR, M., LEISSRING, M. A. & LAFERLA, F. M. 2002. Inclusion body myositis-like phenotype induced by transgenic overexpression of beta APP in skeletal muscle. *Proc Natl Acad Sci U S A,* 99**,** 6334-9.

SUZUKI, N., AOKI, M., MORI-YOSHIMURA, M., HAYASHI, Y. K., NONAKA, I. & NISHINO, I. 2012. Increase in number of sporadic inclusion body myositis (sIBM) in Japan. *J Neurol,* 259**,** 554-6.

TAN, J. A., ROBERTS-THOMSON, P. J., BLUMBERGS, P., HAKENDORF, P., COX, S. R. & LIMAYE, V. 2013. Incidence and prevalence of idiopathic inflammatory myopathies in South Australia: a 30-year epidemiologic study of histology-proven cases. *Int J Rheum Dis,* 16**,** 331-8.

TARATUTO, A. L., MATTEUCCI, M., BARREIRO, C., SACCOLITTI, M. & SEVLEVER, G. 1991. Autosomal dominant neuromuscular disease with cylindrical spirals. *Neuromuscular Disorders,* 1**,** 433-441.

TATEYAMA, M., SAITO, N., FUJIHARA, K., SHIGA, Y., TAKEDA, A., NARIKAWA, K., HASEGAWA, T., TAGUCHI, Y., SAKUMA, R., ONODERA, Y., OHNUMA, A., TOBITA, M. & ITOYAMA, Y. 2003.

Familial inclusion body myositis: a report on two Japanese sisters. *Intern Med,* 42**,** 1035-8.

TAWIL, R. & GRIGGS, R. C. 2002. Inclusion body myositis. *Curr Opin Rheumatol,* 14**,** 653-7.

TERSAR, K., BOENTERT, M., BERGER, P., BONNEICK, S., WESSIG, C., TOYKA, K. V., YOUNG, P. & SUTER, U. 2007. Mtmr13/Sbf2-deficient mice: an animal model for CMT4B2. *Hum Mol Genet,* 16**,** 2991-3001.

THIEL, C., SCHWARZ, M., PENG, J., GRZMIL, M., HASILIK, M., BRAULKE, T., KOHLSCHÜTTER, A., VON FIGURA, K., LEHLE, L. & KÖRNER, C. 2003. A New Type of Congenital Disorders of Glycosylation (CDG-Ii) Provides New Insights into the Early Steps of Dolichol-linked Oligosaccharide Biosynthesis. *Journal of Biological Chemistry,* 278**,** 22498-22505.

THOMAS, P. K., WORKMAN, J. M. & THAGE, O. 1984. Behr's syndrome: A family exhibiting pseudodominant inheritance. *J Neurol Sci,* 64**,** 137-148.

TODD, T. W. & PETRUCELLI, L. 2016. Insights into the pathogenic mechanisms of Chromosome 9 open reading frame 72 (C9orf72) repeat expansions. *J Neurochem***,** n/a-n/a.

TONIN, P., BRUNO, C., CASSANDRINI, D., SAVIO, C., TAVAZZI, E., TOMELLERI, G. & PICCOLO, G. 2009. Unusual presentation of phosphoglycerate mutase deficiency due to two different mutations in PGAM-M gene. *Neuromuscul Disord,* 19**,** 776-8.

TOSCANO, A., TSUJINO, S., VITA, G., SHANSKE, S., MESSINA, C. & DIMAURO, S. 1996. Molecular basis of muscle phosphoglycerate mutase (PGAM-M) deficiency in the Italian kindred. *Muscle Nerve,* 19**,** 1134-7.

TSUJINO, S., SHANSKE, S., SAKODA, S., FENICHEL, G. & DIMAURO, S. 1993. The molecular genetic basis of muscle phosphoglycerate mutase (PGAM) deficiency. *Am J Hum Genet,* 52**,** 472-7.

TSUJINO, S., SHANSKE, S., SAKODA, S., TOSCANO, A. & DIMAURO, S. 1995. Molecular genetic studies in muscle phosphoglycerate mutase (PGAM-M) deficiency. *Muscle Nerve Suppl,* 3**,** S50-3.

