Refining the genetic architecture of inherited cardiomyopathies through case-control analyses

Kate Thomson, Radcliffe Department of Medicine

and Kellogg College, University of Oxford



Thesis submission for the Degree of Doctor of Philosophy (DPhil) in Cardiovascular Medicine

Abstract: "Refining the genetic architecture of inherited cardiomyopathies through case-control analyses."

Kate Thomson, Kellogg College
Submitted for the degree of Doctor of Philosophy (DPhil) in Cardiovascular Medicine.
Trinity Term, 2018.

Inherited cardiomyopathies comprise a clinically and genetically heterogeneous group of heart muscle disorders, which are a major cause of heart disease and cardiac morbidity. Genetic testing for these conditions has been available for over a decade, and the number of genes incorporated into clinical test panels has increased significantly in recent years. However, rather than an increase in genetic diagnoses, this has resulted in a higher proportion of inconclusive results.

In this study, large-scale case-control analyses were employed to compare rare variation in 7855 cardiomyopathy cases and 60,706 reference controls, with the aim of identifying significant excess of rare variation in cases. These analyses highlighted the genes and classes of variant that show the strongest evidence of causality, as well as those that cannot be interpreted reliably. This methodology was used to reassess genes in current clinical test panels and to explore the contribution of more recently implicated genes. Analyses were also undertaken to explore the prevalence of pathogenic variants in the wider population.

These analyses provide critical insight into the genetic architecture of inherited cardiomyopathies, confirming the main causal genes and classes of variant. One novel finding was identified (an excess of rare truncating variants in the *ALKP3* gene in sarcomere-negative HCM cases compared to controls), but in most recently implicated genes, the gene-disease relationship could not be validated and lack of evidence precluded variant interpretation. Collectively, these findings inform evidence-based selection of genes for clinical test panels, reduce uncertainty in interpretation, and increase the clinical utility of testing in these medically important disorders. More broadly, this study demonstrates how the methods described can be used to enhance understanding of rare variation in dominant Mendelian disease genes.

Acknowledgements

This thesis was kindly supervised by Professor Hugh Watkins and Professor Martin Farrall, Radcliffe Department of Medicine, Oxford. I am very grateful for their support and guidance during this project.

Patient recruitment for BRRD project was led by Dr Liz Ormondroyd and supported by colleagues in the ICC service and Department of Cardiovascular Medicine, University of Oxford. Tim Dent assisted in clinical data entry and database management.

Advice on python and bash scripting was very kindly provided by Dr Silvia Salatino, Wellcome Centre for Human Genetics, University of Oxford. UKBB array analyses were assisted by Dr Anuj Goel and Dr Giulia Poloni.

I would like to thank colleagues within the NIHR BRRD team, in particular Dr Kathy Stirrups, Dr Chris Penkett and Hannah Stark who provided support throughout this project.

I am grateful to have had the opportunity to work with Professor Stuart Cook and his team, in particular Dr Roddy Walsh and Dr James Ware.

I am also grateful to Dr Judith Craft, Dr Katja Gehmlich and Dr Liz Ormondroyd for their advice, guidance and support during the preparation of this thesis. Special thanks are also due to Mike Mackley, my DPhil partner in crime.

Last, but not least, I thank my endlessly supportive family, in particular Scott, Jack and Harris. I could not have done this without you!

This thesis is dedicated to my incredible and inspirational sister, Emma.

Commonly used abbreviations

ACMG	American College of Medical Genetics
AD	Autosomal dominant
ARVC	Arrhythmogenic right ventricular cardiomyopathy
BED	Bed file format
BRRD	NIHR BioResource-Rare Diseases
ClinGen	Clinical genome research
DCM	Dilated cardiomyopathy
DNA	Deoxyribonucleic acid
EMR	Electronic medical records
EPV	Expected pathogenic variant
ExAC	Exome aggregation consortium
FET	Fisher's exact test
gnomAD	Genome aggregation database
GRCh37	Genome Reference Consortium Human Build 37
HCM	Hypertrophic cardiomyopathy
ICC	Inherited cardiac condition
KPV	Known pathogenic variant
LMM	Laboratory for Molecular Medicine
LOD	Logarithm of the odds
LQTS	Long QT syndrome
MAF	Minor allele frequency
NGS	Next generation sequencing
OMGL	Oxford Medical Genetics Laboratory

OR	Odds ratio
RNA	Ribonucleic acid
SCD	Sudden cardiac death
SF	Secondary finding
UKBB	UK Biobank
WES	Whole exome sequencing
WES	Whole genome sequencing

CHAPTER 1. INTRODUCTION 1

1.1 BACKGROUND 2

1.1.1 Hypertrophic Cardiomyopathy 2
1.1.1.1 Clinical features 2
1.1.1.2 Genetic basis of HCM 3
1.1.1.3 Clinical genetic testing in HCM 5
1.1.2 Dilated Cardiomyopathy 7
1.1.2.1 Clinical features 7
1.1.2.2 Genetic basis of DCM 7
1.1.2.3 Genetic testing in DCM 9
1.1.3 Arrhythmogenic Right Ventricular Cardiomyopathy 10
1.1.3.1 Clinical features 10
1.1.3.2 Genetic basis of ARVC 11
1.1.3.3 Genetic testing in ARVC 12
1.1.4 Genetic testing in inherited cardiomyopathies 13
1.1.4.1 The challenges of variant interpretation 15
1.1.4.2 Insights from population data 16

1.2 OBJECTIVES 17

CHAPTER 2. REASSESSMENT OF GENES IN CURRENT CARDIOMYOPATHY GENE PANELS USING DATA FROM CLINICAL COHORTS AND THE EXOME AGGREGATION CONSORTIUM. 19

2.1 INTRODUCTION 19

2.2 METHODS 22

2.2.1 Component cohorts 22
2.2.1.1 Oxford Medical Genetics Laboratory (OMGL) 22
2.2.1.2 Laboratory of Molecular Medicine, Partners Healthcare (LMM) 23
2.2.1.3 Variant classification in clinical cohorts 23
2.2.1.4 Exome Aggregation Consortium (ExAC) 24
2.2.2 Deriving a minor allele frequency threshold 24
2.2.3 Calculating rare variant frequency 24
2.2.3.1 Clinical cohorts 26
2.2.3.2 ExAC 26
2.2.4 Case-control analyses 27
2.2.4.1 Assessment of population-specific effects 27
2.2.4.2 Variant distribution 28

2.3 RESULTS 29

2.3.1 Deriving a minor allele frequency threshold 29
2.3.2 Case-control analyses 29
2.3.2.1 HCM 33
2.3.2.2 DCM 34
2.3.2.3 ARVC 35
2.3.2.4 Summary 35
2.3.2.5 Assessment of population-specific effects 50
2.3.2.6 Variant distribution 50

2.4 DISCUSSION 55

2.4.1 Deriving a minor allele frequency threshold 55
2.4.2 Case-control analyses 56
2.4.2.1 Exploring the utility of case control findings 57
2.4.2.1.1 CSRP3 58
2.4.2.1.2 ACTN2 58
2.4.2.1.3 MYBPC3 variants in DCM 59
2.4.2.1.4 ANKRD1 61
2.4.2.2 Variants classes and variant interpretation 63
2.4.2.3 Variant distribution 64
2.4.2.4 Study limitations 65

2.5 CONCLUSIONS 67

CHAPTER 3. ANALYSIS OF 60 PROPOSED HCM GENES USING WHOLE GENOME SEQUENCE DATA IN PATIENTS WITHOUT PATHOGENIC VARIANTS IN CONFIRMED GENES. 69

3.1 INTRODUCTION 69

3.2 METHODS 70

3.2.1 Study participants 70
3.2.2 Clinical genetic testing 71
3.2.3 Reference controls 71
3.2.4 Power calculations 72
3.2.5 Whole genome sequencing 73
3.2.6 Ethnicity 74
3.2.7 Gene selection 74
3.2.8 Coverage data 82
3.2.9 Analysis protocol 83
3.2.10 Homology 85
3.2.11 Comparing variant distribution 86
3.2.12 Variant classification 86

3.3 RESULTS 86

3.3.1 Characteristics of the cohort 86
3.3.2 Ethnicity 90
3.3.3 Gene coverage 90
3.3.4 Regions of homology 92
3.3.4.1 TTN 92
3.3.4.2 FLNC 92
3.3.4.3 MYH6 94
3.3.5 Gene level analyses 101
3.3.5.1 Comparing variant frequency 101
3.3.5.2 Comparing variant distribution 126
3.3.5.3 ALPK3 128
3.3.5.4 DES 133
3.3.5.5 KCNQ1 141
3.3.5.6 TCAP 147
3.3.6 Variant level analyses 148
3.3.7 Case level analyses 149

3.4 DISCUSSION 167

3.4.1 Characteristics of the cohort 167

3.4.2 Ethnicity 167
3.4.3 Gene coverage 168
3.4.4 Regions of Homology 168
3.4.4.1 FLNC 169
3.4.4.2 MYH6 169
3.4.5 Gene level analyses 174
3.4.5.1 ALPK3 174
3.4.5.1.1 Gene structure and protein function 174
3.4.5.1.2 Mouse studies 175
3.4.5.1.3 Role in disease 178
3.4.5.1.4 Histopathological, RNA and functional studies 179
3.4.5.1.5 Heterozygous ALPK3 truncating variants in HCM 180
3.4.5.1.6 Tissue/isoform specific loss of protein function 182
3.4.5.1.7 Co-existing variants 182
3.4.5.1.8 Variant penetrance 183
3.4.5.1.9 False positive result/technical artefact 183
3.4.5.1.10 Clinical review 184
3.4.5.1.11 Conclusions 184
3.4.5.2 DES 184
3.4.5.2.1 Gene structure and protein function 185
3.4.5.2.2 Role in disease 187
3.4.5.2.2.1 DCM 189
3.4.5.2.2.2 ARVC 190
3.4.5.2.3 DES variants in HCM 190
3.4.5.2.3.1 Clinical review 191
3.4.5.2.3.2 Analysis in larger HCM patient cohort 191
3.4.5.2.4 Conclusions 194
3.4.5.3 KCNQ1 194
3.4.5.3.1 Gene structure and protein function 194
3.4.5.3.2 Role in disease 196
3.4.5.3.3 KCNQ1 variants in HCM 197
3.4.5.3.4 Conclusions 198
3.4.5.4 TCAP 198
3.4.5.4.1 Gene structure and protein function 198
3.4.5.4.2 Role in disease 199
3.4.5.4.3 TCAP variants in HCM 199
3.4.5.4.4 Conclusions 201
3.4.6 Variant level analyses 201
3.4.7 Case level analyses 202
3.4.8 Limitations 203
3.4.8.1 Cohort size 203
3.4.8.2 Gene selection 205
3.4.8.3 Clinical heterogeneity 205
3.5 CONCLUSIONS 206
CHAPTER 4. EXPLORING THE POPULATION PREVALENCE OF PATHOGENIC VARIANTS IN INHERITED CARDIAC CONDITION GENES. 209
4.1 INTRODUCTION 209
4.2 METHODS 211

4.2.1 UK BioBank cohort (UKBB) 211
4.2.2 Bioresource rare disease cohort (BRRD) 212
4.2.2.1 BRRD cohort genes analysed 213
4.2.2.2 BRRD cohort variants selected for analysis 214
4.2.2.3 BRRD cohort analysis protocol 214
4.3 RESULTS 216
4.3.1 UKBB cohort 216
4.3.2 BRRD cohort 229
4.4 DISCUSSION 236
4.4.1 UKBB cohort 236
4.4.2 BRRD cohort 238
4.4.2.1 Variant prevalence 238
4.4.2.2 Variant penetrance 239
4.4.2.3 Study limitations 241
4.5 CONCLUSIONS 246
CHAPTER 5. CONCLUSIONS 247
5.1 AIMS 247
5.2 SUMMARY OF FINDINGS 247
5.3 IMPLICATIONS OF FINDINGS 248
5.4 FUTURE WORK 249
5.4.1 Case-control analyses 249
5.4.2 Additional genetic analyses 250
5.4.3 Assessing the penetrance of pathogenic variants detected in BRRD non-HCM cohorts 250
5.4.4 Clinical studies in sarcomere-negative HCM cohort 251
5.5 CONCLUSION 251
BIBLIOGRAPHY 253
APPENDIX I 277

Chapter 1. Introduction

Inherited cardiomyopathies comprise a clinically and genetically heterogeneous group of heart muscle disorders which are a major cause of heart disease and cardiac morbidity, particularly in the young[1]. Collectively, these medically important disorders are relatively common, with an estimated population prevalence of around 3%[2].

Research undertaken over the past three decades has provided critical insight into the molecular basis of these disorders and evolved our understanding of the causal genes and associated phenotypes[1,3]. Many of these genetic discoveries have been translated from the research environment into clinical laboratories, and genetic testing is now an integral component of clinical care, informing the diagnosis, treatment and management of patients and their at risk relatives[4–9].

However, despite this substantial progress, there remain fundamental challenges which limit the clinical utility of genetic testing. This highlights the need for further research to refine the genetic architecture of these conditions and improve the sensitivity of genetic testing.

This study seeks to address some of the key issues that confound genetic analyses in the three most common inherited cardiomyopathies: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM) and arrhythmogenic right ventricular cardiomyopathy (ARVC).

1.1 Background

1.1.1 Hypertrophic Cardiomyopathy

1.1.1.1 Clinical features

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiac disorder, with an estimated population prevalence of approximately 1/500[10] and is the leading cause of sudden death in athletes and young adults <35 years of age[1].

Clinically, HCM is characterised by left ventricular hypertrophy (LVH) in the absence of an obvious cause such as hypertension or valvular heart disease. In the normal heart of an adult male the left ventricle is approximately 1cm thick, a diagnosis of HCM is made when the muscle is over 1.5cm[4]. Thickening can occur anywhere in the left ventricle, but predominantly arises in the inter-ventricular septum (Figure 1.1-1). Myocyte disarray and myofibrillar disorganisation with interstitial and replacement fibrosis are histopathological hallmarks of HCM[11]; these features may be present in up to 40% of the ventricular tissue and are thought to interfere with normal electrical activity in the heart, predisposing to arrhythmia and sudden cardiac death (SCD)[12]. Hypertrophy usually develops during adolescence and is usually apparent by late teens/early twenties, however age of onset and the degree of hypertrophy can be variable. In a small number of patients there is a more progressive disease course, leading to left ventricular dilation and heart failure, this "burnt-out" HCM is morphologically similar to DCM[13].

Making a diagnosis of HCM on clinical grounds alone can be difficult as the clinical features and symptoms of this disorder are variable, even within families, and can appear at any age. Another complicating factor is incomplete penetrance, whereby some

individuals carrying a known familial disease-causing variant do not appear to develop HCM even at advanced age.

Furthermore, LVH is not specific to HCM and can occur in association with other conditions such as infiltrative heart disease (e.g. cardiac amyloidosis), systemic disease (e.g. hypertensive heart disease), and athletes' heart. Other differential diagnoses include inherited conditions, including storage disorders (such as Fabry's Disease[14] and Danon's disease[14]) and mitochondrial disorders[15–17]. Typically, these are multi-system disorders, however they are clinically heterogeneous and can present with isolated LVH indistinguishable from HCM. Clinically, it is important to recognise these conditions as the genetics, clinical management and prognosis is distinct from familial HCM.

1.1.1.2 Genetic basis of HCM

HCM is broadly considered to be an autosomal dominant (AD) condition, although incomplete and age-related penetrance is common. The first causal genes in HCM were identified around 25 years ago through linkage studies in several large pedigrees[18–22]. These genes encode protein components of the cardiac sarcomere, the contractile apparatus of the heart muscle. On the basis of these initial findings, HCM was described as a "disease of the sarcomere" [18].

Pathogenic variants in two sarcomeric genes, *MYH7* and *MYBPC3*, are the most common genetic causes of HCM, and are detected in 10-20% of individuals referred for genetic testing (Table 1.1-1). Variants in six other sarcomeric protein genes (*ACTC1, TNNI3*, *TNNT2, TPM1*, *MYL2* and *MYL3*), are detected in between 1-5% of patients (Table 1.1-1).

The majority of pathogenic variants in the sarcomeric genes are single nucleotide substitutions which cause missense variants in which the altered protein product is

incorporated into the sarcomere and believed to act in a dominant-negative manner. A notable exception is the *MYBPC3* gene where truncating variants (i.e. nonsense, frameshift and splice-site variants) causing premature termination of protein translation and protein haploinsufficiency are prevalent. Recurrent and founder disease-causing variants have been described in some genes[23–26], however the majority of reported variants appear to be novel, or private to a family[27].

Up to 5% of patients have been reported to have more than one sarcomeric gene variant[28–31], although more recent data suggests that number of individuals with more than one clearly pathogenic variant may be lower[32,33]. Patients with more than one definitively pathogenic variant typically present with a more severe phenotype and a younger age of onset compared to single mutation carriers[34–39].

While there is a substantial body of genetic evidence supporting a causal role for variants in sarcomeric genes in HCM, the molecular mechanisms underlying pathogenesis are still unclear[40]. The majority of the functional evidence suggests that the universal outcome of HCM-causing variants in sarcomeric genes is increased myofilament activation[41]. This activation has been proposed to arise through a variety of mechanisms including, i) altered myosin kinetics[42,43], ii) impaired calcium cycling and sensitivity[44], and iii) altered energy homeostasis[45,46]. These perturbations are thought to stimulate downstream signalling pathways, which promote myocyte cell growth, myofibrillar disarray and myocardial fibrosis[1,40].

Pathogenic variants in syndromic disease genes which can mimic HCM (often referred to as phenocopy genes), e.g. Fabry's disease (*GLA*), *PRKAG2*-related cardiomyopathy and Danon's disease (*LAMP2*), together account for approximately 2% of patients in a typical HCM cohort[27].

More recently, the spectrum of genes implicated in HCM has expanded to include non-sarcomeric genes[1,27,47]; for example, genes encoding sarcomeric Z-disk proteins (e.g. *CSRP3*[48]), cytoskeletal proteins (*ACTN2*[49]) and sarcoplasmic reticulum Ca2+-cycling proteins (*PLN*[50]). In some, a causal role in HCM is supported by strong genetic evidence (e.g. linkage and segregation data) or *in vivo* functional data; although, on the basis of current evidence, they appear to be infrequent causes of HCM[27]. However, in many, the gene-disease relationship is less well defined (e.g. *ANKRD1*[51], *MYOZ2*[52], *MYH6*[53] and *TCAP*[54]).

1.1.1.3 Clinical genetic testing in HCM

Clinical genetic testing has been available for individuals with HCM for over 15 years. During this time, advances in technology have transformed testing from low-throughput single gene analyses to high-throughput multi-gene panel tests[55–57]. Genetic testing is now widely available in the UK as an NHS service, and in many other countries, for the well-established sarcomeric genes (Table 1.1-1) and testing for most of these genes is recommended by clinical guidelines[58,59]. A number of the other less commonly associated genes are also included in clinical test panels (https://www.ncbi.nlm.nih.gov/gtr/). With current clinical test panels, putative pathogenic variants are detected in approximately 30-40% of individuals referred for testing.

