Using 'Next-Generation' Sequencing in the Identification of Novel Causes of Inherited Heart Diseases

Rob Hastings, Radcliffe Department of Medicine and Green

Templeton College, University of Oxford

Thesis Submission for the Degree of Doctor of Philosophy (DPhil) in

Cardiovascular Medicine

The author has been kindly supported in this work by a British Heart Foundation Centre of Research Excellence Fellowship Award (grant code HSRNWBY) and Research Pump-Priming Award (grant code HSRNWB11)

Abstract

"Using 'Next-Generation' Sequencing in the Identification of Novel Causes of Inherited Heart Diseases"

Rob Hastings, Green Templeton College

Submitted for the degree of Doctor of Philosophy (DPhil) in Cardiovascular Medicine. Trinity Term, 2013.

Next-generation sequencing methods now allow rapid and cost-effective sequencing of DNA on a scale not previously possible. This offers great opportunities for the research of Mendelian disorders, but also significant challenges. The sequencing of exomes, or whole genomes, has emerged as a powerful clinical research tool, with targeted gene analyses generally being preferred in the clinical diagnostic setting. These methods have been employed here with the aim of identifying novel genetic causes of inherited heart disorders and to gain insights into the utility and limitations of these techniques for clinical diagnosis in these disorders.

Data produced from the introduction of a targeted multi-gene next-generation sequencing test into clinical practice has been studied. Variation within the mitochondrial genome has been analysed to assess the importance of mitochondrial DNA variants in patients with hypertrophic cardiomyopathy. The m.4300A>G mutation is identified as an important cause of this disorder, with other previously cardiomyopathy-associated and novel variants also identified. Such multi-gene tests can facilitate interpretable and phenotype-relevant results, but at the expense of limiting more extensive data acquisition.

Whole-genome sequencing has been performed in five families with different autosomal dominant inherited heart disease phenotypes of unknown genetic aetiology. In two of these likely pathogenic variants were identified, one in the gene encoding titin (*TTN*) and the other in the calcium channel subunit gene *CACNA1C*. *In vitro* studies were undertaken to support the pathogenicity of the *TTN* variant and understand the functional effects of this. In the other three families either multiple candidate gene variants were identified or no clear candidate variant was identified. This highlights the difficulties in interpreting these results, even in carefully selected families. Overall, although the research benefits of exome or genome studies are evident, the interpretation and validation of genetic variant data produced remains highly challenging for clinical diagnosis.

Contents

Chapter 1: Introduction 1

1.1 Objectives 1

1.2 Overview 1

1.3 Background 2

1.3.1 The Study of Mendelian Disorders 2

1.3.2 Inherited Heart Diseases 3

1.3.3 The Historical Basis of Current Genetic Understanding and Technologies 3

1.3.4 Developments in Gene Identification Techniques 4

1.4 Next Generation Sequencing 6

1.4.1 Next Generation Sequencing Platforms 7

1.4.2 Applications of Next Generation Sequencing 9

1.4.3 Advantages of Next-Generation Sequencing in Identifying the Causes of Mendelian Disorders 10

1.4.4 Ethical Considerations 11

1.5 Genomic Variation 13

1.5.1 The Human Genome Project 13

1.5.2 The Reference Genome 14

1.5.3 Genetic Variability Identified by Next Generation Sequencing 14

1.5.4 Identification of Pathogenic Mutations 15

1.6 Genetics of Inherited Heart Disorders 18

1.6.1 Current Genetic Understanding 18

1.6.2 Clinical Diagnostic Testing 20

Chapter 2: Mitochondrial Genomic Variation in Hypertrophic Cardiomyopathy 23

2.1 Background 23

2.1.1 Hypertrophic Cardiomyopathy 23

2.1.2 Mitochondria 25

2.2 Methods 28

2.2.1 Patients 28
2.2.2 Sequencing 29
2.2.3 Bioinformatic Analysis of MtDNA Variants 31
2.3 Results 32
2.3.1 Initial Results 32
2.3.2 Further Patient Cohort 34
2.3.3 Pyrosequencing of the m.4300A>G Variant 34
2.3.4 Haplotype Analysis 35
2.3.5 Validation of Other MtDNA Variants 35
2.3.6 Final Results 36
2.3.7 Clinical Phenotype of Families with the m.4300A>G Variant 36
2.4 Discussion 40
2.4.1 m.4300A>G as a Cause of an HCM Phenotype 40
2.4.2 Other Cardiomyopathy-Associated Variants 42
2.4.3 Novel Variants 43
2.5 Conclusions 43

Chapter 3: Study Design and Methods in Family Studies 45

3.1 Study Design and Strategy 45
3.2 Study Information 49
3.2.1 Patient Recruitment 49
3.2.2 Consent and Ethics 49
3.2.3 Funding 51
3.2.4 WGS500 Project 51
3.3 Methods 52
3.3.1 DNA Extraction 52
3.3.2 Whole Genome Sequencing 52
3.3.3 Single Nucleotide Polymorphism Array 58
3.3.4 Data Analysis 61
3.3.5 Sanger Sequencing 66

Chapter 4: Whole Genome Sequencing in Families with Autosomal Dominant Inherited Heart Diseases 69

4.1 A Family with a Novel Repolarisation Abnormality Phenotype 69

4.1.1 Clinical Information 69

4.1.2 Repolarisation 72

4.1.3 Genetic Studies Performed 75

4.1.4 Results 75

4.1.5 Conclusions 81

4.2 A Family with Autosomal Dominant Dilated Cardiomyopathy 82

4.2.1 Clinical Information 82

4.2.2 Dilated Cardiomyopathy 84

4.2.3 Genetic Studies Performed 85

4.2.4 Results 85

4.2.5 Conclusions 93

4.3 A Family with a Combined Phenotype of Different Cardiomyopathies, Electrophysiological Abnormalities and Septal Defects 94

4.3.1 Clinical Information 94

4.3.2 Complex Cardiac Phenotypes 96

4.3.3 Genetic Studies Performed 97

4.3.4 Results 98

4.3.5 Conclusions 108

4.4 A Family with a Combined Phenotype of Left Ventricular Hypertrophy, Prolonged QT Interval, and Sinoatrial Node Dysfunction 109

4.4.1 Clinical Information 109

4.4.2 Long QT Syndrome and Cardiomyopathy 112

4.4.3 Genetic Studies Performed 113

4.4.4 Results 114

4.4.5 Discussion 116

4.4.6 Future Work 118

4.5 A Family with Autosomal Dominant Left Ventricular Non-Compaction 119

4.5.1 Clinical Information 119

4.5.2 Left Ventricular Non-Compaction 122
4.5.3 Genetic Studies Performed 123
4.5.4 Results 123

Chapter 5: *In Vitro* Studies of the Titin Z1Z2 Region 131

5.1 Titin Overview 131
5.1.1 Structure 132
5.1.2 Isoforms 135
5.1.3 Functions 136
5.1.4 Regulation 139
5.2 The Titin Z1Z2 Region and Telethonin 140
5.2.1 Titin Z1Z2 Structure 140
5.2.2 The Titin Z1Z2/Telethonin Complex 142
5.3 Human Phenotypes Attributable to Mutations in Titin and Telethonin 149
5.3.1 Titin 149
5.3.2 Telethonin 151
5.4 Hypotheses 152
5.5 Methods 154
5.5.1 Expression Constructs 154
5.5.2 Protein Expression and Purification 161
5.5.3 Structural Studies 165
5.5.4 Localisation Studies 168
5.5.5 Binding Studies 170
5.6 Results 171
5.6.1 Structural Studies 171
5.6.2 Localisation Studies 186
5.6.3 Binding Studies 188
5.7 Discussion 191
5.7.1 Genetic Findings 191
5.7.2 Experimental Findings 191
5.7.3 Titin Mutations as a Cause of Left Ventricular Non-Compaction 194
5.7.4 Clinical Significance of the Titin Z1Z2-Telethonin Interaction 196

5.8 Conclusions 197
5.9 Future Work 197

Chapter 6: Discussion and Conclusions 199

6.1 Summary 199
6.2 Discussion 200
6.2.1 Multi-Gene Next Generation Sequencing Tests in Clinical Research and Diagnostics 200
6.2.2 Exome and Genome Studies in Clinical Research and Diagnostics 201
6.2.3 Limitations of Exome Analysis in the Identification of the Cause of Mendelian Disorders 203
6.3 Final Conclusions 210

References 211

Appendix 237

I. Study Documents 237
II. Experimental Protocols and Information 243
III. Sanger Sequencing Primers 252
IV. Supplementary Results 258

Acknowledgements

This thesis was kindly supervised by Professor Hugh Watkins and Dr Katja Gehmlich, Radcliffe Department of Medicine, University of Oxford

Certain elements of the experimental studies have been assisted or performed by other members of the University of Oxford or Oxford University Hospitals NHS Trust Molecular Genetics Laboratory:

- *Library preparation and whole genome sequencing was performed by the Oxford Genomics Centre, Wellcome Trust Centre of Human Genetics (WTCHG), University of Oxford, as part of the WGS500 project.*
- *SNP arrays were performed under supervision of Dr Elham Sadighi-Akha, or performed separately by Dr Alistair Pagnamenta, both members of Dr Sam Knight's group at the Wellcome Trust Centre for Human Genetics, University of Oxford.*
- *Data analysis has been supported by Dr Stefano Lise, Biomedical Research Centre Genomics Unit, Wellcome Trust Centre for Human Genetics (WTCHG), University of Oxford.*
- *Sequencing has been undertaken as part of the clinical sequencing pathway at the Oxford University Hospitals NHS Trust Molecular Genetics Laboratory, Churchill Hospital, Oxford.*
- *Sanger sequencing validation of novel mitochondrial variants was undertaken by Dr Carl Fratter, Oxford University Hospitals NHS Trust Molecular Genetics Laboratory, Churchill Hospital, Oxford.*
- *Protein sequence confirmation by mass spectrometry was performed by Dr Katalin Di Gleria, Weatherall Institute of Molecular Medicine, University of Oxford.*

Commonly Used Abbreviations

1000G	1000 Genomes Project
ACMG	American College of Medical Genetics and Genomics
AF	Atrial Fibrillation
AHA	American Heart Association
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
dbSNP	NCBI Database of Single Nucleotide Polymorphisms
DCM	Dilated Cardiomyopathy
DNA	Deoxyribonucleic Acid
ESC	European Society of Cardiology
ESHG	European Society of Human Genetics
EVS	NHLBI-GO Exome Variant Server
HCM	Hypertrophic Cardiomyopathy
LQTS	Long QT Syndrome
LVH	Left Ventricular Hypertrophy
LVNC	Left Ventricular Non-Compaction
MLP	Muscle LIM protein
MtDB	Human mitochondrial genome database
MtDNA	Mitochondrial DNA
NGS	Next Generation Sequencing
SNP	Single Nucleotide Polymorphism
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WGS500	Collaborative study between the University of Oxford and Illumina to study five-hundred whole genome sequences in human disease
WT	Wild-Type
Z1Z2	The Z1Z2 region of titin, incorporating Z1 and Z2 Ig-like domains

Laboratory Abbreviations

A178D	p.A178D Substitution (in titin)
αα	Amino Acids
APS	Ammonium Persulphate
CD	Circular Dichroism
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal Bovine Serum
GST	Glutathione-S Transferase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Broth
Ni-NTA	Nickel-Nitrilotriacetic Acid
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate-Buffered Saline
PBST	Phosphate-Buffered Saline with Tween
PCR	Polymerase Chain Reaction
PFA	4% Paraformaldehyde in PBS
SDS-PAGE	Sodium-Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis
SOB	Super Optimal Broth
SOC	Super Optimal broth with Catabolite repression
TBE	Tris Borate EDTA
TBST	Tris-Buffered Saline with Tween
TE	Tris-EDTA
Temed	Tetramethylethylenediamine
WT	Wild-Type

Chapter 1:

Introduction

1.1 Objectives

The aim of this thesis is to identify novel causes of inherited heart diseases, and assess the utility of 'next-generation' sequencing strategies for use in clinical diagnostic practice.

1.2 Overview

Following on from the sequencing of the human genome at the beginning of the 21st century genetic sequencing technology has advanced apace. 'Next-generation' sequencing methods have the capability to sequence DNA on a much larger scale, in less time, and at lower cost than previously possible (Dewey et al., 2012). This has facilitated new testing methods into clinical diagnostic testing (Meder et al., 2011), and has led to new capabilities for identifying the causes of Mendelian disorders (Ku et al., 2011).

In the clinical diagnostic setting the ability to sequence more genes at a relatively low cost, particularly for genetically heterogeneous disorders such as inherited heart diseases, is highly attractive. Multi-gene tests utilising next-generation sequencing technology have been shown to provide interpretable and effective clinical diagnostic testing (Meder et al., 2011). Although clinically targeted multi-gene tests are being introduced for diagnostic testing, they are also potentially effective focussed research tools. Chapter 2 describes the use of data, from the introduction of multi-gene NGS testing to clinical practice, to study variants in the mitochondrial genome in patients with hypertrophic cardiomyopathy (HCM).

NGS techniques also allow the sequencing of the majority of genes which code for proteins (exome sequencing), and indeed complete genomes (whole genome sequencing), at a cost and

timescale that makes them practically feasible to use in research, and clinical, testing. Although these tests are only starting to be introduced into clinical diagnostic laboratories, predominantly in the USA, in a clinical research setting they have already allowed the identification of a number of novel causes of Mendelian disorders (Gilissen et al., 2011). In Chapters 3 and 4 the strategy, techniques, and results of whole genome testing, in addition to complementary genetic techniques, to identify the causes of five different autosomal dominantly inherited heart diseases of unknown genetic aetiology is presented. As well as aiming to identify novel causative gene variants for these conditions, this allows an initial interpretation of the utility, and practical issues, of exome or genome sequencing for clinical diagnosis.

1.3 Background

1.3.1 The Study of Mendelian Disorders

Although Mendelian disorders are often individually rare, as a group they are far from infrequent, with a prevalence of around 1% at birth (World Health Organisation data). They are also a significant cause of morbidity and mortality, for example causing approximately 6% of paediatric intensive care hospital admissions (FitzPatrick et al., 1991). An understanding of the underlying genetic cause of a disorder can have a significant clinical impact in terms of diagnosis and genetic counselling of individuals and their families (van Engelen et al., 2013). Increasingly the molecular understanding of Mendelian disorders is also starting to guide therapeutic strategy and development for them (Detaint et al., 2010, Kapsa et al., 2003). Furthermore, these discoveries have been instrumental in the understanding of genetics and molecular biology in health and multifactorial disease (Peltonen et al., 2006), and can therefore have an impact well beyond those with a particular rare disorder.

1.3.2 Inherited Heart Diseases

The term 'inherited heart diseases' can be used to refer to wide variety of familial disorders affecting the heart, including congenital heart disease, inherited disorders affecting the heart or cardiovascular system, cardiac muscle disorders, electrophysiological and/or conduction disorders, and even disorders which may secondarily affect the heart, such as familial hypercholesterolaemia. In this study we focus solely on highly penetrant inherited disorders of cardiac muscle, 'cardiomyopathies', and/or electrophysiological abnormalities.

Inherited cardiomyopathies and electrophysiological disorders are an important cause of morbidity and mortality, and include some of the more common Mendelian disorders. Hypertrophic cardiomyopathy (HCM), for example, affects about 1 in 500 people, and is the leading cause of sudden death in adolescents and young adults (Maron, 2009). Some of the cardiomyopathies and electrophysiological disorders are phenotypically distinct and genetically reasonably well understood, such as HCM and long QT syndrome (LQTS), but many of the others are clinically and genetically more heterogeneous. By the detailed clinical and genetic study of patients and families with some of these disorders, where the genetic aetiology is unknown, it is hoped that novel disease causing genes will be identified that should provide clinical utility for the families concerned, as well as providing further genetic insights into some of these conditions.

1.3.3 The Historical Basis of Current Genetic Understanding and Technologies

The origins of modern genetics lie in the 19th century, with the work of Gregor Mendel and his studies on the inheritance patterns which now bear his name. In the 20th century this work was developed and applied to the genetics of populations and evolution, along with the initial understanding of the role and structure of chromosomes. Key discoveries leading to the modern molecular genetic age occurred. First deoxyribonucleic acid (DNA) was identified as the genetic 'blueprint' by the Avery-MacLeod-McCarty experiment (Avery, 1944), and then

Watson and Crick's discovery of the structure of DNA in 1953 provided insight in to the nature of the genetic code (Watson, 1953). These discoveries formed the foundation for the advances in genetic understanding, and the development of the techniques for identifying, interrogating, and manipulating genes, that followed.

Key to the understanding of the nature and function of genes has been the ability to sequence the genetic code. One of the important developments in making the sequencing of genes practical, in terms of cost, timescale, safety and reliability, was the development of Sanger sequencing (Sanger, 1977). This utilised the incorporation of chain-terminating di-deoxynucleotides, resulting in DNA fragments which could be separated by size to determine the DNA sequence. Since its inception Sanger sequencing has been the main method used for sequencing DNA. Improvements have been made to the initial method to make it more efficient, particularly the incorporation of fluorescence labelling and automation (Smith, 1986). Even with recent advances in alternative techniques Sanger sequencing is still widely employed, and has significant advantages when the scale of sequencing is small and/or quality and read length are important.

1.3.4 Developments in Gene Identification Techniques

Although causative genes have been implicated in over 3500 Mendelian disorders approximately as many phenotypes again, suspected or known to be Mendelian, do not have an identified cause (OMIM, March 2013). Even with the possibility that some of these phenotypes may not represent distinct Mendelian disorders, for example they may be overlapping descriptions of the same condition, there remains a significant number of phenotypes without an identified cause. Traditionally the main techniques used in the identification of the causes of dominantly inherited single gene disorders have been positional cloning, linkage analysis and/or candidate gene sequencing strategies. However, all of these techniques have limitations, and are unable to detect all forms of genomic variation.

Functional cloning

Early gene identification relied upon 'functional cloning', isolating and identifying the relevant protein and then determining the gene sequence and location from this. However, this information was usually unavailable, therefore this approach was restricted to a handful of genes relevant to human disease where the biochemical pathway was understood (Robson et al., 1982).

Linkage analysis and positional cloning

More success was achieved by identifying genes based on their chromosomal location, positional cloning. The earliest gene identified in this manner was that for X-linked chronic granulomatous disease (Royer-Pokora et al., 1986). This was followed by the identification of the genetic causes of several other disorders including cystic fibrosis, Duchenne muscular dystrophy, Huntington's disease, and certain cancer predisposition syndromes. However, the techniques used were incredibly laborious, such that by 1995 only forty-two Mendelian disease genes had been identified by this method (Collins, 1995). Chromosomal locations were typically identified based on linkage analysis in one or more families with a particular disorder (Gusella et al., 1983). Chromosomal regions of interest for were also identified, or refined, by the recognition of chromosomal abnormalities in patients with a particular phenotype (Collins, 1995).

This strategy became increasingly effective as understanding of variation within the genome between individuals improved. The recognition of polymorphic endonuclease restriction sites, microsatellite sequences, and single nucleotide polymorphisms provided better markers for linkage analysis. In addition the chromosomal location of more genes became established, making identification of 'candidate' genes within a linked chromosomal region possible (Collins, 1995). The main difficulty of positional cloning was the requirement to have a

sufficient number of informative DNA samples, or relevant chromosomal abnormality, to adequately narrow down the implicated chromosomal region for gene identification.

Candidate gene analysis

The sequencing of candidate genes, proposed based on prior knowledge regarding a disorder or the relationship to a known genetic cause, has also identified a number of Mendelian diseases. The main limitation of this strategy has been the requirement for prior information to identify a suitable candidate. This prevents the identification of genes causing a phenotype by a novel or unexpected mechanism. An assumption towards pathogenicity of variants identified in this manner, often with limited supporting segregation data within families and/or limited evidence of *in situ* functional effect, has also occurred. This has caused the mis-assignment of certain genes or variants as being pathogenic (Watkins, 2013, Andreasen et al., 2013).

1.4 Next Generation Sequencing

Following on from the initial human genome sequencing projects (see section 1.5.1, page 13) in the last decade technologies have been developed capable of sequencing large amounts of DNA simultaneously, so called 'next-generation' or massively parallel sequencing. The increased sequencing capability of NGS has rapidly reduced the cost of sequencing per base (see Figure 1-1). Improvements in data output have continued to improve rapidly since the inception of these technologies, outpacing Moore's law with an almost thousand-fold increase in data output achievable in 2011 as compared with 2007 (Dewey et al., 2012). This has facilitated the ability to sequence large quantities of DNA sequence, including whole human genomes, in the research and clinical diagnostic settings (Schuler et al., 2013).

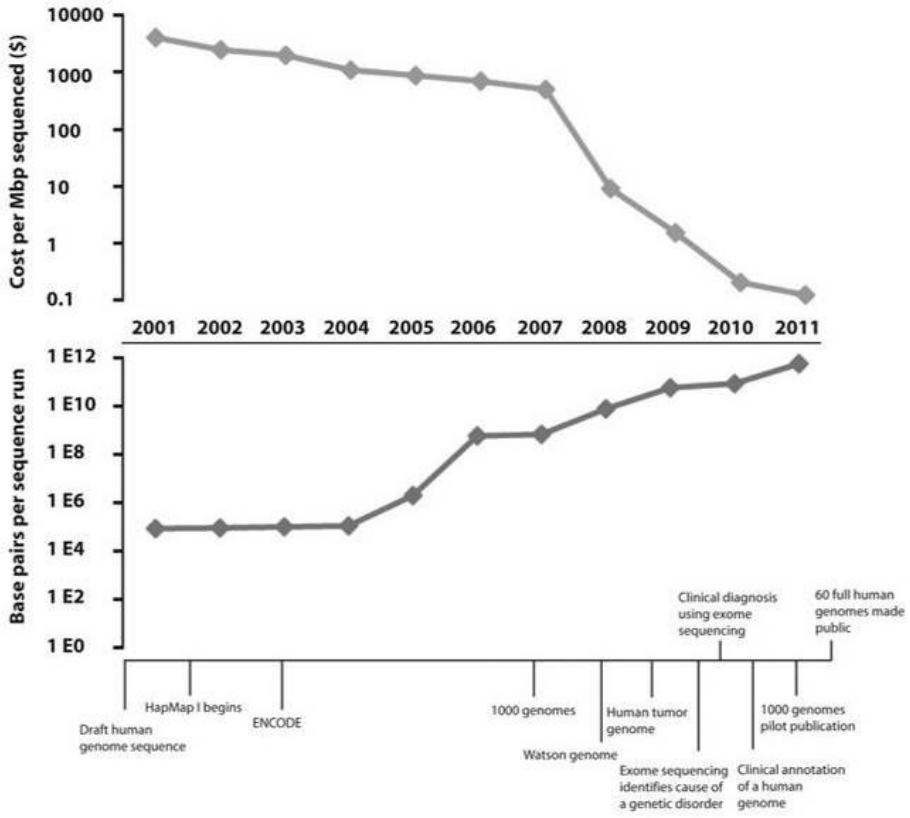


Figure 1-1: Timescale of cost and output changes since the Human Genome Project (Dewey et al., 2012). This shows the rapid reduction in the cost of sequencing, alongside the increased sequencing capabilities per run. Some of the key time points in genome sequencing are shown along the time bar.

1.4.1 Next Generation Sequencing Platforms

There are a number of commercially available 'next-generation sequencing' (NGS) platforms, which are capable of sequencing an entire human genome in less than a week. Different providers use different techniques, with advantages and disadvantages when compared to each other. All are capable of performing simultaneous sequencing reactions, producing thousands to billions of sequencing reads concurrently. The main providers are Roche®-454 Life Sciences®, SOLiD®, and Illumina®/Solexa®.

In this study Illumina® technology has been used for all of the whole genome and exome sequencing (summarised below and discussed in more detail in section 3.3.2, page 52), and Roche-454 Life Sciences technologies have been used in the multi-gene cardiomyopathy panel testing analysed in chapter 2. The Illumina method has its foundation in Sanger sequencing, however reversible dye terminators and immobilised cluster generation allows many reads to

be sequenced in an automated process (Figure 1-2). Billions of relatively short sequence reads are undertaken essentially simultaneously, massively increasing the amount and speed of data production while lowering the cost to the order of $1 per 10 million base pairs read (Illumina data). Roche-454 Life Sciences amplifies DNA attached to individual beads suspended within an emulsion droplet. A pyrosequencing technique is then used, with incorporation of nucleotides in the presence of a luciferase causing the emission of light which can be imaged and interpreted.

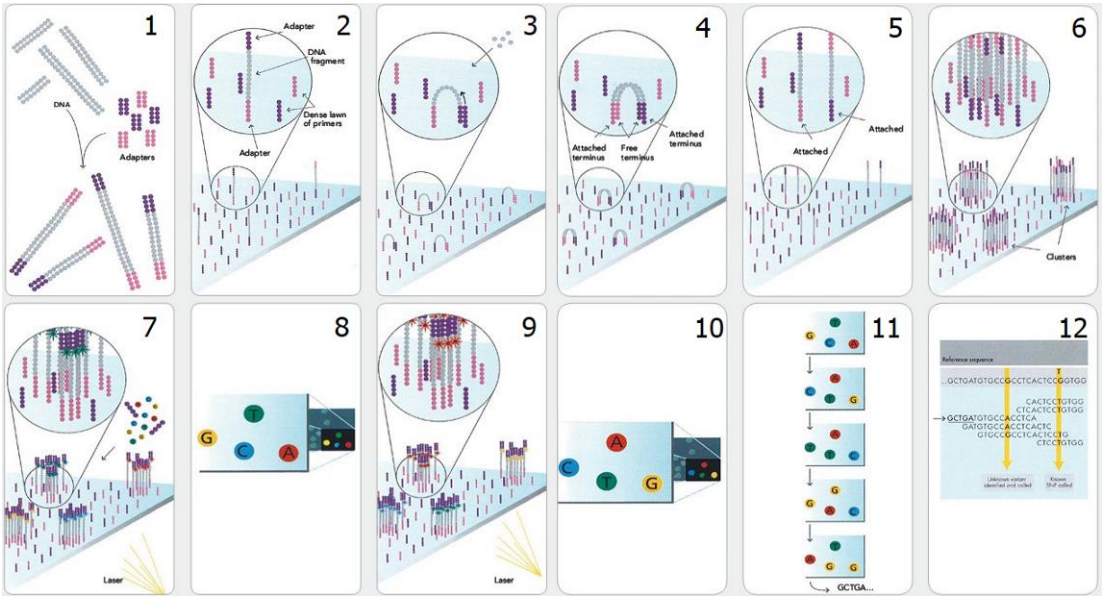


Figure 1-2: Overview of Illumina/Solexa whole genome sequencing:

1) Genomic DNA is fragmented to form a DNA library and small engineered DNA sequences are attached to each end of these fragments.
2) The engineered adaptors ligate to complementary DNA sequences which are attached to a solid flow-cell matrix.
3-6) DNA sequences are then amplified in 'bridges' while attached to the matrix and formed into single stranded amplified DNA clusters.
7) Sequencing is performed by the incorporation of fluorescently labelled end-blocked nucleotides by a DNA polymerase.
8) These nucleotides are excited and the fluorescence imaged and recorded.
9-11) The fluorescent label and block are cleaved to allow incorporation of the next labelled nucleotide. Cycles of nucleotide inclusion, imaging and cleavage allows the determination of a nucleotide sequence, 'sequencing-by-synthesis'.
12) Images are then turned into sequence read data and aligned against a reference sequence.

Following on from these platforms '3rd generation' platforms have been developed by companies such as Oxford Nanopore®, Pacific Biosciences®, Helicos® and Ion Torrent®; but these are not yet in widespread use. This is due to issues particular to each technique and also because the original 'next generation' techniques have made significant improvements since

their inception. For example, with the Illumina method it has become possible to eliminate the polymerase chain reaction (PCR) step from the library preparation. Removing this step reduces artefacts introduced in amplification and improves coverage of AT- and GC-rich regions (Kozarewa et al., 2009).

1.4.2 Applications of Next Generation Sequencing

DNA sequencing by next-generation technologies has already been utilised in studies of 'normal' populations (Abecasis et al., 2010), complex traits (Biesecker, 2012), cancers (Maher, 2009), and learning difficulties/congenital abnormalities (de Ligt et al., 2012, Rauch et al., 2012), as well as Mendelian disorders (Ng et al., 2010b). In terms of sequencing in Mendelian disorders there are three main types of NGS used:

1) Whole genome sequencing (WGS) refers to sequencing of almost the entire genome of an individual of a species. Depending on the species studied there will still be a variable amount of genome sequence which is not covered by current WGS studies, for example due to difficulties with mapping of homologous regions.
2) Exome sequencing (ES) refers to sequencing of the predominantly protein coding sequence, the exons, and a varying degree of regulatory sequence, such as recognised promoters.
3) Targeted sequencing strategies covering a number of selected genes using the same sequencing technologies. There are many strategies and methods included in this category.

Exome sequencing has been used more commonly than whole genome sequencing in the study of Mendelian disease (Ku et al., 2011). This has largely been due to the practical considerations, being significantly cheaper and quicker to sequence. Although only comprising approximately 1% of the genome, the exome harbours over 85% of mutations identified causing Mendelian disease (Cooper et al., 1995). Therefore the compromise of the additional

sequence is only infrequently relevant in identifying a cause. However, whole genome sequencing does have some advantages. The main advantage in that if a causative gene variant is not identified within the exome it allows the opportunity to explore other genomic regions. Even if a causative gene is identified within the exome the study of other regions, for example genome-wide SNPs, may be useful in further study of a condition.

It is of note that even within the exome, whole genome sequencing and exome sequencing are not equivalent. Indeed different exome enrichment libraries are not equivalent (Clark et al., 2011b). Different variants can be identified within the exome depending on whether ES or WGS is used. Certain regions are poorly enriched by exome kits, and are therefore better covered by WGS, whereas other regions are better covered due to the enrichment, and are therefore more easily detected by ES. More false positive variant calls are produced by ES as compared with WGS (Clark et al., 2011b).

Targeted sequencing has been predominantly been used in clinical diagnostic testing, although the ability to test wider gene panels in clinical research has helped expand the clinical phenotype of a number of known genes (Lise et al., 2012). Targeted gene panels may also be used for high-throughput studies of candidate genes.

1.4.3 Advantages of Next-Generation Sequencing in Identifying the Causes of Mendelian Disorders

In more than a half of the approximately seven-thousand Mendelian disorders the cause is unknown (Ku et al., 2011). All of the methods previously used for gene identification in Mendelian disorders have significant drawbacks (previously described in section 1.3.4, page 4). Some of the problems of these techniques can be overcome by using ES/WGS. For example, linkage analysis/positional cloning is of limited power in lower penetrance disorders or when extensive multi-generation families are unavailable (Boehnke, 1994). NGS, with or without other techniques such as single nucleotide polymorphism (SNP) array to provide additional

linkage information, is more resilient to this. The main advantage is the production of a complete list of shared variants, rather than linkage regions in which most of the genes will have no variants of likely pathogenicity. Having variant data, rather than simply linkage data, also allows the clear identification of variants in known or candidate genes at an early point of analysis. Further to this variants can be prioritised with regards to likelihood of pathogenicity on the basis of their properties, such as predicted amino acid change and evolutionary conservation at that position or region. The strategies for prioritising variants in this manner are discussed further in section 1.5.4 (page 15).

1.4.4 Ethical Considerations

The ability of NGS techniques to produce such large amounts of sequencing data in an unbiased manner in an individual obviously has great value in research, and potentially clinical diagnostics, but has raised certain concerns in application. The main issue revolves around the appropriate consent, data sharing, and returning of results to patients, of sequencing information in genes other than those of primary interest. Those using NGS in clinical or research studies have generally each developed local policies, which has led to significant discrepancies in policies relating to these issues between different centres and studies (Jamal et al., 2013). This has resulted in the recent release of guidelines by organisations with activities in this area, such as the American College of Medical Genetics and Genomics (ACMG) and European Society of Human Genetics (ESHG) (Green et al., 2013, van El et al., 2013).

Particularly with clinical testing the main concern regards what should be analysed and reported that is unrelated to the indication for sequencing. Annotating and reporting all information, aside from the practical difficulties, risks importance being attributed to variants of no significance. If no additional information is reported then mutations potentially causative for preventable or treatable disorders would not be acted upon.

The ACMG guidelines (Green et al., 2013) recommend that variants felt highly likely to be pathogenic in a list of designated genes should be reported. This list consists of predominantly cancer predisposition and cardiovascular disorder genes. The considerations taken into account when determining the list included whether those tested may not have an overt clinical phenotype, whether it is confirmable by alternative tests, and whether it may have medical implications for those tested and their families. It is anticipated that approximately 1% of reports will include an 'incidental' finding, although this estimation is based on limited evidence. The recommendations explicitly state that patients should not be able to opt out of being informed of these 'incidental' findings.

These recommendations have not been without immediate criticism. The obligatory analysis of genes, where this would not necessarily occur, and particularly whatever the wishes of the patient, could be considered to be mandatory screening (PHG_Foundation, 2013). The appropriateness and effectiveness of screening asymptomatic individuals has previously been poorly supported, and is particularly controversial in children (Allyse and Michie, 2013). Aside from ethical concerns there are also practical concerns regarding the funding and availability of the additional counselling, analysis, and follow up of reporting these variants (Allyse and Michie, 2013).

The European Guidelines recognise that the generation of 'unsolicited' findings is a disadvantage of using NGS in many clinical settings (van El et al., 2013). Consideration of this possibility is encouraged to form part of pre-test counselling and consent for these techniques. The classification of findings into different categories, with the option of different levels of information being made available, is proposed. Although these guidelines are designed for clinical testing it is reasonable to contemplate many of the same issues in the context of research testing. Consideration of these in relation to this study is discussed in section 3.2.2 (page 49).

1.5 Genomic Variation

The main difficulty in identifying the cause of a Mendelian disorder from WGS/ES data is that there is a high level of 'natural' variation within genes. Across an individual genome millions of single nucleotide, and copy number, variants are identified as compared with the reference genome (Ng et al., 2008). Although genetic variation has long been recognised, it is only with the sequencing of the first human genomes that the extent of this variability has been appreciated (Lander et al., 2001, Venter et al., 2001).

1.5.1 The Human Genome Project

Between 1990 and 2003 the Human Genome Project, an international collaborative project coordinated by the US department of Energy and the National Institutes of Health, and a privately-backed project led by Dr Craig Venter, sequenced the majority of the three billion base pairs of the human genome (Lander et al., 2001, Venter et al., 2001). This has continued to be further refined since then, and now provides an accurate reference sequence of almost the entire genome.

The initial draft human genome sequence was achieved using modified Sanger sequencing methods. Both the International Human Genome Sequencing Consortium (IHGSC) and the Venter group used types of 'shotgun sequencing' methods, although with slight differences due to different strategies for managing the repetitive nature of the human genome. The IHGSC group used 'hierarchical' shot-gun sequencing, whereas the Venter group used predominantly 'whole-genome' shot-gun sequencing. Hierarchical sequencing involves the generation of libraries by incorporating approximately 100kb DNA fragments into large BAC (bacterial artificial chromosome) vectors organised into a physical map, fragments are then amplified in bacterial culture and sheared into 2-3kb fragments. Whole-genome shot-gun sequencing avoids this BAC generation step, the drawback of which is that this then requires additional computational requirements in assembly generation to map reads correctly.

1.5.2 The Reference Genome

These initial sequences had a high level of concordance but have been improved by combination of the data produced, and subsequent sequencing of poorly covered areas, to give a highly accurate human genomic reference sequence (Istrail et al., 2004). However, it must be borne in mind that this 'reference genome' is based on only a few completed *de novo* long-read genome sequences. Although it covers the majority of the genome with high fidelity it provides very limited information on normal variation within the genome, and indeed contains a small number of disease alleles, such as for Factor V Leiden, and a number of 'risk' alleles for complex traits, such as diabetes (Dewey et al., 2012).

1.5.3 Genetic Variability Identified by Next Generation Sequencing

WGS will commonly identify three to five million single nucleotide variants and small insertions or deletions (indels), as compared to the reference genome, in an individual (Biesecker, 2012). The number of reported variants in an exome depends more on the technology, variant calling strategy, and analysis used, but is in the region of thirty to fifty thousand variants (Biesecker, 2012). Although common polymorphisms account for a proportion of these, there is a relatively higher proportion of rare variants than may have been anticipated. Further to this, a number of these variants would be predicted to be of functional significance.

The number of potential loss-of-function (LOF) variants in an individual exome is in the order of twenty to forty nonsense variants (two to seven of which are homozygous), ten to fifteen splice variants (one to three homozygous), about forty indels and frameshift variants (about ten of which are homozygous), and about twenty-five large deletions (about five of which are homozygous) (MacArthur et al., 2012). This is generally lower than estimates from initial results (Abecasis et al., 2010), likely due to the detailed verification of variants identified and clarification of areas of the reference sequence. The majority of these potential LOF variants are rare, and had not previously been reported in dbSNP (MacArthur et al., 2012).

However, there are several reasons why apparently LOF variants may be of no consequence:

- They may be sequencing errors or errors within the reference genome (MacArthur and Tyler-Smith, 2010).
- Gene expression may not be significantly affected by the variant. In one study only a minority of LOF variants, sixteen percent overall and twenty-five percent of those that were predicted to cause nonsense-mediated decay, led to a reduction in gene expression as determined by RNA studies of lymphoblastoid cell lines (MacArthur et al., 2012).
- Protein products may not be significantly altered, for example about one-third are only present in some of the transcripts of the gene (MacArthur et al., 2012).
- Variants may occur in genes of no discernable function. An increased frequency of LOF variants are seen in genes which have effectively been superseded, such as the *VWDE* gene (Von Willebrand Factor D and EDF domains precursor) (MacArthur et al., 2012).
- The variant may alter a neutral phenotype, for example the O allele in the ABO blood group (MacArthur and Tyler-Smith, 2010).
- The variant may be a recessive disease causing mutation but be in heterozygous 'carrier' state. Some of these variants have been identified as being beneficial in the carrier state (Allison, 1954, Gabriel et al., 1994). Indeed certain other LOF variants, not associated with recessive disease, have also been shown to be beneficial (Saleh et al., 2004, Tournamille et al., 1995).

1.5.4 Identification of Pathogenic Mutations

To assist with the filtering of the large number of variants produced by exome or genome sequencing a number of strategies can be used. There are several prediction tools for determining scores of likely pathogenicity, incorporating one or more of the factors discussed below. These have individual strengths but also have marked limitations, and not infrequently

show discrepancies between them (Li et al., 2013). Therefore manual analysis incorporating a variety of sources remains the gold standard at present.

Genomic location

When whole genome data is available it is common to focus the initial search for pathogenic variants in Mendelian disease on the exome, as the majority of causative variants will be identified here (Cooper et al., 1995). Validation of variants outside the exome, including regulatory regions, is also much more difficult due to the more limited experience in understanding the mechanisms of effect of such mutations. This is starting to change as more detailed studies of non-protein coding sequence are undertaken, improving the understanding of possible implications of variation in these regions (Ward and Kellis, 2012).

Predicted function

The vast majority of variants causing highly penetrant Mendelian disorders are LOF alleles (Stenson et al., 2009), i.e. they alter the transcript in such a way as to reduce, remove, or alter its function. Therefore focus is often given to non-synonymous, nonsense or frameshift mutations, along with indels and larger copy number variants (CNVs). As discussed above, identification of LOF variants should not immediately lead to an assumption of significance when identified in NGS data, as most LOF variants identified in such widespread testing will be benign or unrelated to the phenotype in question (MacArthur and Tyler-Smith, 2010). The reverse situation, where an apparently benign (e.g. synonymous) variant causes disease, is less frequent but has also been observed (Buske et al., 2013).

In the case of missense variants the nature of the amino acid change can be helpful in assessing likely pathogenicity. Scores such as SIFT and Polyphen have been developed to predict the tolerance of an amino acid substitution (Kumar et al., 2009). The significance of a specific amino acid change is difficult to predict, and is highly situation dependent, making these tools imprecise for the initial exclusion of variants.

Evolutionary conservation

Conservation through the constraints of evolutionary selection of a particular nucleotide, or amino acid residue, is recognised as important factor in determining the significance of sequence variation (Sidow, 2002). Various scoring systems have been devised to assist with this, such as PhyloP (Pollard et al., 2010). Exome sequencing has confirmed that rare variants are usually highly conserved, and therefore low levels of conservation should raise doubt as to the pathogenicity of a variant. Although conservation scores have been proposed to be useful in identifying significant variants (Cooper et al., 2010) many benign changes will also be attributed high conservation scores, therefore this has limitations.

Population data

Variants that cause high penetrance Mendelian disorders are rarely identified, if at all, in population cohorts. There are several databases of population genetic variation data, including the NCBI database of single nucleotide polymorphisms (dbSNP) (Sherry et al., 2001), 1000 Genome Project data (1000G) (Abecasis et al., 2010), and the NHLBI-GO exome variant server (EVS) (EVS, March 2013). A common filtering strategy in Mendelian disorders is to remove all polymorphisms with a minor allele frequency (MAF) of greater than 1% (Li et al., 2013).

These databases are able to assist with the interpretation of genetic variants identified in an individual, at least in terms of Mendelian disease, however they are not of sufficient size to exclude many rare variants. Another limitation is that they usually have minimal phenotypic information, if any, therefore it is difficult to exclude variants seen rarely as possible causes of Mendelian disease. Disease causing variants may also be included in such databases due to sequencing error. As population genetic variation databases increase in size they will become even more effective in assisting the determination of likely pathogenicity in rare disorders. Genetic population data derived from bio-banks, where phenotypic data is usually significantly more detailed, will be particularly informative in this regard.

Project specific data

As well as more general predictions about the variant concerned, specific knowledge or data can also be incorporated when filtering genomic variant data:

- Knowledge of inheritance pattern and segregation of variants in other family members with a condition, or variants occurring in the same gene in other patients with the particular phenotype, are usually strong filtering mechanisms.
- Predictions as to the likely nature or function of the gene, for example predicting a membrane ion channel to be the cause of an epilepsy phenotype, can also be utilised. This risks turning the data into a candidate gene study, but it may have a role in the very initial stages of variant analysis.
- Tissue-specific gene expression profiles should be useful in filtering data, for example in the case of inherited heart disease limiting genes to those which are expressed in the heart. The limitation of this as a strategy is the incomplete and inconsistent information on expression profiles in these databases (Liang et al., 2006).

1.6 Genetics of Inherited Heart Disorders

1.6.1 Current Genetic Understanding

The knowledge of the familial nature of certain cardiac disorders is longstanding (Braunwald et al., 1964), but the genetic era of inherited heart disease really only started in the early 1990s with the identification of mutations in the β-myosin heavy chain, myosin binding protein C, troponin T, and α-tropomyosin genes as causes of HCM (Geisterfer-Lowrance et al., 1990, Watkins et al., 1995a, Watkins et al., 1995b) and potassium and sodium channel subunit genes as causes of LQTS (Wang et al., 1995, Curran et al., 1995, Wang et al., 1996). These initial gene identification successes utilised linkage analysis in large pedigrees to identify the chromosomal locus for a disease gene, followed by sequencing of candidate genes identified in these regions. A number of other genes have since been implicated in these, and other, inherited

heart diseases. These phenotypes tend to be genetically heterogeneous, for example at least eleven genes have been identified mutations in which can cause HCM (Maron et al., 2012).

Current diagnostic genetic testing in some of these disorders, particularly those which are relatively common and have a highly penetrant discernable phenotype such as HCM and LQTS, have a high level of utility. Testing of all known causative genes in an individual with a clear phenotype and family history has a high chance of identifying a causative mutation, approximately 60% in HCM and 85% in LQTS (Ackerman et al., 2011). This is then useful in genetic counselling of the patient and family, and particularly in the case of LQTS, may have bearing on therapy (Ruan et al., 2008).

There are several possible explanations for the cause of the phenotype in the significant minority of patients with HCM or LQTS who have no known genetic cause. This may be due to other, as yet unidentified, highly penetrant disease alleles that are individually a rare cause of the phenotype. In others the aetiology may be genetic but due to lower penetrance alleles. A number may be phenocopies, i.e. caused by disorders which give a similar phenotype, whether that be genetic or not. In the case of HCM a similar pattern of left ventricular hypertrophy may be caused by non-genetic causes, such as hypertension, the 'athlete's heart', or heart valve disease. Some of these, such as hypertension, may cluster in families due to common genetic and lifestyle predisposition factors, and may therefore give a phenotype of familial HCM (Staessen et al., 2003). Other inherited conditions can also and can be difficult to differentiate from classical HCM, such as Fabry disease (Elliott et al., 2011).

Identifying the unrecognised causes of phenotypes such as HCM or LQTS, is important to allow the appropriate genetic counselling for these patients and their families. In addition the alternative underlying cause may be associated with differences in prognosis, complications, and potentially appropriate therapies. Further understanding of the aetiology of these conditions may also allow further insights into the molecular mechanisms of the disease

phenotype. In the majority of inherited heart disorders the genetic aetiology less well understood, and the clinical heterogeneity is often greater than in HCM or LQTS. Identification of causative genes in these disorders have even more potential to shed significant insights into disease biology.

NGS techniques have already been shown to be effective for identifying novel genetic causes of cardiac disorders (Norton et al., 2011, Haack et al., 2010, Gotz et al., 2011). One of these studies (Norton et al., 2011) used a similar strategy to the one employed in our family studies (described in chapters 3 and 4). They identified mutations in *BAG3*, which encodes a heat-shock protein chaperone, as a cause of DCM. Mutations in *BAG3* as a cause of DCM was then validated in a patient cohort and supportive evidence from a zebrafish model was acquired (Norton et al., 2011). This provides evidence that this strategy is capable of identifying novel genetic causes of inherited heart disease.

1.6.2 Clinical Diagnostic Testing

Targeted gene testing using NGS platforms has been introduced to diagnostic laboratories for a number of disorders. These tests are replacing Sanger sequencing methods as the initial testing in genetically heterogeneous disorders, due to their capability for sequencing multiple genes simultaneously and at a lower cost. They have been shown to be an effective way of providing clinical diagnostic testing in inherited heart disorders, particularly hypertrophic cardiomyopathy (Teekakirikul et al., 2013) and long QT syndrome (Ware et al., 2013).

In 2011 a multi-gene panel test for hypertrophic cardiomyopathy was developed for clinical diagnostic use, by collaboration between the Oxford NHS Molecular Genetics Laboratory and the Biomedical Research Centre Genomics Group at the University of Oxford. The initial panel developed incorporated fifteen 'genes', one of which was the thirty-seven genes of the mitochondrial genome. The other genes included were *MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2, GLA,* and *LAMP2*. These are all proven causes

of HCM, or disorders with a hypertrophy phenotype difficult to distinguish from HCM. Although this gene panel was designed for clinical diagnosis, it also gave the opportunity to address certain clinical research questions. Even though specific mitochondrial mutations have been identified as a cause of cardiomyopathy the frequency, and nature, of mitochondrial variation in HCM has not been well studied (Bates et al., 2012). It is this question that is addressed in the subsequent chapter.

Chapter 2:

Mitochondrial Genomic Variation in Hypertrophic Cardiomyopathy

2.1 Background

Mitochondrial mutations have been identified as a cause of left ventricular hypertrophy as part of multisystem disease (Anan et al., 1995). It has generally been assumed that due to the presence of extra-cardiac features, and often differences in the cardiac phenotype in these disorders, patients with mitochondrial disorders can be distinguished clinically (Keren et al., 2008). However, the frequency and nature of mitochondrial DNA (mtDNA) variation in patient cohorts with isolated HCM has not been systematically studied (Bates et al., 2012). To address this the complete mitochondrial genome was sequenced, as part of a multi-gene next generation sequencing HCM testing panel that was being validated for use as a clinical diagnostic test. Therefore the main aims of studying the mitochondrial genome within this was to determine the frequency and nature of pathogenic mtDNA mutations causing an HCM phenotype, as well as whether it could be practically sequenced and analysed to provide clinically relevant results.

2.1.1 Hypertrophic Cardiomyopathy

Hypertrophic cardiomyopathy (HCM) is one of the commonest inherited cardiac disorders, affecting up to 1 in 500 people (Maron et al., 1995). It is characterised by left ventricular hypertrophy in the absence of pressure overload or infiltrative processes (Gersh et al., 2011). HCM commonly presents during puberty or early adulthood, but may arise at any time from infancy to late-adulthood (Maron and Maron, 2013). Although the clinical severity is very variable, a significant minority develop symptoms due to left ventricular

outflow obstruction, arrhythmias, or systolic dysfunction. Overall mortality is about 1% per year (Maron et al., 2000). The presence of risk factors in an individual (such as syncope, septal thickness >30 mm, family history of sudden death, blunted blood pressure response on exercise, and/or evidence of non-sustained ventricular tachycardia) can identify a higher risk of ventricular arrhythmias and sudden cardiac death (SCD) (Gersh et al., 2011).

HCM has been recognised as a familial disease for over 50 years (Hollman et al., 1960), usually exhibiting autosomal dominant inheritance. In the 1990s mutation in genes encoding proteins of the cardiac sarcomere were identified as the predominant cause of HCM. Mutations in the β-myosin heavy chain (*MYH7*) and cardiac myosin binding protein C (*MYBPC3*) are the most common causes, but at least eleven sarcomere, or associated, genes have been reported as being causative for HCM (Figure 2-1) (Maron et al., 2012). Testing of the main causative sarcomere genes will identify a pathogenic mutation in 50-65% of familial cases, and approximately 30-60% of apparently sporadic cases (Richard et al., 2003, Van Driest et al., 2005). There therefore remains a significant proportion of even familial disease which remains unaccounted for.

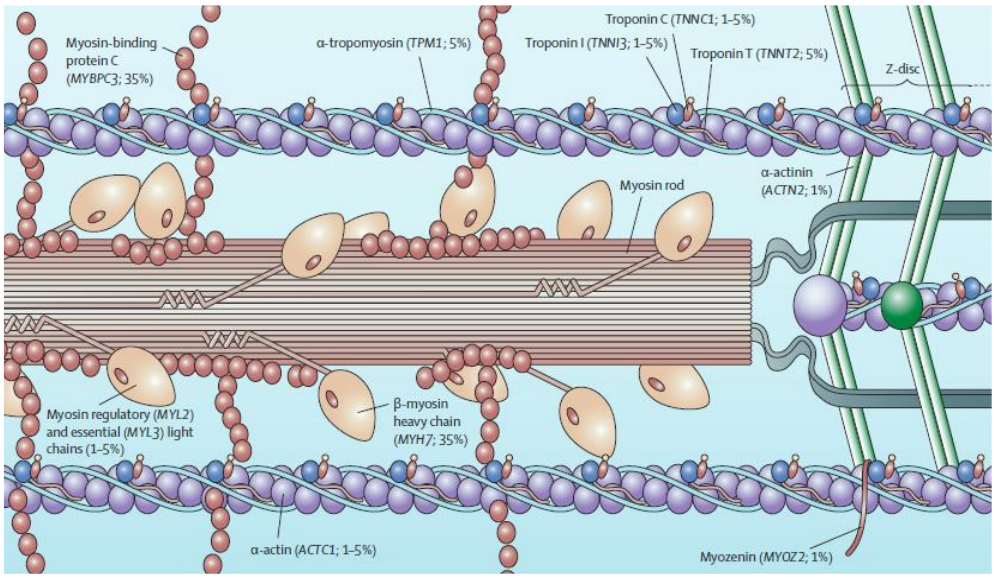


Figure 2-1: Representation of the location of cardiac sarcomere proteins in which mutations cause HCM (gene, approximate frequency) (Maron and Maron, 2013)

A handful of other genes have also been suggested as being causative for HCM in a small number of patients, such as α-myosin heavy chain (*MYH6*) (Carniel et al., 2005), muscle LIM protein (*CSRP3*), titin (*TTN*), telethonin (*TCAP)* (Bos et al., 2006), and those encoding various proteins involved in calcium-handling (Landstrom and Ackerman, 2012). However, the evidence behind the pathogenicity of some of these is less definitive. There are also a number of conditions which can cause an HCM phenotype, usually as part of a broader cardiac and/or multisystem phenotype, such as AMPK disease/*PRKAG2* (Blair et al., 2001), Friedreich ataxia (Boehm et al., 1970), fatty acid oxidation defects (Spiekerkoetter, 2010), Noonan syndrome and related disorders (Tanimura et al., 1977), Danon disease/*LAMP2* (Charron et al., 2004), and Fabry disease (Colucci et al., 1982).

Although previously assumed to be due to compensatory hypertrophy secondary to reduced myocardial contraction, HCM is increasingly now being seen as a more complex disorder of cellular energy (Ashrafian et al., 2003). This impression of HCM as an energy disorder is supported by studies demonstrating that energy deficiency, and altered calcium handling, are key to the cellular pathogenesis of sarcomeric HCM (Ashrafian et al., 2011). In addition most of the HCM phenocopy disorders mentioned, such as AMPK disease (Blair et al., 2001), have clear links to cellular energy homeostasis. As mitochondria are the key 'energy' producing organelle they are therefore a natural candidate for further study as a potential cause of HCM.

2.1.2 Mitochondria

Function

The main role of mitochondria is to produce energy in the form of ATP by oxidative phosphorylation (OXPHOS). The main energy source for cardiac muscle is derived from β-oxidation of fatty acids in the mitochondrial matrix, through the intermediate production of acetyl-CoA (Neely and Morgan, 1974). Acetyl-CoA is used by the tricarboxylic acid (TCA)

cycle, also known as the citric acid or Krebs cycle, to reduce nicotinamide adenine dinucleotide (NAD^+). NADH is then used by the mitochondrial respiratory/electron transport chain to generate a proton gradient across the inner mitochondrial membrane, which is used to generate ATP (Figure 2-2) (Berg et al., 2002).

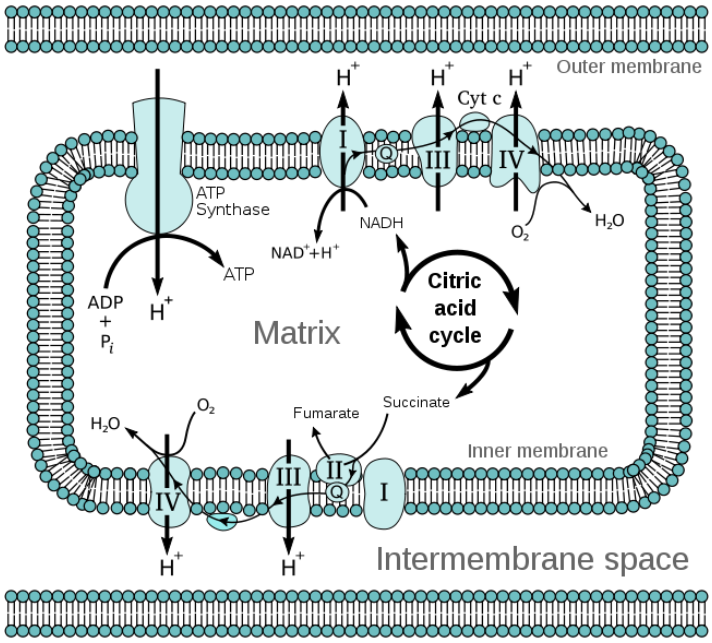


Figure 2-2: Schematic diagram of oxidative phosphorylation in the inner mitochondrial membrane. NADH and succinate produced in the citric acid/tricarboxylic acid cycle are oxidised in the mitochondrial respiratory electron transport chain to generate a proton gradient. This is then used to generate ATP.

Mitochondrial genetics

Although the majority of cellular DNA is within the nucleus there is also DNA within the mitochondria. The mitochondrial genome encodes for some key mitochondrial proteins, although the vast majority of proteins important for mitochondrial function are still encoded by the nuclear genome (Zhang et al., 2008). The thirty-seven mitochondrial DNA (mtDNA) genes encode for thirteen subunits of the electron transfer chain, two ribosomal ribonucleic acids (rRNAs) and twenty-two transfer ribosomal ribonucleic acids (tRNAs) (Figure 2-3) (Anderson et al., 1981, Andrews et al., 1999).

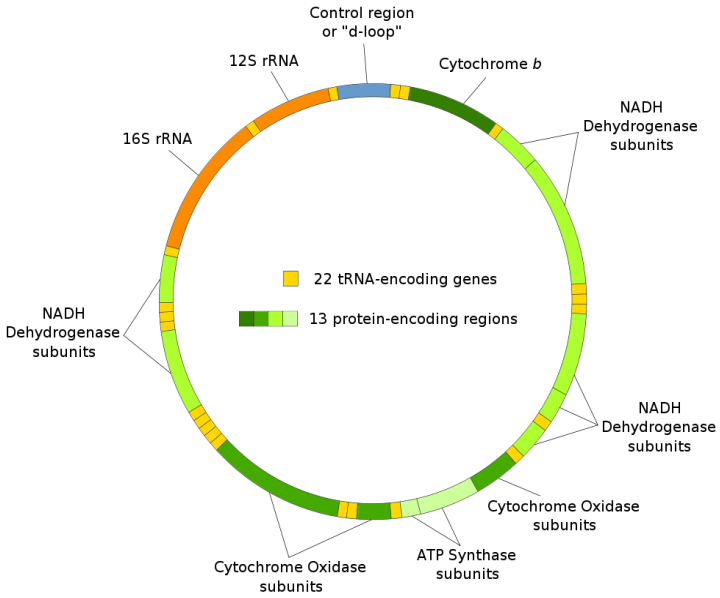


Figure 2-3: A schematic diagram of the genes of the mitochondrial genome

There are several features of mtDNA which give it important distinctions from nuclear DNA. The mitochondrial genome is in multiple copies per cell, with two to eight copies per nucleoid (Legros et al., 2004), one to ten nucleoids per mitochondrion (Satoh and Kuroiwa, 1991) and up to several thousand mitochondria per cell (Berg et al., 2002). An individual cardiomyocyte may therefore typically have between six and ten thousand copies of the mtDNA (Miller et al., 2003a). Genetic variants can therefore occur in a proportion of copies within a cell, when this occurs it is termed heteroplasmy. The proportion of mitochondria with a specific heteroplasmic variant will vary between different cells and tissues, with mutations of mtDNA typically only manifesting a biochemical and clinical phenotype when they reach a certain 'threshold' level in a tissue (Zifa et al., 2007).

Critically mtDNA also exhibits strictly maternal inheritance, with only the ovum retaining mitochondria when the gametes fuse to initiate embryogenesis. In oocyte generation the number of mitochondria numbers are dramatically reduced, before being re-amplified in embryonic development. This can cause a significant shift of variant load through generations, and in different individuals, which may significantly affect the clinical phenotype (Cree et al., 2008). There is also an increased mutation rate in mtDNA, as

compared with nuclear DNA, due to the local generation of oxygen free radicals and lack of a DNA repair system. The potential for this to be of consequence is increased by the relatively low amount of non-coding DNA, and the absence of introns within genes. As mtDNA continues to replicate through life the proportions of variants in a cell can continue to increase, even in post mitotic cells such as cardiomyocytes, which can further affect the load of a mtDNA variant (Bates et al., 2012).

Clinical phenotypes of mitochondrial disorders

Mitochondrial disorders usually display the clinical consequences of pathology of those organs with highest energy requirements (McFarland et al., 2010). Heart muscle is the most energy demanding tissue, with the highest volume density of mitochondria (Goffart et al., 2004). It is therefore unsurprising that it is relatively commonly affected in mitochondrial disorders, with cardiomyopathy and conduction abnormalities being the most common manifestations (Bates et al., 2012). Cardiac involvement is usually seen as part of multisystem disease (Tuppen et al., 2010), but isolated maternally-inherited dilated cardiomyopathy (Arbustini et al., 1998) and left ventricular hypertrophy (Taylor et al., 2003) have been rarely reported.

2.2 Methods

2.2.1 Patients

Lymphocyte-derived DNA was obtained from ninety-one patients who attended the Oxford Inherited Cardiac Conditions Clinic. All patients had been diagnosed with HCM according to standard diagnostic criteria (Gersh et al., 2011). Patients were only excluded if a pathogenic mutation in *MYH7*, *MYBPC3*, *TNNI3*, and *TNNT2* had been identified, or if DNA was unavailable. No selection was done based on family history, or particular clinical features, to more accurately represent testing in a clinical setting. Selected additional samples were

also included in the initial sequencing. Most of these were DNA samples from patients with non-HCM phenotypes but which were of clinical or research interest.

2.2.2 Sequencing

This cohort of patients had full sequencing of the coding regions of the mitochondrial genome as part of a next-generation sequencing multi-gene analysis panel. The other genes sequenced as part of this were *MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2* and *GLA*. The design specifications and results of these other genes will not be discussed here, as they were separately assessed and analysed by members of the Oxford NHS Molecular Genetics Laboratory to determine the utility of this test more generally for use in clinical practice.

Sequencing of this multi-gene panel used the Roche-454 GS FLX sequencing platform. The mechanistic basis of this platform is that fragmented DNA has specialised adapters added that capture individual fragments onto their own beads. These fragments are then clonally amplified within an emulsion. Sequencing occurs in open wells densely packed on to a fibre-optic slide. Sequencing reagents are repetitively added to, and removed from, the wells, with each round incorporating a nucleotide with the release of pyrophosphate and generation of photons (pyrosequencing). These are nucleotide-specific and are recorded to allow identification of each step of the strand synthesis simultaneously in each well (full details (Margulies et al., 2005). This technology has the advantage of longer read lengths than many other NGS platforms can achieve, up to five-hundred base pairs at the time and now potentially over a thousand base pairs, with the main drawback being that it is relatively expensive.

Library preparation, sequencing, and variant calling was performed by Dr Kulvinder Kaur, Biomedical Research Centre Genomics Group, Wellcome Trust Centre for Human Genetics, University of Oxford. Amplification of the mitochondrial genome using long-range PCR was

performed. The two pairs of overlapping mitochondrial-specific primers used were designed using Primer 3 online design tool (Untergasser et al., 2012). One set of primers, GAGGCGGGCATAACACAGCAAGACGA and GGCCCGGAGCGAGGAGAGTAGCAC, amplified a product of 14kb, with the other set, CCGCTTCTGGCCACAGCACTTAAACACATC and GGAGGATGGTGGTCAAGGGACCCCTATCTG, amplifying a 16kb product, both incorporating all coding regions of the mitochondrial genome.

Amplified DNA fragments were quantified using Quant-iT™ PicoGreen® dsDNA reagent, a fluorescent nucleic acid stain for quantitating double-stranded DNA in solution. This quantified value was used to equimolar pool PCR products for each sample. Each of the samples was sheared using a Covaris® ultrasonicator (S-Series) according to the manufacturer's protocol. Libraries were prepared using the GS FLX Titanium Rapid Library Kit™. This involves the addition of sequencing specific adaptors and molecular identifiers (MIDs). The quantity and quality of DNA libraries were assessed using a high sensitivity chip on a Bioanalyzer (Agilent®). Libraries were pooled in to two groups of ten samples per sequencing run. Emulsion based clonal amplification of the pooled DNA libraries was carried out using the GS FLX Titanium MV emPCR (Lib L) kit (Roche®). Beads containing the clonally amplified DNA libraries were sequenced on a Roche-454 GS FLX instrument using a two large region PicoTiterPlate (with ten samples per region) according to manufacturer's guidelines (Margulies et al., 2005). This platform has previously been validated for whole mitochondrial genome analysis in suspected mitochondrial cardiomyopathy, both HCM and DCM, by another group in a small selected cohort (Zaragoza et al., 2010). Each position was covered to a depth of at least 100x. This arbitrary cut-off was used to ensure a reasonable detection of even modest levels of heteroplasmic variants.

2.2.3 Bioinformatic Analysis of MtDNA Variants

Initial assembly of the 454 reads was performed using the Roche-454 GS Mapper Software™. Reads were de-multiplexed and aligned to reference genomic sequence. The mitochondrial sequence reference used was the Revised Cambridge Reference Sequence (GenBank accession number NC_012920; (Andrews et al., 1999)). 'High-confidence differences' were reported by the software, with a variant being reported when identified in both forward and reverse reads at high quality.

Identified mtDNA variants were compared with a compiled list of those variants reported to have caused a phenotype including cardiomyopathy (Table 2-1), the MITOMAP database (Ruiz-Pesini et al., 2007), the Human Mitochondrial Database (HmtDB) (Rubino et al., 2012), the Human Mitochondrial Genome Polymorphism Database (mtSNP) (Tanaka et al., 2004), PhyloTree (van Oven and Kayser, 2009), large population mitochondrial variant data sets (Pereira et al., 2009, Ingman and Gyllensten, 2006), and Genbank accession file information (www.ncbi.nlm.nih.gov/genbank).

Variant	Gene	Load	Phenotype	In Population Databases?
A1555G	MTRNR1	Heteroplasmic	RCM/HCM	Often
C1624T	MTTV	Homoplasmic	Leigh syndrome, HCM/DCM	No
C1628T	MTTV	Homoplasmic	Myopathy, ptosis, ophthalmoplegia, conduction abnormalities, HCM.	Yes
G1644A	MTTV	Homoplasmic	MELAS and HCM	No
G3242A	MTTL1	Both	Infantile cardiomyopathy, lactic acidosis, hypotonia, renal dysfunction	Yes
A3243G	**MTTL1**	**Heteroplasmic**	**Very varied including isolated HCM**	**Yes**
T3252C	MTTL1	Heteroplasmic	MELAS	No
A3260G	**MTTL1**	**Heteroplasmic**	**MELAS/ Myopathy and cardiomyopathy**	**No**
A3280G	MTTL1	Heteroplasmic	Myopathy and HCM	No
C3303T	**MTTL1**	**Both**	**Cardiomyopathy (DCM/HCM) and myopathy**	**No**
T3308C	MTND1	Both	Various phenotypes including LVNC	Frequent
G3337A	MTND1	High heteroplasmic	Various phenotypes with different cardiomyopathies	Yes
T3398C	MTND1	Heteroplasmic	LVNC	Yes
G3460A	MTND1	Both	Various, commonly LHON, including LVNC	Yes
A4269G	MTTI	Heteroplasmic	Multisystem disorder including DCM	No
T4277C	MTTI	Homoplasmic	Hearing impairment, myopathy and HCM	No
G4284A	MTTI	Heteroplasmic	Multisystem disorder including DCM	No

A4295G	MTTI	Heteroplasmic	Isolated HCM	Often
A4300G	**MTTI**	**Homoplasmic**	**Isolated HCM**	**No**
A4316G	MTTI	Heteroplasmic	Infantile cardiomyopathy and hearing loss	Yes
A4317G	MTTI	Heteroplasmic	Infantile DCM/HCM	Yes
C4320T	MTTI	Heteroplasmic	Infantile HCM and encephalopathy	No
4322dupC	MTTI	Heteroplasmic	Isolated DCM	No
C5545T	MTTW	Heteroplasmic	Infantile HCM, hypotonia and lactic acidosis	No
A5814G	MTTC	High heteroplasmic	HCM, myopathy and lactic acidosis	Frequent
A5843G	MTTY	Homoplasmic	Various including nephropathy and DCM	Frequent
G7697A	MTCO2	Both	Various phenotypes including HCM susceptibility	Yes
A8296G	MTTK	Both	Various including HCM	Yes
T8306C	MTTK	Heteroplasmic	Adult-onset multisystem including HCM	No
A8344G	**MTTK**	**Heteroplasmic**	**Variable, classically MERRF, including HCM and DCM**	**Yes**
A8348G	MTTK	Heteroplasmic	HCM progressing to DCM and others	Yes
G8363A	MTTK	High heteroplasmic	Encephalomyopathy, hearing loss, HCM	No
T8528C	MTATP8 /6	Heteroplasmic	Infantile HCM	No
G8529A	MTATP8 /6	Homoplasmic	HCM and neuropathy	No
T8993G	MTATP6	Both	Leigh syndrome including HCM	No
T9997C	MTTG	Heteroplasmic	HCM	Yes
G10677A	MTND4L	Heteroplasmic	Infantile DCM	No
G11778A	MTND4	Homoplasmic	LHON mutation but reported with HCM (with G12192A)	Yes
G12192A	MTTH	Homoplasmic	See above	Frequent
T12297C	MTTL2	Heteroplasmic	DCM	Yes
T12908A	MTND5	?Homoplasmic	Suggested as pathogenic in a child with Pompe including HCM	No
G13708A	MTND5	Homoplasmic	LHON modifier, reported in patient with HCM and WPW	Frequent
T14484C	MTND6	Both	LHON mutation, reported with DCM	Yes
T14849C	MTCYB	Heteroplasmic	Septo-optic dysplasia with HCM	No
G15498A	MTCYB	Heteroplasmic	Histiocytoid cardiomyopathy	Yes

Table 2-1: Mitochondrial DNA variants identified from the medical literature causing a phenotype including cardiomyopathy. Those with good evidence for pathogenicity are highlighted in bold. Abbreviations: DCM = Dilated cardiomyopathy; HCM = hypertrophic cardiomyopathy; LHON = Leber's hereditary optic neuropathy; MELAS = Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes; MERRF = Myoclonic epilepsy with ragged red fibres; RCM = Restrictive cardiomyopathy.

2.3 Results

2.3.1 Initial Results

Prior to analysis using the bioinformatics tools described above, 1910 variants were identified in the ninety-one DNA samples studied (mean- 20.8; range- 2-47 per patient), although the majority of these were common polymorphisms. Rare variants present in population data, and not identified in disease cohorts, have also been excluded on the assumption they are likely to be benign variants. This may be an oversimplification,

particularly with heteroplasmic variants, due to the recognition that definitively pathogenic mtDNA mutations can occur at reasonable levels in population cohorts (Elliott et al., 2008). However, this strategy enriches those variants remaining for likelihood of pathogenicity, albeit at the risk of removing relevant variants from further study.

After comparison with databases of mitochondrial variation several variants of possible significance remained. Four variants previously described as causing cardiomyopathy were identified, along with five novel variants, all identified in separate patients. Of those variants previously reported causing cardiomyopathy two patients had homoplasmic m.4300A>G *MT-TI* variants (2/91; 2.2%), one had a heteroplasmic m.3243A>G *MT-TL1* variant, and another had a homoplasmic m.4295A>G *MT-TI* variant. The five novel variants identified were all homoplasmic, and are described in the table below.

Variant	Gene	Amino Acid Change	Conservation
m.2321A>G	*MT-RNR2*	rRNA gene	
m.13998C>T	*MT-ND5*	Asp (D) → Val (V)	All except playpus
m.8643C>T	*MT-ATP6*	Synonymous (N)	Primates only
m.746A>G	*MT-RNR1*	rRNA gene	
m.3771A>G	*MT-ND1*	Synonymous (L)	Mammals

Table 2-2: Novel variants identified from sequencing the mitochondrial genome in ninety-one patients with HCM

As mentioned previously, in addition to the main HCM cohort certain selected patients were also sequenced. One of these was a family with HCM, but with several family members decompensating with dilatation and heart failure at young ages. Endomyocardial biopsy studies had already indicated non-specific mitochondrial abnormalities (see section 2.3.7, page 36), but clinical diagnostic testing had not identified the cause of these. Sequencing in this family also identified the m.4300A>G variant. This family is therefore also discussed in section 2.3.7 to provide additional clinical information regarding this variant.

Due to the finding of multiple patients with the m.4300A>G variant, further samples were sourced to help determine the significance of this variant in a larger HCM cohort. Testing was performed specifically for this variant, testing of the rest of the mitochondrial genome was not performed in this additional patient cohort.

2.3.2 Further Patient Cohort

Lymphocyte-derived DNA from an additional cohort of four-hundred and sixty-four patients with HCM was obtained. These patient DNA samples were sequential referrals to the Oxford Molecular Genetics Laboratory for HCM genetic testing, referred by clinical geneticists and cardiologists from across the UK. Patients were excluded on the same basis as with the initial cohort (if a causative mutation had been identified or DNA was unavailable), or if they had already been included in this initial cohort (the variant was specifically screened for in this cohort again in any case). Again no selection was done based on family history or clinical information received. None of the probands tested were known to be related.

2.3.3 Pyrosequencing of the m.4300A>G Variant

A specific pyrosequencing assay to test for the m.4300A>G variant was performed in all of the initial cohort to validate the findings from the NGS analysis. The same assay was then used to test for this variant in the further patient cohort. Direct screening and accurate quantification of the m.4300A>G mutation was performed by pyrosequencing using the PyroMark Q24 platform (Qiagen®) according to the manufacturer's protocol. Primers were obtained from Professor Robert Taylor, University of Newcastle, UK. They had been designed using the PyroMark assay design software v2.0 (Qiagen®) to create a mutation-specific pyrosequencing (PSQ) primer trio (biotinylated forward: m.4220-4242, reverse: m.4336-4356, and pyrosequencing primer: m.4302-4320; Genbank accession number NC_012920.1). The allele quantification application of the proprietary Q24 software was

used to accurately calculate mtDNA heteroplasmy levels, reported to have a sensitivity of within 1% of mtDNA mutation load (Alston et al., 2011).

The m.4300A>G mutation was also tested in other cardiomyopathy patient cohorts to determine if it was a significant cause of these. Forty-five patients with HCM due to known sarcomere mutations, one-hundred and forty-two patients with dilated cardiomyopathy (DCM), eleven patients with left ventricular non-compaction (LVNC), and four patients with mixed/unclassified cardiomyopathies were tested.

2.3.4 Haplotype Analysis

The D-loop of mtDNA is a triple stranded region of DNA that forms part of the non-coding control region of the mitochondrial genome. Due to the exclusive maternal inheritance of mtDNA, and variability of parts of the D-loop, the sequence of this region has been used to study evolution. The haplotypes identified from these studies can be used to identify whether a variant identified in several individuals shares the same common ancestor, or is arising separately in different haplotypes. Therefore those patients where the m.4300A>G variant was identified also had Sanger sequencing of the mtDNA D-loop (performed by Charlotte Alston, Newcastle University). The variants identified were ascribed haplotypes using the HaploGrep online algorithm tool (Kloss-Brandstatter et al., 2011, van Oven and Kayser, 2009).

2.3.5 Validation of Other MtDNA Variants

Validation of m.3243A>G, m.4295A>G, and the novel reported variants identified, was undertaken by Dr Carl Fratter, Oxford Molecular Genetics Laboratory, using Sanger sequencing.

2.3.6 Final Results

Pyrosequencing in the further cohort of four-hundred and sixty-four patients identified a further two patients with the m.4300A>G variant. Overall therefore in an unselected cohort of patients with an HCM phenotype four of five-hundred and fifty-five patients (0.7%) had the m.4300A>G variant. A further family with this variant had also been found in the initial sequencing from more stratified testing. In the patient cohorts with different clinical phenotypes none had the m.4300A>G variant.

Of the five patients ascertained with the m4300A>G variant, four different haplogroups were identified. Two patients were in the U4b1b1a haplogroup, with one each of haplogroups V9, K2, and H2a1e1. This indicates that this mutation is arising independently in these patients.

2.3.7 Clinical Phenotype of Families with the m.4300A>G Variant

A summary of the clinical phenotype and investigation findings in each family with the m.4300A>G variant is described below. Family pedigrees are shown in Figure 2-4. The first family is the family already identified as having a likely mitochondrial pathology prior to study, families two to five are those identified by this study in an unselected HCM cohort.

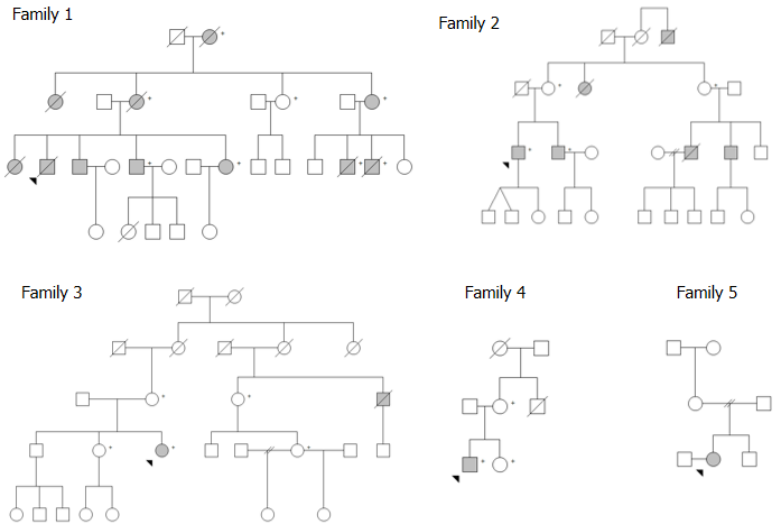


Figure 2-4: Pedigrees of the five families identified with the m.4300A>G variant

Family 1

Presentation: A male died at seventeen years of age with dilated cardiomyopathy, diagnosed shortly before death.

Cardiomyopathy: Where hypertrophy was present, in most family members a concentric pattern was seen although asymmetric hypertrophy was present in a minority. Maximal septal thickness measured was twenty-five millimetres. Five adult family members with the mutation are clinically unaffected. Two patients have had the development of hypertrophy observed in mid- or late adulthood, in the fourth and ninth decades of life.

Deaths: Five family members had progression to dilatation, heart failure and death (one male aged six years and two at seventeen years; and two females aged eight months and fifty-four years). All had hypertrophy greater than twenty millimetres prior to developing progressive dilatation and heart failure. In one family member deterioration was due to documented H1N1 influenza infection. A further female family member died aged four years. Post mortem reported the cause of death as being due to peanut aspiration, although twenty millimetre concentric left ventricular hypertrophy was reported.

Complications: One family member with marked hypertrophy and dilatation had an ICD inserted for primary prevention at thirty-two years of age with appropriate shocks for a VT storm three years later. No extra-cardiac manifestations of mitochondrial disease were reported in the family.