TUCKER, W. S., JR., HUBBARD, W. H., STRYKER, T. D., MORGAN, S. W., EVANS, O. B., FREEMON, F. R. & THEIL, G. B. 1982. A new familial disorder of combined lower motor neuron degeneration and skeletal disorganization. *Trans Assoc Am Physicians,* 95**,** 126-34.

VAN LEEUWEN, F. W., HOL, E. M. & FISCHER, D. F. 2006. Frameshift proteins in Alzheimer's disease and in other conformational disorders: time for the ubiquitin-proteasome system. *J Alzheimers Dis,* 9**,** 319-25.

VAN SLEGTENHORST, M., LEWIS, J. & HUTTON, M. 2000. The molecular genetics of the tauopathies. *Exp Gerontol,* 35**,** 461-71.

VERSCHUUREN J J, B. U., VAN ENGELEN BG, VAN DER HOEVEN JH, HOOGENDIJK J, WINTZEN AR 1997. *Inclusion body myositis,* London, Royal Society of Medicine Press.

VERSHUUREN, J. J., VAN ENGELEN, B. G. M., VAN DER HOEVEN, J. & HOOGENDIJK, J. 1997. *Diagnostic Criteria for Neuromuscular Disorders*.

VISSER, M. 2013. Clinical Features of Muscle Disease. *In:* GOEBEL, H. H., SEWRY, C. & WELLER, R. O. (eds.) *Muscle Disease: Pathology and Genetics.* SECOND ed. UK: Wiley-Blackwell.

VISSER, M. & VAN DER KOOI, A. J. 2014. Idiopathic inflammatory myopathies. *In:* HILTON-JONES, D. & TURNER, M. R. (eds.) *Oxford Textbook of Neuromuscular Disorders.* First ed. UK: Oxford University Press.

VISSING, J., SCHMALBRUCH, H., HALLER, R. G. & CLAUSEN, T. 1999. Muscle phosphoglycerate mutase deficiency with tubular aggregates: effect of dantrolene. *Ann Neurol,* 46**,** 274-7.

VOERMANS, N. C., LAAN, A. E., OOSTERHOF, A., VAN KUPPEVELT, T. H., DROST, G., LAMMENS, M., KAMSTEEG, E. J., SCOTTON, C., GUALANDI, F., GUGLIELMI, V., VAN DEN HEUVEL, L., VATTEMI, G. & VAN ENGELEN, B. G. 2012. Brody syndrome: a clinically heterogeneous entity distinct from Brody disease: a review of literature and a cross-sectional clinical study in 17 patients. *Neuromuscul Disord,* 22**,** 944-54.

WAITE, A. J., BAUMER, D., EAST, S., NEAL, J., MORRIS, H. R., ANSORGE, O. & BLAKE, D. J. 2014. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiol Aging,* 35**,** 1779 e5-1779 e13.

WALTER, M. C., ROSSIUS, M., ZITZELSBERGER, M., VORGERD, M., MULLER-FELBER, W., ERTL-WAGNER, B., ZHANG, Y., BRINKMEIER, H., SENDEREK, J. & SCHOSER, B. 2015. 50 years to diagnosis: Autosomal dominant tubular aggregate myopathy caused by a novel STIM1 mutation. *Neuromuscul Disord*.

WANG, D. W., VANDECARR, D., RUBEN, P. C., GEORGE, A. L. & BENNETT, P. B. 1999. Functional consequences of a domain 1/S6 segment sodium channel mutation associated with painful congenital myotonia. *FEBS Lett,* 448**,** 231-234.

WANG, K., LI, M. & HAKONARSON, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38**,** e164.

WATTS, G. D., THOMASOVA, D., RAMDEEN, S. K., FULCHIERO, E. C., MEHTA, S. G., DRACHMAN, D. A., WEIHL, C. C., JAMROZIK, Z., KWIECINSKI, H., KAMINSKA, A. & KIMONIS, V. E. 2007. Novel VCP mutations in inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia. *Clin Genet,* 72**,** 420-6.