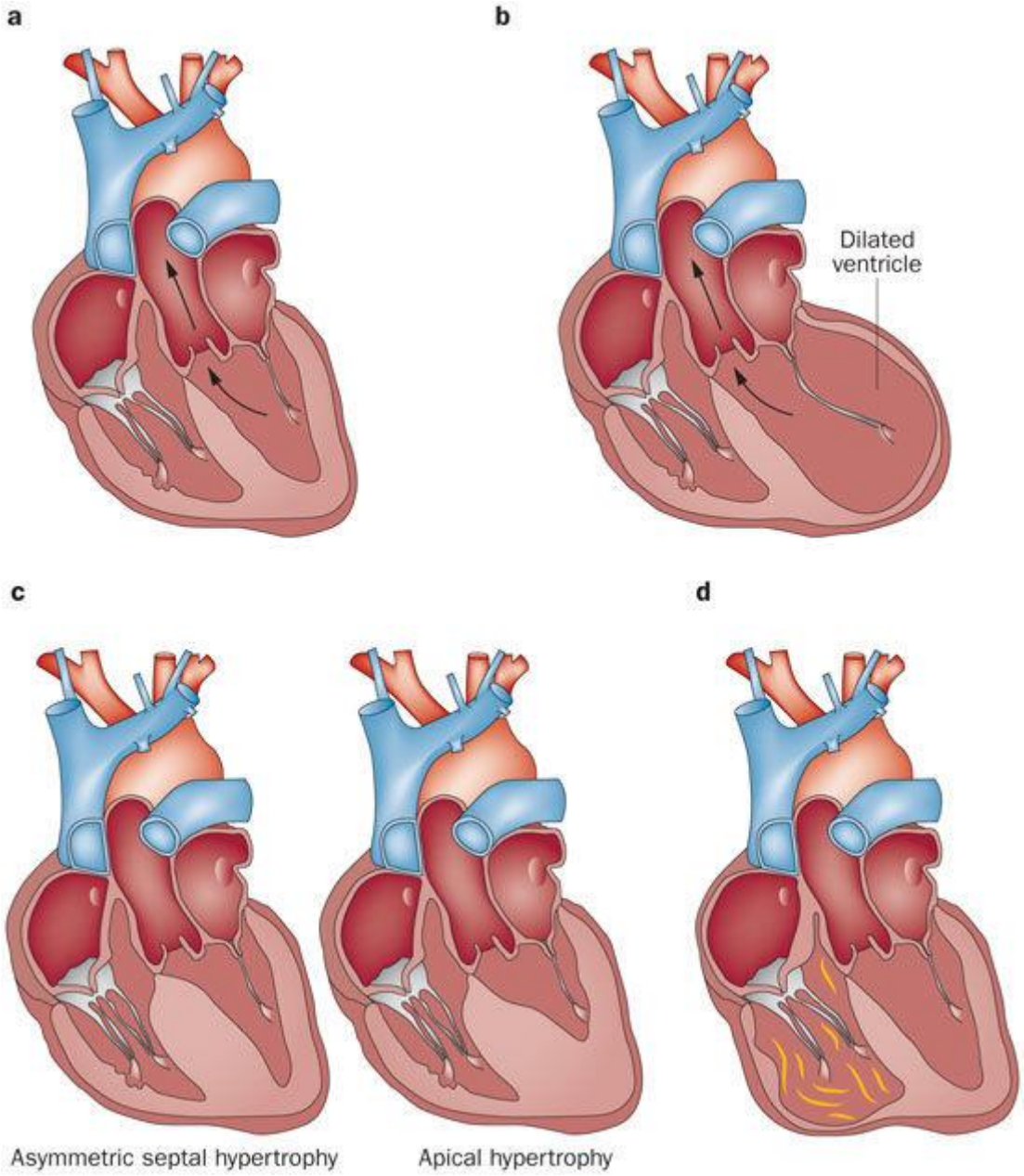


Figure 1.1-1 The three main forms of cardiomyopathy. a The normal heart. b Dilated cardiomyopathy. c Two examples of hypertrophic cardiomyopathy: classic asymmetric septal hypertrophy, most often limited to the basal portion (not shown), and apical hypertrophy. d Arrhythmogenic right ventricular cardiomyopathy, with right-sided involvement only. Reproduced with permission from [60], Copyright Macmillan Publishers Limited, part of Springer Nature.

1.1.2 Dilated Cardiomyopathy

1.1.2.1 Clinical features

DCM is the most clinically heterogeneous cardiomyopathy[1,61]. It is characterised by left ventricular dilatation, systolic dysfunction, myocyte death and myocardial fibrosis. Individuals with DCM typically present with heart failure, although arrhythmia, including SCD, can be the initial disease indication (Figure 1.1-1). The vast majority of cases have a "non-genetic" aetiology (e.g. ischemic heart disease, viral myocarditis, substance abuse), however, in individuals with so called idiopathic DCM, in whom the common "non-genetic" causes have been excluded, approximately a quarter of cases appear to be familial[62]. Estimates relating to the population prevalence of idiopathic DCM are highly variable, ranging between 1 in 5000[63] to 1 in 250[61]. Onset generally occurs in the 4th to 6th decade, but can present at any age.

1.1.2.2 Genetic basis of DCM

DCM is also the most genetically heterogeneous cardiomyopathy. The first causal genes were discovered around 20 years ago[64,65], since then, over 100 genes have been reported to cause DCM either as an isolated phenotype or as part of a syndrome[66]. A number of these have been reported as causal genes in other cardiomyopathies or muscular dystrophies[60,61,66]. The majority of non-syndromic DCM is inherited in an AD manner; however autosomal recessive (AR) and X-linked forms are also reported[61,66]. As in other autosomal dominant cardiomyopathies, incomplete penetrance and variable expressivity is frequently observed[61,66].

The genes implicated in DCM encode proteins which perform a diverse range of structural and functional roles within the cardiac myocyte (e.g. sarcomere, nuclear membrane, desmosome, sarcoplasmic reticulum, cytoskeleton), illustrating the complex

pathophysiology of this disorder (Figure 1.1-2)[60,61]. Although the disease mechanisms of many of these proteins are not yet fully understood, ultimately these aberrations lead to cell death and a decrease in the number of functional cardiomyocytes, which results in impaired contraction and eventually heart failure.

In individuals with AD non-syndromic DCM, truncating variants in the *TTN* gene (TTNtv), which encodes the giant sarcomere protein titin, are the most common genetic defect, detected in up to 25% of cases[67–69]. However, TTNtv are also detected in 2-3% of healthy individuals or unselected control cohorts[67,70,71]. Studies have shown that DCM associated TTNtv tend to cluster predominantly in the A-band portion and exons that are constitutively incorporated into adult cardiac transcripts[69–72]. While variants in these regions are predicted to be pathogenic, and associated with a significantly increased risk of DCM, they are notably less penetrant than AD disease-causing variants in other inherited cardiac condition (ICC) genes. It is recognised that other genetic and environmental factors contribute to the disease aetiology in individuals heterozygous for pathogenic TTNtv[69,73,74].

Variants in the *LMNA* gene, which encodes the nuclear membrane protein isoforms lamin A and lamin C, are one of the most common genetic defects in DCM. Individuals with *LMNA*-associated DCM typically have a distinct phenotype, characterised by progressive conduction system disease, arrhythmia and systolic impairment [75,76]. Pathogenic variants in this gene are detected in around 4–6% of individuals referred for testing[77,78], although the frequency is significantly higher in individuals with familial DCM and conduction disease[79].

Variants in genes encoding sarcomeric proteins (*MYH7*, *TNNT2*, *TNNI3* and *TPM1*) together account for 5-10% of DCM patients referred for genetic testing [61,65,77,80]. In

contrast to the HCM-causing variants in these genes, those implicated in DCM appear to diminish rather than enhance myofibrillar function.[81–85] For example, pathogenic *MYH7* variants implicated in DCM have been shown to depress molecular motor function [86,87] and variants in thin filament genes (*TNNT2* and *TNNI3* and *TPM1*) reduce the calcium sensitivity of contractile regulation[82–85,88].

Pathogenic variants in genes, including protein components of the Z-disk (*MYPN* [89]), the cytoskeleton (*VCL*[90], *DES*[91], *FLNC*[92]), desmosome (*DSP*[93]), and RNA-binding proteins (*RBM20*[94]), appear to account for a smaller proportion of cases. However, accurate data on the contribution of these genes to DCM is lacking[77,95].

1.1.2.3 Genetic testing in DCM

With the advent of high-throughput DNA sequencing techniques, genetic testing for DCM has become more tractable and clinical test panels have expanded to accommodate a large number of implicated genes (https://www.ncbi.nlm.nih.gov/gtr/). However, at present, putative pathogenic variants are detected in only 20-30% of patients and a small number of genes account for the majority cases (Table 1.1-1) [77,96].

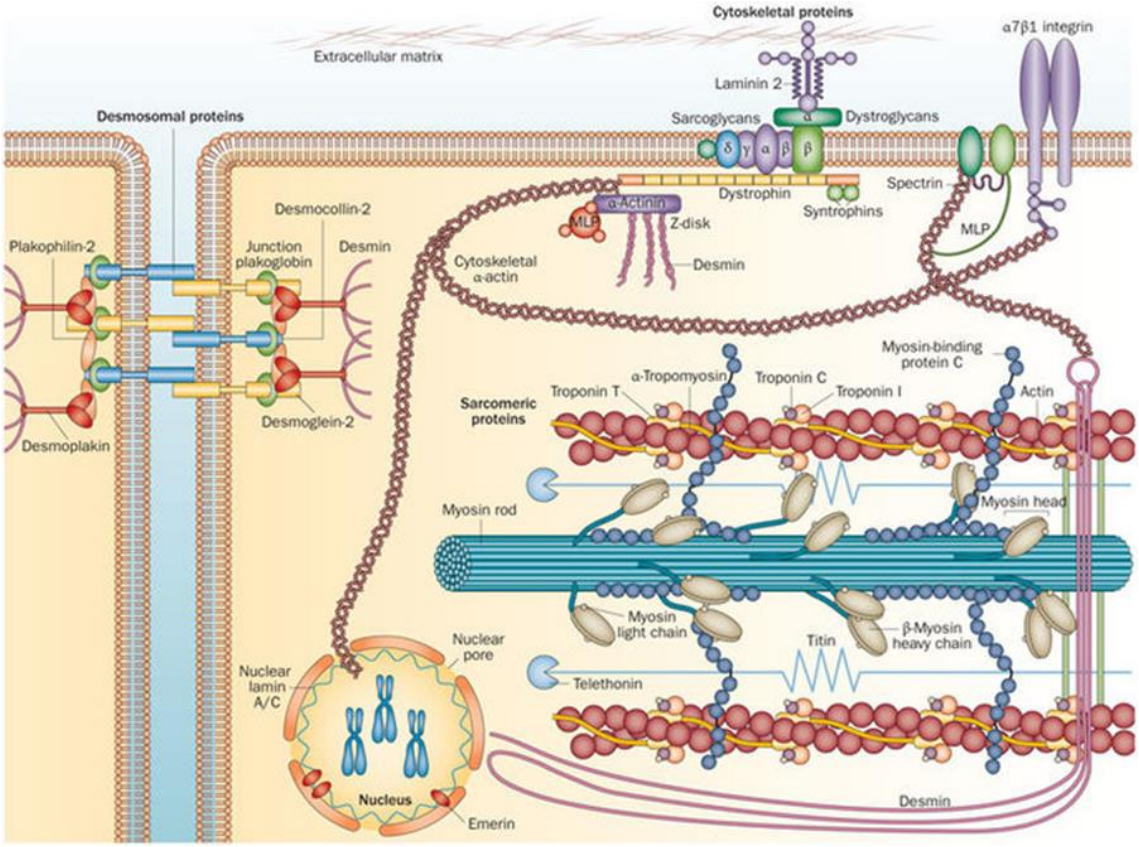


Figure 1.1-2 The main proteins involved in cardiomyopathies. Indicated are the desmosomal proteins, which form the desmosome and connect one cell to another; the sarcomeric proteins, which form the contractile apparatus; the cytoskeletal proteins, which connect the extracellular matrix to the cells, and connect various transmembrane proteins to the sarcomere and the nucleus; and the nuclear envelope proteins, which are involved in anchoring and interacting with cytoskeletal and nucleoskeletal elements involved in movement. Mutations in the encoding genes lead to aberrant function of the respective proteins and to one of the cardiomyopathies. Abbreviation: MLP, cysteine and glycine-rich protein 3 (also known as muscle LIM protein). Reproduced with permission from [60], Copyright Macmillan Publishers Limited, part of Springer Nature.

1.1.3 Arrhythmogenic Right Ventricular Cardiomyopathy

1.1.3.1 Clinical features

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a progressive cardiomyopathy, characterised by electrical and structural dysfunction[97,98]. It is estimated to affect between 1 in 1000 and 1 in 5000 individuals and is associated with cardiac arrhythmia, a high-risk of heart failure and SCD. Histologically, it is characterised by replacement of the myocardium with adipose and fibrous tissue; while these

pathological changes are predominantly observed within the right ventricle, left and biventricular involvement is also observed [99].

The onset of symptoms is usually between 20 and 40 years of age, but can present in childhood[7]. Individuals in the early phase of ARVC tend to present with ventricular arrhythmias, with the structural and functional abnormalities becoming more evident as the disease progresses. Clinical diagnoses can be confounded by variable disease expression and the nonspecific nature of some of the associated features, which can overlap with other cardiac disorders, such as DCM [100,101].

1.1.3.2 Genetic basis of ARVC

Pathogenic variants in genes encoding protein components of the cardiac desmosome were first reported in individuals with AD ARVC in 2005[102] and this cardiomyopathy is now commonly described as a "disorder of the desmosome". Desmosomes are multi-protein complexes that form cell-to-cell junctions and link intermediate filament networks between adjacent cells[103].

Truncating variants in the *PKP2* gene are the most common cause of ARVC, accounting for 20-30% of patients. Variants in the four other desmosomal genes (*DSC2*, *DSG2*, *DSP*, *JUP*) are less frequently associated (Table 1.1-1); the majority of pathogenic variants in these genes are truncating variants [104]. The relatively high background rate of missense variants in the *DSG2* and *DGC2* genes can confound interpretation and lead to high number of uncertain results[105].

Although generally considered to be an autosomal dominant disorder, incomplete penetrance and variable expressivity are common. Also, the presence of more than one putative pathogenic variant in either the same gene (compound heterozygosity) or in

another desmosomal gene (digenic heterozygosity) has been reported, suggesting a more complex heritability in some cases [106,107].

The precise mechanisms by which pathogenic variants in desmosomal proteins cause ARVC are incompletely understood. It has been proposed that loss of desmosomal protein function might disrupt cell to cell adhesion at the intercalated discs (ICs), leaving cells more vulnerable to mechanical stress and prone to detachment and cell death. This may lead to inflammation and loss of myocardium[108,109]. Studies have shown that desmosome dysfunction affects ion channel and gap junction proteins at the ICs, causing abnormal electrical coupling between myocytes[110]. Additionally, remodelling at the ICs has also been reported to alter the Wnt/β-catenin signalling pathway, which is an important regulator of myogenesis versus adipogenesis [111,112].

Pathogenic variants in other non-desmosomal genes have been implicated in ARVC, including genes linked to other inherited cardiac conditions (e.g. *TMEM43*[113], *LMNA*[114,115], *CHD* [116,117], *FLNC*[92], *TTN*[118], *RYR2* [119], *DES* [120,121]. At present, there is very little data relating to the contribution of these genes to ARVC.

1.1.3.3 Genetic testing in ARVC

With current clinical test panels, which largely focus on the five main desmosomal genes, pathogenic variants are detected in up to 40% of individuals referred for genetic testing [98].

Gene	Protein	% of cases with a pathogenic variant
ARVC [97,122]		
PKP2	Plakophilin	20-40
DSP	Desmoplakin	3-15
DSG2	Desmoglein 2	3-20
DSC2	Desmocollin 2	1-8
JUP	Junction plakoglobin	<1
DCM [67,68,123]		
TTN	Titin	13-25
MYH7	β-Myosin heavy chain	~5
LMNA	Lamin A/C	~5
DSP	Desmoplakin	~5
TNNT2	Cardiac Troponin T	2-3
TPM1	α-Tropomyosin 1	~2
TNNI3	Cardiac Troponin I	<1
HCM [124]		
MYBPC3	Myosin binding protein C, cardiac	15-20
MYH7	β-Myosin heavy chain	10-15
TNNI3	Cardiac Troponin I	~2
TNNT2	Cardiac Troponin T	~2
TPM1	α-Tropomyosin 1	~1
MYL2	Myosin light chain 2	~1
MYL3	Myosin light chain 3	~1
ACTC1	Actin, alpha , cardiac muscle 1	~0.5

Table 1.1-1 Summary of most common and well-established causal genes in ARVC, DCM and HCM.

1.1.4 Genetic testing in inherited cardiomyopathies

The clinical heterogeneity and variable expressivity of inherited cardiomyopathies means it can be difficult to make a definitive diagnosis on clinical grounds alone[125]. Another complicating factor is incomplete penetrance, where individuals with a known familial disease-causing variant do not appear to develop cardiomyopathy, even at advanced age[1]. This presents a problem when assessing at-risk relatives, therefore, clinical guidelines recommend that clinical evaluation of phenotype-negative first-degree relatives is undertaken annually during adolescence and subsequently every 5 years[59,126].

The potential clinical utility of genetic testing in individuals with inherited cardiomyopathy is widely recognised [4,7,59,66]. Genetic testing can confirm the clinical diagnosis in an individual, excluding the differential diagnoses. However, the main benefit of genetic testing is that detection of a pathogenic variant in the proband enables cascade screening of at-risk family members. Identifying family members who have inherited the familial variant enables targeting of clinical resources to those at high risk of developing disease and who may be at increased risk of sudden death, and releases others from clinical screening. In the context of family screening, genetic tests can be more cost effective than clinical tests alone [127].

However, the clinical and genetic heterogeneity of these conditions present significant challenges for genetic testing, and in a high proportion of patients referred for testing no pathogenic variant is found. In some cases, this may reflect the fact that an individual's cardiomyopathy does not have a Mendelian basis, and is instead due to the combined interaction of genetic variants of more modest effect (susceptibility variants) and/or environmental factors[1]. In others, it may be that the causative gene was not included in the test panel, or is in an as yet undiscovered gene. Yet while advances in DNA sequencing technology have made it possible to increase the number of genes in clinical test panels, rather than increasing the number of genetic diagnoses, the main outcome has been an increase in the number of uncertain findings[27,56,77].

As a consequence, results interpretation has become a significant problem in inherited cardiomyopathy gene testing, and this is reducing the effectiveness and clinical utility of genetic testing. As genetic testing for these conditions becomes more widely available and routine in clinical practice, there is a clear need to refine test strategies and improve results interpretation.

1.1.4.1 The challenges of variant interpretation

As in most other inherited disorders, one of the main challenges in cardiomyopathy gene analysis is variant interpretation. Multiple lines of evidence are used to assess the likely pathogenicity of variants detected in disease genes, and collectively these data contribute towards a final classification. In recent years, guidelines have been published in efforts to standardise approaches to variant interpretation and classification. These include, i) defining the criteria that can be used to assess pathogenicity, ii) weighting these criteria, and iii) categorising variants into specific classes on the basis of assigned criteria (i.e. pathogenic, likely pathogenic, uncertain significance, likely benign, or benign)[128].