Other: Cardiac biopsy showed myocyte hypertrophy with striking vacuolization and nuclear polymorphism, and irregularly sized mitochondria with abnormal ring-shaped cristae on electron microscopy (Figure 2-5). Cardiomyocytes exhibited a severe defect of the respiratory chain complexes I and IV.

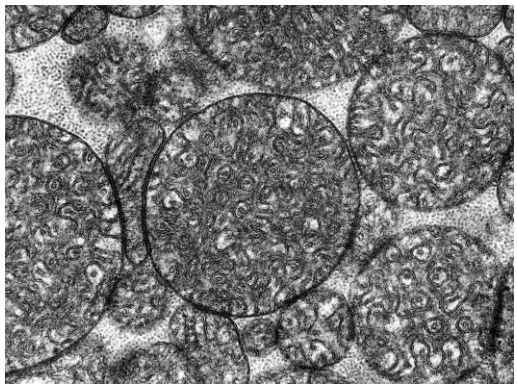


Figure 2-5: Electron microscopy image of an endomyocardial heart biopsy from a member of family 1. This shows irregularly sized mitochondria with abnormal ring-shaped cristae, this is highly suggestive of a primary mitochondrial disorder but is not specific for a particular cause.

Family 2

Presentation: Sudden cardiac death occurring in a forty-eight year old male. On post mortem he was identified as having thirty millimetre asymmetrical septal hypertrophy.

Cardiomyopathy: The hypertrophy seen was either concentric or asymmetrical hypertrophy in different family members. Three family members with the mutation were clinically unaffected (females aged seventy-eight years, sixty-three years, and one who died at eighty years of age from unrelated problems).

Deaths: Prior to diagnosis in the family one male had died in his fourth decade due to 'heart failure' and a female had died aged twenty-two due to 'cardiomyopathy'.

Complications: Various arrhythmias were identified in the family: one patient with SCD, two patients with atrial fibrillation, and a further patient with monomorphic ventricular tachycardia. One male was diagnosed with diabetes at forty-six years of age but no other extra-cardiac mitochondrial features were recognised in the family.

Family 3

Presentation: Incidental diagnosis of asymptomatic concentric hypertrophy up to fifteen millimetres in a thirty-six year old female.

Cardiomyopathy: Four family members with the mutation were clinically unaffected at forty-five, forty-six, seventy and seventy-three years of age. Four other female family

members, who were obligate homoplasmic mutation 'carriers', lived into their ninth or tenth decades.

Deaths: A male relative had been diagnosed with cardiomyopathy at thirty years of age. He is then reported to have died at thirty-nine years of age following an influenza infection.

Complications: No arrhythmias have been recorded in any family members. The proband was noted to have partial congenital deafness of unknown cause, but no other extra-cardiac mitochondrial features were recognised in her or the wider family.

Family 4

Presentation: A sixteen year old male developed exertional breathlessness and chest pain. On investigation of this it was recognised that he had predominantly concentric hypertrophy, but with some asymmetry, up to thirty-four millimetres (Figure 2-6).

Cardiomyopathy: The mother and sister of the proband are both clinically unaffected. Testing of the m.4300A>G, variant in DNA from peripheral lymphocytes, showed heteroplasmy of this variant (measured at twenty-five and twenty-eight percent respectively).

Deaths: None relatable to cardiomyopathy known.

Complications: 24 hour ECGs in the proband showed Wenkebach phenomenon, pauses and nocturnal sinus bradycardia (down to thirty-five beats per minute).

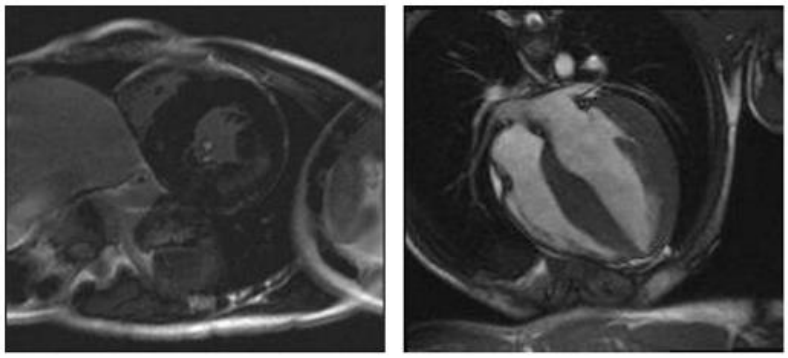


Figure 2-6: Long and short-axis views of delayed T1-weighted post-gadolinium enhancement images from the proband of family 4. This demonstrates marked predominantly concentric left ventricular hypertrophy with late gadolinium enhancement in the postero-infero-lateral wall.

Family 5

Presentation: A twenty year old female was investigated after presenting with exertional chest pain. Echocardiography and cardiac magnetic resonance imaging demonstrated concentric hypertrophy, but with some asymmetry, up to twenty-seven millimetres. On treatment she has remained well over fourteen years of follow up, with no significant change in clinical condition.

Cardiomyopathy: Her mother, with confirmed homoplasmy of the mutation, is clinically unaffected at the age of fifty-six years. Her brother was also clinically unaffected at the age of twenty-four years but has not been followed up since.

Deaths: None relatable to cardiomyopathy known.

2.4 Discussion

2.4.1 m.4300A>G as a Cause of an HCM Phenotype

This study is the first to our knowledge to look at the whole mitochondrial genome systematically in an unselected HCM cohort. It has highlighted the m.4300A>G mutation as a cause of a maternally-inherited hypertrophic cardiomyopathy phenotype. Although this has previously been recognised, it has only been reported in three families to our knowledge (Casali et al., 1995, Casali et al., 1999, Taylor et al., 2003). A prevalence in this cohort of between a half and one percent is modest for a single causative gene, but for an individual mutation it is a significant finding.

Although some patients with the mutation exhibit features that are uncommonly seen in 'classic' sarcomeric HCM, such as rapid decompensation to DCM, overall there is significant overlap in the clinical phenotype. The absence of confirmed extra-cardiac mitochondrial features means that individual patients can be difficult to distinguish on clinical grounds alone. Given the implications of maternal rather than dominant inheritance, and the

emerging differences apparent in clinical course, not recognising these patients has implications for genetic counselling and management.

It is also noticeable that three of the probands (families three to five) with the m.4300A>G mutation were, at least initially, thought to be isolated cases of HCM without known familial involvement. This was typically due to the significant clinical variability associated with this mutation. However, in one family, numbered 4 in section 2.2.1, this was explained by the previously unobserved conversion from heteroplasmy to homoplasmy at the time of inheritance of this mutation. Despite this observed clinical variability the m.4300A>G mutation is absent from population studies of mitochondrial variation (Pereira et al., 2009).

The m.4300A>G mutation is unusual as a pathogenic mtDNA mutation in that it has only been identified as causative for HCM when homoplasmic, rather than heteroplasmic (Taylor et al., 2003). This is not without precedent and is more commonly seen with other tissue-specific disease, such as Leber's Hereditary Optic Neuropathy (Wallace et al., 1999). This study adds further weight that homoplasmy is required by identifying clinically unaffected relatives of a patient where the m.4300A>G mutation is heteroplasmic.

Although cardiomyopathy causing mutations have been identified in several mitochondrial genes, mutations in the mitochondrial isoleucine tRNA gene, *MT-TI* (*tRNA^{Ile}*), appear to give a more consistent HCM phenotype (Kelley et al., 2001), albeit often in association with myopathy. Mutations in the *tRNA^{Ile}* gene are also unusual in being fairly specific for this phenotype, with chronic progressive external ophthalmoplegia being the only other main phenotype commonly reported (Chinnery et al., 1997, Taylor et al., 1998, Silvestri et al., 1996). Mitochondrial tRNA^{Ile} gene variants previously reported as causing HCM include the heteroplasmic variants; m.4269A>G (Taniike et al., 1992), m.4295A>G (Merante et al., 1996), m.4316A>G (Chamkha et al., 2011), m.4317A>G (Tanaka et al., 1990, Obayashi et al., 1992, Tomari et al., 2003, Degoul et al., 1998), m.4320C>T (Santorelli et al., 1995) and

m.4322dupC (Mahjoub et al., 2007); and homoplasmic m.4277T>C (Perli et al., 2012), as well as m.4300GA>G, variants. Interpretation of the significance of mitochondrial variants can be difficult however, and whether all of these are pathogenic has been questioned.

2.4.2 Other Cardiomyopathy-Associated Variants

In the initial cohort of ninety-one patients other potentially significant variants were recognised. The most interpretable finding is the identification of a heteroplasmic m.3243A>G mutation in one patient. This patient had predominantly concentric hypertrophy up to sixteen millimetres, with normal ventricle size and normal ejection fraction. Further study of this family identified mild left ventricular hypertrophy in the mother of the proband, with no features in either of his children (as would be expected). There were no extra-cardiac mitochondrial features in the proband or family. The m.3243A>G is probably the best studied of mitochondrial mutations and can give a variety of phenotypes, including isolated cardiomyopathy (Silvestri et al., 1997). The finding in this patient, who has no indicators of a mitochondrial cause on clinical assessment, adds further support to routine testing for this mutation in HCM.

One patient was homoplasmic for the m.4295A>G variant. This patient is an adult male with moderate asymmetrical septal hypertrophy, no extra-cardiac features, and no family history of cardiomyopathy. Although this variant has been reported previously causing hypertrophic cardiomyopathy (Merante et al., 1996) it is also relatively commonly identified in population studies (Pereira et al., 2009), as well as being reported in other mitochondrial disease phenotypes (Gutierrez Cortes et al., 2012). It remains possible that this gene is a lower penetrance allele promoting hypertrophy, but equally may be a polymorphism. Larger studies in disease cohorts are required to determine whether it is over-represented in these.

2.4.3 Novel Variants

This study has also identified a number of novel homoplasmic variants (m.746A>G, m.2321A>G, m.3771A>G, m.8643C>T, m.13998C>T). Only one of these variants, m.13998C>T, causes a non-synonymous change in a protein-coding gene (*ND5)*. The patient harbouring this variant is an adult male with predominantly concentric hypertrophy, no extra-cardiac mitochondrial features, and no family history of cardiomyopathy. He is of Indian subcontinent background. This may be of relevance as this ethnicity is generally under-represented in the population databases used, giving a higher likelihood that this may be a polymorphism present in this ethnic group. Further studies, for example analysis of mitochondrial structure and respiratory chain function from muscle biopsy, have not yet been undertaken.

Features of all of the other variants make them more likely to be rare benign variants. Two, m.2321A>G and m.3771A>G, were also identified in patients of Indian subcontinent background; in the patient where the m.746A>G variant was found the family history excluded segregation of this variant with the cardiomyopathy phenotype; and the other variant, m.8643C>T, leads to a synonymous change in a poorly conserved residue. This highlights one of the problems with sequencing of the mitochondrial genome, rare variants are relatively common and determining pathogenicity can be challenging.

2.5 Conclusions

This study demonstrates that the m.4300A>G mitochondrial mutation is a significant cause of an HCM-phenotype in an unselected cohort referred for genetic testing. Although some of these patients may be highlighted as being atypical for ‘classical’ HCM, based on their personal or family clinical features, the phenotype seen in many patients cannot be clinically distinguished from typical HCM. The m.3243A>G mutation has already been recognised as a cause of an HCM phenotype, and this is reiterated here. Testing for both of

these mutations should be considered in all patients with HCM where the family history is consistent with maternal inheritance. Further work is required to determine if any of the other variants described are causative for HCM but the likelihood is that most, if not all, are rare benign variants.

For two main reasons it has been decided not to test the whole mitochondrial genome in a clinical setting at present. Firstly, it is difficult to interpret the significance of many of the variants identified, the m.4295A>G variant demonstrating this well. Secondly, due to the relatively large size of the mitochondrial genome and the high coverage required to allow accurate determination of heteroplasmic variants, it is very hard to consistently achieve a satisfactory level of coverage for clinical diagnostic use without compromising other gene testing. For these reasons pyrosequencing assays to specifically test for the m.4300A>G and m.3243A>G mutations are offered as a clinical service in addition to the standard multi-gene panel if requested.

Chapter 3:

Study Design and Methods in Family Studies

Although phenotype specific multi-gene tests using next generation sequencing (NGS) capabilities may be useful in the clinical diagnostic setting (Meder et al., 2011), more expansive testing, looking at the exome or genome, has significant benefits in most instances in the research setting (Ng et al., 2010b). Family studies have long been instrumental in identifying the genetic aetiology of Mendelian disorders (Royer-Pokora et al., 1986), and NGS in families with Mendelian disorders offers advantages over previous techniques used to identify these (as discussed in section 1.4.3, page 10). This chapter introduces the overall study design and discusses the methods used in the initial studies to produce candidate causative variants in the individual families studied, each with differing inherited heart diseases. The subsequent chapters describe the phenotypes of these families in more detail, and the work performed to validate these variant further.

3.1 Study Design and Strategy

As has been discussed in section 1.5.3 (page 14) the large number of genomic variants within an individual can make the identification of the causative mutation in a family difficult. Although there are methods for filtering likely pathogenic variants (section 1.5.4, page 15), the most effective method of identifying the cause is to have the maximum discriminating phenotypic and segregation information available. We have therefore centred our study on families with multiple affected family members, all consistent with autosomal dominant inheritance.

Although all of the families studied have different types of inherited heart muscle and/or electrophysiological disorders, none of the families have a typical phenotype of one of the

genetically better understood disorders in this group, such as HCM or long QT syndrome (LQTS). The phenotypes are discussed in more detail in chapter 4, but they include genetically heterogeneous and less well understood disorders, such as left ventricular non-compaction and dilated cardiomyopathy, along with some rare and unusual phenotypes that are difficult to classify (left ventricular hypertrophy, prolonged QT interval, and sinus node dysfunction; widespread repolarisation abnormalities; and a phenotype with features of a number of cardiomyopathies, electrophysiological abnormalities, and septal defects). Identifying the underlying genetic cause in families such as these may be particularly informative in understanding the molecular biology of some of these individual features in a broader sense.

A broad overview of the study design is illustrated (Figure 3-1) and the steps briefly described below. The methods used are then described in detail in the subsequent sections.

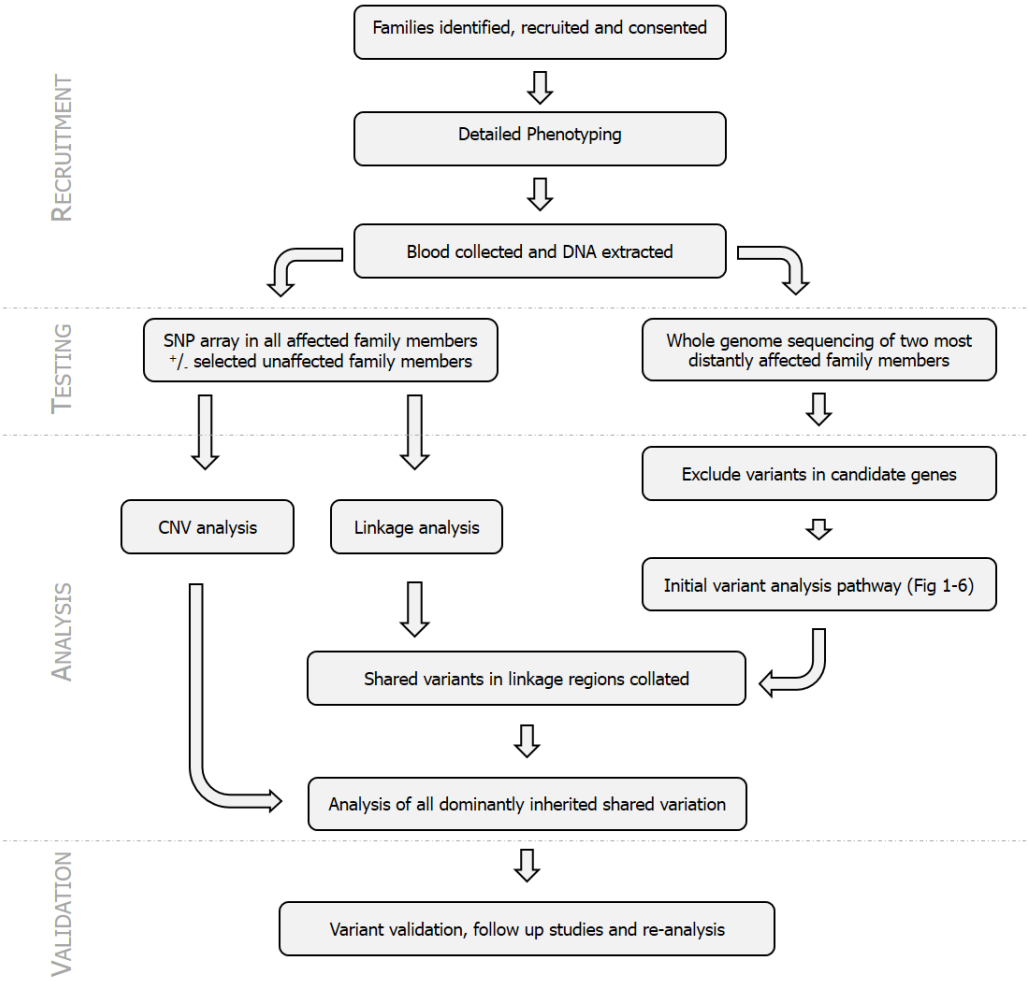


Figure 3-1: Overview of the study strategy used to study families with inherited heart diseases using whole genome sequencing and segregation information from SNP array testing

1) Patient phenotyping

Recruitment is discussed in section 3.2.1 (page 49). Individuals in recruited families were assessed in detail by a cardiologist and a clinical geneticist. This included a clinical history and examination (including assessment for non-cardiac involvement), electrocardiogram (ECG), and echocardiogram. In some patients other specialised tests were also performed, such as cardiac magnetic resonance imaging (CMR), 24 hour ECGs and exercise ECG tests.

2) Obtaining DNA

Blood was collected from all family members seen and DNA extracted (section 3.3.1, page 52) with full informed consent as part of the study (section 3.2.2, page 49).

3) Whole genome sequencing and single nucleotide polymorphism array

Whole genome sequencing (WGS) and single nucleotide polymorphism (SNP) array has been performed in each family contemporaneously.

Whole genome sequencing (full details in section 3.3.2, page 52)

WGS was performed in the two most distantly related family members definitely affected with the condition where DNA was available. It was decided to test two family members, rather than one, for two main reasons. Having two sequences from a family would significantly reduce the identification of variants caused by sequencing error. The exception to this is the possibility of systematic error occurring in multiple samples, although having variant information in all WGS500 study samples would help to detect any systematic error. The other main reason was to provide distinct variant data for comparison between individuals, which should act to further clarify the variants within linked regions.

Single nucleotide polymorphism array (full details in section 3.3.3, page 58)

SNP array was performed in all affected family members where DNA was available to provide maximum linkage data. In some families well phenotyped unaffected adult family members were also assessed. This was to provide further information in assessing likely pathogenicity of identified variants, but variants were not excluded on the basis of this due to the possibility of incomplete penetrance in these individuals.

4) Data analysis (full details in section 3.3.4, page 61)

WGS data was initially analysed for likely pathogenic variants in previously reported and candidate genes for the phenotype. All variants identified were then sifted to reduce artefactual results and polymorphisms. Regions of chromosomal linkage (LOD score>0), suggestive of segregation of variants in these regions, were established in each family from SNP array data. The WGS identified variants were further filtered by removing those located outside these linkage regions in each family. The resultant list

of variants was then manually analysed, and cross-referenced with published literature and databases, to determine likely pathogenicity.

5) **Validation, follow-up work and re-evaluation**

All variants of interest were Sanger sequenced (section 3.3.5, page 66), or had chromosome array comparative hybridisation performed in the case of a larger copy number variant (CNV), for validation. The strategy for further investigation depended on whether one or more variants felt likely to be causal for the phenotype had been identified. Where a potentially causal variant had been identified, the main strategy for investigating these further was to sequence candidate gene(s) in a larger cohort with the same phenotype, and/or by assessing the effect of the variant *in vitro*. If no likely candidates were identified then each step was reassessed, for example considering alternative modes of inheritance and ensuring correct ascertainment of affected status of family members.

3.2 Study Information

3.2.1 Patient Recruitment

Families were predominantly identified and recruited from those seen in the Oxford Inherited Cardiac Conditions Clinic. Other patients and families have been sourced through contact with clinical geneticists and cardiologists from across the UK.

3.2.2 Consent and Ethics

Ethical approval was given by the Oxfordshire Research Ethics Committee B, now the South Central- Oxford B NRES Committee ("Molecular Genetic Studies of Individuals and Families at Risk of Inherited Cardiac Disease"; REC approval number: 09/H0605/3). The main patient information and consent form example are shown in the appendix sections I-a and b (page 237). Further information and consent forms were also approved for alternative situations, such as children or next-of-kin.

Due to the nature of the project being undertaken, with the use of whole genome sequencing rather than specific gene testing, further discussion and consenting of all individuals who had this performed was also undertaken. Particular attention was given to the concerns regarding the generation of data and results given to patients regarding genes not directly related to the question studied (discussed in section 1.4.4, page 11). Due to the study design the chances of incidentally identifying an interpretable variant of significance, unrelated to the cardiac phenotype investigated, was significantly diminished. By only analysing variants shared by affected family members the majority of genetic variation in an individual would not be actively studied, although it would have been sequenced and annotated. In addition the families are well phenotyped such that a dominantly inherited predisposition to, for example, a cancer predisposition syndrome is unlikely not to be implicated in the family history.

Although the study strategy would remove the majority of potential incidental findings, two main questions with regards to the discovery of mutations unrelated to the phenotype remained. The first was whether specific genes should be analysed, further to those studied for the purposes of the project. There is no clear consensus on what should be done with regards to this, nor evidence on the beneficial and/or detrimental effects of reporting additional genetic findings (van El et al., 2013, Green et al., 2013, PHG_Foundation, 2013, Allyse and Michie, 2013). As well as the ethical considerations it would create significant practical issues, such as the interpretation of a potentially significant number of variants of unknown significance, and if pathogenicity could not be resolved whether they then need to be reviewed on an ongoing basis. The other main question was that as the sequencing was being performed as part of the WGS500 study other researchers may study data generated, which could potentially identify other variants of significance. All such studies would be based on coded but anonymous data, although this could be linked to a patient by the clinical team.

After discussion within the department it was decided that our local/study policy was that genes unrelated to the study question would not be specifically analysed. Any truly incidental findings, or those found in the analysis of WGS500 data by other researchers, would be considered on an individual basis. Findings unrelated to their cardiac phenotype would not routinely be returned by the clinical team, but each finding would be considered on an individual basis and acted upon if it was felt this was felt to be of significant medical benefit. Future research and consensus documents are awaited to guide this local policy in the future however. Patients were consented on the basis that their data may be studied to investigate their cardiac phenotype, and anonymised data studied more generally, at any point in the future.

3.2.3 Funding

This work has been kindly funded by the British Heart Foundation. Personal salary and lab consumables were funded by a British Heart Foundation Centre of Research Excellence Fellowship Award (grant code HSRNWBY). Specific funding for next-generation sequencing costs was received from a competitive British Heart Foundation Centre of Research Excellence Research Pump-Priming Award (grant code HSRNWB11)

3.2.4 WGS500 Project

Sequencing has been performed as part of a collaborative project between the University of Oxford and Illumina® to perform whole genome sequencing in five-hundred patients with various medical conditions, the 'WGS500 project'. This project is funded by the Wellcome Trust (grant code 090532/Z/09/Z) and the Medical Research Council (grant code G0900747 91070).

3.3 Methods

3.3.1 DNA Extraction

The majority of samples were manually extracted from blood lymphocytes using the Flexigene DNA Kit™ (Qiagen®) using the manufacturers standard protocol (http://tinyurl.com/nt3f2m2). Some DNA samples had already been extracted using a Qiagen® EZ1 BioRobot™ automated DNA extractor in the Bristol NHS Molecular Genetics Laboratory, or Gentra® Autopure LS™ automated DNA extractor in the Oxford NHS Molecular Genetics Laboratory. DNA was suspended in either Tris-EDTA buffer (TE) (10 mM Tris, 1 mM EDTA; pH 8.0) or sterile filtered water.

3.3.2 Whole Genome Sequencing

3.3.2.1 Sample Preparation

DNA was diluted to 50 ng/μl as determined by Nanodrop™ and/or PicoGreen™ measurement, in either TE buffer or water. A minimum total amount of 1 μg of DNA was used, with quality checked by gel electrophoresis and 260/280 and 260/230 ratios on spectrophotometry (ideally >1.8 and >2.0 respectively).

Library Preparation Steps:

Most of the steps in library preparation were automated using Beckman-Coulter® systems. Agarose gel insert size selection (step 4) was performed manually.

1) DNA fragmentation

   Genomic DNA was fragmented using a Covaris® (S2) ultrasonicator (settings- duty cycle 10, intensity 5, cycles/burst 200).

2) Sample preparation

   Samples were prepared for sequencing using the NEBNext® DNA Sample Prep Master Mix Set for Illumina® Kit (NEB®), following standard manufacturer's guidelines

(http://tinyurl.com/mecxhon) with minor modifications. Overhangs created by the shearing process are blunted using the kit 'end repair mix'. The exonuclease in the kit removes 3' overhangs, and the incorporated polymerase completes the complementary sequence to the 5' overhangs. Each blunt ended fragment has a single adenine added to the 3' end, 'A-tailing', to reduce ligation with other sequence fragments in the adapter ligation step. A thymine residue is present on the 3' end of each adapter to give a complementary overhang for ligation of the fragment to the adapter.

3) Adaptor ligation

Samples are ligated to indexing adaptors, which have eight nucleotide identifier sequences within them. The adaptors are paired to allow paired-end sequencing with insert sizes of up to five-hundred base pairs. Ligation of adapters was performed using 6 µl of Illumina® Adaptors using the proprietary Multiplexing Sample Preparation Oligonucleotide Kit.

4) Product purification

Products of ligation are purified to remove unpaired adapters or those which have ligated with each other, allowing the generation of a library of DNA fragments of an appropriate size for sequencing (300-400bp). Ligated libraries were size selected by electrophoretic mobility, using 2% E-Gel EX gels (Invitrogen®). The distribution of fragments in the purified fraction was determined using a Tapestation 1DK system (Agilent®).

5) Polymerase chain reaction amplification

Polymerase chain reaction (PCR) is used to selectively amplify those fragments which have adapters ligated to each end, but a limited number of PCR cycles are used to try and limit unrepresentative amplification of the library and the introduction of

artefactual sequence changes. Each library was PCR enriched with 25 µM of each of the custom multiplex and index primers:

Multiplex PCR primer 1.0: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT and

Index primer: 5'-CAAGCAGAAGACGGCATACGAGAT[INDEX]CAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT.

An eight base-pair in-house indexing system was used. Four independent PCR reactions per sample were prepared using 25% volume of the pre-PCR library each. After eight cycles of PCR (cycling conditions as per Illumina® recommendations) the four reactions were pooled and purified with AMPure XP beads (Beckman-Coulter®). The final size distribution of the amplified products was determined using the Tapestation 1DK system.

6) Validation

Libraries are quantified by quantitative PCR (qPCR) to create optimum cluster densities across a flow cell (full methods available at http://tinyurl.com/norzr36). The concentration of each library was determined by real-time PCR using the Agilent qPCR Library Quantification Kit and a MX3005P instrument (Agilent®).

3.3.2.2 Sequencing

The created libraries then require cluster generation on the flow cell before being sequenced on the Illumina® HiSeq 2000™ platform.

Cluster generation

Cluster generation is achieved in a highly automated process using proprietary TruSeq™ cluster generation reagents and a cBot™ machine. Full information on this process can be found at http://tinyurl.com/oplkoxe, an overview is illustrated and described below:

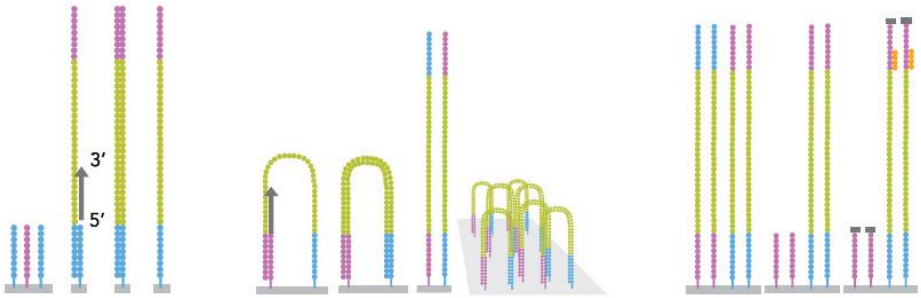


Figure 3-2: Illustration of the Illumina cluster generation method:

Left image - Adaptor sequences hybridise to hundreds of millions of oligonucleotides immobilised on the flow cell surface. A high-fidelity DNA polymerase copies the template from the hybridised primers by 3' extension. Template strands are then denatured to leave the copies attached to the flow cell.

Middle image- Bridge amplification- These copy strands loop over as the paired adaptor attaches to an adjacent oligonucleotide. DNA polymerase then forms complementary sequence to form double-stranded DNA (dsDNA) bridges, this is amplified by repeated cycles of isothermal denaturation and replication to give millions of distinct clonal DNA clusters, each containing about two-thousand DNA strand copies.

Right image- Each dsDNA strand is denatured and the reverse strand is removed by specific base cleavage. Exposed 3' adaptor ends are blocked to prevent interference with sequencing. The sequencing primer (red) is hybridised to the unbound ends of the templates in the clusters. At this point greater than two-hundred million clusters are present on each flow cell for sequencing.

Sequencing

Sequencing was performed on a HiSeq2000™ platform using the manufacturer's protocol (a detailed protocol is at http://tinyurl.com/prrbcbf). This is a highly automated process using proprietary 'sequencing by synthesis (SBS)' reagents.

Flow cells, with amplified DNA clusters, are sequenced using fluorescently labelled deoxynucleoside-triphosphates (dNTPs) in parallel (Figure 3-3: Images 1 and 2). During each sequencing cycle a different dNTP (A, C, T or G) is incorporated, the label then terminates polymerisation, and the fluorescence is triggered by a laser and imaged. The fluorescent label is then enzymatically cleaved, which then allows incorporation of the next nucleotide in the sequence (Figure 3-3: Images 3 to 5). The presence of the different dNTPs leads to natural competition for incorrect incorporation. Base calling occurs at each cycle directly from signal intensity images, giving 'base-by-base' sequencing.

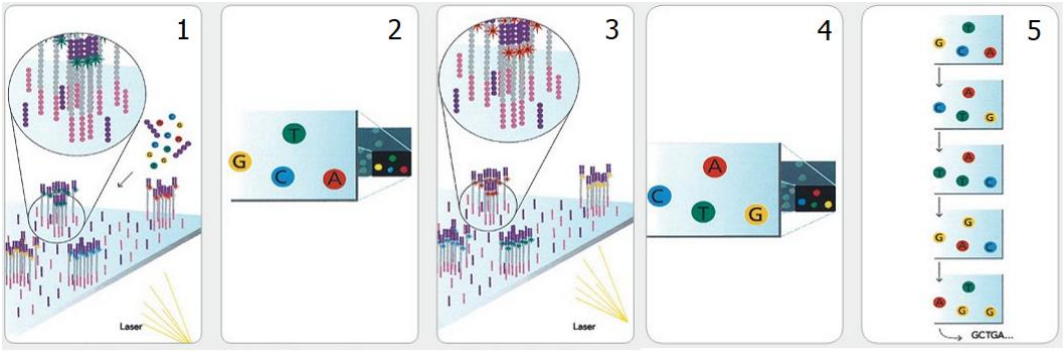


Figure 3-3: Illustration of the Illumina sequencing-by-synthesis method

On standard paired-end flow cells each run can produce up to three-hundred gigabases of data, with up to 1.5 billion clusters on each flow cell and read lengths of up to one-hundred base pairs. On the HiSeq platform two cells can be run at the same time, doubling this total output per run. To obtain this level of information a sequencing run takes up to eleven days. Each flow cell has eight lanes, seven sample lanes and a control lane, with each lane capable of over two-hundred million reads.

In this study version 3 sequencing chemistry used on the HiSeq2000 platform. A PhiX control (Illumina®) library was included to assist in determining sequencing and alignment errors. Paired-end sequencing was performed with one-hundred base pair read lengths and each flow cell was used to sequence three samples, i.e. two full lanes per sample and then the three samples are pooled in the final lane. In most samples this gave a depth of coverage at each base position of at least twenty-five times (25x). Any samples which had regions falling below this level, based on analysis of a single chromosome, were pooled and rerun until this threshold was met.

3.3.2.3 Mapping and Variant Calling

Following the generation of fluorescent images, which capture the incorporation of fluorescently labelled nucleotides to the DNA clusters, these need to be processed to generate variant information.

1) *Base calling and quality scoring*: Base-calling algorithms convert these images of fluorescence (in TIFF format) into data. A quality score is assigned to each base call, the PHRED score (Ewing et al., 1998, Ewing and Green, 1998). A score of 20, typically used as a quality cut off, equates to a 1% chance of a miscalling error. Quality scores are usually recalibrated after alignment by taking into account other information, such as the position in the read and comparison with the reference genome at certain non-polymorphic sites (Phred-type quality score).
2) *Alignment*: Reads are then mapped, or aligned, to the reference genome (in this study NCBI genome build GRCh37/UCSC hg19) to allow identification of variants from the reference. Most of the commonly used tools for this are based on 'hash' tables, derived from the original BLAST algorithm (Altschul et al., 1990). Alignment is more difficult when there is significant variation from the reference, such as with the major histocompatibility complex.
3) *Variant calling and annotation*: This data, stored in the text-based BAM/SAM (Binary Sequence Alignment Map) file format, then requires variants from the reference to be identified, 'called', using specific software. At this point data can be filtered, for example often only bases with a Phred-type quality score >20 are analysed. Different tools will incorporate different information, such as prior knowledge of likely allele frequency, to try and correctly identify variants. This produces a file containing variant information in a text based format, termed the variant call format (VCF) file. These variants are then 'annotated' to add information such as information on cytogenetic position, functional consequence and evolutionary conservation scores, and presence or absence in population data sets.

There are several tools which have been developed for mapping, variant calling and annotation, with different strengths and weaknesses. Analyses have been performed to compare different tools, and combinations of these. These studies have demonstrated that,

although there is a significant overlap in the variants called, there is also a significant minority of different calls (Li and Homer, 2010). As there is no absolute 'gold-standard' reference it is not currently possible to determine which of these are most accurate with complete certainty.

The WGS500 project has used a 'pipeline' incorporating the STAMPY mapping tool (Lunter and Goodson, 2011) and the PLATYPUS variant caller (Rimmer et al., 2012). The available tools have been studied in detail within the WTCHG and it is felt this combination of internally developed tools provides superior accuracy to the others available (Lunter and Goodson, 2011). The annotation of variants in this 'pipeline' is performed using the ANNOVAR tool (Wang et al., 2010b).

3.3.3 Single Nucleotide Polymorphism Array

Single nucleotide polymorphisms (SNPs) are variations of a single nucleotide which vary between individuals. Although they may be relevant for phenotypic or complex traits, by definition these should not be pathogenic variants for Mendelian disease. Over one-and-a-half million common SNPs are recognised (Sachidanandam et al., 2001, Altshuler et al., 2010), with over 50 million SNPs in total reported in the NCBI dbSNP database (Sherry et al., 2001). Many of the identified SNPs are in blocks of linkage disequilibrium, such that in these regions only a minority of 'tagging' SNPs are required to characterise these blocks (Johnson et al., 2001). Analysis of large numbers of these common SNPs, including tagging SNPs, underlies array-based SNP analysis. The ability to sequence large numbers of SNPs has been utilised in several ways, including genotyping in 'normal' populations, genome-wide association (GWAS) complex trait studies, and assisting the study of Mendelian disease (Carlson et al., 2001).

In this study we utilised SNP arrays primarily for providing information on regions of genetic linkage within each family, but also secondarily for the assessment of copy number

variation. There are several commercially available array platforms, but the HumanCytoSNP-12 BeadChip™ (CytoSNP-12) by Illumina® was used in our study. This array detects ~300 000 SNPs in a stratified fashion. There is good coverage across the genome with overrepresentation within regions of interest in terms of human disease, for example reported cytogenetic abnormalities. The method for using Illumina Infinium™ assays, such as the CytoSNP-12, is summarised in Figure 3-4 and described below. As with other 'high-throughput' techniques it is a highly automated process and uses proprietary 'kit' reagents. The methods are described in more detail by Gunderson *et al* (Gunderson et al., 2006).

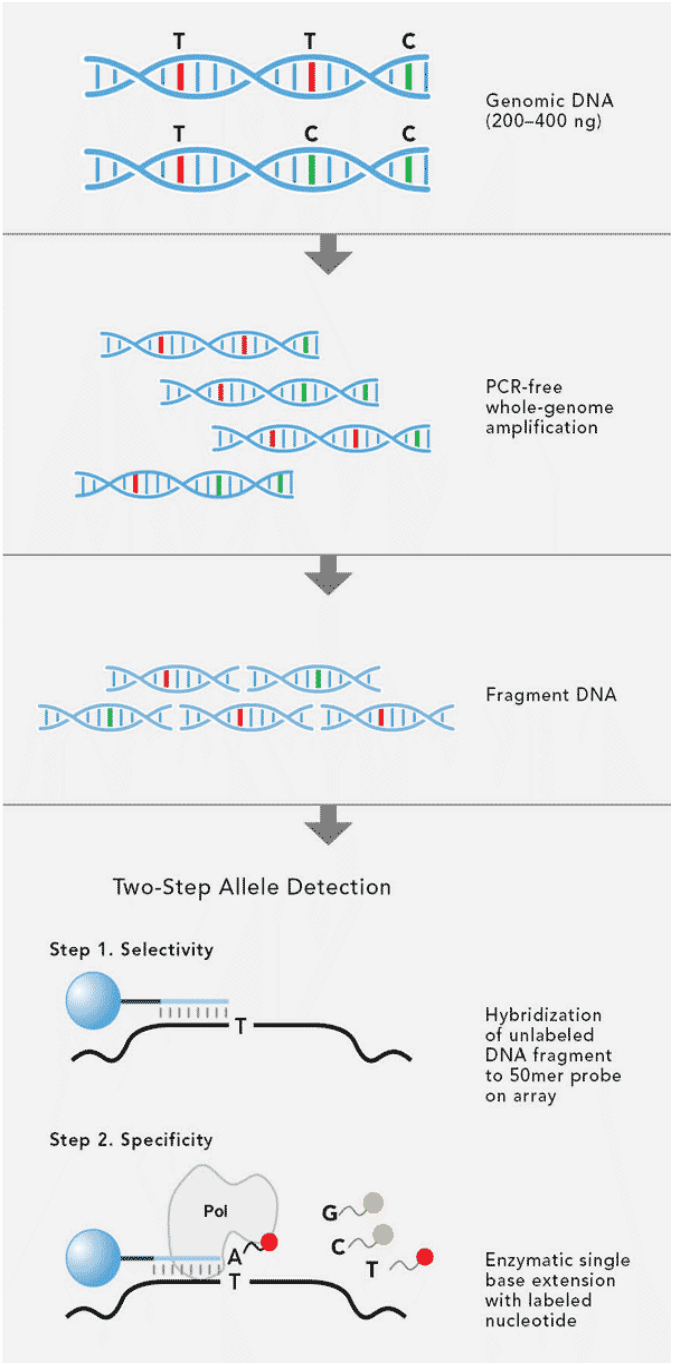


Figure 3-4: Illustration summarising the Illumina Infinium assays used for SNP array analysis.

1) DNA amplification

250 ng of DNA (diluted to 50 ng/µl in TE buffer or water) is amplified by an overnight isothermal reaction to produce at least 1 µg of DNA. The DNA is then fragmented enzymatically and is then precipitated using isopropanol and resuspended.

2) Hybridisation to bead-bound oligonucleotides

A BeadChip is comprised of beads embedded with 80-mer oligonucleotides that are immobilised on a solid array. The 5' thirty bases are used for decoding and the 3' fifty bases are complementary to the locus of interest. The samples are denatured and then hybridised overnight to the BeadChip array in a Te-Flow Through Chamber™ (Tecan®).

3) SNP detection and calling

In the single base extension (SBE) reaction labelled dideoxynucleotides are competitively incorporated into the probe, revealing the base at that position of the template strand. The biotinylated/DNP nucleotides are incorporated during primer extension, these are stained and amplified in an immunohistochemistry-based multilayer sandwich assay. The staining reagent is optimised to give high signal with balance between green and red channels. After staining the arrays are read using a high-resolution confocal scanner (iScan™ system). Genotype calls are made automatically within Illumina® GenomeStudio™ software (mapped to the NCBI genome build reference GRCh37/UCSC reference hg19).

3.3.4 Data Analysis

3.3.4.1 Whole Genome Data Analysis

Initial variant analysis pathway

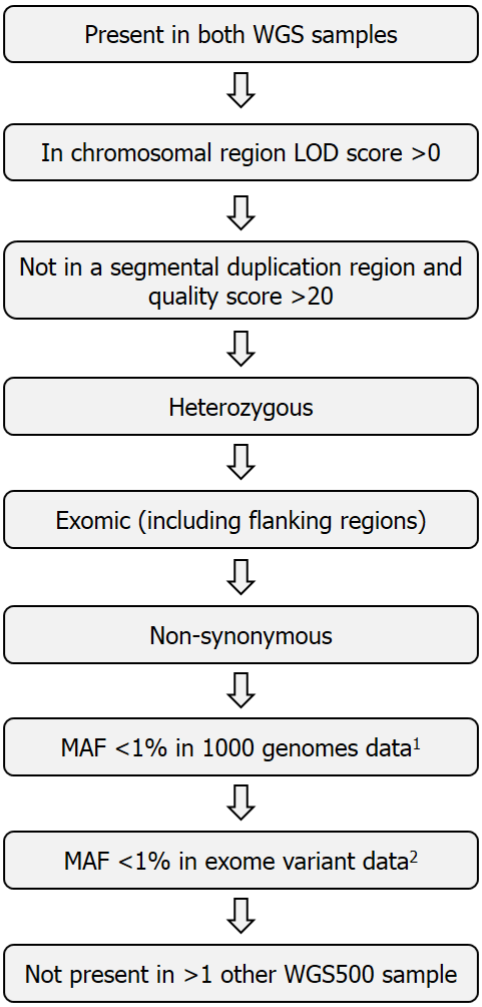
Variant data was annotated using modified ANNOVAR software (Wang et al., 2010b) to give the following information for each variant:

- Nucleotide reference and change
- Chromosomal position reference
- Site and nature of change, e.g. exonic and non-synonymous
- Amino acid change and location in protein
- Quality score and whether located in a segmental duplication region

- Presence in the population databases dbSNP (Sherry et al., 2001), 1000G (Abecasis et al., 2010), and EVS (EVS, March 2013)
- Presence in other WGS500 samples
- SIFT (Kumar et al., 2009), Polyphen2 (Adzhubei et al., 2010), PhyloP (Pollard et al., 2010), and MutationTaster (Schwarz et al., 2010) scores
- Likelihood-ratio (LOD) score

All samples underwent regular re-annotation whenever changes were made to the variant mapping and calling pipeline. This was particularly frequent in the initial stages of development, with some changes to reported annotated variants due to these. Changes in annotation also occurred, including the addition of certain information that could be used in variant filtering. For example the population data from the EVS database (EVS, March 2013) was not initially included. The incorporation of this allowed it to be used in the filtering of population variants, alongside 1000G data (Sherry et al., 2001, Abecasis et al., 2010). Regular updates were also made when more samples from the WGS500 study became available at set 'freezes' in data collection.

An initial analysis pathway was designed to remove low quality/artefactual variants and polymorphisms (Figure 3-5). The scientific reasoning behind each step is discussed in section 1.5.4 (page 15). When combined with linkage information in each family the intention was that this would form a manageable number of variants for manual assessment of likely pathogenicity, but with a low probability of having excluded the pathogenic variant from analysis.



[1]1000 Genomes Project variant database (http://www.1000genomes.org/)
[2]NHLBI-GO Exome Variant Server (http://evs.gs.washington.edu/EVS/)

Figure 3-5: Initial variant analysis pathway to remove likely artefactual variant calls and polymorphisms (MAF = minor allele frequency; LOD = log of odds)

Criteria used in the initial variant analysis pathway:

1. Shared variants

   As two members of each family with inherited heart disease were sequenced, these could then be compared to exclude all variants not shared between these individuals. Variants were also then compared with SNP array data from all affected members of the family. Those variants in chromosomal regions with a LOD score of <0 were excluded (further information is provided in section 3.3.4.2, page 65).

2. Quality threshold

   A low quality score would indicate a low level of confidence that the variant described was genuinely present. A score of 20 is equivalent to a 1% chance a variant has been incorrectly called. Most nucleotide positions were called with much higher quality scores than this. Variants in regions of segmental duplication were also removed as these are well recognised as being likely to be spurious (Bailey et al., 2002).

3. Heterozygosity

   All of our families exhibited inheritance patterns most likely to represent a highly penetrant autosomal dominant allele, therefore heterozygosity of a variant was a stipulated requirement for initial analysis.

4. Exomic

   Although the whole genome was sequenced, only the exome (including flanking regions and known regulatory sequence) was initially analysed. This was due to the high probability that the cause of a highly penetrant dominantly inherited disorder would be in this region (Cooper et al., 1995), as well as the difficulty of interpreting variants in non-gene coding regions (Birney et al., 2007). The advantage of using whole genome sequencing, over exome sequencing, is that it provides more even and complete coverage along with the potential to explore non-coding variants if no plausible coding variants are found.

5. Exclusion of synonymous variants

   Only variants predicted to cause a change in the protein produced were included. Although synonymous variants have been identified to cause disease this is very rare as compared with the large number of these which are benign variants (Buske et al., 2013).

6. Population data

   Variants were removed if they were present at >1% level in a database of population variation (EVS, March 2013, Abecasis et al., 2010). This is a generous threshold for inclusion to ensure a very low risk of excluding a pathogenic variant in this initial automated screen. The threshold within the WGS500 data was lower as we have more information on the phenotypes within this cohort, at one sample (therefore potentially 0.2%). For an autosomal dominant highly-penetrant allele the population frequency is likely to be orders of magnitude less than this. The most common mutation in the most frequently occurring highly-penetrant inherited heart condition, hypertrophic cardiomyopathy, is p.Arg502Trp in myosin binding protein C. HCM occurs in approximately 1 in 500 people, and p.Arg502Trp is the cause in up to 2.5% (Saltzman et al., 2010). Therefore, finding this variant in an unselected cohort would be anticipated to occur in approximately 0.00005% of people. The vast majority of well validated mutations in highly penetrant inherited heart diseases are not reported in these databases, although this is less clear in certain disorders with lower penetrance (Risgaard et al., 2013).

3.3.4.2 SNP Data Analysis

Linkage information

SNP genotype information was provided from the Illumina® GenomeStudio™ software output data file. The small minority of SNPs which had low GenCall (no-call) scores were removed, this wasn't sufficient to significantly affect linkage regions produced due to the inbuilt redundancy at each locus in the densely populated SNP array. Shared alleles/regions of linkage were ascertained from the SNP genotypes using the Merlin analysis tool (http://www.sph.umich.edu/csg/abecasis/merlin/index.html) (Abecasis et al., 2002). This

uses sparse binary trees to provide gene flow information through a pedigree. The visual output linkage/shared allele files are given in the individual family analysis information.

Copy number variant detection

Chromosomal microdeletions or duplications, copy number variants (CNVs) can be detected using SNP array and/or WGS data (Pinkel et al., 1998). Evidence for large scale copy number variation is inferred from regional differences in read depth. Particular software packages can be used to improve detection of these regions. Although the Illumina® GenomeStudio™ software can provide this information samples were reanalysed using Nexus™ software (BioDiscovery®) which is more accurate for this (Darvishi, 2010).

3.3.4.3 Incorporating SNP Data with WGS data

As discussed regions of genetic linkage within families were used within the initial analysis pathway to filter variants produced by WGS (section 3.3.4.1, page 61), using a custom PERL command script generated by Dr Stefano Lise. A LOD score of 0 was used as a cut off, indicating no prediction of linkage between SNPs across affected family members using the parameters set, with a boundary of five kilobases at either side of linked regions. These were automatically calculated by MERLIN pedigree analysis software (Abecasis et al., 2002).

3.3.5 Sanger Sequencing

Variants of interest were validated using Sanger sequencing, this was also used in the analysis of candidate genes identified from this study in wider cohorts with the same phenotype. The specific primers are described in the appropriate individual chapters and/or the appendix section III (pages 252-258). All primers were designed to have a universal M13 sequence tag incorporated (F-GTAAAACGACGGCCAGT, R- CAGGAAACAGCTATGAC). This was used as sequencing was performed in the Molecular Genetics Laboratory, Churchill Hospital, Oxford who use this system on all of their diagnostic and research sequencing to

allow sequencing to be performed in an automated higher throughput manner. Primers were designed using the ExonPrimer PERL script within the UCSC genome browser (produced by the Institute of Human Genetics, Helmholz Center, Munich, Germany; http://ihg.gsf.de), the Primer3 tool (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Untergasser et al., 2012), or rarely manually with sequences checked using a basic local alignment search tool (BLAST). All primers used were validated using test DNA samples.

Samples were sequenced in 96 well plates with 20 μl reaction volumes: 10.2 μl KAPA2G fast HS ready-mix™ (KAPA2G Fast HotStart DNA polymerase™, KAPA2G buffer A, dNTPs, $MgCl_2$ and stabilisers), 2.5 μl each of forward and reverse primers (2 μM), 3.8 μl water, and 1.0 μl DNA (50-100 ng/μl). PCR was performed on a BioRad® Dyad thermocycler. The protocol used was 'Fast 60': denaturation at 95°C for three minutes; then cycles of denaturation at 95°C for ten seconds, annealing at 60°C for ten seconds, and extension at 72°C for one second; then a final extension step for 10 minutes at 72°C. This is then sequenced on an automated pathway using a Biomek NXP Laboratory Automation Workstation (Beckman-Coulter®) and an ABI-3730 sequencer (Applied Biosystems®), and analysed using Mutation Surveyor® software (SoftGenetics®).

Chapter 4:

Whole Genome Sequencing in Families with Autosomal Dominant Inherited Heart Diseases

This chapter describes the clinical phenotypes of the five families studied with different autosomal dominant inherited heart diseases. The results of analysis including whole genome sequencing, as described in chapter 3, are presented and discussed.

4.1 A Family with a Novel Repolarisation Abnormality Phenotype

4.1.1 Clinical Information

This family first presented to clinicians as the now deceased male (III④; Figure 4-3) had learning difficulties. In early childhood he was diagnosed as having a duplication of part of chromosome 2 (q11.2-q21) as the cause for this. At sixteen years of age he developed syncopal episodes and an electrocardiogram (ECG) demonstrated extremely unusual repolarisation changes, with widespread ST segment depression in multiple leads (Figure 4-1). An ECG done at eight years of age had not identified these abnormalities, and was within normal parameters.

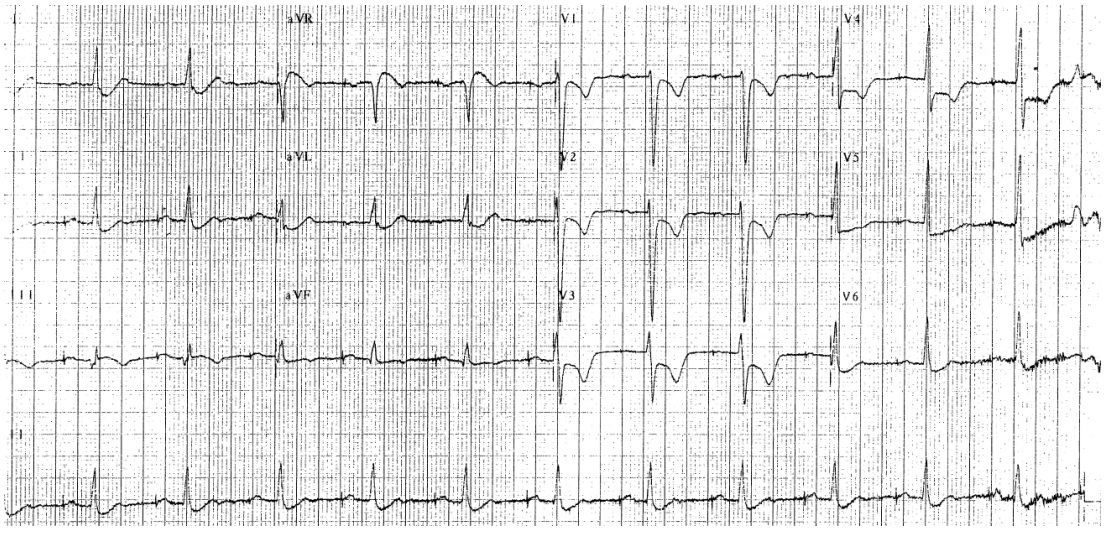


Figure 4-1: ECG demonstrating widespread repolarisation abnormalities in patient III④ across all chest leads.

He subsequently developed difficult to manage episodes of ventricular tachycardia and torsades de pointes. These remained poorly controlled despite maximal anti-arrhythmic medical therapy and insertion of an implantable cardiac defibrillator (ICD). Ventricular dilatation heart and left ventricular dysfunction (left ventricular end-diastolic diameter of 66mm and fractional shortening of 19%) rapidly developed over a period of a few months. He underwent cardiac transplantation but unfortunately died from renal failure in the post-operative period at seventeen years of age.

Cascade screening led to similar ECG changes being identified in his father, uncle, and two half-sisters (ECGs are shown in Figure 4-2, pedigree in Figure 4-3; and further clinical details in Table 4-1). None of these patients have had any symptoms or documented ventricular arrhythmias. His father, II①, and uncle, II②, have both developed atrial fibrillation (AF) between the fifty and sixty years of age. His father has also had mild asymmetrical septal hypertrophy documented from the beginning of his fifth decade, although this has remained relatively static with a maximal septal measurement of fifteen millimetres.

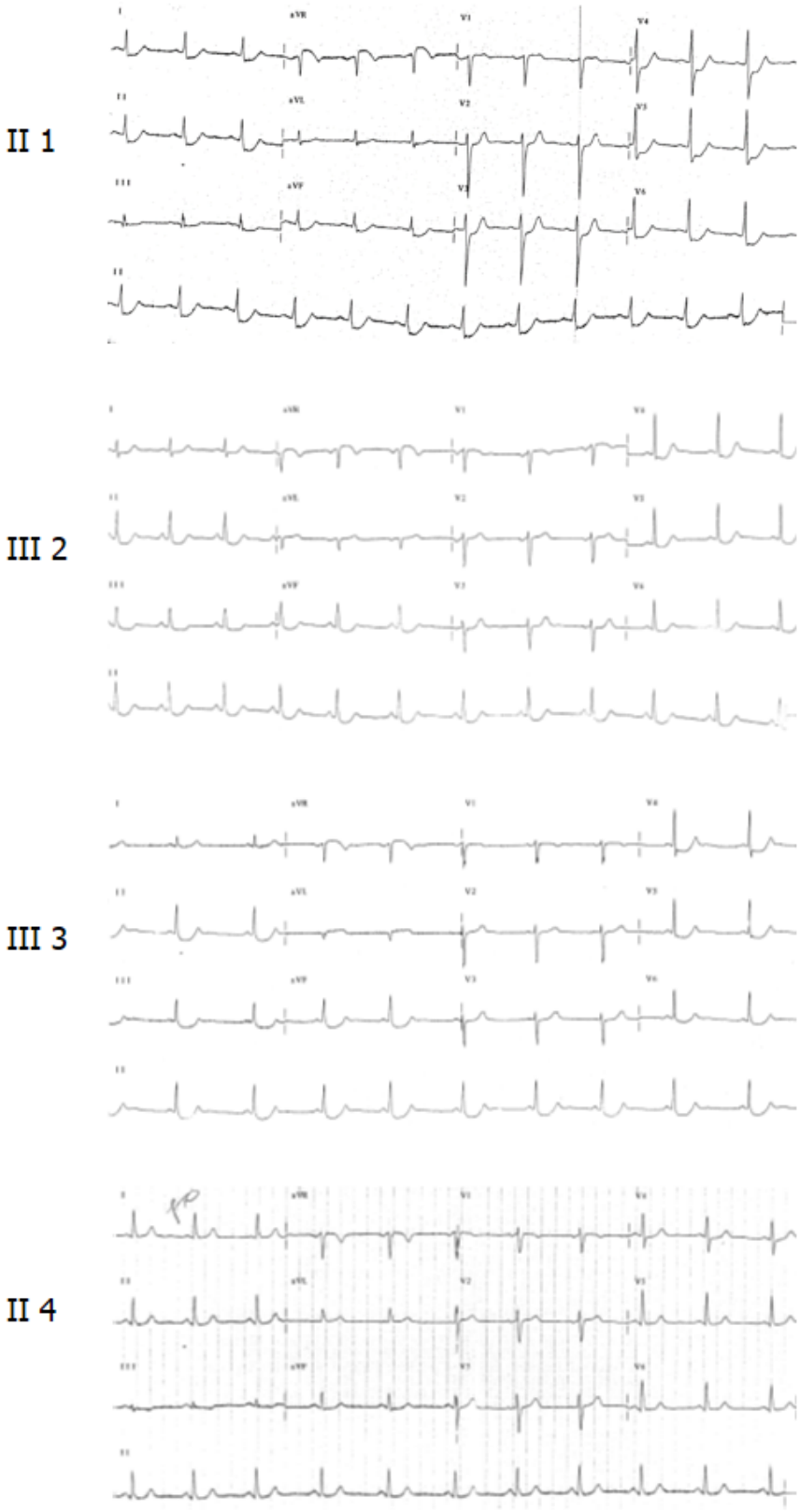


Figure 4-2: ECGs in members of the family with a novel repolarisation abnormality phenotype. II①, III② and III③ all demonstrate widespread repolarisation abnormalities. From the same family, II④, exhibits a normal ECG for comparison.

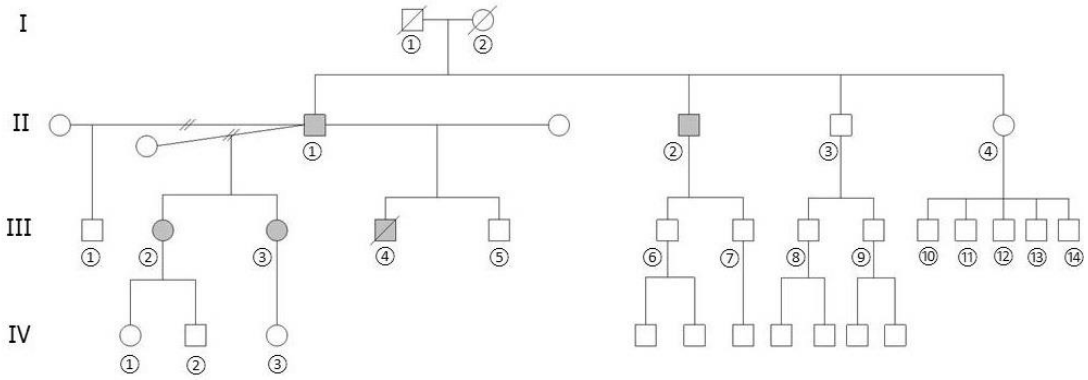


Figure 4-3: Pedigree of family with novel repolarisation abnormalities

Patient	Clinical Information	Summary
II①	Distinctive widespread ST segment depression. Mild cardiac hypertrophy (septum 15mm, posterior wall 11mm) and dilatation (EDD 5.8cm, ESD 4.7cm). Atrial fibrillation.	Affected
II②	Distinctive widespread ST depression. Atrial fibrillation.	Affected
II③	Normal ECG.	Likely unaffected
II④	Normal ECG. Atrial fibrillation in early 60's.	Unclear
III①	Normal ECG.	Likely unaffected
III②	Distinctive widespread ST segment depression. Borderline septal hypertrophy (11mm).	Affected
III③	Distinctive widespread ST segment depression.	Affected
III④	Distinctive widespread ST segment depression from teenage (normal aged eight years). From the age of sixteen years he had recurrent torsades-de-pointes and VF despite ICD, attempted ablation, and medical therapy. He developed left ventricular dilatation and dysfunction and died aged seventeen following cardiac transplantation. Developmental delay and mild dysmorphism (*de novo* 2q21 duplication).	Affected
III⑤	Normal ECG.	Likely unaffected
III⑥	Normal ECG.	Likely unaffected
III⑦	Normal ECG.	Likely unaffected

Table 4-1: Clinical features in members of the family with novel repolarisation abnormalities (those definitely affected are highlighted in black typeface). Acronyms: ECG = electrocardiogram; EDD = end-diastolic diameter; ESD = end-systolic diameter; ECG = electrocardiogram; VF = ventricular fibrillation; ICD = implantable cardiac defibrillator.

4.1.2 Repolarisation

4.1.2.1 Physiology of Repolarisation

Repolarisation is the return of cardiomyocytes to their negative resting potential following the depolarisation of the cardiac action potential. The change in electrochemical gradients, which propagate the action potential, are achieved by fluctuations in sodium, potassium,

chloride and calcium ions through specific ion channels. The channels important in mediating repolarisation are types of voltage-gated potassium channel. The main phase of repolarisation (phase 3) is due to the opening of the rapid delayed rectifier channels (I_{Kr}) and slow delayed rectifier channels (I_{Ks}) (Figure 4-4). The strong inward rectifier channels, I_{K1}, are open throughout and therefore help regulate the threshold and rate of repolarisation, as well as stabilising the resting membrane potential (Nerbonne and Kass, 2005, Ravens and Cerbai, 2008).

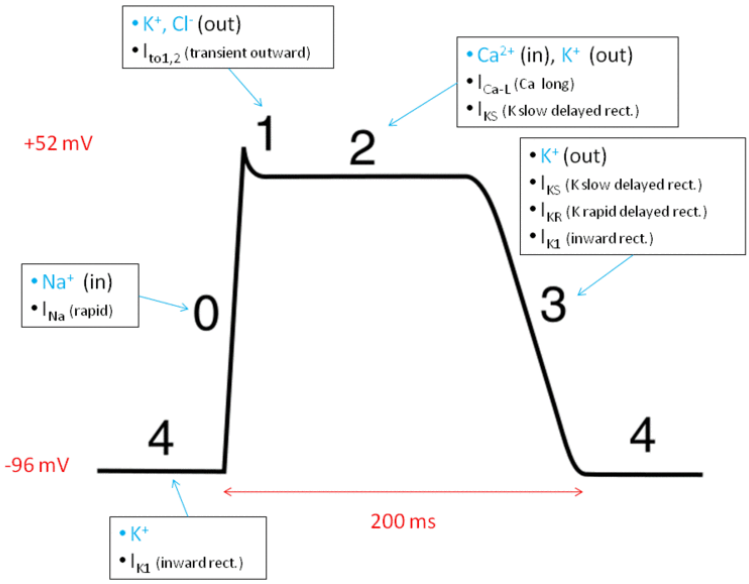


Figure 4-4: Phases of the cardiac action potential. The main depolarisation phase (phase 3) is due to efflux of potassium ions through delayed rectifier and inwardly rectifying channels.

A particular type of inward rectifying channel is the ATP-sensitive K^+ current, I_{KATP}. These channels are formed from the pore-forming K_{IR} subunits, encoded by *KCNJ8* and *KCNJ11*, and regulatory sulphonylurea subunits (SUR 1 and 2), encoded by *ABCC8* and *9*. They are inhibited by ATP and sulphonylureas (Nichols, 2006). Murine ventricular K_{ATP} channels are predominantly composed of Kir6.2 (*KCNJ11*) and SUR2A (*ABCC9*) subunits, but in man both K_{ir} subunits and both sulphonylurea subunits may be important with regional differences in expression (Nichols et al., 2013). I_{KATP} channels are highly expressed in the heart but are thought to be inactive in the normal state, opening when the levels of ATP fall. The roles of

these channels in the heart are still being fully elucidated but they are thought to link cellular metabolism with the membrane potential. They are activated in ischaemia, accelerating repolarisation and shortening the QT interval, and may act to reduce cellular damage. Activation of these channels induces ST elevation in the absence of ischaemia (Flagg et al., 2010).

4.1.2.2 Abnormalities of Repolarisation

Although long QT syndrome can be considered a disorder of repolarisation it has a markedly different phenotype to that seen in this family. The most similar reported group of disorders are the J-wave abnormalities. The QRS-ST junction is termed the J-point and the deflection immediately after the QRS, signifying the start of repolarisation, is termed the J-wave. The best characterised of the 'J-wave syndromes' is Brugada syndrome but phenotypically this family most resembles an early repolarisation syndrome caused by mutations in *KCNJ8,* encoding an α-subunit of the I_{KATP} channel. Patients have variable ECG abnormalities showing widespread J-point deviation with some ST segment deflection, most marked in the inferolateral leads (Figure 4-5) (Haissaguerre et al., 2009, Medeiros-Domingo et al., 2010).

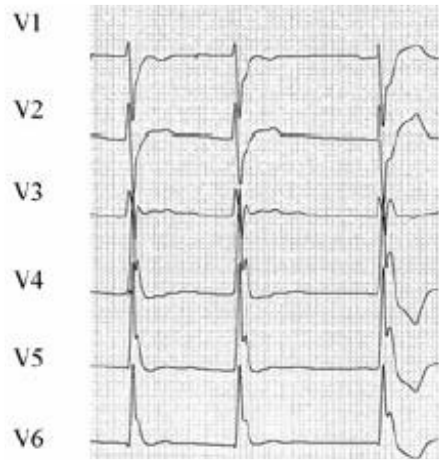


Figure 4-5: ECG showing chest lead traces from a patient with early repolarisation syndrome due to a mutation in KCNJ8 showing deviation of the J-pint and inferolateral ST deflection (Haissaguerre et al., 2009)

4.1.3 Genetic Studies Performed

Prior to enrolling in this study family members had had clinical diagnostic testing of those genes usually associated with LQTS (*KCNQ1, KCNH2, SCN5A, KCNE1,* and *KCNE2*), and due to the presence of hypertrophy also those genes associated with this (*MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2* and *GLA*). Because of the similarities in clinical phenotype to the J wave syndromes genes encoding the K_{ATP} subunits, *KCNJ11*, *ABCC9*, and *KCNJ8* (Primers used are shown in appendix section III-e, page 256), were also sequenced. No variants of likely pathogenicity were identified in any of these genes.

4.1.4 Results

4.1.4.1 WGS with Segregation Data

Family members who had whole genome sequencing (WGS) and/or SNP array are shown in Figure 4-6. Following analysis of the results from these in affected family members, as described previously, seven variants remained (Table 4-2; and marked on linkage plots shown in Figure 4-7).

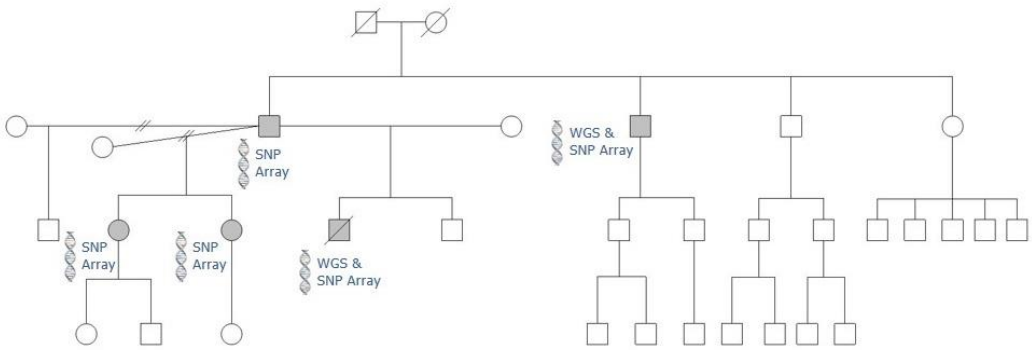


Figure 4-6: Pedigree of family with novel repolarisation abnormalities indicating who had whole genome sequencing and SNP array

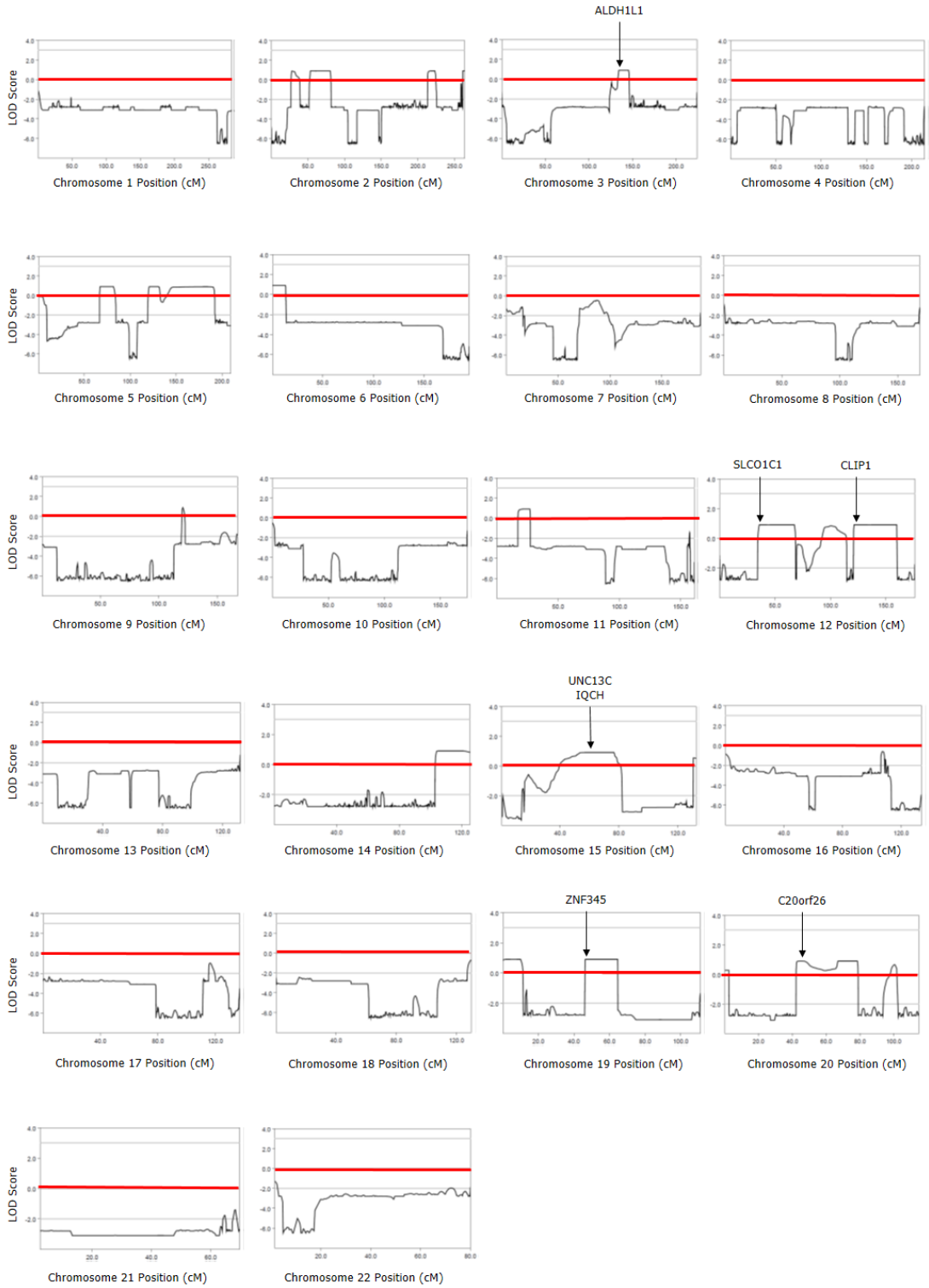


Figure 4-7: Linkage plots for each chromosome from SNP array data analysed in MERLIN software are shown (Abecasis et al., 2002). Genes containing variants identified from WGS studies are highlighted. Red lines indicate a LOD score of 0.

Gene	Variant	Conservation (PhastCons/PhyloP)	Other
ALDH1L1	p.A260P	Primates only (0/0.45)	Present in unaffected adult relative
C20orf26	p.R590S	Human and chimp (0.99/2.38)	SNP at same nucleotide; Present in unaffected adult relative
CLIP1	**p.A1266T**	**Mammals except rat (0.9/0.37)**	
IQCH	p.H466R	Primates, rat, mouse and frog (0/-0.06)	Once in EVS
SLCO1C1	p.I165C	Primates, dog and cat (0.02/1.01)	Present in unaffected adult relative
UNC13C	p.Y2165C	Primates, rat and mouse (0.54/1.42)	Present in unaffected adult relative
ZNF345	p.K457E	Mammals (0.02/0.12)	Present in unaffected adult relative

Table 4-2: Variants identified following WGS and SNP array in family with repolarisation abnormlities phenotype (more likely cause of the phenotype is highlighted in bold)

Summary of the known roles of the seven genes in which variants have been identified:

p.A260P in Aldehyde dehydrogenase 1 member L1 (ALDH1L1)

Encodes 10-formyltetrahydrofolate dehydrogenase (FTHFD) (Hong et al., 1999). Mice lacking FTFHD are phenotypically normal but have reduced fertility (Champion et al., 1994). Expressed widely, including at low levels in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.R590S in the open reading frame C20orf26

An open reading frame of unknown function. It is expressed in several tissues, but has not been identified as being expressed in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.A1266T in Restin (CLIP1)

Restin binds endocytic carrier vesicles to microtubules. It is named restin as it has been recognised as being overexpressed in Reed-Sternberg cells (Bilbe et al., 1992). It was initially thought only to be expressed in non-Hodgkin lymphomas and not in normal tissues (Delabie et al., 1992), but it actually appears to be widely expressed, although at low levels

in many tissues including the heart (Uhlen et al., 2010). Restin has effects on dynactin and end-binding 1 (EB1) at microtubule plus ends (Goodson et al., 2003). These interactions have roles in the action of the dynein motor complex which acts to traffic several proteins to the cell membrane, including potassium channels (Choi et al., 2005) and the connexin hemi-channels of gap junctions which are important for coordinated depolarisation in the heart (Shaw et al., 2007).

p.H466R in IQ motif containing-protein H (IQCH)

This protein contains a carbamate kinase-like domain, an IQ domain, a zinc-dependent exopeptidase domain, and a lactate dehydrogenase domain (Yin et al., 2005). It is thought to play a role in spermatogenesis (Nie et al., 2013). It is predominantly expressed in testes, but is also expressed in some other tissues including the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.I165C in Member 1C1 of the solute carrier organic anion transporter family (SLCO1C1)

SLCO1C1 is a transmembrane transporter of a select group of organic anions including thyroxine (Pizzagalli et al., 2002). It is expressed predominantly in the nervous system, and is not identified as being expressed in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.Y2165C in Homologue of C. elegans UNC13 C (UNC13C)

In *C. elegans* Unc13 is important for synaptic vesicle fusion at neuromuscular junctions (Richmond et al., 1999). UNC13C is predominantly expressed in the brain and knock-out mice have a motor learning deficit (Augustin et al., 2001). It is predominantly expressed in the nervous system, and is not known to be expressed in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.K457E in Zinc-finger protein 345 (ZNF345)

ZNF345 is a zinc-finger protein of unknown function. Fairly widely expressed but not reported to be expressed in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

Interpretation of variant findings

Prior to the study we had hypothesised that the pathogenic mutation would be in a gene encoding a cardiac expressed ion channel or a related protein, but there is no clear candidate fitting this hypothesis. The best candidate, based on known biological function of the proteins produced, is *CLIP1*. Although it is widely expressed, and is involved in the transport of a number of proteins, the recognition of a role in transporting ion channels and gap junction proteins to the cell membrane could account for the repolarisation phenotype seen. The other variants are unlikely causes of the phenotype, predominantly determined on the basis of known unrelated biological functions, but the lack of significant cardiac expression and finding in the unaffected family member and/or at low levels in population data also provide evidence against pathogenicity of these.

3.3.5.1 Follow-Up Studies

Sequencing of the *CLIP1* variant confirmed its presence and segregation in all affected family members (the primers used are described in appendix section III-b, page 254). A degree of further support was provided by the finding that it was also absent in two clinically unaffected adult relatives (II④ and III①).

As this is such a rarely observed phenotype it has been difficult to identify other patients to assist with ascertaining pathogenicity. Through collaboration we have identified a single further, apparently unrelated, patient with a very similar phenotype. Her ECG shows comparable repolarisation abnormalities (Figure 4-8) and she has had multiple syncopal

episodes and ventricular arrhythmias. No other family members are known to have the same phenotype, and there is no family history of cardiac syncope or premature death.