WATTS, G. D., WYMER, J., KOVACH, M. J., MEHTA, S. G., MUMM, S., DARVISH, D., PESTRONK, A., WHYTE, M. P. & KIMONIS, V. E. 2004. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet,* 36**,** 377-81.

WEDATILAKE, Y., PLAGNOL, V., ANDERSON, G., PAINE, S. M. L., CLAYTON, P. T., JACQUES, T. S. & RAHMAN, S. 2015. Tubular aggregates caused by serine active site containing 1 (SERAC1) mutations in a patient with a mitochondrial encephalopathy. *Neuropathol Appl Neurobiol,* 41**,** 399-402.

WEIHL, C. C., BALOH, R. H., LEE, Y., CHOU, T. F., PITTMAN, S. K., LOPATE, G., ALLRED, P., JOCKEL-BALSAROTTI, J., PESTRONK, A. & HARMS, M. B. 2015. Targeted sequencing and identification of genetic variants in sporadic inclusion body myositis. *Neuromuscul Disord*.

WEIHL, C. C., TEMIZ, P., MILLER, S. E., WATTS, G., SMITH, C., FORMAN, M., HANSON, P. I., KIMONIS, V. & PESTRONK, A. 2008. TDP-43

accumulation in inclusion body myopathy muscle suggests a common pathogenic mechanism with frontotemporal dementia. *J Neurol Neurosurg Psychiatry,* 79**,** 1186-9.

WELTER, D., MACARTHUR, J., MORALES, J., BURDETT, T., HALL, P., JUNKINS, H., KLEMM, A., FLICEK, P., MANOLIO, T., HINDORFF, L. & PARKINSON, H. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res,* 42**,** D1001-6.

WILKIE, D. R. 1968. *Muscle,* UK, Edward Arnold.

WILMSHURST, J. M., LILLIS, S., ZHOU, H., PILLAY, K., HENDERSON, H., KRESS, W., MULLER, C. R., NDONDO, A., CLOKE, V., CULLUP, T., BERTINI, E., BOENNEMANN, C., STRAUB, V., QUINLIVAN, R., DOWLING, J. J., AL-SARRAJ, S., TREVES, S., ABBS, S., MANZUR, A. Y., SEWRY, C. A., MUNTONI, F. & JUNGBLUTH, H. 2010. RYR1 mutations are a common cause of congenital myopathies with central nuclei. *Ann Neurol,* 68**,** 717-26.

WILSON, F. C., YTTERBERG, S. R., ST SAUVER, J. L. & REED, A. M. 2008. Epidemiology of sporadic inclusion body myositis and polymyositis in Olmsted County, Minnesota. *J Rheumatol,* 35**,** 445-7.

WOLFE, G. I., BURNS, D. K., KRAMPITZ, D. & BAROHN, R. J. 1997. Cylindrical spirals of myofilamentous origin associated with exertional cramps and rhabdomyolysis. *Neuromuscular Disorders,* 7**,** 536-538.

WONG, E. & CUERVO, A. M. 2010. Autophagy gone awry in neurodegenerative diseases. *Nat Neurosci,* 13**,** 805-11.

WORTMANN, S. B., VAZ, F. M., GARDEITCHIK, T., VISSERS, L. E. L. M., RENKEMA, G. H., SCHUURS-HOEIJMAKERS, J. H. M., KULIK, W., LAMMENS, M., CHRISTIN, C., KLUIJTMANS, L. A. J., RODENBURG, R. J., NIJTMANS, L. G. J., GRUNEWALD, A., KLEIN, C., GERHOLD, J. M., KOZICZ, T., VAN HASSELT, P. M., HARAKALOVA, M., KLOOSTERMAN, W., BARIC, I., PRONICKA, E., UCAR, S. K., NAESS, K., SINGHAL, K. K., KRUMINA, Z., GILISSEN, C., VAN BOKHOVEN, H., VELTMAN, J. A., SMEITINK, J. A. M., LEFEBER, D. J., SPELBRINK, J. N., WEVERS, R. A., MORAVA, E. & DE BROUWER, A. P. M. 2012. Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial function and intracellular cholesterol trafficking and cause dystonia and deafness. *Nat Genet,* 44**,** 797-802.