However, for many of the variants detected in cardiomyopathy disease genes, it can be difficult to obtain the evidence needed to differentiate between pathogenic and benign variation. For example, while segregation studies are considered a useful means of asssessing variant pathogencity, due to incomplete and/or age-related penetrance there are often too few affected relatives to provide strong statistical support for segregation[129].

Also, many of the variants found in cardiomyopathy disease genes are novel missense changes, a class of variants for which interpretation is particularly challenging. *In silico* tools developed to predict the tolerance of missense amino acid substitutions [130–133] do not have the sensitivity or specificity required for use in a clinical setting[134–136]. Also, in many cases there is very limited or no knowledge about the critical functional domains of the proteins, and data from functional studies are rarely available.

Critically, providing a clinically valid and meaningful interpretation requires a thorough understanding of the genetic basis of the disease under investigation, including the

causal genes, the extent of normal and pathogenic variation within these genes, and their associated phenotypes. However, for many cardiomyopathy disease genes, particularly those more recently implicated, this knowledge is lacking. Without this evidence base, reliable variant interpretation is rarely achievable.

These inherent difficulties in variant interpretation mean that many of the variants detected in inherited cardiomyopathy disease genes are classified as variants of uncertain significance (VUS), as there is insufficient evidence to support pathogenicity. These variants (or variants so classified) are clinically un-actionable, and cannot be used for genetic testing of at risk family members.

1.1.4.2 Insights from population data

A more fundamental challenge in rare variant interpretation has been revealed by recent large scale genomic sequencing studies in non-Mendelian disease or population based cohorts[137–139]. Data from these studies has shown that rare variants in numerous inherited disease genes, including those reported to cause inherited cardiomyopathies, are collectively common in the general population, which infers that many are not penetrant pathogenic mutations.

Additionally, these data have shown that many cardiomyopathy gene variants previously reported as pathogenic, are present in the wider population at frequencies incompatible with pathogenicity[140,141]. These findings suggest that some of these previous studies were poorly controlled or relied on insufficient evidence for defining pathogenicity. In a number of cases, this new population data questions the legitimacy of some gene-disease relationships.

Critically, these studies emphasize the importance of taking into consideration background levels of rare variation in disease-genes when investigating variant pathogenicity[142], and caution that rarity alone is not sufficient to confirm pathogenicity.

These studies have also exposed gaps in our understanding about the natural history of pathogenic variants in the wider population. The need to develop this evidence-base has become increasingly pertinent as genomic sequencing is adopted into routine practice, and genetic testing of ICC disease genes is being considered in individuals with no personal or family history of disease[143].

1.2 Objectives

In this thesis, some of the key issues that confound results interpretation in genetic testing in inherited cardiomyopathies were examined. The main objectives were to refine understanding of the causal genes and variants in inherited cardiomyopathies, and to improve variant interpretation in confirmed genes.

i) Chapter 2 describes case-control analyses undertaken to reassess the genes in current cardiomyopathy test panels. Analyses were undertaken using genotype data from 7,855 cardiomyopathy cases, and up to 60,706 individuals from the Exome Aggregation Consortium (ExAC). These analyses provided an opportunity to determine whether the ExAC resource was a suitable reference cohort for "rare variant burden testing" in dominant Mendelian disease. Through these comparative rare variant analyses, the aim was to define the genes, regions of genes, and classes of variant that can be most reliably interpreted, and also to examine the hypothesis that some of

the genes on current cardiomyopathy test panels are not disease-causing, or at least, rare variants in these genes are not interpretable.

ii) In Chapter 3, this case-control approach was extended to assess the contribution of a further 60 genes implicated in HCM. These analyses were performed using whole genome sequencing data from a series of 230 sarcomere gene-negative HCM patients, and 7,253 reference controls. These analyses allowed me to explore the hypothesis that many of the genes recently implicated in HCM do not play a substantial role in this condition.

iii) Finally, in Chapter 4, analyses were undertaken to refine understanding about prevalence and penetrance of rare variants in cardiomyopathy genes in the wider population. The prevalence of pathogenic variants in a number of key inherited cardiac condition genes was evaluated in two large unselected reference cohorts. These analyses enabled me to examine the hypothesis that pathogenic variants in ICC genes are more frequent in the general population that would be expected given current estimates of disease prevalence and penetrance.

Chapter 2. Reassessment of genes in current cardiomyopathy gene panels using data from clinical cohorts and the Exome Aggregation Consortium.

2.1 Introduction

In recent years, high throughput DNA sequencing methods have enabled us to explore rare genetic variation in individuals with Mendelian disease and the wider population. During this time we have seen a dramatic increase in the number of genes implicated in Mendelian disorders[144], and a corresponding rise in the number of genes incorporated into clinical genetic test panels[145] or analyses being extended to exome and genome sequencing[146].

However, it is becoming clear that our ability to interpret the results of these diagnostic tests has not kept pace. Rather than an increase in high confidence genetic diagnoses, we are seeing a higher proportion of inconclusive results, with more patients reported to have a variant of uncertain significance (VUS)[77,147,148]. These clinically un-actionable findings limit the utility and reduce the cost effectiveness of genetic testing[127].

This increase in variants classified as VUS is in part due to the fact that for some genes, particularly those implicated through candidate gene studies, definitive evidence of disease causation is lacking (e.g. small case numbers, small control cohort, no segregation studies or functional evidence); as such, novel variants in these genes cannot be interpreted reliably.

A more pervasive problem, exposed by recent large-scale population based genomic sequencing studies, is that very rare variation in Mendelian disease genes is

considerably more common in the general population than previously anticipated [149,150]. As the population is explored in greater depth, it has become clear that some 'rare' variants are not rare at all, and that many of the rare variants in Mendelian disease genes formally thought to be pathogenic for this reason, are most likely benign. This has not only led us to question the clinical significance of individual variants previously reported as pathogenic, but also whether some previously implicated genes are truly disease-causing.

The challenges of variant interpretation in Mendelian disorders are particularly well illustrated by inherited cardiomyopathies: Hypertrophic Cardiomyopathy (HCM), Dilated Cardiomyopathy (DCM) and Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC). Collectively, these clinically and genetically heterogeneous disorders are relatively common and medically important.[1] While clinical genetic testing for these conditions has been available for over a decade, the number of genes reported as disease-causing has increased significantly in recent years, fuelling the trend towards large inclusive test panels[56,57,151]. However, the proportion of un-interpretable inconclusive results is increasing[56,77,124].

Additionally, analysis of cardiomyopathy genes within population based cohorts (1000 Genomes Project[152] and Exome Sequencing Project [ESP] [150]) has shown a higher than expected frequency of very rare coding variants and variants previously defined as disease-causing mutations [140,141,153]. Leading us question the validity of some previously reported variants and certain gene-disease relationships.

To improve the specificity of diagnostic testing in these conditions, there is a clear need to re-evaluate the genes in current clinical test panels. A key first step of this re-evaluation is in-depth comparison of rare variation in patient cohorts and large

population based cohorts. Penetrant pathogenic variants must be rare in the general population; therefore, in true causal genes there should be a significant excess of very rare variation in cardiomyopathy cohorts compared to controls. As such, case control analyses should enable us to highlight the genes most commonly and robustly associated with a particular disorder, and which consequently, are suitable for inclusion in a clinical diagnostic test panels.

The 1000 Genomes and ESP projects highlighted the importance of population data for rare variant interpretation in cardiomyopathy and other Mendelian disease genes. However, neither dataset was sufficiently powered to accurately measure levels of variation in the range that would be considered compatible with pathogenicity (i.e. less than 1/10,000). The Exome Aggregation Consortium (ExAC)[138], which provides data from 60,706 human exomes, was the first publically available dataset large enough to reliably measure background levels of very rare variation in cardiomyopathy disease genes and to allow reliable re-evaluation of previously implicated genes. Whilst not healthy controls, the component cohorts are from studies of non-Mendelian diseases and population controls that would not be expected to be enriched for inherited cardiac conditions.

Using data from 7855 cardiomyopathy cases and up to 60,706 ExAC reference samples, case control analyses were undertaken to explore rare variation in the genes in current cardiomyopathy clinical test panels. Data from cardiomyopathy cases was used to empirically derive an upper bound for the allele frequency of confirmed cardiomyopathy-causing mutations. The number of variants below this threshold was compared between cases and ExAC. Analyses were also undertaken to explore differences in rare variation distribution between cases and controls.

These analyses provide an opportunity to explore the hypothesis that some, but not all the genes on current cardiomyopathy clinical test panels are disease-causing. Importantly, by defining the genes, regions of genes, and/or classes of variants that can be reliably interpreted in a clinical setting, it may be possible to reduce uncertainty in results interpretation and improve the diagnostic yield of testing.

2.2 Methods

2.2.1 Component cohorts

Table 2.2-1 provides an overview of the cohorts analysed in this study.

2.2.1.1 Oxford Medical Genetics Laboratory (OMGL)

This cohort comprised index cases, with a clinical diagnosis of HCM, DCM or ARVC made by a consultant cardiologist, referred for genetic testing from Clinical Genetics centres across the UK. Data on patient ethnicity was not available. All samples received for diagnostic genetic testing of HCM, DCM or ARVC genes were eligible and analysis was undertaken in a routine clinical setting using clinical consent.

Genotype data was obtained from 3267 individuals with HCM, 559 with DCM and 361 with ARVC. At the time of this study, the clinical test panels included 16 genes for HCM, 28 genes for DCM and 8 genes for ARVC. Not all patients were analysed for every gene. Variants were classified according to national guidelines as "highly likely to be pathogenic" (Class 5), "likely to be pathogenic" (Class 4), or "variant of unknown significance" (VUS) (Class 3).[154] Variant data was extracted from the anonymised contents of an in-house variant database.

2.2.1.2 Laboratory of Molecular Medicine, Partners Healthcare (LMM)

Data from the Laboratory of Molecular Medicine, Partners Healthcare (LMM), was downloaded from the supplemental files from published HCM[27] (18 genes sequenced in 632 to 2912 patients) and DCM[77] (46 genes sequenced in 121 to 756 patients) cohorts. No published ARVC data was available from this laboratory.

The LMM HCM cohort comprised unrelated cases referred for HCM clinical genetic testing. Individuals with an unclear clinical diagnosis of HCM, or with left ventricular hypertrophy due to an identified syndrome such as Fabry or Danon disease, or unaffected individuals with a family history of HCM were excluded.

The LMM DCM cohort comprised individual cases referred for DCM clinical genetic testing. According to the published report, all patients had DCM or clinical features consistent with DCM based on the medical and family history information provided by ordering providers. Additionally, any cases with confirmed diagnoses of other cardiomyopathies, structural heart disease, congenital heart disease or syndromic or environmental causes were not included.

Variants were classified as “pathogenic”, “likely pathogenic”, “VUS favour pathogenic” or “other” (other VUS, likely benign) according to the LMM's clinical grade variant classification criteria.[77]

2.2.1.3 Variant classification in clinical cohorts

OMGL and LMM used similar clinical guidelines and employed equivalent approaches for variant classification in line with published guidelines[128], however different names are used for each class. In this study, the term “Pathogenic” (P), includes OMGL Class 5, and LMM Pathogenic; “Likely pathogenic” (LP), includes OMGL Class 4 and LMM Likely

pathogenic; and "Variant of Uncertain significance" (VUS), includes OMGL Class 3 and the LMM VUS favour pathogenic and other VUS.

2.2.1.4 Exome Aggregation Consortium (ExAC)

Variant Call Format (VCF) data for each gene of interest was downloaded from the Exome Aggregation Consortium (ExAC), Cambridge, MA (http://exac.broadinstitute.org) [version 0.3, Jan 2015]. Only genes with a median sequence depth of >20x and only high quality (PASS filter) variants were included in analyses. In addition, the total number of ExAC samples per gene was based on the mean coverage at the variant sites.

2.2.2 Deriving a minor allele frequency threshold

Variant data from the clinical cases series was reviewed to determine the most frequent confirmed pathogenic variants in these cohorts. The upper 95% CI of the binomial distribution of the most frequently occurring pathogenic variant in these two large cases series was used to define an upper bound rare variant threshold to apply in these analyses.

2.2.3 Calculating rare variant frequency

Only likely protein-altering variants in designated canonical transcripts were analysed, i.e. missense, in-frame insertions/deletions, frameshift, nonsense and variants affecting the splice donor and acceptor regions (first and last two bases of each intron). Analyses were performed on all protein-altering variants and separately on variants predicted to be non-truncating (missense and in frame insertions and deletions) and variants predicted to be truncating (frameshift, nonsense, splice donor/acceptor).

Dataset	Cohort Type	Reference	No. individuals sequenced	No. genes tested	Methods used to generate sequence data
OMGL HCM	Clinical	This study	807 - 3267	16	High-resolution DNA melting (WAVE dHPLC, LightScanner®) Sanger sequencing Targeted NGS
LMM HCM	Clinical	[27]	632 - 2912	18	DNA microarray (Cardiochip) Sanger sequencing. Targeted NGS
Combined HCM	Clinical	-	632 - 6179	20	As detailed above
OMGL DCM	Clinical	This study	304 - 559	28	High-resolution DNA melting (WAVE dHPLC, LightScanner®) Sanger sequencing Targeted NGS
LMM DCM	Clinical	[123]	121 - 756	46	DNA microarray (Cardiochip) Sanger sequencing Targeted NGS
Combined DCM	Clinical	-	121 - 1315	48	As detailed above
OMGL ARVC	Clinical	This study	93 - 361	8	High-resolution DNA melting (WAVE dHPLC, LightScanner®) Sanger sequencing Targeted NGS
ExAC	Database	[138]	Up to 60706	all	Exome sequencing

Table 2.2-1 Overview of the data sets analysed in this study.
OMGL= Oxford Medical Genetics Laboratories. LMM=Laboratory for Molecular Medicine (LMM). ExAC=Exome Aggregation Consortium.

2.2.3.1 Clinical cohorts

The frequency of rare variation in the cardiomyopathy cohorts was calculated by dividing the sum of rare variants identified in cardiomyopathy cases (rare variants defined as an allele frequency below the selected MAF threshold), by the total number of patients analysed for each gene. In total, after combining data from both clinical laboratories and excluding genes poorly covered in ExAC, 20 genes sequenced in 632 - 6179 HCM patients, 46 genes sequenced in 121 - 1315 DCM patients and 8 genes sequenced in 93 - 361 ARVC patients were analysed (Table 2.2-1). Fisher's Exact Test was performed to compare the proportion of cases with rare putative pathogenic variants in each gene between the Oxford and LMM cohorts.

2.2.3.2 ExAC

For each gene, the overall frequency of ExAC variants with a minor allele frequency (MAF) below the selected threshold was calculated by dividing the sum of the adjusted allele count by the mean of the total adjusted alleles. To minimise potential biases that might arise due to false positive results arising from technical artefacts, the denominator for reference controls (i.e. the total number of ExAC samples per gene) was adjusted based on the mean coverage across rare variant sites of interest, and only high quality (PASS filter) variants in uniformly well-sequenced genes were included in analyses.

As ethnicity data was not available for most of the cardiomyopathy datasets, the full ExAC dataset, comprising individuals from a range of major global ancestry groups with a majority of European descent (52% non-Finnish European), was selected for comparison.

Bar charts were created to show the proportion of individuals in each case cohort with a rare variant in each gene, alongside the proportion in ExAC. Plots were created using GraphPad software V6.0.

2.2.4 Case-control analyses

For the purposes of these analyses, two simplifying assumptions were made: (1) that the frequency of rare benign variants was equivalent in cases and ExAC and (2) that the frequency of pathogenic variants in ExAC is sufficiently low so as not to affect this comparison.

For each gene and variant class, the proportion of cases with a rare variant was compared to that observed in ExAC using Fisher's exact test. The odds ratio (OR), with 95% confidence intervals was also calculated. For each gene set, the significance threshold was adjusted for multiple testing using Bonferroni correction (HCM $p<0.0025$, DCM $p<0.001$, ARVC $p<0.006$). Where no rare variants were detected in cases, a Haldane contingency correction[155] was applied (0.5 added to all cells in 2x2 table).

Statistical analyses were undertaken using Stata statistical analysis software (StataCorp. 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP).

2.2.4.1 Assessment of population-specific effects

To assess for any confounding effects of population stratification between cases and controls, the LMM DCM cohort (the only cardiomyopathy dataset for which individual, self-reported ethnicity data was available) was compared to ExAC across all case and control samples and separately for "white" cases (n=456 individuals) versus the non-Finnish European subset of ExAC (n=33,370). Discrepancies were defined as any gene which had a significant case excess in one of the analyses but not the other.

2.2.4.2 Variant distribution

The linear distribution of rare missense variants in the HCM, DCM and ARVC genes was compared between cases and controls. Statistical comparison of variant distribution between cases and ExAC was undertaken using the Kolomogorov-Smirnov Test (KS)[1]. Kernel Density Plots (KDP)[2] were generated to provide an overview of the spread and distribution of rare missense variants along the length of each gene in each cohort. Statistical analysis and plot generation was performed in R version 3.2.0. Additionally, non-random mutation cluster (NMC)[3] analysis was used to identify clusters of amino acid variants across the protein sequence in each cohort.[156] These analyses were performed using the iPAC Bioconductor R package, https://www.bioconductor.org/packages/devel/bioc/html/iPAC.html).

[1] The KS test is used to determine if two datasets differ significantly. This is a non-parametric test and so makes no assumption about the distribution of data. The null hypothesis for this test is that there is no difference in data distribution between datasets. Cumulative Fraction plots are made for each dataset; these provide a graphical display of how the data is distributed. The maximum vertical deviation between the two curves is reflected in the D statistic and the corresponding P value suggests a significant difference.

[2] KDP can be used as an alternative to histograms, providing a more effective way to illustrate the distribution of a variable. This method overcomes the problems encountered when using histograms where the shape of the distribution can be very sensitive to bin size. KDP give a smoothed version of the data. This method involves creating a Probability Density Function (PDF) for each data point (a kernel) and then summing the kernels to generate a kernel density estimate across a region. Essentially the more data points in a region, the higher the kernel density, and vice versa for regions with few data points. As these plots don't rely on defining the correct bins for the dataset they are less prone to data skewing and give a more accurate overview of data distribution.

[3] NMC is a probability model derived using order statistics which assume that the location of amino acid variants on a protein follows a uniform distribution. This method provides a statistical measure of the differences between pair-wise order statistics. Variant clusters are identified by considering if pairwise variants are closer together than expected by chance alone.

2.3 Results

2.3.1 Deriving a minor allele frequency threshold

The single most common confirmed pathogenic variant in both the OMGL and LMM cohorts was *MYBPC3* c.1504C>T (p.Arg502Trp), found in 104/6179 HCM cases (1.7%, 95%CI 1.4-2.0%). Using the upper 95% CI of the binomial distribution of *MYBPC3* c.1504C>T p.(Arg502Trp) in this case cohort (2%), and considering a disease prevalence of 1 in 500[10], the estimated population allele frequency of this variant was calculated to be approximately $2x10^{-5}$. This variant was observed 3 times in ExAC (MAF $2.5x10^{-5}$), which was equivalent to this population frequency estimate. A minor allele frequency (MAF) threshold of $1x10^{-4}$ was selected for analyses. Applying this threshold, where the ExAC allele number was ≥80,000, it is expected that all variants with a population MAF of $2x10^{-5}$ will be retained for analysis.