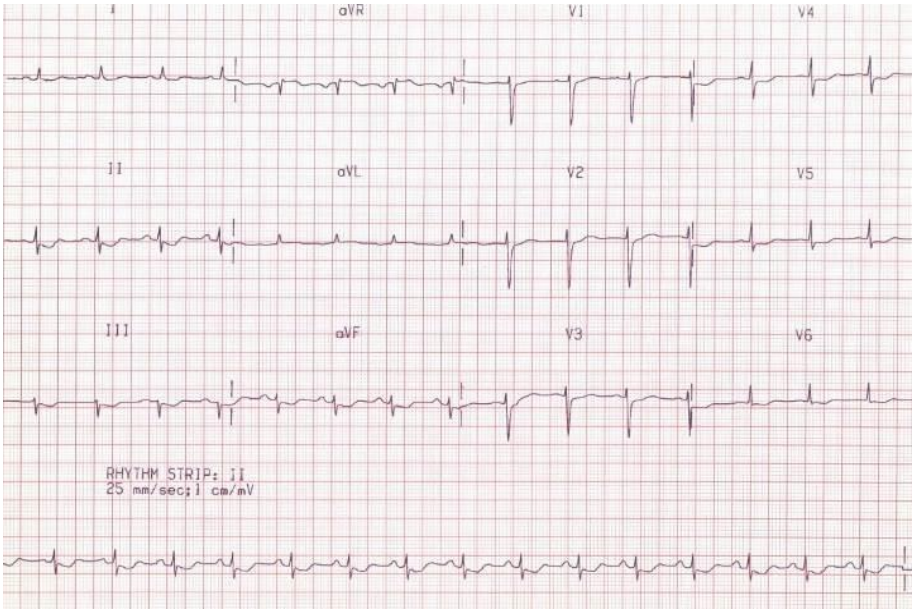


Figure 4-8: ECG from patient with similar widespread ST depression

Sanger sequencing of all coding exons of *CLIP1* in this patient did not identify any variants other than known polymorphisms. Exome sequencing was undertaken to adequately assess the other variants identified in our family, along with other candidate genes. This was performed by the High-Throughput Sequencing at the Wellcome Trust Centre for Human Genetics, University of Oxford in a similar manner to the strategy for whole genome sequencing described in section 3.3.2 (page 52).

The data produced from exome sequencing has been analysed in a targeted manner. As well as all of the genes identified by the study in our family described above (*ALDH1L1*, *C20orf26*, *CLIP1*, *IQCH*, *SLCO1C1*, *UNC13C*, *ZNF345*), selected other genes which may cause a repolarisation or arrhythmic phenotype, based on current knowledge, have been analysed (*KCNQ1*, *KCNH2*, *KCNE1*, *2*, *3* and *5/1L*; *KCNJ2*, *5*, *8* and *11*; *KCNIP3*; *KCND3*, *KCNG2*, *ANCC8* and *9*, *AKAP9*, *SCN5A*, *SCN7A*, *SCN1B*, *RYR2*, *CASQ2*, *PLN*, *CACNA1C*, *CACNA2D1*, *CACNB2*, *ATP2A1* and *2*, *PKP2*, *DSG2*, *DSP*, *DSC2*). Although no variants in the

genes highlighted by our study were identified, a novel missense variant in *RYR2*, p.E4659G, has been identified.

RYR2 localises to the sarcoplasmic reticulum and forms a tetrameric calcium channel that regulates the release of stored calcium to initiate contraction in response to cell excitation. The classical phenotype caused by mutations in *RYR2* is catecholaminergic polymorphic ventricular tachycardia (CPVT) (Laitinen et al., 2001); although arrhythmogenic right ventricular cardiomyopathy (ARVC) type 2 (Tiso et al., 2001) and LQTS (Kauferstein et al., 2011) phenotypes have also been reported rarely (the location of pathogenic mutations is shown in Figure 4-9). This variant is not reported in the WGS500 samples, population databases, or as a disease causing mutation. Although ventricular arrhythmias are within the known spectrum of *RYR2* mutations, fixed repolarisation abnormalities have not been reported.

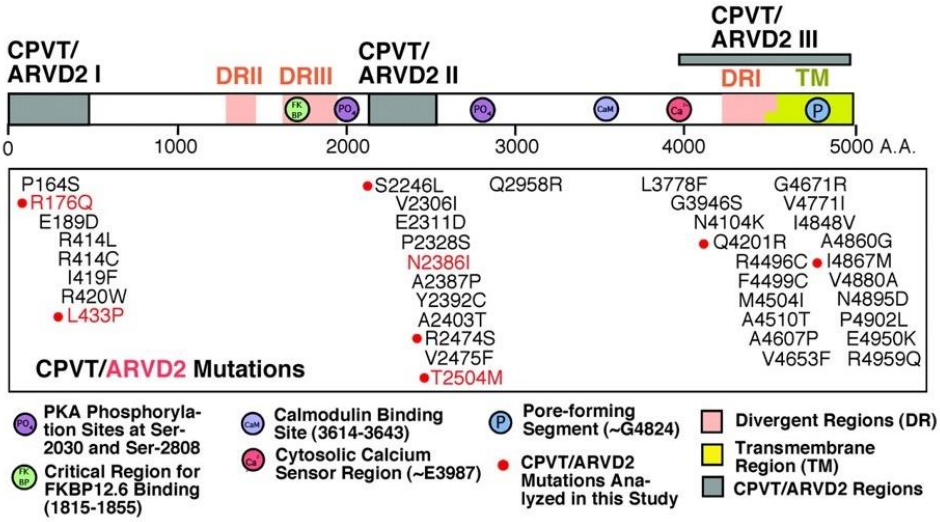


Figure 4-9: A schematic of the distribution of disease causing mutations in *RYR2* (Jiang et al., 2005)

4.1.5 Conclusions

The variant in *CLIP1* remains a possibility for causing the phenotype seen, although further evidence for this has not been obtained. Due to the relative paucity of evidence surrounding the cardiac roles of this protein this has not been pursued further at present.

The presence of such a rare phenotype significantly limits the ability to identify further patients with a similar phenotype.

4.2 A Family with Autosomal Dominant Dilated Cardiomyopathy

4.2.1 Clinical Information

This family exhibits a dilated cardiomyopathy (DCM) phenotype with very variable age of onset and clinical course. Patient II⑥ (pedigree is shown in Figure 4-11) presented with peri-partum cardiomyopathy at thirty years of age. She remained well for the next two decades, although over the past ten years she has developed progressively worsening dilatation and heart failure. The condition was only recognised as being familial after patient IV④ presented in infancy with DCM. She has had a very variable clinical course, with predominantly good health but has recently required cardiac transplantation at the age of twelve years (pre-transplantation echocardiogram image is shown in Figure 4-10).

Cascade screening in the family after identification of cardiomyopathy in this girl identified other family members with DCM. Her mother, III②, was identified as having DCM, and despite having been well up to that point then developed rapidly progressive dilatation and heart failure, leading to death over only a few years from diagnosis. Clinical details of other family members are provided in Table 4-3.



Figure 4-10: Echocardiogram of IV④ prior to cardiac transplantation (aged twelve years) showing significant dilatation of the left ventricle

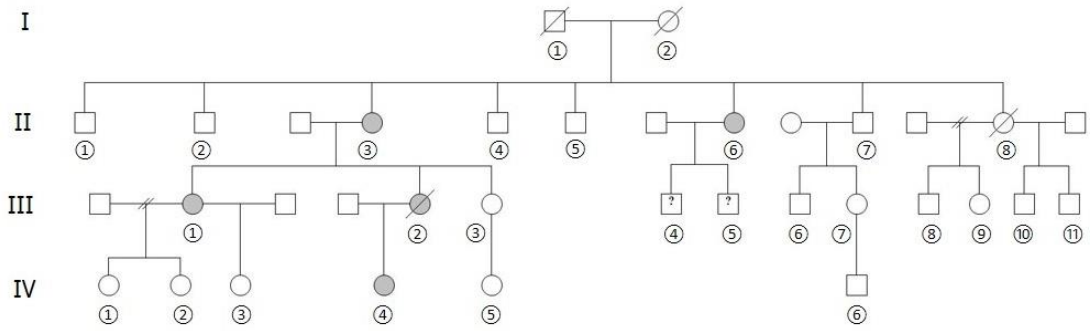


Figure 4-11: Pedigree of family with dilated cardiomyopathy

Patient	Clinical Information	Summary
I①	Died from emphysema aged 73.	Unknown
I②	Died aged 78 of non-cardiac causes.	Unknown
II①	No information known	Unknown
II②	No information known	Unknown
II③	SOB and orthopnoea from age 40. Chest pain but normal coronaries on angiogram. Severe mitral regurgitation due to LV dilatation, mild LV and LA dilatation and mild systolic impairment on echocardiogram.	Affected
II④	Asymptomatic and normal screening at ages 46 and 54.	Probably unaffected
II⑤	Asymptomatic and normal screening at age 46.	Probably unaffected
II⑥	Diagnosed with postpartum cardiomyopathy at age 29 although no discernable dilatation or dysfunction on echo. TWI in anterior chest leads. On review at age 51 had significant LV dilatation and reduced ejection fraction, with mild MR and a dilated LA.	Affected
II⑦	Asymptomatic, not screened.	Unknown
II⑧	Died in road traffic accident aged 35.	Unknown
III①	LV dimensions variably measured as high normal to mildly dilated, and from normal to mild systolic dysfunction. Large voltages on ECG.	Probably affected
III②	Died aged 29 of heart failure. Diagnosed aged 19 with significant LV dilatation and moderate-severe impairment. Significant variability in LV function in subsequent years with general improvement. Symptoms often more significant than cardiac function would suggest. Rapid deterioration in the year before death following an episode of gastroenteritis.	Affected
III③	Asymptomatic. Borderline LV dilatation during pregnancy but all other imaging has shown LV dimensions at high normal levels.	Possibly affected
III④	Asymptomatic. LV dimensions and septal thickness at upper limit of normal.	Possibly affected
III⑤	Asymptomatic. LV dimensions at upper limit of normal.	Possibly affected
III⑥-III⑪	Asymptomatic. Not screened.	Unknown
IV①	Asymptomatic. Normal screening at age 20.	Possibly unaffected
IV②	Asymptomatic. LV dimensions at upper limit of normal with slightly thinned walls.	Possibly affected

IV③	Asymptomatic and unscreened in infancy.	Unknown
IV④	Presented in infancy with severe DCM and heart failure. Improvement and relatively stable through early childhood. Deterioration in year leading to cardiac transplantation at age 12.	Affected
IV⑤	Asymptomatic and unscreened in infancy.	Unknown
IV⑥	Asymptomatic and unscreened in early childhood.	Unknown

Table 4-3: Clinical features of members of the family with likely autosomal dominant dilated cardiomyopathy (definitely affected family members in black typeface). Abbreviations: DCM = dilated cardiomyopathy; ECG = electrocardiogram; LV = left ventricle; LA = left atrium; MR = mitral regurgitation; SOB = shortness of breath; TWI = T-wave inversion.

4.2.2 Dilated Cardiomyopathy

Dilated cardiomyopathy (DCM) is the second most common of the cardiomyopathies, affecting approximately one in two-and-a-half thousand people (Codd et al., 1989). It is a primary disorder of heart muscle characterised by left ventricular dilatation and reduced function. Dilated cardiomyopathy is one of the main causes of heart failure, and may also cause other complications such as arrhythmias and thromboembolism (Luk et al., 2009). Although most commonly presenting in early adulthood there is a bimodal distribution with the other peak in incidence occurring in infancy (Daubeney et al., 2006).

DCM can have a variety of causes including myocarditis, toxins (such as alcohol and certain drugs), infiltrative disorders (i.e. amyloid and sarcoid), ischaemic, valvular or structural heart disease, chronic anaemia or hypertension, peripartum cardiomyopathy (which is itself a heterogeneous disorder some of which has a monogenic basis), and genetic causes (Luk et al., 2009).

Up to a half of patients with DCM can be shown to have familial disease, even in these patients the causal mutation is only identified in about a quarter (Ackerman et al., 2011). There is considerable genetic heterogeneity with mutations in genes encoding proteins of the sarcomere, Z-disc, ion channels, and calcium-handling mechanisms all recognised to cause highly penetrant dominantly-inherited DCM (Karkkainen and Peuhkurinen, 2007). Recently truncating mutations in *TTN*, the gene encoding titin, have been recognised as a

significant cause of DCM, with early studies suggesting this may be the one gene identified to date that may account for a significant proportion of this condition (Herman et al., 2012). Certain mutations in the lamin A/C gene cause DCM with conduction disease (Fatkin et al., 1999). Mutations causing dysfunction of the dystrophin complex and mitochondrial function cause DCM with generalised myopathy, with or without additional features (Karkkainen and Peuhkurinen, 2007). Multisystem lower-penetrance genetic disorders such as haemochromatosis may also give a DCM phenotype (Luk et al., 2009). Due to this heterogeneity genetic testing in DCM tends to be restricted to certain sarcomeric genes in familial DCM, and specific genes when other features indicate.

4.2.3 Genetic Studies Performed

Clinical diagnostic testing of sarcomeric genes which can cause different cardiomyopathies (*MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2* and *GLA*) and also the mitochondrial 3243A>G mutation, was performed prior to enrolment in this study, with no likely pathogenic variants identified.

4.2.4 Results

4.2.4.1 WGS with Segregation Data

Family members who had whole genome sequencing and/or SNP array are shown in Figure 4-12. Following analysis of variants produced from WGS and SNP array in definitively affected family members, as described previously, nine variants remained (Table 4-4; chromosome linkage plots with genes in which variants have been identified are indicated in Figure 4-13).

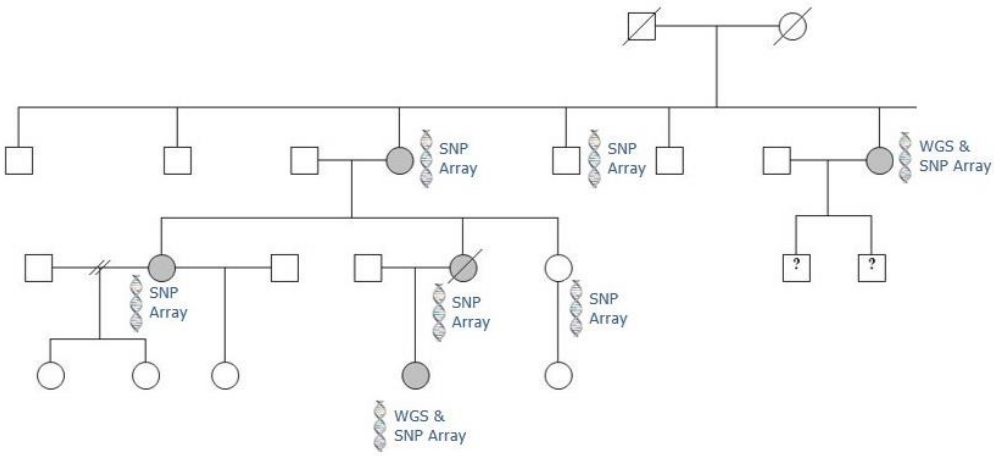


Figure 4-12: Partial pedigree indicating which family members in the family with a DCM phenotype had WGS and/or SNP array

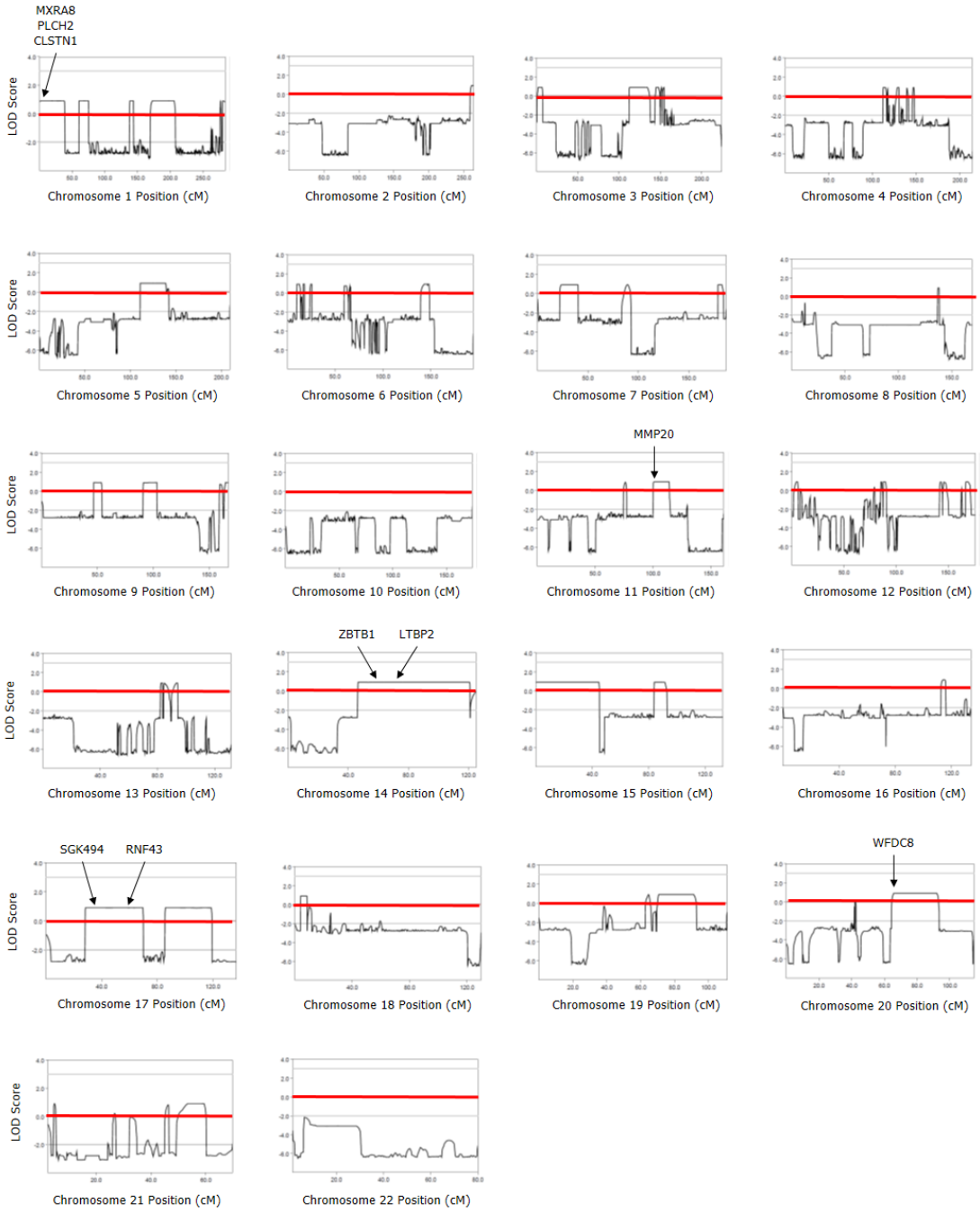


Figure 4-13: Linkage plots for each chromosome from SNP array data analysed in MERLIN software are shown (Abecasis et al., 2002). Genes containing variants identified from WGS studies are highlighted. Red lines indicate a LOD score of 0.

Gene	Variant	Conservation (PhastCons/PhyloP)	Other
CLSTN1	p.R416W	All higher organisms (1.0/2.06)	Once in EVS; Present in clinically unaffected family member
LTBP2	p.A1727T	All mammals (0.4/3.03)	Once in EVS; Present in clinically unaffected family member
MMP20	p.G80C	All mammals (0.99/3.19)	Present in clinically unaffected family member
MXRA8	**p.D431H**	**All except tetraodon (0.98/2.55)**	
PLCH2	p.T993M	Primates only (0.00/-1.89)	
RNF43	**p.F375L**	**Mammals (0.99/1.09)**	Present in clinically unaffected family member
SGK494	p.R252H	All except yeast (0.71/-0.12)	Present in clinically unaffected family member
WFDC8	p.C229S	All mammals (0.98/2.79)	Present in clinically unaffected family member
ZBTB1	p.605_606 del	Highly conserved region	Present in clinically unaffected family member

Table 4-4: Variants identified after WGS and SNP array in affected family members in family with autosomal dominant DCM. Those variants felt more likely to be the cause of the phenotype are highlighted in bold typeface.

Summary of the known roles of the nine genes in which variants have been identified:

p.R416W in Calsyntenin 1 (CLSTN1)

CLSTN1 is a postsynaptic protein of neurons and is involved in axonal transport (Vogt et al., 2001, Vagnoni et al., 2012). It is predominantly expressed in the nervous system but is widely expressed at low levels, including the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.A1727T in Latent transforming growth factor-beta binding protein (LTBP2)

LTBP2 is an extracellular matrix protein which is a member of the transforming growth factor binding protein family, including fibrillin-1 (Moren et al., 1994). Homozygous mutations in this gene have been identified causing various ophthalmic abnormalities and Weill-Marchesani syndrome (Kumar et al., 2010, Haji-Seyed-Javadi et al., 2012). This protein is expressed at highest levels in parts of the eye but is reported to be widely expressed at low levels, including in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.G80C in Matrix metallopeptidase 20 (MMP20)

Matrix metallopeptidases are a group structurally similar zinc-dependent endopeptidases that degrade components of the extracellular matrix. MMP20 has been implicated in dental health and abnormalities (Llano et al., 1997). It is expressed in dental issues only (Husar Bioinformatics Lab, April 2013).

p.D431H in Matrix-remodelling associated protein 8 (MXRA8)

MXRA8, also known as limitin or limitrin, is an immunoglobulin superfamily protein. Little is known about the function of MXRA8, although it has been shown to be upregulated in the skeletal muscle of cattle which exhibit increased fat marbling (Clark et al., 2011a). It is expressed in the heart, among a number of other tissues (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.T993M in Phospholipase C-eta 2 (PLCH2)

PLCH2 is a phosphoinositide-specific phospholipase C that cleaves phosphatidylinositol 4,5-bisphosphate to second messengers of the inositol lipid signalling pathways, playing a role in intracellular calcium release and protein kinase C activation (Stewart et al., 2005). The mouse homologue is expressed only in neuronal tissue (Nakahara et al., 2005). PLCH2 is also not known to be expressed in the heart (Liu et al., 2008, Husar Bioinformatics Lab, April 2013).

p.F375L in Ring-finger protein 43 (RNF43)

RNF43 is a ubiquitin ligase that induces endocytosis on Wnt receptors, therefore acting as a tumour suppressor (Koo et al., 2012). It is overexpressed in certain tumour lines but is also endogenously widely expressed, including at low levels in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008). Although identified for its roles in carcinogenesis this pathway also has roles in embryonic development and other aspects of cell proliferation, migration and fate (Heuberger and Birchmeier, 2010). Suppression of the Wnt/β-catenin

signalling pathway has been observed in arrhythmogenic right ventricular cardiomyopathy (Garcia-Gras et al., 2006), a cardiomyopathy that is usually distinct from DCM but has some similar features.

p.R252H in a homologue of a mouse protein kinase domain-containing protein (SGK494)

SGK494 is an uncharacterised protein kinase domain-containing protein (Manning et al., 2002). It is widely expressed, particularly in the GI tract, but also at low levels in the heart (Uhlen et al., 2010).

p.C229S in Whey acidic protein 4-disulphide core domain 8 (WFDC8)

WFDC8 is a serine protease inhibitor with roles in the immune function of the male reproductive tract (Ferreira et al., 2011, Rajesh et al., 2011). It is expressed in the testes, and is not known to be expressed in the heart (Liu et al., 2008).

p.605_606del in Zinc-finger and btb domain containing protein 1 (ZBTB1)

ZBTB1 is a transcriptional regulator of lymphoid T-cell development (Siggs et al., 2012). It is widely expressed, including in the heart (Liu et al., 2008).

Interpretation of variant findings

Of the nine variants identified in affected family members none are clearly related to the multitude of known causes of DCM. Indeed based on known biological function none really connect to a DCM phenotype. RNF43 is connected to the Wnt signalling pathway. This pathway has diverse roles and has been associated with a different form of cardiomyopathy, this is therefore highlighted as a possible cause of the phenotype despite this connection to cardiomyopathy being rather tenuous. Almost all of the identified variants are in highly conserved residues. The variants in *CLSTN1* and *LTBP2* are identified in population data, which is evidence against pathogenicity. Most of the variants are

identified in an unaffected family member, this is further evidence against pathogenicity but not conclusively so.

One of the two variants identified as being a possible cause of this phenotype is a variant in *MXRA8*. Although little is known about this gene one thing that makes it potentially more interesting is that it is located at 1p36.33. This locus is commonly deleted in a recurrent chromosomal deletion syndrome of 1p36 which has a particularly high prevalence of cardiomyopathy, both LVNC and DCM. Although many of the features of 1p36 microdeletion syndrome are consistent the specific chromosomal region deletion is quite variable. Most are terminal deletions but no common breakpoint is recognised, with breakpoints being identified between 0.5 and over 10 Mb from the telomere (Gajecka et al., 2007). The critical deleted region in those with cardiomyopathy is unknown. The genes present in the more common terminal deletions are shown in Figure 4-14.

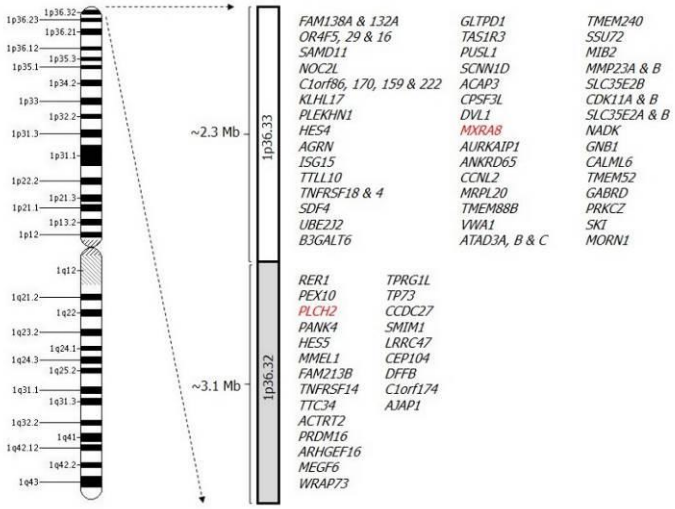


Figure 4-14: Genes in the 1p36 terminal region highlighting the identified candidate gene *MXRA8* in red. *PLCH2* is also highlighted as a variant in this gene was also identified in this family, although known functions make it a less plausible candidate for causing DCM.

4.2.4.2 Follow-Up Studies and Analysis

MXRA8

The variant in *MXRA8* was confirmed to be present and segregating in all affected family members (the chromatogram from Sanger sequencing to confirm the mutation is shown in Figure 4-15). It was also shown to be absent in a clinically unaffected adult family member, II④. Through collaboration with Professor Stuart Cook (Imperial College, London) DNA from seventy-nine patients with dilated cardiomyopathy was obtained. Full sequencing of all exons of *MXRA8* was performed (the primers used are described in appendix section III-f, page 256), although exons three to five failed to amplify with the test samples. The quantity of DNA provided was limited such that repeat testing was not possible. In exons sequenced no variants, other than known polymorphisms, were identified in this DCM cohort.

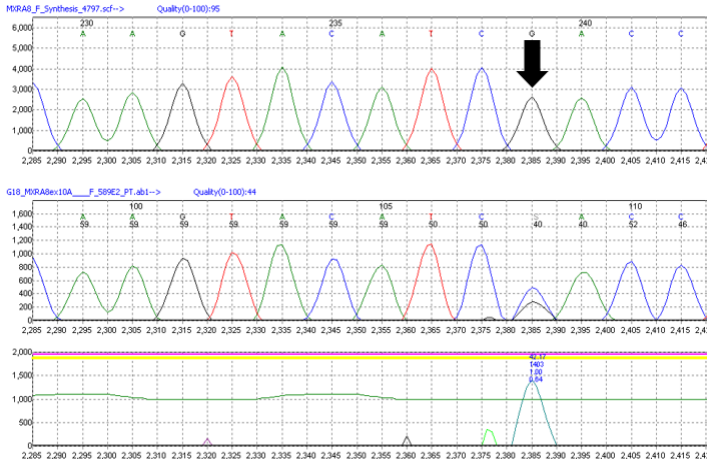


Figure 4-15: Chromatogram confirming presence of the identified *MXRA8* variant in the family with a DCM phenotype (produced by Mutation Surveyor software). Top line shows the reference sequence and the middle line shows the test sequence using the forward primer.

Recently an alternative gene is the 1p36 critical region, *PRDM16*, has been identified as a cause of DCM (Arndt et al., 2013). Although it is not inconceivable that more than one gene in this region may be responsible for the same phenotype, this makes the chromosomal location of *MXRA8* a less compelling factor to support a possible role in pathogenesis. Sequencing data on *PRDM16* from whole genome sequencing has been reanalysed in this

family and no exomic variants identified. Seven Intronic variants are identified, three not reported in other WGS500 samples or population data (GRCh37: 1_3033036:C>T, 1_3275150:G>A, 1_3345247:C>A). None of these are in highly conserved regions strongly suggestive of regulatory function.

Mitochondrial genome analysis

As the inheritance pattern in this family is consistent with maternal/mitochondrial inheritance, and mitochondrial mutations have been recognised as a cause of DCM (Arbustini et al., 1998), mitochondrial DNA variants were studied from both the WGS data and from specific mitochondrial genome sequencing, performed as in section 2.2.2 (page 29). There was a high level of concordance between the two data sets, but only known polymorphisms in the mitochondrial genome were identified.

Non-coding variant analysis

Due to the lack of recognition of a high probability causative variant from those identified some of the non-coding genomic sequence was analysed. Due to the difficulty in interpreting the majority of non-coding sequence we targeted highly conserved known regulatory regions and sequence within five kilobases of putative DCM associated genes. The only segregating variant in these regions, using the same filters as for exomic variants, was on chromosome 14 (14_104,146,002:G>A; NCBI genome build GRCh37). The nearest gene to this is *KLC1*, which encodes kinesin light-chain 1. There is no biological function evidence to associate this gene with a DCM phenotype.

4.2.5 Conclusions

These studies leave this family without a clear candidate from the gene variants identified. Although one of these may be the cause, just with insufficient evidence for pathogenicity based on current understanding, this raises the possibility that the cause is not amongst

these. Potential reasons for the failure of exome sequence analysis to identify a causative mutation are discussed further in section 6.2.3.1 (page 204).

4.3 A Family with a Combined Phenotype of Different Cardiomyopathies, Electrophysiological Abnormalities and Septal Defects

4.3.1 Clinical Information

This family first came to clinical attention when patient II③ (Figure 4-17) died at fifteen years of age following cardiac transplantation for dilated cardiomyopathy. Her sister, II②, presented at a similar time due to syncopal episodes and was diagnosed with long QT syndrome (LQTS). At thirty years of age she developed atypical chest pain, shortness of breath on exertion, and syncopal episodes. Investigation at this time demonstrated that, in addition to the prolonged QT interval, she also had other ECG abnormalities; partial right-bundle branch block, ST-elevation in leads V1-3, dominant R-wave in V1, PR interval at the upper limit of normal (Figure 4-16). She also had evidence of mild asymmetrical left ventricular hypertrophy on echocardiography, with a maximal septal wall thickness of fourteen millimetres.

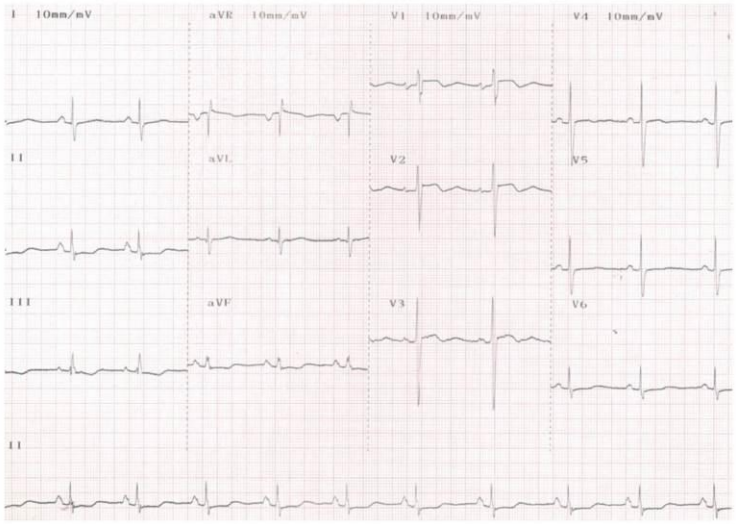


Figure 4-16: ECG from patient II 2 showing prolonged QT interval, partial RBBB, ST-elevation V1-3, dominant R-wave in V1, and PR interval at upper limit of normal

This led to cascade screening initially in her sister, II①, and then subsequently to each of their children (pedigree shown as Figure 4-17). The other affected family members have evidence of varying combinations of cardiomyopathy, electrophysiological and congenital heart disease phenotypes. Full clinical information are given in Table 4-5. In addition to the cardiac phenotype some have varying degrees of learning difficulties, mild dysmorphism and possible epilepsy/seizures.

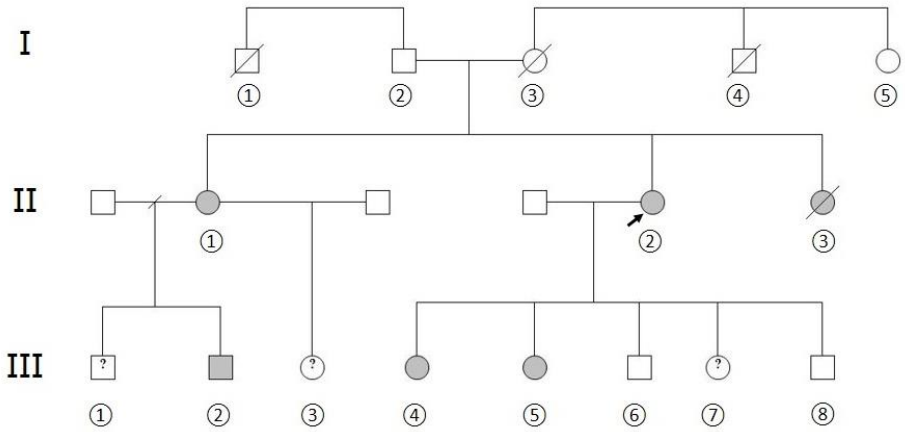


Figure 4-17: Pedigree of family with differing cardiomyopathies, electrophysiological abnormalities and septal defects. Although the inheritance pattern is suggestive of an autosomal dominant condition it is also consistent with X-linked or maternal inheritance.

Patient	Clinical Information	Summary
I①	Died aged 24, cause unknown.	Unknown
I②	Asymptomatic. Not screened.	Unknown
I③	Sudden death aged 42. Subarachnoid haemorrhage given as cause of death.	Unknown
I④	Died in his 60s, cause unknown.	Unknown
I⑤	Ischaemic heart disease. Not screened.	Unknown
II①	Asymptomatic. ECG shows an unusual bifid P-wave, ST elevation in anterior chest leads, dominant R-wave in V1, partial RBBB, QT interval upper limit of normal, widespread TWI. Asymmetrical hypertrophy (up to 19mm), hypercontractility and mildly dilated LA.	Affected
II②	Diagnosed with LQTS as a teenager. Atypical chest pain, syncope and palpitations from late 20s. Prolonged QT interval and some Brugada-like features on ECG (QTc up to 500mS; partial RBBB, ST-elevation V1-3, dominant R-wave in V1, PR interval at upper limit of normal). Mild asymmetric hypertrophy (up to 14mm) and mildly dilated LA (4.5cm). Possible epilepsy.	Affected
II③	Died aged 15 following cardiac transplantation for dilated cardiomyopathy.	Affected
III①	Sees cardiologist but unable to get consent to obtain notes.	Possibly affected
III②	Infantile DCM. Secundum ASD.	Affected
III③	Dilated LA and small mid-muscular VSD. Normal ECG with QTc of 420 at age 10.	Possibly affected
III④	On ECG the QT interval is mildly prolonged (QTc up to 470mS), abnormal and prolonged repolarisation, high voltages, abnormal intraventricular conduction. Echocardiogram- mild septal hypertrophy and non-compaction at LV apex and posterior wall, abnormal diastolic function. Learning difficulties and mild dysmorphism (possibly due to anticonvulsants in pregnancy).	Affected
III⑤	Single syncopal episode on exertion. Mildly prolonged QT interval (up to 470mS, partial RBBB, dominant R-wave in V1). Borderline septal hypertrophy (up to 11mm) and some suggestions of non-compaction. Mild learning difficulties, febrile convulsions and possibly absence seizures.	Affected
III⑥	ECG normal in childhood.	Unknown
III⑦	Asymptomatic. Borderline prolonged QT (QTc 450mS).	Possibly affected
III⑧	Asymptomatic. Unscreened in early childhood.	Unknown

Table 4-5: Clinical features of family with differing cardiomyopathies, electrophysiological abnormalities and septal defects (definitely affected family members in black typeface). Abbreviations: ECG = electrocardiogram; LA = left atrium; RBBB = right bundle branch block; DCM = dilated cardiomyopathy; ASD = atrial septal defect; VSD = ventricular septal defect

4.3.2 Complex Cardiac Phenotypes

The combination of cardiac abnormalities within a single family is highly unusual. In terms of the cohort seen in the specialised Oxford Inherited Heart Disorders Clinic the variety of cardiac phenotypes seen is unprecedented. In different family members there are one or

more elements of hypertrophic cardiomyopathy, dilated cardiomyopathy, non-compaction cardiomyopathy, long QT syndrome, repolarisation/J-wave syndromes, abnormal conduction, and septal defects. There is also the suggestion of non-cardiac features, including developmental delay/learning difficulty, dysmorphism, and seizures, although these are all mild and not reliably segregating with the cardiac phenotype in the family. Although recognised connections between certain elements of the cardiac phenotype are known, for example between DCM and LQTS (discussed further in section 4.4.2, page 112), it is difficult to link such a diversity of clinical phenotypes.

Particularly given the presence of non-cardiac features the possibility that this was due to a chromosomal or mitochondrial disorder was considered. Chromosomal disorders typically give more extensive, and often variable, phenotypes due to deletion or duplication of multiple genes. Phenotypes then result from the combination of effects of dose-sensitive genes in the affected region. The inheritance pattern in this family is consistent with mitochondrial inheritance, and different cardiomyopathies, structural heart disease, conduction abnormalities, learning difficulties and seizures are all reported in mitochondrial disorders (Bates et al., 2012). An X-linked inheritance pattern is also consistent with the pedigree, although the clinical phenotype is not consistent with a defined X-linked disorder.

4.3.3 Genetic Studies Performed

Due to the complex nature of the phenotype seen a number of genetic tests had already been performed in members of this family in a clinical diagnostic setting, without a cause being identified. This included tests of a hypertrophy phenotype (*MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2* and *GLA*), long QT syndrome (*KCNQ1, KCNH2, SCN5A, KCNE1,* and *KCNE2*), mitochondrial disorders (m.3243A>G, *POLG1*, and *PEO1*), as well as G-banded karyotype analysis.

4.3.4 Results

4.3.4.1 WGS with Segregation Data

Family members who had whole genome sequencing or SNP array are shown in Figure 4-18. Following analysis of variants produced from WGS and SNP array in affected family members, as described previously, eight variants remained (Table 4-6; chromosome linkage plots with genes in which variants have been identified are shown in Figure 4-19).

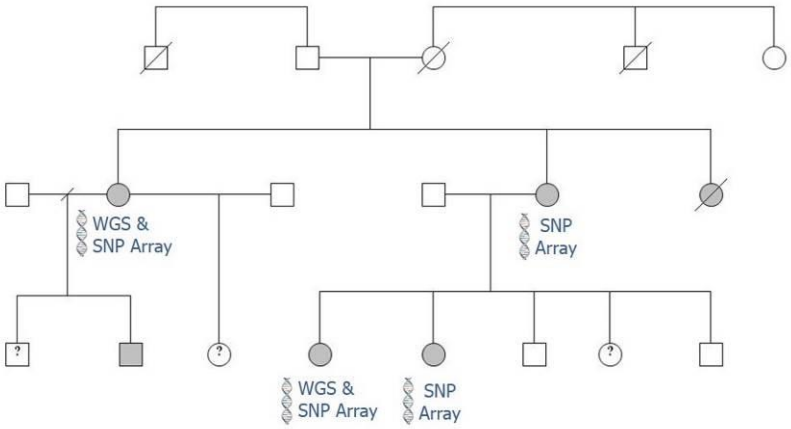


Figure 4-18: Pedigree diagram showing who in the family had whole genome sequencing and/or SNP array analysis.

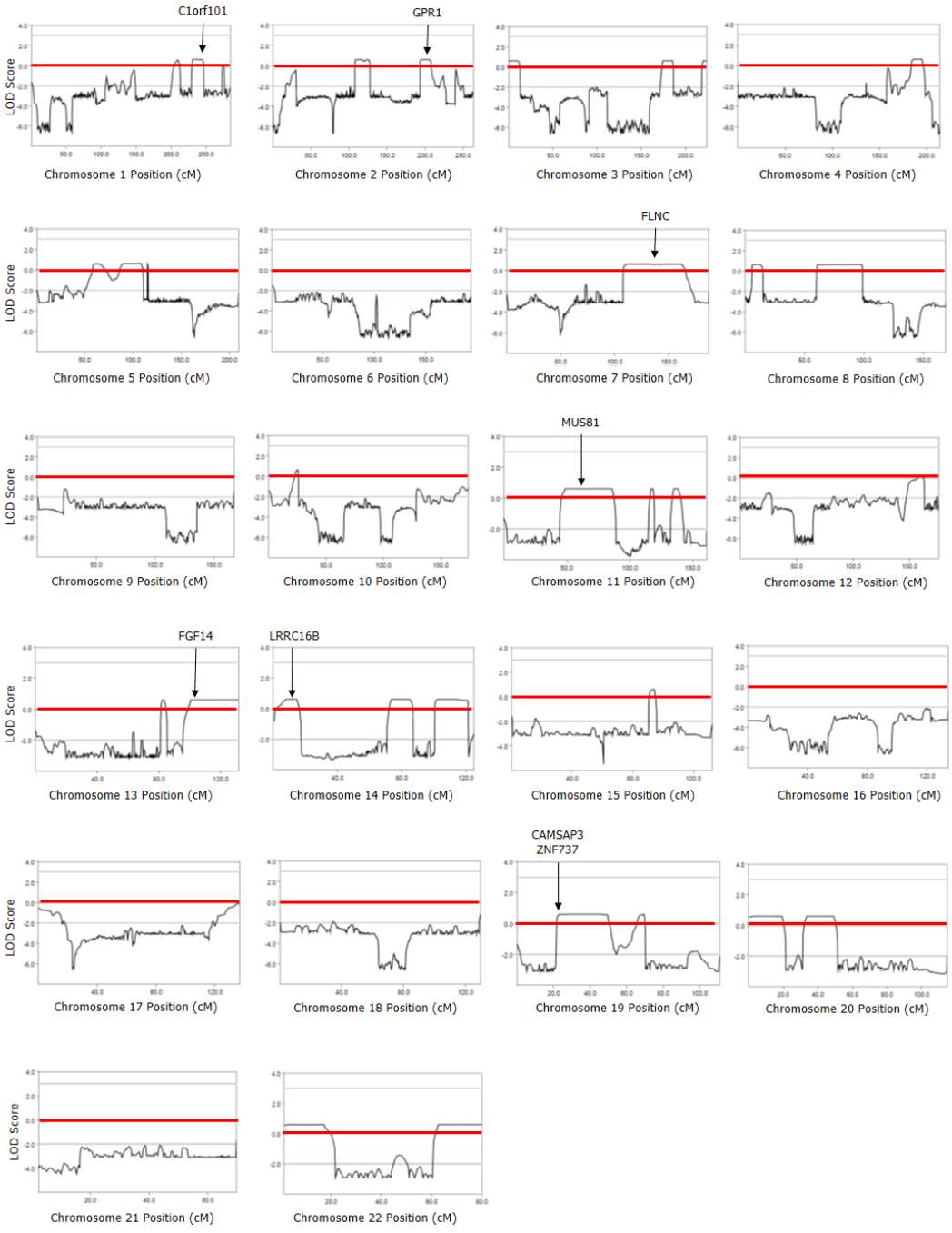


Figure 4-19: Linkage plots for each chromosome from SNP array data analysed in MERLIN software are shown (Abecasis et al., 2002). Genes containing variants identified from WGS studies are highlighted. Red lines indicate a LOD score of 0.

Gene	Variant	Conservation (PhastCons/PhyloP)
C1orf101	p.Y687H	Mammals (0.01/-0.04)
CAMSAP3	p.E182V	Mammals except dog (0.98/1.34)
FGF14	**p.V218F**	**Mammals (1/5.69)**
FLNC	**p.W2131C**	**All species (1/5.53)**
GPR1	p.I107N	Primates, dog and cow (0/-0.06)
LRRC16B	p.E286K	Mammals (1/4.81)
MUS81	p.K400T	Mammals (1/4.32)
ZNF737	p.T497I	Primates (0/-0.12)

Table 4-6: Information on segregating variants identified in affected family members in a family with a combined phenotype of different cardiomyopathy and electrophysiological abnormalities. Stronger candidates for causing the phenotype are highlighted in bold typeface. No unaffected family members were tested as those available for study were all children. Due to age-related penetrance in inherited heart diseases these family members would therefore be genetically relatively uninformative.

Summary of the known roles of the eight genes in which variants have been identified:

p.Y687H in C1orf101

Chromosome 1 open-reading frame. No information available on function. Highly expressed in the testes and ovaries, not identified as being expressed in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.E182V in Calmodulin-regulated spectrin-associated protein family member 3 (CAMSAP3)

CAMSAP3 appears to maintain the integrity of the zona adherens in epithelial cells (Meng et al., 2008). It is widely expressed but with low expression in the heart (Nagase et al., 2000).

p.V218F in Fibroblast growth factor 14 (FGF14)

FGF14 is a member of the fibroblast growth factor homologous factors (FHFs), a subset of the fibroblast growth factor family incorporating FGFs 11-14. Mutations in *FGF14* are the cause of dominant spinocerebellar ataxia-27 (SCA27), patients may also have mild learning difficulties (van Swieten et al., 2003). This is in keeping with the finding that it is predominantly expressed in the central nervous system. Although patients with SCA27 have

not been reported to have a cardiac phenotype, roles of the FGF14 homologue in murine cardiomyocytes have been established (Wei et al., 2011).

FHFs are intracellular proteins that bind and modulate voltage-gated sodium channels such as $Na_v1.5$ (encoded by SCN5A). All of the FHFs are expressed in the developing mouse heart with Fgf13 being the predominant FHF of the adult mouse heart, although Fgf14 is expressed at low levels. Knockdown of Fgf13 in adult mouse ventricular cardiomyocytes led to a reduced sodium channel current density and delayed recovery from inactivation (Wei et al., 2011). The relative roles of the FHFs in the human heart have not been investigated.

p.W2131C in Filamin C (FLNC)

FLNC acts to crosslink actin filaments and interact with a number of diverse cellular proteins. It is upregulated in the early stages of myocyte differentiation, and locates to the periphery of the Z-disk, interacting with myotilin and FATZ/calsarcin/myozenin (van der Ven et al., 2000b, Faulkner et al., 2000), suggesting a role in myofibrillogenesis (proteins of the Z-disk are discussed in more detail in section 5.2.2.3, page 143). It also interacts with the sarcoglycans, transmembrane proteins which link the cytoskeleton to the extracellular matrix as part of the dystrophin glycoprotein complex (Thompson et al., 2000).

FLNC is highly expressed cardiac and skeletal muscle. Heterozygous and homozygous mutations in *FLNC* cause myofibrillar and limb-girdle type skeletal myopathies. A significant minority of these have cardiac involvement, including cardiomyopathy and conduction abnormalities (Kley et al., 2007). Mutations in some of the interacting proteins of filamin C, such as the sarcoglycans, can also give myopathy and cardiomyopathy phenotypes (Thompson et al., 2000).

The residue substituted by this variant is in a region between the filamin 19 (Ig-like) repeat and an intra-domain insert before the filamin 20 repeat. Filamin repeats 18/19 (and 20/21)

adopt an unusual conformation with repeat 18 folding over repeat 19 to regulate an integrin binding site (Heikkinen et al., 2009). Widely expressed isoforms of the homologous filamins A and B lack this region (including residues 2127-2167). The loss of this region affects filamin binding with integrins and migfilin (Pentikainen et al., 2011). The integrins are transmembrane receptors which are important for the attachment of a cell to its surroundings. Migfilin is a widely expressed component of actin-cytoskeleton-membrane junctions, linking cell adhesion structures to the actin cytoskeleton to regulate actin remodelling, cell morphology and motility (Wu, 2005).

p.I107N in G protein-coupled receptor 1 (GPR1)

G protein-coupled receptors mediate the actions of endogenous peptides in the brain (Marchese et al., 1994). Mutations in other GPR genes have been reported to cause central endocrine disorders (Spiegel, 1996). Although it is predominantly expressed in the central nervous system and endocrine tissues it is more widely expressed, including in the heart (Liu et al., 2008, Husar Bioinformatics Lab, April 2013).

p.E286K in Leucine-rich repeat-containing protein 16B (LRRC16B)

LRRC16B is a leucine-rich protein which is very highly expressed in the developing brain and also in certain cancers (Hsu et al., 2011). Its functions are largely unknown.

p.K400T in S. Cerevisiae endonuclease homologue (MUS81)

This is an endonuclease which is widely expressed in different tissues and appears to have a role in resolving Holliday junctions (Chen et al., 2001). Mus81-null mice were phenotypically normal but were more susceptible to DNA damage caused by mitomycin C (McPherson et al., 2004).

p.T497I in Zinc-finger protein 737 (ZNF737)

ZNF737 is a zinc-finger protein which may be involved in transcriptional regulation. It is widely expressed, particularly in T-lymphoid cells (Bellefroid et al., 1993), but little detail is known about its function.

Interpretation of variant findings

Based on known biological function and expression the variants in the *FGF14* and *FLNC* emerge as the stronger candidates for causing a cardiac phenotype. Both of these variants are in highly conserved residues, although other variants identified are also in conserved residues.

4.3.4.2 Copy Number Analysis

Due to the complex phenotype, with extra-cardiac features, chromosomal rearrangements were studied. Chromosomal copy number variation was also studied in other families as part of wider study of WGS500 data, and no CNVs identified, but was not specifically requested. Copy number variation was determined using both the SNP array data, analysed in the Nexus software, along with a clinical diagnostic array CGH analysis. Both of these strategies detected the same two 17q duplications in the family. These two duplications, a 430 kb duplication at 17q11.2 (17:26036063-26469240 in NCBI genome build 36.1/hg18) and a 1.69 Mb duplication at 17q12 (17:31635465-33323906 in NCBI genome build 36.1/hg18), were segregating independently in the family (Figure 4-20). Similar duplications of 17q12 have been reported previously (Sharp et al., 2006), whereas the 17q11.2 duplication is not reported, although deletions of this region have been (Kluwe et al., 2004).

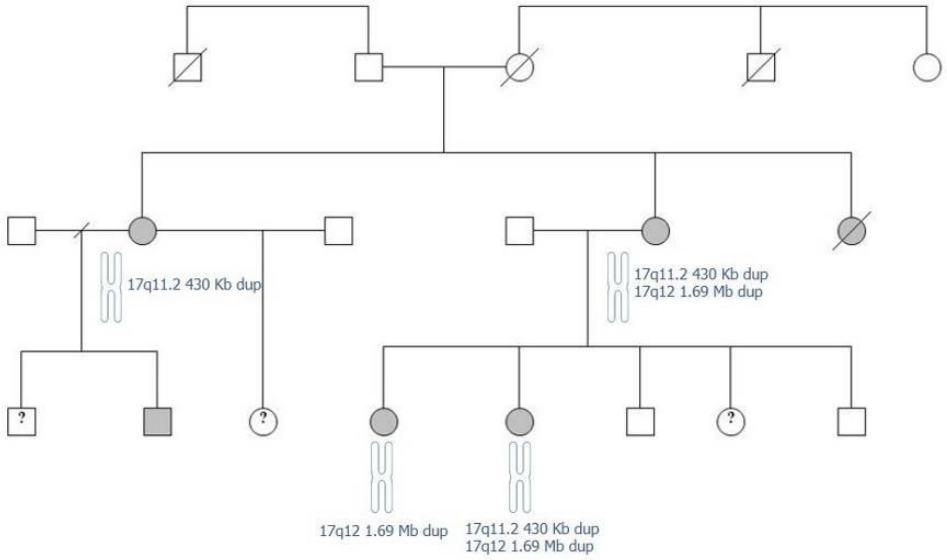


Figure 4-20: Segregation of chromosomal duplications in family with differing cardiomyopathies, electrophysiological abnormalities and septal defects. Neither of the duplications segregates with the whole phenotype in the family.

17q11.2 duplication

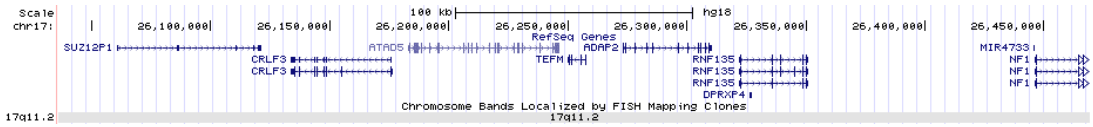


Figure 4-21: Genes in the 17q11.2 duplication region depicted in the UCSC genome browser

Gene	Summary
SUZ12P1	Suppressor of zeste 12 homologue pseudogene.
CRLF3	Encodes the cytokine receptor-like factor 3. Overexpression HEK cells induced cell cycle arrest (Yang et al., 2009).
ATAD5	Encodes a DNA damage response protein. Somatic mutations identified in certain cancers. Knockout in mice is lethal, heterozygosity leads to genomic instability (Bell et al., 2011).
C17orf42	Open reading frame.
ADAP2	Contains an ARF-GAP-like zinc finger motif which locates to the plasma membrane. Weak expression in the heart (Hanck et al., 2004).
RNF135	Ring-finger protein. Mutations identified in cohort with overgrowth syndromes (Douglas et al., 2007).
NF1	Encodes neurofibromin. Heterozygous mutations and deletions cause neurofibromatosis type 1 (NF1) (Upadhyaya et al., 1992).

Table 4-7: Genes incorporated in the 17q11.2 duplication

The duplication of 17q11.2 identified has not been reported previously. Large deletions of this region are identified as a cause of approximately 5% of NF1 (Kluwe et al., 2004). In those with deletions, typically of 1.4 Mb, a more severe phenotype is usually seen (Pasmant et al., 2010). The significance of this duplication is therefore uncertain.

17q12 duplication

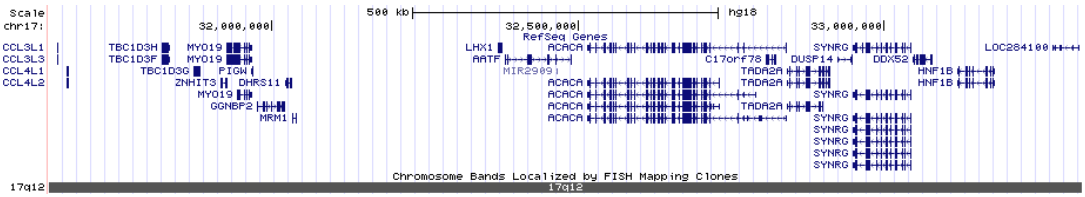


Figure 4-22: Genes in the 17q12 duplication region depicted in the UCSC genome browser

Gene	Summary
ZNHIT3	Thyroid hormone receptor interactor (Lee et al., 1995).
MYO19	Myosin motor protein which interacts with mitochondria and is thought to mediate the movement and/or tethering of mitochondria to actin (Quintero et al., 2009).
PIGW	An acyltransferase which is involved in the synthesis of glycosylphosphatidylinositol, a complex glycolipid that anchors proteins to the cell surface (Murakami et al., 2003).
GGNBP2	Gametogenetin-binding protein 2.
DHRS11	A 17β-hydroxysteroid dehydrogenase homologue.
MRM1	Mitochondrial rRNA methyltransferase homologue (*S. Cerevisiae*).
LHX1	LIM homeobox gene involved in the regulation of neurological development (Kania et al., 2000).
AATF	Apoptosis-antagonising transcription factor, down-regulated in many tumours. Highly expressed in the heart (Di Padova et al., 2003).
ACACA	α form of acetyl-CoA carboxylase that catalyses long-chain fatty acid synthesis (Abu-Elheiga et al., 2005).
TADA2L	Subunit of the histone acetyltransferase complex that enhances RNA polymerase II mediated gene transcription (Candau et al., 1996).
DUSP14	Dual specificity phosphatase which modulate certain signalling pathways (ERK, JNK, p38) (Marti et al., 2001).
AP1GBP1	Interacts with the AP1 complex which associates with the trans-Golgi network linking selected proteins to the clathrin lattice (Page et al., 1999).
DDX52	Dead box polypeptide homologue.
HNF1B	Hepatocyte nuclear factor 1-β. Heterozygous mutations cause a syndrome of diabetes and renal cysts (Horikawa et al., 1997).

Table 4-8: Genes incorporated in the 17q12 duplication

17q12 duplications cause a recurrent definable syndrome (Sharp et al., 2006, Bierhals et al., 2013). Duplications at this locus have been reported ranging from 300 kb to 2.1 Mb, with most of a similar size order and location as in our family. Clinical phenotype is somewhat variable but reported features include developmental delay, behavioural disturbance, epilepsy/seizures, structural brain, gastrointestinal, renal and genitourinary abnormalities

(Bierhals et al., 2013). Cardiac abnormalities are not reported. 17q12 duplications are found as rare copy number variants in the normal population, but are enriched in populations with learning difficulties and certain other congenital abnormalities (Mefford et al., 2007).

4.3.4.3 Mitochondrial Genome Analysis

As many of the clinical features in this family were consistent with mitochondrial dysfunction, and the inheritance pattern being consistent with maternal inheritance, variants in the mitochondrial genome were studied. The mitochondrial genome was analysed in both the WGS data and by specific sequencing of the mitochondrial genome, as in section 2.2.2 (page 29).

Once variants had been excluded by their presence in population databases (Ingman and Gyllensten, 2006, Kogelnik et al., 1996, Pereira et al., 2009) only two variants were identified. Both sets of data identified a homoplasmic synonymous variant in the mitochondrial cytochrome b gene, m.15745C>T. The fact that this change is synonymous means it is more likely to be a rare benign variant. Blood lactate levels were normal but more specific analysis, such as the study of the mitochondrial respiratory chain in muscle, was not possible. The other variant identified, m.3862T>C, was only evident in the Roche-454 sequencing data, and this suggested a relatively low level of heteroplasmy. Sanger and pyrosequencing of this specific region were undertaken and did not identify this variant at any level, it is therefore more likely to be an artefact of the original sequencing.

4.3.4.4 Follow-Up Studies

The *FLNC* and *FGF14* variants were confirmed in those family members who had WGS or SNP array studies (the primers used are reported in appendix section III-d, page 256; Chromatogram from Sanger sequencing confirmation of the FLNC variant is shown in Figure

4-23). It was hoped that further clinical and genetic segregation information would be obtained on family members III①, ②, and ③, but unfortunately this was not possible.

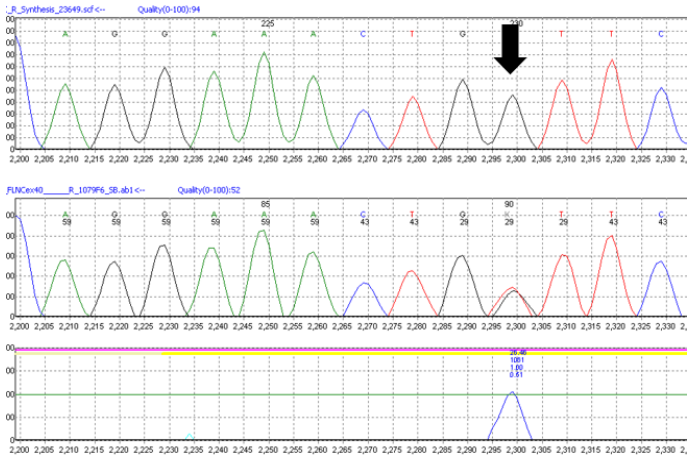


Figure 4-23: Chromatogram from Sanger sequencing confirming the presence of the *FLNC* variant identified from whole genome sequencing in a family with a highly varied cardiac phenotype. The reference sequence is shown at the top with the test sequence from the forward primer read below, the variant was also confirmed with the reverse primer (analysed and imaged using MutationSurveyor software).

4.3.5 Conclusions

The genetic findings in this family are as difficult to interpret as the clinical phenotype. The most certain finding is that of the 17q12 duplication. The finding of the 17q12 duplication in one branch of our family (II②, III③ and III④; pedigree on page 95) may explain at least part of their phenotype. This branch do have learning difficulties and seizures, the most frequently reported features with such duplications (Bierhals et al., 2013). One individual with this duplication, III③, has also been reported as having mild facial dysmorphism, along with some mild skeletal features (talipes equinovarus and clinodactyly), tooth abnormalities and 'small' kidneys on IVP noted when she was younger which again may be due to this duplication. However, it is notable that neither duplication identified segregated with the cardiac phenotype in the family.

The role of other variants found is uncertain. The variants in *FLNC* and *FGF14* are noted to be possible causes of a cardiac phenotype based on the known biological functions of the encoded proteins. Determining the contribution of these different variants in the family is

challenging and would be greatly aided by further family segregation information as it emerges in the future. The very unusual phenotype means that a cohort with a similar phenotype is not available for further interrogation of these genes. As further information becomes available about these genes, and more genetic population data becomes available allowing improved determination of rare benign variants, further investigation of these variants may become warranted.

4.4 A Family with a Combined Phenotype of Left Ventricular Hypertrophy, Prolonged QT Interval, and Sinoatrial Node Dysfunction

4.4.1 Clinical Information

The proband in the family, patient II② (Figure 4-25), presented with atypical chest pain at forty-seven years of age. On investigation she was noted to have a prolonged corrected QT interval (QTc) on ECG (Figure 4-24) and was commenced on β-blocker therapy. The maximal QT interval measured on twenty-four hour ECG was 0.84 seconds (QTc 0.69secs). On the twenty-four hour tape she also had nocturnal bradycardias down to thirty-seven beats per minute, and sinus pauses of up to two seconds. Her echocardiogram was normal.

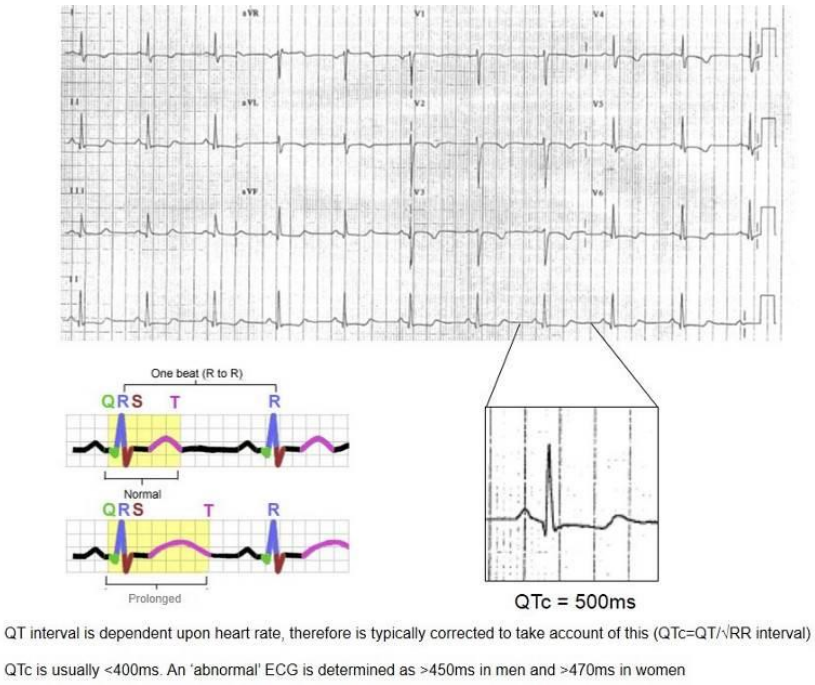


Figure 4-24: ECG from patient II② showing QT prolongation, with a corrected QT interval (QTc) of 500 milliseconds, and T-wave abnormalities. QT interval is dependent on heart rate therefore this is typically corrected for. QTc is usually <400 milliseconds, although 'normal' measurements are considered to be up to 450 milliseconds in men and 470 milliseconds in women.

She was clinically diagnosed as having long QT syndrome (LQTS) type three, although the T-wave abnormalities are not typical of this. Usually in type three LQTS the T-waves are of normal morphology although may be small. On the basis of the significantly prolonged QTc a DDD pacemaker was inserted at fifty-three years of age. Pacemaker interrogation showed frequent brief episodes of atrioventicular nodal re-entry tachycardia (AVNRT) causing dizziness and palpitations but no syncope that was ablated with good response.

Cascade screening led to the identification of a prolonged QT interval (II⑥, II⑧, III③, and III⑦) and sinus bradycardia and pauses (all affected) in other family members (the pedigree for this family is shown in Figure 4-25, and further clinical details on all family members are documented in Table 4-9). Patients II⑥, II⑧, II⑨, and III③ were also identified as having evidence of mild concentric or asymmetrical septal hypertrophy (maximal septal measurement of sixteen millimetres; Figure 4-26). Other than the proband all family members were asymptomatic.

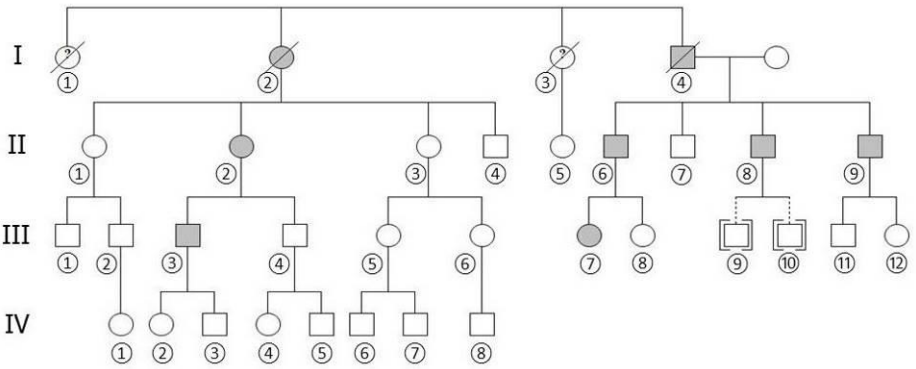


Figure 4-25: Pedigree of family with a combined phenotype of left ventricular hypertrophy, prolonged QT interval, and sinoatrial node dysfunction.

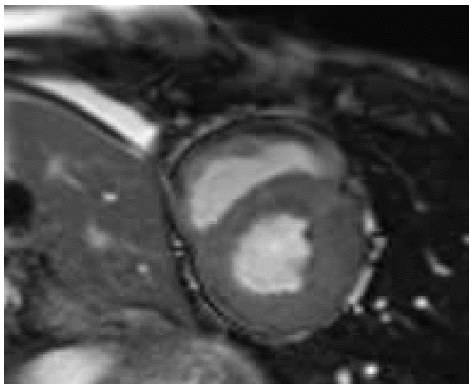


Figure 4-26: Cardiac magnetic resonance short-axis view T1-weighted post-gadolinium enhancement image from patient II⑧ showing mild asymmetric left ventricular hypertrophy.

Patient	Clinical Information	Summary
I①	PPM. No further information.	Possibly affected
I②	Died aged 61 from melanoma. No known heart problems.	Implicitly affected
I③	PPM. No further information.	Possibly affected
I④	Syncopal episodes and PPM inserted in 50s. Died in 60s due to cerebrovascular accident.	Affected
II①	Normal electrocardiogram.	Possibly unaffected
II②	Asymptomatic. Normal echocardiogram. Prolonged QT (500ms, up to 700ms nocturnal), lateral ST depression and T-wave inversion. AVNRT treated with ablation. Bradycardias and pauses (up to 2.0secs), PPM inserted.	Affected
II③	Normal electrocardiogram.	Possibly unaffected
II④	Normal electrocardiogram.	Possibly unaffected
II⑥	Asymptomatic. Mild concentric LVH (11mm). Mildly prolonged QT interval (470mS). Bradycardias and pauses (up to 3.1secs), PPM inserted.	Affected
II⑦	Asymptomatic. Not wanted screening.	Unknown
II⑧	Occasional pre-syncopal episodes. Mildly prolonged QT interval (470mS) and lateral T-wave inversion. Discrete area of septal hypertrophy to 16mm and increased trabeculations. Bradycardias and pauses (up to 2.6secs), PPM inserted.	Affected
II⑨	Asymptomatic. Septal hypertrophy (up to 15mm). QT interval 360mS. Bradycardias (down to 34bpm).	Affected
III③	Asymptomatic. QTc of 500mS with partial RBBB and lateral ST depression. Mild concentric hypertrophy (11mm).	Affected
III④	Asymptomatic. Not wanted screening.	Unknown
III⑦	Asymptomatic. QTc up to 510mS. Hyperdynamic systolic function. Nocturnal bradycardias down to 25bpm. PPM inserted.	Affected
III⑧	Asymptomatic. QTc of 360mS, lateral T-wave inversion. Normal echocardiogram. Nocturnal bradycardias (down to 38bpm).	Likely affected
III⑪	Asymptomatic. Not yet screened.	Unknown
III⑫	Asymptomatic. Not yet screened.	Unknown

Table 4-9: Clinical summary of family with a combined phenotype of left ventricular hypertrophy, prolonged QT interval, and sinoatrial node dysfunction (definitely affected family members in black typeface). Abbreviations: PPM = permanent pacemaker; AVNRT = atrioventricular nodal re-entry tachycardia; LVH = left ventricular hypertrophy; RBBB = right bundle branch block.

4.4.2 Long QT Syndrome and Cardiomyopathy

LQTS and HCM are usually clinically distinct disorders, and the overlap seen within this family is very unusual. There is also the additional finding of marked sinus bradycardia and pauses. Bradycardia is reported in long QT syndrome, particularly in children (Kugler, 1991), but is not typically as pronounced as in this family. It is not a feature associated with HCM.

The combination of cardiomyopathy and prolonged QT interval in an individual, or family, can occur in one of two ways. Firstly, a prolonged QT interval may be secondary to the underlying cardiomyopathy, thought to be due to electrophysiological instability caused by the structural change (Dritsas et al., 1992, Martin et al., 1994). This usually only causes mild prolongation of the QT interval, and is usually associated with the clinical severity of the cardiomyopathy. Secondly, in rare cases, both prolonged QT and cardiomyopathy are independent but due to the same cause. For example, mutations in *SCN5A* can cause both cardiomyopathy and prolonged QT. Although DCM and LQTS due to *SCN5A* mutations are usually seen separately they have been reported in the same individual and family (Shi et al., 2008, Abriel, 2010).

LQTS, without additional features, affects approximately 1 in 2000 people (Schwartz et al., 2009). Without treatment it is associated with a significant risk of ventricular arrhythmias and sudden death, particularly in patients with risk factors such as syncopal episodes. This risk is significantly reduced by treatment with β-blockers, particularly in type one and two LQTS, and implantable defibrillators may also be indicated (Schwartz et al., 2012). LQTS is usually autosomal dominantly inherited. The majority of patients have loss-of-function mutations in the potassium channel subunit genes *KCNQ1* or *KCNH2*, or gain-of-function mutations in the sodium channel subunit gene *SCN5A*. However heterozygous mutations in at least eleven different genes, encoding other ion channel subunits or interacting proteins, can cause dominantly inherited LQTS (Schwartz et al., 2012).

The main clinical features and genetic aetiology of classical HCM have been discussed in section 2.1.1 (page 23).

4.4.3 Genetic Studies Performed

Prior to enrolment in this study clinical diagnostic testing had already been undertaken. Due to the overlap of hypertrophy and prolonged QT testing of a number genes had been

analysed: *MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2* and *GLA* as causes of hypertrophy, and *KCNQ1, KCNH2, SCN5A, KCNE1,* and *KCNE2* as causes of LQTS. No likely pathogenic variants were identified.

4.4.4 Results

4.4.4.1 WGS with Segregation Data

Family members who had whole genome sequencing and/or SNP array are shown in Figure 4-27. Following analysis of variants produced from WGS and SNP array in definitively affected family members, as described previously, two variants remained (Table 4-10).

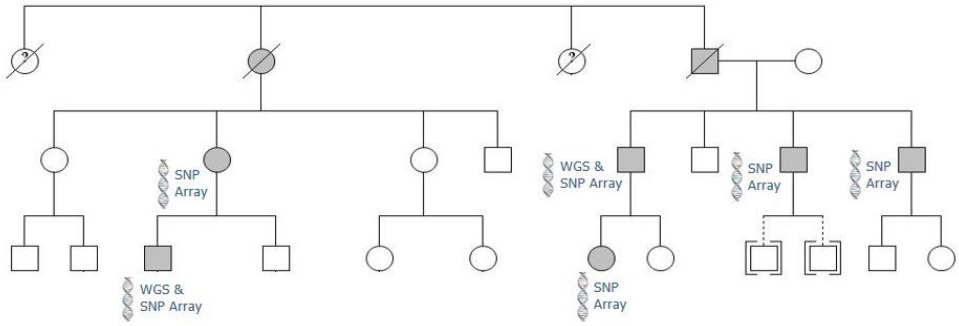


Figure 4-27: Pedigree diagram indicating who in the family had whole genome sequencing and/or SNP array analysis.

Gene	Variant	Conservation (PhastCons/PhyloP)
DUSP7	p.S396L	Human and chimp only (1, 6.1)
CACNA1C	p.R518C	All species (1, 3.1)

Table 4-10: Variants identified following WGS and SNP array in family with a combined phenotype of left ventricular hypertrophy, prolonged QT interval, and sinoatrial node dysfunction.

Summary of the known roles of the two genes in which variants have been identified:

p.S396L in Dual specificity phosphatase 7 (DUSP7)

The dual specificity phosphatases are a large heterogeneous subgroup of phosphatases. DUSP7 dephosphorylates certain mitogen-activated protein kinases that have roles in cell proliferation and programmed cell death (Patterson et al., 2009). It is fairly widely

expressed, but not reportedly at significant levels in the heart (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.R518C in $Ca_v1.2$, the α-1C subunit of the L-type calcium channel (CACNA1C)

The L-type calcium channels mediate the influx of calcium into cardiomyocytes, and are key for excitation-contraction coupling. This links electrical excitation of the cell membrane to the contractile mechanisms of the heart, with the triggering of calcium-mediated calcium release from the sarcoplasmic reticulum. Phenotypes caused by mutations in *CACNA1C* are discussed in section 4.4.5. This variant changes a polar to a hydrophobic residue in a highly conserved region.

Interpretation of variant findings

The variant in the *CACNA1C* gene is a very strong candidate for the phenotype seen, and is discussed below. This is further supported by it being in a highly conserved residue, and the lack of other variants identified. Although the variant in DUSP7 cannot be completely excluded the fact that this is a poorly conserved residue, and the known unrelated function and expression profile, makes pathogenicity highly unlikely.

4.4.4.2 Follow-Up Studies

Sanger sequencing of exon twelve of *CACNA1C* confirmed the variant segregated through all affected family members (chromatogram is shown in Figure 4-28). Through collaboration with the Clinical Genetics Department in Dublin, Ireland, a family with a similar phenotype was identified. This family showed features of hypertrophy, sudden death in early adulthood, prolonged QT interval, and evidence of conduction abnormalities, although not sinoatrial node dysfunction. A further family from our clinic was identified with a combination of hypertrophy and prolonged QT, although again without sinoatrial node dysfunction. All forty-eight exons of *CACNA1C* were Sanger sequenced in both of these

families but no causative mutation identified (*CACNA1C* primers used are shown in the appendix section III-a, page 252).

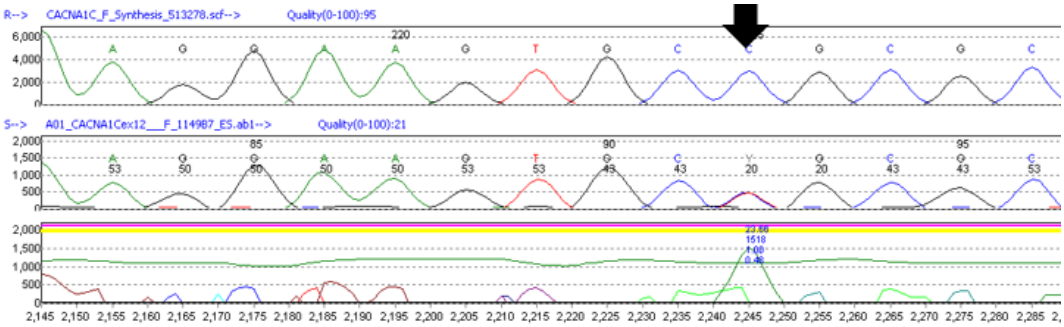


Figure 4-28: Chromatogram from Sanger sequencing confirming the presence of the *CACNA1C* variant, identified from whole genome sequencing, in a family with a phenotype of prolonged QT interval, left ventricular hypertrophy and sinus node dysfunction. The reference sequence is shown at the top with the test sequence from the forward primer read below, the variant was also confirmed with the reverse primer (analysed and imaged using MutationSurveyor software).

4.4.5 Discussion

Whole genome sequencing has identified a variant in *CACNA1C* in a family with a combined phenotype of prolonged QT interval, left ventricular hypertrophy, and sinoatrial node dysfunction. This is a very strong candidate for the phenotype seen. *CACNA1C* encodes an α-1, pore-forming, subunit of L-type voltage-gated calcium channels. L-type calcium channels are responsible for the influx calcium into cells when there is membrane depolarisation. In cardiomyocytes this links electrical excitation to the contractile mechanisms of the heart, by triggering calcium-mediated calcium release from the sarcoplasmic reticulum.

There are four genes encoding α-1 subunits of L-type channels, *CACNA1S* ($Ca_v1.1$), *CACNA1C* ($Ca_v1.2$), *CACNA1D* ($Ca_v1.3$), and *CACNA1F* ($Ca_v1.4$), each with a different expression pattern (Birnbaumer et al., 1994). In all of these genes heterozygous mutations have been identified causing Mendelian disorders (Ptacek et al., 1994, Baig et al., 2011, Strom et al., 1998), with mutations in *CACNA1C* causing Timothy syndrome (Splawski et al., 2004), LQTS (Boczek et al., 2013), and Brugada syndrome with short QT interval (Antzelevitch et al., 2007).