WU, M. C., LEE, S., CAI, T., LI, Y., BOEHNKE, M. & LIN, X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet,* 89**,** 82-93.

WU, X., RUSH, J. S., KARAOGLU, D., KRASNEWICH, D., LUBINSKY, M. S., WAECHTER, C. J., GILMORE, R. & FREEZE, H. H. 2003. Deficiency of UDP-GlcNAc:Dolichol Phosphate N-Acetylglucosamine-1 Phosphate Transferase (DPAGT1) Causes a Novel Congenital Disorder of Glycosylation Type Ij. *Hum Mutat,* 22**,** 144-150.

XI, Z., VAN BLITTERSWIJK, M., ZHANG, M., MCGOLDRICK, P., MCLEAN, JESSE R., YUNUSOVA, Y., KNOCK, E., MORENO, D., SATO, C., MCKEEVER, PAUL M., SCHNEIDER, R., KEITH, J., PETRESCU, N., FRASER, P., TARTAGLIA, MARIA C., BAKER, MATTHEW C., GRAFF-RADFORD, NEILL R., BOYLAN, KEVIN B., DICKSON, DENNIS W., MACKENZIE, IAN R., RADEMAKERS, R., ROBERTSON, J., ZINMAN,

L. & ROGAEVA, E. 2015. Jump from Pre-mutation to Pathologic Expansion in C9orf72. *The American Journal of Human Genetics,* 96**,** 962-970.

XU, J.-W., LIU, F.-C., LI, W., ZHAO, Y.-Y., ZHAO, D.-D., LUO, Y.-B., LU, J.-Q. & YAN, C.-Z. 2015. Cylindrical Spirals in Skeletal Muscles Originate From the Longitudinal Sarcoplasmic Reticulum. *Journal of Neuropathology & Experimental Neurology*.

YAMAMOTO, H., SAHASHI, K., MIZUNO, Y., IBI, T. & SOBUE, G. 1982. [A case of a mitochondrial myopathy with cylindrical spirals (author's transl)]. *臨床神経学,* 22**,** 244-250.

YANG, C. C., ALVAREZ, R. B., ENGEL, W. K. & ASKANAS, V. 1996. Increase of nitric oxide synthases and nitrotyrosine in inclusion-body myositis. *Neuroreport,* 8**,** 153-8.

YANG, J., LEE, S. H., GODDARD, M. E. & VISSCHER, P. M. 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics,* 88**,** 76-82.

YU, C. E., SELTMAN, H., PESKIND, E. R., GALLOWAY, N., ZHOU, P. X., ROSENTHAL, E., WIJSMAN, E. M., TSUANG, D. W., DEVLIN, B. & SCHELLENBERG, G. D. 2007. Comprehensive analysis of APOE and selected proximate markers for late-onset Alzheimer's disease: patterns of linkage disequilibrium and disease/marker association. *Genomics,* 89**,** 655-65.

YUNIS, E. J. & SAMAHA, F. J. 1971. Inclusion body myositis. *Lab Invest,* 25**,** 240-8.

ZHANG, S. L., YEROMIN, A. V., HU, J., AMCHESLAVSKY, A., ZHENG, H. & CAHALAN, M. D. 2011. Mutations in Orai1 transmembrane segment 1 cause STIM1-independent activation of Orai1 channels at glycine 98 and channel closure at arginine 91. *Proceedings of the National Academy of Sciences,* 108**,** 17838-17843.

ZOLTOWSKA, K., WEBSTER, R., FINLAYSON, S., MAXWELL, S., COSSINS, J., MULLER, J., LOCHMULLER, H. & BEESON, D. 2013. Mutations in GFPT1 that underlie limb-girdle congenital myasthenic syndrome result in reduced cell-surface expression of muscle AChR. *Hum Mol Genet,* 22**,** 2905-13.

ZONDERVAN, K. T. & CARDON, L. R. 2007. Designing candidate gene and genome-wide case-control association studies. *Nat. Protocols,* 2**,** 2492-2501.