2.3.2 Case-control analyses

There were no significant differences in the proportion of cases with a rare variant in the HCM and DCM panel genes within the OMGL and LMM clinical genetics laboratories, therefore data from these cohorts was combined (Table S3A). A summary of the genes included in the analyses and the sample numbers from each cohort is shown in Table 2.3-1.

Gene	Ensembl transcript	Refseq transcript	HCM clinical - OMGL	HCM clinical - LMM	DCM clinical - OMGL	DCM clinical - LMM	ARVC clinical - OMGL
ABCC9	ENST00000261201	NM_005691				590	
ACTC1	ENST00000290378	NM_005159	1535	2650	355	748	
ACTN2	ENST00000366578	NM_001103	807	632	305	590	
ANK2	ENST00000264366						
ANKRD1	ENST00000371697	NM_014391	807		305	121	
CASQ2	ENST00000261448	NM_001232				121	
CAV3	ENST00000343849	NM_033337				121	
CRYAB	ENST00000526180	NM_001885			304	121	
CSRP3	ENST00000533783	NM_003476	1535	632	355	590	
<CTF1>	*ENST00000279804*	*NM_001330*				*590*	
DES	ENST00000373960	NM_001927			304	590	93
DSC2	ENST00000280904	NM_024422			304	123	351
DSG2	ENST00000261590	NM_001943			304	123	354
DSP	ENST00000379802	NM_004415			304	123	352
DTNA	ENST00000348997	NM_032978				121	
EMD	ENST00000369842	NM_000117				590	
FHL1	ENST00000370690	NM_001449	1535		355		
FHL2	ENST00000358129	NM_201555			304	121	
GLA	ENST00000218516	NM_000169	1535	2165	355	186	
JUP	ENST00000393930	NM_002230			304	121	94
KCNE1	ENST00000399289	NM_000219					
KCNE2	ENST00000290310	NM_172201					
KCNH2	ENST00000262186	NM_000238					
KCNJ2	ENST00000535240	NM_000891					
KCNQ1	ENST00000155840	NM_000218					
LAMA4	ENST00000424408	NM_002290				121	
LAMP2	ENST00000200639	NM_002294	839	2451	309	223	
LDB3	ENST00000361373	NM_007078				740	

Gene	Ensembl transcript	Refseq transcript	HCM clinical - OMGL	HCM clinical - LMM	DCM clinical - OMGL	DCM clinical - LMM	ARVC clinical - OMGL
LMNA	ENST00000368300	NM_170707			304	740	93
MYBPC3	ENST00000545968	NM_000256	3267	2912	405	756	
MYH6	ENST00000356287	NM_002471				121	
MYH7	ENST00000355349	NM_000257	3200	2912	559	756	
MYL2	ENST00000228841	NM_000432	1535	2650	355	188	
MYL3	ENST00000395869	NM_000258	1535	2650	355	188	
MYLK2	ENST00000375994	NM_033118				121	
MYOZ2	ENST00000307128	NM_016599		632		121	
NEXN	ENST00000334785	NM_144573		632		156	
PDLIM3	ENST00000284770	NM_014476					
PKP2	ENST00000070846	NM_004572			304	123	361
PLN	ENST00000357525	NM_002667	1535	632	355	740	
PRKAG2	ENST00000287878	NM_016203	1535	2438	355	191	
<RBM20>	*ENST00000369519*	*NM_001134363*				*156*	
RYR2	ENST00000366574	NM_001035				121	
SCN5A	ENST00000333535	NM_198056			304		
SGCD	ENST00000337851	NM_000337				590	
TAZ	ENST00000299328	NM_000116				740	
TCAP	ENST00000309889	NM_003673				590	
TGFB3	ENST00000238682	NM_003239					
TMEM43	ENST00000306077	NM_024334			304	123	94
TNNC1	ENST00000232975	NM_003280		632		156	
TNNI3	ENST00000344887	NM_000363	3135	2912	483	756	
TNNT2	ENST00000367318	NM_001001430	3191	2912	498	756	
TPM1	ENST00000403994	NM_001018005	1535	2912	355	756	
TTN	ENST00000589042	NM_001267550			304	156	
TTR	ENST00000237014	NM_000371		632		121	

Gene	Ensembl transcript	Refseq transcript	HCM clinical - OMGL	HCM clinical - LMM	DCM clinical - OMGL	DCM clinical - LMM	ARVC clinical - OMGL
VCL	ENST00000211998	NM_014000				590	

Table 2.3-1 Genes and transcripts analysed in this study and the number of patients sequenced in each disease cohort. The *CTF1* and *RBM20* DCM panel genes were removed from analyses due to poor sequence coverage in ExAC

Variants identified in cases (n=7855) were collated by disease and gene. Overall, the yield of pathogenic (P) and likely pathogenic (LP) variants was 32% in HCM, 13% in DCM (although TTN was only sequenced in a third of samples), and 36% in ARVC.

2.3.2.1 HCM

The proportion of HCM cases and ExAC reference controls with a rare variant (ExAC MAF < $1x10^{-4}$) in each of the 20 HCM genes is shown in Figure 2.3-1. The results of Fishers Exact Test and Odds Ratios analyses are shown in Table 2.3-2 and Figure 2.3-3.

A high proportion of HCM cases had a rare variant in one of the two major HCM genes, *MYBPC3* (19.0%) and *MYH7* (14.2%). A smaller proportion (≤2% cases per gene) had a rare variant in one of the other well-characterised HCM genes (*ACTC1, FHL1, MYL2, MYL3, PLN TNNI3, TNNT2, TPM1)* or genes which cause metabolic disorders which can mimic HCM (*GLA, LAMP2, PRKAG2*).

There was strong evidence ($p<0.0001$) supporting a case excess of rare non-truncating variants in *ACTC1*, *FHL1*, *GLA*, *MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *TNNI3*, *TNNT2*, *TPM1*. There was also evidence to suggest that non-truncating variants in *PRKAG2* were in excess in HCM cases ($p =0.0002$). The proportion of HCM cases with a truncating variant in *LAMP2*, *MYBPC3*, *PLN* and *TNNT2* was significantly higher than in ExAC reference controls ($p<0.0001$).

In the *ACTN2, ANKRD1, CSRP3, MYOZ2, NEXN, TNNC1*, and *TTR* genes, no evidence was found to suggest that there was a difference in the proportion of HCM cases with a rare variant compared to ExAC controls.

2.3.2.2 DCM

The proportion of DCM cases and ExAC reference controls with a rare variant in each of the DCM genes is shown in Figure 2.3-2. The results of Fishers Exact Test and Odds Ratios analyses are shown in Table 2.3-3 and Figure 2.3-3. The *CTF1* and *RBM20* genes were removed from analyses due to very poor sequence coverage in ExAC; therefore, case control analyses were performed in 46 genes in total. Data relating to non-truncating *TTN* variants were not available from the clinical laboratories; therefore for this gene, case control analyses were performed for truncating variants only.

In the DCM cases, truncating variants in *TTN* were found in the highest proportion of cases (14.6%). The proportion of cases with a rare variant in other well-characterised DCM genes was more modest (*MYH7* 5.3%, *LMNA* 4.4%, *TNNT2* 2.9%, and *TPM1* 1.9%). In several genes, no rare variants were found in cases (*ANKRD1, CAV3, FHL1, GLA, MYL2, MYOZ2* and *TTR*).

These analyses found strong evidence ($p < 0.0001$) to suggest that there was a case excess of rare truncating variants in the *DSP, LMNA, TTN* and *VCL* genes, and rare non-truncating variants in *MYH7*, *LMNA*, *TNNT2* and *TPM1*. There was also some evidence to suggest that there was a case excess of rare non–truncating variants in the *TCAP* (p=0.0004) and *TNNC1* (p=0.0001) genes, and truncating variants in *SCN5A* (p=0.001). However, notably, the *TNNC1* gene had only been analysed in 156 cases.

In the remaining 36 genes, no evidence was found to suggest there was a difference in the proportion of cases with a rare variant compared to ExAC. Although, in 10 of these genes (*CASQ2*, *CAV3*, *DTNA*, *LAMA4*, *MYH6*, *MYLK2*, *MYOZ2*, *NEXN, RYR2*, *TTR*), case numbers were very low (n=121 to n=156).

2.3.2.3 ARVC

The proportion of ARVC cases and ExAC reference controls with a rare variant in each of the ARVC genes is shown in Figure 2.3-1. The results of Fishers Exact Test and Odds Ratios analyses in ARVC genes are shown in Table 2.3-4 and Figure 2.3-3.

A high proportion of ARVC cases had a rare truncating variant in the *PKP2* gene (27%). Approximately 12% of ARVC cases had a rare variant in *DSP*. Variants in the two other commonly tested ARVC genes, *DSG2* and *DSC2*, were detected in 2% and 3% of cases respectively.

There was strong evidence ($p < 0.0001$) that rare truncating variants in *PKP2, DSG2*, and *DSP* were in excess in ARVC cases compared to ExAC. There was also evidence to suggest that truncating variants in *DSC2* were in excess in cases ($p = 0.005$).

The proportion of ARVC cases with non-truncating variants in *DSG2*, *DSP* was significantly higher than in controls ($p = 0.001$)

Fewer than 100 ARVC cases had undergone testing for *JUP*, *LMNA*, *TMEM43* and *DES*; therefore, no robust conclusions can be made about the role of rare variation in these genes in this ARVC cohort.

2.3.2.4 Summary

Table 2.2-1 provides an overview of the results of case control analyses within each cohort.

Hypertrophic cardiomyopathy (HCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**OR**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
ACTC1	All protein-altering	22	4163	39	60159	8.15	4.83	13.76	<0.0001
ACTN2	All protein-altering	10	1429	656	59732	0.64	0.34	1.19	0.1941
ANKRD1	All protein-altering	4	803	214	60086	1.40	0.52	3.77	0.5408
CSRP3	All protein-altering	11	2156	197	60450	1.57	0.85	2.88	0.1766
FHL1	All protein-altering	14	1521	74	60194	7.49	4.22	13.29	<0.0001
GLA	All protein-altering	27	3673	59	59345	7.39	4.68	11.67	<0.0001
LAMP2	All protein-altering	21	3269	119	59868	3.23	2.03	5.15	<0.0001
MYBPC3	All protein-altering	1176	5003	929	46226	11.70	10.68	12.81	<0.0001
MYH7	All protein-altering	869	5243	845	59626	11.70	10.60	12.91	<0.0001
MYL2	All protein-altering	46	4139	107	60409	6.27	4.43	8.88	<0.0001
MYL3	All protein-altering	37	4148	110	60491	4.91	3.38	7.13	<0.0001
MYOZ2	All protein-altering	1	631	155	60281	0.62	0.09	4.41	1.0000
NEXN	All protein-altering	8	624	448	58829	1.68	0.83	3.40	0.1584
PLN	All protein-altering	7	2160	29	60446	6.75	2.96	15.44	0.0002
PRKAG2	All protein-altering	42	3931	307	57440	2.00	1.45	2.76	0.0001
TNNC1	All protein-altering	0	632	35	59121	1.32	0.08	21.49	1.0000
TNNI3	All protein-altering	135	5912	123	53733	9.98	7.80	12.75	<0.0001
TNNT2	All protein-altering	119	5984	138	56786	8.18	6.39	10.47	<0.0001
TPM1	All protein-altering	66	4381	51	58676	17.33	12.01	25.01	<0.0001
TTR	All protein-altering	2	630	85	60550	2.26	0.56	9.21	0.2264
ACTC1	Truncating	0	4185	2	60196	2.88	0.14	59.93	1.0000
ACTN2	Truncating	1	1438	6	60380	7.00	0.84	58.17	0.1520
ANKRD1	Truncating	1	806	31	60268	2.41	0.33	17.69	0.3466
CSRP3	Truncating	3	2164	22	60625	3.82	1.14	12.77	0.0536
FHL1	Truncating	1	1534	1	60268	117.83	4.80	2893.68	0.0248

Hypertrophic cardiomyopathy (HCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**OR**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
GLA	Truncating	0	3700	0	59404	Not computed			
LAMP2	Truncating	13	3277	13	59986	494.17	29.37	8315.01	<0.0001
MYBPC3	Truncating	566	5613	40	47113	118.77	86.08	163.86	<0.0001
MYH7	Truncating	5	6107	29	60441	1.71	0.66	4.41	0.2356
MYL2	Truncating	3	4182	14	60502	3.10	0.89	10.79	0.0932
MYL3	Truncating	0	4185	2	60599	2.90	0.14	60.33	1.0000
MYOZ2	Truncating	0	632	12	60424	3.82	0.23	64.61	1.0000
NEXN	Truncating	1	631	38	59239	2.47	0.34	18.02	0.3388
PLN	Truncating	4	2163	4	60471	27.96	6.99	111.86	<0.0001
PRKAG2	Truncating	2	3971	7	57740	4.15	0.86	20.01	0.1103
TNNC1	Truncating	0	632	2	59154	18.70	0.90	390.03	1.0000
TNNI3	Truncating	6	6041	22	53834	2.43	0.99	6.00	0.0571
TNNT2	Truncating	11	6092	17	56906	6.04	2.83	12.91	<0.0001
TPM1	Truncating	0	4447	2	58725	2.64	0.13	55.02	1.0000
TTR	Truncating	0	632	3	60631	13.69	0.71	265.40	1.0000
ACTC1	Non-truncating	22	4163	37	60161	8.59	5.06	14.58	<0.0001
ACTN2	Non-truncating	9	1430	650	59738	0.58	0.30	1.12	0.1172
ANKRD1	Non-truncating	3	804	183	60117	1.23	0.39	3.84	0.7398
CSRP3	Non-truncating	8	2159	175	60472	1.28	0.63	2.60	0.4189
FHL1	Non-truncating	13	1522	74	60194	6.95	3.85	12.55	<0.0001
GLA	Non-truncating	27	3673	59	59345	7.39	4.68	11.67	<0.0001
LAMP2	Non-truncating	8	3282	119	59868	1.23	0.60	2.51	0.5454
MYBPC3	Non-truncating	610	5569	889	46266	5.70	5.12	6.34	<0.0001
MYH7	Non-truncating	864	5248	816	59655	12.04	10.89	13.30	<0.0001

Hypertrophic cardiomyopathy (HCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**OR**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
MYL2	Non-truncating	43	4142	93	60423	6.74	4.69	9.70	<0.0001
MYL3	Non-truncating	37	4148	108	60493	5.00	3.43	7.27	<0.0001
MYOZ2	Non-truncating	1	631	143	60293	0.67	0.09	4.78	1.0000
NEXN	Non-truncating	7	625	410	58867	1.61	0.76	3.41	0.2192
PLN	Non-truncating	3	2164	25	60450	3.35	1.01	11.11	0.0713
PRKAG2	Non-truncating	40	3933	300	57447	1.95	1.40	2.71	0.0002
TNNC1	Non-truncating	0	632	33	59123	1.40	0.09	22.80	1.0000
TNNI3	Non-truncating	129	5918	101	53755	11.60	8.93	15.07	<0.0001
TNNT2	Non-truncating	108	5995	121	56803	8.46	6.52	10.98	<0.0001
TPM1	Non-truncating	66	4381	49	58678	18.04	12.45	26.14	<0.0001
TTR	Non-truncating	2	630	82	60553	2.34	0.58	9.55	0.2150

Table 2.3-2 Case-control analyses Hypertrophic Cardiomyopathy (HCM) genes. Odds ratios and Fisher's Exact test comparing HCM cases to ExAC controls (Fisher's Exact test level of significance = 0.05, with Bonferroni correction for 20 tests = 0.0025). For cells with zero values, 0.5 was added to all cells before calculating the odds ratio (highlighted in grey)[155].

Dilated cardiomyopathy (DCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
ABCC9	All protein-altering	8	582	528	59925	1.56	0.77	3.15	0.1847
ACTC1	All protein-altering	5	1098	39	60159	7.02	2.76	17.86	0.0011
ACTN2	All protein-altering	8	887	656	59732	0.82	0.41	1.65	0.7440
ANKRD1	All protein-altering	0	426	214	60086	0.33	0.02	5.28	0.4119
CASQ2	All protein-altering	2	119	304	59650	3.30	0.81	13.40	0.1268
CAV3	All protein-altering	0	121	83	60179	2.97	0.18	48.09	1.0000
CRYAB	All protein-altering	2	423	105	50387	2.27	0.56	9.22	0.2248
CSRP3	All protein-altering	4	941	197	60450	1.30	0.48	3.52	0.5559
DES	All protein-altering	9	885	241	50887	2.15	1.10	4.19	0.0429
DSC2	All protein-altering	2	425	579	59485	0.48	0.12	1.94	0.4509
DSG2	All protein-altering	3	424	774	58893	0.54	0.17	1.68	0.3877
DSP	All protein-altering	32	395	1876	57744	2.49	1.73	3.59	<0.0001
DTNA	All protein-altering	1	120	263	59956	1.90	0.26	13.65	0.4120
EMD	All protein-altering	1	589	98	58965	1.02	0.14	7.34	0.6265
FHL1	All protein-altering	0	355	74	60194	1.14	0.07	18.37	1.0000
FHL2	All protein-altering	1	424	184	59044	0.76	0.11	5.41	1.0000
GLA	All protein-altering	0	541	59	59345	0.92	0.06	14.92	1.0000
JUP	All protein-altering	2	423	669	55227	0.39	0.10	1.57	0.2550
LAMA4	All protein-altering	6	115	1363	58378	2.23	0.98	5.09	0.0599
LAMP2	All protein-altering	2	530	119	59868	1.90	0.47	7.70	0.2878
LDB3	All protein-altering	11	729	653	57513	1.33	0.73	2.42	0.3760
LMNA	All protein-altering	46	998	320	50946	7.34	5.35	10.06	<0.0001
MYBPC3	All protein-altering	29	1132	929	46226	1.27	0.88	1.85	0.2002
MYH6	All protein-altering	3	118	1487	58841	1.01	0.32	3.17	1.0000
MYH7	All protein-altering	69	1246	845	59626	3.91	3.04	5.03	<0.0001
MYL2	All protein-altering	1	542	107	60409	1.04	0.15	7.48	0.6192