Timothy syndrome (TS) is a multisystem disorder that is characterised by cardiac abnormalities (long QT syndrome, congenital heart disease, HCM and/or DCM), facial dysmorphism, syndactyly, developmental delay and autism, immune disorder, and hypoglycaemia (Lo et al., 2005). TS mutations are gain-of-function mutations in the equivalent residue in two differentially spliced exons, p.G406R and p.G402S, in exons 8 and 8A respectively (Splawski et al., 2005). The variant identified in the family studied, p.R518C, is located within the same intracellular linker region as these mutations, albeit adjacent to the next transmembrane domain (Figure 4-29). Mutations causing dominantly inherited LQTS have only recently been reported but there is early evidence of a predilection for the II-III linker region, again causing a gain-of-function (Boczek et al., 2013). Although mutations in this gene have been reported to cause overlapping cardiac disorders the phenotype of this family is distinct.

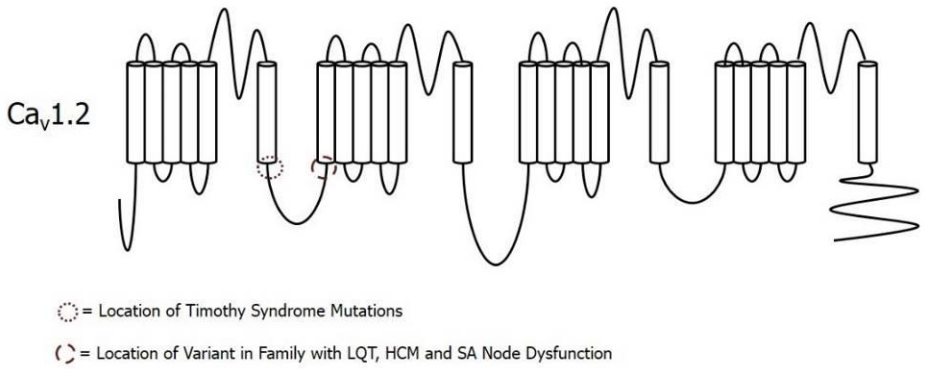


Figure 4-29: Schematic representation of the location of Timothy mutations and the p.R518C variant in $Ca_v1.2$ identified from whole genome sequencing in a family with a combined phenotype of prolonged QT interval, left ventricular hypertrophy, and sinus node dysfunction.

Aside from the occasional finding of left ventricular hypertrophy in TS, there is further evidence for the role of calcium channels in hypertrophy. Calcium is an important regulator of several hypertrophic pathways, for example calcineurin/NFAT and CaMK/HDAC (Heineke and Molkentin, 2006). Although influx of calcium through L-type calcium channels does not cause hypertrophy under physiological conditions, an increase in influx under pathological

conditions is proposed as an important source of calcium activating these pathways (Chen et al., 2011).

A connection between a variant in *CACNA1C* and sinus bradycardia and pauses can also be explained, as the L-type calcium channel is the principal channel governing depolarisation of sinoatrial node cardiomyocytes. The situation is slightly more complex than this though, as both $Ca_v1.2$ and $Ca_v1.3$ are expressed at the sinoatrial node (Christel et al., 2012). A recessive 3 base-pair insertion in the *CACNA1D* gene has been identified causing bradycardia, along with deafness (Baig et al., 2011). Interestingly this mutation is in the equivalent linker-region of $Ca_v1.3$ (p.403_404insGly of exon 8A) as the variant found in our family in $Ca_v1.2$. The relative roles of these two isoforms at the sinoatrial node are yet to be established.

4.4.6 Future Work

Further families with mixed phenotypes incorporating LQTS, LVH and/or sinoatrial node dysfunction are being sought, although we know from clinical practice that these families are very rare. Identification of mutations in *CACNA1C* in additional patients with a similar phenotype as this family would be strong supportive evidence for the pathogenicity of this variant. In addition it may provide further information on possible genotype-phenotype correlation of mutations in this gene.

Further evidence of pathogenicity can also be gathered from showing that this variant has a functional effect on the function of the L-type calcium channel. Family members are being recruited for further study in the department to create inducible pluripotent stem (iPS) cells from skin fibroblasts. This will allow study of L-type calcium channels *in vitro*, using cardiomyocytes derived from patient cells. Collaborations are also being established to generate a transgenic zebrafish with this variant in the homologous gene. This would allow the effects on phenotype and channel function to be studied in this model. It is anticipated

that this variant will be causing gain-of-function of this channel, as with the mutations causing TS and isolated LQTS. Further work to determine the different cellular effects of these mutations to understand what is mediating the differences in phenotype will be illuminating.

4.5 A Family with Autosomal Dominant Left Ventricular Non-Compaction

4.5.1 Clinical Information

The main clinical phenotype identified in this family is left ventricular non-compaction (LVNC) (Figure 4-30), with ventricular dilatation being identified in some (Figure 4-31). The initial presentation of cardiomyopathy in the family was a man, II⑪ (Figure 4-32), who died suddenly at twenty years of age due to rapidly decompensating congestive cardiac failure, having previously been well (pedigree is shown in Figure 4-32). At that time this was not recognised as a familial condition. The proband was his sister, II⑩, who presented at sixty-one years of age with a non-ST elevation myocardial infarct due to coronary embolus. LVNC, with mild left ventricular dilatation, was diagnosed in her at this time. Cascade family screening identified the same condition in several other family members (clinical details of family members are summarised in Table 4-11).

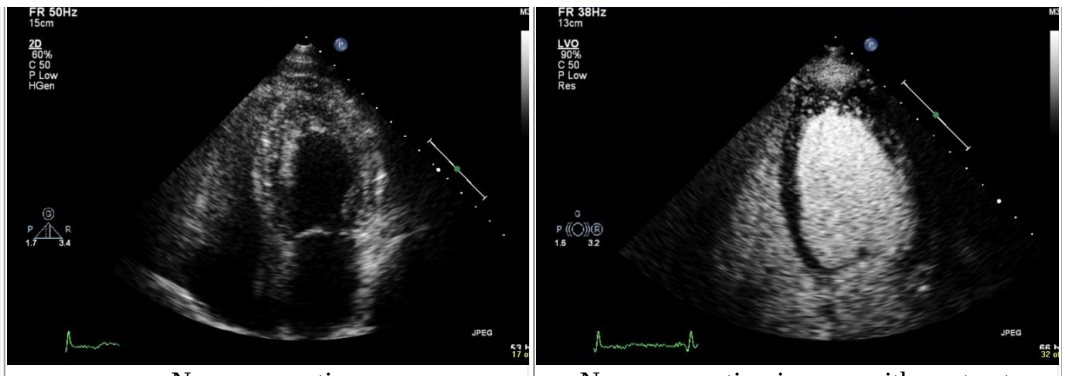


Figure 4-30: Echocardiogram images showing the characteristic 'spongy' appearance of non-compaction in patient II⑩, with and without contrast

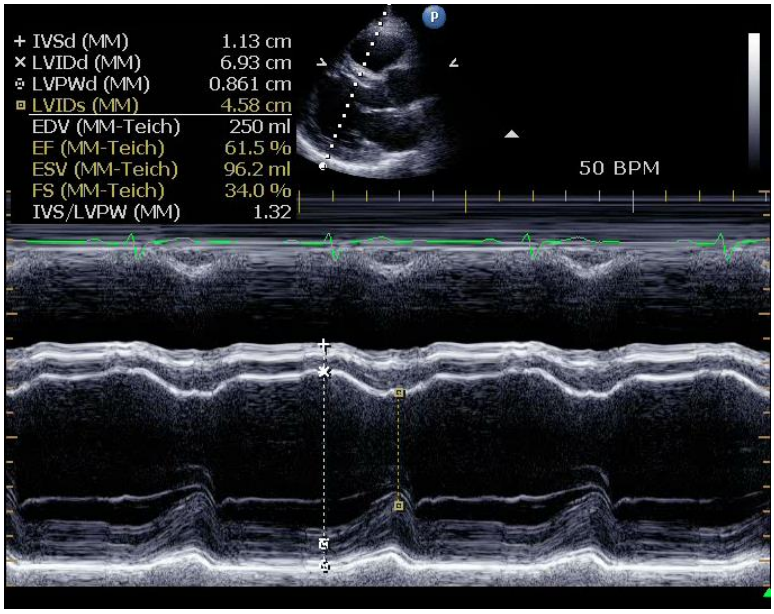


Figure 4-31: Echocardiogram image from patient II⑫ showing significant dilatation (LVIDd- left ventricle internal dimensions in diastole- 6.93cm; <5.6cm normal; LVIDs- left ventricle internal dimensions in systole- 4.58cm; <4.0cm normal), but maintaining a thickened myocardium (IVSd- interventricular septal diameter- 1.13cm; 0.6-1.1cm normal), typical for LVNC and unlike the thinned myocardium seen in DCM.

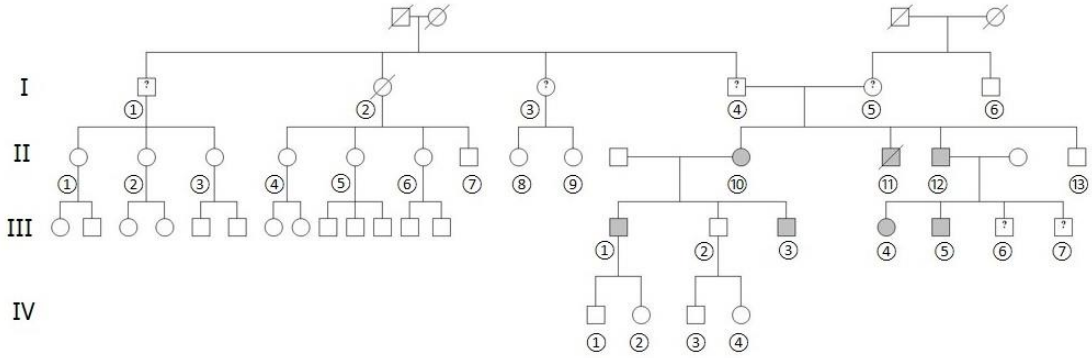


Figure 4-32: Pedigree of family with autosomal dominant LVNC.

Patient	Clinical Information	Summary
I①	Mild concentric hypertrophy (1.2cm) but in the setting of longstanding hypertension and mild aortic stenosis. Atrial fibrillation and bradycardia, pacemaker *in situ.*	Possibly affected
I②	Died of cancer aged 82. No known heart problems.	Unknown
I③	Mild concentric hypertrophy. Unable to have CMR.	Possibly affected
I④	Moderate global hypertrophy up to 16mm with increased trabeculation not meeting diagnostic criteria for LVNC. CVA aged 78. History of hypertension.	Probably affected
I⑤	Mild concentric hypertrophy with no evidence of LVNC, had been normal aged 72. TIA aged 79. Pacemaker inserted aged 72 due to symptomatic frequent ventricular ectopics of RV outflow origin.	Possibly affected
II①	Borderline LV hypertrophy and relatively small LV cavity. Small Q-wave in lead II on ECG.	Possibly affected

II②	Normal echocardiogram. Unable to have CMR.	Probably unaffected
II③	Asymptomatic. Not screened.	Unknown
II④	Asymptomatic. Normal screening.	Probably unaffected
II⑤	Asymptomatic. Normal screening.	Probably unaffected
II⑥	Asymptomatic. Not screened.	Unknown
II⑦	Asymptomatic. CMR shows mildly increased trabeculations of the lateral wall but not diagnostic for LVNC.	Possibly affected
II⑧	Asymptomatic. Normal screening including CMR.	Probably unaffected
II⑨	Asymptomatic. Not screened.	Unknown
II⑩	**Presented aged 61 with a NSTEMI believed to be due to coronary embolus. Contrast echocardiogram demonstrated a mildly dilated LV and LA with regional inferolateral hypokinesia and apical hypertrophy with trabeculation meeting LVNC criteria. Echocardiogram and ECG had been normal when investigated aged 48.**	**Affected**
II⑪	**Presented aged 20 in congestive heart failure and died suddenly in hospital. Post mortem showed a 680g heart with evidence of dilatation and hypertrophy.**	**Affected**
II⑫	**Diagnosed aged 38 on family screening with a dilated (7cm) and hypertrophied (1.2cm) left ventricle with reduced function. Marked trabeculations but not reaching diagnostic criteria for LVNC on CMR along with RV dilatation and delayed hyperenhancement in the basal anterior enhancement.**	**Affected**
II⑬	Asymptomatic. Normal screening.	Probably unaffected
III①	**LV dilatation (6.8/4.1) and mildly impaired systolic function with relative hypertrophy (1cm), LA dilatation; diagnosed aged 19. Unable to have MRI to clarify level of trabeculation. Symptomatic SVTs from age 11.**	**Affected**
III②	Asymptomatic. Normal screening at age 22. No follow up since.	Possibly unaffected
III③	**Asymptomatic. CMR showed increased trabeculation meeting diagnostic criteria for LVNC. Left ventricle cavity size and contractile function normal.**	**Affected**
III④	**Asymptomatic. CMR- Low normal systolic function. Prominent hypertrabeculation of lateral and inferior walls meeting diagnostic criteria for LVNC. No hypertrophy. Echo- distal septal hypertrophy up to 12mm becoming more clearly hypertrabeculated aged 24. ECG T-wave inversion lead 3.**	**Affected**
III⑤	**Asymptomatic. CMR- Borderline LV dilatation and mildly impaired systolic function. Hypertrabeculation meeting LVNC criteria with focal hypokinesia and wall thinning.**	**Affected**
III⑥	Asymptomatic. CMR shows LV volumes at upper limit of normal and mildly impaired function with some apical trabeculations.	Possibly affected
III⑦	Asymptomatic. CMR shows LV volume at higher end of normal. Wall thickness 11mm. Normal function and no hypertrabeculation. Late gadolinium enhancement at inferior LV/RV junction.	Possibly affected

Table 4-11: Clinical Information on members of the family with autosomal dominant LVNC(definitely affected family members in black typeface). Abbreviations: CMR = cardiac magnetic resonance imaging; LVNC = left ventricular non-compaction; CVA = cerebrovascular accident; TIA = transient ischaemic attack; RV = right ventricle; LV = left ventricle; NSTEMI = non-ST elevation myocardial infarct; LA = left atrium; SVT = supraventricular tachycardia; ECG = electrocardiogram.

4.5.2 Left Ventricular Non-Compaction

LVNC is a rare form of cardiomyopathy characterised by a 'spongy' myocardium with increased trabeculations. Diagnosis of LVNC is based on the morphological pattern of increased trabeculation of the left ventricular wall, with deep inter-trabecular recesses communicating with the ventricular cavity (Figure 4-33). A ratio of non-compacted to compacted myocardium of greater than two is considered abnormal. The diagnosis of LVNC is challenging because trabeculation of the myocardium is a normal finding in a relatively high proportion of the population (Boyd et al., 1987). Early reports suggested LVNC had a very poor prognosis with significant risk of embolic events, cardiac failure, and arrhythmia/sudden death (Oechslin et al., 2000). These can certainly all occur but subsequent studies (Oechslin and Jenni, 2011) suggest these early reports overstate these risks for the majority of patients.

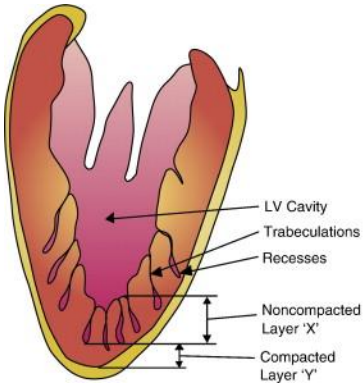


Figure 4-33: Cartoon of a cross-section of the left ventricle showing increased trabeculations and the measurement of compacted and non-compacted layers of the myocardium (Sarma et al., 2010). A ratio of non-compacted to compacted myocardium of >2 is considered to be abnormal.

There has been much speculation as to the underlying causes of non-compaction, but it is increasingly being considered a heterogeneous disorder which can either be a congenital developmental abnormality, or a later adverse modelling response of the heart more akin to the other main cardiomyopathy types. In the majority of patients with LVNC the aetiology is unknown (Ackerman et al., 2011). In both paediatric and adult populations it is recognised as a primary genetic disorder in the majority, with approximately two-thirds

being familial (Hoedemaekers et al., 2010). Heterozygous mutations in *MYBPC3*, *MYH7*, *ACTC*, *TNNT2* and *TPM1* have been identified causing LVNC with high penetrance, accounting for up to a third of apparently sporadic LVNC and up to a half of familial LVNC (Probst et al., 2011, Klaassen et al., 2008).

Occasional patients with LVNC, often with left ventricular dilatation and impairment, have been found to have mutations in Cypher/*ZASP*/*LDB3*, a component of the skeletal and cardiac Z-line (Vatta et al., 2003). Cardiac calcium-handling proteins have also been implicated in LVNC, but have not yet been identified as causing LVNC in a significant number of patients. Recently a NOTCH pathway regulatory gene, *MIB1*, has been identified as a further autosomal dominantly inherited cause of LVNC in a small number of patients (Luxan et al., 2013). In addition a locus for isolated LVNC has been reported at 11p15 but no specific gene has yet been identified (Sasse-Klaassen et al., 2004).

4.5.3 Genetic Studies Performed

Due to the finding of mutations in genes of the sarcomere in LVNC a full clinical diagnostic panel of hypertrophic cardiomyopathy causing genes (*MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1, PLN, CSRP3, FHL1, PRKAG2* and *GLA*) had been undertaken prior to including the family in this study.

4.5.4 Results

4.5.4.1 WGS with Segregation Data

Family members who had whole genome sequencing or SNP array are shown in Figure 4-34. Following analysis of variants produced from WGS and SNP array in affected family members, as described previously, five variants remained (Table 4-12; chromosome linkage plots with genes in which variants have been identified are shown in Figure 4-35).

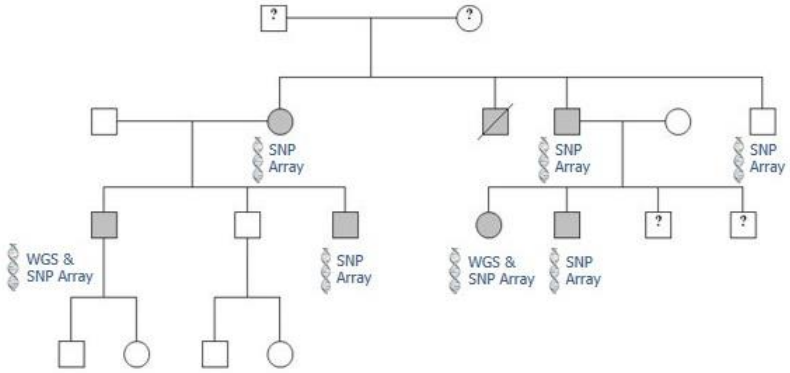


Figure 4-34: Pedigree showing who in the family with autosomal dominant LVNC had whole genome sequencing and/or SNP array analysis.

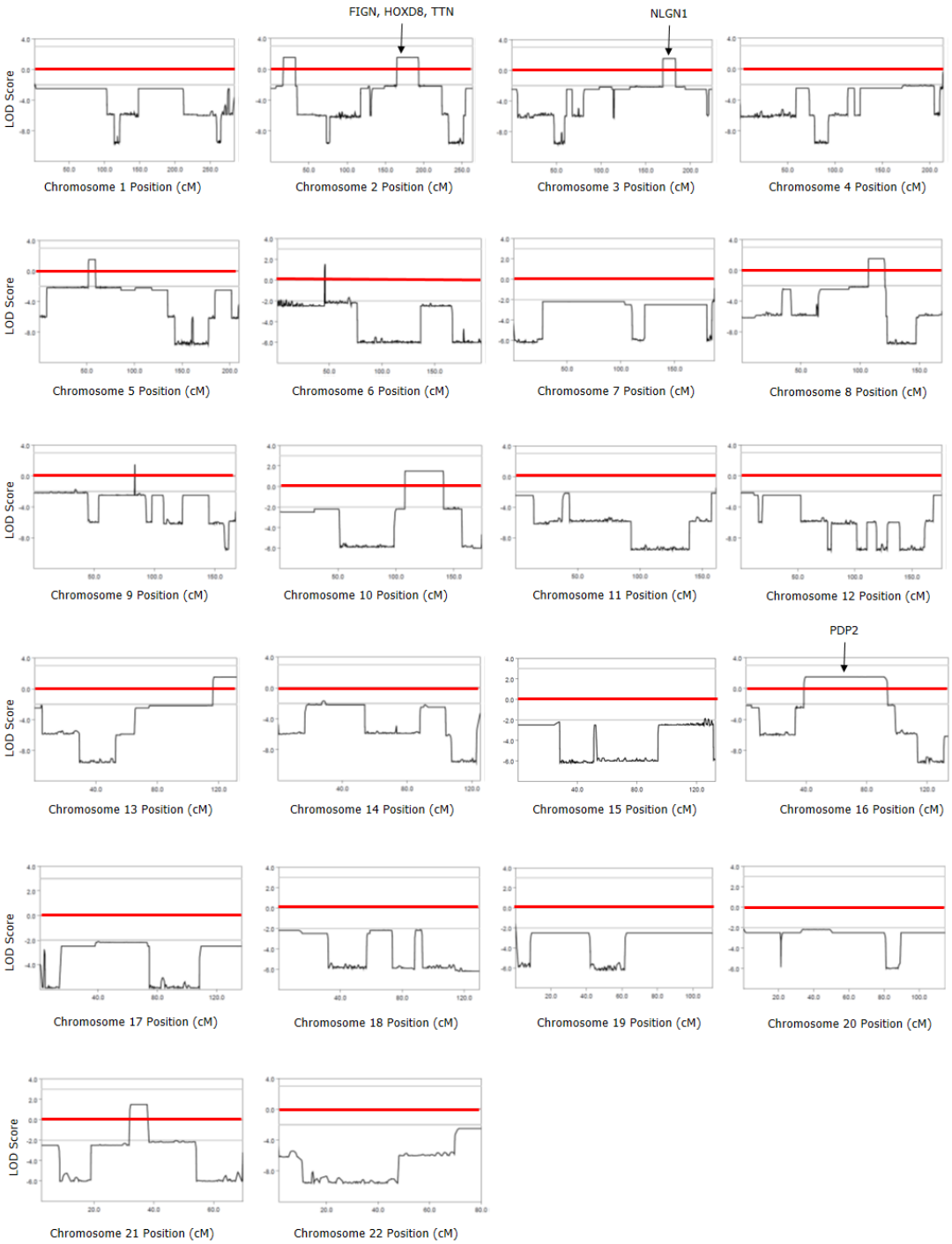


Figure 4-35: Linkage plots for each chromosome from SNP array data analysed in MERLIN software are shown (Abecasis et al., 2002). Genes containing variants identified from WGS studies are highlighted. Red lines indicate a LOD score of 0.

Gene	Variant	Conservation (PhastCons/PhyloP)	Other
FIGN	p.T473I	All species except dog (1, 4.56)	Seen once in EVS
HOXD8	p.A134V	All species (1, 5.05)	
TTN	**p.A178D**	**All species (1, 4.64)**	
NLGN1	p.N568S	All mammals (1, 5.05)	Not present in unaffected family member
PDP2	**p.E316K**	**All species (1, 6.26)**	

Table 4-12: Variants identified following WGS and SNP array in affected members of family with autosomal dominant LVNC. Stronger candidate variants for causing the phenotype are highlighted in bold typeface.

Summary of the known roles of the five genes in which variants have been identified:

p.T473I in Fidgetin (FIGN)

Mutations have been recognised in the mouse homologue for *FIGN*, a member of the ATPase associated with diverse cellular activities (AAA) family that has roles in forming the mitotic spindle (Cox et al., 2000, Mukherjee et al., 2012). These mice have been well phenotypically studied and noted to have skeletal, ophthalmic and vestibular/audiological problems which lead to the behaviours for which the protein is named, but no cardiac phenotype (Truslove, 1956). Despite fidgetin being widely expressed no significant cardiac expression is reported (Husar Bioinformatics Lab, April 2013, Liu et al., 2008).

p.A134V in Homeobox D8 (HOXD8)

Homeobox genes are a family of transcription factor genes important in segmentation in embryonic development. Homeobox genes typically have very localised roles, and HOXD8 only appears to be significantly expressed in the urogenital tracts (Redline et al., 1992).

p.A178D in Titin (TTN)

Titin is a key structural and signalling protein of the sarcomere in striated muscle, it is highly expressed in skeletal and cardiac muscle. Heterozygous mutations in titin have been previously reported as a cause of other cardiomyopathies, particularly DCM (Herman et al.,

2012), and skeletal myopathies (Pfeffer et al., 2012, Hackman et al., 2002). This variant in relation to the known biological functions of titin is discussed in detail in the following chapter.

p.N568S in Neuroligin 1 (NLGN1)

Neuroligin 1 has a known function in forming neural synapse junctions (Varoqueaux et al., 2006). It is expressed in most tissues, including cardiac muscle, but is most highly expressed in areas of the brain. Dysfunction has been implicated in patients with autism and/or learning difficulties (Millson et al., 2012, Glessner et al., 2009).

p.E316K in Pyruvate dehydrogenase phosphatase catalytic subunit 2 (PDP2)

Pyruvate dehydrogenase phosphatase catalytic subunit-2 (*PDP2*) is a gene with high homology to the *PDP1* gene on chromosome 8. PDP1 assists in catalysing the phosphorylation and activation of the mitochondrial pyruvate dehydrogenase multi-enzyme complex. Homozygous mutations in *PDP1* have been identified in a small number of patients with a 'mitochondrial' phenotype (lactic acidosis, neurological problems and infantile death) (Cameron et al., 2009). A similar phenotype is also seen in pyruvate dehydrogenase deficiency itself, with no cardiac involvement described (Patel et al., 2012). PDP2 is expressed in several tissues including the heart, with highest expression in the brain and nervous system and smooth muscle (Husar Bioinformatics Lab, April 2013).

Interpretation of variant findings

The variant in *TTN* is the clear candidate for the phenotype seen due to its known function and association as a cause of a variety of cardiomyopathies. Although most of the other variants are in genes it is easy to dismiss on the basis of known roles it is hard to completely dismiss the variant in *PDP2*. It would be anticipated from what is known that if mutations in this gene were to cause a mitochondrial phenotype it would be a more severe multisystem

disorder, due to an effect on pyruvate dehydrogenase. And if this were to be the case it would be expected to be a recessive, rather than dominant, disorder.

4.5.4.2 Follow-Up Studies

The variants in *TTN* and *PDP2* have been confirmed with Sanger sequencing (chromatogram confirming the titin variant is shown in Figure 4-36). The segregation of these two variants in the family has also been checked, including in the unaffected adult family member (II⑬) and the father who is probably affected but has not been included due to coexistent heart disease complicating the clinical phenotype (I⑤). Both variants were found to segregate together in all of these individuals, being absent in II⑬ and present in I⑤ (see Figure 4-37).

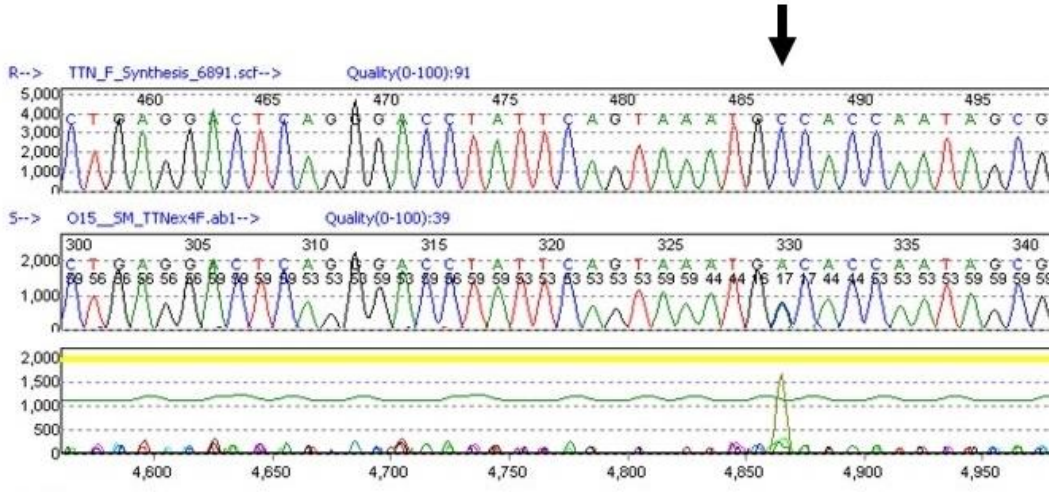


Figure 4-36: Chromatogram from Sanger sequencing confirming the presence of the *TTN* variant, identified from whole genome sequencing, in a family with an autosomal dominant LVNC phenotype of prolonged QT interval. The reference sequence is shown at the top with the test sequence from the forward primer read below, the variant was also confirmed with the reverse primer (analysed and imaged using MutationSurveyor software).

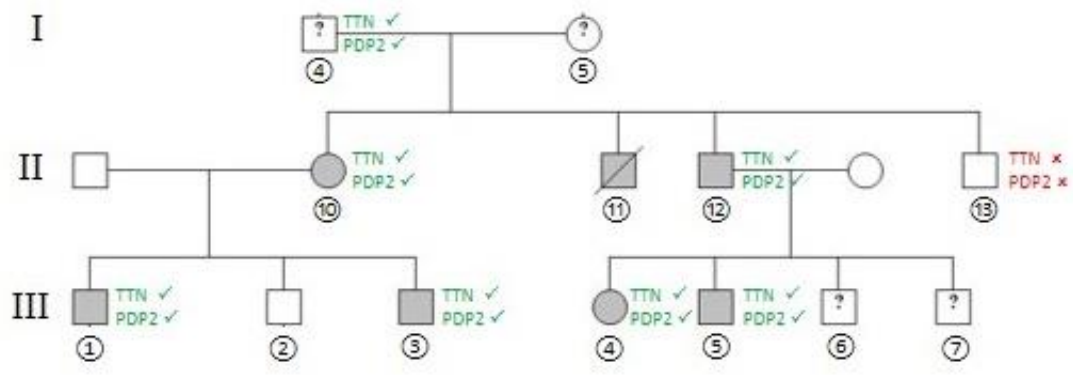


Figure 4-37: Focussed pedigree showing segregation of *TTN* and *PDP2* variants (tick-present, cross-absent; refers to patient to the left)

Thirteen patients with LVNC who attend the Oxford Inherited Cardiac Conditions Clinic were identified and consented into this study to allow research sequencing. Four of these patients have clearly familial disease, and in seven patients the condition was apparently sporadic. The region encoding the titin Z1Z2 region (exons 2-4), and the single large coding exon of *PDP2,* were sequenced in this cohort (primers are shown in appendix sections III-g and h, page 257). No variants, other than known polymorphisms, were identified from this sequencing in any of these patients.

Further detailed funcitonal studies of the titin variant have been performed and are described in the subsequent chapter. The variant in titin is a far stronger candidate for causing LVNC and has therefore been focussed on. After excluding elevated blood lactate levels further studies of PDP2 have not been undertaken.

Chapter 5:

In Vitro Studies of the Titin Z1Z2 Region

From the sequencing and segregation data presented in chapter 3, a heterozygous variant in the gene encoding titin, c.533G>T/p.A178D, has been identified as the likely cause for the left ventricular non-compaction (LVNC) phenotype observed in this family. This is highly plausible due to the recognition that mutations in other sarcomeric protein genes are known to cause LVNC (Klaassen et al., 2008) and other missense mutations in the titin gene have previously been identified causing other forms of myopathy and cardiomyopathy (Granzier and Labeit, 2004).

This variant at position 178 changes alanine to aspartate, a residue within the Z2 Ig-like domain of titin. The Z2 domain localises adjacent to the Z-disk and is an important binding site for another sarcomeric protein, telethonin (also known as T-cap). This chapter will give a broad overview of the known functions of titin, concentrating on the Z2 domain and its relationships with telethonin and other proteins of the Z-disk. Work performed to ascertain the effects of the p.A178D variant *in vitro* will be presented.

5.1 Titin Overview

Titin is encoded by the *TTN* gene, located on the long arm of chromosome 2 at 2q24.3 (Pelin et al., 1997). The 283kb genomic segment encompassing the gene contains 363 exons with a coding capacity of 114 414 base pairs (Bang et al., 2001). This sheer size of the titin gene made it technically very challenging to study through sequencing, prior to next-generation methods. This size also makes the interpretation of variants found extremely difficult, with over 20 missense variants typically identified in each individual (EVS, March 2013). Titin is highly expressed in striated muscle, and is the third most abundant protein

after myosin and actin. It is the largest protein known, with a molecular weight of up to 4MDa and a length of up to 1µm.

5.1.1 Structure

This rope-like protein, with 244 individually folded domains linked by unstructured peptide sequences, spans a half sarcomere and links the Z-disk to the M-line. The N-terminus of titin is anchored in the Z-disk of the sarcomere, followed by an elastic I-band region, a thick filament binding A-band, and then the C-terminus is embedded in the M-line (Figure 5-1). Each of these regions has a different structural composition to allow local variation in mechanical, binding, and catalytic capabilities.

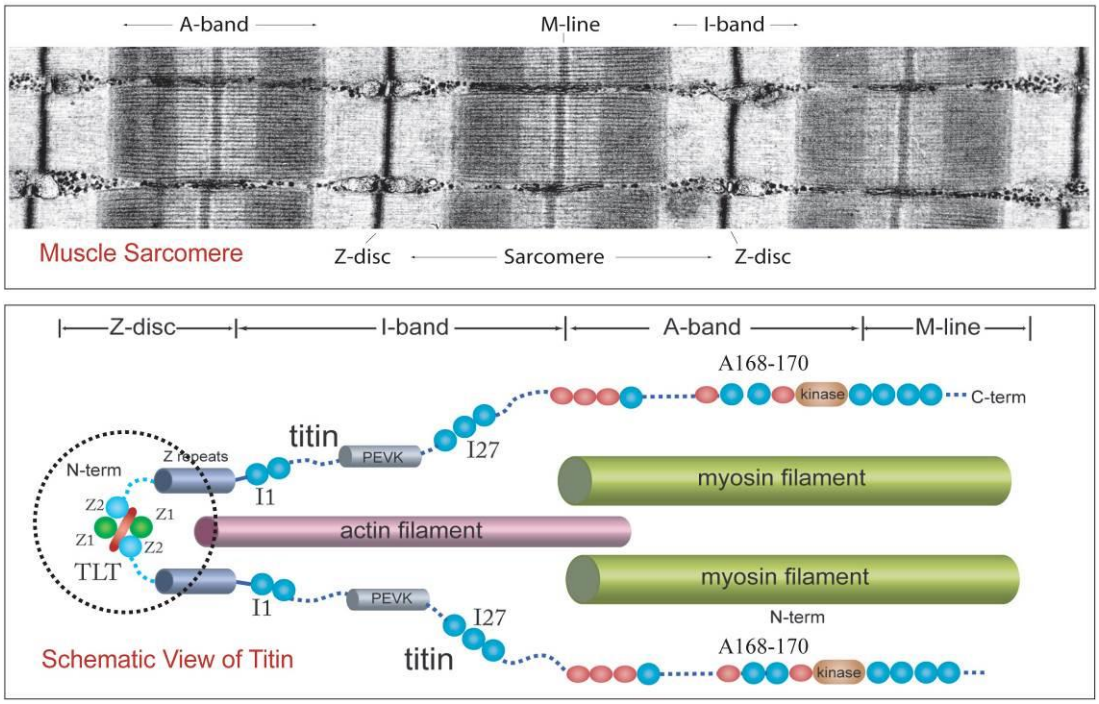


Figure 5-1: Schematic representation of titin across the sarcomere (Lee et al., 2007). The Z1Z2 is highlighted with a dashed circle.

Titin Z-disk region

The Z-disk region of titin is predominantly composed of immunoglobulin (Ig)-like domains. Ig-like domains are typically about 100 amino acids and consist of two β-sheets. Although some of these domains are ubiquitous to all isoforms of titin, others vary between different

muscles and different species (Peckham et al., 1997). The most N-terminal domains are the Ig-like Z1 and Z2 domains (Z1Z2) (discussed in detail in section 5.2.1, page 140). The Z1Z2 region is followed by proline-rich 'Zis' sequences. This region has been shown to bind with nebulin and filamin in smooth and skeletal muscle (Figure 5-2) (Labeit et al., 2006), although the relevance of these interactions to cardiac muscle is unclear. There is then a variable region of 'Z-repeats' which embeds titin in the Z-disk, linking titin to the thin filament and the plasma membrane through α-actinin (Ohtsuka et al., 1997). Typically there are between 200 and 700 residues of the Z-repeat sequence traversing the Z-disk. The number of repeats affects the size and composition of the Z-disk, which in turn affects sarcomere 'stiffness' (Peckham et al., 1997, Knoll et al., 2011a). Following on from the Z-repeats are further Ig-like Z domains, including a putative obscurin binding site (Young et al., 2001).

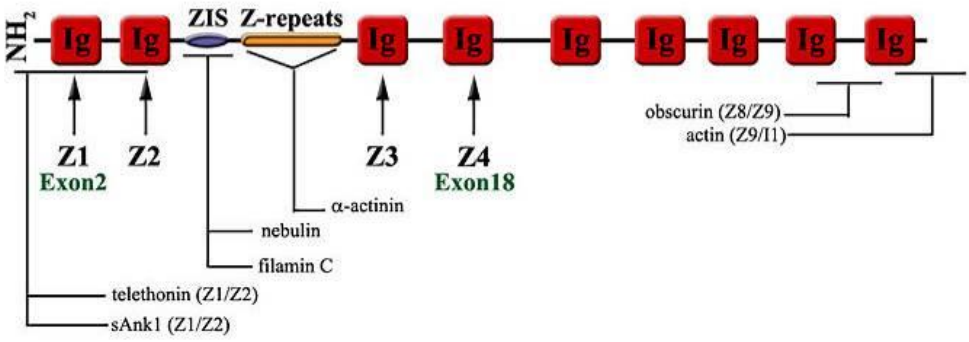


Figure 5-2: Proposed titin Z-disk region binding partners (Linke, 2008)

Titin I-band region

It is the I-band region that provides much of the spring-like quality of titin. At either end of the I-band are highly conserved tandem repeats of Ig-like domains. Between these domains is a highly variable region incorporating the N2 region, composed of Ig-like domains, and then the proline, glutamate, valine, and lysine-rich (PEVK) domains. The N2 region is one of the main differentiators of titin isoforms (Figure 5-3; discussed further in section 5.1.2, page 135). Skeletal muscle isoforms predominantly have an N2A sequence,

whereas cardiac muscle isoforms have N2B or N2BA sequences (Linke, 2008). The following PEVK domains are random coils of variable length and composition (Linke et al., 2002).

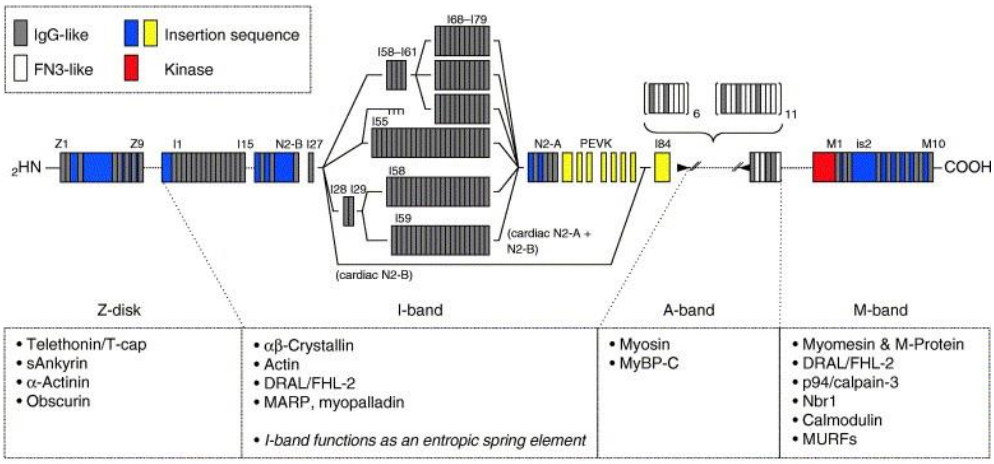


Figure 5-3: The regions of titin including binding partners and the main differential splice isoforms of the I-band (Lange et al., 2006)

Titin A-band region

The A-band region constitutes the majority of the titin protein. Here it aligns with the thick filament and provides stiffness to the myosin backbone (Tskhovrebova and Trinick, 2003). This region is relatively inextensible and is predominantly composed of fibronectin type-III domains, with further Ig-like and other domains interspersed. The Ig-like and fibronectin domains are generally much more resistant to unfolding. Most parts of the A-band region are highly conserved and have long-range 'super' repeats in their structure, these are of the appropriate length to complement the thick filament (Labeit et al., 1992). The main interactions of this region are with myosin and myosin binding-protein C (Labeit et al., 1992, Freiburg and Gautel, 1996).

Titin M-line region

This region is again composed predominantly of Ig-like domains. These are important for sarcomere assembly and stability, binding to myomesin which anchors the M-line (Nave et al., 1989). One of the key features of this region, and indeed of titin generally, is a serine-threonine titin kinase (TK) domain. The full functions of this domain are still being

established, but it has roles in myofibrillogenesis (through phosphorylation of substrates (Mayans et al., 1998) and by direct protein-protein interactions (Weinert et al., 2006)), sarcomere maintenance (Gotthardt et al., 2003), altering signalling factors in response to mechanical conformational changes (Lange et al., 2005), and altering contractility in response to calcium signalling (Ottenheijm et al., 2009). Other binding partners of this region include obscurin (Fukuzawa et al., 2008), and other signalling proteins, such as MuRF1 which may regulate the kinase domain (Centner et al., 2001)

5.1.2 Isoforms

Unlike most sarcomeric proteins, where there are multiple highly homologous genes expressed at different stages of development and in different tissues, titin is encoded by a single gene (*TTN*). However, this gene is subject to differential splicing to give it developmental and tissue specific variation (Granzier and Labeit, 2004). Expression of different titin isoforms allows control over aspects of the function of titin, providing adaption of its properties in different muscles and at different stages of development. The ration between different isoforms of titin may also change in pathological states of muscle (Lahmers et al., 2004).

Only a small number of titin domains are ubiquitous to all isoforms. The domains which are maintained are generally those within inextensible regions, these typically have regulatory or signalling functions and are less important for the mechanical properties of titin (Granzier and Labeit, 2004). The domains of the I-band are the most variable between isoforms, as these have the most significant effects on the mechanical properties of titin (discussed further in section 5.1.3.2, page 137) (Bang et al., 2001).

Cardiac isoforms of titin can be divided into N2B and N2BA isoforms (Figure 5-3). N2B isoforms incorporate the N2B sequence, but not a stretch of other I-band regions, leading to a shorter ‘stiffer’ isoform overall. The N2B sequence is encoded by exon 49 and is

present in all cardiac isoforms, but is not present in non-cardiac isoforms. The N2B region acts to lower passive force levels at the upper limit of the physiological sarcomere length. This limits energy loss by limiting Ig-like domain unfolding, and therefore reduces the requirement for refolding (Granzier and Labeit, 2004). It is also a substrate for protein kinase A, with phosphorylation reducing passive tension which is thought to allow a mechanism for adrenergic control of diastolic function to allow more complete diastolic ventricular filling (Yamasaki et al., 2002). N2BA isoforms incorporate the N2A element, a longer PEVK segment, and further Ig-like domains, in addition to the N2B element, giving a longer more compliant isoform (Bang et al., 2001). The PEVK elements in N2BA isoforms have E-rich motifs which are thought to be key in making these variants calcium responsive (Granzier and Labeit, 2004). In adult left ventricle N2B isoforms predominate to give a high level of passive stiffness, with about two-thirds of isoforms being N2B and one-third being N2BA (Neagoe et al., 2002).

5.1.3 Functions

5.1.3.1 Structural Functions

In striated muscle titin has a key role as the 'molecular ruler', regulating sarcomere filament length and acting as the backbone for sarcomere assembly (Whiting et al., 1989). This hypothesis was initially proposed due to the location and structure of titin, but further support has come from the recognition that titin is correctly located very early in myofibrillogenesis (Ehler et al., 1999), and that titin knockout *in vitro* causes abnormal myofibril arrangement (van der Ven et al., 2000a). The link of titin with the thick filament through the A-band, and the ends anchoring in the Z-disk and M-line, are likely key for this (Tskhovrebova and Trinick, 2003).

5.1.3.2 Mechanical Functions

In mature muscle titin provides passive elasticity and limits the range of motion of the sarcomere, acting as a molecular 'spring' (Maruyama et al., 1977). The elastic force of titin alone generates about half of the passive tension of resting muscle, with most of the rest being due to the extracellular collagen network (Makarenko et al., 2004). The whole of titin is important in this role, but it is the I-band region which contributes most. The I-band of titin region is highly folded when no force is present, and then as force is applied it elongates due to unfolding of the PEVK and N2 regions. This exerts passive tension to resist over-extending, and applies a restoring force when sarcomeres shorten (LeWinter et al., 2007). There is also capacity for further elongation by the unfolding of Ig-like and fibronectin domains throughout titin, allowing further reversible extension such that sarcomeres can actually be extended beyond the capabilities of the I-band (Tskhovrebova and Trinick, 2003).

5.1.3.3 Signalling Functions

Due to its size, structure, location, and protein interactions, titin is ideally positioned to integrate biomechanical information, with the TK domain playing a key role in sensing strain (Puchner et al., 2008), and then coordinate an appropriate response, through a number of protein interactions that then act within a complex signalling network (Figure 5-4) (Voelkel and Linke, 2011).

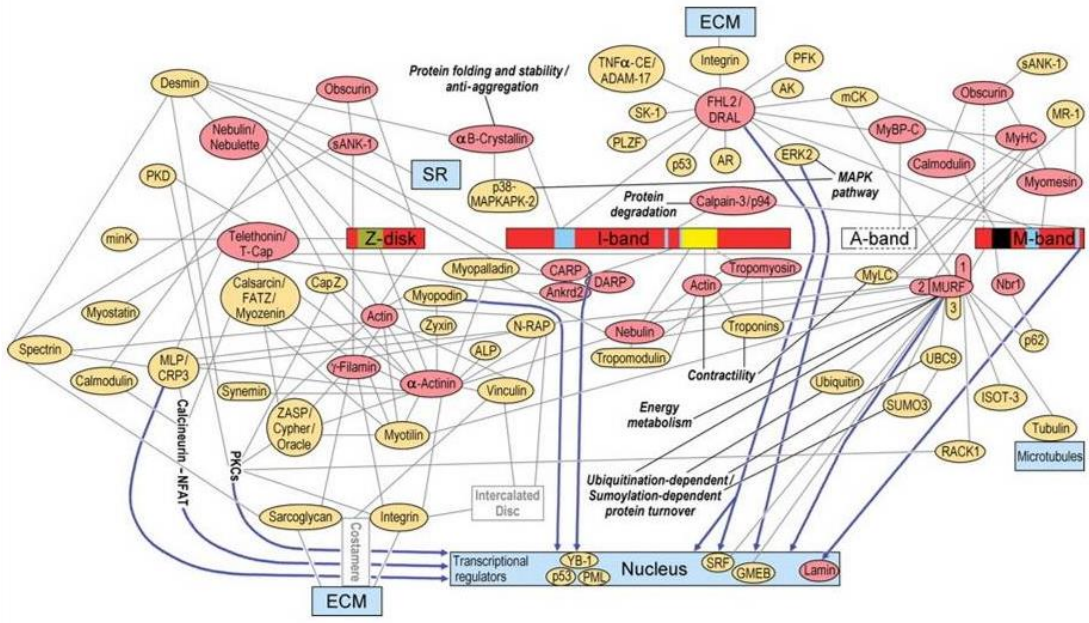


Figure 5-4: Proposed protein interactions of different regions of titin (direct interactions are highlighted with a red background, indirect interactions have a yellow background). This allows the integration of biomechanical information to be communicated to a number of signalling pathways (Linke, 2008).

Several of these pathways link titin to hypertrophy regulation and the response to biomechanical stress. The roles of the Z1Z2 region of titin in these pathways, through its interaction with telethonin (Faulkner et al., 2001) and as part of the putative titin-telethonin-muscle LIM protein (MLP) stretch sensor complex (Knoll et al., 2002), are discussed in more detail in section 5.2.2.3 (page 143) (Knoll et al., 2002). The N2B and M-line regions of titin have been shown to interact with the four-and-a-half LIM domain proteins (FHL) 1 and 2 (Lange et al., 2002, Sheikh et al., 2008), which again have been shown to be important in the hypertrophic response (Kong et al., 2001). The N2A region binds to certain muscle ankyrin repeat proteins, and in turn myopalladin. These have been shown to be regulators of gene expression and may also be important in the stress response (Miller et al., 2003b).

Some of these pathways also connect titin to the regulation of protein turnover. There are links to the ubiquitin-proteosome degradation system, for example through the titin Z1Z2-telethonin complex interacting with mouse double minute 2 (MDM2) (Tian et al., 2006) and

muscle ring finger protein 1 (MuRF1) (McElhinny et al., 2002) and direct titin M-line interactions with muscle ring finger proteins (Centner et al., 2001). Furthermore the N2B region of titin interacts with the calcium-dependant calpain protease, calpain-1, which appears to be important for dissociating sarcomeric proteins to allow the UPS to act (Willis et al., 2009). There are also links with heat-shock protein pathways, such as with α-B crystalline, which may play a role in preventing stress induced sarcomere degradation (Kruger and Linke, 2011).

5.1.3.4 Other Functions

Non-muscle roles of titin are more controversial, and often evidence for these is conflicting. Some studies suggest that certain small isoforms are expressed in other tissues, such as connective tissue (Qi et al., 2008). It has been reported that there is a nuclear isoform of titin involved in regulation and maintenance of chromosomes, chromosome spindles, and nuclear membranes (Machado and Andrew, 2000, Zastrow et al., 2006, Fabian et al., 2007). Here it was proposed to bind to actin and myosin intra-nuclear filaments, but this was not identified in studies in chromosomes of rapidly dividing cells (Takata et al., 2007). A nuclear localisation sequence has been proposed at amino acids 200-206, with overexpression of the Z1Z2-Zis region in an osteoblastic cell line leading to increased cellular proliferation and reduced aggregation due to activation of the Wnt/β-catenin pathway (Qi et al., 2008). Telethonin is suggested, but not tested, as a blocker of this nuclear localisation sequence in cardiac myocytes (Qi et al., 2008).

5.1.4 Regulation

Phosphorylation has been recognised as one of the key regulatory mechanisms of titin. Phosphorylation sites are recognised most notably the TK domain at the M-line, but also within the Z-disk and I-band regions (Kruger and Linke, 2011). The TK domain has been

established as the hub of an M-line protein signalling network important for adaptation to mechanical stress, gene expression, and protein turnover (Lange et al., 2005) (Figure 5-5).

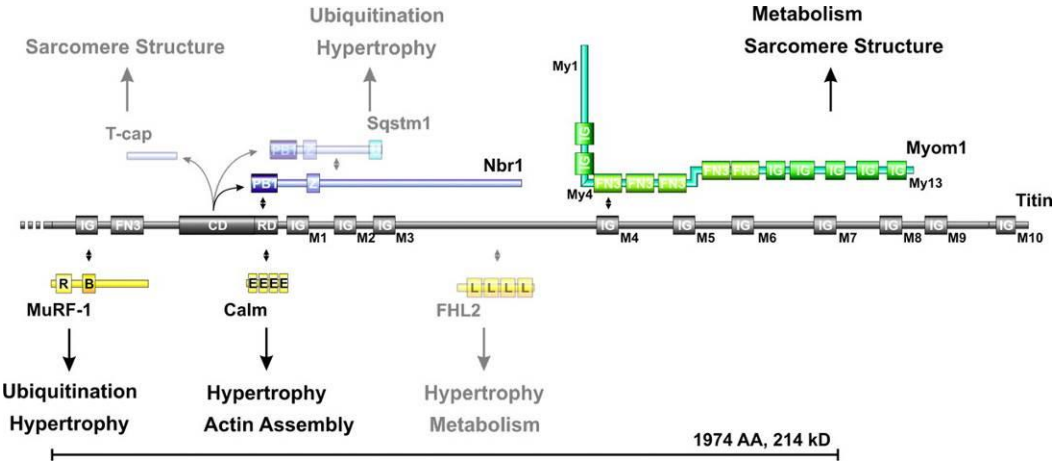


Figure 5-5: Signalling functions of the M-line region of titin including the titin kinase domain (CD + RD) (Weinert et al., 2006)

Some of the N-terminal Zis sequences contain SPXR motifs, which have been shown to be phosphorylated by ERK 1/2 and Cdc2 *in vitro* (Gautel et al., 1996). Phosphorylation at these sites may be important for integration of titin in the Z-disk and myofibrillogenesis (Kruger and Linke, 2011). In the I-band PEVK region, serine residues can be phosphorylated by calcium dependent PKC to increase passive tension (Hidalgo et al., 2009). Phosphorylation of a serine residue in the cardiac N2B sequence by PKA/PKG reduces passive tension and may be important in rapid regulation of this (Kruger et al., 2009). Phosphorylation of N2A also occurs and may modulate signalling pathways (Kruger et al., 2009).

5.2 The Titin Z1Z2 Region and Telethonin

5.2.1 Titin Z1Z2 Structure

As has been discussed the titin variant identified, p.A178D (A178D), is within the Z2 domain, an Ig-like domain at the very N-terminus of titin. It is part of the Z1Z2 domain tandem (Z1Z2) which is composed of 194 residues (Z1 being 98 residues, Z2 93 residues, and a linker of 3 residues). It was initially thought that the N-terminus of titin was embedded in the Z-disk (Gautel et al., 1996), however it was clarified that titin completely

traverses the Z-disk with the Z-repeats embedding within the Z-disk and the Z1Z2 region extends beyond to the I-Z junction of the adjacent sarcomere (Gregorio et al., 1998).

Crystallographic data has shown that Z1Z2 has two conformational states under crystallography conditions. Depending on the angle of the link between the two domains either a V-shape or a linear configuration are adopted (Figure 5-6: A and B) (Lee et al., 2007). However, nuclear magnetic resonance (NMR) residual dipolar coupling and small-angle X-ray scattering (SAXS) have demonstrated that Z1Z2 has an intermediary structure to these configurations in solution (Figure 5-6: C) (Marino et al., 2006, Lee et al., 2007).

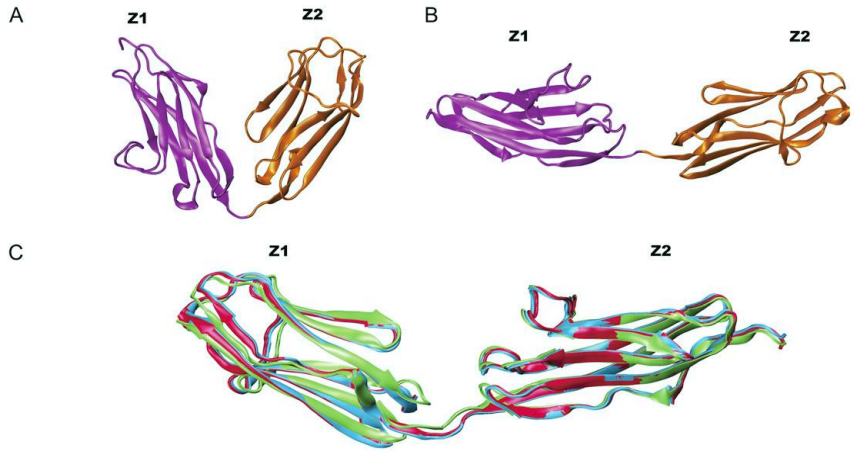


Figure 5-6: Schematic of the conformations of the Z1Z2 domains. A shows the closed and B the open conformations adopted under crystallography conditions. C demonstrates the intermediate conformation identified in solution by NMR and SAXS (Lee et al., 2007).

Molecular dynamics simulations predict that the linker region between these domains can bend, and possibly twist, to adopt differing stable tertiary structures at relatively low forces. In addition the Z2 domain, but not Z1, may exhibit an intermediate confirmation on stretching at relatively low forces (Garcia-Manyes et al., 2012, Lee et al., 2007). In isolation Z1Z2 is a relatively weak point of titin, unfolding at very modest forces, but on binding to its ligand telethonin this region is greatly stabilised (Bertz et al., 2009, Kollar et al., 2010, Lee et al., 2006).

5.2.2 The Titin Z1Z2/Telethonin Complex

5.2.2.1 Telethonin

Telethonin is a relatively small protein, 167 amino acids, with a unique β-sheet structure (Lee et al., 2006). It is encoded by the *TCAP* gene on chromosome 17q12, and is one of the most abundant transcripts in striated muscle (Valle et al., 1997). There is evidence for its involvement in myofibril assembly and maintenance, and signalling pathways (Mason et al., 1999, Schroder et al., 2001, Mayans et al., 1998, Mues et al., 1998, Knoll et al., 2002). In mature sarcomeres telethonin localises to the Z-disk, where it binds to titin and a number of other proteins (Faulkner et al., 2001).

5.2.2.2 Titin Z1Z2-Telethonin Binding

Binding studies have demonstrated that telethonin binds specifically to Z1Z2, but not to either domain individually (Mues et al., 1998). The conformation of Z1Z2 affects its ability to bind telethonin, with the semi-extended conformation of Z1Z2 favouring this interaction (Marino et al., 2006). Only the N-terminal 90 residues of telethonin are required for this interaction, and *in vitro* additional C-terminal residues may inhibit this (Gregorio et al., 1998, Zou et al., 2003, Pinotsis et al., 2006). *In vitro* telethonin is unstable, forming aggregates in solution, but becomes stable in complex with Z1Z2 (Zou et al., 2003, Pinotsis et al., 2006).

Two Z1Z2 regions bind to telethonin in an antiparallel 'sandwich', with telethonin forming an elongated structure between two Z1Z2 regions with opposing polarity (Zou et al., 2006, Zou et al., 2003, Mues et al., 1998, Gregorio et al., 1998). Z1Z2 and telethonin are both predominantly composed of β-sheets. When these proteins interact the β-sheets align and form multiple cross-linking hydrogen bonds between them (shown in Figure 5-7 image B) (Zou et al., 2006, Lee et al., 2006). These bonds are particularly focussed on parts of the interaction between the Z2 domain and telethonin, as this point is placed under most stress

due to the orientation of this complex *in situ* (Bertz et al., 2009). This complex is highly resistant to stretching forces, indeed this interaction is reported to be the strongest known biological bond (Bertz et al., 2009).

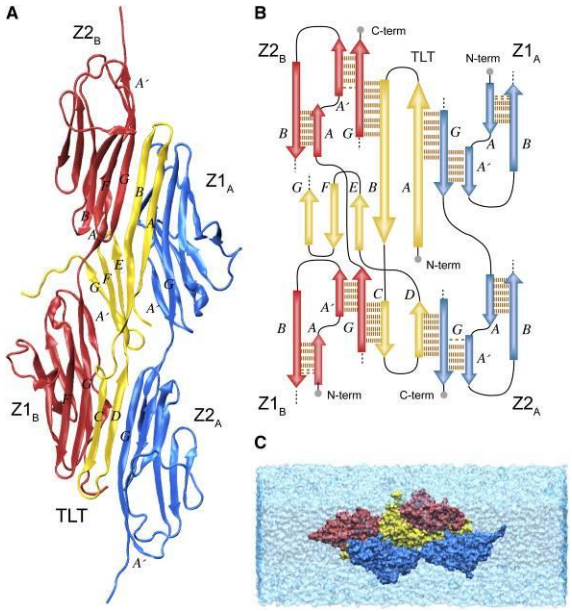


Figure 5-7: Structure of the titin Z1Z2-telethonin (TLT) complex. A and B show different representations of the β-sheet interactions between Z1Z2 and telethonin. C shows a surface view of the solvated Z1Z2-telethonin complex (Lee et al., 2006).

5.2.2.3 Roles of the Titin Z1Z2-Telethonin Complex

The Z-disk region of titin is ideally positioned to integrate multiple biomechanical inputs, not just from itself but also tension transmitted through thin and thick filaments and through links with the cell membrane, sarcoplasmic reticulum (SR), t-tubules, and ion channels. Titin can then communicate these inputs to the protein signalling network associating at the Z-disk (Granzier and Labeit, 2004). The observed strength of the interaction between Z1Z2 and telethonin would suggest it is of functional importance. However, the biological effects of disrupting this interaction *in vitro* and *in situ* are very variable depending on the situation observed.

Sarcomere assembly and maintenance

Certain structures of the Z-disk, such as α-actinin and Z-line titin, appear very early in myofibrillogenesis and are likely important for sarcomere assembly (Ehler et al., 1999). This dense protein network is highly resistant to mechanical stress (Gautel, 2011), enabling the transmission and transduction of forces across it (Knoll et al., 2011a). The main role of the Z1Z2-telethonin complex was initially believed to be in anchoring titin, to allow titin to act as the backbone for sarcomere assembly and integrity. Some *in vitro* experiments give support for this hypothesis, with overexpression or depletion of components of the complex causing disruption of the sarcomere structure in cultured cardiomyocytes (Gregorio et al., 1998, Markert et al., 2008, Zou et al., 2006). However, telethonin remains diffusely distributed until the final stages of myofibril maturation, arguing against it being important in early sarcomere assembly (Wang et al., 2005, Zhang et al., 2009). Disruption of this complex *in situ*, be that animal models or human phenotypes, demonstrates limited evidence of an effect on myofibrillogenesis.

Depletion of the telethonin homologue in zebrafish leads to a spontaneous early muscular dystrophy phenotype, but with different models exhibiting conflicting evidence regarding the degree of sarcomere disorganisation (Zhang et al., 2009, Sadikot et al., 2010). Telethonin knockout mice have normal muscle development in the first year and normal force generation capabilities, although an increase in passive stiffness, before developing a relatively mild later-onset generalised dystrophy phenotype (Markert et al., 2010). At three months of age telethonin knockout mice demonstrate are largely equivalent to the wild-type mice, including exhibiting normal Z-disk architecture and localisation of the titin N-terminus (Knoll et al., 2011b). What is evident in these mice is a reduction in myofibril fusion and maintenance, although it is suggested that this may be due to signalling effects rather a direct structural effect (Markert et al., 2008).

Homozygous nonsense mutations in telethonin cause a relatively mild limb-girdle muscular dystrophy with cardiac involvement (Moreira et al., 2000, Olive et al., 2008), and although muscle immunofluorescence shows an absence of sarcomeric telethonin there is otherwise normal sarcomeric architecture (Vainzof et al., 2002). Other human phenotypes where the Z1Z2-telethonin is reported to be disrupted are discussed in more detail as part of section 5.3 (page 149).

Signalling pathways

The Z-disk is one of the most densely packed and complex cellular structures in eukaryotic cells. It acts as a link between predominantly structural proteins (such as actin, titin, α-actinin, and nebulin/nebulette) and an extensive and complex network of signalling proteins that associate around the Z-disk (Figure 5-8). In addition to telethonin some of the key signalling proteins that localise to the Z-disk include muscle LIM protein (MLP), cypher/ZASP, filamin, calsarcin/myozenin/FATZ, FHL 1 and 2, Bcl2-associate athanogene (BAG3), α-B crystalline, calpain, MuRFs, ALP, myopalladin, calcineurin, and of course telethonin (Faulkner et al., 2001, Knoll et al., 2011a). Many of these proteins are mobile within the sarcomere and localise in response to mechanical stress (Lange et al., 2006). Other proteins may associate with the Z-disk at times, or as one of many roles, such that over two-hundred proteins are reported to be associated with the Z-disk in the NCBI database (NCBI, June 2013).

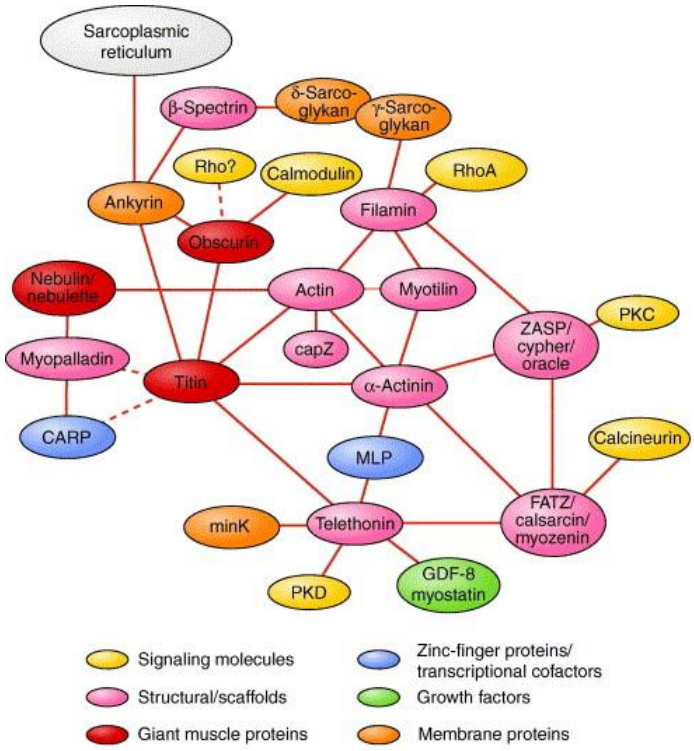


Figure 5-8: Schematic map demonstrating the complex network of interacting proteins of the Z-disk (Lange et al., 2006)

Many of the Z-disk associated proteins which interact with the Z1Z2-telethonin complex appear to have roles in regulating apoptosis/proliferation and atrophy/hypertrophy pathways. Some of these pathways may have greater physiological relevance than previously thought, with the relatively recent finding that cardiomyocytes continue to turnover through life rather than being static (Bergmann et al., 2009). The relevance of telethonin in these pathways is supported by the finding that telethonin knockout mice rapidly develop cardiac hypertrophy and failure after aortic banding, with myocytes showing an increased rate of apoptosis. Inhibition of p53, a critical controller of the cell cycle and initiator of apoptosis, reduced the observed phenotype in these mice (Knoll et al., 2011b).

The Z1Z2-telethonin complex, in combination with MLP, forms a putative stretch sensor (Knoll et al., 2002). MLP, encoded by *CSRP3*, is reported to have roles in sarcomere formation and maintenance, calcium sensitivity, myocyte differentiation and hypertrophy

(Knoll et al., 2011a, Arber et al., 1994). Mutations in *CSRP3* have been reported causing HCM (Geier et al., 2003) and DCM (Knoll et al., 2002), and MLP knockout mice are reported to show left ventricular dilatation (Arber et al., 1997). However, there is evidence that this complex may not be essential for normal function. Some variants reported to disrupt this complex, in both telethonin (p.E13del) and MLP (p.W4R) (Bos et al., 2006, Knoll et al., 2002, Knoll et al., 2011b), have been identified at significant levels in the normal population (Andersen et al., 2009, Perrot et al., 2006, Marziliano et al., 2006, EVS, March 2013). This raises questions over the pathogenicity of these variants, and therefore the importance of this protein complex *in situ*.

Telethonin has also been shown to interact with other proteins implicated in hypertrophy and/or apoptotic pathways: calsarcin-1 (Frey and Olson, 2002), myostatin (Nicholas et al., 2002), bone morphogenetic protein-10 (BMP10) (Nakano et al., 2007), Siva (Mihatsch et al., 2009) and the E3 ubiquitin ligases MuRF1 and 2 and MDM2 (Tian et al., 2006, Witt et al., 2005).

Links to calcium homeostasis and ion trafficking

The Z1Z2-telethonin complex may also have a role linking the contractile apparatus with the SR, t-tubules, and specific ion channels. Depletion of the telethonin homologue in zebrafish disrupts the sarcomere-membrane interaction and leads to abnormal t-tubule formation (Zhang et al., 2009, Sadikot et al., 2010). Telethonin binds to MinK, the β-subunit of the slow activating component of the delayed rectifier potassium channel, which also binds to the t-tubules (Furukawa et al., 2001). The interaction between telethonin and MinK may act to provide a mechanism by which stretch mechanisms can affect potassium channel activity in muscle cells (Furukawa et al., 2001). In smooth muscle telethonin also appears to interact with sodium channels (Mazzone et al., 2008), however the relevance of this interaction for cardiac muscle has not been demonstrated.

5.2.2.4 Other Proposed Interactions of the Titin Z1Z2 Region

Apart from the main ligand of Z1Z2, telethonin, an interaction with small ankyrin 1.5 (sAnk1.5) has also been proposed. Ankyrins are proteins that help to organise the attachment of membrane proteins with the spectrin-actin membrane skeleton (Bennett and Chen, 2001). *ANK1* encodes ankyrinR proteins. In striated muscle there is a group of small muscle-specific ANK1 isoforms, where transcription is initiated at an alternative promoter site between exons 39 and 40 (Gallagher and Forget, 1998), with sAnk1.5 the most abundant of these. These small ankyrins lack the membrane and spectrin-binding domains, which suggest they may have specialist intracellular roles.

The evidence for an interaction between Z1Z2 and sAnk1.5 is conflicting. Immunostaining of adult skeletal muscle has suggested that it targets to specific domains of the SR and localises to the M-band and the Z-disk (Zhou et al., 1997). However, *in vitro* studies using cultured myotubes show exclusive localisation at the M-band (Bagnato et al., 2003). An initial report which showed sAnk1.5 binding to Z1Z2 (Kontrogianni-Konstantopoulos and Bloch, 2003) has not been replicated (Armani et al., 2006). The difference in localisation pattern has been explained by the same discrepancy in the localisation pattern of one of sAnk1.5s confirmed binding partners, obscurin. It is also worth noting that in cardiac muscle ankyrin 2 isoforms predominate, rather than the ankyrin 1 isoforms of skeletal muscle, therefore the relevance of this to cardiac muscle is questionable in any case.

5.3 Human Phenotypes Attributable to Mutations in Titin and Telethonin

Mutations in titin, telethonin, and other Z-disk proteins have been reported to cause various cardiomyopathy and myopathy phenotypes (Gerull et al., 2002, Itoh-Satoh et al., 2002, Satoh et al., 1999, Hackman et al., 2002, Hayashi et al., 2004, Moreira et al., 2000, Knoll et al., 2002, Geier et al., 2003). The majority are rare causes of these phenotypes, and many have been studied simply on the basis of their presence in the sarcomere rather than due to strong genetic data (Watkins, 2013). Most are missense variants and often an alteration in binding between Z-disk components is put forward as supportive of pathogenicity. In some of these cases, however, the same variants have also been found in controls or population data, casting doubt on their pathogenicity.

5.3.1 Titin

DCM

Truncating mutations in *TTN* have been recognised as a significant cause of DCM (Gerull et al., 2006, Gerull et al., 2002, Herman et al., 2012, Itoh-Satoh et al., 2002, Yoskovitz et al., 2012). Due to the size of *TTN* missense variants are commonly identified in population cohorts, with an average of approximately twenty-three in each person (EVS, March 2013). This means the interpretation of these variants can be very difficult, and therefore their importance in cardiomyopathy phenotypes is not yet known. Occasionally missense variants have been reported segregating in extensive families with DCM (Itoh-Satoh et al., 2002, Gerull et al., 2002) suggesting that some variants may well be important. Even the interpretation of truncating mutations can be difficult, with the occasional finding of such mutations not segregating with the DCM phenotype in a family, and they too are present in population cohorts (Norton et al., 2013, EVS, March 2013, Golbus et al., 2012).

One of these DCM associated titin missense mutations, p.V54M, has been shown to disrupt the titin-telethonin complex (Itoh-Satoh et al., 2002, Matsumoto et al., 2005). However,

due to the low numbers of segregating missense mutations in this condition the interpretation of the significance of this particular interaction is difficult. Interestingly the majority of pathogenic *TTN* mutations occur in regions also present in skeletal muscle isoforms, but patients do not show signs of a more generalised myopathy. Homozygosity of dominant mutations does not appear to be consistent with live-born animal models (Gotthardt et al., 2003) and is not seen in large families (Gerull et al., 2002). Recessive mutations have rarely been reported, for example a patient with homozygous splicing mutations was reported to have DCM and myopathy (Herman et al., 2012).

HCM

Although the evidence that mutations in titin cause DCM is far stronger, mutations in titin have also been ascribed as a cause of HCM. The role of *TTN* mutations in HCM is still in question however, with no extensive pedigrees of segregating mutations identified and studies of larger cohorts of patients with HCM not identifying titin variants at a higher rate than the normal population (Herman et al., 2012). It has been suggested that mutations causing HCM and DCM may have opposing actions. For example the p.R740L variant, reported as causing HCM, acts to increase binding to α-actinin (Satoh et al., 1999), while the nearby DCM mutation, p.A743V, reduces the binding to α-actinin (Itoh-Satoh et al., 2002). The *TTN* p.S3799Y variant found in two sisters with HCM (Itoh-Satoh et al., 2002) increased binding to FHL2, whereas a nearby DCM causing mutation reduced the same interaction (Matsumoto et al., 2005). However, the pathological mechanisms by which altering binding with these various proteins causes cardiomyopathy are unknown.

ARVC

A study of families with ARVC identified eight missense variants in titin, more commonly in those with conduction disease, with the best supporting segregation data being for the p.T2896I variant (Taylor et al., 2011). Study of this mutation suggested it destabilises the Ig-

like domain-10 (Anderson et al., 2013). Clinical and genetic overlap between DCM and ARVC phenotypes has been recognised (van der Zwaag et al., 2012), but the majority of the patients in this study lacked significant left ventricular involvement and met diagnostic criteria for ARVC (Marcus et al., 2010). As far as I am aware this is the only significant study of titin in ARVC therefore it will be interesting to see if this finding is reproduced more widely.

Myopathies

Rare identification of heterozygous or homozygous, truncating or missense, mutations in *TTN* can give various myopathies including tibial muscular dystrophy (Hackman et al., 2002), limb-girdle muscular dystrophy type 2J (Udd, 1992), congenital myopathy with dilated cardiomyopathy (Carmignac et al., 2007, Hackman et al., 2008), and variable myopathy with early respiratory failure (Pfeffer et al., 2012). None of the dominant mutations causing cardiomyopathy have been identified as a cause of myopathy and vice versa.

5.3.2 Telethonin

Cardiomyopathies

Heterozygous missense variants in telethonin have also been reported to cause different types of cardiomyopathy, but the supporting evidence for pathogenicity is generally poor. The variants reported as being associated with HCM are p.E13del, p.R70W, p.P90L[1], p.R106C[2], p.T137I, and p.R153H (Hayashi et al., 2004, Bos et al., 2006, Andersen et al., 2009), and those associated with DCM are p.R18Q, p.E49K, and p.E132Q (Hayashi et al., 2004, Hershberger et al., 2008). Evidence for these mutations is variable, for example p.R153H is found in two of thirteen-thousand samples in the NHLBI GO exome variant database (EVS, March 2013) and, as discussed in section 5.1.3.3 (page 137), p.E13del has

[1] Found in conjunction with a variant in *MYBPC3* (p.Q998R)

[2] Found in conjunction with a variant in *TNNI3* (p.R145W)

been identified in control groups as frequently as cases in other studies (Andersen et al., 2009, Perrot et al., 2006, Marziliano et al., 2006). It has been suggested that HCM-causing variants act by increasing binding with titin and other Z-disk components, whereas DCM-causing variants decrease binding with these (Hayashi et al., 2004). Interpretation of the significance of affecting these interactions is difficult however, particularly in those variants also reported in the normal population.

Myopathies

Homozygous telethonin nonsense mutations have been reported causing limb-girdle muscular dystrophy type 2G (LGMD2G) (Moreira et al., 2000, Olive et al., 2008) and congenital muscular dystrophy (Ferreiro et al., 2011). LGMD2G is relatively mild compared with other recessive forms of this disorder, but cardiac involvement occurs in a significant proportion (Moreira et al., 2000, Vainzof et al., 2002). Muscle immunofluorescence in these patients has shown an absence of sarcomeric telethonin, but normal staining of other muscle proteins including titin, and normal sarcomeric architecture on ultrastructural analysis (Vainzof et al., 2002).

5.4 Hypotheses

Given the identification of mutations in titin, telethonin, and other Z-disk proteins as causes of various types of cardiomyopathy and myopathy, it seems reasonable to hypothesise that the p.A178D variant in the Z2 domain of titin is the cause of LVNC in this family. However, the existing literature on many of the identified missense variants in sarcomeric genes, and in particular titin, demonstrates how complex assigning pathogenicity to these can be. As has been discussed the Z1Z2 region of titin forms part of the complex protein network at the Z-disk. The findings regarding the functions of this region of titin, and its protein interactions, are conflicting. Determining the effects A178D has on the structure and function of Z1Z2 is therefore of clinical and academic interest.

Demonstrating that A178D is disruptive to the known functions of the titin Z2 domain would provide further evidence of pathogenicity. From what is known of this region of titin there are a few mechanisms by which this would seem most likely to occur. These form the basis of the initial hypotheses and lines of investigation.

Hypothesis 1: A178D disrupts the stability and conformation of the Z1Z2 region of titin

Ig-like domains, such as Z2, are composed of folded β-sheets. If A178D causes unfolding of this domain this could impair the stability of this region of the protein.

Hypothesis 2: A178D causes mislocalisation of the titin N-terminus and ligands

The mechanism by which the N-terminus of titin localises to its position adjacent to the Z-disk is unknown. A178D may affect the mechanism by which this occurs, therefore preventing appropriate localisation. The localisation of ligands, such as telethonin, may also be secondarily affected by this.