Dilated cardiomyopathy (DCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
MYL3	All protein-altering	0	543	110	60491	0.50	0.03	8.11	1.0000
MYLK2	All protein-altering	2	119	448	56778	2.13	0.52	8.64	0.2456
MYOZ2	All protein-altering	0	121	155	60281	1.60	0.10	25.77	1.0000
NEXN	All protein-altering	5	151	448	58829	4.35	1.78	10.65	0.0071
PKP2	All protein-altering	7	420	793	57613	1.21	0.57	2.56	0.5287
PLN	All protein-altering	4	1091	29	60446	7.64	2.68	21.78	0.0027
PRKAG2	All protein-altering	2	544	307	57440	0.69	0.17	2.77	1.0000
RYR2	All protein-altering	6	115	1871	51821	1.45	0.64	3.29	0.3207
SCN5A	All protein-altering	7	297	1343	55077	0.97	0.46	2.05	1.0000
SGCD	All protein-altering	4	586	171	50554	2.02	0.75	5.46	0.1437
TAZ	All protein-altering	3	737	48	51273	4.35	1.35	13.99	0.0360
TCAP	All protein-altering	7	583	134	55099	4.94	2.30	10.60	0.0008
TMEM43	All protein-altering	3	424	439	59782	0.96	0.31	3.01	1.0000
TNNC1	All protein-altering	3	153	35	59121	33.12	10.08	108.84	0.0001
TNNI3	All protein-altering	8	1231	123	53733	2.84	1.39	5.82	0.0098
TNNT2	All protein-altering	36	1218	138	56786	12.16	8.39	17.63	<0.0001
TPM1	All protein-altering	21	1090	51	58676	22.17	13.29	36.98	<0.0001
TTN	All protein-altering	not assessed							
TTR	All protein-altering	0	121	85	60550	2.91	0.18	47.24	1.0000
VCL	All protein-altering	13	577	604	58789	2.19	1.26	3.82	0.0111
ABCC9	Truncating	1	589	44	60408	2.33	0.32	16.95	0.3542
ACTC1	Truncating	1	1102	2	60196	27.31	2.47	301.44	0.0530
ACTN2	Truncating	0	895	6	60380	5.19	0.29	92.14	1.0000
ANKRD1	Truncating	0	426	31	60268	2.24	0.14	36.72	1.0000

Dilated cardiomyopathy (DCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
CASQ2	Truncating	1	120	27	59926	18.50	2.49	137.21	0.0549
CAV3	Truncating	0	121	8	60254	29.17	1.67	508.25	1.0000
CRYAB	Truncating	0	425	5	50486	10.79	0.60	195.39	1.0000
CSRP3	Truncating	1	944	22	60625	2.92	0.39	21.68	0.2993
DES	Truncating	2	892	16	51110	7.16	1.64	31.20	0.0376
DSC2	Truncating	0	427	16	60046	4.26	0.25	71.06	1.0000
DSG2	Truncating	0	427	43	59623	1.60	0.10	26.08	1.0000
DSP	Truncating	12	415	42	59570	41.01	21.44	78.46	<0.0001
DTNA	Truncating	0	121	6	60213	38.12	2.14	680.47	1.0000
EMD	Truncating	0	590	0	59063	Not computed			
FHL1	Truncating	0	355	0	60268	Not computed			
FHL2	Truncating	0	425	6	59222	10.71	0.60	190.36	1.0000
GLA	Truncating	0	541	0	59404	Not computed			
JUP	Truncating	0	425	10	55881	6.25	0.37	106.90	1.0000
LAMA4	Truncating	0	121	60	59676	4.06	0.25	66.02	1.0000
LAMP2	Truncating	0	532	0	59986	Not computed			
LDB3	Truncating	0	740	29	58135	1.33	0.08	21.80	1.0000
LMNA	Truncating	16	1028	8	51257	99.72	42.58	233.53	<0.0001
MYBPC3	Truncating	0	1161	40	47113	0.50	0.03	8.15	1.0000
MYH6	Truncating	0	121	76	60248	3.24	0.20	52.59	1.0000
MYH7	Truncating	1	1314	29	60441	1.59	0.22	11.65	0.4756
MYL2	Truncating	0	543	14	60502	3.84	0.23	64.43	1.0000
MYL3	Truncating	0	543	2	60599	22.30	1.07	465.06	1.0000
MYLK2	Truncating	0	121	8	57216	27.70	1.59	482.62	1.0000
MYOZ2	Truncating	0	121	12	60424	19.89	1.17	337.89	1.0000

Dilated cardiomyopathy (DCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
NEXN	Truncating	1	155	38	59239	10.06	1.37	73.71	0.0975
PKP2	Truncating	0	427	43	58362	1.57	0.10	25.53	1.0000
PLN	Truncating	1	1094	4	60471	13.82	1.54	123.74	0.0858
PRKAG2	Truncating	0	546	7	57740	7.04	0.40	123.49	1.0000
RYR2	Truncating	0	121	59	53629	3.71	0.23	60.34	1.0000
SCN5A	Truncating	3	301	34	56381	16.53	5.05	54.10	0.0010
SGCD	Truncating	1	589	10	50715	8.61	1.10	67.37	0.1195
TAZ	Truncating	1	739	1	51320	69.45	4.34	1111.37	0.0282
TCAP	Truncating	0	590	14	55219	3.22	0.19	54.12	1.0000
TMEM43	Truncating	0	427	24	60195	2.87	0.17	47.33	1.0000
TNNC1	Truncating	0	156	2	59154	75.60	3.61	1581.11	1.0000
TNNI3	Truncating	0	1239	22	53834	0.97	0.06	15.92	1.0000
TNNT2	Truncating	0	1254	17	56906	1.30	0.08	21.56	1.0000
TPM1	Truncating	2	1109	2	58725	52.95	7.45	376.27	0.0020
TTN	Truncating	67	393	501	58112	19.77	15.04	26.00	<0.0001
TTR	Truncating	0	121	3	60631	71.29	3.66	1387.60	1.0000
VCL	Truncating	4	586	19	59372	21.33	7.23	62.89	<0.0001
ABCC9	Non-truncating	7	583	484	59969	1.49	0.70	3.15	0.2487
ACTC1	Non-truncating	4	1099	37	60161	5.92	2.11	16.63	0.0062
ACTN2	Non-truncating	8	887	650	59738	0.83	0.41	1.67	0.7436
ANKRD1	Non-truncating	0	426	183	60117	0.38	0.02	6.17	0.6431
CASQ2	Non-truncating	1	120	277	59677	1.80	0.25	12.90	0.4298
CAV3	Non-truncating	0	121	75	60187	3.28	0.20	53.24	1.0000
CRYAB	Non-truncating	2	423	100	50392	2.38	0.59	9.69	0.2095

Dilated cardiomyopathy (DCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
CSRP3	Non-truncating	3	942	175	60472	1.10	0.35	3.45	0.7558
CTF1	Non-truncating	1	589	24	60416	4.27	0.58	31.65	0.2157
DES	Non-truncating	7	887	225	50903	1.79	0.84	3.80	0.1248
DSC2	Non-truncating	2	425	563	59501	0.50	0.12	2.00	0.4491
DSG2	Non-truncating	3	424	731	58936	0.57	0.18	1.78	0.5021
DSP	Non-truncating	20	407	1834	57786	1.55	0.99	2.43	0.0660
DTNA	Non-truncating	1	120	257	59962	1.94	0.27	13.97	0.4049
EMD	Non-truncating	1	589	98	58965	1.02	0.14	7.34	0.6265
FHL1	Non-truncating	0	355	74	60194	1.14	0.07	18.37	1.0000
FHL2	Non-truncating	1	424	178	59050	0.78	0.11	5.60	1.0000
GLA	Non-truncating	0	541	59	59345	0.92	0.06	14.92	1.0000
JUP	Non-truncating	2	423	659	55237	0.40	0.10	1.59	0.2533
LAMA4	Non-truncating	6	115	1303	58438	2.34	1.03	5.33	0.0503
LAMP2	Non-truncating	2	530	119	59868	1.90	0.47	7.70	0.2878
LDB3	Non-truncating	11	729	624	57542	1.39	0.76	2.54	0.2778
LMNA	Non-truncating	30	1014	312	50954	4.83	3.31	7.06	<0.0001
MYBPC3	Non-truncating	29	1132	889	46266	1.33	0.92	1.94	0.1278
MYH6	Non-truncating	3	118	1411	58917	1.06	0.34	3.34	0.7623
MYH7	Non-truncating	68	1247	816	59655	3.99	3.09	5.14	<0.0001
MYL2	Non-truncating	1	542	93	60423	1.20	0.17	8.62	0.5684
MYL3	Non-truncating	0	543	108	60493	0.51	0.03	8.26	1.0000
MYLK2	Non-truncating	2	119	440	56786	2.17	0.53	8.80	0.2393
MYOZ2	Non-truncating	0	121	143	60293	1.73	0.11	27.94	1.0000
NEXN	Non-truncating	4	152	410	58867	3.78	1.39	10.25	0.0241
PKP2	Non-truncating	7	420	750	57656	1.28	0.60	2.71	0.5117

Dilated cardiomyopathy (DCM)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
PLN	Non-truncating	3	1092	25	60450	6.64	2.00	22.03	0.0132
PRKAG2	Non-truncating	2	544	300	57447	0.70	0.17	2.84	1.0000
RYR2	Non-truncating	6	115	1812	51880	1.49	0.66	3.40	0.3076
SCN5A	Non-truncating	4	300	1309	55111	0.56	0.21	1.51	0.3361
SGCD	Non-truncating	3	587	161	50564	1.61	0.51	5.04	0.4418
TAZ	Non-truncating	2	738	47	51274	2.96	0.72	12.19	0.1538
TCAP	Non-truncating	7	583	120	55113	5.51	2.56	11.87	0.0004
TMEM43	Non-truncating	3	424	415	59805	1.02	0.33	3.19	0.7711
TNNC1	Non-truncating	3	153	33	59123	35.13	10.66	115.77	0.0001
TNNI3	Non-truncating	8	1231	101	53755	3.46	1.68	7.12	0.0033
TNNT2	Non-truncating	36	1218	121	56803	13.88	9.52	20.22	<0.0001
TPM1	Non-truncating	19	1092	49	58678	20.84	12.23	35.51	<0.0001
TTN	Non-truncating	not assessed							
TTR	Non-truncating	0	121	82	60553	3.02	0.19	48.98	1.0000
VCL	Non-truncating	9	581	585	58808	1.56	0.80	3.02	0.2023

Table 2.3-3 Case-control analyses Dilated Cardiomyopathy (DCM) genes. Odds ratios and Fisher's Exact test comparing DCM cases to ExAC controls (Fisher's Exact test level of significance = 0.05, with Bonferroni correction for 48 tests = 0.001).For cells with zero values, 0.5 was added to all cells before calculating the odds ratio (highlighted in grey)[155]. The *CTF1* and *RBM20* genes were removed from analyses due to poor sequence coverage in ExAC.

Arrhythmogenic right ventricular cardiomyopathy (ARVC)									
Gene	**Variant Class**	**Cases with**	**Cases without**	**Controls with**	**Controls without**	**All Odds ratio**	**CI lower**	**CI upper**	**Fishers exact 2-sided**
DES	All protein-altering	3	90	241	50887	7.04	2.21	22.39	0.0101
DSC2	All protein-altering	9	342	579	59485	2.70	1.39	5.27	0.0080
DSG2	All protein-altering	17	337	774	58893	3.84	2.35	6.28	<0.0001
DSP	All protein-altering	43	309	1876	57744	4.28	3.10	5.91	<0.0001
JUP	All protein-altering	8	86	669	55227	7.68	3.71	15.91	<0.0001
LMNA	All protein-altering	0	93	320	50946	0.85	0.05	13.72	1.0000
PKP2	All protein-altering	101	260	793	57613	28.22	22.19	35.89	<0.0001
TMEM43	All protein-altering	0	94	439	59782	0.72	0.04	11.61	1.0000
DES	Truncating	0	93	16	51110	16.56	0.99	278.14	1.0000
DSC2	Truncating	2	349	16	60046	21.51	4.93	93.89	0.0048
DSG2	Truncating	5	349	43	59623	19.87	7.82	50.45	<0.0001
DSP	Truncating	21	331	42	59570	89.98	52.71	153.62	<0.0001
JUP	Truncating	0	94	10	55881	28.16	1.64	484.02	1.0000
LMNA	Truncating	0	93	8	51257	32.25	1.85	562.77	1.0000
PKP2	Truncating	95	266	43	58362	484.73	331.54	708.71	<0.0001
TMEM43	Truncating	0	94	24	60195	13.00	0.78	215.33	1.0000
DES	Non-truncating	3	90	225	50903	7.54	2.37	24.00	0.0084
DSC2	Non-truncating	7	344	563	59501	2.15	1.01	4.57	0.0504
DSG2	Non-truncating	12	342	731	58936	2.83	1.58	5.05	0.0017
DSP	Non-truncating	22	330	1834	57786	2.10	1.36	3.24	0.0018
JUP	Non-truncating	8	86	659	55237	7.80	3.76	16.16	<0.0001
LMNA	Non-truncating	0	93	312	50954	0.87	0.05	14.08	1.0000
PKP2	Non-truncating	6	355	750	57656	1.30	0.58	2.92	0.4767
TMEM43	Non-truncating	0	94	415	59805	0.76	0.05	12.29	1.0000

Table 2.3-4 Case-control analyses Arrhythmogenic right ventricular cardiomyopathy (ARVC) genes. Odds ratios and Fisher's Exact test comparing ARVC cases to ExAC controls (Fisher's Exact test level of significance = 0.05, with Bonferroni correction for 8 tests = 0.006).For cells with zero values, 0.5 was added to all cells before calculating the odds ratio (highlighted in grey)[155].

		Non-truncating				Truncating			
Cohort	**Total no. genes**	**Genes with case excess**	**Mean OR (SD)**	**Min OR**	**Max OR**	**Genes with case excess**	**Mean OR (SD)**	**Min OR**	**Max OR**
HCM	20	**11/20 (55%)** *ACTC1, FHL1* *GLA, MYBPC3* *MYH7,MYL2* *MYL3,PRKAG2* *TNNI3,TNNT2* *TPM1*	8.41 (± 4.28)	1.95 (*PRKAG2*)	18.04 (*TPM1*)	**4/20 (20%)** *LAMP2* *MYBPC3* *PLN* *TNNT2*	161.73 (±226.93)	6.04 (*TNNT2*)	494.17 (*LAMP2*)
DCM	46	**6/46(12.5%)** *LMNA* *MYH7* *TCAP* *TNNC1** *TNNT2* *TPM1*	14.03 (±12.24)	3.99 (*MYH7*)	35.13 (*TNNC1**)	**5 (10%)** *DSP* *LMNA* *SCN5A* *TTN* *VCL*	39.67 (± 34.91)	16.53 (*SCN5A*)	99.72 (*LMNA*)
ARVC	8	**4/8 (50%)** *DES* *DSG2* *DSP* *JUP^*	5.07 (± 3.02)	2.1 (*DSP*)	7.8 (*DES*)	**4(50%)** *DSC2* *DSG2* *DSP* *PKP2*	154.02 (± 222.88)	19.87 (*DSG2*)	484.73 (*PKP2*)

Table 2.3-5 Summary of case control analyses. Results shown for truncating and non-truncating variants. Genes with case excess= number of genes (and %) in which there was evidence to suggest that the proportion of cases with a rare variant was higher than in ExAC reference controls (Bonferroni corrected p value HCM p ≤0.0025, DCM p≤0.001, ARVC p≤0.006). Mean OR across genes with a case excess (±SD). Min OR=gene with lowest OR. Max OR=gene with highest OR. * Data available from 156 DCM cases. ^ Data available from 94 cases.

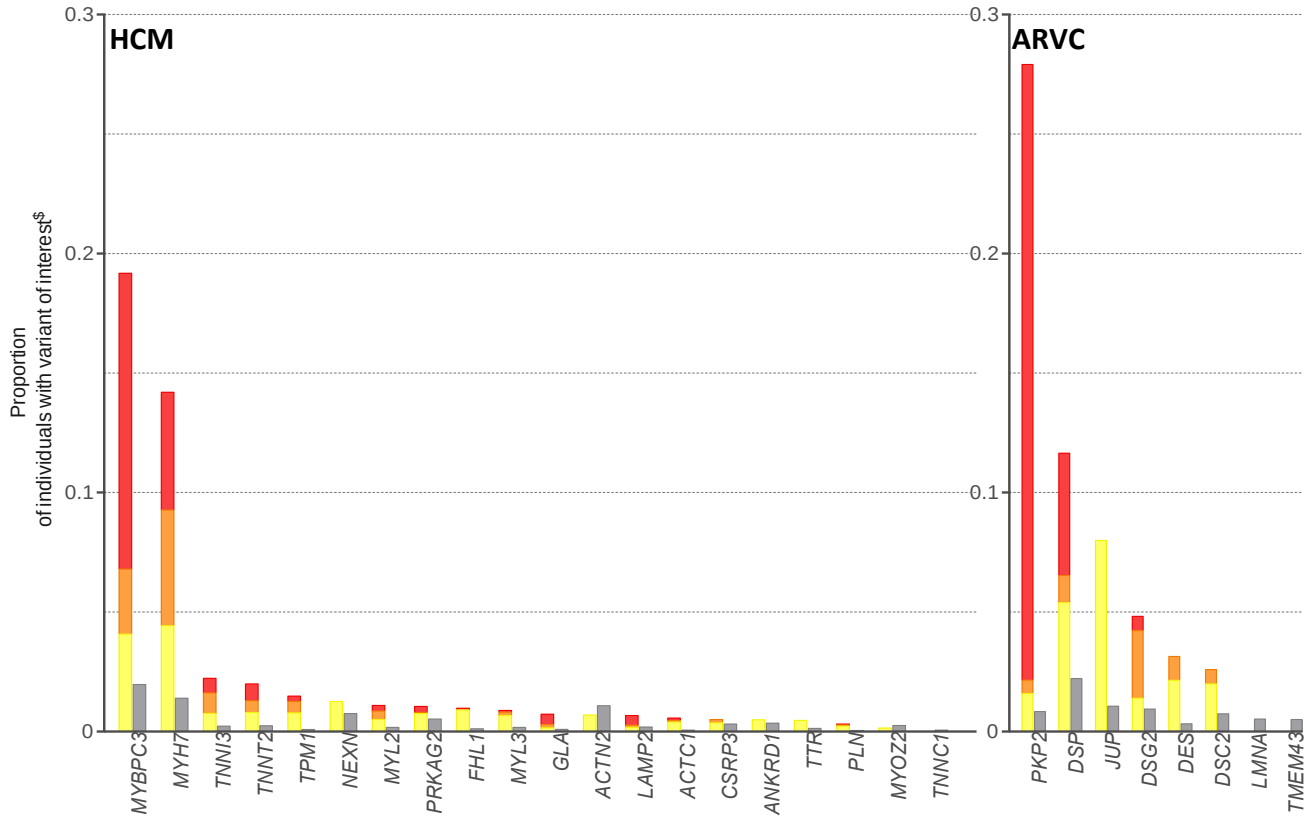


Figure 2.3-1 Proportion of individuals with a variant of interest in HCM and ARVC genes in combined clinical cohorts (data from OMGL and LMM) compared to ExAC (grey columns). $ Variant of interest is defined as a single nucleotide change or small insertion/deletion variant detected in the coding region ±2bp, with an ExAC minor allele frequency (MAF) of<$1x10^{-4}$. Clinical cohorts: HCM n=632 to 6179, ARVC n=93 to 361. Proportions are separated into reported pathogenicity class (red = pathogenic (P), orange = likely pathogenic (LP), yellow = variant of uncertain significance (VUS). ExAC: n=mean of total adjusted allele count for rare variant carriers. For HCM genes n ranged from 47153 to 60647, for ARVC genes n was 51126 to 60221.