Hypothesis 3: A178D impairs the functions of the titin Z1Z2 region

A structural change in the Z2 domain, as described in the first hypothesis, could significantly affect the normal function of this region of titin. The main interaction of Z1Z2 is with telethonin. As has been discussed this interaction has been postulated as being of importance in a number of functions. Structural changes may act to disturb the binding interface between these ligands exerting pathological effects through this mechanism. The proposed interaction with sAnk1.5 may also be of importance, although the evidence for this is limited.

To prove or refute these hypotheses a variety of biochemical and biophysical studies of the effects of A178D have been performed.

5.5 Methods

5.5.1 Expression Constructs

Plasmid vectors are used to enable protein expression of the incorporated DNA sequence in a host cell (either bacterial or mammalian cell cultures). They may also incorporate other sequences useful for this process, such as inducible promoters, protein tags, and antibiotic resistance genes. A full list of titin, telethonin and ankyrin plasmid constructs are summarised in Table 5-1. Primers used are summarised in Table 5-2. Equipment, manufactured kit reagents and buffers are listed in the appendix section II)

The following methods were used in the production of expression constructs, and are referred to in the descriptions of the individual constructs:

- PCR using Phusion® high fidelity polymerase (NEB®). The manufacturer's protocol (http://tinyurl.com/lso5vvk) was used in all cases. Primers and template used are given in the description of each construct and summarised in Table 5-2. The experimentally determined optimum PCR conditions, in terms of kit buffer used (HF or GC) and annealing temperature are given for each construct.
- Analysis and purification of DNA products was achieved on agarose mini-gels. Gels constituted of 1.5% agarose in Tris/Borate/EDTA (TBE) were cast in a standard kit (BioRad®). 6x DNA loading dye (Fermentas ThermoScientific®) was added to the PCR product solution in a 1:6 ratio before loading in to the wells of the gel. Samples were run alongside 2 μl of GeneRuler 1kb DNA ladder (Fermentas ThermoScientific®). Electrophoresis used a BioRad® PowerPac and mini-agarose gel system run at a constant 80 mV. These were imaged using a ChemiDoc™ XRS imaging system (BioRad®).
- DNA was extracted from agarose gels using the QIAquick gel extraction kit (Qiagen®) using the manufacturer's protocol (pages 18-19 http://tinyurl.com/ltsqshb).

- Plasmids and DNA inserts were individually cleaved at appropriate restriction sites (detailed in the description of each construct and summarised in Table 5-1) using restriction enzymes from New England Biolabs (NEB®) on a heat block at 37°C for 5 hours. The complementary insert and plasmid were ligated using T4 DNA ligase (NEB®) at 16°C for 16 hours.
- The full protocol for transformation of the constructs into XL10 Gold ultracompetent cells (Stratagene/Agilent Technologies®) is given in appendix section II-a (page 244).
- Plasmid DNA was purified from bacterial transformations using the Qiaprep spin miniprep kit (Qiagen®) according to the manufacturer's protocol (http://tinyurl.com/kwc76b8), and from larger bacterial cultures using the Qiaprep spin midiprep kit (Qiagen®) according to the manufacturer's protocol (http://tinyurl.com/kwc76b8).
- Sequencing was performed by Source Bioscience®, the primer sequence used for each construct is given in Table 5-1, and analysed using BioEdit® software.
- The A178D substitution was introduced using the Quickchange II XL site-directed mutagenesis kit (Agilent Technologies®) according to the manufacturer's protocol (http://tinyurl.com/lpuv5oo). The primers used were designed using the manufacturer's primer design program (http://www.genomics.agilent.com/primerDesignProgram.jsp). Forward primer 5'-GGACCTATTCAGTAAATGACACCAATAGCGTTGGAAG and reverse primer 5'-CTTCCAACGCTATTGGTGTCATTTACTGAATAGGTCC.

Protein Product	Plasmid Vector	Insert	Restriction Sites	Protein Tag (C or N terminal)	Sequencing Primer	Expression	Resistance
Z1Z2 WT-GFP	p.EGFP.N1	Titin αα1-196 WT	XhoI/BamHI	GFP (C)	CMVf	Mammalian	Kanamycin
Z1Z2 A178D-GFP	p.EGFP.N1	Titin αα1-196 A178D	XhoI/BamHI	GFP (C)	CMVf	Mammalian	Kanamycin
Z1Z2 WT-6H	p.ET-6H	Titin αα1-196 WT	XhoI/MluI	6H/EEF (C)	T7	Bacterial	Ampicillin
Z1Z2 A178D-6H	p.ET-6H	Titin αα1-196 A178D	XhoI/MluI	6H/EEF (C)	T7	Bacterial	Ampicillin
Z2 WT-6H	p.ET-6H	Titin αα98-195 WT	XhoI/MluI	6H/EEF (C)	T7	Bacterial	Ampicillin
Z2 A178D-6H	p.ET-6H	Titin αα98-195 A178D	XhoI/MluI	6H/EEF (C)	T7	Bacterial	Ampicillin
Telethonin 1-167-GST	pEBG	Telethonin αα1-167	BamHI/NotI	GST (N)	pGEX5	Mammalian	Ampicillin
Telethonin 1-90-GST	pEBG	Telethonin αα1-90	BamHI/NotI	GST (N)	pGEX5	Mammalian	Ampicillin
sAnk1.5-GST	pEBG	sAnk1.5 αα29-155	BamHI/NotI	GST (N)	pGEX5	Mammalian	Ampicillin

Table 5-1: Plasmid constructs with product abbreviation and key features

Purpose	Strand	Primer
Titin αα1-196 cloning	Forward	5′-TTTctcgagATGACAACTCAAGCACCGACG
	Reverse	5′-TTTggatccCGACCTTGAACCAGTAATTCAGCAG
Titin αα98-195 cloning	Forward	5′-TTTctcgagCAAAGCTGAGACAGCACCAC
	Reverse	5′-TTTacgcgtTTCACCTTGAACCAGTAATTCAGC
Titin A178D mutagenesis	Forward	5′-GGACCTATTCAGTAAATGACACCAATAGCGTTGGAAG
	Reverse	5′- CTTCCAACGCTATTGGTGTCATTTACTGAATAGGTCC
Telethonin αα1-90 cloning	Forward	5′-TTTggatccATGGCTACCTCAGAGCTGAGC
	Reverse	5′-TTTgcggccgcTTACGGCAGTACCCGCTGGTAG
Telethonin αα1-167 cloning	Forward	5′-TTTggatccATGGCTACCTCAGAGCTGAGC
	Reverse	5′-TTTgcggccgcTCAGCCTCTCTGTGCTTCCTG
sAnk1.5 αα29-155 cloning	Forward	5′-TTTggatccGTCAGGGGGTCCCTGTGCTTTG
	Reverse	5′-TTTgcggccgcTCACTGTTTCCCCCTTTTCAGGC

Table 5-2: Primers used in the production of expression constructs

5.5.1.1 Titin αα1-196 WT pEGFP.N1

The titin Z1Z2 region sequence (encoding amino acids 1 to 196) was cloned and amplified from a cardiac muscle cDNA library with the Phusion high fidelity polymerase using the manufacturer's protocol, optimised to an annealing temperature of 70°C and using kit GC buffer. The forward primer used was 5′-TTTctcgagATGACAACTCAAGCACCGACG, incorporating the XhoI restriction site (CTCGAG), and reverse primer 5′-TTTggatccCGACCTTGAACCAGTAATTCAGCAG, incorporating the BamHI restriction site GGATCC. PCR products were gel purified on a 1.5% agarose-TBE mini-gel, and DNA was extracted using the QIAquick gel extraction kit following the manufacturer's protocol. The product was subcloned into the pEGFP.N1 vector (Clontech®; plasmid map in Figure 5-9), by digestion using XhoI and BamHI restriction enzymes and ligation using T4 DNA ligase as described.

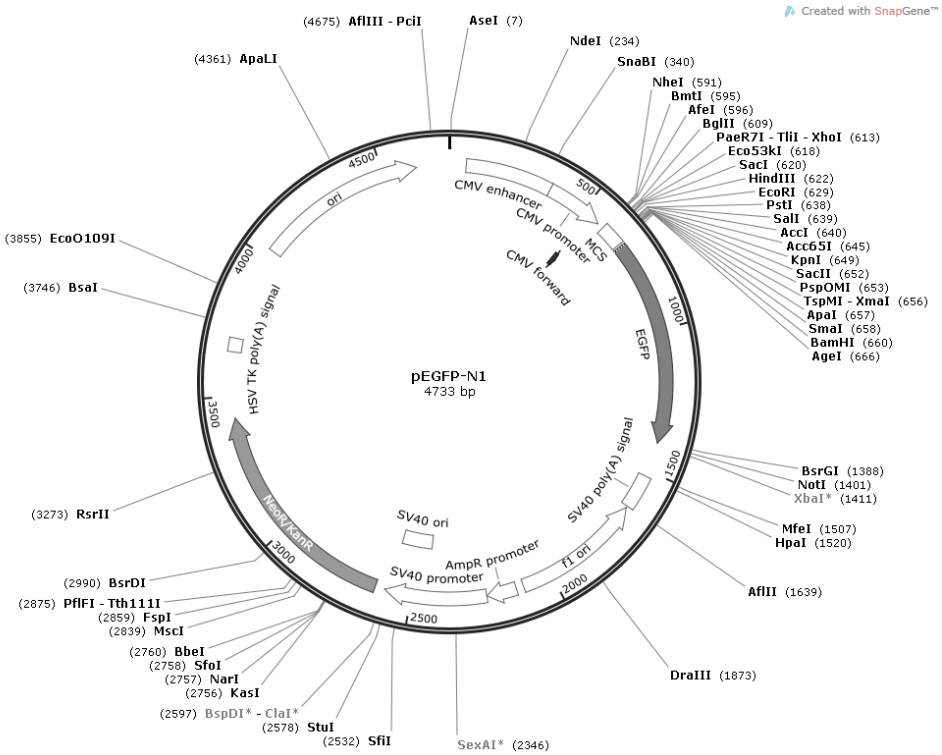


Figure 5-9: pEGFP.N1 plasmid map

The plasmid construct was transformed into XL10 Gold ultracompetent cells (Stratagene/Agilent Technologies®) and colonies grown on kanamycin-agar plates incubated at 37°C overnight. Individual colonies were grown overnight in Luria broth (LB)-kanamycin overnight in an incubator-shaker (New Brunswick Scientific®) at 37°C/250 rpm. Plasmid DNA was purified using the Qiaprep spin miniprep kit according to the manufacturer's protocol. Eluted DNA was digested with XhoI and BamHI and run on a minigel as before to determine which colonies had successfully incorporated the plasmid construct. These were then sequenced to confirm identity, and successful colonies were amplified in larger LB-kanamycin cultures overnight at 37°C/250 rpm. Plasmid DNA was purified with the Qiaprep spin midiprep kit according to the manufacturer's protocol.

5.5.1.2 Titin αα1-196 A178D pEGFP.N1

The c.533G>T/p.A178D substitution was introduced to the titin αα1-196 WT pEGFP.N1 construct using the Quickchange II XL site-directed mutagenesis kit as described above.

5.5.1.3 Titin αα1-196 WT pET-6H

Titin Z1Z2 sequence subcloned into the pET-6H vector was kindly provided by Professor Mathias Gautel, King's College, London, after failed attempts to subclone this sequence into an alternative pET vector. This vector incorporates a T7 promoter and hexa-histidine (6H) and EEF tags. The plasmid construct was transformed, cultured and purified as described previously.

5.5.1.4 Titin αα1-196 A178D pET-6H

The c.533G>T/p.A178D substitution was introduced into the titin αα1-196 WT pET-6H construct using the Quickchange II XL site-directed mutagenesis kit as before.

5.5.1.5 Titin αα98-195 WT pET-6H

A single, Z2, domain coding sequence construct was also produced and incorporated into the pET-6H vector. The single-domain sequence was cloned from a 1:250 dilution of the titin αα1-196 WT pEGFP construct using the Phusion high fidelity polymerase. Phusion PCR was optimised to use an annealing temperature of 60^{o}C and GC buffer. The forward primer used was 5'-TTTctcgagCAAAGCTGAGACAGCACCAC, incorporating the XhoI restriction site CTCGAG, and the reverse primer was 5'-TTTacgcgtTTCACCTTGAACCAGTAATTCAGC, incorporating the MluI restriction site ACGCGT. The titin insert from the titin Z1Z2 pET-6H construct was excised by digestion with the XhoI and MluI restriction enzymes and the vector was gel purified. The Z2 sequence was subcloned in to the same restriction sites, and the plasmid construct was then transformed, cultured, and purified as before.

5.5.1.6 Titin αα98-195 A178D pET-6H

The c.533G>T/p.A178D substitution was introduced to the titin αα98-195 WT pET-6H construct using the Quickchange II XL site-directed mutagenesis kit.

5.5.1.7 Telethonin αα1-90 pEBG

The sequence for the region of telethonin required for titin Z1Z2 binding (amino acids 1-90) was cloned from a cardiac muscle cDNA library with Phusion polymerase, optimised to use an annealing temperature of 60^{o}C and kit HF buffer. The forward primer used was 5'-TTTggatccATGGCTACCTCAGAGCTGAGC, incorporating the BamHI restriction site, and the reverse primer was 5'-TTTgcggccgcTTACGGCAGTACCCGCTGGTAG, incorporating the NotI restriction site. Gel purified PCR products of the appropriate molecular weight were then subcloned into the pEBG vector (plasmid map in Figure 5-10) at the BamHI and NotI restriction sites. The pEBG plasmid incorporates a GST tag and lac promoter. The plasmid construct was transformed into XL10 Gold ultracompetent cells, colonies grown on ampicillin-agar plates and then in Luria broth (LB) with ampicillin. The telethonin insert was sequenced to confirm identity using the pGEX5' primer. Plasmid DNA was purified using the Qiaprep spin miniprep kit and colonies with the appropriate constructs were amplified in larger LB-ampicillin cultures and purified with the Qiaprep spin midiprep kit.

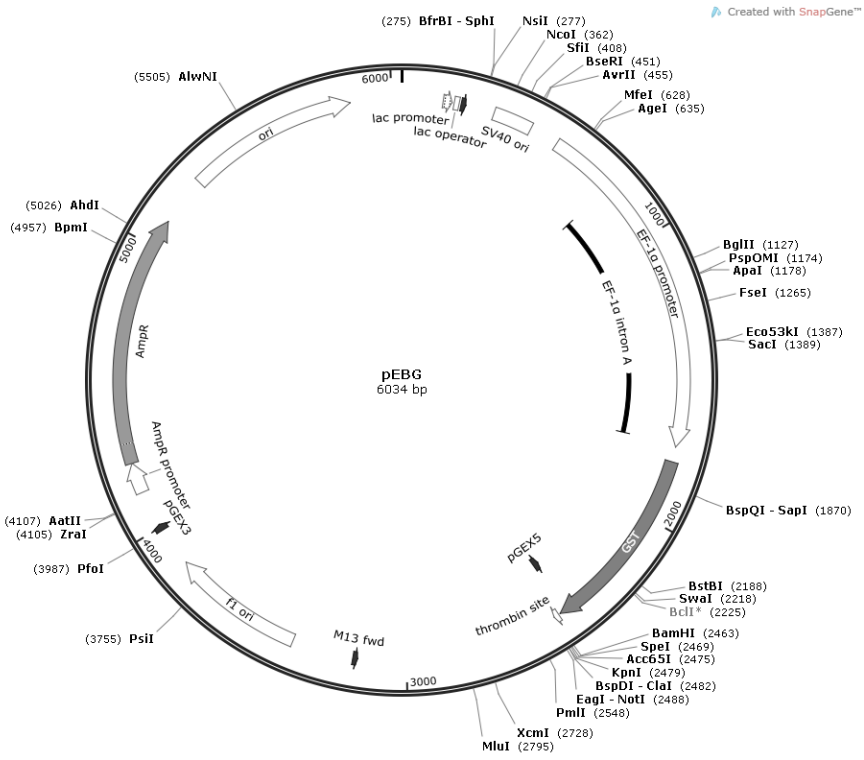


Figure 5-10: pEBG plasmid map

5.5.1.8 Telethonin αα1-167 pEBG

The full telethonin sequence (amino acids 1-167) was cloned from a cardiac muscle cDNA library with Phusion polymerase, optimised to use an annealing temperature of 60°C and kit GC buffer. The forward primer used was 5'- TTTggatccATGGCTACCTCAGAGCTGAGC, incorporating the BamHI restriction site, and the reverse primer was 5'-TTTgcggccgcTCAGCCTCTCTGTGCTTCCTG, incorporating the NotI restriction site. Gel purified PCR products of the appropriate molecular weight were subcloned into pEBG, transformed into XL10 Gold, culture amplified and purified as in section 5.5.1.7.

5.5.1.9 sAnk1.5 29-155 pEBG

The cytoplasmic sAnk1.5 sequence (amino acids 1-167) was cloned from a skeletal muscle cDNA library with Phusion polymerase, optimised to use an annealing temperature of 60°C and kit GC buffer. The forward primer used was 5'-TTTggatccGTCAGGGGGTCCCTGTGCTTTG, incorporating the BamHI restriction site, and the reverse primer was 5'-TTTgcggccgcTCACTGTTTCCCCCTTTTCAGGC, incorporating the NotI restriction site. Gel purified PCR products of the appropriate molecular weight were subcloned into pEBG, transformed into XL10 Gold, culture amplified and purified as in section 5.5.1.7.

5.5.2 Protein Expression and Purification

Proteins were expressed using the constructs above in bacterial and mammalian cell cultures. Bacterially expressed proteins were purified by use of their fusion protein tags, GST and hexa-histidine. Full protocols for the expression of plasmid constructs in *Escherichia coli* BL21 competent cells, test (mini) and larger-scale (maxi) protein expressions, Ni-NTA purification, and gel filtration chromatography, are described in appendix sections II-a to e, page 248.

5.5.2.1 Protein Expression from pET-6H Constructs in *Escherichia coli*

pET-6H plasmid constructs were transformed into *Escherichia coli* (*E. coli*) BL21 competent cells. They were cultured in 100 µl SOC medium (Sigma-Aldrich®) for 1 hour at 37°C/250 rpm. Cultures were then grown overnight on chloramphenicol and ampicillin-agar plates at 37°C. Individual colonies were selected and cultured in 3 ml of LB medium (supplemented with chloramphenicol and ampicillin) overnight at 37°C/250 rpm. Test protein expressions were performed from these cultures using Isopropyl β-D-1-thiogalactopyranoside (IPTG; Fermentas®) to induce expression. IPTG is a mimic of a lactose metabolite which releases the repressor from the *lac*/*tac* operator to trigger expression of the desired protein. Expressions were then analysed on a Coomassie stained polyacrylamide gel to ensure appropriate protein expression (SDS-PAGE protocol is described in appendix II-g, page 249).

Large-scale protein expressions were then performed from the same cultures. Larger cultures in SOC medium (SOB medium made to manufacturer's instructions; with 16 ml 1M glucose, 4 ml 1M $MgSO_4$, and 4 ml 1M $MgCl_2$ added per 800 ml medium) were incubated at 37°C/250 rpm until an optical density at a wavelength of 600nm of 0.7 was reached. IPTG was added to a final concentration of 0.5 mM to induce expression. This optical density represents an approximation for cells being at a period of exponential growth and therefore most effective for protein expression. These cultures were then incubated at 18°C/250 rpm for four hours and centrifuged in aliquots at 5000 rpm for twelve minutes to form cell pellets.

A 500 µl aliquot of the expression culture was analysed on a 13% polyacrylamide gel (using BioRad casting apparatus and electrophoresis chambers, 80 mV for stacking and 120 mV for resolving) and Coomassie stained to ensure appropriate protein expression (SDS-PAGE full protocol is provided in appendix II-g, page 249). In A178D cultures the presence of the amino acid substitution was confirmed by mass spectrometry performed by Dr Katalin Di

Gleria, Weatherall Institute of Molecular Medicine, University of Oxford using the matrix-assisted laser desorption/ionisation (MALDI) technique. The proteins were then purified by Ni-NTA purification and gel filtration chromatography (section 5.5.2.3 and 5.5.3.2; pages 163 and 165).

5.5.2.2 Protein Expression from pEGFP and pEBG Constructs in COS1 Cell Cultures

COS1 cells are an immortalised African Green Monkey kidney fibroblast-like cell line (ATCC® CRL-1650™). These were used for protein expression from pEGFP.N1 and pEBG constructs. Cell cultures were grown in Cellstar® filter-cap tissue culture flasks (Greiner Bio-one®) at 37°C/5% CO_2 in high glucose Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich®) containing 10% heat-inactivated fetal bovine serum (FBS; Gibco®), 1% L-glutamine (Sigma-Aldrich®), and 1% penicillin and streptomycin (Sigma-Aldrich®).

For GST pull-down assays (section 5.6.3.1, page 188) cells were seeded at 1:3.5 density from confluent cultures into 9 cm diameter dishes (Falcon®) and incubated for 24 hours at 37°C/5% CO_2. The cells were then transfected using 5 µg of each appropriate construct DNA and 17.5 µl Escort IV transfection reagent (Sigma-Aldrich®) in 500 µl Opti-MEM® reduced-serum medium (Gibco/Life Technologies®) and incubated for 48 hours. Cultures were then washed with PBS, harvested into PBS with protease inhibitors (Complete Protease Inhibitor Cocktail Tablets, Roche®), and centrifuged at 3000 rpm for five minutes to produce a cell pellet.

5.5.2.3 Protein Purification of Bacterially Expressed Proteins

Hexa-histidine fusion proteins

His-tagged fusion proteins, i.e. those subcloned into the pET-6H vector, expressed in *E. coli* BL21 cells were purified by Nickel-NTA purification (full protocol in appendix section II-d, page 247). This is possible due to the high affinity of the poly-histidine tag with nickel ions,

which are bound to agarose beads allowing the poly-histidine-tagged fusion proteins to be isolated from solution.

Cell pellets were resuspended in Nickel-NTA buffer with 10 mM imidazole and protease inhibitors. Imidazole reduces non-specific binding and therefore improves the purity of the resulting protein. The cells were lysed using lysozyme (1 mg/ml), deoxyribonuclease (DNase; 5 µg/ml), and sonicated. After centrifugation the supernatant was incubated with Ni-NTA agarose beads (Qiagen®). These were then repeatedly washed with Ni-NTA buffer with 20 mM imidazole to remove non-bound proteins. Histidine-tagged fusion proteins were eluted from the beads in aliquots using Ni-NTA buffer with 250 mM imidazole. The elution fractions were spotted onto nitrocellulose paper and stained with Ponceau. Those fractions containing a significant amount of protein, as judged by the degree of Ponceau staining, were merged and kept on ice. Purity of the protein preparation was verified on SDS-PAGE. Further purification, as well as analysis, was performed using gel filtration chromatography (section 5.5.3.2, page 165). Protein concentrations were calculated from nanodrop measurements of absorbance using the Beer-Lambert law:

$$\text{Concentration (mg/ml)} = \frac{\text{Absorbance at 280nm}}{\text{Extinction Coefficient}} \text{ x Molecular Weight}$$

Molecular weights and extinction coefficients of the relevant proteins are shown in Table 5-3.

Protein	Molecular Weight (kDa)	Extinction Coefficient
TTN Z1Z2	20.9	12,950
TTN Z2	9.7	5960

Table 5-3: The molecular weights and extinction coefficients of titin Z1Z2 and Z2. These were calculated from their amino acid sequence using the ProtParam tool on the ExPASy Bioinformatics Resource Portal (http://web.expasy.org/protparam/) (Gasteiger et al., 2005)

5.5.3 Structural Studies

5.5.3.1 *In Silico* Analysis

The predicted effects of the mutation can be better determined by analysing it further using structuring modelling tools. Pymol (PyMol, June 2013) and Wincoot (Emsley et al., 2010) have been used to image the titin-telethonin interaction and possible effects of the alanine to aspartate substitution at position 178. Protein data bank (pdb) files ascertained from crystallography studies of the titin Z1Z2 region alone (PDB2A38) (Marino et al., 2006) and in combination with telethonin (PDB1YA5) (Zou et al., 2006, Pinotsis et al., 2006) were analysed.

5.5.3.2 Gel Filtration Chromatography

Gel filtration chromatography separates proteins on the basis of their physical properties when run in a buffer over a resin matrix in a column. This was used both for further purification of titin hexa-histidine fusion proteins in PBS buffer, and for their analysis (full protocol in appendix section II-d, page 247). An Amersham Biosciences® Hiload 16/60 superdex 75 prep grade column was used on an Äkta FPLC system for this. The concentration of these proteins was then re-established as previously described, and concentration was achieved using Vivaspin® (MWCO 5000) centrifugal concentrators when required using manufacturer's guidelines.

5.5.3.3 Polyacrylamide Gel Electrophoresis and Western Blotting

PAGE

Polyacrylamide gel electrophoresis (PAGE) separates proteins on the basis of their physical properties. Samples were either run in their native state (Native-PAGE), retaining their higher order structure, or after the addition of the anionic detergent sodium dodecyl sulphate (SDS-PAGE), which disrupts most non-covalent bonds and confers a negative charge on the proteins. Full procedures for Native-PAGE and SDS-PAGE are given in

appendix sections II-f and g respectively (page 249). Protein bands can then either be directly imaged by Coomassie stain or transferred to nitrocellulose for antibody staining and visualised by Western blotting.

Western blotting

Blots were transferred using standard wet transfer methods using the BioRad system (at 40 mA overnight), or dry transfer using the iBlot™ system (Invitrogen®) following the manufacturer's guidelines (http://tinyurl.com/nztdx2r). The blots were stained with Ponceau and then blocked with 5% milk in TBST (20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween 20; pH 7.5).

Primary antibodies diluted in 5% milk/TBST were applied overnight at 4°C. Antibodies used are shown in Table 5-4. Blots were then washed thoroughly with TBST and secondary antibodies, diluted in 1% milk/TBST, were applied for one hour at room temperature. Horseradish peroxidase secondary antibodies were used (Table 5-4), to allow effective imaging of protein bands using chemiluminescence after application of SuperSignal West Pico Chemiluminescent Substrate (ThermoScientific®). Bands were imaged using light-sensitive film (GE Healthcare®) and developed and fixed using an imager (Konica Minolta®).

Protein	Primary Antibody	Dilution	Secondary Antibody	Dilution	ECL Substrate
GFP	αGFP monoclonal mouse (Roche® 12600500)	1:400	anti-mouse HRP (GE Healthcare® NA931)	1:2000	Pico
GST	αGST polyclonal goat (GE Healthcare® 27-4577-01)	1:20 000	anti-goat HRP (ThermoScientific® 31402)	1:10 000	Pico

Table 5-4: Primary and secondary antibodies for western-blotting

5.5.3.4 Thermolysin Proteolysis Assays

1 µl of thermolysin (x100; Sigma-Aldrich®) was added to 50 µl aliquots of titin Z1Z2 domain protein fractions post purification (~0.1 mg/ml in PBS), and incubated on a heat block at 37°C. 12.5 µl SDS-SB was added to stop reaction after a set time period (0, 5, 10, 20, 60, 120

or 240 minutes, and 24 hours for each protein fraction). These were than visualised by separation by SDS-PAGE and Coomassie staining.

5.5.3.5 Thermal Denaturation Assays

The Protein Thermal Shift™ Dye Kit (Applied Biosystems®) was used using the standard protein melt protocol (5.0 µl of kit buffer, 2.5 µl SyproOrange dye at 8x concentration; 12.5 µl protein in buffer) in MicroAmp® 96-well plates. Total protein per well of between 0.8 and 4 µg were used. The fluorescence is measured using a real time PCR machine (StepOne™ system, Applied Biosystems®), set to a 0.5°C rise every 2 minutes from 25°C to 99°C and measurement at each temperature increment. Analysis was performed using the manufacturer's software and guidelines. PBS buffer was used as standard but stability optimisation was undertaken using the PACT suite of buffers with varying anions/cations and pH (full details at http://tinyurl.com/p9une56 and in the appendix section II-h, page 250) (Newman et al., 2005).

5.5.3.6 Circular Dichroism Spectroscopy

Circular dichroism spectroscopy is a method for assessing the secondary structure and folding of proteins based on the unequal absorption of left and right-handed circularly polarised light waves. Different structural elements of a protein have characteristic CD spectra which can be used to infer the relative composition of a protein (Greenfield, 2006). This technique was therefore used to help assess the conformational differences between wild-type and mutant titin proteins.

Titin-6H fusion proteins were dialysed (with multiple buffer changes including overnight at 4°C) into 15 mM potassium phosphate buffer pH 7.4, to reduce the optical activity of this. Protein concentrations in this buffer were inferred from nanodrop measurements as before, and where required concentrated using Vivaspin centrifugation concentrators.

Samples were analysed in a cuvette with an optical path of one millimetre on a Chirascan™ CD spectrometer (AppliedPhotophysics®) according to manufacturer's guidelines and using manufacturer's software. Settings used were- wavelength 185-260 nm in 0.5 nm steps, measured for 1 second per point with blank potassium phosphate buffer auto-subtracted at room temperature. Protein concentration was optimised to 0.1 mg/ml. For Z1Z2 construct purified fractions CD spectra were also produced at increasing temperatures, between 30 and 90^{o}C, to assess changes during thermal denaturation.

To predict protein structure from CD spectra a variety of algorithms for deconvoluting these curves have been devised. These try to overcome the issue that the secondary structure motifs give different spectra depending on their combinations and context (Chen and Yang, 1971). The DichroWeb online CD spectra analysis tool used can run several of these available algorithms (Whitmore and Wallace, 2004).

5.5.4 Localisation Studies

5.5.4.1 Neonatal Rat Cardiomyocyte Cell Culture and Transfection

Cultures of neonatal rat cardiomyocytes (NRC) muscle rat were obtained from the ventricles of new-born Sprague-Dawley® rats using a neonatal heart dissociation kit according to manufacturer's guidelines (Miltenyi Biotec®). NRCs were maintained at 37°C/5% CO_2 in maintenance medium (156 ml DMEM, 38.8 ml Medium 199, 8 ml horse serum, 4 ml L-glutamine, 2 ml Pen/Strep). For immunofluorescence 350 000 cells were seeded into 35 mm tissue culture plates coated with collagen solution. After 1 day NRC cells were transfected with 3 µg plasmid construct DNA using 3.5 µl Escort III transfection reagent (Sigma-Aldrich®) in 100 µl Opti-MEM, vortexed and incubated at RT for >10mins before being added to 800 µl transfection medium (21 ml Medium 199, 73 ml DBSS-K medium, 4 ml horse serum, 2 ml L-glutamine). After 4 hours this was changed back to

maintenance medium. After 24-48 hours the cells were washed in PBS and then fixed in 4% PFA for 10 minutes at room temperature.

5.5.4.2 Antibody Staining for Immunofluorescence

Transfected NRC/C2C12 cells were permeabilised using 0.2% Triton (Sigma-Aldrich®) for 10 minutes. The plates were then washed with PBST (PBS with 0.1% Tween 20) and blocked with 10% normal goat serum for 1 hour. The primary antibodies were diluted into 1% bovine serum albumin (BSA) in PBS buffer and applied to the plates, then left in a humid environment at 4°C overnight. Primary antibodies used were: T12, a murine monoclonal antibody to the Z-disk/I-region border of titin, diluted 1:10 (Boehringer®) (Furst et al., 1988); 653, a polyclonal rabbit antibody to α-actinin, diluted 1:200; EA53, a murine monoclonal antibody to α-actinin, diluted 1:300 (Sigma-Aldrich®) (Fridlianskaia et al., 1989); and B4, a murine monoclonal antibody to myomesin, diluted 1:10 (Grove et al., 1984). T12 and B4 antibodies were kindly provided by Dr Elisabeth Ehler, Randall Division of Cell and Molecular Biophysics, King's College, London.

The plates were washed thoroughly with PBT (phosphate-buffered saline with 0.1% Tween 20) before secondary antibodies, diluted in 1% BSA, were applied and left in a humid environment for 1 hour. Secondary antibodies used were goat anti-mouse Cy3/red 1:500 and goat anti-rabbit Cy5/far-red 1:100 (both Jackson ImmunoResearch®). The plates were then washed thoroughly with PBT and mounted with DAPI containing mounting solution (Slow Fade Gold; Invitrogen®) under a cover slip on glass slides. They were then imaged using confocal microscopy (Leica® SP5 using a 63x magnification oil immersion lens).

5.5.5 Binding Studies

GST pull-downs

GST pull-downs utilise the high affinity of GST for glutathione. The telethonin-GST fusion proteins therefore bind strongly to glutathione beads, and in turn the titin-GFP protein will bind to the telethonin. Therefore an assessment of the presence of titin-GFP wild-type and mutant proteins will give an indication of any difference in binding of these with telethonin. The full protocol for these pull-downs is provided in appendix section II-i (page 251).

COS1 cells pellets co-expressing GFP and GST fusion proteins (from section 5.5.2.2, page 163) were analysed (the combinations of these used for study and as control experiments is shown in Table 5-5). Cell pellets resuspended in GST pull-down buffer (10 mM TRIS, 80 mM NaCl, 0.5% NP 40). Following further centrifugation aliquots for use as lysate/loading controls were removed. The rest of the supernatant was combined with glutathione beads in suspension and incubated and agitated on ice for one hour. The beads were then thoroughly washed with GST buffer to remove any unbound proteins. Proteins are then dissociated by being boiled and sonicated in SDS buffer. They are then run on an SDS-PAGE and Western blotting for GFP and GST tags performed as described in section 5.5.3.3 (page 165). Band density is assessed using Image J software (Schneider et al., 2012).

	GFP Fusion Protein	GST Fusion Protein
Experimental Study	Titin αα1-196 WT	Telethonin αα1-90
	Titin αα1-196 WT	Telethonin αα1-167
	Titin αα1-196 WT	sAnk1.5 αα29-155
	Titin αα1-196 A178D	Telethonin αα1-90
	Titin αα1-196 A178D	Telethonin αα1-167
	Titin αα1-196 A178D	sAnk1.5 αα29-155
Control	GFP Alone	Telethonin αα1-90
	GFP Alone	Telethonin αα1-167
	GFP Alone	sAnk1.5 αα29-155
	Titin αα1-196 WT	GST Alone
	Titin αα1-196 A178D	GST Alone

Table 5-5: Combinations of GFP and GST fusion proteins used in GST pull-downs to study the effects of A178D on the binding interactions of the titin Z1Z2 region with telethonin and small ankyrin 1.5.

5.6 Results

5.6.1 Structural Studies

5.6.1.1 *In Silico* Analysis

The protein analysis tool PyMol (PyMol, June 2013) was used to visualise the location and environment of residue 178 within the titin Z2 domain. Protein data bank (pdb) files ascertained from crystallography studies of the titin Z1Z2 region alone (PDB2A38) (Marino et al., 2006) and in combination with telethonin (PDB1YA5) (Zou et al., 2006, Pinotsis et al., 2006) were analysed.

Figure 5-11 shows the β-sheet conformation of the Z1 and Z2 domains of titin and of telethonin, and the anti-parallel 'sandwich' interaction between them. The alanine residue at position 178 in the Z2 domain is shown to be in the hydrophobic core, other amino acids contributing to this include Pro104, Phe106, Val125, Val127, Pro133, Val135, Val176 and Ala185. The substitution of aspartate in place of alanine introduces a larger polar amino acid into this hydrophobic core. Therefore, this is predicted to destabilise this region and affect the folding of the Z2 domain.

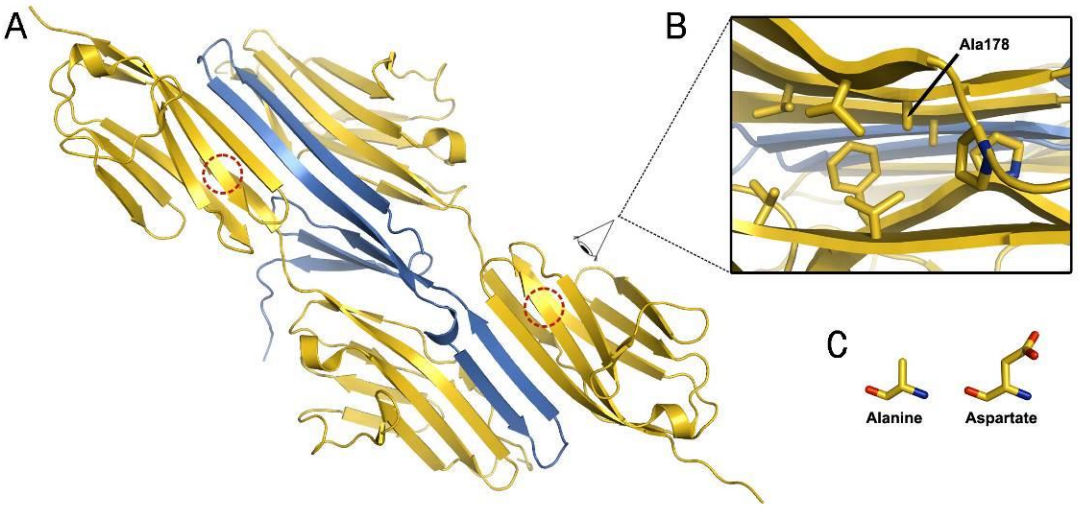


Figure 5-11: Diagram showing the location of the A178D titin variant:

A- Cartoon representation of the crystal structure of two titin Z1Z2 regions (gold) in complex with telethonin (blue). Dotted red circles indicate the location of residue 178. The eye icon indicates the view shown in panel B.
B- A view of the hydrophobic environment immediately surrounding the alanine at position 178 between β-sheets.
C- Stick view of alanine and aspartate to demonstrate the increased size of the side-chain and the introduction of two charged (oxygen) atoms.
Figure drawn using Pymol (www.pymol.org/).

5.6.1.2 Titin Fragment Ni-NTA Purification and Gel Filtration Chromatography

Titin fragments during purification

A two-step purification protocol was employed to purify recombinant titin Z1Z2 expressed in *E. coli* BL21 culture. The first step involves Ni-NTA purification, utilising the affinity of the hexa-histidine tag to nickel to purify the fusion protein. Subsequent further purification is performed by gel filtration chromatography to remove bacterial protein contaminants, based on their different elution profiles determined by molecular weight and conformation. It was observed that, despite expression levels being comparable, the same volume of *E. coli* BL21 culture consistently produced significantly more wild-type protein than mutant protein post Ni-NTA purification (Figure 5-12). This has many possible explanations, but the most likely would be reduced solubility of the mutant protein, a greater proportion of the mutant protein being contained in bacterial inclusion bodies, or the formation of protein aggregates. The differences in protein quantity can also be inferred from gel filtration chromatography elution peak areas (Figure 5-13).

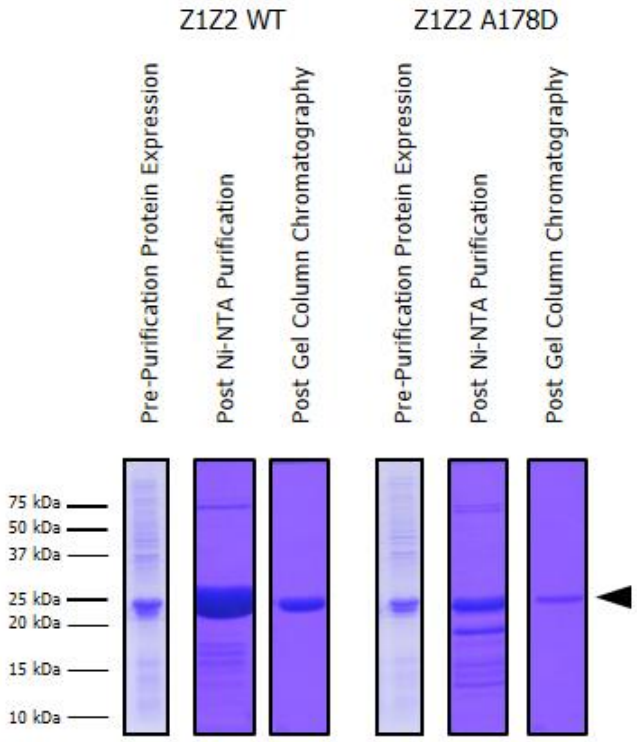


Figure 5-12: Titin Z1Z2 wild-type (WT) and mutant (A178D) protein fractions (band level marked) after initial expression, Ni-NTA and then gel filtration purification visualised by Coomassie stained SDS-PAGE. Expression of the WT and A178D proteins is comparable but significantly more WT is present after Ni-NTA purification than A178D. This is relatively maintained after gel filtration purification.

Gel filtration chromatography elution profiles

Gel filtration chromatography separates proteins on the basis of physical properties, with proteins with greater molecular weight, or more space-occupying conformations, eluting more rapidly. The elution profiles from gel column chromatography of Z1Z2 WT and A178D proteins show marked distinctions between the two. The wild-type protein elutes in a single fraction, whereas the mutant protein elutes in two distinct fractions. This implies the mutant protein is in two different conformations under these conditions, assuming the same size of protein (Figure 5-13). Both of these A178D fractions elute more quickly than the wild-type, with elution peaks at approximately 45 ml for A178D fraction 1, 60 ml for A178D fraction 2, and 70 ml for the WT fraction. This suggests an increase in the volume of space occupied by the same protein sequence (apart from the single amino acid substitution), most likely through partial or complete unfolding of the protein, or an increase in molecular weight of the complex, most likely by protein multimerisation.

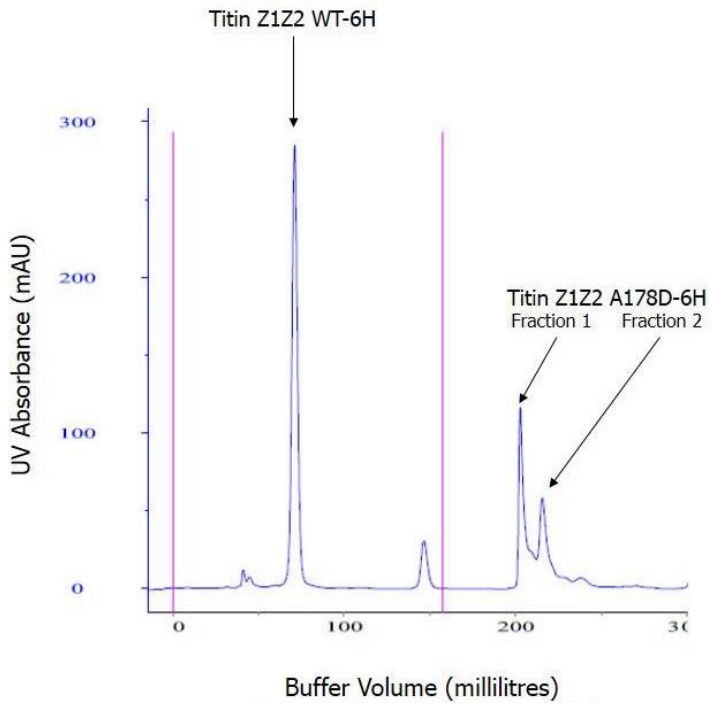


Figure 5-13: Elution curves from gel filtration chromatography of Z1Z2 WT and A178D protein fractions. The WT elutes in a single fraction, whereas A178D elutes in two distinct fractions. Both of the A178D fractions elute more rapidly than the WT (purple vertical lines indicating the sample injection times). As compared with the wild-type the mutant protein fractions are significantly smaller (total elution of protein following Ni-NTA purification of 100 ml of expression culture for the WT and 400 ml of expression culture for the A178D are shown). Run at 1 ml/min, with a total column volume of 120 ml.

Although WT and A178D fractions were usually purified independently, running a mixture of both wild-type and mutant proteins did not result in any changes to the elution profile (Figure 5-14). It can be inferred from this that there is an absence of significant interaction between the different protein fractions in solution.

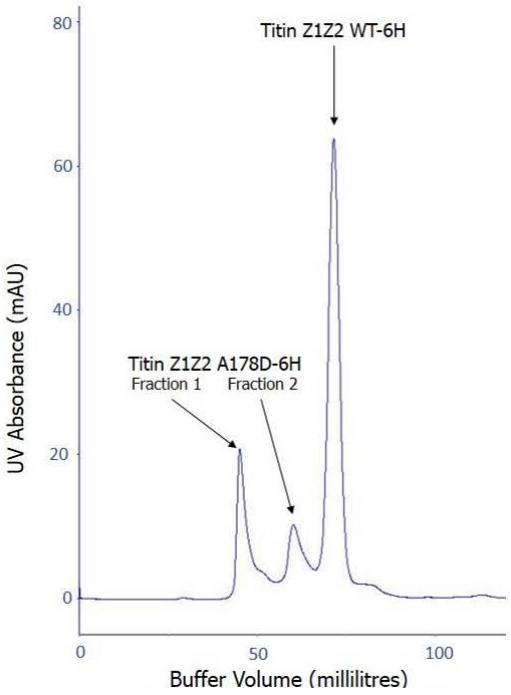


Figure 5-14: Elution curves from gel filtration chromatography of a mixture of 40 μg of Z1Z2 WT and 40 μg of Z1Z2 A178D post Ni-NTA purificatiion. A larger, single, peak is seen for the WT fraction. Two distinct, earlier eluting, smaller, peaks are seen for the A178D fractions. Run at 1 ml/min, with a total column volume of 120 ml.

Elution curves of the single-domain protein fragments, Z2 WT and A178D, gave the same pattern of elution profile, but at later time points (Figure 5-15). As compared with the two-domain protein fragments the elution peak of the WT fraction is significantly later, occurring at 85 ml as compared to 70 ml for Z1Z2. Although the mutant protein fractions also elute more slowly this is less marked, fraction one eluting at 48ml as compared with 45ml and fraction two eluting at 65 ml as compared with 60 ml. The general pattern of later elution of the single domain proteins is expected due to the lower molecular weight of these. The relative reduction in this difference seen with the mutant protein fractions, may suggest a greater proportion of unfolded protein as compared to the equivalent two-domain protein fraction.

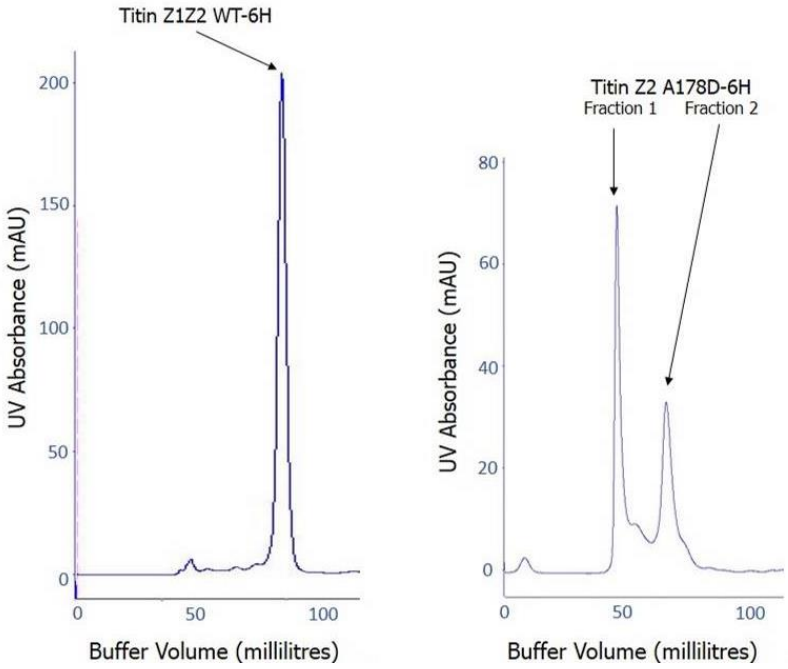


Figure 5-15: Elution curves from gel filtration chromatography of bacterially expressed Z2 WT (left) and A178D (right) post Ni-NTA purification. This shows the same pattern of elution curves as the Z1Z2 protein fractions, with a single WT fraction and two, earlier eluting, A178D fractions. Run at 1 ml/min, with a total column volume of 120 ml.

SDS-PAGE of purified Z1Z2 fractions

Purified protein fractions were studied under denaturing conditions in SDS-PAGE. This demonstrated that the protein fractions for each fragment type all have comparable apparent molecular weight (Figure 5-16). The observed molecular weight of the main Z1Z2 bands, of approximately 24 kDa, is in agreement with the calculated molecular weight of 22 kDa. The higher molecular weight band seen is a known bacterial contaminant. This has a histidine rich region, resulting in co-purification on the Ni-NTA matrix, and elutes at the same time as the first Z1Z2 A178D fraction on gel filtration chromatography. There is also a lower molecular weight band predominantly evident in the A178D fractions. Mass spectrometry has confirmed this is a Z1Z2 degradation fragment, missing the distal part of the Z2 domain, however the exact sequence could not be determined.

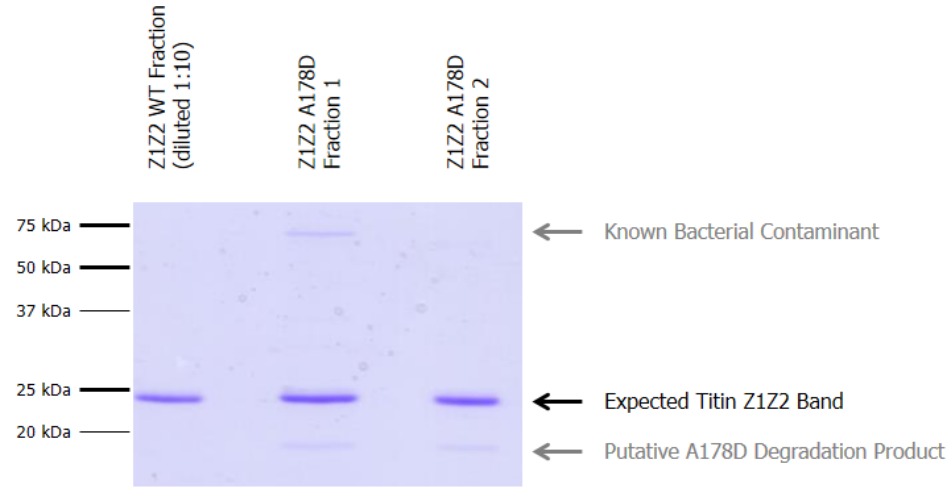


Figure 5-16: Coomassie stained 13% SDS PAGE of TTN WT and A178D fractions post gel filtration chromatography. All titin Z1Z2 fractions show a main band at approximately 24 kDa, which is consistent with the calculated mass of 22kDa. The higher molecular weight band see in the Z1Z2 A178D first fraction lane is a known bacterial contaminant with a histidine-rich region. The lower molecular weight band in both Z1Z2 A178D lanes is a putative Z1Z2 degradation product.

The observed molecular weight of the Z2 fractions is also broadly consistent with the calculated mass (approximately 11 kDa). However, there is a marked discrepancy in the electrophoretic mobility of the WT and A178D fractions, with the A178D fraction showing lower electrophoretic mobility. On reanalysis of the Z1Z2 fractions there is a suggestion that the A178D fractions may also demonstrate this, however this effect may be masked by the higher molecular weight (Figure 5-17). Mass spectrometry of the second Z2 A178D fraction gel band confirmed its identity as titin Z2, but could not determine the full sequence or total molecular weight.

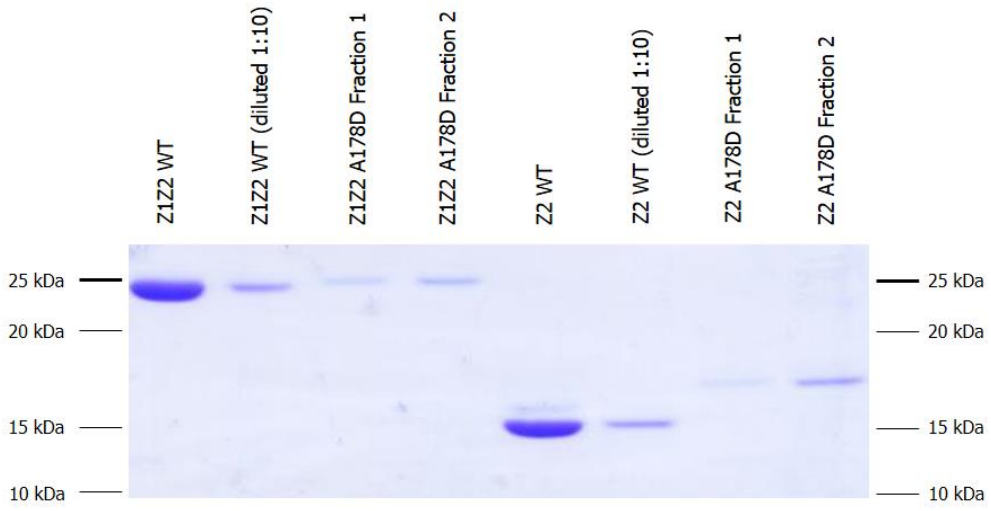


Figure 5-17: Coomassie stained 13% SDS-PAGE of titin Z1Z2 and Z2 fractions after gel filtration chromatography. Observed molecular weights are in agreement with the calculated molecular weights of 22 kDa for Z1Z2 proteins and 11 kDa for Z2 proteins. The A178D fractions have reduced electrophoretic mobility when compared with the equivalent WT fraction, particularly with the Z2 fractions.

Native-PAGE of purified Z1Z2 fractions

Native-PAGE was used to analyse the titin fractions under non-denaturing conditions. Z1Z2 WT and the second A178D fraction show similar electrophoretic mobility, implying they are of approximate size and conformation. The first fractions of both one- and two-domain A178D proteins were retained at the top of the gel. This suggests they were forming multimeric protein aggregates. The second Z2 A178D fraction has higher electrophoretic mobility than the wild-type, opposite to the mobility differences seen with SDS-PAGE.

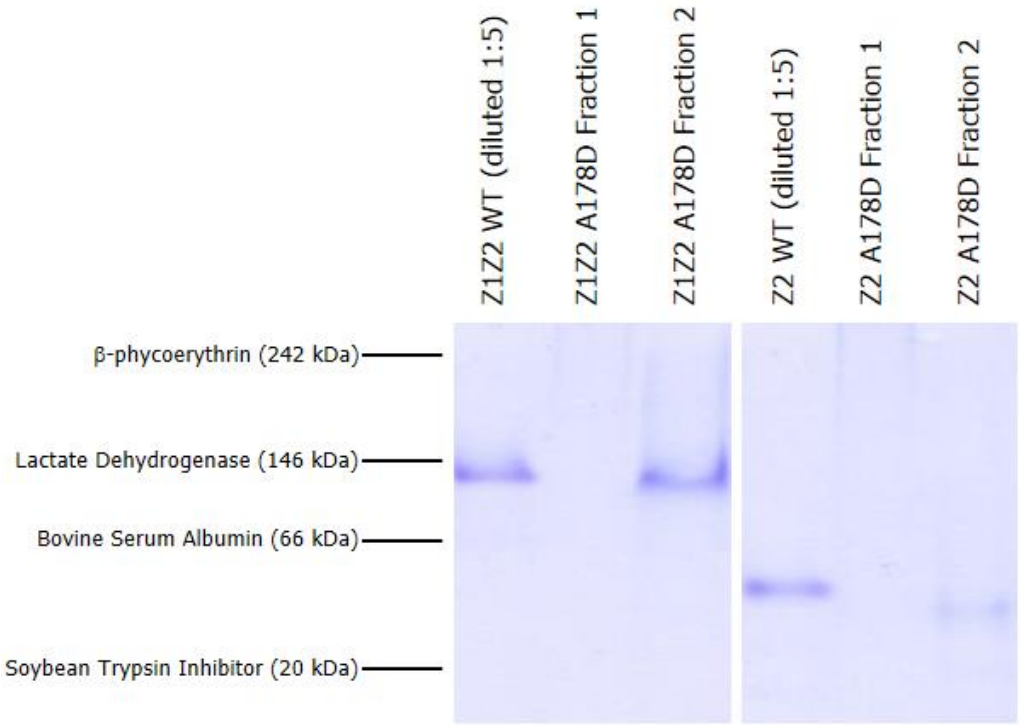


Figure 5-18: Titin Z1Z2 and Z2 WT and A178D post-purification protein fractions visualised using Coomassie staining on an 8% Native-PAGE. WT and A178D fraction 2 proteins have similar electrophoretic mobility, with a slight increase in mobility of the mutant protein fraction. The first fractions of A178D proteins are retained at the top of the gel. NB Although NativeMark™ marker bands are shown molecular weights are not directly comparable.

5.6.1.3 Thermolysin Proteolysis Assays

Thermolysin is a thermostable proteolytic enzyme that preferentially hydrolyses peptide bonds on the N-terminal side of hydrophobic amino acids. Native folded proteins are often resistant to proteolysis due to the lack of exposure of the amino acid sequences needed for protease action. It can therefore be used to identify conformational changes in proteins that expose hydrophobic residues. The purified Z1Z2 WT protein fraction remained intact, with no evidence of significant thermolysin-catalysed proteolysis up to forty-eight hours (time point not shown) (Figure 5-19). Purified Z1Z2 A178D protein fractions rapidly underwent hydrolysis to a stable protein fragment with an estimated molecular weight of approximately 13 kDa.

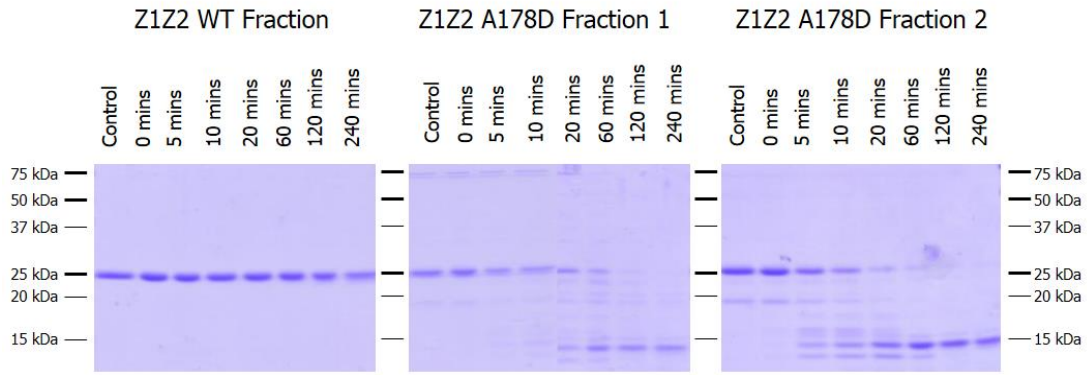


Figure 5-19: Visualisation of thermolysin proteolysis of titin Z1Z2 WT and A178D post-purification protein fractions at 37°C at different time points on a Coomassie stained 13% SDS-PAGE. The WT protein fraction remains intact, with no significant degradation observed. The A178D protein fractions are rapidly degraded to a stable 13 kDa product.

5.6.1.4 Thermal Denaturation Assays

To determine differences in stability between wild-type and mutant titin protein fractions thermal denaturation studies were performed. Proteins are denatured with increasing temperature depending on the strength and number of the interactions between amino acids maintaining their structure. The temperature at which a protein denatures can be determined by increasing the temperature in the presence of SYPRO® Orange fluorescent dye. The fluorescence of SYPRO Orange is quenched in an aqueous environment, but as a protein denatures and unfolds it exposes hydrophobic residues leading to an increase in fluorescence, which can then be quantified.

Melt curves for purified Z1Z2 and Z2 WT protein fractions demonstrate comparable thermal stability of both, with denaturation midpoint (Tm) values of ~63°C calculated for each (Fluorescence and derivative plots for the Z1Z2 WT fraction are shown in Figure 5-20; Fluorescence plots for Z1Z2 WT and Z2 WT are A and D respectively in Figure 5-21).

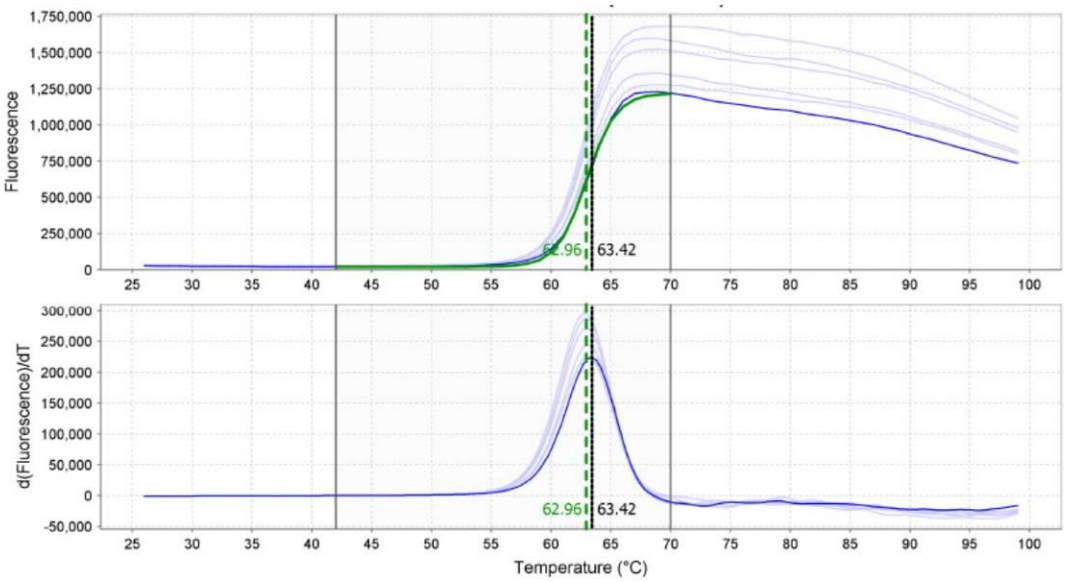


Figure 5-20: Fluorescence and derivative plots for the titin Z1Z2 WT post-purification fraction in PBS. These demonstrate a denaturation midpoint (Tm) level of approximately 63°C (green value is the Boltzmann Tm and black value is the derivative Tm) averaged from six samples.

In contrast the fluorescence plots for the A178D fractions demonstrate a high level of fluorescence at room temperature, implying exposed hydrophobic residues (B, C, E and F in Figure 5-21). Although it is not possible to obtain accurate measures of denaturation temperature for Z1Z2 A178D fractions, there is a suggestion of a fluorescence curve on top of the baseline level of fluorescence occurring at a lower temperature than the Z1Z2 or Z2 WT denaturation temperature. In the Z2 A178D fractions there is significant fluorescence curve from the baseline. Attempts to stabilise the Z1Z2 A178D fractions using the ninety-six buffer conditions of the PACT suite (see methods section 5.5.3.5, page 167, and appendix section II-h, page 250) did not reduce fluorescence at baseline significantly enough to allow more accurate quantification of the changes in fluorescence observed (data not shown).

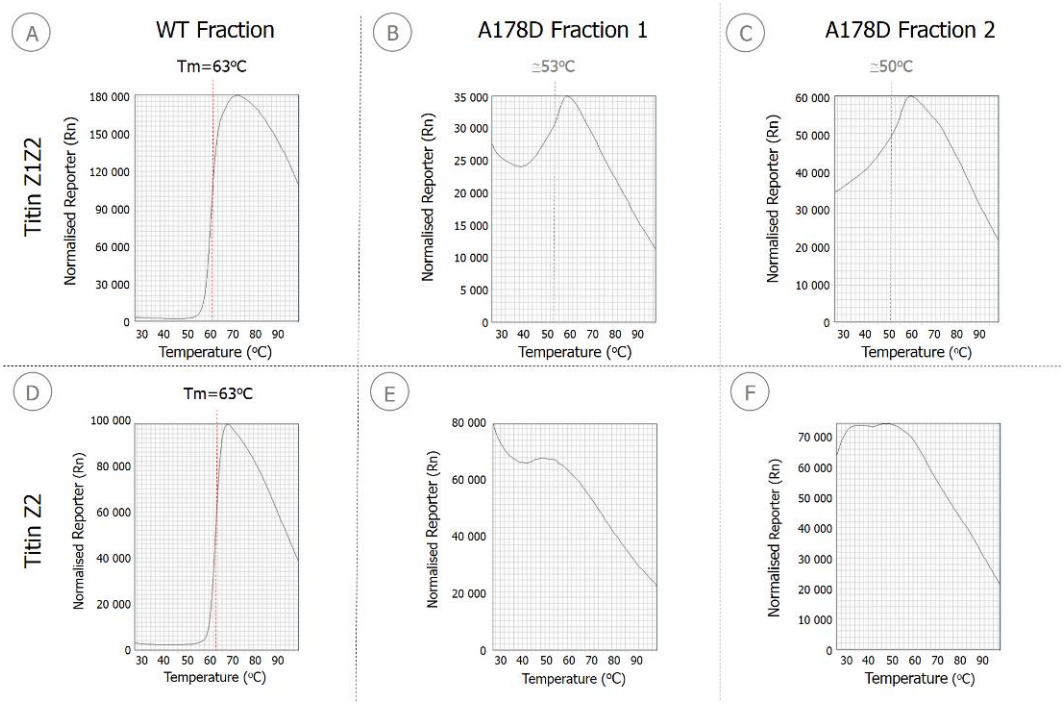


Figure 5-21: Visualisation of fluorescence plots of titin Z1Z2 and Z2 post-purification WT and A178D protein fractions with SYPRO Orange at increasing temperatures (x-axis - temperature, y-axis - fluorescence). Z1Z2 and Z2 WT fraction have negligible baseline fluorescence and clear denaturation curves, with Tm measurements of both being approximately 63ºC. All A178D fractions have significant baseline fluorescence preventing the accurate determination of denaturation temperatures.

5.6.1.5 Circular Dichroism Spectroscopy

Post-purification Z1Z2 WT and A178D protein fractions at room temperature

Circular Dichroism (CD) spectroscopy has been used to gain further information on the folding of the one- and two-domain post-purification titin protein fractions, and more clearly visualise the effects of thermal denaturation. The CD spectrum of the Z1Z2 WT protein fraction (Figure 5-22, blue) is consistent with a protein composed predominantly of β-sheets (Edlich and Muhle-Goll, 2003). The spectrum seen is comparable to other titin Ig-like domains (Figure 5-23) (Politou et al., 1995).

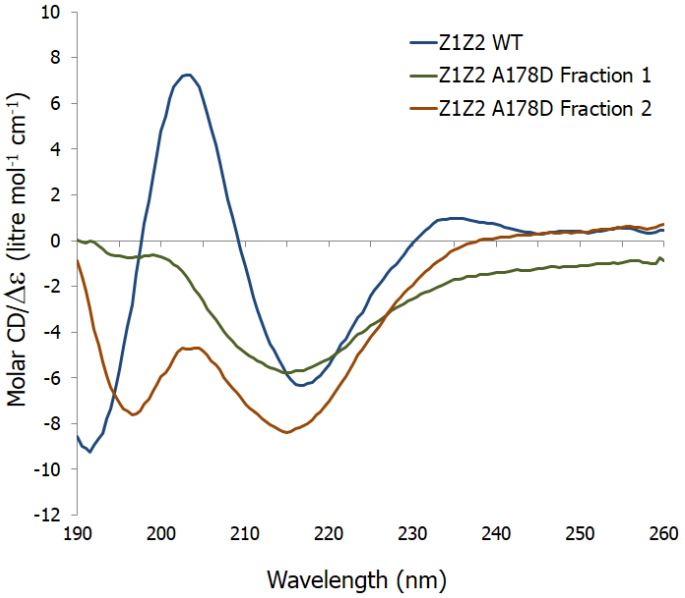


Figure 5-22: CD spectra of purified Z1Z2 WT and A178D protein fractions (0.1 mg/ml) at 25°C with a path length of 1 mm. The WT fraction gives a pattern typical for a protein consisting predominantly of β-sheet motifs. The A178D fractions have similarities to the WT but the marked differences suggest alternative structures.

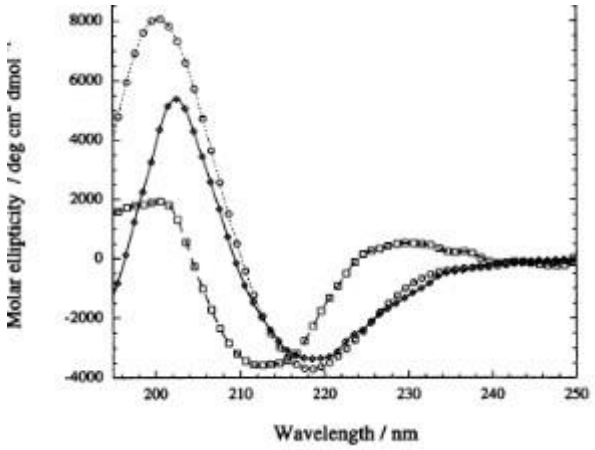


Figure 5-23: CD spectra of other titin Ig-like domains (modules I27- dashed line with squares, A2- solid line with diamonds, M5- dotted line with circles) (Politou et al., 1995). Spectra of I27 and A2 are comparable to the purified Z1Z2 WT protein spectrum.

The purified Z1Z2 A178D fractions show deviations from the WT spectrum at room temperature, with abrogation of the typical β-sheet curves. The algorithms available on the Dichroweb online CD spectra analysis tool were used to interpret the nature of the changes in spectra observed (Whitmore and Wallace, 2004). The interpreted secondary structures of the protein fractions using these algorithms had significant variations between them (a full table of results is shown in appendix section IV-a, page 258). The K2D neuronal network prediction tool (Andrade et al., 1993) was used for analysis as it gave predictions of Z1Z2

WT structure closest to those previously observed. This interprets the A178D protein fractions as having similar secondary structures, with a lower proportion of β-sheet and higher proportion of α-helix when compared to the WT protein fraction (Table 5-6).

	Wild-Type	A178D Fraction 1	A178D Fraction 2
Alpha-helix	0.03	0.21	0.18
Beta Sheet	0.5	0.28	0.33
Random Coil	0.47	0.51	0.49

Table 5-6: Predicted secondary structure of titin Z1Z2 post-purification fractions using the K2D neuronal network prediction tool to interpret the CD spectra (Andrade et al., 1993). This interprets the purified A178D fractions as having a lower proportion of β-sheet and a higher proportion of α-helix motifs, as compared with the WT fraction.

Thermal denaturation CD spectra of purified Z1Z2 WT and A178D protein fractions

Thermal denaturation CD spectroscopy of the purified Z1Z2 WT protein fraction shows reduction of the β-sheet signature with increasing temperatures between 60 and 70°C (blue coloured spectra in Figure 5-24). This suggests protein unfolding at this temperature range, which is in agreement with the thermal denaturation studies in section 5.6.1.4 (page 180). The purified Z1Z2 A178D fractions (green and red coloured spectra in Figure 5-24) also show evidence of denaturation in the same temperature range. All Z1Z2 fractions reach comparable spectra by 70°C, indicating a fully unfolded/denatured state at this temperature.

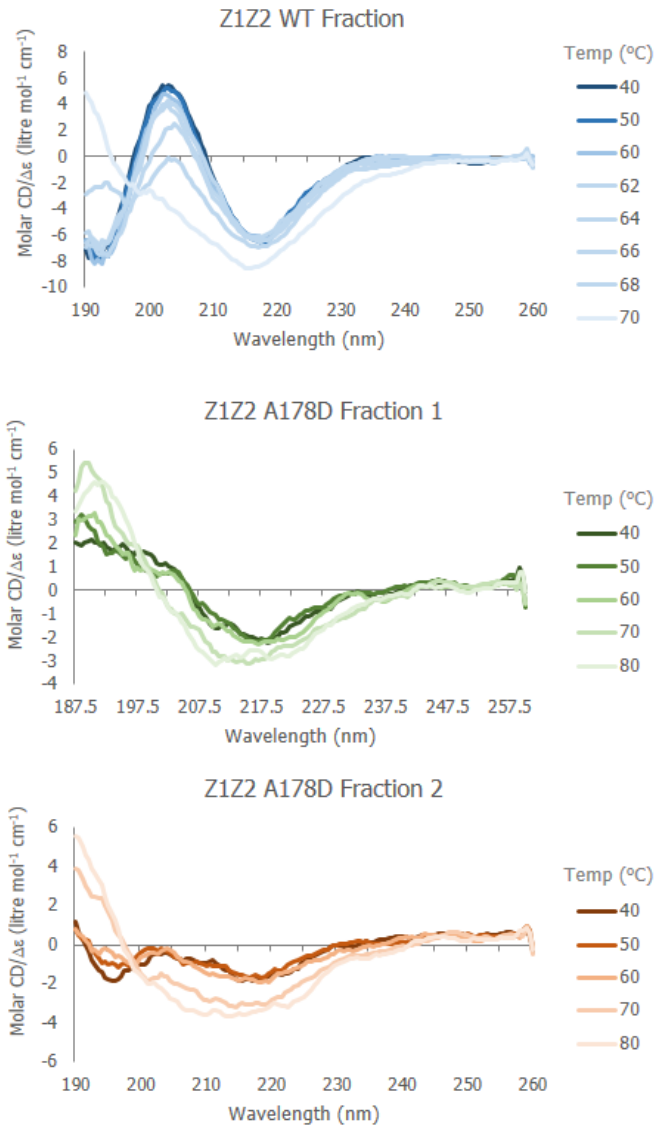


Figure 5-24: CD spectra of the purified titin Z1Z2 WT and A178D protein fractions (0.1 mg/ml) at increasing temperatures (see legends) with an optical path of 1mm. The Z1Z2 WT fraction (top; blue) shows loss of the β-motif signature pattern between 60 and 70°C. The Z1Z2 A178D fractions also show changes over the same temperature range, suggesting further unfolding at these temperatures.

Post-purification Z2 WT and A178D protein fractions at room temperature

The CD spectrum of the purified Z2 WT protein fraction shows the same β-sheet signature as seen with the Z1Z2 WT fraction. The spectrum of the purified Z2 A178D fraction shows an absence of any characteristic features, with little similarity to the WT fraction (Figure 5-25).

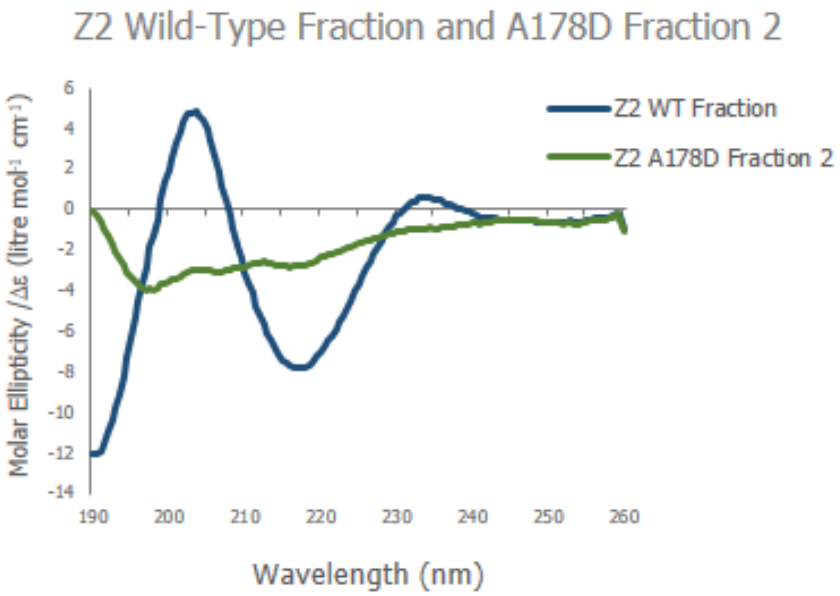


Figure 5-25: CD spectra of the purified titin Z2 WT protein fraction and second A178D fraction (0.1 mg/ml) at room temperature with an optical path of 1 mm. The Z2 WT fraction spectrum exhibits the β-sheet motif signature, whereas the A178D fraction has no characteristic features.

5.6.2 Localisation Studies

5.6.2.1 Localisation of Titin Z1Z2 in Transfected Neonatal Rat Cardiomyocyte and COS1 Cells

Neonatal rat cardiomyocytes form sarcomeres *in vitro*. Transfection of both of these cell types shows localisation of expressed titin Z1Z2 wild-type and A178D GFP fusion proteins to the Z-disk (Figure 5-26). However, GFP alone shows this targeting to the Z-disks to a similar extent, therefore the importance of the localisation of the titin Z1Z2-GFP fusion proteins remains unclear. No aggregation of Z1Z2 A178D is observed. In contrast to Z1Z2 WT and GFP, Z1Z2 A178D is excluded from the nucleus of all of these cell types, most clearly demonstrated in the COS1 cells (Figure 5-27).

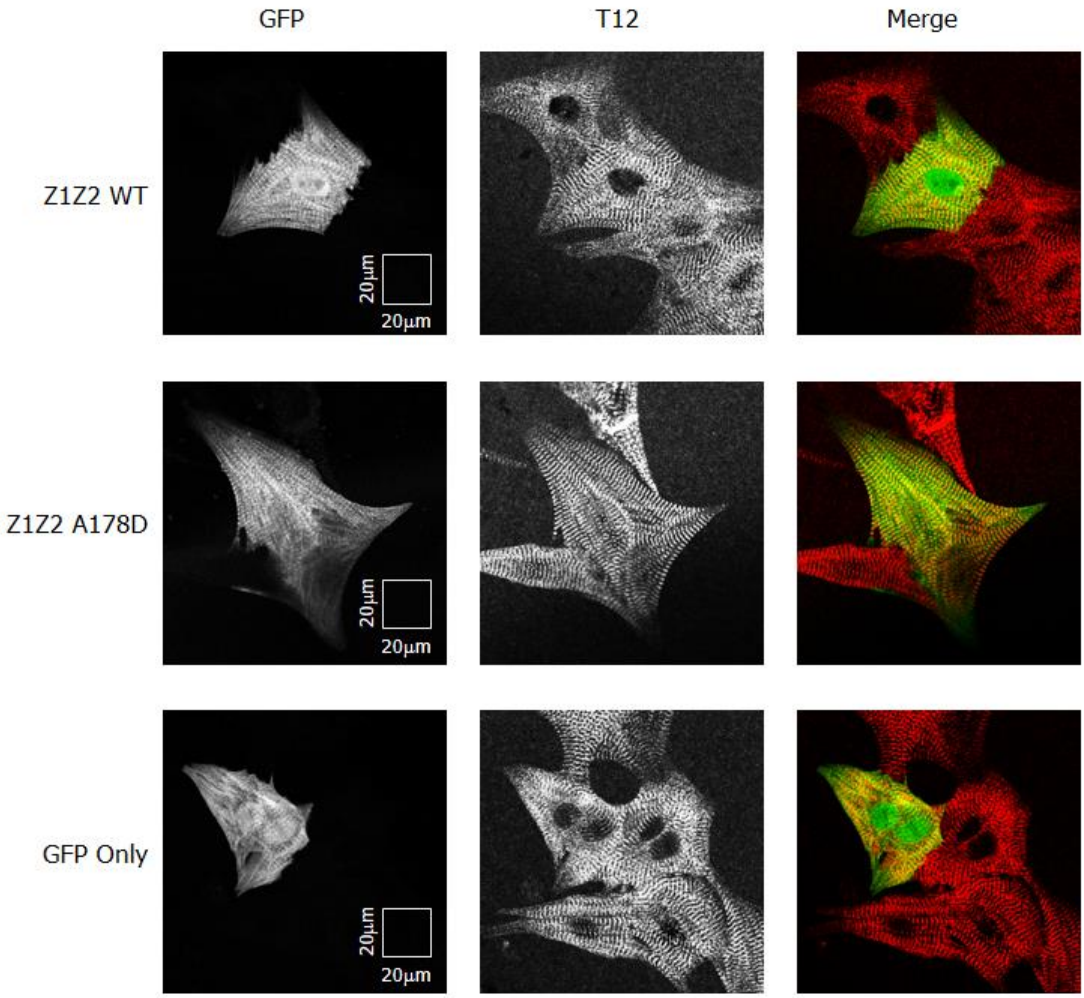


Figure 5-26: Immunofluorescence confocal microscopy images of neonatal rat cardiomyocytes transfected with titin Z1Z2 wild-type (WT) and mutant (A178D) pEGFP plasmid constructs, and pEGFP alone. GFP is shown in the left images and in green in the merged images. T12, a monoclonal murine antibody for a titin epitope close to the Z-disk, followed by an anti-mouse Cy3 antibody are shown in the middle images and shown in red in the merged images. Co-localisation of all three proteins at the Z-disk is demonstrated by yellow in the merged images. Exclusion of GFP signal from the nucleus can be observed in the titin Z1Z2 A178D construct, but not the wild-type or GFP alone.

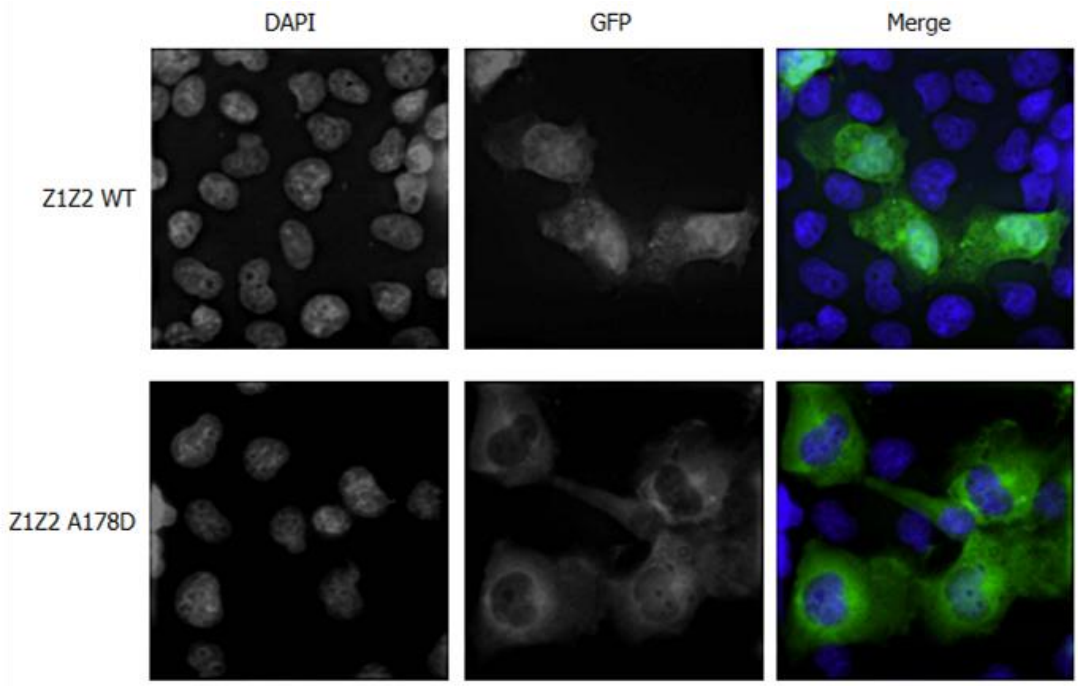


Figure 5-27: Immunofluorescence confocal microscopy images of COS1 cells transfected with titin Z1Z2 wild-type (WT) and mutant (A178D) pEGFP plasmid constructs. DAPI staining highlighting the nucleus is shown in the left images and in blue in the merged images. GFP is shown in the middle images and in green in the merged images. As compared with the wild-type the A178D protein is observed to be excluded from the nucleus.

5.6.3 Binding Studies

5.6.3.1 GST Pull-downs

Titin Z1Z2 WT/A178D and telethonin 1-90/1-167

Titin Z1Z2-GFP and telethonin-GST fusion proteins were expressed in a COS1 cell line as described in the methods section 5.5.2.2 (page 163). The interaction between titin and telethonin proteins was assessed using GST pull-downs and visualised by Western blotting using appropriate antibodies (Figure 5-28). As expected the titin Z1Z2 region binds to the short fragment of telethonin (amino acids 1-90) as well as to full length telethonin (amino acids 1-167). The binding affinity between titin Z1Z2 and the longer telethonin protein appears reduced when compared with the shorter construct (Figure 5-29), as has been documented in previous studies (Pinotsis et al., 2006), although telethonin expression is also reduced.

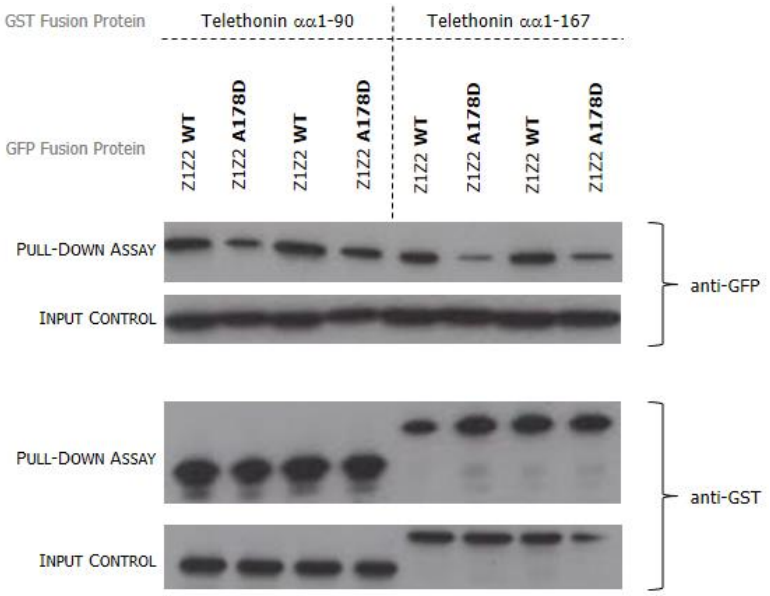


Figure 5-28: Visualisation of titin Z1Z2-telethonin GST pull-downs by immunoblotting of anti-GFP and anti-GST primary antibodies followed by species specific horse radish peroxidase secondary antibodies. Compared with the wild-type there is a reduction in binding of the mutant titin protein to telethonin. For both WT and A178D proteins there is a reduction in binding with the 167 amino acid telethonin protein, when compared with the 90 amino acid telethonin protein, although a reduction in telethonin expression is also seen.

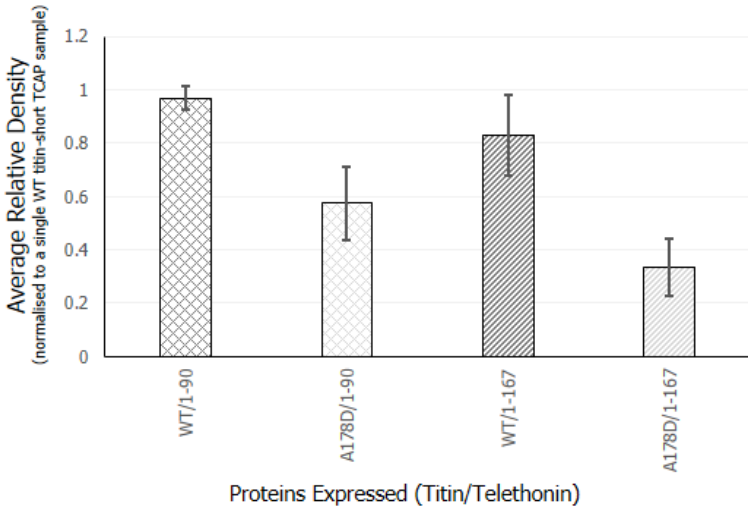


Figure 5-29: Normalised densitometry of the Western blot image shown in Figure 5-28. Densitometry was performed using ImageJ software (Schneider et al., 2012). This demonstrates the relative reduction in band density, and therefore by inference a reduction in titin-telethonin binding, between WT and A178D titin and between the shorter and longer telethonin proteins.

Control experiments were also performed using combinations of expressed titin-GFP or telethonin-GST fusion proteins with GFP or GST proteins alone, expressed in place of the relevant other test sample. These confirmed that no non-specific binding with GST or GFP was evident (Figure 5-30).

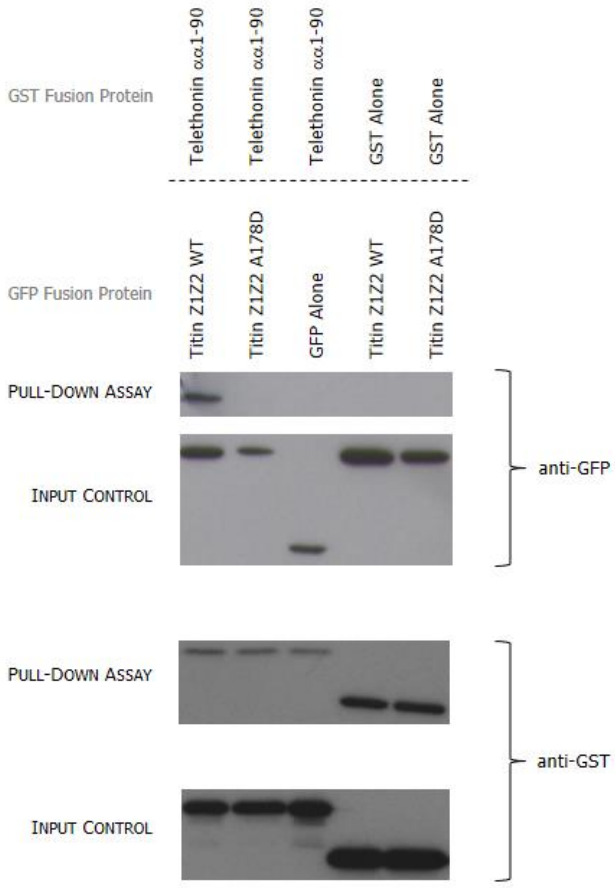


Figure 5-30: Control experiment GST pull-downs. These show no binding between GFP alone and telethonin, and GST alone with WT or A178D titin fragments. Experimental details as in Figure 5-28

Titin Z1Z2 and sAnk1.5

GST pull-down assays were performed to study the proposed interaction between the titin Z1Z2 region and small ankyrin 1.5. Under the experimental conditions used no interaction was observed (Figure 5-31). This is consistent with the findings of other studies (Armani et al., 2006) since the initial report of a proposed direct interaction (Kontrogianni-Konstantopoulos and Bloch, 2003).

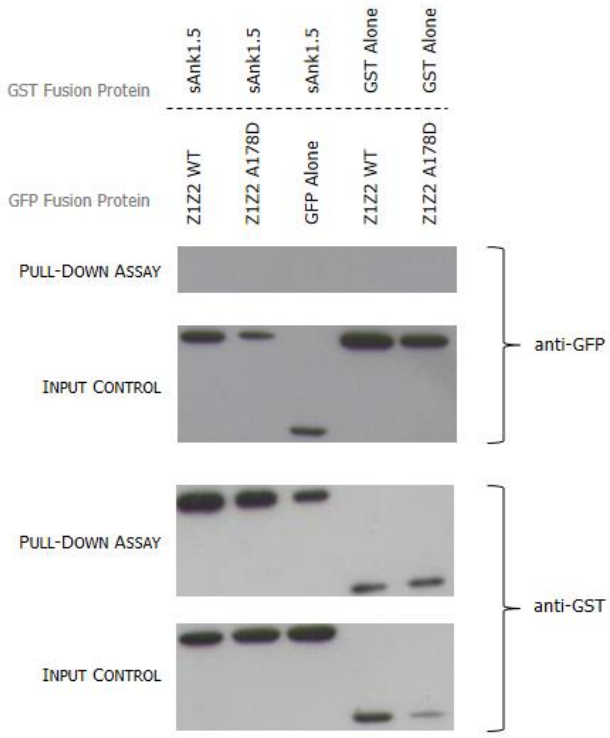


Figure 5-31: Visualisation of titin Z1Z2-sAnk1.5 GST pull-downs by immunoblotting of anti-GFP and anti-GST primary antibodies followed by species specific horse radish peroxidase secondary antibodies. This demonstrates no detectable interaction between titin Z1Z2 and sAnk1.5 under these conditions.

5.7 Discussion

5.7.1 Genetic Findings

Analysis of the alanine to aspartate substitution at position 178, p.A178D, in the Z2 domain of titin was undertaken after this variant was identified segregating in a family with left ventricular non-compaction (LVNC). Full details of this study and results are discussed in section 4.5 (page 119).

5.7.2 Experimental Findings

Overall the *in vitro* studies performed indicate that the substitution of an aspartate residue for an alanine residue at position 178 has a significant effect on the structure of the titin Z1Z2 region, causing unfolding of the Z2 domain. There is evidence that this structural change impairs the interaction of this region with its main ligand, telethonin.

Structure

Under experimental conditions both one- and two-domain purified titin protein A178D fractions adopt two conformations. The first of these fractions is in a larger conformation, as characterised by faster elution on gel filtration chromatography. Despite having the same electrophoretic mobility under denaturing conditions, in the native state these fractions are retained at the top of the gel. This suggests that this fraction is forming multimeric protein aggregates in solution.

The second A178D fractions exhibit more similarities in structural characteristics to the wild-type fractions, but again with a larger conformation as characterised by faster elution on gel filtration. Gel electrophoresis shows similar electrophoretic mobility as the wild-type proteins, although with subtle differences in both denatured and native states. The exact reasons for these differences remain undetermined but are assumed to due to changes in biophysical properties from an alteration in conformational state.

Thermolysin proteolysis demonstrates that the second Z1Z2 A178D protein fraction has unfolded, at least partially, to expose the hydrophobic amino acids at the domain core. This proteolysis of the A178D fraction is only partial, with a stable protein of approximately half the molecular weight of the initial protein remaining at the end-point of this experiment. Given what is known of the structure of this region (Figure 5-11, page 172), and the disruptive amino acid substitution occurring in the Z2 domain, it can be inferred that this represents a stable Z1 domain. This is supported by data from CD spectroscopy.

CD spectroscopy of the one-domain, Z2, A178D protein fraction at room temperature produces a relatively featureless spectrum, indicative of a largely unfolded/denatured protein, unlike the clear β-sheet signature observed with the wild-type. Two-domain, Z1Z2, A178D fractions have spectrums suggesting maintained β-sheet motif, but loss of quantity of this compared to the wild-type. The room temperature second Z1Z2 A178D fraction has

strong resemblances to the spectrum of the Z1Z2 WT fraction at temperatures between 60 and 70°C (Figure 5-32). As determined by this analysis, and fluorescence thermal denaturation studies, at these temperatures the wild-type protein is denaturing (with an experimentally determined Tm of 63°C). Therefore the similarity of the wild-type fraction at 68°C (shown as the light blue spectrum in Figure 5-32), when the majority of the protein is unfolding, to the mutant fraction at room temperature (shown as the green spectrum in Figure 5-32) is further evidence of significant unfolding of Ig-like domains of the mutant protein *in vitro*.

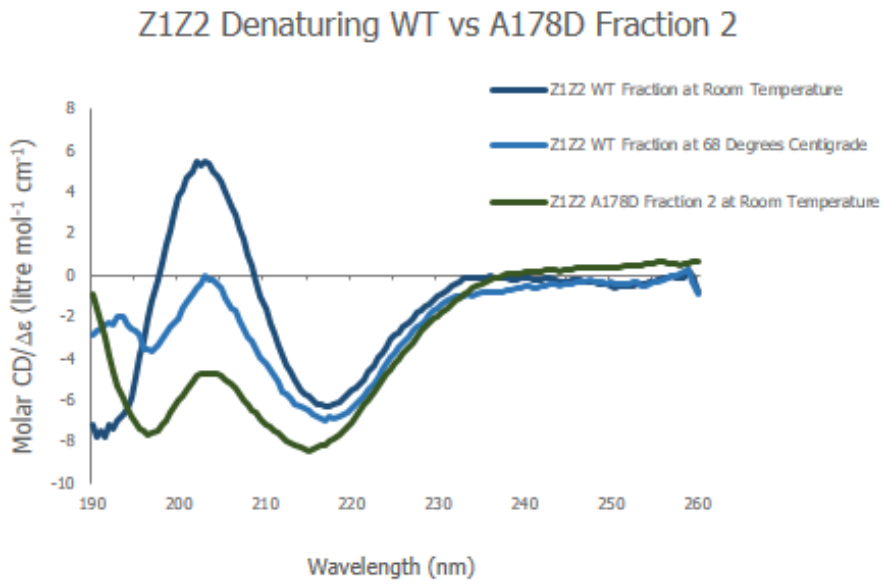


Figure 5-32: CD spectra of titin Z1Z2 WT fraction at room temperature and 68°C and the second titin Z1Z2 A178D fraction at room temperature. This shows a high level of concordance between the denaturing WT fraction, at 68°C, and the A178D fraction at room temperature.

Localisation

Titin Z1Z2 WT-GFP and titin Z1Z2 A178D-GFP protein constructs both localised to the Z-disk in cardiomyocytes. However, as GFP protein alone had a similar distribution under the conditions used, the ability to infer significant conclusions from this is limited. It is notable that no aggregation of the mutant protein is observed, implying it is stable within these cell cultures. One difference observed was that, unlike the WT-GFP and GFP proteins, the A178D-GFP protein was excluded from the nucleus in these transfected cells. Although this

is almost certainly of no pathological relevance, as this region would usually be part of the giant structural protein titin, it implies a conformational difference of the mutant protein preventing passage across the nuclear pores. The size of the titin Z1Z2-GFP fusion proteins (approximately 48 kDa) are about the threshold size for passive movement through the nuclear pores (Rodriguez et al., 2004), therefore a change to a higher volume conformation could prevent this.

Binding

GST pull-down assays have been used to determine the interaction between the Z1Z2 region of titin and telethonin. The binding of wild-type Z1Z2 protein and telethonin is confirmed. Although an interaction between the mutant, A178D, Z1Z2 protein is also observed, this appears significantly reduced when compared with the wild-type protein. Förster/Fluorescence resonance energy transfer (FRET) studies have been performed by Mr Andrea Ghisleni in the laboratory of Professor Mathias Gautel, Randall Division of Cell and Molecular Biophysics and Cardiovascular Division, King's College, London to study this interaction further. The full details of this have not yet been discussed but A178D, introduced into a Z1Z3 construct, has been shown to significantly reduce the affinity of titin for telethonin (WT FRET%- 31.9, SD 20.78; A178D FRET%- 6.11, SD 13.72).

Although binding of the titin Z1Z2 region with small ankyrin 1.5 has been proposed (Kontrogianni-Konstantopoulos and Bloch, 2003) there was no evidence of a significant interaction of the Z1Z2 region of titin with this protein under the conditions used.

5.7.3 Titin Mutations as a Cause of Left Ventricular Non-Compaction

The nature and pathogenesis of LVNC is poorly understood, with the majority of even familial cases having no genetic cause identified (Ackerman et al., 2011). Mutations in titin have not previously been reported to cause LVNC, although mutations in other genes encoding proteins of the sarcomere have (Klaassen et al., 2008). Mutations in titin have

been recognised as a cause of other forms of cardiomyopathy, most notably DCM (Herman et al., 2012).

This association of titin mutations and DCM is particularly notable because many of the families with LVNC, including the one described here, have some features of DCM in addition to LVNC. It would therefore be unsurprising if significant aetiological overlap between LVNC and DCM was observed. If this does occur this finding would have significant implications for the understanding of LVNC. Establishing why some patients have cardiac remodelling to give a non-compacted LVNC phenotype, and others purely have thinning of the myocardium as seen with a DCM phenotype, may provide mechanistic insight in to the processes of cardiac remodelling.

Due to the difficulty in interpreting titin missense variants, because of their high frequency in the general population, relatively few reports of DCM causing missense mutations have been reported (Itoh-Satoh et al., 2002) when compared with truncating mutations (Herman et al., 2012, Gerull et al., 2006, Gerull et al., 2002, Yoskovitz et al., 2012). The missense variants reported by Itoh-Satoh (Itoh-Satoh et al., 2002) also occur in the Z-disk region of titin, with one of these variants being predicted to disrupt the interaction of the Z1Z2 region with telethonin. However, it is difficult to draw definite inferences from such small data sets. Further information on this will be gained with the more frequent analysis of *TTN* in DCM patients now starting to be undertaken clinically with the increased sequencing capabilities of NGS.

Sequencing of the Z1Z2 region of *TTN* in our cohort of consented LVNC patients did not identify any likely pathogenic variants (section 4.5.4.2, page 128). However, due to the small size of this cohort (n=13) this is not informative in ruling out mutations in this region as a significant cause of this condition. Particularly given the observed clinical and genetic heterogeneity in LVNC, larger cohorts will be required to determine the significance of titin

mutations in this condition in general, and whether the titin Z-disk or titin-telethonin interaction is specifically important.

5.7.4 Clinical Significance of the Titin Z1Z2-Telethonin Interaction

Evidence for the consequences of impairing the titin-telethonin interaction *in situ* is inconsistent. Simply from a biomechanical point of view the high mechanical strength of the bond formed between the titin Z1Z2 region and telethonin would imply it is of physiological importance (Lee et al., 2006). This is supported by the finding of myofibrillar disruption when this interaction is hindered *in vitro* (Gregorio et al., 1998), and variably in zebrafish studies (Zhang et al., 2009). However, disrupting this interaction in a mouse model, by knockout of telethonin, leads to only a relatively mild late-onset myopathic phenotype (Markert et al., 2010). In clinical studies mutations in titin (Itoh-Satoh et al., 2002) and telethonin (Hayashi et al., 2004) genes impairing the titin Z1Z2-telethonin interaction have been reported to cause dilated cardiomyopathy. However, this is tempered by the observation that other variants reported to completely disrupt this interaction, such as the p.E13del variant in telethonin, are found at a significant level (approximately 2%) in the normal population (Perrot et al., 2006, Marziliano et al., 2006, EVS, March 2013).

It is evident that despite the molecular strength and significance of the titin-telethonin interaction *in vitro*, the clinical effects of mutations which diminish this interaction are less dramatic than may have been predicted from these. The phenotype seen in the family studied is in keeping with this relatively conserved function of the sarcomere. The non-compaction phenotype observed typically did not develop until adolescence or adulthood, and in some family members late onset and/or indolent clinical course have been observed. Assuming pathogenicity of this titin variant, ascertaining what effects this is causing *in situ*

will be informative in determining the importance and roles of the titin-telethonin interaction.

5.8 Conclusions

Genetic studies have identified a titin p.A178D variant segregating in a family with autosomal dominant left ventricular non-compaction. *In vitro* studies have shown this variant causes unfolding of the titin Z2 domain and impairs titin-telethonin binding. Although together the genetic and experimental data make a strong case for pathogenicity of this variant this remains unproven. Assuming this variant is causative for the phenotype observed, the molecular mechanism of pathogenicity of this variant is as yet undetermined.

5.9 Future Work

There are several further strands of analysis that can provide further support for pathogenicity for the p.A178D titin variant, and if so help understand the mechanisms by which it causes an autosomal dominant non-compaction phenotype. Further *in vitro* work to clarify some of the experimental findings presented, and provide further information, are planned. Adenoviral titin Z1Z2 constructs to infect cardiomyocyte cell cultures have been generated. These studies will ascertain whether A178D has an effect on stability of the titin protein, which may lead to premature degradation and a haploinsufficiency phenotype or result in an altered protein with diminished function of the wild-type which could lead to dominant negative effects on the functions of titin. These can also be used to better assess the effects of the variant on the appropriate localisation titin and telethonin, which may play a role in pathogenicity. However, such *in vitro* studies will still fail to clearly assess the effects of A178D *in situ*.

Although not currently planned the introduction of A178D into a transgenic animal model, typically mouse or zebrafish, would allow more detailed study of the effects of the variant

in situ. Causation of a cardiomyopathy phenotype in such a model would provide further supportive evidence of pathogenicity and would provide access to muscle tissue for more mechanistic studies. However, many of the same questions can be studied by analysis of patient muscle tissue. Muscle samples have been obtained from one family member harbouring the titin variant and further samples are planned.

Analysis of muscle tissue should allow further study in a few key ways. Firstly the presence of the variant can be confirmed at the level of the mRNA and protein *in situ*, and quantification of the relative levels of wild-type to mutant can be undertaken. Whether the variant affects the functions of this region can be determined, for example by studying the appropriate localisation of telethonin. Any effects on titin more generally can also be studied, such as causing premature degradation or a shift in isoform expression. It will also allow analysis of the effects on muscle mechanics, such as whether the variant affects the passive stiffness of the myofibrils. This muscle biopsy should also allow further study of pyruvate dehydrogenase and mitochondrial function in muscle, to determine whether there is evidence for a pathogenic role of the *PDP2* variant identified in the same family. The ethical approval for this study only includes study of lymphocyte DNA, therefore further approval is being sought to allow these further investigations using muscle tissue.

In addition collaborations have been set up to obtain DNA from additional patients with LVNC. Full sequencing of titin, as part of a wider inherited heart conditions panel, has also been established using a Haloplex® capture array method. This should allow the determination of whether titin mutations are seen as more than a rare finding in LVNC, and if so whether there appears to be any regional specificity for these mutations. The family concerned will also continue to be followed-up to provide further information on the phenotype and if the condition can be clearly diagnosed in any other family members this will provide further genetic segregation information and statistical power.

Chapter 6:

Discussion and Conclusions

6.1 Summary

This thesis has addressed the applications of next-generation sequencing to clinical research in inherited heart diseases. Both targeted gene analysis and genome sequencing have been used with the aim of identifying novel causes of different inherited heart disorders.

The role of mitochondrial genetic variants in hypertrophic cardiomyopathy (HCM) was analysed from data produced following the introduction of a multi-gene next-generation sequencing (NGS) panel for clinical diagnostic use in this condition. The m.4300A>G mutation was identified in five families with an HCM-like phenotype. This mutation had previously been reported in three families with left ventricular hypertrophy (Taylor et al., 2003). Identification of the families in this study has provided better delineation of the clinical phenotype caused, and demonstrated that this mutation is one of the most significant single mutation causes of an HCM phenotype. Five novel mitochondrial variants have also been identified in other patients with HCM. Although there is evidence that most of these are rare polymorphisms, one or more of these may represent previously unrecognised causes of this phenotype.

Whole genome sequencing has been performed in members of five families with different autosomal dominant inherited heart diseases of unknown genetic aetiology. This was done to identify novel causes of these disorders, to provide clinical utility and biological insights, and also to make a preliminary assessment of the utility of exome sequencing in clinical diagnostic testing. In two of these families likely pathogenic variants have been identified,

a variant in the *TTN* gene in a family with left ventricular non-compaction, and a variant in *CACNA1C* in a family with a combined phenotype of left ventricular hypertrophy, prolonged QT interval, and sinoatrial node dysfunction. In one family, with an unusual phenotype incorporating elements of different cardiomyopathy types and electrophysiological abnormalities, multiple possibly contributory genetic variants were found. In this family two different chromosome 17 duplications, a 430 kb duplication at 17q11.2 and 1.6 Mb duplication at 17q12, a novel mitochondrial variant, m.15745C>T, and two plausibly causal autosomal variants, in *FGF14* and *FLNC,* were identified. In the other two families no variants with good evidence for pathogenicity were recognised.

Follow-up studies were performed in four of these families. Where possible DNA from other patients with similar phenotypes has been obtained, and sequencing of the study-identified candidate genes undertaken. However, to date no supportive evidence for these variants has been acquired in this manner. The possible effects of the identified *TTN* variant, and support for it being pathogenic, have been obtained from *in vitro* biochemical and biophysical experiments. This variant, p.A178D in the titin Z2 domain, causes a significant structural protein change, with evidence of unfolding of the Z2 Ig-like domain, and a reduction in binding between this region of titin and its main ligand, telethonin, is observed.

6.2 Discussion

6.2.1 Multi-Gene Next Generation Sequencing Tests in Clinical Research and Diagnostics

The use of multi-gene NGS panels allows rapid and cost-efficient clinical diagnostic testing in specific inherited heart diseases (Meder et al., 2011, Ware et al., 2013). These techniques are therefore establishing themselves as the standard technique for such disorders in this setting. NGS has the capability to sequence DNA on an even greater scale, such as exome

sequencing, for relatively little additional cost. Therefore, in the research environment exome sequencing has tended to be used in preference to multi-gene tests.

The analysis of variation in the mitochondrial genome demonstrates that, in addition to clinical diagnostic use, such tests can be used to address research questions in a focussed manner. Indeed, from the whole genome sequencing studies the genes in which likely pathogenic variants have been found, *TTN* and *CACNA1C*, are already known causes of inherited heart diseases, albeit for different phenotypes (Boczek et al., 2013, Splawski et al., 2004, Itoh-Satoh et al., 2002, Gerull et al., 2002, Herman et al., 2012). Therefore, it could be argued, that both of these would have been detected by a gene panel which incorporated all known inherited heart disease associated genes. In a research setting there are benefits to having the additional exomic, or genomic, sequence. The lack of plausible variants in other genes is in itself supportive of likely pathogenicity of variants identified. In addition the possibility to identify variants in genes beyond known 'candidates' may provide opportunities for insights into currently unrecognised pathological mechanisms important for a particular phenotype.

6.2.2 Exome and Genome Studies in Clinical Research and Diagnostics

In the clinical diagnostic setting, where most patients have a phenotype with at least partially known genetic aetiology, the requirement for sequencing genes beyond those already known to cause the particular phenotype is usually less. For many of the more common inherited heart diseases, such as HCM, testing of recognised genetic causes will identify the mutation responsible in the majority of patients (Ackerman et al., 2011), meaning that further genetic sequencing is unrequired. In those patients where a causative mutation is not found further analysis of exomic variants is usually uninterpretable. Sequencing of all of the protein-coding genes in an individual will identify several thousand variants, many of which will be rare, and several will be predicted to be loss-of-function

alleles (MacArthur et al., 2012). Even on the assumption that the unrecognised aetiology is one of these variants there is no practical way of identifying which one, and proving pathogenicity, within a routine diagnostic service.

In clinical research testing exome, or genome, analysis has been effectively used to identify the causes of Mendelian disorders of unknown genetic aetiology (Gilissen et al., 2011). These successes raise the question as to whether exome sequencing should be becoming available for diagnostic testing. Although the interface between research and diagnostic genetic testing may be overlapping, there are important distinctions. Diagnostic testing requires the test to be practical in a diagnostic service setting and the results produced must be of proven clinical utility, rather than just academic interest. The utility of such testing as a clinical service is currently unproven. Although the causes of a number of Mendelian disorders have been identified using exome sequencing in research testing these are unlikely to be representative of more routine clinical diagnostic testing. Typically for clinical research highly selected families or phenotypes are studied at the outset of a new technology being introduced. Even with this select group the failure to find a supported gene-association is unlikely to be reported, therefore the 'success' rate of exome studies in Mendelian disorders is unknown.

Research exome testing is often individualised to the situation studied, and results interpreted in light of particular knowledge regarding a family or phenotype. The resulting genetic studies often form the basis for further studies to prove pathogenicity, rather than providing immediately clinically actionable findings. Functional studies have increasingly become the rate-limiting step in the study of genetic variants, as sequencing and analysis has become quicker, with a need for a high-throughput *in vitro* testing system. In a clinical service setting a standardised testing, analysing and reporting method is likely to be employed. The need to report all variants, probably without family or disease specific

knowledge, has major resource implications. At present most clinical diagnostic laboratories focus on a number of genes and types of disorder, this allows greater expertise to develop in the interpretation of the variation within these particular genes. This will be much more challenging in reporting exome sequencing findings appropriately. In addition genetic diagnostic laboratories are not staffed, skilled, or funded to be able to provide further studies of pathogenicity beyond computer predictions of significance. Further barriers include the requirements for long-term storage of high numbers of large data sets produced, and the lack of current clarity about the appropriate consenting and return of results to patients, particularly with regards to findings unrelated to the clinical phenotype (van El et al., 2013, Green et al., 2013).

Assuming the cost of exome sequencing continues to fall (Dewey et al., 2012) there may be an argument for performing exome sequencing as the standard diagnostic test, but only analysing the variants in genes known to be relevant to the clinical question. Although only initially providing the same clinical utility as targeted gene tests initially this would have certain advantages. Analyses could be returned to either when further genes are recognised in association with a phenotype, or if the study of other disease related genes or variants is desired in an individual. The data produced across a phenotype cohort would, in time, provide a powerful clinical research tool. However, concerns of consenting and confidentiality of this information, the ethics of not returning potentially medically relevant information to the patient, and also data storage requirements, would remain.

6.2.3 Limitations of Exome Analysis in the Identification of the Cause of Mendelian Disorders

The families studied by whole genome sequencing were selected on the basis of having a highly penetrant dominantly-inherited phenotype, with multiple affected family members. From this it can be assumed that the conditions in these families are likely to be caused by

heterozygous mutations in individual genes. Despite this, in two families no candidate variants with a high probability of being pathogenic were identified. In the families with a probably causal variant identified further studies are still required to confirm a causative role for these. Both the difficulty of proving pathogenicity of variants identified and the inability to identify likely causative variants are major limitations of implementing exome sequencing successfully into a clinical setting.

6.2.3.1 Difficulties Identifying Causative Variants

The possible explanations for the inability to recognise a causal variant from exome sequencing include:

- *The causal variant is present but without evidence for pathogenicity*

  Determination of likely pathogenicity incorporates a number of factors. The known biological function is key in this. Therefore if the biological function for the protein product of a gene is unknown, or only an unrelated role is identified, a variant is unlikely to be recognised as a candidate for pathogenicity. Therefore, the variant may reside in the filtered variant list but not be recognised as being disease causing.

- *The causal variant was identified but was filtered out as being non-pathogenic*

  Depending on the nature and stringency of the analysis filters used a greater or lesser number of variants will be removed. Filtering is a mechanism for focussing analysis on a manageable number of variants, but every filtering step has the potential to remove the causal variant. Commonly applied filters are discussed in more detail in section 1.5.4 (page 15), but typically include criteria such as quality score, frequency of the variant in population data, and the exclusion of variants not predicted to alter protein sequence (Li et al., 2012). All of these filters are based on presumptions about the causative variant. Although these are designed to be correct in most situations they are not

specific for ruling out benign variants, for example variants causing a synonymous change may infrequently be the cause of disease (Buske et al., 2013).

- *The causal variant was outside the exome*

  Although the exome is highly enriched for variants causative for Mendelian disease these may also occur outside this. Intronic regions, including regulatory elements, may all harbour causal variants for Mendelian disease (Cooper et al., 1995). If whole genome sequencing has been performed most of these areas are available for interrogation. Due to the scale of these regions as compared with the exome and the currently limited knowledge regarding their function (Birney et al., 2007), interpretation of the large number of variants within them (Abecasis et al., 2012) is challenging.

- *The causal variant was not identified due to being incorrectly sequenced, called or annotated*

  Whole genome sequencing and exome sequencing have high levels of coverage of the exome sequence but they are not complete. Different methods, such as the exome enrichment kit or whole genome sequencing platform used, will lead to different levels of coverage in regions of the exome (Clark et al., 2011b). Coverage will often be lower due to specific sequence characteristics, with GC-rich areas often being poorly sequenced for example. There are also particular genomic regions, such as repetitive areas, which may be misaligned to the reference genome more frequently. Several such problems are clearly demonstrated by a recent report, regarding the eventual identification of a single nucleotide insertion in a variable-number tandem repeat sequence in the *MUC1* gene as a cause of medullary cystic kidney disease. This was initially missed by various NGS techniques. It was excluded by exome and capture based libraries due to its low complexity and the fact that it was not annotated as coding

sequence, and the WGS sequencing data covering this region was poor due to the high GC level and homopolymer repeats (Kirby et al., 2013).

- *One or more family members have been incorrectly ascribed as affected*

  Phenotyping is key to being able to appropriately interpret the likely causality of variants. Depending on the phenotype the incorrect attribution of status as affected or unaffected may be more or less likely. With the inherited heart disorders studied it is not infrequent to identify family members with borderline changes, on cardiac measurements for example, who in the context of the family may be more likely to be reported as affected. In the DCM family studied in this thesis (section 4.2, page 82) a further family member was initially studied as being affected. She had borderline left ventricular dilatation measured by echocardiogram during pregnancy, so in the context of the family she was assumed to be affected. Although she agreed to take part in the study, she did not wish to come to the Inherited Cardiac Conditions Clinic. Therefore it took some time to appreciate that subsequent investigation, with cardiac MRI and further echocardiograms performed by her general cardiologist, had been entirely normal. This therefore casted significant doubt on the diagnosis in her, and led to her being reclassified as 'unknown' rather an affected.

  A further confounding issue in determining whether an individual is affected occurs in the rare situation when more than one inherited condition with similar phenotype may be present in a family. The family with the combined phenotype of variable cardiomyopathy and electrophysiological abnormalities (section 4.3, page 94) may well demonstrate this phenomenon. Due to the number of family members with rare cardiac features seen in the family it was assumed a single cause was most likely (whether that was a chromosomal abnormality, single-gene cause, or possibly linked co-segregating

variants). Although the genetic results are not validated, they are suggestive that there may be multiple causal variants accounting for different aspects of the phenotype seen.

Another pitfall in determining whether an individual is affected with a condition is in those disorders where phenocopies may occur, for example hypertrophy due to hypertension in a member of a family with HCM. The phenotypes studied in this thesis would seem less susceptible to this issue but it remains a possibility.

- *Incorrect assumptions on the mode of inheritance were made*

  The segregation of disease within a family may be consistent with more than one inheritance pattern. This is obviously important in determining whether a causal variant will be heterozygous, homozygous, X-linked (hemizygous) or in the mitochondrial genome. In two families studied here (sections 4.2 and 4.3) dominant inheritance was felt to be the most likely inheritance pattern, however they were also compatible with X-linked and mitochondrial inheritance. In these families the alternative possible inheritance patterns were acknowledged, but this could easily have been overlooked and the causal variant missed.

6.2.3.2 Difficulties Proving Pathogenicity of Variants

Identification of genetic variants by linkage analysis in extensive, well phenotyped, families have generally been shown to be robust over time. However, some disease genes and variants have been reported without good supportive genetic and/or functional data. These have a higher risk of subsequently being recognised as not being causative (Andreasen et al., 2013, Watkins, 2013). The types of supportive evidence for pathogenicity of an identified variant have typically included: absence in a few hundred control samples, conservation across species, and evidence of functional effect at the protein level *in vitro*. Increasingly it is recognised that individually, and even in combination, these factors may

be far from conclusive. For example the finding from exome and genome studies of a much higher frequency of rare variants than anticipated (Abecasis et al., 2012), means that absence in relatively small control group is of less note than had been presumed. This has been discussed in this thesis in relation to the p.W4R variant in *CSRP3* and p.E13del in *TCAP* (Bos et al., 2006, Knoll et al., 2002). These variants were initially reported as a cause of cardiomyopathy but were later identified as being common in the general population as larger cohorts of population data were studied (section 5.2.2.3, page 143) (Andersen et al., 2009, Perrot et al., 2006, Marziliano et al., 2006, EVS, March 2013).

The extensive pedigrees that were used to identify disease genes using linkage analysis and positional cloning have generally already had causal variants identified. This means large families are only rarely available for the study of the causes of Mendelian diseases by NGS. Most families with Mendelian disorders being studied with exome sequencing therefore tend to have fewer genetically informative family members for study. In such families the number of segregating candidate variants following exome analysis produces a modest list, often of fewer than ten variants. Due to the size of these families the variants are typically within one of several chromosomal linkage regions each with modest odds likelihood (LOD) scores. The genetic data is therefore useful in identifying likely causal variants, but alone is rarely able to identify a single causative variant with strong statistical support for its pathogenicity in the families studied.

It is notable that many of the successes of exome sequencing in dominant Mendelian disorders have not often come from studying families of the size of those reported in chapter 4. They have tended to come from extensive pedigrees that give informative genetic linkage to a single small locus (Wang et al., 2010a, Norton et al., 2011, Weedon et al., 2011), from identifying *de novo* mutations (Ng et al., 2010a, Vissers et al., 2010, Simpson et al., 2011, Clayton-Smith et al., 2011), and/or from studying several families with

a homogeneous disorder (Simpson et al., 2011, Ng et al., 2010a). Only a small number of published studies have used similar families and testing strategies as those employed here. In such families the successes have usually identified a mutation in a gene with known relevant biological mechanism (Regalado et al., 2011, Austin et al., 2012) or in one that is already known to cause a related phenotype (Bonnefond et al., 2012). In these families burden of proof of pathogenicity has come from the techniques employed in this study, namely identification in cohorts with the same phenotype and/or supporting evidence of a functional effect, with transgenic animal studies being performed in some.

By its very nature exome sequencing will identify variants segregating with a phenotype in any family studied. Without high-quality supportive data for pathogenicity a large number of variants may be incorrectly attributed as being disease causing. Many of the measures of disease likelihood, such as predicted function or conservation, may be useful screening tools but are only limited supporting evidence for pathogenicity. Proof of pathogenicity of a variant following exome studies is generally through a weight of additional evidence, although none are truly conclusive.

Identification of variants in a candidate gene in other patients with the same phenotype is commonly performed. In most of the families this was somewhat compromised by the varying degrees of novelty of the phenotypes observed. In itself it is notable that although the more delineated inherited heart disease phenotypes, such as HCM and LQTS, make up the vast majority of clinical practice, it is very rare to identify families with multiple affected family members that do not link to a known genetic cause.

Following the recognition of a variant in a gene as a candidate cause for a phenotype, the study of that gene in patients with the same phenotype can provide good supporting data. However this strategy can have limitations. A good example of this is with one of the genes identified in this study, titin. Titin missense variants are very common in the general

population, averaging approximately twenty-three per individual, and more notably on average one of these will be rare (EVS, March 2013). Even truncating mutations, which are identified as a relatively common cause of DCM (Herman et al., 2012), are comparatively frequent in population data (EVS, March 2013, Norton et al., 2013). *In situ* evidence of effect of the variant in causing the phenotype can provide further good evidence for pathogenicity, whether this is in patients, patient tissue, or animal models.

6.3 Final Conclusions

We are reaching an era where it is possible to cost-effectively sequence exomes, and whole genomes, on a large scale. Although exome sequencing has proven to be a powerful clinical research tool, there are significant barriers to successfully applying it to clinical diagnosis. While sequencing is now relatively easy, managing and interpreting the data produced to inform clinical management remains a significant challenge, even in the context of highly-penetrant Mendelian disease. Identifying a causal variant in genetic disorders is mainly limited by the large number of common and rare variants within the genome, and the difficulties in establishing a burden of proof for pathogenicity. With the UK government announcing funding for the sequencing of 100 000 clinical genomes the amount of clinical genomic data produced in the next few years looks set to increase exponentially. This provides the opportunity to try and tackle these biological and bioinformatic challenges, and has the potential to make genomic sequencing a powerful clinical tool. The development of validated clinically-relevant bioinformatic algorithms and high-throughput *in vitro* test systems are highlighted by this study as being required to make clinical exome testing effective.

References

ABECASIS, G. R., ALTSHULER, D., AUTON, A., BROOKS, L. D., DURBIN, R. M., GIBBS, R. A., HURLES, M. E. & MCVEAN, G. A. 2010. A map of human genome variation from population-scale sequencing. *Nature,* 467, 1061-73.

ABECASIS, G. R., AUTON, A., BROOKS, L. D., DEPRISTO, M. A., DURBIN, R. M., HANDSAKER, R. E., KANG, H. M., MARTH, G. T. & MCVEAN, G. A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491, 56-65.

ABECASIS, G. R., CHERNY, S. S., COOKSON, W. O. & CARDON, L. R. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet,* 30, 97-101.

ABRIEL, H. 2010. Cardiac sodium channel Na(v)1.5 and interacting proteins: Physiology and pathophysiology. *J Mol Cell Cardiol,* 48, 2-11.

ABU-ELHEIGA, L., MATZUK, M. M., KORDARI, P., OH, W., SHAIKENOV, T., GU, Z. & WAKIL, S. J. 2005. Mutant mice lacking acetyl-CoA carboxylase 1 are embryonically lethal. *Proc Natl Acad Sci U S A,* 102, 12011-6.

ACKERMAN, M. J., PRIORI, S. G., WILLEMS, S., BERUL, C., BRUGADA, R., CALKINS, H., CAMM, A. J., ELLINOR, P. T., GOLLOB, M., HAMILTON, R., HERSHBERGER, R. E., JUDGE, D. P., LE MAREC, H., MCKENNA, W. J., SCHULZE-BAHR, E., SEMSARIAN, C., TOWBIN, J. A., WATKINS, H., WILDE, A., WOLPERT, C. & ZIPES, D. P. 2011. HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart Rhythm,* 8, 1308-39.

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.

ALLISON, A. C. 1954. Protection afforded by sickle-cell trait against subtertian malareal infection. *Br Med J,* 1, 290-4.

ALLYSE, M. & MICHIE, M. 2013. Not-so-incidental findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing. *Trends Biotechnol*.

ALSTON, C. L., HE, L., MORRIS, A. A., HUGHES, I., DE GOEDE, C., TURNBULL, D. M., MCFARLAND, R. & TAYLOR, R. W. 2011. Maternally inherited mitochondrial DNA disease in consanguineous families. *Eur J Hum Genet,* 19, 1226-9.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215, 403-10.

ALTSHULER, D. M., GIBBS, R. A., PELTONEN, L., DERMITZAKIS, E., SCHAFFNER, S. F., YU, F., BONNEN, P. E., DE BAKKER, P. I., DELOUKAS, P., GABRIEL, S. B., GWILLIAM, R., HUNT, S., INOUYE, M., JIA, X., PALOTIE, A., PARKIN, M., WHITTAKER, P., CHANG, K., HAWES, A., LEWIS, L. R., REN, Y., WHEELER, D., MUZNY, D. M., BARNES, C., DARVISHI, K., HURLES, M., KORN, J. M., KRISTIANSSON, K., LEE, C., MCCARROL, S. A., NEMESH, J., KEINAN, A., MONTGOMERY, S. B., POLLACK, S., PRICE, A. L., SORANZO, N., GONZAGA-JAUREGUI, C., ANTTILA, V., BRODEUR, W., DALY, M. J., LESLIE, S., MCVEAN, G., MOUTSIANAS, L., NGUYEN, H., ZHANG, Q., GHORI, M. J., MCGINNIS, R., MCLAREN, W., TAKEUCHI, F., GROSSMAN, S. R., SHLYAKHTER, I., HOSTETTER, E. B., SABETI, P. C., ADEBAMOWO, C. A., FOSTER, M. W., GORDON, D. R., LICINIO, J., MANCA, M. C., MARSHALL, P. A., MATSUDA, I., NGARE, D., WANG, V. O., REDDY, D., ROTIMI, C. N., ROYAL, C. D., SHARP, R. R., ZENG, C., BROOKS, L. D. & MCEWEN, J. E. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature,* 467, 52-8.

ANAN, R., NAKAGAWA, M., MIYATA, M., HIGUCHI, I., NAKAO, S., SUEHARA, M., OSAME, M. & TANAKA, H. 1995. Cardiac involvement in mitochondrial diseases. A study on 17 patients with documented mitochondrial DNA defects. *Circulation,* 91, 955-61.

ANDERSEN, P. S., HAVNDRUP, O., HOUGS, L., SORENSEN, K. M., JENSEN, M., LARSEN, L. A., HEDLEY, P., THOMSEN, A. R., MOOLMAN-SMOOK, J., CHRISTIANSEN, M. & BUNDGAARD, H. 2009. Diagnostic yield, interpretation, and clinical utility of mutation screening of sarcomere encoding genes in Danish hypertrophic cardiomyopathy patients and relatives. *Hum Mutat,* 30, 363-70.

ANDERSON, B. R., BOGOMOLOVAS, J., LABEIT, S. & GRANZIER, H. 2013. Single molecule force spectroscopy on titin implicates immunoglobulin domain stability as a cardiac disease mechanism. *J Biol Chem,* 288, 5303-15.

ANDERSON, S., BANKIER, A. T., BARRELL, B. G., DE BRUIJN, M. H., COULSON, A. R., DROUIN, J., EPERON, I. C., NIERLICH, D. P., ROE, B. A., SANGER, F., SCHREIER, P. H., SMITH, A. J., STADEN, R. & YOUNG, I. G. 1981. Sequence and organization of the human mitochondrial genome. *Nature,* 290, 457-65.

ANDRADE, M. A., CHACON, P., MERELO, J. J. & MORAN, F. 1993. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng,* 6, 383-90.

ANDREASEN, C., NIELSEN, J. B., REFSGAARD, L., HOLST, A. G., CHRISTENSEN, A. H., ANDREASEN, L., SAJADIEH, A., HAUNSO, S., SVENDSEN, J. H. & OLESEN, M. S. 2013. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur J Hum Genet*.

ANDREWS, R. M., KUBACKA, I., CHINNERY, P. F., LIGHTOWLERS, R. N., TURNBULL, D. M. & HOWELL, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet,* 23, 147.

ANTZELEVITCH, C., POLLEVICK, G. D., CORDEIRO, J. M., CASIS, O., SANGUINETTI, M. C., AIZAWA, Y., GUERCHICOFF, A., PFEIFFER, R., OLIVA, A., WOLLNIK, B., GELBER, P., BONAROS, E. P., JR., BURASHNIKOV, E., WU, Y., SARGENT, J. D., SCHICKEL, S., OBERHEIDEN, R., BHATIA, A., HSU, L. F., HAISSAGUERRE, M., SCHIMPF, R., BORGGREFE, M. & WOLPERT, C. 2007. Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. *Circulation,* 115, 442-9.

ARBER, S., HALDER, G. & CARONI, P. 1994. Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation. *Cell,* 79, 221-31.

ARBER, S., HUNTER, J. J., ROSS, J., JR., HONGO, M., SANSIG, G., BORG, J., PERRIARD, J. C., CHIEN, K. R. & CARONI, P. 1997. MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell,* 88, 393-403.

ARBUSTINI, E., DIEGOLI, M., FASANI, R., GRASSO, M., MORBINI, P., BANCHIERI, N., BELLINI, O., DAL BELLO, B., PILOTTO, A., MAGRINI, G., CAMPANA, C., FORTINA, P., GAVAZZI, A., NARULA, J. & VIGANO, M. 1998. Mitochondrial DNA mutations and mitochondrial abnormalities in dilated cardiomyopathy. *Am J Pathol,* 153, 1501-10.

ARMANI, A., GALLI, S., GIACOMELLO, E., BAGNATO, P., BARONE, V., ROSSI, D. & SORRENTINO, V. 2006. Molecular interactions with obscurin are involved in the localization of muscle-specific small ankyrin1 isoforms to subcompartments of the sarcoplasmic reticulum. *Exp Cell Res,* 312, 3546-58.

ARNDT, A. K., SCHAFER, S., DRENCKHAHN, J. D., SABEH, M. K., PLOVIE, E. R., CALIEBE, A., KLOPOCKI, E., MUSSO, G., WERDICH, A. A., KALWA, H., HEINIG, M., PADERA, R. F., WASSILEW, K., BLUHM, J., HARNACK, C., MARTITZ, J., BARTON, P. J., GREUTMANN, M., BERGER, F., HUBNER, N., SIEBERT, R., KRAMER, H. H., COOK, S. A., MACRAE, C. A. & KLAASSEN, S. 2013. Fine Mapping of the 1p36 Deletion Syndrome Identifies Mutation of PRDM16 as a Cause of Cardiomyopathy. *Am J Hum Genet*.

ASHRAFIAN, H., MCKENNA, W. J. & WATKINS, H. 2011. Disease pathways and novel therapeutic targets in hypertrophic cardiomyopathy. *Circ Res,* 109, 86-96.

ASHRAFIAN, H., REDWOOD, C., BLAIR, E. & WATKINS, H. 2003. Hypertrophic cardiomyopathy:a paradigm for myocardial energy depletion. *Trends Genet,* 19, 263-8.

AUGUSTIN, I., KORTE, S., RICKMANN, M., KRETZSCHMAR, H. A., SUDHOF, T. C., HERMS, J. W. & BROSE, N. 2001. The cerebellum-specific Munc13 isoform Munc13-3 regulates cerebellar synaptic transmission and motor learning in mice. *J Neurosci,* 21, 10-7.

AUSTIN, E. D., MA, L., LEDUC, C., BERMAN ROSENZWEIG, E., BORCZUK, A., PHILLIPS, J. A., 3RD, PALOMERO, T., SUMAZIN, P., KIM, H. R., TALATI, M. H., WEST, J., LOYD, J. E. & CHUNG, W. K. 2012. Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. *Circ Cardiovasc Genet,* 5, 336-43.

BAGNATO, P., BARONE, V., GIACOMELLO, E., ROSSI, D. & SORRENTINO, V. 2003. Binding of an ankyrin-1 isoform to obscurin suggests a molecular link between the sarcoplasmic reticulum and myofibrils in striated muscles. *J Cell Biol,* 160, 245-53.

BAIG, S. M., KOSCHAK, A., LIEB, A., GEBHART, M., DAFINGER, C., NURNBERG, G., ALI, A., AHMAD, I., SINNEGGER-BRAUNS, M. J., BRANDT, N., ENGEL, J., MANGONI, M. E., FAROOQ, M., KHAN, H. U., NURNBERG, P., STRIESSNIG, J. & BOLZ, H. J. 2011. Loss of Ca(v)1.3 (CACNA1D) function in a human channelopathy with bradycardia and congenital deafness. *Nat Neurosci,* 14, 77-84.

BAILEY, J. A., GU, Z., CLARK, R. A., REINERT, K., SAMONTE, R. V., SCHWARTZ, S., ADAMS, M. D., MYERS, E. W., LI, P. W. & EICHLER, E. E. 2002. Recent segmental duplications in the human genome. *Science,* 297, 1003-7.

BANG, M. L., CENTNER, T., FORNOFF, F., GEACH, A. J., GOTTHARDT, M., MCNABB, M., WITT, C. C., LABEIT, D., GREGORIO, C. C., GRANZIER, H. & LABEIT, S. 2001. The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circ Res,* 89, 1065-72.

BATES, M. G., BOURKE, J. P., GIORDANO, C., D'AMATI, G., TURNBULL, D. M. & TAYLOR, R. W. 2012. Cardiac involvement in mitochondrial DNA disease: clinical spectrum, diagnosis, and management. *Eur Heart J*.

BELL, D. W., SIKDAR, N., LEE, K. Y., PRICE, J. C., CHATTERJEE, R., PARK, H. D., FOX, J., ISHIAI, M., RUDD, M. L., POLLOCK, L. M., FOGOROS, S. K., MOHAMED, H., HANIGAN, C. L., ZHANG, S., CRUZ, P., RENAUD, G., HANSEN, N. F., CHERUKURI, P. F., BORATE, B., MCMANUS, K. J., STOEPEL, J., SIPAHIMALANI, P., GODWIN, A. K., SGROI, D. C., MERINO, M. J., ELLIOT, G., ELKAHLOUN, A., VINSON, C., TAKATA, M., MULLIKIN, J. C., WOLFSBERG, T. G., HIETER, P., LIM, D. S. & MYUNG, K. 2011. Predisposition to cancer caused by genetic and functional defects of mammalian Atad5. *PLoS Genet,* 7, e1002245.

BELLEFROID, E. J., MARINE, J. C., RIED, T., LECOCQ, P. J., RIVIERE, M., AMEMIYA, C., PONCELET, D. A., COULIE, P. G., DE JONG, P., SZPIRER, C. & ET AL. 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *EMBO J,* 12, 1363-74.

BENNETT, V. & CHEN, L. 2001. Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. *Curr Opin Cell Biol,* 13, 61-7.

BERG, J., TYMOCZKO, J. & STRYER, L. 2002. Oxidative Phosphorylation. *Biochemistry.* 5th ed.: WH Freeman.

BERGMANN, O., BHARDWAJ, R. D., BERNARD, S., ZDUNEK, S., BARNABE-HEIDER, F., WALSH, S., ZUPICICH, J., ALKASS, K., BUCHHOLZ, B. A., DRUID, H., JOVINGE, S. & FRISEN, J. 2009. Evidence for cardiomyocyte renewal in humans. *Science,* 324, 98-102.

BERTZ, M., WILMANNS, M. & RIEF, M. 2009. The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proc Natl Acad Sci U S A,* 106, 13307-133310.

BIERHALS, T., MADDUKURI, S. B., KUTSCHE, K. & GIRISHA, K. M. 2013. Expanding the phenotype associated with 17q12 duplication: case report and review of the literature. *Am J Med Genet A,* 161A, 352-9.

BIESECKER, L. G. 2012. Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. *Genet Med,* 14, 393-8.

BILBE, G., DELABIE, J., BRUGGEN, J., RICHENER, H., ASSELBERGS, F. A., CERLETTI, N., SORG, C., ODINK, K., TARCSAY, L., WIESENDANGER, W. & ET AL. 1992. Restin: a novel intermediate filament-associated protein highly expressed in the Reed-Sternberg cells of Hodgkin's disease. *EMBO J,* 11, 2103-13.

BIRNBAUMER, L., CAMPBELL, K. P., CATTERALL, W. A., HARPOLD, M. M., HOFMANN, F., HORNE, W. A., MORI, Y., SCHWARTZ, A., SNUTCH, T. P., TANABE, T. & ET AL. 1994. The naming of voltage-gated calcium channels. *Neuron,* 13, 505-6.

BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGO, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., ADZHUBEI, I., GREENBAUM, J. A., ANDREWS, R. M., FLICEK, P., BOYLE, P. J., CAO, H., CARTER, N. P., CLELLAND, G. K., DAVIS, S., DAY, N., DHAMI, P., DILLON, S. C., DORSCHNER, M. O., FIEGLER, H., GIRESI, P. G., GOLDY, J., HAWRYLYCZ, M., HAYDOCK, A., HUMBERT, R., JAMES, K. D., JOHNSON, B. E., JOHNSON, E. M., FRUM, T. T., ROSENZWEIG, E. R., KARNANI, N., LEE, K., LEFEBVRE, G. C., NAVAS, P. A., NERI, F., PARKER, S. C., SABO, P. J., SANDSTROM, R., SHAFER, A., VETRIE, D., WEAVER, M., WILCOX, S., YU, M., COLLINS, F. S., DEKKER, J., LIEB, J. D., TULLIUS, T. D., CRAWFORD, G. E., SUNYAEV, S., NOBLE, W. S., DUNHAM, I., DENOEUD, F., REYMOND, A., KAPRANOV, P.,

ROZOWSKY, J., ZHENG, D., CASTELO, R., FRANKISH, A., HARROW, J., GHOSH, S., SANDELIN, A., HOFACKER, I. L., BAERTSCH, R., KEEFE, D., DIKE, S., CHENG, J., HIRSCH, H. A., SEKINGER, E. A., LAGARDE, J., ABRIL, J. F., SHAHAB, A., FLAMM, C., FRIED, C., HACKERMULLER, J., HERTEL, J., LINDEMEYER, M., MISSAL, K., TANZER, A., WASHIETL, S., KORBEL, J., EMANUELSSON, O., PEDERSEN, J. S., HOLROYD, N., TAYLOR, R., SWARBRECK, D., MATTHEWS, N., DICKSON, M. C., THOMAS, D. J., WEIRAUCH, M. T., GILBERT, J., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447, 799-816.

BLAIR, E., REDWOOD, C., ASHRAFIAN, H., OLIVEIRA, M., BROXHOLME, J., KERR, B., SALMON, A., OSTMAN-SMITH, I. & WATKINS, H. 2001. Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis. *Hum Mol Genet,* 10, 1215-20.

BOCZEK, N. J., BEST, J. M., TESTER, D. J., GIUDICESSI, J. R., MIDDHA, S., EVANS, J. M., KAMP, T. J. & ACKERMAN, M. J. 2013. Exome Sequencing and Systems Biology Converge to Identify Novel Mutations in the L-Type Calcium Channel, CACNA1C, Linked to Autosomal Dominant Long QT Syndrome. *Circ Cardiovasc Genet*.

BOEHM, T. M., DICKERSON, R. B. & GLASSER, S. P. 1970. Hypertrophic subaortic stenosis occurring in a patient with Friedreich's ataxia. *Am J Med Sci,* 260, 279-84.

BOEHNKE, M. 1994. Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *Am J Hum Genet,* 55, 379-90.

BONNEFOND, A., PHILIPPE, J., DURAND, E., DECHAUME, A., HUYVAERT, M., MONTAGNE, L., MARRE, M., BALKAU, B., FAJARDY, I., VAMBERGUE, A., VATIN, V., DELPLANQUE, J., LE GUILCHER, D., DE GRAEVE, F., LECOEUR, C., SAND, O., VAXILLAIRE, M. & FROGUEL, P. 2012. Whole-exome sequencing and high throughput genotyping identified KCNJ11 as the thirteenth MODY gene. *PLoS One,* 7, e37423.

BOS, J. M., POLEY, R. N., NY, M., TESTER, D. J., XU, X., VATTA, M., TOWBIN, J. A., GERSH, B. J., OMMEN, S. R. & ACKERMAN, M. J. 2006. Genotype-phenotype relationships involving hypertrophic cardiomyopathy-associated mutations in titin, muscle LIM protein, and telethonin. *Mol Genet Metab,* 88, 78-85.

BOYD, M. T., SEWARD, J. B., TAJIK, A. J. & EDWARDS, W. D. 1987. Frequency and location of prominent left ventricular trabeculations at autopsy in 474 normal human hearts: implications for evaluation of mural thrombi by two-dimensional echocardiography. *J Am Coll Cardiol,* 9, 323-6.

BRAUNWALD, E., LAMBREW, C. T., ROCKOFF, S. D., ROSS, J., JR. & MORROW, A. G. 1964. Idiopathic Hypertrophic Subaortic Stenosis. I. A Description of the Disease Based Upon an Analysis of 64 Patients. *Circulation,* 30, SUPPL 4:3-119.

BUSKE, O. J., MANICKARAJ, A., MITAL, S., RAY, P. N. & BRUDNO, M. 2013. Identification of deleterious synonymous variants in human genomes. *Bioinformatics*.

CAMERON, J. M., MAJ, M., LEVANDOVSKIY, V., BARNETT, C. P., BLASER, S., MACKAY, N., RAIMAN, J., FEIGENBAUM, A., SCHULZE, A. & ROBINSON, B. H. 2009. Pyruvate dehydrogenase phosphatase 1 (PDP1) null mutation produces a lethal infantile phenotype. *Hum Genet,* 125, 319-26.

CANDAU, R., MOORE, P. A., WANG, L., BARLEV, N., YING, C. Y., ROSEN, C. A. & BERGER, S. L. 1996. Identification of human proteins functionally conserved with the yeast putative adaptors ADA2 and GCN5. *Mol Cell Biol,* 16, 593-602.

CARLSON, C. S., NEWMAN, T. L. & NICKERSON, D. A. 2001. SNPing in the human genome. *Curr Opin Chem Biol,* 5, 78-85.

CARMIGNAC, V., SALIH, M. A., QUIJANO-ROY, S., MARCHAND, S., AL RAYESS, M. M., MUKHTAR, M. M., URTIZBEREA, J. A., LABEIT, S., GUICHENEY, P., LETURCQ, F., GAUTEL, M., FARDEAU, M., CAMPBELL, K. P., RICHARD, I., ESTOURNET, B. & FERREIRO, A. 2007. C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy. *Ann Neurol,* 61, 340-51.

CARNIEL, E., TAYLOR, M. R., SINAGRA, G., DI LENARDA, A., KU, L., FAIN, P. R., BOUCEK, M. M., CAVANAUGH, J., MIOCIC, S., SLAVOV, D., GRAW, S. L., FEIGER, J., ZHU, X. Z., DAO, D., FERGUSON, D. A., BRISTOW, M. R. & MESTRONI, L. 2005. Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation,* 112, 54-9.

CASALI, C., D'AMATI, G., BERNUCCI, P., DEBIASE, L., AUTORE, C., SANTORELLI, F. M., COVIELLO, D. & GALLO, P. 1999. Maternally inherited cardiomyopathy: clinical and molecular characterization of a large kindred harboring the A4300G point mutation in mitochondrial deoxyribonucleic acid. *J Am Coll Cardiol,* 33, 1584-9.

CASALI, C., SANTORELLI, F. M., D'AMATI, G., BERNUCCI, P., DEBIASE, L. & DIMAURO, S. 1995. A novel mtDNA point mutation in maternally inherited cardiomyopathy. *Biochem Biophys Res Commun,* 213, 588-93.

CENTNER, T., YANO, J., KIMURA, E., MCELHINNY, A. S., PELIN, K., WITT, C. C., BANG, M. L., TROMBITAS, K., GRANZIER, H., GREGORIO, C. C., SORIMACHI, H. & LABEIT, S. 2001. Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain. *J Mol Biol,* 306, 717-26.

CHAMKHA, I., MKAOUAR-REBAI, E., ALOULOU, H., CHABCHOUB, I., KIFAGI, C., FENDRI-KRIAA, N., KAMMOUN, T., HACHICHA, M. & FAKHFAKH, F. 2011. A novel m.3395A>G missense mutation in the mitochondrial ND1 gene associated with the new tRNA(Ile) m.4316A>G mutation in a patient with hypertrophic cardiomyopathy and profound hearing loss. *Biochem Biophys Res Commun,* 404, 504-10.

CHAMPION, K. M., COOK, R. J., TOLLAKSEN, S. L. & GIOMETTI, C. S. 1994. Identification of a heritable deficiency of the folate-dependent enzyme 10-formyltetrahydrofolate dehydrogenase in mice. *Proc Natl Acad Sci U S A,* 91, 11338-42.

CHARRON, P., VILLARD, E., SEBILLON, P., LAFORET, P., MAISONOBE, T., DUBOSCQ-BIDOT, L., ROMERO, N., DROUIN-GARRAUD, V., FREBOURG, T., RICHARD, P., EYMARD, B. & KOMAJDA, M. 2004. Danon's disease as a cause of hypertrophic cardiomyopathy: a systematic survey. *Heart,* 90, 842-6.

CHEN, X., NAKAYAMA, H., ZHANG, X., AI, X., HARRIS, D. M., TANG, M., ZHANG, H., SZETO, C., STOCKBOWER, K., BERRETTA, R. M., ECKHART, A. D., KOCH, W. J., MOLKENTIN, J. D. & HOUSER, S. R. 2011. Calcium influx through Cav1.2 is a proximal signal for pathological cardiomyocyte hypertrophy. *J Mol Cell Cardiol,* 50, 460-70.

CHEN, X. B., MELCHIONNA, R., DENIS, C. M., GAILLARD, P. H., BLASINA, A., VAN DE WEYER, I., BODDY, M. N., RUSSELL, P., VIALARD, J. & MCGOWAN, C. H. 2001. Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Mol Cell,* 8, 1117-27.

CHEN, Y. H. & YANG, J. T. 1971. A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. *Biochem Biophys Res Commun,* 44, 1285-91.

CHINNERY, P. F., JOHNSON, M. A., TAYLOR, R. W., DURWARD, W. F. & TURNBULL, D. M. 1997. A novel mitochondrial tRNA isoleucine gene mutation causing chronic progressive external ophthalmoplegia. *Neurology,* 49, 1166-8.

CHOI, W. S., KHURANA, A., MATHUR, R., VISWANATHAN, V., STEELE, D. F. & FEDIDA, D. 2005. Kv1.5 surface expression is modulated by retrograde trafficking of newly endocytosed channels by the dynein motor. *Circ Res,* 97, 363-71.

CHRISTEL, C. J., CARDONA, N., MESIRCA, P., HERRMANN, S., HOFMANN, F., STRIESSNIG, J., LUDWIG, A., MANGONI, M. E. & LEE, A. 2012. Distinct localization and modulation of Cav1.2 and Cav1.3 L-type Ca2+ channels in mouse sinoatrial node. *J Physiol,* 590, 6327-42.

CLARK, D. L., BOLER, D. D., KUTZLER, L. W., JONES, K. A., MCKEITH, F. K., KILLEFER, J., CARR, T. R. & DILGER, A. C. 2011a. Muscle gene expression associated with increased marbling in beef cattle. *Anim Biotechnol,* 22, 51-63.

CLARK, M. J., CHEN, R., LAM, H. Y., KARCZEWSKI, K. J., EUSKIRCHEN, G., BUTTE, A. J. & SNYDER, M. 2011b. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol,* 29, 908-14.

CLAYTON-SMITH, J., O'SULLIVAN, J., DALY, S., BHASKAR, S., DAY, R., ANDERSON, B., VOSS, A. K., THOMAS, T., BIESECKER, L. G., SMITH, P., FRYER, A., CHANDLER, K. E., KERR, B., TASSABEHJI, M., LYNCH, S. A., KRAJEWSKA-WALASEK, M., MCKEE, S., SMITH, J., SWEENEY, E., MANSOUR, S., MOHAMMED, S., DONNAI, D. & BLACK, G. 2011. Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. *Am J Hum Genet,* 89, 675-81.

CODD, M. B., SUGRUE, D. D., GERSH, B. J. & MELTON, L. J., 3RD 1989. Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted County, Minnesota, 1975-1984. *Circulation,* 80, 564-72.

COLLINS, F. S. 1995. Positional cloning moves from perditional to traditional. *Nat Genet,* 9, 347-50.
COLUCCI, W. S., LORELL, B. H., SCHOEN, F. J., WARHOL, M. J. & GROSSMAN, W. 1982. Hypertrophic obstructive cardiomyopathy due to Fabry's disease. *N Engl J Med,* 307, 926-8.
COOPER, D., KRAWCZAK, M. & ANTONORAKIS, S. 1995. The nature and mechanisms of human gene mutation. *In:* SCRIVER, C., BEAUDET., SLY, W. & VALLE, D. (eds.) *The Metabolic and Molecular Bases of Inherited Disease.* 7th ed.: McGraw-Hill.
COOPER, G. M., GOODE, D. L., NG, S. B., SIDOW, A., BAMSHAD, M. J., SHENDURE, J. & NICKERSON, D. A. 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods,* 7, 250-1.
COX, G. A., MAHAFFEY, C. L., NYSTUEN, A., LETTS, V. A. & FRANKEL, W. N. 2000. The mouse fidgetin gene defines a new role for AAA family proteins in mammalian development. *Nat Genet,* 26, 198-202.
CREE, L. M., SAMUELS, D. C., DE SOUSA LOPES, S. C., RAJASIMHA, H. K., WONNAPINIJ, P., MANN, J. R., DAHL, H. H. & CHINNERY, P. F. 2008. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat Genet,* 40, 249-54.
CURRAN, M. E., SPLAWSKI, I., TIMOTHY, K. W., VINCENT, G. M., GREEN, E. D. & KEATING, M. T. 1995. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell,* 80, 795-803.
DARVISHI, K. 2010. Application of Nexus copy number software for CNV detection and analysis. *Curr Protoc Hum Genet,* Chapter 4, Unit 4 14 1-28.
DAUBENEY, P. E., NUGENT, A. W., CHONDROS, P., CARLIN, J. B., COLAN, S. D., CHEUNG, M., DAVIS, A. M., CHOW, C. W. & WEINTRAUB, R. G. 2006. Clinical features and outcomes of childhood dilated cardiomyopathy: results from a national population-based study. *Circulation,* 114, 2671-8.
DE LIGT, J., WILLEMSEN, M. H., VAN BON, B. W., KLEEFSTRA, T., YNTEMA, H. G., KROES, T., VULTO-VAN SILFHOUT, A. T., KOOLEN, D. A., DE VRIES, P., GILISSEN, C., DEL ROSARIO, M., HOISCHEN, A., SCHEFFER, H., DE VRIES, B. B., BRUNNER, H. G., VELTMAN, J. A. & VISSERS, L. E. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med,* 367, 1921-9.
DEGOUL, F., BRULE, H., CEPANEC, C., HELM, M., MARSAC, C., LEROUX, J., GIEGE, R. & FLORENTZ, C. 1998. Isoleucylation properties of native human mitochondrial tRNAIle and tRNAIle transcripts. Implications for cardiomyopathy-related point mutations (4269, 4317) in the tRNAIle gene. *Hum Mol Genet,* 7, 347-54.
DELABIE, J., SHIPMAN, R., BRUGGEN, J., DE STROOPER, B., VAN LEUVEN, F., TARCSAY, L., CERLETTI, N., ODINK, K., DIEHL, V., BILBE, G. & ET AL. 1992. Expression of the novel intermediate filament-associated protein restin in Hodgkin's disease and anaplastic large-cell lymphoma. *Blood,* 80, 2891-6.
DETAINT, D., AEGERTER, P., TUBACH, F., HOFFMAN, I., PLAUCHU, H., DULAC, Y., FAIVRE, L. O., DELRUE, M. A., COLLIGNON, P., ODENT, S., TCHITCHINADZE, M., BOUFFARD, C., ARNOULT, F., GAUTIER, M., BOILEAU, C. & JONDEAU, G. 2010. Rationale and design of a randomized clinical trial (Marfan Sartan) of angiotensin II receptor blocker therapy versus placebo in individuals with Marfan syndrome. *Arch Cardiovasc Dis,* 103, 317-25.
DEWEY, F. E., PAN, S., WHEELER, M. T., QUAKE, S. R. & ASHLEY, E. A. 2012. DNA sequencing: clinical applications of new DNA sequencing technologies. *Circulation,* 125, 931-44.
DI PADOVA, M., BRUNO, T., DE NICOLA, F., IEZZI, S., D'ANGELO, C., GALLO, R., NICOSIA, D., CORBI, N., BIROCCIO, A., FLORIDI, A., PASSANANTI, C. & FANCIULLI, M. 2003. Che-1 arrests human colon carcinoma cell proliferation by displacing HDAC1 from the p21WAF1/CIP1 promoter. *J Biol Chem,* 278, 36496-504.
DOUGLAS, J., CILLIERS, D., COLEMAN, K., TATTON-BROWN, K., BARKER, K., BERNHARD, B., BURN, J., HUSON, S., JOSIFOVA, D., LACOMBE, D., MALIK, M., MANSOUR, S., REID, E., CORMIER-DAIRE, V., COLE, T. & RAHMAN, N. 2007. Mutations in RNF135, a gene within the NF1 microdeletion region, cause phenotypic abnormalities including overgrowth. *Nat Genet,* 39, 963-5.
DRITSAS, A., SBAROUNI, E., GILLIGAN, D., NIHOYANNOPOULOS, P. & OAKLEY, C. M. 1992. QT-interval abnormalities in hypertrophic cardiomyopathy. *Clin Cardiol,* 15, 739-42.
EDLICH, C. & MUHLE-GOLL, C. 2003. 1H, 13C, and 15N backbone assignment of the first two Ig domains Z1Z2 of the giant muscle protein Titin. *J Biomol NMR,* 27, 283-4.

EHLER, E., ROTHEN, B. M., HAMMERLE, S. P., KOMIYAMA, M. & PERRIARD, J. C. 1999. Myofibrillogenesis in the developing chicken heart: assembly of Z-disk, M-line and the thick filaments. *J Cell Sci,* 112 (Pt 10), 1529-39.

ELLIOTT, H. R., SAMUELS, D. C., EDEN, J. A., RELTON, C. L. & CHINNERY, P. F. 2008. Pathogenic mitochondrial DNA mutations are common in the general population. *Am J Hum Genet,* 83, 254-60.

ELLIOTT, P., BAKER, R., PASQUALE, F., QUARTA, G., EBRAHIM, H., MEHTA, A. B. & HUGHES, D. A. 2011. Prevalence of Anderson-Fabry disease in patients with hypertrophic cardiomyopathy: the European Anderson-Fabry Disease survey. *Heart,* 97, 1957-60.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66, 486-501.

EVS March 2013. NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [March, 2013]. http://evs.gs.washington.edu/EVS/.

EWING, B. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* 8, 186-94.

EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res,* 8, 175-85.

FABIAN, L., XIA, X., VENKITARAMANI, D. V., JOHANSEN, K. M., JOHANSEN, J., ANDREW, D. J. & FORER, A. 2007. Titin in insect spermatocyte spindle fibers associates with microtubules, actin, myosin and the matrix proteins skeletor, megator and chromator. *J Cell Sci,* 120, 2190-204.