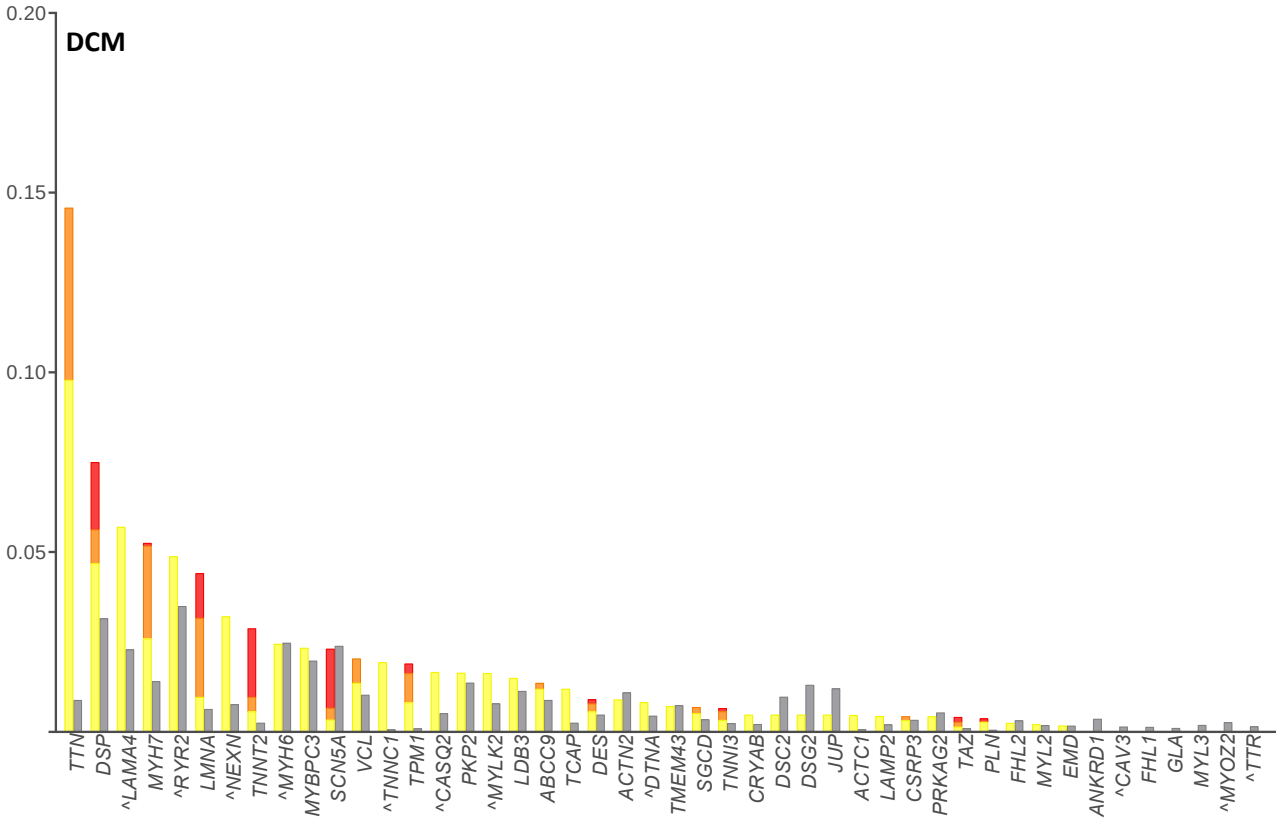


Figure 2.3-2 Proportion of individuals with a variant of interest in DCM genes in combined clinical cohorts (data from OMGL and LMM) compared to ExAC (grey columns). Variant of interest defined as a single nucleotide change or small insertion/deletion variant detected in the coding region ±2bp, with an ExAC minor allele frequency (MAF) of<$1x10^{-4}$. Clinical cohorts: DCM n=121 to 1315. Proportions are separated into reported pathogenicity class (red = pathogenic (P), orange = likely pathogenic (LP), yellow = variant of uncertain significance (VUS). ^ = genes analyzed in fewer than 200 cases. ExAC: n=mean of total adjusted allele count for rare variant carriers. For DCM genes n was 42697 to 60647. CTF1 and RBM20 were removed from analysis due to poor coverage in ExAC.

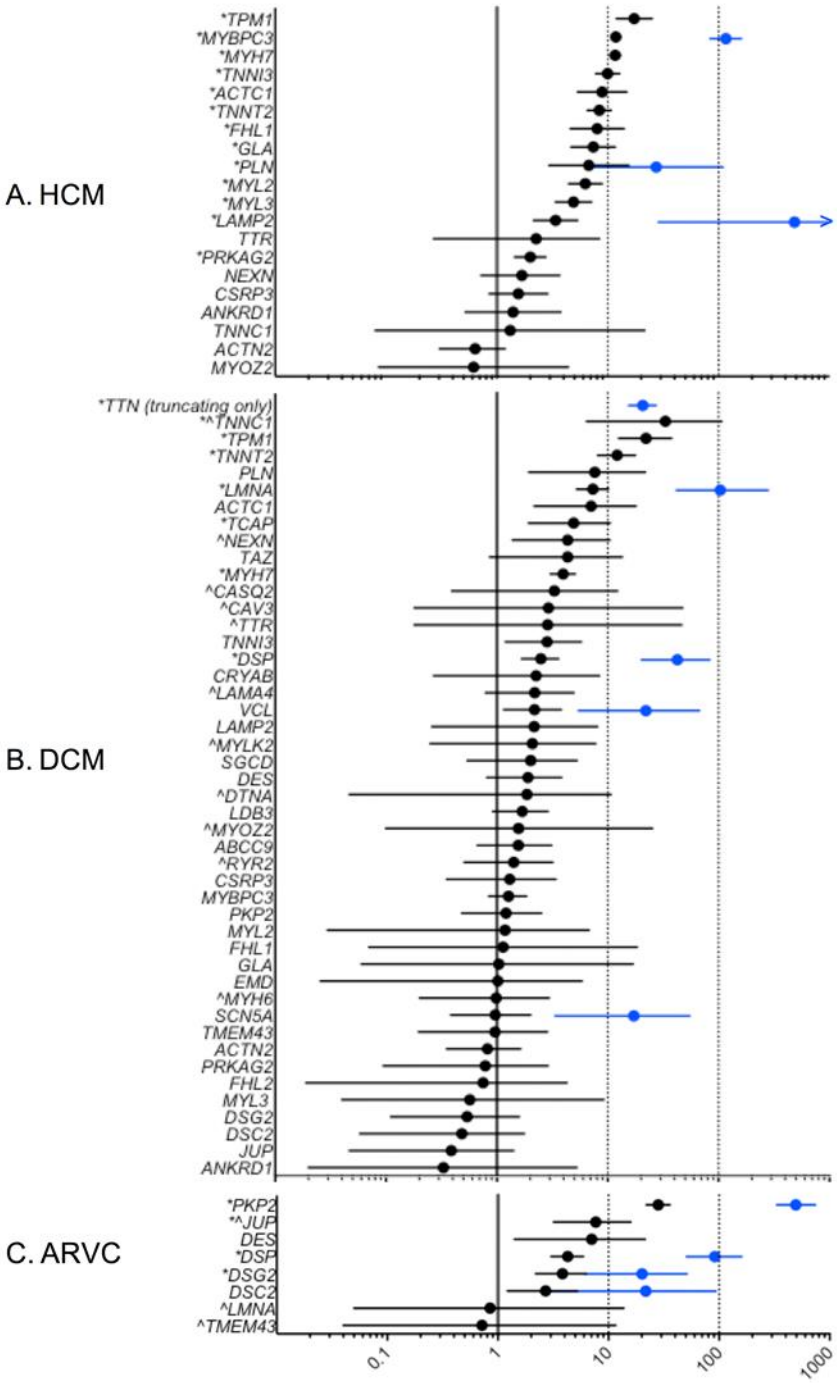


Figure 2.3-3 Odds ratios (OR) with 95% CI for each gene tested in the HCM (n=632 to 6179), DCM (n=121 to 1315) and ARVC (n=93 to 361) clinical cohorts compared to ExAC reference samples (n=mean of total adjusted allele count for rare variant carriers:HCM genes n=47153 to 60647, DCM genes n=42697 to 60647 and ARVC genes n=51126 to 60221). Data plotted (log10 scale) for all protein-altering variants (black) and truncating variants (blue). For truncating variants, OR with 95% CI have been plotted for genes where a statistically significant case excess was observed for this variant type on FET.
*= statistically significant Fishers exact test (FET) (p=0.05 with Bonferroni correction, HCM p ≤0.0025, DCM p ≤0.001 and ARVC p ≤0.006.) ^ =genes analysed in fewer than 200 cases. CTF1 and RBM20 were removed from analysis due to poor coverage in ExAC.

2.3.2.5 Assessment of population-specific effects

There was full correlation between the genes that were significantly enriched in DCM cases when comparing all case and ExAC control samples against the subset of DCM cases annotated as "white" with ExAC European (Non-Finnish).

2.3.2.6 Variant distribution

The results of the Kolmogorov-Smirnov (KS) tests are presented for genes in the HCM (Table 2.3-6), DCM (Table 2.3-8) and ARVC (Table 2.3-7) panels. Evidence supporting a difference in the linear distribution of rare missense variants was found for two genes in the HCM cohort, *MYH7* ($p= 1.48 \times 10^{-14}$) and *TNNI3* ($p=7.83 \times 10^{-8}$). No evidence was found for any of the other genes analysed (i.e. p values greater than a Bonferroni corrected p-value threshold).

Kernel Density Plots illustrate the differences in variant distribution between cases and ExAC (*MYH7* Figure 2.3-4 and *TNNI3* Figure 2.3-5). These findings were further supported by Non-random Mutation Cluster analyses (NMC) which identified significant variant clusters in both genes in HCM cases and not in ExAC (Table 2.3-9). In the *MYH7* gene, a distinct cluster incorporating a different region of the protein was detected in ExAC (Table 2.3-9).

Gene	Kolmogorov_Smirnov test statistic (D-value)	p-value
ACTN2	0.27649	0.5162
ANKRD1	0.34545	0.8768
ACTC1	0.18323	0.932
FHL1	0.39726	0.1245
GLA	0.40027	0.05753
MYBPC3	0.062581	0.7179
CSRP3	0.42884	0.2525
LAMP2	0.2	0.9866
MYH7	0.29877	***1.51E-14***
MYL2	0.17154	0.7265
MYL3	0.34284	0.04624
PLN	0.57143	0.3953
PRKAG2	0.29804	0.02156
TNNI3	0.63496	***2.44E-10***
TNNT2	0.34773	0.002587
TPM1	0.14338	0.7106

Table 2.3-6 Results of Kolmogorov-Smirnov analyses. HCM genes. Bonferroni corrected p-value=0.003 (0.05/16). *MYOZ2*, *TNNC1*, *TTR* genes not included in analyses due to low numbers of case variants.

Gene	Kolmogorov_Smirnov test statistic D-value)	p-value
DES	0.41667	0.6881
DSC2	0.22857	0.8669
DSG2	0.41032	0.1033
DSP	0.12323	0.9
JUP	0.423	0.1208
PKP2	0.27681	0.7568

Table 2.3-7 Results of Kolmogorov-Smirnov analyses. ARVC genes. Bonferroni corrected p-value=0.008 (0.05/6). TMEM43 and LMNA genes not included in analyses due to low numbers of case variants.

Gene	Kolmogorov_Smirnov test statistic D-value)	p-value
ABCC9	0.28373	0.6411
ACTC1	0.27174	0.8995
ACTN2	0.33721	0.3382
CSRP3	0.397	0.7502
DES	0.37255	0.4024
DSP	0.32007	0.07975
LAMA4	0.42264	0.2384
LDB3	0.2869	0.4663
LMNA	0.29167	0.05283
MYBPC3	0.23656	0.141
MYH7	0.21804	0.01571
NEXN	0.97768	0.2972
PKP2	0.40911	0.201
PLN	0.57143	0.6172
RYR2	0.23602	0.8937
SCN5A	0.56777	0.5441
SGCD	0.40824	0.7188
TCAP	0.54688	0.2102
TNNI3	0.55487	0.03995
TNNT2	0.375	0.1278
TPM1	0.35	0.07626
VCL	0.37138	0.2339

Table 2.3-8 Results of Kolmogorov-Smirnov analyses. DCM genes. Bonferroni corrected p-value=0.002 (0.05/22). Several genes were not included in analyses due to a very low number of (or no) rare variants detected in cases.

Gene	Cohort	Codons	p-value	Protein domains
MYH7	HCM	167-969	$1.4x10^{-14}$	Myosin Motor IQ Actin binding
MYH7	ExAC	1247-1931	$8.32x10^{-9}$	Coiled coil
TNNI3	HCM	123-209	$1.8x10^{-9}$	TnC and actin-tropomyosin binding

Table 2.3-9 Results of non-random mutation cluster analysis[157]. Protein domains as defined by UniProt (www.uniprot.org).

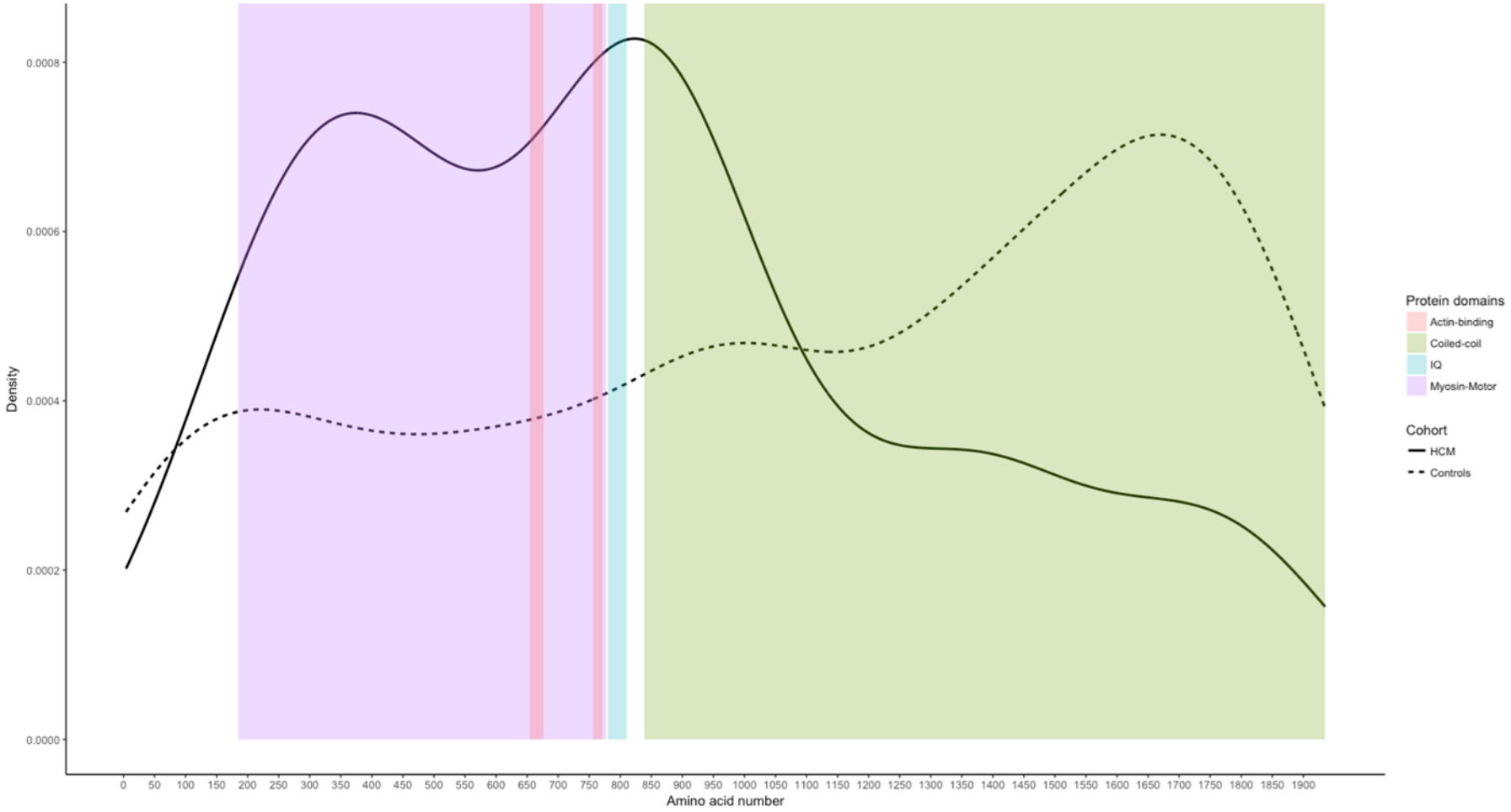


Figure 2.3-4 Kernel Density Plot showing distribution of rare missense variants in the *MYH7* gene in HCM cases (solid black line) and ExAC (dashed line). Protein domains annotated (see Legend).

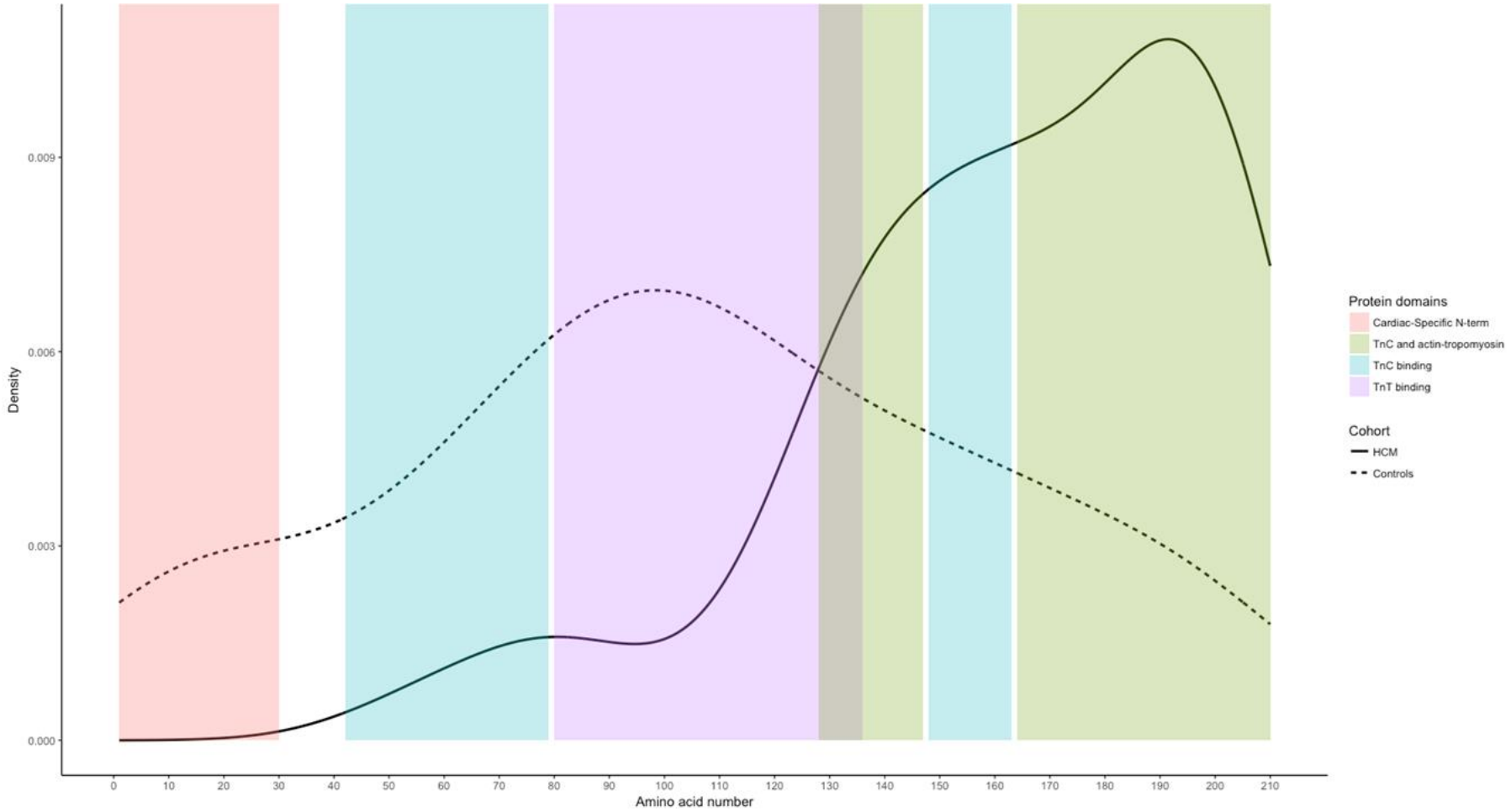


Figure 2.3-5 Kernel Density Plot showing distribution of rare missense variants in the *TNNI3* gene in HCM cases (solid black line) and ExAC (dashed line). Protein domains annotated (see Legend).