FATKIN, D., MACRAE, C., SASAKI, T., WOLFF, M. R., PORCU, M., FRENNEAUX, M., ATHERTON, J., VIDAILLET, H. J., JR., SPUDICH, S., DE GIROLAMI, U., SEIDMAN, J. G., SEIDMAN, C., MUNTONI, F., MUEHLE, G., JOHNSON, W. & MCDONOUGH, B. 1999. Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. *N Engl J Med,* 341, 1715-24.

FAULKNER, G., LANFRANCHI, G. & VALLE, G. 2001. Telethonin and other new proteins of the Z-disc of skeletal muscle. *IUBMB Life,* 51, 275-82.

FAULKNER, G., PALLAVICINI, A., COMELLI, A., SALAMON, M., BORTOLETTO, G., IEVOLELLA, C., TREVISAN, S., KOJIC, S., DALLA VECCHIA, F., LAVEDER, P., VALLE, G. & LANFRANCHI, G. 2000. FATZ, a filamin-, actinin-, and telethonin-binding protein of the Z-disc of skeletal muscle. *J Biol Chem,* 275, 41234-42.

FERREIRA, Z., HURLE, B., ROCHA, J. & SEIXAS, S. 2011. Differing evolutionary histories of WFDC8 (short-term balancing) in Europeans and SPINT4 (incomplete selective sweep) in Africans. *Mol Biol Evol,* 28, 2811-22.

FERREIRO, A., MEZMEZIAN, M., OLIVE, M., HERLICOVIEZ, D., FARDEAU, M., RICHARD, P. & ROMERO, N. B. 2011. Telethonin-deficiency initially presenting as a congenital muscular dystrophy. *Neuromuscul Disord,* 21, 433-8.

FITZPATRICK, D. R., SKEOCH, C. H. & TOLMIE, J. L. 1991. Genetic aspects of admissions to a paediatric intensive care unit. *Arch Dis Child,* 66, 639-41.

FLAGG, T. P., ENKVETCHAKUL, D., KOSTER, J. C. & NICHOLS, C. G. 2010. Muscle KATP channels: recent insights to energy sensing and myoprotection. *Physiol Rev,* 90, 799-829.

FREIBURG, A. & GAUTEL, M. 1996. A molecular map of the interactions between titin and myosin-binding protein C. Implications for sarcomeric assembly in familial hypertrophic cardiomyopathy. *Eur J Biochem,* 235, 317-23.

FREY, N. & OLSON, E. N. 2002. Calsarcin-3, a novel skeletal muscle-specific member of the calsarcin family, interacts with multiple Z-disc proteins. *J Biol Chem,* 277, 13998-4004.

FRIDLIANSKAIA, II, GONCHAROVA, E. I., BORISOV, A. B., KRYLOVA, T. A. & PINAEV, G. P. 1989. [Monoclonal antibodies to the muscle isoform of alpha-actinin--a marker for the study of the differentiation of skeletal and cardiac muscles]. *Tsitologiia,* 31, 1234-7.

FUKUZAWA, A., LANGE, S., HOLT, M., VIHOLA, A., CARMIGNAC, V., FERREIRO, A., UDD, B. & GAUTEL, M. 2008. Interactions with titin and myomesin target obscurin and obscurin-like 1 to the M-band: implications for hereditary myopathies. *J Cell Sci,* 121, 1841-51.

FURST, D. O., OSBORN, M., NAVE, R. & WEBER, K. 1988. The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z line extends close to the M line. *J Cell Biol,* 106, 1563-72.

FURUKAWA, T., ONO, Y., TSUCHIYA, H., KATAYAMA, Y., BANG, M. L., LABEIT, D., LABEIT, S., INAGAKI, N. & GREGORIO, C. C. 2001. Specific interaction of the potassium channel beta-subunit minK with the sarcomeric protein T-cap suggests a T-tubule-myofibril linking system. *J Mol Biol,* 313, 775-84.

GABRIEL, S. E., BRIGMAN, K. N., KOLLER, B. H., BOUCHER, R. C. & STUTTS, M. J. 1994. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science,* 266, 107-9.

GAJECKA, M., MACKAY, K. L. & SHAFFER, L. G. 2007. Monosomy 1p36 deletion syndrome. *Am J Med Genet C Semin Med Genet,* 145C, 346-56.

GALLAGHER, P. G. & FORGET, B. G. 1998. An alternate promoter directs expression of a truncated, muscle-specific isoform of the human ankyrin 1 gene. *J Biol Chem,* 273, 1339-48.

GARCIA-GRAS, E., LOMBARDI, R., GIOCONDO, M. J., WILLERSON, J. T., SCHNEIDER, M. D., KHOURY, D. S. & MARIAN, A. J. 2006. Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhythmogenic right ventricular cardiomyopathy. *J Clin Invest,* 116, 2012-21.

GARCIA-MANYES, S., BADILLA, C. L., ALEGRE-CEBOLLADA, J., JAVADI, Y. & FERNANDEZ, J. M. 2012. Spontaneous dimerization of titin protein Z1Z2 domains induces strong nanomechanical anchoring. *J Biol Chem,* 287, 20240-7.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. R., APPEL, R. D. & BAIROCH, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. *In:* WALKER, J. M. (ed.) *The Proteomics Protocols Handbook.* Humana Press.

GAUTEL, M. 2011. The sarcomeric cytoskeleton: who picks up the strain? *Curr Opin Cell Biol,* 23, 39-46.

GAUTEL, M., GOULDING, D., BULLARD, B., WEBER, K. & FURST, D. O. 1996. The central Z-disk region of titin is assembled from a novel repeat in variable copy numbers. *J Cell Sci,* 109 (Pt 11), 2747-54.

GEIER, C., PERROT, A., OZCELIK, C., BINNER, P., COUNSELL, D., HOFFMANN, K., PILZ, B., MARTINIAK, Y., GEHMLICH, K., VAN DER VEN, P. F., FURST, D. O., VORNWALD, A., VON HODENBERG, E., NURNBERG, P., SCHEFFOLD, T., DIETZ, R. & OSTERZIEL, K. J. 2003. Mutations in the human muscle LIM protein gene in families with hypertrophic cardiomyopathy. *Circulation,* 107, 1390-5.

GEISTERFER-LOWRANCE, A. A., KASS, S., TANIGAWA, G., VOSBERG, H. P., MCKENNA, W., SEIDMAN, C. E. & SEIDMAN, J. G. 1990. A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell,* 62, 999-1006.

GERSH, B. J., MARON, B. J., BONOW, R. O., DEARANI, J. A., FIFER, M. A., LINK, M. S., NAIDU, S. S., NISHIMURA, R. A., OMMEN, S. R., RAKOWSKI, H., SEIDMAN, C. E., TOWBIN, J. A., UDELSON, J. E. & YANCY, C. W. 2011. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation,* 124, e783-831.

GERULL, B., ATHERTON, J., GEUPEL, A., SASSE-KLAASSEN, S., HEUSER, A., FRENNEAUX, M., MCNABB, M., GRANZIER, H., LABEIT, S. & THIERFELDER, L. 2006. Identification of a novel frameshift mutation in the giant muscle filament titin in a large Australian family with dilated cardiomyopathy. *J Mol Med (Berl),* 84, 478-83.

GERULL, B., GRAMLICH, M., ATHERTON, J., MCNABB, M., TROMBITAS, K., SASSE-KLAASSEN, S., SEIDMAN, J. G., SEIDMAN, C., GRANZIER, H., LABEIT, S., FRENNEAUX, M. & THIERFELDER, L. 2002. Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat Genet,* 30, 201-4.

GILISSEN, C., HOISCHEN, A., BRUNNER, H. G. & VELTMAN, J. A. 2011. Unlocking Mendelian disease using exome sequencing. *Genome Biol,* 12, 228.

GLESSNER, J. T., WANG, K., CAI, G., KORVATSKA, O., KIM, C. E., WOOD, S., ZHANG, H., ESTES, A., BRUNE, C. W., BRADFIELD, J. P., IMIELINSKI, M., FRACKELTON, E. C., REICHERT, J., CRAWFORD, E. L., MUNSON, J., SLEIMAN, P. M., CHIAVACCI, R., ANNAIAH, K., THOMAS, K., HOU, C., GLABERSON, W., FLORY, J., OTIENO, F., GARRIS, M., SOORYA, L., KLEI, L., PIVEN, J., MEYER, K. J., ANAGNOSTOU, E., SAKURAI, T., GAME, R. M., RUDD, D. S., ZURAWIECKI, D., MCDOUGLE, C. J., DAVIS, L. K., MILLER, J., POSEY, D. J., MICHAELS, S., KOLEVZON, A., SILVERMAN, J. M., BERNIER, R., LEVY, S. E., SCHULTZ, R. T., DAWSON, G., OWLEY, T.,

MCMAHON, W. M., WASSINK, T. H., SWEENEY, J. A., NURNBERGER, J. I., COON, H., SUTCLIFFE, J. S., MINSHEW, N. J., GRANT, S. F., BUCAN, M., COOK, E. H., BUXBAUM, J. D., DEVLIN, B., SCHELLENBERG, G. D. & HAKONARSON, H. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature,* 459, 569-73.

GOFFART, S., VON KLEIST-RETZOW, J.-C. & WIESNER, R. J. 2004. Regulation of mitochondrial proliferation in the heart: power-plant failure contributes to cardiac failure in hypertrophy. *Cardiovasc Res,* 64, 198-207.

GOLBUS, J. R., PUCKELWARTZ, M. J., FAHRENBACH, J. P., DELLEFAVE-CASTILLO, L. M., WOLFGEHER, D. & MCNALLY, E. M. 2012. Population-based variation in cardiomyopathy genes. *Circ Cardiovasc Genet,* 5, 391-9.

GOODSON, H. V., SKUBE, S. B., STALDER, R., VALETTI, C., KREIS, T. E., MORRISON, E. E. & SCHROER, T. A. 2003. CLIP-170 interacts with dynactin complex and the APC-binding protein EB1 by different mechanisms. *Cell Motil Cytoskeleton,* 55, 156-73.

GOTTHARDT, M., HAMMER, R. E., HUBNER, N., MONTI, J., WITT, C. C., MCNABB, M., RICHARDSON, J. A., GRANZIER, H., LABEIT, S. & HERZ, J. 2003. Conditional expression of mutant M-line titins results in cardiomyopathy with altered sarcomere structure. *J Biol Chem,* 278, 6059-65.

GOTZ, A., TYYNISMAA, H., EURO, L., ELLONEN, P., HYOTYLAINEN, T., OJALA, T., HAMALAINEN, R. H., TOMMISKA, J., RAIVIO, T., ORESIC, M., KARIKOSKI, R., TAMMELA, O., SIMOLA, K. O. J., PAETAU, A., TYNI, T. & SUOMALAINEN, A. 2011. Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am J Hum Genet,* 88, 635-42.

GRANZIER, H. L. & LABEIT, S. 2004. The giant protein titin: a major player in myocardial mechanics, signaling, and disease. *Circ Res,* 94, 284-95.

GREEN, R., BERG, J., GRODY, W., KALIA, S., KORF, B., MARTIN, C., MCGUIRE, A., NUSSBAUM, R., O'DANIEL, J., ORMOND, K., REHM, H., WATSON, M., WILLIAMS, M. & BIESECKER, L. 2013. ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing.

GREENFIELD, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc,* 1, 2876-90.

GREGORIO, C. C., TROMBITAS, K., CENTNER, T., KOLMERER, B., STIER, G., KUNKE, K., SUZUKI, K., OBERMAYR, F., HERRMANN, B., GRANZIER, H., SORIMACHI, H. & LABEIT, S. 1998. The NH2 terminus of titin spans the Z-disc: its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. *J Cell Biol,* 143, 1013-27.

GROVE, B. K., KURER, V., LEHNER, C., DOETSCHMAN, T. C., PERRIARD, J. C. & EPPENBERGER, H. M. 1984. A new 185,000-dalton skeletal muscle protein detected by monoclonal antibodies. *J Cell Biol,* 98, 518-24.

GUNDERSON, K. L., STEEMERS, F. J., REN, H., NG, P., ZHOU, L., TSAN, C., CHANG, W., BULLIS, D., MUSMACKER, J., KING, C., LEBRUSKA, L. L., BARKER, D., OLIPHANT, A., KUHN, K. M. & SHEN, R. 2006. Whole-genome genotyping. *Methods Enzymol,* 410, 359-76.

GUSELLA, J. F., WEXLER, N. S., CONNEALLY, P. M., NAYLOR, S. L., ANDERSON, M. A., TANZI, R. E., WATKINS, P. C., OTTINA, K., WALLACE, M. R., SAKAGUCHI, A. Y. & ET AL. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature,* 306, 234-8.

GUTIERREZ CORTES, N., PERTUISET, C., DUMON, E., BORLIN, M., HEBERT-CHATELAIN, E., PIERRON, D., FELDMANN, D., JONARD, L., MARLIN, S., LETELLIER, T. & ROCHER, C. 2012. Novel mitochondrial DNA mutations responsible for maternally inherited nonsyndromic hearing loss. *Hum Mutat,* 33, 681-9.

HAACK, T. B., DANHAUSER, K., HABERBERGER, B., HOSER, J., STRECKER, V., BOEHM, D., UZIEL, G., LAMANTEA, E., INVERNIZZI, F., POULTON, J., ROLINSKI, B., IUSO, A., BISKUP, S., SCHMIDT, T., MEWES, H. W., WITTIG, I., MEITINGER, T., ZEVIANI, M. & PROKISCH, H. 2010. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat Genet,* 42, 1131-4.

HACKMAN, P., MARCHAND, S., SARPARANTA, J., VIHOLA, A., PENISSON-BESNIER, I., EYMARD, B., PARDAL-FERNANDEZ, J. M., HAMMOUDA EL, H., RICHARD, I., ILLA, I. & UDD, B. 2008. Truncating mutations in C-terminal titin may cause more severe tibial muscular dystrophy (TMD). *Neuromuscul Disord,* 18, 922-8.

HACKMAN, P., VIHOLA, A., HARAVUORI, H., MARCHAND, S., SARPARANTA, J., DE SEZE, J., LABEIT, S., WITT, C., PELTONEN, L., RICHARD, I. & UDD, B. 2002. Tibial muscular dystrophy is a

titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *Am J Hum Genet,* 71, 492-500.

HAISSAGUERRE, M., CHATEL, S., SACHER, F., WEERASOORIYA, R., PROBST, V., LOUSSOUARN, G., HORLITZ, M., LIERSCH, R., SCHULZE-BAHR, E., WILDE, A., KAAB, S., KOSTER, J., RUDY, Y., LE MAREC, H. & SCHOTT, J. J. 2009. Ventricular fibrillation with prominent early repolarization associated with a rare variant of KCNJ8/KATP channel. *J Cardiovasc Electrophysiol,* 20, 93-8.

HAJI-SEYED-JAVADI, R., JELODARI-MAMAGHANI, S., PAYLAKHI, S. H., YAZDANI, S., NILFORUSHAN, N., FAN, J. B., KLOTZLE, B., MAHMOUDI, M. J., EBRAHIMIAN, M. J., CHELICH, N., TAGHIABADI, E., KAMYAB, K., BOILEAU, C., PAISAN-RUIZ, C., RONAGHI, M. & ELAHI, E. 2012. LTBP2 mutations cause Weill-Marchesani and Weill-Marchesani-like syndrome and affect disruptions in the extracellular matrix. *Hum Mutat,* 33, 1182-7.

HANCK, T., STRICKER, R., SEDEHIZADE, F. & REISER, G. 2004. Identification of gene structure and subcellular localization of human centaurin alpha 2, and p42IP4, a family of two highly homologous, Ins 1,3,4,5-P4-/PtdIns 3,4,5-P3-binding, adapter proteins. *J Neurochem,* 88, 326-36.

HAYASHI, T., ARIMURA, T., ITOH-SATOH, M., UEDA, K., HOHDA, S., INAGAKI, N., TAKAHASHI, M., HORI, H., YASUNAMI, M., NISHI, H., KOGA, Y., NAKAMURA, H., MATSUZAKI, M., CHOI, B. Y., BAE, S. W., YOU, C. W., HAN, K. H., PARK, J. E., KNOLL, R., HOSHIJIMA, M., CHIEN, K. R. & KIMURA, A. 2004. Tcap gene mutations in hypertrophic cardiomyopathy and dilated cardiomyopathy. *J Am Coll Cardiol,* 44, 2192-201.

HEIKKINEN, O. K., RUSKAMO, S., KONAREV, P. V., SVERGUN, D. I., IIVANAINEN, T., HEIKKINEN, S. M., PERMI, P., KOSKELA, H., KILPELAINEN, I. & YLANNE, J. 2009. Atomic structures of two novel immunoglobulin-like domain pairs in the actin cross-linking protein filamin. *J Biol Chem,* 284, 25450-8.

HEINEKE, J. & MOLKENTIN, J. D. 2006. Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat Rev Mol Cell Biol,* 7, 589-600.

HERMAN, D. S., LAM, L., TAYLOR, M. R., WANG, L., TEEKAKIRIKUL, P., CHRISTODOULOU, D., CONNER, L., DEPALMA, S. R., MCDONOUGH, B., SPARKS, E., TEODORESCU, D. L., CIRINO, A. L., BANNER, N. R., PENNELL, D. J., GRAW, S., MERLO, M., DI LENARDA, A., SINAGRA, G., BOS, J. M., ACKERMAN, M. J., MITCHELL, R. N., MURRY, C. E., LAKDAWALA, N. K., HO, C. Y., BARTON, P. J., COOK, S. A., MESTRONI, L., SEIDMAN, J. G. & SEIDMAN, C. E. 2012. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med,* 366, 619-28.

HERSHBERGER, R. E., PARKS, S. B., KUSHNER, J. D., LI, D., LUDWIGSEN, S., JAKOBS, P., NAUMAN, D., BURGESS, D., PARTAIN, J. & LITT, M. 2008. Coding sequence mutations identified in MYH7, TNNT2, SCN5A, CSRP3, LBD3, and TCAP from 313 patients with familial or idiopathic dilated cardiomyopathy. *Clin Transl Sci,* 1, 21-6.

HEUBERGER, J. & BIRCHMEIER, W. 2010. Interplay of cadherin-mediated cell adhesion and canonical Wnt signaling. *Cold Spring Harb Perspect Biol,* 2, a002915.

HIDALGO, C., HUDSON, B., BOGOMOLOVAS, J., ZHU, Y., ANDERSON, B., GREASER, M., LABEIT, S. & GRANZIER, H. 2009. PKC phosphorylation of titin's PEVK element: a novel and conserved pathway for modulating myocardial stiffness. *Circ Res,* 105, 631-8, 17 p following 638.

HOEDEMAEKERS, Y. M., CALISKAN, K., MICHELS, M., FROHN-MULDER, I., VAN DER SMAGT, J. J., PHEFFERKORN, J. E., WESSELS, M. W., TEN CATE, F. J., SIJBRANDS, E. J., DOOIJES, D. & MAJOOR-KRAKAUER, D. F. 2010. The importance of genetic counseling, DNA diagnostics, and cardiologic family screening in left ventricular noncompaction cardiomyopathy. *Circ Cardiovasc Genet,* 3, 232-9.

HOLLMAN, A., GOODWIN, J. F., TEARE, D. & RENWICK, J. W. 1960. A family with obstructive cardiomyopathy (asymmetrical hypertrophy). *Br Heart J,* 22, 449-56.

HONG, M., LEE, Y., KIM, J. W., LIM, J. S., CHANG, S. Y., LEE, K. S., PAIK, S. G. & CHOE, I. S. 1999. Isolation and characterization of cDNA clone for human liver 10-formyltetrahydrofolate dehydrogenase. *Biochem Mol Biol Int,* 47, 407-15.

HORIKAWA, Y., IWASAKI, N., HARA, M., FURUTA, H., HINOKIO, Y., COCKBURN, B. N., LINDNER, T., YAMAGATA, K., OGATA, M., TOMONAGA, O., KUROKI, H., KASAHARA, T., IWAMOTO, Y. & BELL, G. I. 1997. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet,* 17, 384-5.

HSU, C. C., CHIANG, C. W., CHENG, H. C., CHANG, W. T., CHOU, C. Y., TSAI, H. W., LEE, C. T., WU, Z. H., LEE, T. Y., CHAO, A., CHOW, N. H. & HO, C. L. 2011. Identifying LRRC16B as an oncofetal

gene with transforming enhancing capability using a combined bioinformatics and experimental approach. *Oncogene,* 30, 654-67.

HUSAR BIOINFORMATICS LAB, D. O. M. B., GERMAN CANCER RESEARCH CENTER, DKFZ, HEIDELBERG April 2013. Tissue Distribution Databases. http://genome.dkfz-heidelberg.de/menu/tissue_db/index.html.

INGMAN, M. & GYLLENSTEN, U. 2006. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res,* 34, D749-51.

ISTRAIL, S., SUTTON, G. G., FLOREA, L., HALPERN, A. L., MOBARRY, C. M., LIPPERT, R., WALENZ, B., SHATKAY, H., DEW, I., MILLER, J. R., FLANIGAN, M. J., EDWARDS, N. J., BOLANOS, R., FASULO, D., HALLDORSSON, B. V., HANNENHALLI, S., TURNER, R., YOOSEPH, S., LU, F., NUSSKERN, D. R., SHUE, B. C., ZHENG, X. H., ZHONG, F., DELCHER, A. L., HUSON, D. H., KRAVITZ, S. A., MOUCHARD, L., REINERT, K., REMINGTON, K. A., CLARK, A. G., WATERMAN, M. S., EICHLER, E. E., ADAMS, M. D., HUNKAPILLER, M. W., MYERS, E. W. & VENTER, J. C. 2004. Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc Natl Acad Sci U S A,* 101, 1916-21.

ITOH-SATOH, M., HAYASHI, T., NISHI, H., KOGA, Y., ARIMURA, T., KOYANAGI, T., TAKAHASHI, M., HOHDA, S., UEDA, K., NOUCHI, T., HIROE, M., MARUMO, F., IMAIZUMI, T., YASUNAMI, M. & KIMURA, A. 2002. Titin mutations as the molecular basis for dilated cardiomyopathy. *Biochem Biophys Res Commun,* 291, 385-93.

JAMAL, S. M., YU, J. H., CHONG, J. X., DENT, K. M., CONTA, J. H., TABOR, H. K. & BAMSHAD, M. J. 2013. Practices and Policies of Clinical Exome Sequencing Providers: Analysis and Implications. *Am J Med Genet A*.

JIANG, D., WANG, R., XIAO, B., KONG, H., HUNT, D. J., CHOI, P., ZHANG, L. & CHEN, S. R. 2005. Enhanced store overload-induced Ca2+ release and channel sensitivity to luminal Ca2+ activation are common defects of RyR2 mutations linked to ventricular tachycardia and sudden death. *Circ Res,* 97, 1173-81.

JOHNSON, G. C., ESPOSITO, L., BARRATT, B. J., SMITH, A. N., HEWARD, J., DI GENOVA, G., UEDA, H., CORDELL, H. J., EAVES, I. A., DUDBRIDGE, F., TWELLS, R. C., PAYNE, F., HUGHES, W., NUTLAND, S., STEVENS, H., CARR, P., TUOMILEHTO-WOLF, E., TUOMILEHTO, J., GOUGH, S. C., CLAYTON, D. G. & TODD, J. A. 2001. Haplotype tagging for the identification of common disease genes. *Nat Genet,* 29, 233-7.

KANIA, A., JOHNSON, R. L. & JESSELL, T. M. 2000. Coordinate roles for LIM homeobox genes in directing the dorsoventral trajectory of motor axons in the vertebrate limb. *Cell,* 102, 161-73.

KAPSA, R., KORNBERG, A. J. & BYRNE, E. 2003. Novel therapies for Duchenne muscular dystrophy. *Lancet Neurol,* 2, 299-310.

KARKKAINEN, S. & PEUHKURINEN, K. 2007. Genetics of dilated cardiomyopathy. *Ann Med,* 39, 91-107.

KAUFERSTEIN, S., KIEHNE, N., ERKAPIC, D., SCHMIDT, J., HAMM, C. W., BRATZKE, H., PITSCHNER, H. F., KUNISS, M. & NEUMANN, T. 2011. A novel mutation in the cardiac ryanodine receptor gene (RyR2) in a patient with an unequivocal LQTS. *Int J Cardiol,* 146, 249-50.

KELLEY, S. O., STEINBERG, S. V. & SCHIMMEL, P. 2001. Fragile T-stem in disease-associated human mitochondrial tRNA sensitizes structure to local and distant mutations. *J Biol Chem,* 276, 10607-11.

KEREN, A., SYRRIS, P. & MCKENNA, W. J. 2008. Hypertrophic cardiomyopathy: the genetic determinants of clinical disease expression. *Nat Clin Pract Cardiovasc Med,* 5, 158-68.

KIRBY, A., GNIRKE, A., JAFFE, D. B., BARESOVA, V., POCHET, N., BLUMENSTIEL, B., YE, C., AIRD, D., STEVENS, C., ROBINSON, J. T., CABILI, M. N., GAT-VIKS, I., KELLIHER, E., DAZA, R., DEFELICE, M., HULKOVA, H., SOVOVA, J., VYLET'AL, P., ANTIGNAC, C., GUTTMAN, M., HANDSAKER, R. E., PERRIN, D., STEELMAN, S., SIGURDSSON, S., SCHEINMAN, S. J., SOUGNEZ, C., CIBULSKIS, K., PARKIN, M., GREEN, T., ROSSIN, E., ZODY, M. C., XAVIER, R. J., POLLAK, M. R., ALPER, S. L., LINDBLAD-TOH, K., GABRIEL, S., HART, P. S., REGEV, A., NUSBAUM, C., KMOCH, S., BLEYER, A. J., LANDER, E. S. & DALY, M. J. 2013. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet,* 45, 299-303.

KLAASSEN, S., PROBST, S., OECHSLIN, E., GERULL, B., KRINGS, G., SCHULER, P., GREUTMANN, M., HURLIMANN, D., YEGITBASI, M., PONS, L., GRAMLICH, M., DRENCKHAHN, J. D., HEUSER, A.,

BERGER, F., JENNI, R. & THIERFELDER, L. 2008. Mutations in sarcomere protein genes in left ventricular noncompaction. *Circulation,* 117, 2893-901.

KLEY, R. A., HELLENBROICH, Y., VAN DER VEN, P. F., FURST, D. O., HUEBNER, A., BRUCHERTSEIFER, V., PETERS, S. A., HEYER, C. M., KIRSCHNER, J., SCHRODER, R., FISCHER, D., MULLER, K., TOLKSDORF, K., EGER, K., GERMING, A., BRODHERR, T., REUM, C., WALTER, M. C., LOCHMULLER, H., KETELSEN, U. P. & VORGERD, M. 2007. Clinical and morphological phenotype of the filamin myopathy: a study of 31 German patients. *Brain,* 130, 3250-64.

KLOSS-BRANDSTATTER, A., PACHER, D., SCHONHERR, S., WEISSENSTEINER, H., BINNA, R., SPECHT, G. & KRONENBERG, F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat,* 32, 25-32.

KLUWE, L., SIEBERT, R., GESK, S., FRIEDRICH, R. E., TINSCHERT, S., KEHRER-SAWATZKI, H. & MAUTNER, V. F. 2004. Screening 500 unselected neurofibromatosis 1 patients for deletions of the NF1 gene. *Hum Mutat,* 23, 111-6.

KNOLL, R., BUYANDELGER, B. & LAB, M. 2011a. The sarcomeric Z-disc and Z-discopathies. *J Biomed Biotechnol,* 2011, 569628.

KNOLL, R., HOSHIJIMA, M., HOFFMAN, H. M., PERSON, V., LORENZEN-SCHMIDT, I., BANG, M. L., HAYASHI, T., SHIGA, N., YASUKAWA, H., SCHAPER, W., MCKENNA, W., YOKOYAMA, M., SCHORK, N. J., OMENS, J. H., MCCULLOCH, A. D., KIMURA, A., GREGORIO, C. C., POLLER, W., SCHAPER, J., SCHULTHEISS, H. P. & CHIEN, K. R. 2002. The cardiac mechanical stretch sensor machinery involves a Z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell,* 111, 943-55.

KNOLL, R., LINKE, W. A., ZOU, P., MIOCIC, S., KOSTIN, S., BUYANDELGER, B., KU, C. H., NEEF, S., BUG, M., SCHAFER, K., KNOLL, G., FELKIN, L. E., WESSELS, J., TOISCHER, K., HAGN, F., KESSLER, H., DIDIE, M., QUENTIN, T., MAIER, L. S., TEUCHER, N., UNSOLD, B., SCHMIDT, A., BIRKS, E. J., GUNKEL, S., LANG, P., GRANZIER, H., ZIMMERMANN, W. H., FIELD, L. J., FAULKNER, G., DOBBELSTEIN, M., BARTON, P. J., SATTLER, M., WILMANNS, M. & CHIEN, K. R. 2011b. Telethonin deficiency is associated with maladaptation to biomechanical stress in the mammalian heart. *Circ Res,* 109, 758-69.

KOGELNIK, A. M., LOTT, M. T., BROWN, M. D., NAVATHE, S. B. & WALLACE, D. C. 1996. MITOMAP: a human mitochondrial genome database. *Nucleic Acids Res,* 24, 177-9.

KOLLAR, V., SZATMARI, D., GRAMA, L. & KELLERMAYER, M. S. 2010. Dynamic strength of titin's Z-disk end. *J Biomed Biotechnol,* 2010, 838530.

KONG, Y., SHELTON, J. M., ROTHERMEL, B., LI, X., RICHARDSON, J. A., BASSEL-DUBY, R. & WILLIAMS, R. S. 2001. Cardiac-specific LIM protein FHL2 modifies the hypertrophic response to beta-adrenergic stimulation. *Circulation,* 103, 2731-8.

KONTROGIANNI-KONSTANTOPOULOS, A. & BLOCH, R. J. 2003. The hydrophilic domain of small ankyrin-1 interacts with the two N-terminal immunoglobulin domains of titin. *J Biol Chem,* 278, 3985-91.

KOO, B. K., SPIT, M., JORDENS, I., LOW, T. Y., STANGE, D. E., VAN DE WETERING, M., VAN ES, J. H., MOHAMMED, S., HECK, A. J., MAURICE, M. M. & CLEVERS, H. 2012. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature,* 488, 665-9.

KOZAREWA, I., NING, Z., QUAIL, M. A., SANDERS, M. J., BERRIMAN, M. & TURNER, D. J. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods,* 6, 291-5.

KRUGER, M., KOTTER, S., GRUTZNER, A., LANG, P., ANDRESEN, C., REDFIELD, M. M., BUTT, E., DOS REMEDIOS, C. G. & LINKE, W. A. 2009. Protein kinase G modulates human myocardial passive stiffness by phosphorylation of the titin springs. *Circ Res,* 104, 87-94.

KRUGER, M. & LINKE, W. A. 2011. The giant protein titin: a regulatory node that integrates myocyte signaling pathways. *J Biol Chem,* 286, 9905-12.

KU, C. S., NAIDOO, N. & PAWITAN, Y. 2011. Revisiting Mendelian disorders through exome sequencing. *Hum Genet,* 129, 351-70.

KUGLER, J. D. 1991. Sinus nodal dysfunction in young patients with long QT syndrome. *Am Heart J,* 121, 1132-6.

KUMAR, A., DUVVARI, M. R., PRABHAKARAN, V. C., SHETTY, J. S., MURTHY, G. J. & BLANTON, S. H. 2010. A homozygous mutation in LTBP2 causes isolated microspherophakia. *Hum Genet,* 128, 365-71.

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.

LABEIT, S., GAUTEL, M., LAKEY, A. & TRINICK, J. 1992. Towards a molecular understanding of titin. *EMBO J,* 11, 1711-6.

LABEIT, S., LAHMERS, S., BURKART, C., FONG, C., MCNABB, M., WITT, S., WITT, C., LABEIT, D. & GRANZIER, H. 2006. Expression of distinct classes of titin isoforms in striated and smooth muscles by alternative splicing, and their conserved interaction with filamins. *J Mol Biol,* 362, 664-81.

LAHMERS, S., WU, Y., CALL, D. R., LABEIT, S. & GRANZIER, H. 2004. Developmental control of titin isoform expression and passive stiffness in fetal and neonatal myocardium. *Circ Res,* 94, 505-13.

LAITINEN, P. J., BROWN, K. M., PIIPPO, K., SWAN, H., DEVANEY, J. M., BRAHMBHATT, B., DONARUM, E. A., MARINO, M., TISO, N., VIITASALO, M., TOIVONEN, L., STEPHAN, D. A. & KONTULA, K. 2001. Mutations of the cardiac ryanodine receptor (RyR2) gene in familial polymorphic ventricular tachycardia. *Circulation,* 103, 485-90.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409, 860-921.

LANDSTROM, A. P. & ACKERMAN, M. J. 2012. Beyond the cardiac myofilament: hypertrophic cardiomyopathy- associated mutations in genes that encode calcium-handling proteins. *Curr Mol Med,* 12, 507-18.

LANGE, S., AUERBACH, D., MCLOUGHLIN, P., PERRIARD, E., SCHAFER, B. W., PERRIARD, J. C. & EHLER, E. 2002. Subcellular targeting of metabolic enzymes to titin in heart muscle may be mediated by DRAL/FHL-2. *J Cell Sci,* 115, 4925-36.

LANGE, S., EHLER, E. & GAUTEL, M. 2006. From A to Z and back? Multicompartment proteins in the sarcomere. *Trends Cell Biol,* 16, 11-8.

LANGE, S., XIANG, F., YAKOVENKO, A., VIHOLA, A., HACKMAN, P., ROSTKOVA, E., KRISTENSEN, J., BRANDMEIER, B., FRANZEN, G., HEDBERG, B., GUNNARSSON, L. G., HUGHES, S. M., MARCHAND, S., SEJERSEN, T., RICHARD, I., EDSTROM, L., EHLER, E., UDD, B. & GAUTEL, M. 2005. The kinase domain of titin controls muscle gene expression and protein turnover. *Science,* 308, 1599-603.

LEE, E. H., GAO, M., PINOTSIS, N., WILMANNS, M. & SCHULTEN, K. 2006. Mechanical strength of the titin Z1Z2-telethonin complex. *Structure,* 14, 497-509.

LEE, E. H., HSIN, J., MAYANS, O. & SCHULTEN, K. 2007. Secondary and tertiary structure elasticity of titin Z1Z2 and a titin chain model. *Biophys J,* 93, 1719-35.

LEE, J. W., CHOI, H. S., GYURIS, J., BRENT, R. & MOORE, D. D. 1995. Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. *Mol Endocrinol,* 9, 243-54.

LEGROS, F., MALKA, F., FRACHON, P., LOMBES, A. & ROJO, M. 2004. Organization and dynamics of human mitochondrial DNA. *J Cell Sci,* 117, 2653-62.

LEWINTER, M. M., WU, Y., LABEIT, S. & GRANZIER, H. 2007. Cardiac titin: structure, functions and role in disease. *Clin Chim Acta,* 375, 1-9.

LI, H. & HOMER, N. 2010. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform,* 11, 473-83.

LI, M. X., GUI, H. S., KWAN, J. S., BAO, S. Y. & SHAM, P. C. 2012. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res,* 40, e53.

LI, M. X., KWAN, J. S., BAO, S. Y., YANG, W., HO, S. L., SONG, Y. Q. & SHAM, P. C. 2013. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet,* 9, e1003143.

LIANG, S., LI, Y., BE, X., HOWES, S. & LIU, W. 2006. Detecting and profiling tissue-selective genes. *Physiol Genomics,* 26, 158-62.

LINKE, W. A. 2008. Sense and stretchability: the role of titin and titin-associated proteins in myocardial stress-sensing and mechanical dysfunction. *Cardiovasc Res,* 77, 637-48.

LINKE, W. A., KULKE, M., LI, H., FUJITA-BECKER, S., NEAGOE, C., MANSTEIN, D. J., GAUTEL, M. & FERNANDEZ, J. M. 2002. PEVK domain of titin: an entropic spring with actin-binding properties. *J Struct Biol,* 137, 194-205.

LISE, S., CLARKSON, Y., PERKINS, E., KWASNIEWSKA, A., SADIGHI AKHA, E., SCHNEKENBERG, R. P., SUMINAITE, D., HOPE, J., BAKER, I., GREGORY, L., GREEN, A., ALLAN, C., LAMBLE, S., JAYAWANT, S., QUAGHEBEUR, G., CADER, M. Z., HUGHES, S., ARMSTRONG, R. J., KANAPIN, A., RIMMER, A., LUNTER, G., MATHIESON, I., CAZIER, J. B., BUCK, D., TAYLOR, J. C., BENTLEY, D., MCVEAN, G., DONNELLY, P., KNIGHT, S. J., JACKSON, M., RAGOUSSIS, J. & NEMETH, A. H. 2012. Recessive mutations in SPTBN2 implicate beta-III spectrin in both cognitive and motor development. *PLoS Genet,* 8, e1003074.

LIU, X., YU, X., ZACK, D. J., ZHU, H. & QIAN, J. 2008. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics,* 9, 271.

LLANO, E., PENDAS, A. M., KNAUPER, V., SORSA, T., SALO, T., SALIDO, E., MURPHY, G., SIMMER, J. P., BARTLETT, J. D. & LOPEZ-OTIN, C. 1997. Identification and structural and functional characterization of human enamelysin (MMP-20). *Biochemistry,* 36, 15101-8.

LO, A. N. S. M., WILDE, A. A., VAN ERVEN, L. & BLOM, N. A. 2005. Syndactyly and long QT syndrome (CaV1.2 missense mutation G406R) is associated with hypertrophic cardiomyopathy. *Heart Rhythm,* 2, 1365-8.

LUK, A., AHN, E., SOOR, G. S. & BUTANY, J. 2009. Dilated cardiomyopathy: a review. *J Clin Pathol,* 62, 219-25.

LUNTER, G. & GOODSON, M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res,* 21, 936-9.

LUXAN, G., CASANOVA, J. C., MARTINEZ-POVEDA, B., PRADOS, B., D'AMATO, G., MACGROGAN, D., GONZALEZ-RAJAL, A., DOBARRO, D., TORROJA, C., MARTINEZ, F., IZQUIERDO-GARCIA, J. L., FERNANDEZ-FRIERA, L., SABATER-MOLINA, M., KONG, Y. Y., PIZARRO, G., IBANEZ, B., MEDRANO, C., GARCIA-PAVIA, P., GIMENO, J. R., MONSERRAT, L., JIMENEZ-BORREGUERO, L. J. & DE LA POMPA, J. L. 2013. Mutations in the NOTCH pathway regulator MIB1 cause left ventricular noncompaction cardiomyopathy. *Nat Med*.

MACARTHUR, D. G., BALASUBRAMANIAN, S., FRANKISH, A., HUANG, N., MORRIS, J., WALTER, K., JOSTINS, L., HABEGGER, L., PICKRELL, J. K., MONTGOMERY, S. B., ALBERS, C. A., ZHANG, Z. D., CONRAD, D. F., LUNTER, G., ZHENG, H., AYUB, Q., DEPRISTO, M. A., BANKS, E., HU, M., HANDSAKER, R. E., ROSENFELD, J. A., FROMER, M., JIN, M., MU, X. J., KHURANA, E., YE, K., KAY, M., SAUNDERS, G. I., SUNER, M. M., HUNT, T., BARNES, I. H., AMID, C., CARVALHO-SILVA, D. R., BIGNELL, A. H., SNOW, C., YNGVADOTTIR, B., BUMPSTEAD, S., COOPER, D. N., XUE, Y., ROMERO, I. G., WANG, J., LI, Y., GIBBS, R. A., MCCARROLL, S. A., DERMITZAKIS, E. T., PRITCHARD, J. K., BARRETT, J. C., HARROW, J., HURLES, M. E., GERSTEIN, M. B. & TYLER-SMITH, C. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science,* 335, 823-8.

MACARTHUR, D. G. & TYLER-SMITH, C. 2010. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet,* 19, R125-30.

MACHADO, C. & ANDREW, D. J. 2000. D-Titin: a giant protein with dual roles in chromosomes and muscles. *J Cell Biol,* 151, 639-52.

MAHER, B. 2009. Exome sequencing takes centre stage in cancer profiling. *Nature,* 459, 146-7.

MAHJOUB, S., STERNBERG, D., BOUSSAADA, R., FILAUT, S., GMIRA, F., MECHMECH, R., JARDEL, C. & ARAB, S. B. 2007. A novel mitochondrial DNA tRNAIle (m.4322dupC) mutation associated with idiopathic dilated cardiomyopathy. *Diagn Mol Pathol,* 16, 238-42.

MAKARENKO, I., OPITZ, C. A., LEAKE, M. C., NEAGOE, C., KULKE, M., GWATHMEY, J. K., DEL MONTE, F., HAJJAR, R. J. & LINKE, W. A. 2004. Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. *Circ Res,* 95, 708-16.

MANNING, G., WHYTE, D. B., MARTINEZ, R., HUNTER, T. & SUDARSANAM, S. 2002. The protein kinase complement of the human genome. *Science,* 298, 1912-34.

MARCHESE, A., DOCHERTY, J. M., NGUYEN, T., HEIBER, M., CHENG, R., HENG, H. H., TSUI, L. C., SHI, X., GEORGE, S. R. & O'DOWD, B. F. 1994. Cloning of human genes encoding novel G protein-coupled receptors. *Genomics,* 23, 609-18.

MARCUS, F. I., MCKENNA, W. J., SHERRILL, D., BASSO, C., BAUCE, B., BLUEMKE, D. A., CALKINS, H., CORRADO, D., COX, M. G., DAUBERT, J. P., FONTAINE, G., GEAR, K., HAUER, R., NAVA, A., PICARD, M. H., PROTONOTARIOS, N., SAFFITZ, J. E., SANBORN, D. M., STEINBERG, J. S., TANDRI, H., THIENE, G., TOWBIN, J. A., TSATSOPOULOU, A., WICHTER, T. & ZAREBA, W. 2010. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Circulation,* 121, 1533-41.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437, 376-80.

MARINO, M., ZOU, P., SVERGUN, D., GARCIA, P., EDLICH, C., SIMON, B., WILMANNS, M., MUHLE-GOLL, C. & MAYANS, O. 2006. The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. *Structure,* 14, 1437-47.

MARKERT, C. D., MEANEY, M. P., VOELKER, K. A., GRANGE, R. W., DALLEY, H. W., CANN, J. K., AHMED, M., BISHWOKARMA, B., WALKER, S. J., YU, S. X., BROWN, M., LAWLOR, M. W., BEGGS, A. H. & CHILDERS, M. K. 2010. Functional muscle analysis of the Tcap knockout mouse. *Hum Mol Genet,* 19, 2268-83.

MARKERT, C. D., NING, J., STALEY, J. T., HEINZKE, L., CHILDERS, C. K., FERREIRA, J. A., BROWN, M., STOKER, A., OKAMURA, C. & CHILDERS, M. K. 2008. TCAP knockdown by RNA interference inhibits myoblast differentiation in cultured skeletal muscle cells. *Neuromuscul Disord,* 18, 413-22.

MARON, B. J. 2009. Sudden death in hypertrophic cardiomyopathy. *J Cardiovasc Transl Res,* 2, 368-80.

MARON, B. J., GARDIN, J. M., FLACK, J. M., GIDDING, S. S., KUROSAKI, T. T. & BILD, D. E. 1995. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. *Circulation,* 92, 785-9.

MARON, B. J. & MARON, M. S. 2013. Hypertrophic cardiomyopathy. *Lancet,* 381, 242-55.

MARON, B. J., MARON, M. S. & SEMSARIAN, C. 2012. Genetics of hypertrophic cardiomyopathy after 20 years: clinical perspectives. *J Am Coll Cardiol,* 60, 705-15.

MARON, B. J., OLIVOTTO, I., SPIRITO, P., CASEY, S. A., BELLONE, P., GOHMAN, T. E., GRAHAM, K. J., BURTON, D. A. & CECCHI, F. 2000. Epidemiology of hypertrophic cardiomyopathy-related death: revisited in a large non-referral-based patient population. *Circulation,* 102, 858-64.

MARTI, F., KRAUSE, A., POST, N. H., LYDDANE, C., DUPONT, B., SADELAIN, M. & KING, P. D. 2001. Negative-feedback regulation of CD28 costimulation by a novel mitogen-activated protein kinase phosphatase, MKP6. *J Immunol,* 166, 197-206.

MARTIN, A. B., GARSON, A., JR. & PERRY, J. C. 1994. Prolonged QT interval in hypertrophic and dilated cardiomyopathy in children. *Am Heart J,* 127, 64-70.

MARUYAMA, K., MATSUBARA, S., NATORI, R., NONOMURA, Y. & KIMURA, S. 1977. Connectin, an elastic protein of muscle. Characterization and Function. *J Biochem,* 82, 317-37.

MARZILIANO, N., PILOTTO, A., GRASSO, M., PASOTTI, M. & ARBUSTINI, E. 2006. Deletion of Glu at codon 13 of the TCAP gene encoding the titin-cap-telethonin is a rare polymorphism in a large Italian population. *Mol Genet Metab,* 89, 286-7.

MASON, P., BAYOL, S. & LOUGHNA, P. T. 1999. The novel sarcomeric protein telethonin exhibits developmental and functional regulation. *Biochem Biophys Res Commun,* 257, 699-703.

MATSUMOTO, Y., HAYASHI, T., INAGAKI, N., TAKAHASHI, M., HIROI, S., NAKAMURA, T., ARIMURA, T., NAKAMURA, K., ASHIZAWA, N., YASUNAMI, M., OHE, T., YANO, K. & KIMURA, A. 2005. Functional analysis of titin/connectin N2-B mutations found in cardiomyopathy. *J Muscle Res Cell Motil,* 26, 367-74.

MAYANS, O., VAN DER VEN, P. F., WILM, M., MUES, A., YOUNG, P., FURST, D. O., WILMANNS, M. & GAUTEL, M. 1998. Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature,* 395, 863-9.

MAZZONE, A., STREGE, P. R., TESTER, D. J., BERNARD, C. E., FAULKNER, G., DE GIORGIO, R., MAKIELSKI, J. C., STANGHELLINI, V., GIBBONS, S. J., ACKERMAN, M. J. & FARRUGIA, G. 2008. A mutation in telethonin alters Nav1.5 function. *J Biol Chem,* 283, 16537-44.

MCELHINNY, A. S., KAKINUMA, K., SORIMACHI, H., LABEIT, S. & GREGORIO, C. C. 2002. Muscle-specific RING finger-1 interacts with titin to regulate sarcomeric M-line and thick filament structure and may have nuclear functions via its interaction with glucocorticoid modulatory element binding protein-1. *J Cell Biol,* 157, 125-36.

MCFARLAND, R., TAYLOR, R. W. & TURNBULL, D. M. 2010. A neurological perspective on mitochondrial disease. *Lancet Neurol,* 9, 829-40.

MCPHERSON, J. P., LEMMERS, B., CHAHWAN, R., PAMIDI, A., MIGON, E., MATYSIAK-ZABLOCKI, E., MOYNAHAN, M. E., ESSERS, J., HANADA, K., POONEPALLI, A., SANCHEZ-SWEATMAN, O., KHOKHA, R., KANAAR, R., JASIN, M., HANDE, M. P. & HAKEM, R. 2004. Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science,* 304, 1822-6.

MEDEIROS-DOMINGO, A., TAN, B. H., CROTTI, L., TESTER, D. J., ECKHARDT, L., CUORETTI, A., KROBOTH, S. L., SONG, C., ZHOU, Q., KOPP, D., SCHWARTZ, P. J., MAKIELSKI, J. C. & ACKERMAN, M. J. 2010. Gain-of-function mutation S422L in the KCNJ8-encoded cardiac K(ATP) channel Kir6.1 as a pathogenic substrate for J-wave syndromes. *Heart Rhythm,* 7, 1466-71.

MEDER, B., HAAS, J., KELLER, A., HEID, C., JUST, S., BORRIES, A., BOISGUERIN, V., SCHARFENBERGER-SCHMEER, M., STAHLER, P., BEIER, M., WEICHENHAN, D., STROM, T. M., PFEUFER, A., KORN, B., KATUS, H. A. & ROTTBAUER, W. 2011. Targeted Next-Generation Sequencing for the Molecular Genetic Diagnostics of Cardiomyopathies. *Circ Cardiovasc Genet*.

MEFFORD, H. C., CLAUIN, S., SHARP, A. J., MOLLER, R. S., ULLMANN, R., KAPUR, R., PINKEL, D., COOPER, G. M., VENTURA, M., ROPERS, H. H., TOMMERUP, N., EICHLER, E. E. & BELLANNE-CHANTELOT, C. 2007. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet,* 81, 1057-69.

MENG, W., MUSHIKA, Y., ICHII, T. & TAKEICHI, M. 2008. Anchorage of microtubule minus ends to adherens junctions regulates epithelial cell-cell contacts. *Cell,* 135, 948-59.

MERANTE, F., MYINT, T., TEIN, I., BENSON, L. & ROBINSON, B. H. 1996. An additional mitochondrial tRNA(Ile) point mutation (A-to-G at nucleotide 4295) causing hypertrophic cardiomyopathy. *Hum Mutat,* 8, 216-22.

MIHATSCH, K., NESTLER, M., SALUZ, H. P., HENKE, A. & MUNDER, T. 2009. Proapoptotic protein Siva binds to the muscle protein telethonin in cardiomyocytes during coxsackieviral infection. *Cardiovasc Res,* 81, 108-15.

MILLER, F. J., ROSENFELDT, F. L., ZHANG, C., LINNANE, A. W. & NAGLEY, P. 2003a. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic Acids Res,* 31, e61.

MILLER, M. K., BANG, M. L., WITT, C. C., LABEIT, D., TROMBITAS, C., WATANABE, K., GRANZIER, H., MCELHINNY, A. S., GREGORIO, C. C. & LABEIT, S. 2003b. The muscle ankyrin repeat proteins: CARP, ankrd2/Arpp and DARP as a family of titin filament-based stress response molecules. *J Mol Biol,* 333, 951-64.

MILLSON, A., LAGRAVE, D., WILLIS, M. J., ROWE, L. R., LYON, E. & SOUTH, S. T. 2012. Chromosomal loss of 3q26.3-3q26.32, involving a partial neuroligin 1 deletion, identified by genomic microarray in a child with microcephaly, seizure disorder, and severe intellectual disability. *Am J Med Genet A,* 158A, 159-65.

MOREIRA, E. S., WILTSHIRE, T. J., FAULKNER, G., NILFOROUSHAN, A., VAINZOF, M., SUZUKI, O. T., VALLE, G., REEVES, R., ZATZ, M., PASSOS-BUENO, M. R. & JENNE, D. E. 2000. Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. *Nat Genet,* 24, 163-6.

MOREN, A., OLOFSSON, A., STENMAN, G., SAHLIN, P., KANZAKI, T., CLAESSON-WELSH, L., TEN DIJKE, P., MIYAZONO, K. & HELDIN, C. H. 1994. Identification and characterization of LTBP-2, a novel latent transforming growth factor-beta-binding protein. *J Biol Chem,* 269, 32469-78.

MUES, A., VAN DER VEN, P. F., YOUNG, P., FURST, D. O. & GAUTEL, M. 1998. Two immunoglobulin-like domains of the Z-disc portion of titin interact in a conformation-dependent way with telethonin. *FEBS Lett,* 428, 111-4.

MUKHERJEE, S., DIAZ VALENCIA, J. D., STEWMAN, S., METZ, J., MONNIER, S., RATH, U., ASENJO, A. B., CHARAFEDDINE, R. A., SOSA, H. J., ROSS, J. L., MA, A. & SHARP, D. J. 2012. Human Fidgetin is a microtubule severing the enzyme and minus-end depolymerase that regulates mitosis. *Cell Cycle,* 11, 2359-66.

MURAKAMI, Y., SIRIPANYAPINYO, U., HONG, Y., KANG, J. Y., ISHIHARA, S., NAKAKUMA, H., MAEDA, Y. & KINOSHITA, T. 2003. PIG-W is critical for inositol acylation but not for flipping of glycosylphosphatidylinositol-anchor. *Mol Biol Cell,* 14, 4285-95.

NAGASE, T., KIKUNO, R., ISHIKAWA, K., HIROSAWA, M. & OHARA, O. 2000. Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res,* 7, 143-50.

NAKAHARA, M., SHIMOZAWA, M., NAKAMURA, Y., IRINO, Y., MORITA, M., KUDO, Y. & FUKAMI, K. 2005. A novel phospholipase C, PLC(eta)2, is a neuron-specific isozyme. *J Biol Chem,* 280, 29128-34.

NAKANO, N., HORI, H., ABE, M., SHIBATA, H., ARIMURA, T., SASAOKA, T., SAWABE, M., CHIDA, K., ARAI, T., NAKAHARA, K., KUBO, T., SUGIMOTO, K., KATSUYA, T., OGIHARA, T., DOI, Y., IZUMI, T. & KIMURA, A. 2007. Interaction of BMP10 with Tcap may modulate the course of hypertensive cardiac hypertrophy. *Am J Physiol Heart Circ Physiol,* 293, H3396-403.

NAVE, R., FURST, D. O. & WEBER, K. 1989. Visualization of the polarity of isolated titin molecules: a single globular head on a long thin rod as the M band anchoring domain? *J Cell Biol,* 109, 2177-87.

NCBI June 2013. http://www.ncbi.nlm.nih.gov/guide/.

NEAGOE, C., KULKE, M., DEL MONTE, F., GWATHMEY, J. K., DE TOMBE, P. P., HAJJAR, R. J. & LINKE, W. A. 2002. Titin isoform switch in ischemic human heart disease. *Circulation,* 106, 1333-41.

NEELY, J. R. & MORGAN, H. E. 1974. Relationship between carbohydrate and lipid metabolism and the energy balance of heart muscle. *Annu Rev Physiol,* 36, 413-59.

NERBONNE, J. M. & KASS, R. S. 2005. Molecular physiology of cardiac repolarization. *Physiol Rev,* 85, 1205-53.

NEWMAN, J., EGAN, D., WALTER, T. S., MEGED, R., BERRY, I., BEN JELLOUL, M., SUSSMAN, J. L., STUART, D. I. & PERRAKIS, A. 2005. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr,* 61, 1426-31.

NG, P. C., LEVY, S., HUANG, J., STOCKWELL, T. B., WALENZ, B. P., LI, K., AXELROD, N., BUSAM, D. A., STRAUSBERG, R. L. & VENTER, J. C. 2008. Genetic variation in an individual human exome. *PLoS Genet,* 4, e1000160.

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.

NICHOLAS, G., THOMAS, M., LANGLEY, B., SOMERS, W., PATEL, K., KEMP, C. F., SHARMA, M. & KAMBADUR, R. 2002. Titin-cap associates with, and regulates secretion of, Myostatin. *J Cell Physiol,* 193, 120-31.

NICHOLS, C. G. 2006. KATP channels as molecular sensors of cellular metabolism. *Nature,* 440, 470-6.

NICHOLS, C. G., SINGH, G. K. & GRANGE, D. K. 2013. KATP channels and cardiovascular disease: suddenly a syndrome. *Circ Res,* 112, 1059-72.

NIE, D. S., LIU, Y., JUAN, H. & YANG, X. 2013. Overexpression of human SPATA17 protein induces germ cell apoptosis in transgenic male mice. *Mol Biol Rep,* 40, 1905-10.

NORTON, N., LI, D., RAMPERSAUD, E., MORALES, A., MARTIN, E. R., ZUCHNER, S., GUO, S., GONZALEZ, M., HEDGES, D. J., ROBERTSON, P. D., KRUMM, N., NICKERSON, D. A. & HERSHBERGER, R. E. 2013. Exome Sequencing and Genome-Wide Linkage Analysis in 17 Families Illustrates the Complex Contribution of TTN Truncating Variants to Dilated Cardiomyopathy. *Circ Cardiovasc Genet*.

NORTON, N., LI, D., RIEDER, M. J., SIEGFRIED, J. D., RAMPERSAUD, E., ZUCHNER, S., MANGOS, S., GONZALEZ-QUINTANA, J., WANG, L., MCGEE, S., REISER, J., MARTIN, E., NICKERSON, D. A. & HERSHBERGER, R. E. 2011. Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am J Hum Genet,* 88, 273-82.

OBAYASHI, T., HATTORI, K., SUGIYAMA, S., TANAKA, M., TANAKA, T., ITOYAMA, S., DEGUCHI, H., KAWAMURA, K., KOGA, Y., TOSHIMA, H. & ET AL. 1992. Point mutations in mitochondrial DNA in patients with hypertrophic cardiomyopathy. *Am Heart J,* 124, 1263-9.

OECHSLIN, E. & JENNI, R. 2011. Left ventricular non-compaction revisited: a distinct phenotype with genetic heterogeneity? *Eur Heart J,* 32, 1446-56.

OECHSLIN, E. N., ATTENHOFER JOST, C. H., ROJAS, J. R., KAUFMANN, P. A. & JENNI, R. 2000. Long-term follow-up of 34 adults with isolated left ventricular noncompaction: a distinct cardiomyopathy with poor prognosis. *J Am Coll Cardiol,* 36, 493-500.

OHTSUKA, H., YAJIMA, H., MARUYAMA, K. & KIMURA, S. 1997. Binding of the N-terminal 63 kDa portion of connectin/titin to alpha-actinin as revealed by the yeast two-hybrid system. *FEBS Lett,* 401, 65-7.

OLIVE, M., SHATUNOV, A., GONZALEZ, L., CARMONA, O., MORENO, D., QUEREDA, L. G., MARTINEZ-MATOS, J. A., GOLDFARB, L. G. & FERRER, I. 2008. Transcription-terminating mutation in telethonin causing autosomal recessive muscular dystrophy type 2G in a European patient. *Neuromuscul Disord,* 18, 929-33.

OMIM March 2013. Online Mendelian Inheritance in Man. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). . World Wide Web URL: http://omim.org/.

OTTENHEIJM, C. A., HIDALGO, C., ROST, K., GOTTHARDT, M. & GRANZIER, H. 2009. Altered contractility of skeletal muscle in mice deficient in titin's M-band region. *J Mol Biol,* 393, 10-26.

PAGE, L. J., SOWERBY, P. J., LUI, W. W. & ROBINSON, M. S. 1999. Gamma-synergin: an EH domain-containing protein that interacts with gamma-adaptin. *J Cell Biol,* 146, 993-1004.

PASMANT, E., SABBAGH, A., SPURLOCK, G., LAURENDEAU, I., GRILLO, E., HAMEL, M. J., MARTIN, L., BARBAROT, S., LEHEUP, B., RODRIGUEZ, D., LACOMBE, D., DOLLFUS, H., PASQUIER, L., ISIDOR, B., FERKAL, S., SOULIER, J., SANSON, M., DIEUX-COESLIER, A., BIECHE, I., PARFAIT, B., VIDAUD, M., WOLKENSTEIN, P., UPADHYAYA, M. & VIDAUD, D. 2010. NF1 microdeletions in neurofibromatosis type 1: from genotype to phenotype. *Hum Mutat,* 31, E1506-18.

PATEL, K. P., O'BRIEN, T. W., SUBRAMONY, S. H., SHUSTER, J. & STACPOOLE, P. W. 2012. The spectrum of pyruvate dehydrogenase complex deficiency: clinical, biochemical and genetic features in 371 patients. *Mol Genet Metab,* 106, 385-94.

PATTERSON, K. I., BRUMMER, T., O'BRIEN, P. M. & DALY, R. J. 2009. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem J,* 418, 475-89.

PECKHAM, M., YOUNG, P. & GAUTEL, M. 1997. Constitutive and variable regions of Z-disk titin/connectin in myofibril formation: a dominant-negative screen. *Cell Struct Funct,* 22, 95-101.

PELIN, K., RIDANPAA, M., DONNER, K., WILTON, S., KRISHNARAJAH, J., LAING, N., KOLMERER, B., MILLEVOI, S., LABEIT, S., DE LA CHAPELLE, A. & WALLGREN-PETTERSON, C. 1997. Refined localisation of the genes for nebulin and titin on chromosome 2q allows the assignment of nebulin as a candidate gene for autosomal recessive nemaline myopathy. *Eur J Hum Genet,* 5, 229-34.

PELTONEN, L., PEROLA, M., NAUKKARINEN, J. & PALOTIE, A. 2006. Lessons from studying monogenic disease for common disease. *Hum Mol Genet,* 15 Spec No 1, R67-74.

PENTIKAINEN, U., JIANG, P., TAKALA, H., RUSKAMO, S., CAMPBELL, I. D. & YLANNE, J. 2011. Assembly of a filamin four-domain fragment and the influence of splicing variant-1 on the structure. *J Biol Chem,* 286, 26921-30.

PEREIRA, L., FREITAS, F., FERNANDES, V., PEREIRA, J. B., COSTA, M. D., COSTA, S., MAXIMO, V., MACAULAY, V., ROCHA, R. & SAMUELS, D. C. 2009. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet,* 84, 628-40.

PERLI, E., GIORDANO, C., TUPPEN, H. A. L., MONTOPOLI, M., MONTANARI, A., ORLANDI, M., PISANO, A., CATANZARO, D., CAPARROTTA, L., MUSUMECI, B., AUTORE, C., MOREA, V., DI MICCO, P., CAMPESE, A. F., LEOPIZZI, M., GALLO, P., FRANCISCI, S., FRONTALI, L., TAYLOR, R. W. & D'AMATI, G. 2012. Isoleucyl-tRNA synthetase levels modulate the penetrance of a homoplasmic m.4277T>C mitochondrial tRNAIle mutation causing hypertrophic cardiomyopathy. *Hum Mol Genet,* 21, 85-100.

PERROT, A., POSCH, M. G. & OSTERZIEL, K. J. 2006. Deletion of Glu at codon 13 in the TCAP gene encoding the Z-disc protein titin-cap/telethonin is a rare non-synonymous polymorphism. *Mol Genet Metab,* 88, 199-200.

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.

PHG_FOUNDATION. 2013. Response to the ACMG statement on report of incidental findings generated by genome wide sequencing.

PINKEL, D., SEGRAVES, R., SUDAR, D., CLARK, S., POOLE, I., KOWBEL, D., COLLINS, C., KUO, W. L., CHEN, C., ZHAI, Y., DAIRKEE, S. H., LJUNG, B. M., GRAY, J. W. & ALBERTSON, D. G. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet,* 20, 207-11.

PINOTSIS, N., PETOUKHOV, M., LANGE, S., SVERGUN, D., ZOU, P., GAUTEL, M. & WILMANNS, M. 2006. Evidence for a dimeric assembly of two titin/telethonin complexes induced by the telethonin C-terminus. *J Struct Biol,* 155, 239-50.

PIZZAGALLI, F., HAGENBUCH, B., STIEGER, B., KLENK, U., FOLKERS, G. & MEIER, P. J. 2002. Identification of a novel human organic anion transporting polypeptide as a high affinity thyroxine transporter. *Mol Endocrinol,* 16, 2283-96.

POLITOU, A. S., THOMAS, D. J. & PASTORE, A. 1995. The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys J,* 69, 2601-10.

POLLARD, K. S., HUBISZ, M. J., ROSENBLOOM, K. R. & SIEPEL, A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res,* 20, 110-21.

PROBST, S., OECHSLIN, E., SCHULER, P., GREUTMANN, M., BOYE, P., KNIRSCH, W., BERGER, F., THIERFELDER, L., JENNI, R. & KLAASSEN, S. 2011. Sarcomere gene mutations in isolated left ventricular noncompaction cardiomyopathy do not predict clinical phenotype. *Circ Cardiovasc Genet,* 4, 367-74.

PTACEK, L. J., TAWIL, R., GRIGGS, R. C., ENGEL, A. G., LAYZER, R. B., KWIECINSKI, H., MCMANIS, P. G., SANTIAGO, L., MOORE, M., FOUAD, G. & ET AL. 1994. Dihydropyridine receptor mutations cause hypokalemic periodic paralysis. *Cell,* 77, 863-8.

PUCHNER, E. M., ALEXANDROVICH, A., KHO, A. L., HENSEN, U., SCHAFER, L. V., BRANDMEIER, B., GRATER, F., GRUBMULLER, H., GAUB, H. E. & GAUTEL, M. 2008. Mechanoenzymatics of titin kinase. *Proc Natl Acad Sci U S A,* 105, 13385-90.

PYMOL June 2013. The PyMOL Molecular Graphics System. Version 1.2r3pre ed.: Schrödinger, LLC.

QI, J., CHI, L., LABEIT, S. & BANES, A. J. 2008. Nuclear localization of the titin Z1Z2Zr domain and role in regulating cell proliferation. *Am J Physiol Cell Physiol,* 295, C975-85.

QUINTERO, O. A., DIVITO, M. M., ADIKES, R. C., KORTAN, M. B., CASE, L. B., LIER, A. J., PANARETOS, N. S., SLATER, S. Q., RENGARAJAN, M., FELIU, M. & CHENEY, R. E. 2009. Human Myo19 is a novel myosin that associates with mitochondria. *Curr Biol,* 19, 2008-13.

RAJESH, A., MADHUBABU, G. & YENUGU, S. 2011. Identification and characterization of Wfdc gene expression in the male reproductive tract of the rat. *Mol Reprod Dev,* 78, 633-41.

RAUCH, A., WIECZOREK, D., GRAF, E., WIELAND, T., ENDELE, S., SCHWARZMAYR, T., ALBRECHT, B., BARTHOLDI, D., BEYGO, J., DI DONATO, N., DUFKE, A., CREMER, K., HEMPEL, M., HORN, D., HOYER, J., JOSET, P., ROPKE, A., MOOG, U., RIESS, A., THIEL, C. T., TZSCHACH, A., WIESENER, A., WOHLLEBER, E., ZWEIER, C., EKICI, A. B., ZINK, A. M., RUMP, A., MEISINGER, C.,

GRALLERT, H., STICHT, H., SCHENCK, A., ENGELS, H., RAPPOLD, G., SCHROCK, E., WIEACKER, P., RIESS, O., MEITINGER, T., REIS, A. & STROM, T. M. 2012. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet,* 380, 1674-82.

RAVENS, U. & CERBAI, E. 2008. Role of potassium currents in cardiac arrhythmias. *Europace,* 10, 1133-7.

REDLINE, R. W., WILLIAMS, A. J., PATTERSON, P. & COLLINS, T. 1992. Human HOX4E: a gene strongly expressed in the adult male and female urogenital tracts. *Genomics,* 13, 425-30.

REGALADO, E. S., GUO, D. C., VILLAMIZAR, C., AVIDAN, N., GILCHRIST, D., MCGILLIVRAY, B., CLARKE, L., BERNIER, F., SANTOS-CORTEZ, R. L., LEAL, S. M., BERTOLI-AVELLA, A. M., SHENDURE, J., RIEDER, M. J., NICKERSON, D. A. & MILEWICZ, D. M. 2011. Exome sequencing identifies SMAD3 mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. *Circ Res,* 109, 680-6.

RICHARD, P., CHARRON, P., CARRIER, L., LEDEUIL, C., CHEAV, T., PICHEREAU, C., BENAICHE, A., ISNARD, R., DUBOURG, O., BURBAN, M., GUEFFET, J. P., MILLAIRE, A., DESNOS, M., SCHWARTZ, K., HAINQUE, B. & KOMAJDA, M. 2003. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation,* 107, 2227-32.

RICHMOND, J. E., DAVIS, W. S. & JORGENSEN, E. M. 1999. UNC-13 is required for synaptic vesicle fusion in C. elegans. *Nat Neurosci,* 2, 959-64.

RIMMER, A., MATHIESON, I., LUNTER, G. & MCVEAN, G. 2012. Platypus: An Integrated Variant Caller. www.well.ox.ac.uk/platypus.

RISGAARD, B., JABBARI, R., REFSGAARD, L., HOLST, A., HAUNSO, S., SADJADIEH, A., WINKEL, B., OLESEN, M. & TFELT-HANSEN, J. 2013. High prevalence of genetic variants previously associated with Brugada syndrome in new exome data. *Clin Genet*.

ROBSON, K. J., CHANDRA, T., MACGILLIVRAY, R. T. & WOO, S. L. 1982. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proc Natl Acad Sci U S A,* 79, 4701-5.

RODRIGUEZ, M. S., DARGEMONT, C. & STUTZ, F. 2004. Nuclear export of RNA. *Biol Cell,* 96, 639-55.

ROYER-POKORA, B., KUNKEL, L. M., MONACO, A. P., GOFF, S. C., NEWBURGER, P. E., BAEHNER, R. L., COLE, F. S., CURNUTTE, J. T. & ORKIN, S. H. 1986. Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. *Nature,* 322, 32-8.

RUAN, Y., LIU, N., NAPOLITANO, C. & PRIORI, S. G. 2008. Therapeutic strategies for long-QT syndrome: does the molecular substrate matter? *Circ Arrhythm Electrophysiol,* 1, 290-7.

RUBINO, F., PIREDDA, R., CALABRESE, F. M., SIMONE, D., LANG, M., CALABRESE, C., PETRUZZELLA, V., TOMMASEO-PONZETTA, M., GASPARRE, G. & ATTIMONELLI, M. 2012. HmtDB, a genomic resource for mitochondrion-based human variability studies. *Nucleic Acids Res,* 40, D1150-9.

RUIZ-PESINI, E., LOTT, M. T., PROCACCIO, V., POOLE, J. C., BRANDON, M. C., MISHMAR, D., YI, C., KREUZIGER, J., BALDI, P. & WALLACE, D. C. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res,* 35, D823-8.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C., KAKOL, J. M., STEIN, L. D., MARTH, G., SHERRY, S., MULLIKIN, J. C., MORTIMORE, B. J., WILLEY, D. L., HUNT, S. E., COLE, C. G., COGGILL, P. C., RICE, C. M., NING, Z., ROGERS, J., BENTLEY, D. R., KWOK, P. Y., MARDIS, E. R., YEH, R. T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R. H., MCPHERSON, J. D., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W. J., REICH, D., HIGGINS, J., DALY, M. J., BLUMENSTIEL, B., BALDWIN, J., STANGE-THOMANN, N., ZODY, M. C., LINTON, L., LANDER, E. S. & ALTSHULER, D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409, 928-33.

SADIKOT, T., HAMMOND, C. R. & FERRARI, M. B. 2010. Distinct roles for telethonin N-versus C-terminus in sarcomere assembly and maintenance. *Dev Dyn,* 239, 1124-35.

SALEH, M., VAILLANCOURT, J. P., GRAHAM, R. K., HUYCK, M., SRINIVASULA, S. M., ALNEMRI, E. S., STEINBERG, M. H., NOLAN, V., BALDWIN, C. T., HOTCHKISS, R. S., BUCHMAN, T. G., ZEHNBAUER, B. A., HAYDEN, M. R., FARRER, L. A., ROY, S. & NICHOLSON, D. W. 2004. Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. *Nature,* 429, 75-9.

SALTZMAN, A. J., MANCINI-DINARDO, D., LI, C., CHUNG, W. K., HO, C. Y., HURST, S., WYNN, J., CARE, M., HAMILTON, R. M., SEIDMAN, G. W., GORHAM, J., MCDONOUGH, B., SPARKS, E., SEIDMAN, J. G., SEIDMAN, C. E. & REHM, H. L. 2010. Short communication: the cardiac myosin binding protein C Arg502Trp mutation: a common cause of hypertrophic cardiomyopathy. *Circ Res,* 106, 1549-52.

SANTORELLI, F. M., MAK, S. C., VAZQUEZ-ACEVEDO, M., GONZALEZ-ASTIAZARAN, A., RIDAURA-SANZ, C., GONZALEZ-HALPHEN, D. & DIMAURO, S. 1995. A novel mitochondrial DNA point mutation associated with mitochondrial encephalocardiomyopathy. *Biochem Biophys Res Commun,* 216, 835-40.

SARMA, R. J., CHANA, A. & ELKAYAM, U. 2010. Left ventricular noncompaction. *Prog Cardiovasc Dis,* 52, 264-73.

SASSE-KLAASSEN, S., PROBST, S., GERULL, B., OECHSLIN, E., NURNBERG, P., HEUSER, A., JENNI, R., HENNIES, H. C. & THIERFELDER, L. 2004. Novel gene locus for autosomal dominant left ventricular noncompaction maps to chromosome 11p15. *Circulation,* 109, 2720-3.

SATOH, M. & KUROIWA, T. 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res,* 196, 137-40.

SATOH, M., TAKAHASHI, M., SAKAMOTO, T., HIROE, M., MARUMO, F. & KIMURA, A. 1999. Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochem Biophys Res Commun,* 262, 411-7.

SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9, 671-5.

SCHRODER, R., REIMANN, J., IAKOVENKO, A., MUES, A., BONNEMANN, C. G., MATTEN, J. & GAUTEL, M. 2001. Early and selective disappearance of telethonin protein from the sarcomere in neurogenic atrophy. *J Muscle Res Cell Motil,* 22, 259-64.

SCHULER, B. A., PRISCO, S. Z. & JACOB, H. J. 2013. Using whole exome sequencing to walk from clinical practice to research and back again. *Circulation,* 127, 968-70.

SCHWARTZ, P. J., CROTTI, L. & INSOLIA, R. 2012. Long-QT syndrome: from genetics to management. *Circ Arrhythm Electrophysiol,* 5, 868-77.

SCHWARTZ, P. J., STRAMBA-BADIALE, M., CROTTI, L., PEDRAZZINI, M., BESANA, A., BOSI, G., GABBARINI, F., GOULENE, K., INSOLIA, R., MANNARINO, S., MOSCA, F., NESPOLI, L., RIMINI, A., ROSATI, E., SALICE, P. & SPAZZOLINI, C. 2009. Prevalence of the congenital long-QT syndrome. *Circulation,* 120, 1761-7.

SCHWARZ, J. M., RODELSPERGER, C., SCHUELKE, M. & SEELOW, D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods,* 7, 575-6.

SHARP, A. J., HANSEN, S., SELZER, R. R., CHENG, Z., REGAN, R., HURST, J. A., STEWART, H., PRICE, S. M., BLAIR, E., HENNEKAM, R. C., FITZPATRICK, C. A., SEGRAVES, R., RICHMOND, T. A., GUIVER, C., ALBERTSON, D. G., PINKEL, D., EIS, P. S., SCHWARTZ, S., KNIGHT, S. J. & EICHLER, E. E. 2006. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet,* 38, 1038-42.

SHAW, R. M., FAY, A. J., PUTHENVEEDU, M. A., VON ZASTROW, M., JAN, Y. N. & JAN, L. Y. 2007. Microtubule plus-end-tracking proteins target gap junctions directly from the cell interior to adherens junctions. *Cell,* 128, 547-60.

SHEIKH, F., RASKIN, A., CHU, P. H., LANGE, S., DOMENIGHETTI, A. A., ZHENG, M., LIANG, X., ZHANG, T., YAJIMA, T., GU, Y., DALTON, N. D., MAHATA, S. K., DORN, G. W., 2ND, BROWN, J. H., PETERSON, K. L., OMENS, J. H., MCCULLOCH, A. D. & CHEN, J. 2008. An FHL1-containing complex within the cardiomyocyte sarcomere mediates hypertrophic biomechanical stress responses in mice. *J Clin Invest,* 118, 3870-80.

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.

SHI, R., ZHANG, Y., YANG, C., HUANG, C., ZHOU, X., QIANG, H., GRACE, A. A., HUANG, C. L. & MA, A. 2008. The cardiac sodium channel mutation delQKP 1507-1509 is associated with the expanding phenotypic spectrum of LQT3, conduction disorder, dilated cardiomyopathy, and high incidence of youth sudden death. *Europace,* 10, 1329-35.

SIDOW, A. 2002. Sequence first. Ask questions later. *Cell,* 111, 13-6.

SIGGS, O. M., LI, X., XIA, Y. & BEUTLER, B. 2012. ZBTB1 is a determinant of lymphoid development. *J Exp Med,* 209, 19-27.

SILVESTRI, G., BERTINI, E., SERVIDEI, S., RANA, M., ZACHARA, E., RICCI, E. & TONALI, P. 1997. Maternally inherited cardiomyopathy: a new phenotype associated with the A to G AT nt.3243 of mitochondrial DNA (MELAS mutation). *Muscle Nerve,* 20, 221-5.

SILVESTRI, G., SERVIDEI, S., RANA, M., RICCI, E., SPINAZZOLA, A., PARIS, E. & TONALI, P. 1996. A novel mitochondrial DNA point mutation in the tRNA(Ile) gene is associated with progressive external ophtalmoplegia. *Biochem Biophys Res Commun,* 220, 623-7.

SIMPSON, M. A., IRVING, M. D., ASILMAZ, E., GRAY, M. J., DAFOU, D., ELMSLIE, F. V., MANSOUR, S., HOLDER, S. E., BRAIN, C. E., BURTON, B. K., KIM, K. H., PAULI, R. M., AFTIMOS, S., STEWART, H., KIM, C. A., HOLDER-ESPINASSE, M., ROBERTSON, S. P., DRAKE, W. M. & TREMBATH, R. C. 2011. Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nat Genet,* 43, 303-5.

SPIEGEL, A. M. 1996. Mutations in G proteins and G protein-coupled receptors in endocrine disease. *J Clin Endocrinol Metab,* 81, 2434-42.

SPIEKERKOETTER, U. 2010. Mitochondrial fatty acid oxidation disorders: clinical presentation of long-chain fatty acid oxidation defects before and after newborn screening. *J Inherit Metab Dis,* 33, 527-32.