2.4 Discussion

Using data from two large clinical cardiomyopathy case series and ExAC, comparative analyses were undertaken to assess the strength of evidence supporting the contribution of rare variants in genes in current test panels and disease. The results of these analyses resoundingly validate the long-established genes and variant classes in these disorders. Also, they illustrate that in validated genes, there is scope to further refine variant interpretation and increase the diagnostic yield of clinical testing. Critically, these analyses highlight genes which appear to show little or no disease association and in which rare variants are uninterpretable.

2.4.1 Deriving a minor allele frequency threshold

The availability of case data from two large clinical cohorts provided a reliable estimate of the frequency of the most common pathogenic variant across inherited cardiomyopathy cases (*MYBPC3* p.Arg502Trp). Using this data, along with estimates of the population prevalence[10], a minor allele frequency (MAF) threshold of $1x10^{-4}$ was hypothesised to ensure these analyses focused on variants which are at a frequency compatible with pathogenicity. While not all variants below this threshold are expected to be pathogenic, variants more frequent than this in the general population would not be expected to be pathogenic.

Key to these analyses was the availability of the ExAC database, which was the first publically available reference dataset large enough to enable the assessment of variant alleles prevalent in the population at a range of 1:10,000-100,000, which are rare enough to be causal for low prevalence Mendelian diseases.

Previous rare variant analyses in cardiomyopathy genes applied more lenient thresholds for variant filtering, at least an order of magnitude greater than the MAF applied in this study.[56,141] Taking into consideration estimates of the population prevalence and genetic heterogeneity of these disorders, those MAF stringency thresholds were too liberal and analyses would inevitably have included variants that are too common to be disease causing. This has undoubtedly led to erroneous conclusions about the link between some genes and disease[56,126,151], the proportion of affected individuals with more than one pathogenic variant[56,126,151], and the frequency of putative pathogenic cardiomyopathy variants in the background population[141].

2.4.2 Case-control analyses

As expected, in each cardiomyopathy cohort, a high proportion of cases had a variant in one of the core well established genes, which form the basis of most clinical laboratory tests. Many of these genes have been implicated in these conditions for a number of years, having been identified through linkage studies in large pedigrees (e.g. *MYH7* in HCM[19,20] and DCM[158], *MYBPC3* in HCM [21], *TPM1*[18] and *TNNT2*[18] in HCM, *DSP* in ARVC [159]), or inferred from animal models of disease (e.g. *PKP2* in ARVC [160]), others are supported by more recent large case control and human genetic studies (e.g. *TTN* in DCM[67,161,162]).

For these genes, the knowledge base is generally more extensive and there is more evidence to support variant interpretation. This is illustrated by the higher proportion of variants classified by the clinical laboratories as "Likely" or "Highly Likely" pathogenic (Figure 2.3-1 and Figure 2.3-2). Notably, in several of these core genes, even variants classified as VUS, showed a several-fold case excess over the background in ExAC, indicating that there is scope to refine interpretation and improve diagnostic yield in these genes.

Beyond these core genes, the proportion of cases with a rare variant was lower, and the number of variants classified as VUS was higher (Figure 2.3-1 and Figure 2.3-2). This was particularly apparent in the DCM cohort, reflecting the clinical and genetic heterogeneity of this condition[61].

Integral to these analyses was the ability to assess the proportion of cases with a rare variant within the context of background levels of rare variation in ExAC (MAF ≤ $1x10^{-4}$). This not only confirmed whether the observed differences between cases and ExAC controls were statistically significant, but gave a clear indication of the strength of the association between rare variants in each gene and the corresponding cardiomyopathy. This gene level information is critical for variant interpretation, as our ability to evaluate the pathogenicity of the novel rare variants found in cases depends on the signal to noise ratio.[142,163]

2.4.2.1 Exploring the utility of case control findings

For a number of genes, particularly in the DCM panel, data was only available from a small number of cases, limiting the conclusions that can be drawn from these analyses. Data from larger cases series will be required to explore the contribution of rare variants in these genes in each cardiomyopathy.

In several genes, despite data from a reasonable number of cases, no evidence was found to suggest that the proportion of individuals with a rare variant was different in cases compared to ExAC. However, to understand the implications of these findings, and the conclusions that can be drawn from them, it is important to assess this data within the context of existing evidence for each gene-disease relationship. Some illustrative examples are provided below.

2.4.2.1.1 CSRP3

In some genes, it is possible that only a very small proportion of rare variants are pathogenic, and that these analyses were not sufficiently powered to detect very low level of enrichment in cases. The *CSRP3* gene is a key example here. This gene, which encodes the Muscle LIM protein (MLP) Z-disk protein, was first reported as a causal HCM gene in 2003[48] and is now almost universally included in HCM clinical gene panels.

While a number of putative pathogenic variants have been reported in this gene[164], few have been robustly implicated in disease; however, there are some notable exceptions. The pathogenicity of the p.Cys58Gly variant, which was identified in a pedigree showing genome wide linkage to a region encompassing the *CSRP3* gene, is supported by strong segregation data (LOD >3) and *in vitro* functional evidence [165].

There is also evidence supporting the pathogenicity of the p.Lys44Pro variant, which has been reported in several HCM cases[48,166–168] and appears to be significantly enriched in HCM cases compared to ExAC in this study (3/2167 HCM cases, 0/60706 ExAC samples, $p<1x10^{-4}$). However, there remains some uncertainty surrounding this variant, as segregation analyses are limited and in some pedigrees, putative pathogenic variants in other sarcomere gene variants were also detected.[167,168]

Therefore, there is relatively convincing evidence to suggest that specific variants in *CSRP3* are disease-causing; however, the results from this study suggest that these are in the minority, implying that rare variants in this gene should be interpreted with caution.

2.4.2.1.2 ACTN2

Similarly, much of the evidence supporting a role for *ACTN2*, which encodes the Z-disk protein Alpha Actinin 2, relates to two missense variants, reported in three case studies.

[49,169,170] The p.Ala119Thr variant has been detected on the same haplotype in two pedigrees, one identified through genome wide linkage studies[49] and the other by exome sequencing[169]. In these pedigrees, this variant showed segregation (LOD 2.82) across a spectrum of cardiac phenotypes, including both structural and arrhythmogenic features. In another pedigree, the p.Met288Thr variant, was reported to segregate with a similarly complex and atypical spectrum of phenotypes (LOD=4.2)[170]. As yet, there appears to be very little evidence supporting the pathogenicity of other rare variants in this gene [171] and there was no evidence of rare variant enrichment in the HCM or DCM cohorts in this study case cohort. However, the previously reported variants which show strong evidence of pathogenicity appear to cause an atypical phenotype, which may not be represented in these clinical cohorts; this may explain the lack of signal in this study.

Therefore, as illustrated in the examples above, lack of evidence showing that the proportion of cases with a rare variant is higher than in controls, does not necessarily exclude the possibility that a gene is disease-causing. However, it does suggest that the majority of rare variants in these genes are not penetrant pathogenic mutations. This information is important for clinical laboratories, as in the absence of additional evidence allowing discrimination between benign and pathogenic variation (i.e. segregation or functional data), rare variants in these genes are largely uninterpretable and likely to yield a high number of uncertain results.

2.4.2.1.3 MYBPC3 variants in DCM

Critically, for some genes, when the results of these analyses are considered in the context of previously published data, there appears to be very little evidence supporting their involvement in cardiomyopathy. One key example is the *MYBPC3* gene, which has been implicated in DCM in a number of candidate gene studies[126,172,173]. Pathogenic

variants in *MYBCP3* are one of the most common causes of HCM[1,124,174], and the majority of pathogenic variants in this gene are predicted to lead to premature termination of translation. On the basis of evidence from *in vivo* and *in vitro* functional studies, haploinsufficiency has been proposed as the main disease mechanism in HCM [175–179]. However, functional evidence supporting a causal relationship between heterozygous truncating variants *MYBPC3* and DCM is lacking [176,180–182], and while a number of rare variants in *MYBPC3* have been implicated in DCM, the majority are missense changes, unsupported by segregation or functional evidence [173,183–187].

In the DCM cohort analysed in this study, which appears to be the largest case series reported to date (n=1161), no truncating variants were detected in *MYBPC3*. Additionally, there was no significant evidence that the proportion of DCM cases with a rare non-truncating variant was higher than in ExAC controls (OR:1.27 [95% CI 0.88-1.85], p=0.2002). This finding is consistent with our understanding of the pathobiology of rare variants in this gene[176,178,180–182], but challenges previous reports in the literature[95,172,173,183,186,188].

It now seems likely that previous studies, published before the availability of ExAC, had not adequately controlled for the relatively high background rate of rare missense variation in this gene. Indeed, many of the missense variants previously reported in individuals with DCM are at a frequency that seems incompatible with pathogenicity [95,172,173,183,186,188]. Consequently, this has led to erroneous conclusions about the role of this gene in DCM.

An argument could be made that some of the individuals with pathogenic variants in *MYBPC3* could develop end-stage HCM, with progression to a DCM phenotype. However, such cases appear to be rare[189] and arise in families with a predominant HCM

phenotype[187], so would be unlikely to account for the relatively high numbers of *MYBPC3*-related cases reported in primary DCM cohorts.

Taking into consideration the findings from this study, and re-evaluating the findings from previous studies in light of ExAC data, there seems to be very little evidence to support inclusion of the *MYBPC3* gene in DCM clinical test panels.

2.4.2.1.4 ANKRD1

The findings in this study also call into question the proposed link between *ANKRD1* and HCM. The *ANKRD1* gene, which encodes the cardiac ankyrin repeat protein (CARP), a Z-disc component, was first implicated in HCM in 2009 through candidate gene studies.[51] In that study, three variants (p.Pro52Ala, p.Thr123Met and p.Ile280Val) were detected in 384 unrelated HCM probands; these variants were not found in the 700 ethnically matched controls used in this study (400 Japanese and 300 U.S.). As none of the probands had a family history of HCM, segregation studies were not possible. While some *in vitro* data was presented to suggest that these variants might alter some aspects of protein function, the results of these analyses do not confirm variant pathogenicity.[51] However, all three variants have been detected in the ExAC cohort[4], and two of them are at a frequency which seems incompatible with pathogenicity (p.Pro52Ala and p.Thr123Met). The third variant, p.Ile280Val, has been classified by LMM as "Likely benign" due to lack of conservation among species. Therefore, it seems unlikely that these variants are in fact pathogenic.

[4] ***ANKRD1* variants in ExAC** (http://exac.broadinstitute.org/gene/ENSG00000148677)
p.Pro52Ala 6 of 8,256 South Asian (MAF 0.04%), 4 of 5,787 Latino (MAF 0.03%), 3 of 33,368 European (Non-Finnish) (MAF 0.004%)
p.Thr123Met 20 of 33,209 European (MAF 0.03%), 4 of 5114 African (MAF 0.04%) and 4 of 8239 South Asian (MAF 0.02%).
p.Ile280Val 4 of 33,220 European (Non-Finnish), (MAF 0.006%).

Inevitably, as this gene is on a number of clinical test panels, other missense variants have been reported in individuals with HCM and DCM, however due to lack of additional evidence, the overwhelming majority have been classified as variants of uncertain significance. With no convincing evidence in the literature, and considering the findings from this study of ~800 HCM cases, it seems likely that this gene has been erroneously implicated in this condition. Consequently, there appears to be no clinical utility in including this gene in HCM test panels.

As the above examples illustrate, some of the genes included in current clinical test panels may not in fact be disease-causing. In a clinical setting, testing these genes increases the potential for harm. At best, due to lack of evidence supporting pathogenicity, analyses of these genes will increase the proportion of uncertain results which are not clinically actionable. Of greater concern is the potential for mis-interpretation, leading to a false positive result. These could have serious implications for at-risk family members who may undergo genetic testing for the familial variant, leading to misdiagnosis and inappropriate therapy in those who test positive, or inappropriate discharge from follow-up in those who test negative.

An argument is often made that variants in candidate genes, even if not causal, could be contributing as modifiers[107,190,191]. While this remains possible, in the absence of any significant over-representation in cases, the results of this study do not support this hypothesis. Further analyses, including variants across a range of MAF's and larger numbers of cases would be needed to investigate possible modifier effects.

2.4.2.2 Variants classes and variant interpretation

These analyses also highlight the classes of variants that show the strongest disease association in each gene. This provides crucial insight into the pathobiology of each gene and cardiomyopathy, information which is essential for variant interpretation.

For example, in genes where heterozygous loss-of-function of protein function is disease-causing (e.g. *MYBPC3* in HCM[178], *LMNA* in DCM[75] and *PKP2* in ARVC[160,192]), truncating variants are associated with very high ORs, reflecting their high disease-related penetrance and consequently rarity in reference control populations. As such, the pathogenicity of novel truncating variants detected in these genes can be predicted with a high degree of certainty.

TTN truncating variants in DCM are an exception here. As illustrated in recent large cohort studies, this class of variant is significantly enriched in DCM cohorts compared to controls; however, the background "noise" is relatively high, so only a subset these variants appear to have disease-causing potential[67–69]. As such, at the gene level, the OR for *TTN* truncating variants is comparatively low (OR 19.77 95% CI [15.04-26.00], p<0.0001), indicating that a more considered approach to variant interpretation should be employed.

In the majority of genes encoding myofilament proteins (e.g. *MYH7*, *TNNT2, TNNI3, TPM1, ACTC1*), missense changes are the most common class of pathogenic variant[124,193,194]. In these genes, the mutant protein incorporates into the sarcomere and has a dominant-negative effect. This can have opposing effects on myofibrillar function and cause distinct phenotypes[1,195]. For example, variants causing HCM are typically activating, increasing myofilament sensitivity to calcium, leading to increased contractility and excessive energy use [41,196]. In contrast, variants causing DCM appear to

decrease the myofibrillar response to calcium, leading to diastolic dysfunction [85–87]. Consequently, in these genes, protein-truncating variants are not expected to be pathogenic and therefore divergent results are expected for truncating and non-truncating variants. This was evident in the *MYH7* gene in HCM, where an OR of 12 [10.9-13.3] was observed for non-truncating variants, but only 1 [0.5-4.5] for truncating variants, consistent with our understanding of the disease mechanisms.

In general, at the gene level, ORs for missense variants are typically modest (minimum OR ~2 for *PRKAG2* in HCM and maximum OR ~35 for *TPM1* in DCM, Table 2.3-5), which reflects the fact that this class of variant has the potential to elicit a wider range of effects on the protein, not all of which are pathogenic. As such, background levels of this class of variant are generally higher, making interpretation of novel missense variants more challenging.

In providing more reliable estimates of the signal to noise ratio in each gene, the results of this study could enhance missense variant interpretation and increase the yield of genetic testing in validated cardiomyopathy genes. For example, in genes where the OR for missense variants are relatively high, such as *TPM1* and *ACTC1,* there is a higher probability that a novel rare missense variant, detected in an affected individual, is pathogenic. This empirical data, in combination with existing and emerging computational approaches, which integrate evolutionary, biochemical, and functional information to predict the deleteriousness of protein altering variants[197], could allow better discrimination between benign and pathogenic variation.

2.4.2.3 Variant distribution

Additional analyses were undertaken to determine whether variant interpretation in validated genes could be further refined by comparing the linear distribution of

missense variants between cases and controls. Significant differences in variant distribution were identified in two genes, *MYH7* ($1.51x10^{-14}$) and *TNNI3* ($2.44x10^{-10}$), in HCM cases compared to ExAC. Additionally, there was strong statistical evidence indicating that rare missense variants in these genes clustered in the well-characterised, functionally important domains of the proteins [198–202].

Although in a number of genes, low variants counts precluded these analyses, the results in *MYH7* and *TNNI3* show the potential utility of these methods and indicate that future analyses, incorporating data from larger cases series, could highlight functionally important regions in less well-characterised genes and proteins. Furthermore, where data relating to tertiary protein structure is available, this could be integrated into these analyses, increasing power by enabling non-linear amino acid clusters to be identified.[203] Similarly, newly developed methods, leveraging ExAC data to identify missense-constrained regions within genes, will have potential utility here.[204]

2.4.2.4 Study limitations

The cardiomyopathy case cohorts used in these analyses are representative of those encountered by clinical diagnostic laboratories, rather than a clinically selected subset enriched for familial disease[174]. However, this cohort was considered appropriate for these analyses, which aimed to explore the contribution of rare variants in each gene in individuals referred for clinical genetic testing, and to assess variant interpretability in this setting. Additionally, while the ExAC cohort does not necessarily represent a healthy control population, the component cohorts are not expected to be enriched for inherited cardiomyopathies.[138] Therefore, this cohort was the largest available reference control population based on whole exome sequencing for these analyses.

Another key point is that variation in the analytical sensitivity of the sequencing methods used to generate the clinical case and ExAC data could impact on the accuracy of these results. Sensitivity data is not available for very rare variants in ExAC[138], however, exome sequencing data might be less sensitive at detecting variants than the targeted clinical gene panel tests[205–207]. To adjust for this, genes which had low sequencing coverage were not analysed (e.g. *RBM20* and *CTF1*). Also, to minimize the inclusion of false positive variant calls, only high confidence variant calls were included. As such, the estimates of rare variant frequency reported in this study are considered to be broadly accurate and sufficiently robust to support the stated conclusions.

One limitation of these analyses was that ethnicity data was not available for the OMGL cohort, or from LMM HCM patients; therefore, it was not possible to confirm the extent to which the cohorts used in this study were matched by race. However, as these analyses focused on the aggregate burden of multiple very rare variants, potential confounding effects due to individual population-specific variants in cases or ExAC are expected to be limited. Notably, there was no difference in the findings of case control analyses when comparing the LMM DCM cohort (for which self-reported ethnicity data was available) to all ExAC populations and to the European (Non-Finnish) subset, suggesting that population stratification effects were not a significant issue in this study.