SPLAWSKI, I., TIMOTHY, K. W., DECHER, N., KUMAR, P., SACHSE, F. B., BEGGS, A. H., SANGUINETTI, M. C. & KEATING, M. T. 2005. Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc Natl Acad Sci U S A,* 102, 8089-96; discussion 8086-8.

SPLAWSKI, I., TIMOTHY, K. W., SHARPE, L. M., DECHER, N., KUMAR, P., BLOISE, R., NAPOLITANO, C., SCHWARTZ, P. J., JOSEPH, R. M., CONDOURIS, K., TAGER-FLUSBERG, H., PRIORI, S. G., SANGUINETTI, M. C. & KEATING, M. T. 2004. Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell,* 119, 19-31.

STAESSEN, J. A., WANG, J., BIANCHI, G. & BIRKENHAGER, W. H. 2003. Essential hypertension. *Lancet,* 361, 1629-41.

STENSON, P. D., MORT, M., BALL, E. V., HOWELLS, K., PHILLIPS, A. D., THOMAS, N. S. & COOPER, D. N. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med,* 1, 13.

STEWART, A. J., MUKHERJEE, J., ROBERTS, S. J., LESTER, D. & FARQUHARSON, C. 2005. Identification of a novel class of mammalian phosphoinositol-specific phospholipase C enzymes. *Int J Mol Med,* 15, 117-21.

STROM, T. M., NYAKATURA, G., APFELSTEDT-SYLLA, E., HELLEBRAND, H., LORENZ, B., WEBER, B. H., WUTZ, K., GUTWILLINGER, N., RUTHER, K., DRESCHER, B., SAUER, C., ZRENNER, E., MEITINGER, T., ROSENTHAL, A. & MEINDL, A. 1998. An L-type calcium-channel gene mutated in incomplete X-linked congenital stationary night blindness. *Nat Genet,* 19, 260-3.

TAKATA, H., UCHIYAMA, S., NAKAMURA, N., NAKASHIMA, S., KOBAYASHI, S., SONE, T., KIMURA, S., LAHMERS, S., GRANZIER, H., LABEIT, S., MATSUNAGA, S. & FUKUI, K. 2007. A comparative proteome analysis of human metaphase chromosomes isolated from two different cell lines reveals a set of conserved chromosome-associated proteins. *Genes Cells,* 12, 269-84.

TANAKA, M., INO, H., OHNO, K., HATTORI, K., SATO, W., OZAWA, T., TANAKA, T. & ITOYAMA, S. 1990. Mitochondrial mutation in fatal infantile cardiomyopathy. *Lancet,* 336, 1452.

TANAKA, M., TAKEYASU, T., FUKU, N., LI-JUN, G. & KURATA, M. 2004. Mitochondrial genome single nucleotide polymorphisms and their phenotypes in the Japanese. *Ann N Y Acad Sci,* 1011, 7-20.

TANIIKE, M., FUKUSHIMA, H., YANAGIHARA, I., TSUKAMOTO, H., TANAKA, J., FUJIMURA, H., NAGAI, T., SANO, T., YAMAOKA, K., INUI, K. & ET AL. 1992. Mitochondrial tRNA(Ile) mutation in fatal cardiomyopathy. *Biochem Biophys Res Commun,* 186, 47-53.

TANIMURA, A., HAYASHI, I., ADACHI, K., NAKASHIMA, T., OTA, K. & TOSHIMA, H. 1977. Noonan syndrome with hypertrophic obstructive cardiomyopathy. *Acta Pathol Jpn,* 27, 225-30.

TAYLOR, M., GRAW, S., SINAGRA, G., BARNES, C., SLAVOV, D., BRUN, F., PINAMONTI, B., SALCEDO, E. E., SAUER, W., PYXARAS, S., ANDERSON, B., SIMON, B., BOGOMOLOVAS, J., LABEIT, S., GRANZIER, H. & MESTRONI, L. 2011. Genetic Variation in Titin in Arrhythmogenic Right Ventricular Cardiomyopathy-Overlap Syndromes. *Circulation*.

TAYLOR, R. W., CHINNERY, P. F., BATES, M. J., JACKSON, M. J., JOHNSON, M. A., ANDREWS, R. M. & TURNBULL, D. M. 1998. A novel mitochondrial DNA point mutation in the tRNA(Ile) gene: studies in a patient presenting with chronic progressive external ophthalmoplegia and multiple sclerosis. *Biochem Biophys Res Commun,* 243, 47-51.

TAYLOR, R. W., GIORDANO, C., DAVIDSON, M. M., D'AMATI, G., BAIN, H., HAYES, C. M., LEONARD, H., BARRON, M. J., CASALI, C., SANTORELLI, F. M., HIRANO, M., LIGHTOWLERS, R. N., DIMAURO, S. & TURNBULL, D. M. 2003. A homoplasmic mitochondrial transfer ribonucleic acid mutation as a cause of maternally inherited hypertrophic cardiomyopathy. *J Am Coll Cardiol,* 41, 1786-96.

TEEKAKIRIKUL, P., KELLY, M. A., REHM, H. L., LAKDAWALA, N. K. & FUNKE, B. H. 2013. Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era. *J Mol Diagn,* 15, 158-70.

THOMPSON, T. G., CHAN, Y. M., HACK, A. A., BROSIUS, M., RAJALA, M., LIDOV, H. G., MCNALLY, E. M., WATKINS, S. & KUNKEL, L. M. 2000. Filamin 2 (FLN2): A muscle-specific sarcoglycan interacting protein. *J Cell Biol,* 148, 115-26.

TIAN, L. F., LI, H. Y., JIN, B. F., PAN, X., MAN, J. H., ZHANG, P. J., LI, W. H., LIANG, B., LIU, H., ZHAO, J., GONG, W. L., ZHOU, T. & ZHANG, X. M. 2006. MDM2 interacts with and downregulates a sarcomeric protein, TCAP. *Biochem Biophys Res Commun,* 345, 355-61.

TISO, N., STEPHAN, D. A., NAVA, A., BAGATTIN, A., DEVANEY, J. M., STANCHI, F., LARDERET, G., BRAHMBHATT, B., BROWN, K., BAUCE, B., MURIAGO, M., BASSO, C., THIENE, G., DANIELI, G. A. & RAMPAZZO, A. 2001. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum Mol Genet,* 10, 189-94.

TOMARI, Y., HINO, N., NAGAIKE, T., SUZUKI, T. & UEDA, T. 2003. Decreased CCA-addition in human mitochondrial tRNAs bearing a pathogenic A4317G or A10044G mutation. *J Biol Chem,* 278, 16828-33.

TOURNAMILLE, C., COLIN, Y., CARTRON, J. P. & LE VAN KIM, C. 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet,* 10, 224-8.

TRUSLOVE, G. M. 1956. The anatomy and development of the fidget mouse. *J. Genet,* 54, 64-86.

TSKHOVREBOVA, L. & TRINICK, J. 2003. Titin: properties and family relationships. *Nat Rev Mol Cell Biol,* 4, 679-89.

TUPPEN, H. A., BLAKELY, E. L., TURNBULL, D. M. & TAYLOR, R. W. 2010. Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta,* 1797, 113-28.

UDD, B. 1992. Limb-girdle type muscular dystrophy in a large family with distal myopathy: homozygous manifestation of a dominant gene? *J Med Genet,* 29, 383-9.

UHLEN, M., OKSVOLD, P., FAGERBERG, L., LUNDBERG, E., JONASSON, K., FORSBERG, M., ZWAHLEN, M., KAMPF, C., WESTER, K., HOBER, S., WERNERUS, H., BJORLING, L. & PONTEN, F. 2010. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol,* 28, 1248-50.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40, e115.

UPADHYAYA, M., SHEN, M., CHERRYSON, A., FARNHAM, J., MAYNARD, J., HUSON, S. M. & HARPER, P. S. 1992. Analysis of mutations at the neurofibromatosis 1 (NF1) locus. *Hum Mol Genet,* 1, 735-40.

VAGNONI, A., PERKINTON, M. S., GRAY, E. H., FRANCIS, P. T., NOBLE, W. & MILLER, C. C. 2012. Calsyntenin-1 mediates axonal transport of the amyloid precursor protein and regulates Abeta production. *Hum Mol Genet,* 21, 2845-54.

VAINZOF, M., MOREIRA, E. S., SUZUKI, O. T., FAULKNER, G., VALLE, G., BEGGS, A. H., CARPEN, O., RIBEIRO, A. F., ZANOTELI, E., GURGEL-GIANNETI, J., TSANACLIS, A. M., SILVA, H. C., PASSOS-BUENO, M. R. & ZATZ, M. 2002. Telethonin protein expression in neuromuscular disorders. *Biochim Biophys Acta,* 1588, 33-40.

VALLE, G., FAULKNER, G., DE ANTONI, A., PACCHIONI, B., PALLAVICINI, A., PANDOLFO, D., TISO, N., TOPPO, S., TREVISAN, S. & LANFRANCHI, G. 1997. Telethonin, a novel sarcomeric protein of heart and skeletal muscle. *FEBS Lett,* 415, 163-8.

VAN DER VEN, P. F., BARTSCH, J. W., GAUTEL, M., JOCKUSCH, H. & FURST, D. O. 2000a. A functional knock-out of titin results in defective myofibril assembly. *J Cell Sci,* 113 (Pt 8), 1405-14.

VAN DER VEN, P. F., WIESNER, S., SALMIKANGAS, P., AUERBACH, D., HIMMEL, M., KEMPA, S., HAYESS, K., PACHOLSKY, D., TAIVAINEN, A., SCHRODER, R., CARPEN, O. & FURST, D. O. 2000b. Indications for a novel muscular dystrophy pathway. gamma-filamin, the muscle-specific filamin isoform, interacts with myotilin. *J Cell Biol,* 151, 235-48.

VAN DER ZWAAG, P. A., VAN RIJSINGEN, I. A., ASIMAKI, A., JONGBLOED, J. D., VAN VELDHUISEN, D. J., WIESFELD, A. C., COX, M. G., VAN LOCHEM, L. T., DE BOER, R. A., HOFSTRA, R. M., CHRISTIAANS, I., VAN SPAENDONCK-ZWARTS, K. Y., LEKANNE DIT DEPREZ, R. H., JUDGE, D. P., CALKINS, H., SUURMEIJER, A. J., HAUER, R. N., SAFFITZ, J. E., WILDE, A. A., VAN DEN BERG, M. P. & VAN TINTELEN, J. P. 2012. Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. *Eur J Heart Fail,* 14, 1199-207.

VAN DRIEST, S. L., OMMEN, S. R., TAJIK, A. J., GERSH, B. J. & ACKERMAN, M. J. 2005. Yield of genetic testing in hypertrophic cardiomyopathy. *Mayo Clin Proc,* 80, 739-44.

VAN EL, C. G., CORNEL, M. C., BORRY, P., HASTINGS, R. J., FELLMANN, F., HODGSON, S. V., HOWARD, H. C., CAMBON-THOMSEN, A., KNOPPERS, B. M., MEIJERS-HEIJBOER, H., SCHEFFER, H., TRANEBJAERG, L., DONDORP, W. & DE WERT, G. M. 2013. Whole-genome sequencing in health care. *Eur J Hum Genet,* 21, 580-4.

VAN ENGELEN, K., BAARS, M. J., FELIX, J. P., POSTMA, A. V., MULDER, B. J. & SMETS, E. M. 2013. The value of the clinical geneticist caring for adults with congenital heart disease: Diagnostic yield and patients' perspective. *Am J Med Genet A,* 161, 1628-37.

VAN OVEN, M. & KAYSER, M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat,* 30, E386-94.

VAN SWIETEN, J. C., BRUSSE, E., DE GRAAF, B. M., KRIEGER, E., VAN DE GRAAF, R., DE KONING, I., MAAT-KIEVIT, A., LEEGWATER, P., DOOIJES, D., OOSTRA, B. A. & HEUTINK, P. 2003. A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]. *Am J Hum Genet,* 72, 191-9.

VAROQUEAUX, F., ARAMUNI, G., RAWSON, R. L., MOHRMANN, R., MISSLER, M., GOTTMANN, K., ZHANG, W., SUDHOF, T. C. & BROSE, N. 2006. Neuroligins determine synapse maturation and function. *Neuron,* 51, 741-54.

VATTA, M., MOHAPATRA, B., JIMENEZ, S., SANCHEZ, X., FAULKNER, G., PERLES, Z., SINAGRA, G., LIN, J. H., VU, T. M., ZHOU, Q., BOWLES, K. R., DI LENARDA, A., SCHIMMENTI, L., FOX, M., CHRISCO, M. A., MURPHY, R. T., MCKENNA, W., ELLIOTT, P., BOWLES, N. E., CHEN, J., VALLE, G. & TOWBIN, J. A. 2003. Mutations in Cypher/ZASP in patients with dilated cardiomyopathy and left ventricular non-compaction. *J Am Coll Cardiol,* 42, 2014-27.

VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W., MURAL, R. J., SUTTON, G. G., SMITH, H. O., YANDELL, M., EVANS, C. A., HOLT, R. A., GOCAYNE, J. D., AMANATIDES, P., BALLEW, R. M., HUSON, D. H., WORTMAN, J. R., ZHANG, Q., KODIRA, C. D., ZHENG, X. H., CHEN, L., SKUPSKI, M., SUBRAMANIAN, G., THOMAS, P. D., ZHANG, J., GABOR MIKLOS, G. L., NELSON, C., BRODER, S., CLARK, A. G., NADEAU, J., MCKUSICK, V. A., ZINDER, N., LEVINE, A. J., ROBERTS, R. J., SIMON, M., SLAYMAN, C., HUNKAPILLER, M., BOLANOS, R., DELCHER, A., DEW, I., FASULO, D., FLANIGAN, M., FLOREA, L., HALPERN, A., HANNENHALLI, S., KRAVITZ, S., LEVY, S., MOBARRY, C., REINERT, K., REMINGTON, K., ABU-THREIDEH, J., BEASLEY, E., BIDDICK, K., BONAZZI, V., BRANDON, R., CARGILL, M., CHANDRAMOULISWARAN, I., CHARLAB, R., CHATURVEDI, K., DENG, Z., DI FRANCESCO, V., DUNN, P., EILBECK, K., EVANGELISTA, C., GABRIELIAN, A. E., GAN, W., GE, W., GONG, F., GU, Z., GUAN, P., HEIMAN, T. J., HIGGINS, M. E., JI, R. R., KE, Z., KETCHUM, K. A., LAI, Z., LEI, Y., LI, Z., LI, J., LIANG, Y., LIN, X., LU, F., MERKULOV, G. V., MILSHINA, N., MOORE, H. M., NAIK, A. K., NARAYAN, V. A., NEELAM, B., NUSSKERN, D., RUSCH, D. B., SALZBERG, S., SHAO, W., SHUE, B., SUN, J., WANG, Z., WANG, A., WANG, X., WANG, J., WEI, M., WIDES, R., XIAO, C., YAN, C., et al. 2001. The sequence of the human genome. *Science,* 291, 1304-51.

VISSERS, L. E. L. M., DE LIGT, J., GILISSEN, C., JANSSEN, I., STEEHOUWER, M., DE VRIES, P., VAN LIER, B., ARTS, P., WIESKAMP, N., DEL ROSARIO, M., VAN BON, B. W. M., HOISCHEN, A., DE VRIES, B. B. A., BRUNNER, H. G. & VELTMAN, J. A. 2010. A de novo paradigm for mental retardation. *Nat Genet,* 42, 1109-12.

VOELKEL, T. & LINKE, W. A. 2011. Conformation-regulated mechanosensory control via titin domains in cardiac muscle. *Pflugers Arch,* 462, 143-54.

VOGT, L., SCHRIMPF, S. P., MESKENAITE, V., FRISCHKNECHT, R., KINTER, J., LEONE, D. P., ZIEGLER, U. & SONDEREGGER, P. 2001. Calsyntenin-1, a proteolytically processed postsynaptic membrane protein with a cytoplasmic calcium-binding domain. *Mol Cell Neurosci,* 17, 151-66.

WALLACE, D. C., BROWN, M. D. & LOTT, M. T. 1999. Mitochondrial DNA variation in human evolution and disease. *Gene,* 238, 211-30.

WANG, J., SHANER, N., MITTAL, B., ZHOU, Q., CHEN, J., SANGER, J. M. & SANGER, J. W. 2005. Dynamics of Z-band based proteins in developing skeletal muscle cells. *Cell Motil Cytoskeleton,* 61, 34-48.

WANG, J. L., YANG, X., XIA, K., HU, Z. M., WENG, L., JIN, X., JIANG, H., ZHANG, P., SHEN, L., GUO, J. F., LI, N., LI, Y. R., LEI, L. F., ZHOU, J., DU, J., ZHOU, Y. F., PAN, Q., WANG, J., LI, R. Q. & TANG, B. S. 2010a. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain,* 133, 3510-8.

WANG, K., LI, M. & HAKONARSON, H. 2010b. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38, e164.

WANG, Q., CURRAN, M. E., SPLAWSKI, I., BURN, T. C., MILLHOLLAND, J. M., VANRAAY, T. J., SHEN, J., TIMOTHY, K. W., VINCENT, G. M., DE JAGER, T., SCHWARTZ, P. J., TOUBIN, J. A., MOSS, A. J., ATKINSON, D. L., LANDES, G. M., CONNORS, T. D. & KEATING, M. T. 1996. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat Genet,* 12, 17-23.

WANG, Q., SHEN, J., SPLAWSKI, I., ATKINSON, D., LI, Z., ROBINSON, J. L., MOSS, A. J., TOWBIN, J. A. & KEATING, M. T. 1995. SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell,* 80, 805-11.

WARD, L. D. & KELLIS, M. 2012. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol,* 30, 1095-106.

WARE, J. S., JOHN, S., ROBERTS, A. M., BUCHAN, R., GONG, S., PETERS, N. S., ROBINSON, D. O., LUCASSEN, A., BEHR, E. R. & COOK, S. A. 2013. Next generation diagnostics in inherited arrhythmia syndromes : a comparison of two approaches. *J Cardiovasc Transl Res,* 6, 94-103.

WATKINS, H. 2013. Assigning a causal role to genetic variants in hypertrophic cardiomyopathy. *Circ Cardiovasc Genet,* 6, 2-4.

WATKINS, H., CONNER, D., THIERFELDER, L., JARCHO, J. A., MACRAE, C., MCKENNA, W. J., MARON, B. J., SEIDMAN, J. G. & SEIDMAN, C. E. 1995a. Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. *Nat Genet,* 11, 434-7.

WATKINS, H., MCKENNA, W. J., THIERFELDER, L., SUK, H. J., ANAN, R., O'DONOGHUE, A., SPIRITO, P., MATSUMORI, A., MORAVEC, C. S., SEIDMAN, J. G. & ET AL. 1995b. Mutations in the genes for cardiac troponin T and alpha-tropomyosin in hypertrophic cardiomyopathy. *N Engl J Med,* 332, 1058-64.

WEEDON, M. N., HASTINGS, R., CASWELL, R., XIE, W., PASZKIEWICZ, K., ANTONIADI, T., WILLIAMS, M., KING, C., GREENHALGH, L., NEWBURY-ECOB, R. & ELLARD, S. 2011. Exome sequencing identifies a DYNC1H1 mutation in a large pedigree with dominant axonal Charcot-Marie-Tooth disease. *Am J Hum Genet,* 89, 308-12.

WEI, E. Q., BARNETT, A. S., PITT, G. S. & HENNESSEY, J. A. 2011. Fibroblast growth factor homologous factors in the heart: a potential locus for cardiac arrhythmias. *Trends Cardiovasc Med,* 21, 199-203.

WEINERT, S., BERGMANN, N., LUO, X., ERDMANN, B. & GOTTHARDT, M. 2006. M line-deficient titin causes cardiac lethality through impaired maturation of the sarcomere. *J Cell Biol,* 173, 559-70.

WHITING, A., WARDALE, J. & TRINICK, J. 1989. Does titin regulate the length of muscle thick filaments? *J Mol Biol,* 205, 263-8.

WHITMORE, L. & WALLACE, B. A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res,* 32, W668-73.

WILLIS, M. S., SCHISLER, J. C., PORTBURY, A. L. & PATTERSON, C. 2009. Build it up-Tear it down: protein quality control in the cardiac sarcomere. *Cardiovasc Res,* 81, 439-48.

WITT, S. H., GRANZIER, H., WITT, C. C. & LABEIT, S. 2005. MURF-1 and MURF-2 target a specific subset of myofibrillar proteins redundantly: towards understanding MURF-dependent muscle ubiquitination. *J Mol Biol,* 350, 713-22.

WU, C. 2005. Migfilin and its binding partners: from cell biology to human diseases. *J Cell Sci,* 118, 659-64.

YAMASAKI, R., WU, Y., MCNABB, M., GREASER, M., LABEIT, S. & GRANZIER, H. 2002. Protein kinase A phosphorylates titin's cardiac-specific N2B domain and reduces passive tension in rat cardiac myocytes. *Circ Res,* 90, 1181-8.

YANG, F., XU, Y. P., LI, J., DUAN, S. S., FU, Y. J., ZHANG, Y., ZHAO, Y., QIAO, W. T., CHEN, Q. M., GENG, Y. Q., CHE, C. Y., CAO, Y. L., WANG, Y., ZHANG, L., LONG, L., HE, J., CUI, Q. C., CHEN, S. C., WANG, S. H. & LIU, L. 2009. Cloning and characterization of a novel intracellular protein p48.2 that negatively regulates cell cycle progression. *Int J Biochem Cell Biol,* 41, 2240-50.

YIN, L. L., LI, J. M., ZHOU, Z. M. & SHA, J. H. 2005. Identification of a novel testis-specific gene and its potential roles in testis development/spermatogenesis. *Asian J Androl,* 7, 127-37.

YOSKOVITZ, G., PELED, Y., GRAMLICH, M., LAHAT, H., RESNIK-WOLF, H., FEINBERG, M. S., AFEK, A., PRAS, E., ARAD, M., GERULL, B. & FREIMARK, D. 2012. A novel titin mutation in adult-onset familial dilated cardiomyopathy. *Am J Cardiol,* 109, 1644-50.

YOUNG, P., EHLER, E. & GAUTEL, M. 2001. Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly. *J Cell Biol,* 154, 123-36.

ZARAGOZA, M. V., FASS, J., DIEGOLI, M., LIN, D. & ARBUSTINI, E. 2010. Mitochondrial DNA variant discovery and evaluation in human Cardiomyopathies through next-generation sequencing. *PLoS One,* 5, e12295.

ZASTROW, M. S., FLAHERTY, D. B., BENIAN, G. M. & WILSON, K. L. 2006. Nuclear titin interacts with A- and B-type lamins in vitro and in vivo. *J Cell Sci,* 119, 239-49.

ZHANG, J., LI, X., MUELLER, M., WANG, Y., ZONG, C., DENG, N., VONDRISKA, T. M., LIEM, D. A., YANG, J.-I., KORGE, P., HONDA, H., WEISS, J. N., APWEILER, R. & PING, P. 2008. Systematic characterization of the murine mitochondrial proteome using functionally validated cardiac mitochondria. *Proteomics,* 8, 1564-75.

ZHANG, R., YANG, J., ZHU, J. & XU, X. 2009. Depletion of zebrafish Tcap leads to muscular dystrophy via disrupting sarcomere-membrane interaction, not sarcomere assembly. *Hum Mol Genet,* 18, 4130-40.

ZHOU, D., BIRKENMEIER, C. S., WILLIAMS, M. W., SHARP, J. J., BARKER, J. E. & BLOCH, R. J. 1997. Small, membrane-bound, alternatively spliced forms of ankyrin 1 associated with the sarcoplasmic reticulum of mammalian skeletal muscle. *J Cell Biol,* 136, 621-31.

ZIFA, E., GIANNOULI, S., THEOTOKIS, P., STAMATIS, C., MAMURIS, Z. & STATHOPOULOS, C. 2007. Mitochondrial tRNA mutations: clinical and functional perturbations. *RNA Biol,* 4, 38-66.

ZOU, P., GAUTEL, M., GEERLOF, A., WILMANNS, M., KOCH, M. H. & SVERGUN, D. I. 2003. Solution scattering suggests cross-linking function of telethonin in the complex with titin. *J Biol Chem,* 278, 2636-44.

ZOU, P., PINOTSIS, N., LANGE, S., SONG, Y. H., POPOV, A., MAVRIDIS, I., MAYANS, O. M., GAUTEL, M. & WILMANNS, M. 2006. Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. *Nature,* 439, 229-33.

Appendix

I. Study Documents

a. *Patient information sheet (main form for adult patients)*

Oxford Radcliffe Hospitals NHS
NHS Trust
UNIVERSITY OF OXFORD
DEPARTMENT OF CARDIOVASCULAR MEDICINE
LEVEL 6 WEST WING, JOHN RADCLIFFE HOSPITAL, OXFORD OX3 9DU

TEL: 01865 234676 MOBILE: 07812 391582 E-MAIL: liz.ormondroyd@cardiov.ox.ac.uk

PARTICIPANT INFORMATION SHEET

Study name: Molecular Genetic Studies of Individuals and Families at Risk of Inherited Cardiac Disease

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study.

Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

PART 1

What is an inherited heart condition?
In some cases the heart loses its efficiency as a pump. In other cases, the heart rhythm is disturbed, leading to irregular heart beats, called arrhythmias. These effects can be caused by a change, like a spelling mistake, in one of the genes that controls heart development.

How is inherited heart disease different from other heart disorders?
- Inherited heart disease can, and often does, occur in the young.
- Various effective treatments are available for inherited heart disease.
- Sometimes several people in the family have the condition, but not everyone is affected in the same way.

What is the purpose of the study?
To identify new genetic changes (mutations) that might cause inherited heart conditions. If we find a gene change we will then analyse it to find out how it is inherited and what effect it might have on the function of the heart.

Why have I been invited?
You have been chosen because you have been diagnosed with an inherited heart disorder which may be caused by a change in a gene involved in heart function. We would like to look for the gene change that would explain why you developed the heart condition.

Do I have to take part?
It is up to you to decide. We will describe the study, go through this information sheet and discuss what it would mean if we found a gene change that would explain your heart condition. We will then ask you to sign a consent form to show you have agreed to take part. You are free to withdraw at any time, without giving a reason. This would not affect the standard of care you receive.

What would happen to me if I take part?
If you have already given a blood sample for genetic testing, we would ask your permission to get some of the sample from the laboratory where it is stored. If you have not already given a blood sample you would be asked to provide one (this can be done at your GP's surgery or next out patient appointment) which would be sent to us. The study co-ordinator would ask you about your family history, and for your permission to look at your genetic and/or cardiology hospital notes. We may ask you if any other members of your family would like to take part, because samples from other family members can help to identify the gene change that is causing the condition in your family.

If we find a gene change that explains the heart condition, we would write to your hospital consultant and they would inform you, if you wish.

What would I have to do?
Our study co-ordinator, Liz Ormondroyd will contact you to discuss the study. If we need to get a blood sample we will arrange with you the most convenient way of doing this. We would not require you to have any extra tests. We may ask you to talk to your relatives about taking part in the study, although you could still take part even if they don't want to.

What is being tested?
We would extract DNA (the genetic material) from your sample and then analyse the DNA in detail. This is a long process which might takes several months or years, and there is no guarantee that we would find a gene change. If we found a gene change that would account for the heart condition, we would tell your hospital doctor who will inform you and offer counseling support.

Are there any possible risks from taking part?
It is possible that, if a definite genetic cause for your heart condition were found, you might find this upsetting. We will discuss this fully with you with you before you decide whether to take part. Genetic counselling support is also available at your genetics referral centre.

What are the possible benefits?
At present we have only limited knowledge about how individual gene changes work and what kind of heart problems they might cause. We hope that in the future this knowledge will allow us to tailor advice and treatment to individual patients, depending on the specific gene change they have. The information generated may also be of benefit to other family members who know they are affected with the heart condition, or who wish to know whether they should also be seen by a heart doctor.

What happens when the research study stops?
We will continue to look for a gene change to explain your heart condition over the period of the study, which is 5 years.

Will my taking part in the study be kept confidential?
Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. Your GP will be informed only if you wish.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

PART 2

What will happen if I don't want to carry on with the study?
You are free to withdraw from the study at any time. We can ensure that if your samples and information are used for future research it is entirely anonymously, or we can destroy any identifiable samples or information we hold about you.

What if I have a complaint?
If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions (01865 234676).
You may also contact the University of Oxford Clinical Trials and Research Governance office on 01865 743002.

What if something goes wrong?
Compensation for harm arising from an accidental injury and occurring as a consequence of your participation in the study will be covered by the University of Oxford. If you are harmed and this is due to someone's negligence then you may have grounds for legal action for compensation against the University of Oxford (in respect of any harm arising out of the participation in the Clinical Study) or the NHS (in respect of any harm which has resulted from the clinical procedure being undertaken).

Will my taking part in the study be kept confidential?
If you take part in the study, this will be indicated on your hospital medical records and your GP would also be informed if you agree to this. Some parts of your medical records and the data collected from the study would be looked at by authorised persons from the University of Oxford, to check that the study is being carried out correctly. They may also be looked at by authorised persons from the NHS Trust. All investigators have a duty of confidentiality to you as a research participant and nothing that could reveal your identity would be disclosed outside the research site. The data collected from the study will be recorded anonymously and you would not be identifiable from this.

What will happen to any samples I give?
All samples will be retained in a secure environment for future analysis and will be stored in an anonymous format with a link to your identity that will be known only to the study co-ordinator and the doctor in charge of the study, at the University of Oxford under the custodianship of the Department of Cardiovascular Medicine. After the study, samples will either be destroyed or further ethical approval will be obtained for future studies.

Will any other genetic tests be done?
No, we will only look at genes that might cause a heart problem.

What will happen to the results of the research study?
We anticipate that the results will be published in a scientific journal for the benefit of the wider medical community and other people with a similar heart condition. However, individual participants will not be identified in any publication and your personal and clinical details will remain strictly confidential. Any scientific publications arising from the study will be available on request to all participants. You would have no legal right to a share of any profits that may arise from the research.

Who is organising and funding the research?
The investigators are Professor Hugh Watkins (Oxford University) and Dr Edward Blair (Churchill Hospital, Oxford). The research is funded by Oxford Biomedical Research Centre and the British Heart Foundation.

Who has reviewed the study?
The Oxfordshire B Research Ethics Committee has reviewed and approved the study. The reference number is: 09/H0605/3.

Where can I find independent information about taking part in research?
You can contact local branches of the NHS Patient Advisory Liaison Service (PALS). Here is their website: http://www.pals.nhs.uk/

Further information and contact details
If you wish to know more about any aspect of the study, please contact Liz on 01865 234676, email: liz.ormondroyd@cardiov.ox.ac.uk.

Thank you for reading this information sheet.

b. Consent form (main form for adult patients)



UNIVERSITY OF OXFORD
DEPARTMENT OF CARDIOVASCULAR MEDICINE

Oxford Radcliffe Hospitals NHS

CONSENT FORM (Version 2, 21.01.09)

Study Full Title	Molecular Genetic Studies of Individuals and Families at Risk of Inherited Cardiac Disease
Patient ID	
Researchers	Professor Hugh Watkins, Dr Edward Blair

Please initial box

1. I confirm that I have read and understand the information sheet (Version 2, 21.01.09) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by authorized individuals from the University of Oxford, or from the NHS Trust, where it is relevant to my taking part in this research. I permit these individuals access to my records. ☐
4. I understand that samples I have given previously for DNA analysis AND/OR new blood will be obtained during the procedure involved in this study. Samples will be considered a gift to the University of Oxford. I give permission for this. ☐
5. I understand that my DNA will be stored by the Department of Cardiovascular Medicine for future use, and that it may be used in any future ethically approved studies in inherited heart disease. ☐
6. I agree to my GP being informed of my participation in the study. ☐
7. If under the care of a Clinical Geneticist and/or Cardiologist I agree to him/her/them being informed of my participation in, and any results of this research. ☐
8. I wish to be informed of any results of this research. ☐
9. In view of the long term nature of the study, in the event of results becoming available after my death, individuals that I would like to be informed are listed overleaf, with contact details. ☐
10. I agree to take part in the above study. ☐

Name of participant	Signature	Date
Name of person taking consent	Signature	Date

PLEASE TURN OVER

When completed, 1 for patient; 1 for researcher site file; 1 (original) to be kept in medical notes

DOB:

Address:

Cardiologist Name (if known):
Hospital:

Geneticist Name (if known):
Hospital:

Next of kin name	Address	Relationship

Please Return Completed Form to:

Dr Rob Hastings
Department of Cardiovascular Medicine
Level 6, West Wing
John Radcliffe Hospital
Oxford
OX3 9DU

II. Experimental Protocols and Information

Table of manufactured kits and laboratory equipment used in titin studies:

	Equipment	Manufacturer
Kits	Phusion High Fidelity PCR Kit	NEB
	Protein Thermal Shift Dye Kit	Applied Biosystems
	Qiaquick Gel Extraction Kit	Qiagen
	Qiaprep Spin Miniprep Kit	Qiagen
	Qiaprep Spin Midiprep Kit	Qiagen
	Quickchange II XL site-directed mutagenesis kit	Agilent
Equipment	Innova 44 incubator-shaker	New Brunswick Scientific
	Heraeus Pico 17 Eppendorf centrifuge	ThermoScientific
	Rotanta 460RTable centrifuge	Hettich
	Avanti J26 XP Ultracentrifuge	Beckman-Coulter
	Sonopuls sonicator	Bandolin
	ChemiDoc XRS system	BioRad
	Tetrad 2 thermal cycler	BioRad
	UV-sensitive film imager	Konica-Minolta
	Ultrospec 2000 UV spectrophotometer	Pharmacia Biotech
	iBlot blotting system	Invitrogen
	Chirascan Circular Dichroism spectrometer	AppliedPhotophysics
	Hiload 16/60 superdex 75 prep grade column	Amersham Bioscience
	FPLC system	Äkta
	Mini Protean 3 electrophoresis system and PowerPac basic	BioRad
	StepOne real-time PCR system	Applied Biosystems
	Waterbath	Grant
	Roller mixer	Stuart
	Rotamax 120 rotator shaker	Heidolph
	Balance	Denver Instument
	Innova CO-170 incubator	New Brunswick Scientific
	DMIL light microscope	Leica
	Tissue culture hood	Scanlaf
	SP5 confocal microscope	Leica

Table of buffers used in titin studies:

Buffer	Constituents	pH
CD spectroscopy buffer	15 mM potassium phosphate buffer	7.4
GST pull-down buffer	10 mM Tris, 80 mM NaCl and 0.5% NP40	7.4
Native anode buffer	12.1 g Tris per litre H_2O	7.8
Native cathode buffer	6.4 g Tris, 5.1 g glycine per litre H_2O	8.9
Native sample buffer	4 ml 100% Glycerol, 2.5 ml 1 M Tris-HCl pH 6.8, 150µl of 1% bromophenol blue dissolved in 100% Ethanol, 3.35ml H_2O	6.8
Nickel-NTA buffer	50 mM NaH_2PO_4 and 300 mM NaCl	8.0
PBS	0.137 M NaCl and 0.05 M NaH_2PO_4	7.4
PBST	PBST (PBS with 0.1% Tween 20)	
PFA	4% formaldehyde in PBS	-
SDS-PAGE running buffer	25 mM Tris, 192 mM glycine, 0.1% SDS	-
SDS-PAGE sample buffer	50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA, 0.02 % bromophenol blue	6.8
SOC	SOB made up to manufacturer's guidelines, plus 16ml glucose, 4ml 1 M $MgSO_4$, and 4ml 1 M $MgCl_2$ per 800ml	-
TBE	National Diagnostics (0.089 M Tris-Borate pH 8.3 and 2 mM Na_2EDTA)	8.3
TBST	20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween 20	7.5
TE	10 mM Tris, 1 mM EDTA	8.0

a. Transforming plasmid constructs into XL10 Gold Ultracompetent Cells

- Pre-chill 14 ml BD Falcon polypropylene round bottom tubes on ice and pre-warm water bath (42°C)
- Bunsen burner on to create laminar air flow
- Add 4 µl 2-mercaptoethanol to XL10 Gold bacteria aliquot
- Divide bacteria between Falcon tubes
- Incubate on ice for 10 minutes
- Add 7.5 µl of plasmid with ligated product
- Incubate on ice for 30 minutes
- Place tubes in 42°C water bath for 30 seconds
- Incubate on ice for 5 minutes
- Add 0.9 ml warm SOC medium
- Incubate at 37°C for 1 hour shaking at 225-250 rpm
- Transfer to 1.5 ml Eppendorf tubes and centrifuge at <5000 rpm for 3 minutes

- Remove most of the supernatant and resuspend pellet
- Grow on antibiotic infused agar plates overnight at 37°C, place 10% on one side of plate and 90% on the other side of the plate, distribute with loop spreader; then transfer to cold room
- Transfer a few individual colonies to Falcon tube with 3 ml of LB culture medium and 3 µl of appropriate antibiotic using a yellow pipette tip; 37°C shaking at 225-250 rpm overnight
- Transforming Plasmid Constructs into E. coli BL21 Competent Cells
- Pre-chill 1.5 ml Eppendorfs on ice and pre-warm water bath (42°C)
- Bunsen burner on to create laminar air flow
- Divide BL21 bacteria between Eppendorfs (10-20 µl per tube)
- Add 1-2 µl of plasmid with ligated product
- Incubate on ice for 20-30 minutes
- Place tubes in 42°C water bath for 30 seconds
- Incubate on ice for 5 minutes
- Add 100 µl warm SOC medium
- Incubate at 37°C for 45-60 minutes shaking at 225-250 rpm
- Grow on Chloramphenicol and ampicillin-agar plates overnight at 37°C, place 10% on one side of plate and 90% on the other side of the plate, distribute with loop spreader; then transfer to cold room

b. Test bacterial protein expression

- Overnight culture of individual colonies picked using yellow pipette tip, add to 3 ml LB medium and 3 µl chloramphenicol and ampicillin
- Incubate overnight at 37°C shaking at 250 rpm
- Also make up equivalent media but without colonies added and keep in cold room overnight
- Next day add 1 ml of the overnight culture to the 3 ml media made up the day before (store the rest of the overnight culture in the cold room)
- Incubate for 1 hour at 37°C shaking at 250 rpm
- Remove 500µl to use as control
- Add IPTG to a concentration of 0.4 mM (kept at 0.4 M so for example 3.5 µl in 3.5 ml), which triggers protein expression
- Incubate for 3 hours at 37°C shaking at 250 rpm

- Remove 500 µl to test (throw away the rest)
- With the two 500 µl samples centrifuge at 6000 rpm for 5 minutes
- Remove supernatant and dry inverted
- Add 100 µl 1xSDS-SB and resuspend pellet
- Boil on heat block at 100^{o}C for 3 minutes and sonicate
- Run on 13% gel (SDS PAGE- resolving gel- water 1.27 ml, TGB 1 ml, 30% AA 1.73 ml, Temed 4 µl, APS 10% 40 µl)
- Coomassie stain (put in stain box on shaker for 15 minutes, wash with water and put in destain for several hours/overnight, wash with water and look at on lightbox/computer scan)

c. Scaled-up bacterial protein expression

- Make up SOB medium (800 ml per bottle) and leave out on bench overnight
- Overnight culture of 3 individual colonies picked using yellow pipette tip, add to 3 ml LB medium with 3 µl chloramphenicol and ampicillin
- If mini-protein expression has been performed take 3 x 3 µl of overnight culture rather than individual colony
- Incubate overnight at 37^{o}C shaking at 25 rpm
- Convert SOB to SOC by adding 16 ml glucose, 4 ml 1 M $MgSO_4$, and 4 ml 1 M $MgCl_2$ per 800 ml
- Add 800 µl chloramphenicol and ampicillin to each bottle
- For each expression put 200 ml SOC in a medium sized (500 ml) autoclaved plastic conical flask and 400 ml in a large (2l) flask
- Take 1-2 ml SOC to use as a blank when doing optical density
- Add 3 ml overnight culture to smaller flask and 6 ml to larger flask
- Incubate in shaker at 37^{o}C at 250 rpm until optical density is 0.8-1.0 measured on UV spectrophotometer at a wavelength of 600 nm
- Optical density measured, blank using SOC control to zero. Measure first after about 2 hours. Likely to reach correct OD after about 3+ hours. Increase of about 0.3 every 30 minutes as a rule of thumb.
- Once correct optical density reached add IPTG to 0.5 mM (250 µl in 200 ml culture)
- Incubate at low temperature (18^{o}C typically used) on shaker at 250 rpm for 4 hours. Keep temp down by adding 2 2l Erlenmeyer flasks with ice, and replace when melted.

- Spin larger flasks down into larger plastic screw-top flasks in large centrifuge cooled to 4°C at 5000 rpm for 12 minutes. Smaller flasks into 2x 50 ml Falcon in chilled usual centrifuge at max for 12 minutes (refill each falcon with culture). Remove supernatant back into a flask and dry. Freeze at -20°C.
- Run 500 ml sample on gel and Coomassie stain to ensure correct expression

d. Nickel-NTA purification of hexa-histidine tagged proteins

- Uses proteins expressed in E. coli BL21 cells (pellet stored at -20°C)
- Cool ultracentrifuge to 4°C and work on ice
- Prepare buffers (stored in cold room)- stock buffers are Nickel-NTA buffer (50 mM NaH_2PO_4 and 300 mM NaCl at pH 8.0) with either no imidazole or 500 mM imidazole (imidazole competes with protein binding), mixing these two can give different required concentrations of imidazole
- Prepare Ni-NTA wash (20 mM imidazole)
- Resuspend pellets after thawing for 15 minutes on ice
- falcon tubes in 5 ml Ni-NTA buffer with 10 mM imidazole and protein inhibitor tablet (EDTA free, 1 per 10 ml)
- Add lysozyme to a final concentration of 1 mg lysozyme per ml (0.5 ml)
- If viscous add 5 µg/ml DNase (2.5 µl)
- Incubate for 30 minutes on ice
- Set up chromatography column, wash with water then PBS buffer
- Sonicate samples (60% power, door closed, sonicator not touching bottom or sides, three bursts of 10 seconds) still in falcon tubes then transfer to screw-top small ultracentrifuge tubes
- Centrifuge lysate at 20 000 rpm JA25.50 4°C for 20 minutes
- During this time make up buffers for chromatography
- 1 litre filtered MQ water and 1 litre filtered PBS
- Filter using vacuum filter next to bench
- Put 1 ml Ni-NTA agarose beads washed (with Ni-NTA wash) in small falcon tubes using disposable pipette dropper or pipette with tip cut off and add supernatant
- Incubate for 1 hour on rotating shaker in cold room
- Wash column (funnel shaped) with washing buffer
- Centrifuge bead/supernatant mix at 1500 rpm for 3 minutes at 4°C and remove supernatant

- Wash beads twice with 10 ml washing buffer and centrifuge as above
- Resuspend beads in 500 µl of washing buffer and pour suspension into columns (filters inserted to hold beads), washing out falcon with 5-10 ml of wash
- Once all of the wash has run through the column elute the protein with 5x250 µl of elution buffer (Ni-NTA buffer with 250 mM imidazole), and collect the fractions in labelled Eppendorfs
- Store protein solutions on ice
- Spot 1 µl of the elution fraction onto nitrocellulose and stain with Ponceau for 5 minutes, destain with water and dry the membrane
- Aggregate those protein solutions with significant protein content (if 5th blot still has significant protein consider further elution step)
- Keep aggregated protein solutions on ice in cold room
- Store beads in column (stoppered bottom the top), washed with water then stored in 20% ethanol

e. *Gel filtration chromatography*

- Amersham Biosciences® Hiload 16/60 superdex 75 prep grade column was used on an Äkta FPLC system
- Wash ethanol from column using filtered water over at least two hours
- Wash chromatography column with PBS buffer slowly overnight at 0.15 ml/min
- Centrifuge protein solutions at max rate for 10 minutes in cold room
- Transfer supernatant to flat bottomed small tubes
- Run chromatography column
- Set up collecting tubes
- Autozero UV
- Increase flow-rate to 1 ml/min (max pressure 0.33 MPa)
- Ensure valve is on load and flush 2ml loop with 2 volumes worth of buffer
- inject sample into loop and flush through dead space with buffer
- Change valve to inject
- Set up Manual peak fractionation to 12 mm tubes, row by row, 1 ml, collect from first tube and insert (not execute)
- Set peak fractionation parameters UV to level and 30mAU then insert then execute
- After about 5 minutes return valve setting to load
- Between samples can do peak frac stop and then reinsert parameters

- To store run through column volume of water (120 ml) then 20% ethanol for storage

f. Native-PAGE

- 8% Resolving gel (x2):
- 4.8 ml H_2O, 2.6 ml 30% acrylamide, 2.6 ml 1.5M Tris pH 8.8, 120 µl of ammonium persulphate (APS) and 8ul of tetramethylethylenediamine (Temed)
- Gels cast on BioRad® casting plates (using a dropping pipette, a layer of butanol, and allow to set for approximately 20 minutes.
- Pour off butanol from top of the running gel and add stacking gel
- Stacking gel:
- 2.8 ml H_2O, 666 µl 30% Acrylamide, 250 µl 0.5M Tris pH 6.8, 50 µl APS and 6µl Temed
- Place comb into the wells and leave to set for approximately 40 minutes
- Load samples using native loading buffer
- Native buffer (stock):
- 4 ml 100% Glycerol, 2.5 ml 1M Tris-HCl pH 6.8, 150 µl of 1% bromophenol blue dissolved in 100% Ethanol, 3.35 ml H_2O
- Marker used: NativeMark™ (Invitrogen®)
- Run in BioRad system at constant 25mA
- Buffers:
- Cathode (inner chamber) - 6.4 g Tris, 5.1 g glycine per litre H_2O, pH 8.9
- Anode (outer chamber) - 12.1 g Tris per litre H_2O, pH 7.8

g. SDS-PAGE

- 12% Resolving gel (x2):
- 2.6 ml H_2O, 2 ml Protogel® Resolving Buffer (1.5M Tris-HCl, 0.4% SDS, pH8.8; National Diagnostics®), 3.2 ml 30% acrylamide (National Diagnostics®), 8µl TEMED, and 80µl APS
- Gels cast on BioRad plates, layer of isopropanol, and leave to set (~20 minutes)
- Stacking gel:
- 2.6 ml H_2O, 1 ml Protogel® Stacking Buffer (0.5M Tris-HCl, 0.4% SDS, pH6.8; National Diagnostics®), 0.4 ml 30% acrylamide, 8µl TEMED and 80µl APS
- Insert combs and leave to set (~40 minutes)

- Boil samples in SDS-buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA, 0.02 % bromophenol blue) at 100^{o}C for 5 minutes
- Marker used: Precision plus™ all blue marker (BioRad®)
- Run in BioRad system (80mV for stacking, 120mV for resolving)
- SDS-PAGE Running Buffer: 25 mM Tris, 192 mM glycine, 0.1% SDS

h. PACT suite buffers for protein stability optimisation

1 0.1M SPG buffer pH 4 25% w/v PEG 1500 135701
2 0.1M SPG buffer pH 5 25% w/v PEG 1500 135702
3 0.1M SPG buffer pH 6 25% w/v PEG 1500 135703
4 0.1M SPG buffer pH 7 25% w/v PEG 1500 135704
5 0.1M SPG buffer pH 8 25% w/v PEG 1500 135705
6 0.1M SPG buffer pH 9 25% w/v PEG 1500 135706
7 0.2M Sodium chloride 0.1M Sodium acetate pH 5 20% w/v PEG 6000 135707
8 0.2M Ammonium chloride 0.1M Sodium acetate pH 5 20% w/v PEG 6000 135708
9 0.2M Lithium chloride 0.1M Sodium acetate pH 5 20% w/v PEG 6000 135709
10 0.2M Magnesium chloride 0.1M Sodium acetate pH 5 20% w/v PEG 6000 135710
11 0.2M Calcium chloride 0.1M Sodium acetate pH 5 20% w/v PEG 6000 135711
12 0.01M Zinc chloride 0.1M Sodium acetate pH 5 20% w/v PEG 6000 135712
13 0.1M MIB buffer pH 4 25% w/v PEG 1500 135713
14 0.1M MIB buffer pH 5 25% w/v PEG 1500 135714
15 0.1M MIB buffer pH 6 25% w/v PEG 1500 135715
16 0.1M MIB buffer pH 7 25% w/v PEG 1500 135716
17 0.1M MIB buffer pH 8 25% w/v PEG 1500 135717
18 0.1M MIB buffer pH 9 25% w/v PEG 1500 135718
19 0.2M Sodium chloride 0.1M MES pH 6 20% w/v PEG 6000 135719
20 0.2M Ammonium chloride 0.1M MES pH 6 20% w/v PEG 6000 135720
21 0.2M Lithium chloride 0.1M MES pH 6 20% w/v PEG 6000 135721
22 0.2M Magnesium chloride 0.1M MES pH 6 20% w/v PEG 6000 135722
23 0.2M Calcium chloride 0.1M MES pH 6 20% w/v PEG 6000 135723
24 0.01M Zinc chloride 0.1M MES pH 6 20% w/v PEG 6000 135724
25 0.1M PCB buffer pH 4 25% w/v PEG 1500 135725
26 0.1M PCB buffer pH 5 25% w/v PEG 1500 135726
27 0.1M PCB buffer pH 6 25% w/v PEG 1500 135727
28 0.1M PCB buffer pH 7 25% w/v PEG 1500 135728
29 0.1M PCB buffer pH 8 25% w/v PEG 1500 135729
30 0.1M PCB buffer pH 9 25% w/v PEG 1500 135730
31 0.2M Sodium chloride 0.1M Hepes pH 7 20% w/v PEG 6000 135731
32 0.2M Ammonium chloride 0.1M Hepes pH 7 20% w/v PEG 6000 135732
33 0.2M Lithium chloride 0.1M Hepes pH 7 20% w/v PEG 6000 135733
34 0.2M Magnesium chloride 0.1M Hepes pH 7 20% w/v PEG 6000 135734
35 0.2M Calcium chloride 0.1M Hepes pH 7 20% w/v PEG 6000 135735
36 0.01M Zinc chloride 0.1M Hepes pH 7 20% w/v PEG 6000 135736
37 0.1M MMT buffer pH 4 25% w/v PEG 1500 135737
38 0.1M MMT buffer pH 5 25% w/v PEG 1500 135738
39 0.1M MMT buffer pH 6 25% w/v PEG 1500 135739
40 0.1M MMT buffer pH 7 25% w/v PEG 1500 135740
41 0.1M MMT buffer pH 8 25% w/v PEG 1500 135741
42 0.1M MMT buffer pH 9 25% w/v PEG 1500 135742
43 0.2M Sodium chloride 0.1M Tris pH 8 20% w/v PEG 6000 135743
44 0.2M Ammonium chloride 0.1M Tris pH 8 20% w/v PEG 6000 135744
45 0.2M Lithium chloride 0.1M Tris pH 8 20% w/v PEG 6000 135745

46 0.2M Magnesium chloride 0.1M Tris pH 8 20% w/v PEG 6000 135746
47 0.2M Calcium chloride 0.1M Tris pH 8 20% w/v PEG 6000 135747
48 0.01M Zinc chloride 0.1M Tris pH 8 20% w/v PEG 6000 135748
49 0.2M Sodium fluoride 20% w/v PEG 3350 135749
50 0.2M Sodium bromide 20% w/v PEG 3350 135750
51 0.2M Sodium iodide 20% w/v PEG 3350 135751
52 0.2M Potassium thiocyanate 20% w/v PEG 3350 135752
53 0.2M Sodium nitrate 20% w/v PEG 3350 135753
54 0.2M Sodium formate 20% w/v PEG 3350 135754
55 0.2M Sodium acetate 20% w/v PEG 3350 135755
56 0.2M Sodium sulphate 20% w/v PEG 3350 135756
57 0.2M Potassium/sodium tartrate 20% w/v PEG 3350 135757
58 0.2M Sodium/potassium phosphate 20% w/v PEG 3350 135758
59 0.2M Sodium citrate 20% w/v PEG 3350 135759
60 0.2M Sodium malonate 20% w/v PEG 3350 135760
61 0.2M Sodium fluoride 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135761
62 0.2M Sodium bromide 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135762
63 0.2M Sodium iodide 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135763
64 0.2M Potassium thiocyanate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135764
65 0.2M Sodium nitrate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135765
66 0.2M Sodium formate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135766
67 0.2M Sodium acetate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135767
68 0.2M Sodium sulphate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135768
69 0.2M Potassium/sodium tartrate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135769
70 0.2M Sodium/potassium phosphate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135770
71 0.2M Sodium citrate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135771
72 0.2M Sodium malonate 0.1M Bis Tris propane pH 6.5 20% w/v PEG 3350 135772
73 0.2M Sodium fluoride 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135773
74 0.2M Sodium bromide 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135774
75 0.2M Sodium iodide 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135775
76 0.2M Potassium thiocyanate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135776
77 0.2M Sodium nitrate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135777
78 0.2M Sodium formate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135778
79 0.2M Sodium acetate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135779
80 0.2M Sodium sulphate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135780
81 0.2M Potassium/sodium tartarte 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135781
82 0.2M Sodium/potassium phosphate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135782
83 0.2M Sodium citrate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135783
84 0.2M Sodium malonate 0.1M Bis Tris propane pH 7.5 20% w/v PEG 3350 135784
85 0.2M Sodium fluoride 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135785
86 0.2M Sodium bromide 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135786
87 0.2M Sodium iodide 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135787
88 0.2M Potassium thiocyanate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135788
89 0.2M Sodium nitrate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135789
90 0.2M Sodium formate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135790
91 0.2M Sodium acetate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135791
92 0.2M Sodium sulphate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135792
93 0.2M Potassium/sodium tartrate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135793
94 0.2M Sodium/potassium phosphate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135794
95 0.2M Sodium citrate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135795
96 0.2M Sodium malonate 0.1M Bis Tris propane pH 8.5 20% w/v PEG 3350 135796

i. GST pull-downs using mammalian expressed fusion proteins

- Resuspend pellet in 500 µl GST buffer
- Incubate on rocker on ice horizontal for 20 minutes
- Centrifuge at max speed for 15 minutes at 4^{o}c

- Transfer supernatant to new Eppendorf
- Lysate control- 40 µl in to separate 500 µl Eppendorf with 10 µl 5 fold SDS-SB
- Wash glutathione beads with GST buffer (equal quantities of beads and buffer), when pipetting beads use yellow tips with the ends cut off
- Add 30 µl of bead suspension to supernatant and incubate for 1 hour on ice on rocker
- During this time make up 2 SDS-PAGE gels as per protocol
- Centrifuge briefly to max speed for a few seconds and remove supernatant
- Wash beads 3 times with GST buffer, vortex and brief centrifuge between washes, aspirating off the supernatant leaving the beads undisturbed
- Remove final aspirate after a further centrifuge using loading pipette tips to completely remove liquid
- Add 25 µl SDS-SB buffer diluted to 2-fold to each sample
- Boil samples and lysate controls on heat block at 100^{o}c for 3 minutes and then sonicate, repeat brief centrifuge
- SDS-PAGE (as g)

III. Sanger Sequencing Primers

* All have M13 universal tags.

a. CACNA1C

CACNA1C_Exon1_F	GTAAAACGACGGCCAGTGGGGTGTTTTCACATTTCTTC
CACNA1C_Exon1_R	CAGGAAACAGCTATGACAGGAACCCGGTGGAAAAG
CACNA1C_Exon2_F	GTAAAACGACGGCCAGTTGAATCTGGGGTCCAGAGAG
CACNA1C_Exon2_R	CAGGAAACAGCTATGACGTAGCCCAGCACAGCTGAC
CACNA1C_Exon3_F	GTAAAACGACGGCCAGTTCAGAATACAAGGGGCTTGG
CACNA1C_Exon3_R	CAGGAAACAGCTATGACATGCCAAACACACGTCTCAC
CACNA1C_Exon4_F	GTAAAACGACGGCCAGTTCCAAACTTTCCAAATCCCC
CACNA1C_Exon4_R	CAGGAAACAGCTATGACTTAAGGAAAAGTCCCGCAAC
CACNA1C_Exon5_F	GTAAAACGACGGCCAGTGCTGGGCAAAAGAAAACC
CACNA1C_Exon5_R	CAGGAAACAGCTATGACAAGGACAGCAGAAGGTGGTG
CACNA1C_Exon6_F	GTAAAACGACGGCCAGTTTGCTGGAAGATTGCATGAG
CACNA1C_Exon6_R	CAGGAAACAGCTATGACAAGGTGCTGGGAGCGATAG
CACNA1C_Exon7_F	GTAAAACGACGGCCAGTATCCCATCAACCTCATCCTG
CACNA1C_Exon7_R	CAGGAAACAGCTATGACGGAGGAGAAAGGGAGGTGTC
CACNA1C_Exon8_F	GTAAAACGACGGCCAGTCTGCGTGGGTCAGTGTCTC

CACNA1C_Exon8_R	CAGGAAACAGCTATGACATCCCAGGTTCATCAGAGCC
CACNA1C_Exon9_F	GTAAAACGACGGCCAGTCTCGGTGCTCCCTCCTTC
CACNA1C_Exon9_R	CAGGAAACAGCTATGACGGTCAAAGTTCTGTGCTGACC
CACNA1C_Exon10_F	GTAAAACGACGGCCAGTTTCCCTTGTCTCCCCACC
CACNA1C_Exon10_R	CAGGAAACAGCTATGACCCACTGTGATTTCCAGATGC
CACNA1C_Exon11_F	GTAAAACGACGGCCAGTAAATTTCCCTGGGACTGTTG
CACNA1C_Exon11_R	CAGGAAACAGCTATGACGCTGAAGACAGAGCTGGTGG
CACNA1C_Exon12_F	GTAAAACGACGGCCAGTATCTTGGGTTGGAGGAAACC
CACNA1C_Exon12_R	CAGGAAACAGCTATGACGGTTATGCCCTCCCCTG
CACNA1C_Exon13_F	GTAAAACGACGGCCAGTTATTTCCTTTCCCTGCCTCC
CACNA1C_Exon13_R	CAGGAAACAGCTATGACTCCCTTGCCCTCTGGAAG
CACNA1C_Exon14_F	GTAAAACGACGGCCAGTTGGGGAGGACAGCAAGG
CACNA1C_Exon14_R	CAGGAAACAGCTATGACGGTAAAAGGGTGCAGCTTCC
CACNA1C_Exon15_F	GTAAAACGACGGCCAGTTTTTGGGTTTGCTGTTGGTC
CACNA1C_Exon15_R	CAGGAAACAGCTATGACCCGTTTGGCACTGTGCTG
CACNA1C_Exon16_F	GTAAAACGACGGCCAGTCCCACCAAAACCCCAAAC
CACNA1C_Exon16_R	CAGGAAACAGCTATGACGTCAAGCCTCTCTCTGCCAC
CACNA1C_Exon17_F	GTAAAACGACGGCCAGTTTCCCTCCTACACTGTTCCC
CACNA1C_Exon17_R	CAGGAAACAGCTATGACGAAGGCTGTGGCCCTCC
CACNA1C_Exon18_F	GTAAAACGACGGCCAGTCTTCAAGGCTACTGCAAGCC
CACNA1C_Exon18_R	CAGGAAACAGCTATGACTTTAGAATTTAGATCTCTGCCCC
CACNA1C_Exon19_F	GTAAAACGACGGCCAGTAAGTGGTAAAATAAGTGGGAGTGC
CACNA1C_Exon19_R	CAGGAAACAGCTATGACGGTTCAGGTAAAATACAGGCAAC
CACNA1C_Exon20_F	GTAAAACGACGGCCAGTACTGGTGCTTCCCCTTGTC
CACNA1C_Exon20_R	CAGGAAACAGCTATGACTGTCAGCAACAGAAGCTGTG
CACNA1C_Exon21_F	GTAAAACGACGGCCAGTCTCTTCCCGTCCTGCTTTTC
CACNA1C_Exon21_R	CAGGAAACAGCTATGACGACAGACAGCGGCAGAGAC
CACNA1C_Exon22_F	GTAAAACGACGGCCAGTGAGGGGTCTCTGGGCTAGG
CACNA1C_Exon22_R	CAGGAAACAGCTATGACGGTCTGGTTGTCCTCTCTGG
CACNA1C_Exon23_F	GTAAAACGACGGCCAGTTCACCTGTCAGGACATTCCC
CACNA1C_Exon23_R	CAGGAAACAGCTATGACTGACCATCATCTCCTTCCAAC
CACNA1C_Exon24_F	GTAAAACGACGGCCAGTCTACCTCCTGGAAAGGCTCC
CACNA1C_Exon24_R	CAGGAAACAGCTATGACTCCCCAGGAATGCTGATG
CACNA1C_Exon25_F	GTAAAACGACGGCCAGTGCTTTTGGATTGACTCATTGATTAC
CACNA1C_Exon25_R	CAGGAAACAGCTATGACCTTTCTGGTCAAGCCTCCAG
CACNA1C_Exon26_F	GTAAAACGACGGCCAGTCAGGCGTGAAGGAAGATGG
CACNA1C_Exon26_R	CAGGAAACAGCTATGACTAACAAAACTTGGGGATGGG
CACNA1C_Exon27_F	GTAAAACGACGGCCAGTATCCCTGGAGCAGTGGTG
CACNA1C_Exon27_R	CAGGAAACAGCTATGACCAGGCTGCCATTCCCTC
CACNA1C_Exon28_F	GTAAAACGACGGCCAGTACATCCCACACTTCTCTGCC
CACNA1C_Exon28_R	CAGGAAACAGCTATGACCCATGGCACACTCCCTTC

CACNA1C_Exon29_F	GTAAAACGACGGCCAGTGAGGAGCGACCTCCCTG
CACNA1C_Exon29_R	CAGGAAACAGCTATGACCTTCAATACCTGCACCCCAC
CACNA1C_Exon30_F	GTAAAACGACGGCCAGTCTGATGGTGGTGTTGCTCTG
CACNA1C_Exon30_R	CAGGAAACAGCTATGACAAAAGAAGGTAAACAGAGAGAGTGAG
CACNA1C_Exon31_F	GTAAAACGACGGCCAGTTTGTTTTGTCCTTTCTTGTTGG
CACNA1C_Exon31_R	CAGGAAACAGCTATGACAGGAAACAAACCGGGAAAAG
CACNA1C_Exon32_F	GTAAAACGACGGCCAGTTACCGGGCATCTTCATGG
CACNA1C_Exon32_R	CAGGAAACAGCTATGACTCTGGAGTTCGGGGTGG
CACNA1C_Exon33_F	GTAAAACGACGGCCAGTCCCTCCTGTTCTCACCCC
CACNA1C_Exon33_R	CAGGAAACAGCTATGACTGTGTCAGCTGTGTTCTGGG
CACNA1C_Exon34_F	GTAAAACGACGGCCAGTAGAGACAGGGATGCGGC
CACNA1C_Exon34_R	CAGGAAACAGCTATGACCTGGGGAGAGACAGAGCG
CACNA1C_Exon35_F	GTAAAACGACGGCCAGTATGGTGGGAAATTAGGACCC
CACNA1C_Exon35_R	CAGGAAACAGCTATGACATTCTACCACCAGGCCACAG
CACNA1C_Exon36_F	GTAAAACGACGGCCAGTCCCTCCTTGGCATCTGTAGG
CACNA1C_Exon36_R	CAGGAAACAGCTATGACTCAGACTGCCCTGGAACTG
CACNA1C_Exon37_F	GTAAAACGACGGCCAGTGAGGCAGGTGTGTAGGAAGG
CACNA1C_Exon37_R	CAGGAAACAGCTATGACCCTCCTGCAGGTATGGTCAG
CACNA1C_Exon38_F	GTAAAACGACGGCCAGTTGCTGGCAGAGACCGTG
CACNA1C_Exon38_R	CAGGAAACAGCTATGACGGATCACAAGAAATGGGCAC
CACNA1C_Exon39_F	GTAAAACGACGGCCAGTGTGTTCTGCGGTCCCCTAAG
CACNA1C_Exon39_R	CAGGAAACAGCTATGACCTGCCGAGCTCCTCTCTG
CACNA1C_Exon40_F	GTAAAACGACGGCCAGTCTTACGCAGAGGGACCCAG
CACNA1C_Exon40_R	CAGGAAACAGCTATGACAGCTGGAGGAGAGGCAGG
CACNA1C_Exon41_F	GTAAAACGACGGCCAGTTCAACTGAATTCCCCTGCTC
CACNA1C_Exon41_R	CAGGAAACAGCTATGACACTGGAGGTCAGAGGCAATG
CACNA1C_Exon42_F	GTAAAACGACGGCCAGTGGAGCTTGGAGGAAAGGG
CACNA1C_Exon42_R	CAGGAAACAGCTATGACGACGTTCCCCGCAAAAC
CACNA1C_Exon43_F	GTAAAACGACGGCCAGTCTGTGGAGGCTGCTCTCTG
CACNA1C_Exon43_R	CAGGAAACAGCTATGACCTGTCTCCTGAGGGTTGCAG
CACNA1C_Exon47_F	GTAAAACGACGGCCAGTAGGGGAAAATAATGGCATGG
CACNA1C_Exon47_R	CAGGAAACAGCTATGACCATGGCTTTCCGCACAC
CACNA1C_Exon48_F	GTAAAACGACGGCCAGTGGCTCCAGGAGTAATGTTCC
CACNA1C_Exon48_R	CAGGAAACAGCTATGACTCACGAGAACAGTGAGGCAC

b. CLIP1

CLIP1_Exon2_F	GTAAAACGACGGCCAGTTGGTTGGGGAAAGATAGTGG
CLIP1_Exon2_R	CAGGAAACAGCTATGACGGGTGACAGAGTTAGATTCCAT
CLIP1_Exon3_1_F	GTAAAACGACGGCCAGTCTACTCCTGTGGGCTGTTGG
CLIP1_Exon3_1_R	CAGGAAACAGCTATGACGTCTGCAGGCCATTAGCTTC
CLIP1_Exon3_2_F	GTAAAACGACGGCCAGTTGGCAGGAGTTCGGTATTTC

CLIP1_Exon3_2_R	CAGGAAACAGCTATGACCCACTCACTGGTGTTTATATCTCAG
CLIP1_Exon4_F	GTAAAACGACGGCCAGTGTAGGGACCACAAACGATTAAG
CLIP1_Exon4_R	CAGGAAACAGCTATGACCTTGGTGAGACTGGCCTGAG
CLIP1_Exon5_F	GTAAAACGACGGCCAGTGAGTTGGGAGAACCTTTTGG
CLIP1_Exon5_R	CAGGAAACAGCTATGACGGGCCTTAATCCTGCCTG
CLIP1_Exon6_F	GTAAAACGACGGCCAGTAAGCCTTTCCTAATCCTGGC
CLIP1_Exon6_R	CAGGAAACAGCTATGACAAGTCTAGCTCCTCGGTGCC
CLIP1_Exon7_F	GTAAAACGACGGCCAGTTTAGAGTAACGCCCACCCAC
CLIP1_Exon7_R	CAGGAAACAGCTATGACTTCCTTTGGTTCTAAGGGTTG
CLIP1_Exon8_F	GTAAAACGACGGCCAGTGCCTACCATGCCCAACG
CLIP1_Exon8_R	CAGGAAACAGCTATGACATTATTTCGCCCGTGTTCTG
CLIP1_Exon9_F	GTAAAACGACGGCCAGTCACATATTGAAGCTTTATTGGAATC
CLIP1_Exon9_R	CAGGAAACAGCTATGACGCATTTTAATCTGTGATCAATCTTTG
CLIP1_Exon10_F	GTAAAACGACGGCCAGTCGCTGTCTCTCCTCTCGTG
CLIP1_Exon10_R	CAGGAAACAGCTATGACGTGTCCACCCTTCACATGC
CLIP1_Exon11_1_F	GTAAAACGACGGCCAGTAGTTTGCTGCCTGAGATGG
CLIP1_Exon11_1_R	CAGGAAACAGCTATGACTCCATCTCTTTAGCATGGGC
CLIP1_Exon11_2_F	GTAAAACGACGGCCAGTAGGCGATGGAAGAACTGAAG
CLIP1_Exon11_2_R	CAGGAAACAGCTATGACTCAACTGGGGACATCAAGC
CLIP1_Exon12_F	GTAAAACGACGGCCAGTAGGGGAGACCCTGTGTCAG
CLIP1_Exon12_R	CAGGAAACAGCTATGACAAGCAGTGCGTTACCAAAGG
CLIP1_Exon13_F	GTAAAACGACGGCCAGTTGACCCACATGTAGTAATTGGTG
CLIP1_Exon13_R	CAGGAAACAGCTATGACACTCCACAGGTCAGCTCCAG
CLIP1_Exon14_F	GTAAAACGACGGCCAGTGGGTTGTAAGACAGCCAAGC
CLIP1_Exon14_R	CAGGAAACAGCTATGACTGTCTCGAATACGGGAAAGC
CLIP1_Exon15_F	GTAAAACGACGGCCAGTCCAGAGCAAGAATACACTCAGTC
CLIP1_Exon15_R	CAGGAAACAGCTATGACCTTGCCGCAGAGCCTAAC
CLIP1_Exon16&17_F	GTAAAACGACGGCCAGTAAAACAAAACAAAATTTCTTGCC
CLIP1_Exon16&17_R	CAGGAAACAGCTATGACTTTCTAAGCACCCTTCCTGC
CLIP1_Exon18_F	GTAAAACGACGGCCAGTTCCTCACCTAACGGTGTTGG
CLIP1_Exon18_R	CAGGAAACAGCTATGACCCTGAGTCAGTGACCAAACATTC
CLIP1_Exon19_F	GTAAAACGACGGCCAGTTTCATTCATGTCACTGGAAAATG
CLIP1_Exon19_R	CAGGAAACAGCTATGACATGTTGCTTCTGGGGTATGC
CLIP1_Exon20_F	GTAAAACGACGGCCAGTGTCTCCCACAGTTTCCTTGC
CLIP1_Exon20_R	CAGGAAACAGCTATGACCAACAGCAGCAGCACCC
CLIP1_Exon21_F	GTAAAACGACGGCCAGTTTTAGCTGTCCTGGAGCCTG
CLIP1_Exon21_R	CAGGAAACAGCTATGACTGAATCTAAGAAGGTCACTTGTTTTC
CLIP1_Exon22&23_F	GTAAAACGACGGCCAGTTGTACACTGACCAGTCTTTCGG
CLIP1_Exon22&23_R	CAGGAAACAGCTATGACTACTAGAAAGGGCTCCTCGG
CLIP1_Exon24_F	GTAAAACGACGGCCAGTTTTTGAACACAGGCCCTTTC
CLIP1_Exon24_R	CAGGAAACAGCTATGACCTAGAGGAATGCATTACCAATTC

CLIP1_Exon25_F	GTAAAACGACGGCCAGTATCTCAGCATGCCTGTCCTC
CLIP1_Exon25_R	CAGGAAACAGCTATGACTGATGAAACTAAGGTTCCAATCAG
CLIP1_Exon26_F	GTAAAACGACGGCCAGTTGCAAGTGTTCTTGTCTCTTCAG
CLIP1_Exon26_R	CAGGAAACAGCTATGACACGTTGTGTCAATGCGAGTG
CLIP1_Exon25_F	GTAAAACGACGGCCAGTTGCAAGTGTTCTTGTCTCTTCAG
CLIP1_Exon25_R	CAGGAAACAGCTATGACACGTTGTGTCAATGCGAGTG

c. *FGF14*

FGF14_Exon5_F	GTAAAACGACGGCCAGTCCATTCTCACACCTAACCAAGC
FGF14_Exon5_R	CAGGAAACAGCTATGACGGAGCAGGAATGTCTGGTG

d. *FLNC*

FLNC_Exon40_F	GTAAAACGACGGCCAGTTCCACCCTGCTTCCTCAC
FLNC_Exon40_R	CAGGAAACAGCTATGACCCAGCATCAAGCCTGGTC

e. *KCNJ8*

KCNJ8_Exon2_F	GTAAAACGACGGCCAGTGACGGAGAGGCAGGTGAG
KCNJ8_Exon2_R	CAGGAAACAGCTATGACGGATAAGAGAAAGGGCATTTTGAC
KCNJ8_Exon3_1_F	GTAAAACGACGGCCAGTAGCTCAGCTTGGCATATTTC
KCNJ8_Exon3_1_R	CAGGAAACAGCTATGACGGGTGATGCCAGTAGTTTCAACC
KCNJ8_Exon3_2_F	GTAAAACGACGGCCAGTAACCCAATCGAGAGCAATAAC
KCNJ8_Exon3_2_R	CAGGAAACAGCTATGACGAGTGCCCAGCATAAACCG

f. *MXRA8*

MXRA8_Exon1_F	GTAAAACGACGGCCAGTCTCTGCTCCCACACTGCC
MXRA8_Exon1_R	CAGGAAACAGCTATGACTGAAGACGTCACAGCCTCT
MXRA8_Exon2_F	GTAAAACGACGGCCAGTTGGGACTACAGGAGGGACTC
MXRA8_Exon2_R	CAGGAAACAGCTATGACCAAGATAGGAGGGGCTGAGG
MXRA8_Exon3_F	GTAAAACGACGGCCAGTCCTTACAGACATCCCTCCCC
MXRA8_Exon3_R	CAGGAAACAGCTATGACCGTCCGTCTCCTCCACC
MXRA8_Exon4_F	GTAAAACGACGGCCAGTACTTCTCGCTGCTCATCCG
MXRA8_Exon4_R	CAGGAAACAGCTATGACCTCGCCGTCCCAGTAGG
MXRA8_Exon5_F	GTAAAACGACGGCCAGTACGGCCGTGAGTGCTTC
MXRA8_Exon5_R	CAGGAAACAGCTATGACCCTCGGGCAAGGGACAG
MXRA8_Exon6-8_F	GTAAAACGACGGCCAGTGCTGTCCCTTGCCCGAG
MXRA8_Exon6-8_R	CAGGAAACAGCTATGACTGTAATCTGTGGGAGGAGAGG
MXRA8_Exon9_F	GTAAAACGACGGCCAGTGGGCTGGGAAGTGCATC
MXRA8_Exon9_R	CAGGAAACAGCTATGACCATCAGGAGATGCCCCG
MXRA8_Exon10_F	GTAAAACGACGGCCAGTCTCACCTGGGACCAACGAT
MXRA8_Exon10_R	CAGGAAACAGCTATGACGCTATGCTGCCACCAAGC

g. PDP2

PDP2_Exon2_1F	5′-GTAAAACGACGGCCAGTGAATGGCATGAACCTGGG
PDP2_Exon2_1R	5′-CAGGAAACAGCTATGACAATCTTGTGGGTTGTCTCGC
PDP2_Exon2_2F	5′-GTAAAACGACGGCCAGTTGTCAAGTACTGTGTCCTACTG
PDP2_Exon2_2R	5′-CAGGAAACAGCTATGACTCATGCTTTCCATAGCTCCC
PDP2_Exon2_3F	5′-GTAAAACGACGGCCAGTAATTCCCCAGTGGAGGACC
PDP2_Exon2_3R	5′-CAGGAAACAGCTATGACAGTCACCAGCATTTGCCAC
PDP2_Exon2_4F	5′-GTAAAACGACGGCCAGTCTGGAAGATGAGGTGACAAGG
PDP2_Exon2_4R	5′-CAGGAAACAGCTATGACATCTGAGGCCAGCACAAGG
PDP2_Exon2_5F	5′-GTAAAACGACGGCCAGTAGGCCCTCAACATTTACCAG
PDP2_Exon2_5R	5′-CAGGAAACAGCTATGACTGCGCTTTTAATAGTAAGGTTTG

h. TTN

TTN_Exon2_F	5′-GTAAAACGACGGCCAGTGACCCTTAGCTGGGACACC
TTN_Exon2_R	5′-CAGGAAACAGCTATGACGTCAAGTCCTGCAGCAACG
TTN_Exon3_F	5′-GTAAAACGACGGCCAGTCAAGGCAGTGAACATGATGG
TTN_Exon3_R	5′-CAGGAAACAGCTATGACTTTGCTGGAGATGTCCTCTG
TTN_Exon4_F	5′-GTAAAACGACGGCCAGTTCACCTGGTTTTGGAATTGG
TTN_Exon4_R	5′-CAGGAAACAGCTATGACGGCCCCATTTAGACACAAAC

IV. Supplementary Results

a. CD spectroscopy deconvolutions of purified titin fractions at room temperature using algorithms available on the Dichoweb online analysis tool

		Wild-Type	A178D Fraction 1	A178D Fraction 2
SELCON3	*Helix 1*	-0.001	0.038	0.065
	Helix 2	-0.026	0.059	0.073
	Strand 1	0.123	0.214	0.06
	Strand 2	0.134	0.124	0.09
	Turns	0.218	0.22	0.209
	Unordered	0.587	0.331	0.508
CONTIN	*Helix 1*	0	0.022	0
	Helix 2	0	0.071	0.008
	Strand 1	0.298	0.215	0.246
	Strand 2	0.142	0.128	0.129
	Turns	0.138	0.237	0.215
	Unordered	0.421	0.327	0.403
CDSSTR	*Helix 1*	-0.03	0.02	-0.01
	Helix 2	0.08	0.06	0.04
	Strand 1	0.2	0.22	0.26
	Strand 2	0.11	0.13	0.15
	Turns	0.2	0.24	0.24
	Unordered	0.43	0.33	0.31
K2D	*Alpha-helix*	0.03	0.21	0.18
	Beta Sheet	0.5	0.28	0.33
	Random Coil	0.47	0.51	0.49
	Least squares distance	306.8	260.4	127
	Maximum error	>0.22	>0.22	>0.22