It is also acknowledged that the MAF threshold applied in these analyses was selected on the basis of data from cardiomyopathy cohorts from a broadly Northern European referral population. Therefore, it is possible that these analyses could have excluded more common pathogenic founder variants in specific populations, where the genetic architecture of cardiomyopathy is not well defined.

It is also important to consider that these analyses only focused on single nucleotide changes and small insertion and deletions within coding regions and splice sites of the selected genes. Other classes of genetic variation, such as larger structural variants, or non-coding variants were not investigated. Despite lack of supporting evidence in the literature[208–210], it is possible that this class of variant may be causal in some haploinsufficient genes (e.g. *MYBPC3* and *PKP2*); however, the OMGL and LMM laboratories do not routinely test for these variants, so comparative analyses were not possible.

As individual level data was not available from ExAC, it was not possible to assess the impact of co-inheritance of variants, which have previously been proposed to account for a subset of cardiomyopathy cases[28,30,32]. In addition, this study did not evaluate more common variants (ExAC MAF >$1x10^{-4}$), which could be mechanistically informative but are likely to have smaller effects.[211]

2.5 Conclusions

These analyses demonstrate that large-scale comparison of rare variation in cardiomyopathy disease genes using data from patient cohorts and a very large reference control population can highlight the genes, regions of genes, and/or classes of variants which can be reliably interpreted in a clinical setting.

For validated disease genes, there is clear potential to increase the yield of correctly interpreted, actionable variants. However, these analyses highlight that it is equally important to recognize that many of the genes implicated in these disorders, and also a significant proportion of variants, may not be disease causing.

The findings in this study enhance our understanding of the genetic architecture of these inherited cardiomyopathies. Crucially, they provide much needed genetic evidence which will inform ongoing gene curation efforts which aim to systematically and comprehensively assess the genes and variants implicated in inherited cardiomyopathies[212–214].

Importantly, these methods could be applied to other dominant Mendelian disease genes. As clinical genetic testing moves to large gene panels and whole exome and genome sequencing, studies such as these will be essential to improve our understanding of gene and variant pathogenicity and ensure we are able to deliver reliable genomic interpretation.

Chapter 3. Analysis of 60 proposed HCM genes using whole genome sequence data in patients without pathogenic variants in confirmed genes.

3.1 Introduction

Despite decades of research into the genetic basis of Hypertrophic cardiomyopathy (HCM) and the widespread adoption of genetic testing for this disorder, in up to 60% of individuals referred for testing, the genetic basis of their phenotype remains elusive[124,215]. Understanding the genetic architecture of HCM continues to be a priority in both research and clinical domains; as such, clinical test panels constantly expand to accommodate the increasing number of genes claimed to cause this condition[56,145,216]. However, for many of these newly reported genes, lack of robust evidence supporting a causal role in HCM increases uncertainty in test interpretation [27,56,215].

To minimise uncertainty and the potential for misdiagnoses, it is widely recognised that there is an urgent need to define explicitly the causal genes in this disorder and rectify previous erroneous findings. However, as shown in a recent initiative led by ClinGen [213], which sought to systematically review and curate the evidence for over 50 genes implicated in HCM, in the majority of cases the purported link to HCM remains equivocal due to lack of empirical data. This emphasises the need for concerted and ongoing efforts to generate additional evidence to enable definitive assessment of each purported gene-disease relationship.

As described in Chapter 2, analyses using data from large clinical cardiomyopathy case series and reference control cohorts enabled evaluation of the genes currently included

in clinical gene panels, and highlighted the genes and classes of variants that can be reliably interpreted in a clinical setting [215,217].

Here, using whole genome sequencing data from 230 sarcomere gene-negative HCM cases and 7,253 reference controls, analyses of a further 60 genes implicated in HCM were undertaken. Additionally, variant level analyses were performed to review the evidence for pathogenicity and classify each variant according to clinical guidelines[128] and case level analyses to evaluate the clinical utility of genetic testing in these selected genes.

The aim of these analyses was to explore the role of these genes in HCM, and to examine the outcomes of testing large numbers of un-validated genes in a clinical setting. The results presented here contribute much needed empirical evidence about gene-disease relationships in HCM and will inform future gene curation efforts.

3.2 Methods

3.2.1 Study participants

Patients with HCM were recruited via the Oxford Medical Genetics Inherited Cardiac Conditions Clinic from the period May 2014 to September 2016. Eligibility criteria were as follows:

- Clinical diagnosis of HCM made in a specialist Inherited Heart Disease Clinic within Oxford University Hospitals NHS Foundation Trust, Royal Brompton & Harefield NHS Foundation Trust, Guy's and St Thomas' NHS Foundation Trust, or The Newcastle upon Tyne Hospitals NHS Foundation Trust.
- Age 18 -70 years, or >70 with a family history of HCM.

- Absence of “highly likely” or “likely” pathogenic variants within well-established HCM genes following clinical genetic testing.

For the majority of cases (referrals from Oxford, Guy’s and St Thomas’s and Newcastle), informed consent was given for genomic sequencing and analysis of demographic, clinical and family history data through the NIHR Bioresource for Rare Disease study (BRRD) (REC ref 13/EE/0325). Samples received from the Royal Brompton & Harefield NHS Foundation Trust were included as anonymous ‘legacy’ samples.

3.2.2 Clinical genetic testing

Prior to recruitment, genetic testing of a minimum of 13 HCM genes was undertaken by the Oxford Medical Genetics Laboratory (OMGL), a UKAS-accredited clinical diagnostic laboratory, or the Royal Brompton & Harefield NHS Foundation Trust. This included testing of eight well-established HCM genes (*MYBPC3*, *MYH7*, *TNNI3*, *TNNT2*, *MYL2*, *MYL3*, *ACTC1*, *TPM1*), two genes for common differential diagnoses (*PRKAG2*, *GLA*), and other more rarely associated, but validated, HCM genes (*CSRP3*, *FHL1*, *PLN*)[27,215,217]. Analysis was undertaken using a custom designed Haloplex Target Enrichment System (Agilent Technologies, Santa Clara, California), followed by high throughput DNA sequencing on an Illumina MiSeq (Illumina Inc, Great Chesterford, UK). Variants were classified according to clinical guidelines[218]. Variants of interest (variants classified either as “highly likely” to be pathogenic, “likely” to be pathogenic, or VUS) were confirmed by Sanger sequencing.

3.2.3 Reference controls

Rare variant data from 7,253 unrelated individuals recruited to other rare disease projects within the BRRD and sequenced as part of the same whole genome sequencing

project, were used as a reference control cohort (Table 3.2-1). Although formal exclusion of HCM was not performed within this control cohort, the prevalence of HCM within it is not expected to exceed that of the general population.

Code	Cohort	Number
A	Bleeding and Platelet Disorders	851
B	Ehler Danlos and ED-like Syndromes	13
C	Genomics England RD Pilot	1284
E	Pulmonary Arterial Hypertension	942
F	Primary Immune Disorders	957
G	Specialist Pathology: Evaluating exomes in Diagnostics (RD and ID)	1063
H	Thrombogenomics (BPD)	49
K	BPD with Autism Spectrum Disorder	49
L	Stem Cell and Myeloid Disorders	144
M	Primary Membranoproliferative Glomerulonephritis	152
N	BPD Co-segregation	40
P	Controls for WGS protocol testing	5
R	Multiple Primary Malignant Tumours	498
S	Steroid Resistant Nephrotic Syndrome	231
T	Intrahepatic Cholestsis of Pregnancy	260
V	Neuropathic Pain Disorders	165
W	Mixed (non HCM)	362
Y	Cerebral Small Vessel Disease	134
Z	Leber Hereditary Optic Neuropathy	54
	Total	7253

Table 3.2-1 Bioresource Rare Disease study-component cohorts.

3.2.4 Power calculations

Power calculations are shown in Table 3.2-2. Pre-study power calculations were based on analysis of 60 genes in 300 HCM cases, across a range of estimates relating to the proportion of controls with a rare variant in a gene selected for analysis (0.01, 0.004 and 0.001), and a 5 or 10 fold increase in the proportion of cases with a rare variant compared to controls. Power calculations were subsequently revised as fewer samples (n=230) were available for analysis.

Case n	Control n	Significance level (α)	Proportion of controls with rare variant	Power 2-fold	Power 5-fold	Power 10-fold
230	7253	0.0008	0.01	3%	85%	100%
230	7253	0.0008	0.004*	1%	32%	93%
230	7253	0.0008	0.001	0%	2%	13%

Table 3.2-2 Power calculations. Significance level= 0.0008 (Bonferroni correction for 60 genes (p= 0.05/60). Proportion of controls carrying a rare variant (rare defined as ExAC MAF <0.0001). The 0.004 figure is based on mean value from previous analysis of HCM genes (Chapter 2). Power= Power to reject the null hypothesis based on case and control numbers, p-value ≤ 0.0008, considering the proportion of cases with rare variants is either 5-fold or 10-fold greater than in controls. Power calculations were calculated using Stata scripts and run in Stata V10.0.

3.2.5 Whole genome sequencing

Whole genome sequencing was undertaken by Illumina on behalf of the NIHR Bioresource Rare Disease (BRRD) project. Genomic sequencing data was available from 7253 unrelated individuals recruited to other rare disease projects within the BRRD.

DNA samples were processed in batches of 96 prepared to a minimum concentration of 30ng/ul and volume 110ul. DNA fragmentation was undertaken using Covaris LE220 (Covaris Inc., Woburn, MA, USA) to obtain an average size of 450 base pair (bp) DNA fragments. Samples were processed using the Illumina TruSeq DNA PCR-Free Sample Preparation kit (Illumina Inc., San Diego, CA, USA). Seqeuncing libraries were checked using the Roche LightCycler 480 II (Roche Diagnostics Corporation, Indianapolis, IN, USA) with KAPA Library Quantification Kit (Kapa Biosystems, Inc, Wilmington, MA, USA). Sequencing was undertaken using either the Illumina HiSeq 2500 or HiSeqX instrument. Over the course of the project 3 read lengths were used: 100bp (377 samples), 125bp (1138 samples) and 150bp (6016 samples). Reads were aligned to the genome (GRCh37) using Isaac aligner (version 01.14) (Illumina Inc, Great Chesterford,

UK). Single nucleotide variants (SNVs) and indels were identified using Isaac variant caller (version 2.0.17).

3.2.6 Ethnicity

Data on ethnicity was generated by the BRRD bioinformatics core team. Estimates were calculated using Principal Component Analysis (PCA) using data from 34,535 SNPs with the 1000 Genomes study [219] used as a reference set to define populations. The admixed American population was excluded from analysis. Labelling was based on four populations only (European, African, South Asian and East Asian). PC-AIR from GENESIS package [220] was implemented for PCA analysis. PCA analysis was run on the 1000G reference set. Then, using the PC values from 1000G, WGS samples were projected (predicted) on the same space. Once the PC values were obtained, each sample was assigned to a cluster, using likelihood estimation from multi-dimensional normal distribution. This approach uses the first 5 principal components from the PCA analysis.

3.2.7 Gene selection

A range of resources were used to select genes for these analyses (Table 3.2-3). As the size of the case cohort was smaller than anticipated, to increase power, analyses were focused on a small selected set of genes with the aim of selecting those which were either already included in prominent commercial HCM clinical test panels, or under review by clinical gene curation initiatives (ClinGen HCM gene curation expert working group and Genomics England Panel App).

A total of 60 genes were selected (Table 3.2-4). A brief summary of the function of the protein encoded by each gene is provided in Appendix I. Review of current clinical

genetic test registries found that 67% of the selected genes are in current HCM clinical test panels and 78% are in familial cardiomyopathy clinical test panels (Figure 3.2-1).

Resource	Link	Date	Comments
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/omim	Oct 2016	Search term: Hypertrophic cardiomyopathy
GeneReviews	http://www.ncbi.nlm.nih.gov/	Oct 2016	Search term: Hypertrophic cardiomyopathy
OrphaNet	http://www.orpha.net	Oct 2016	Search term: Hypertrophic cardiomyopathy
PubMed	http://www.ncbi.nlm.nih.gov/pubmed	Oct 2016	Search terms Hypertrophic cardiomyopathy Hypertrophic cardiomyopathy & genetics
LMM	http://personalizedmedicine.partners.org/Laboratory-For-Molecular-Medicine/Tests/Default.aspx	Nov 2016	Hypertrophic cardiomyopathy gene panel
GeneDx HCM and Combined Cardiac panel	https://www.genedx.com/test-catalog/medical-specialty/cardiology/	Nov 2016	In combined panel selected genes only where they had stated that there had has been a publication suggesting link with HCM. (25 genes)
Illumina ICC panel	Pau et al (2016) J of Cardiovascular Translational Research 9(1):3-11.	Oct 2016	47 HCM genes
Genomics England Panel App	https://www.genomicsengland.co.uk/about-genomics-england/panelapp/	Oct 2016	Search term: Hypertrophic cardiomyopathy
ClinGen HCM gene curation	https://www.clinicalgenome.org/working-groups/clinical-domain/sub-groups/cardiovascular/hypertrophic-cardiomyopathy-gene-curation/	June 2016	Member of working group
String:-Protein network analysis	https://string-db.org/	Oct 2016	ClinGen list used to seed String-max 5 interactions
PathCards	https://pathcards.genecards.org/card/hypertrophic_cardiomyopathy_(hcm)	Oct 2016	Search term: Hypertrophic cardiomyopathy (118 genes returned)
GeneCards		Oct 2016	Search term: Hypertrophic cardiomyopathy (1,194 genes returned)
Malacards	http://www.malacards.org/	Oct 2016	Search term: Hypertrophic cardiomyopathy (51 genes returned)
GeneTests	https://www.genetests.org	Oct 2016	Search terms: Hypertrophic Cardiomyopathy Test Type: Multi-Gene panelTest

Resource	Link	Date	Comments
			Method:Sequencing, NGS All Locations: US labs and European Labs. NB. This website was retired on Dec 1 2017. This is now replaced by the Genetic Test Registry (see below).
Genetic Test Registry	https://www.ncbi.nlm.nih.gov/gtr/	Dec 2017	Reviewed to update status of genes selected for analysis in December 2016.

Table 3.2-3 Range of resources used to review and collate genes for case control analyses

		Selection criteria												
	Gene	**OMGL HCM**	**LMM HCM**	**GeneDx HCM**	**Gene DX Pan cardiac panel.**	**Illumina Cardio**	**OMIM-HCM**	**Orphanet-HCM**	**Orphanet syndromic**	**Genomics England**	**Protein network analysis String- 5**	**ClinGen review list**	**HGNC**	**OMIM**
1	*ABCC9*											x	60	601439
2	*ACADVL*									x			92	609575
3	*ACTA1*					x				x		x	129	102610
4	*AGL*									x			321	610860
5	*ALPK3*										x	x	17574	n/a
6	*ATP5E*									x			838	606153
7	*BAG3*											x	939	603883
8	*BRAF*									x			1097	164757
9	*CACNA1C*				x	x				x			1390	114205
10	*CALR3*					x	x			x		x	20407	611414
11	*CASQ2*				x	x				x		x	1513	114251
12	*CAV3*			x	x	x	x			x		x	1529	601253
13	*COA5*									x			33848	613920
14	*COX15*					x					x		2263	603646
15	*CRYAB*					x				x			2389	123590
16	*DES*				x					x		x	2770	125660
17	*DSP*										x	x	3052	125647

		Selection criteria												
	Gene	**OMGL HCM**	**LMM HCM**	**GeneDx HCM**	**Gene DX Pan cardiac panel.**	**Illumina Cardio**	**OMIM-HCM**	**Orphanet-HCM**	**Orphanet syndromic**	**Genomics England**	**Protein network analysis String- 5**	**ClinGen review list**	**HGNC**	**OMIM**
18	*FLNC*	x					x			x		x	3756	102565
19	*FOXRED1*									x			26927	613622
20	*FXN*					x				x			3951	606829
21	*GAA*					x				x		x	4065	606800
22	*GLB1*									x			4298	611458
23	*GUSB*									x			4696	611499
24	*HRAS*									x			5173	190020
25	*JPH2*			x		x	x			x		x	14202	605267
26	*KCNQ1*					x				x		x	6294	607542
27	*KLF10*					x				x			11810	601878
28	*LDB3*					x	x			x		x	15710	605906
29	*LMNA*					x				x			6636	150330
30	*MAP2K1*									x			6840	176872
31	*MAP2K2*				x					x			6842	601263
32	*MRPL3*									x			10379	607118
33	*MYH6*					x	x	x		x		x	7576	160710
34	*MYLK2*				x	x	x			x		x	16243	606566

		Selection criteria												
	Gene	**OMGL HCM**	**LMM HCM**	**GeneDx HCM**	**Gene DX Pan cardiac panel.**	**Illumina Cardio**	**OMIM-HCM**	**Orphanet-HCM**	**Orphanet syndromic**	**Genomics England**	**Protein network analysis String- 5**	**ClinGen review list**	**HGNC**	**OMIM**
35	*MYO6*					x			x	x			7605	600970
36	*MYOM1*					x				x		x	7613	603508
37	*MYOZ2*		x		x	x	x	x				x	1330	605602
38	*MYPN*				x	x	x			x		x	23246	608517
39	*NEXN*				x	x	x	x		x		x	29557	613121
40	*NRAS*									x			7989	164790
41	*OBSCN*										x	x	15719	608616
42	*PDLIM3*				x	x				x		x	20767	605889
43	*PTPN11*		x	x	x				x	x		x	9644	176876
44	*RAF1*		x		x				x	x		x	9829	164760
45	*RYR2*			x	x							x	10484	180902
46	*SCO2*									x			10604	604272
47	*SHOC2*									x			15454	602775
48	*SLC25A3*									x			10989	600370
49	*SLC25A4*					x				x			10990	103220
50	*SOS1*				x				x	x			11187	182530
51	*TCAP*			x		x	x	x		x		x	11610	604488

		Selection criteria												
	Gene	**OMGL HCM**	**LMM HCM**	**GeneDx HCM**	**Gene DX Pan cardiac panel.**	**Illumina Cardio**	**OMIM-HCM**	**Orphanet-HCM**	**Orphanet syndromic**	**Genomics England**	**Protein network analysis String- 5**	**ClinGen review list**	**HGNC**	**OMIM**
52	*TMEM70*									x			26050	612418
53	*TMPO*										x	x	11875	188380
54	*TNNC1*	x	x	x		x	x	x		x		x	11943	191040
55	*TNNC2*										x	x	11944	191039
56	*TRIM63*					x				x		x	16007	606131
57	*TSFM*									x			12367	604723
58	*TTN*						x			x		x	12403	188840
59	*TTR*	x		x		x				x		x	12405	176300
60	*VCL*			x		x	x			x		x	12665	193065

Table 3.2-4 Genes selected for case-control analyses and source of information. HGNC=HUGO Gene Nomenclature Committee ID (https://www.genenames.org/). OMIM=Online Mendelian Inheritance in Man ID (https://www.omim.org/).

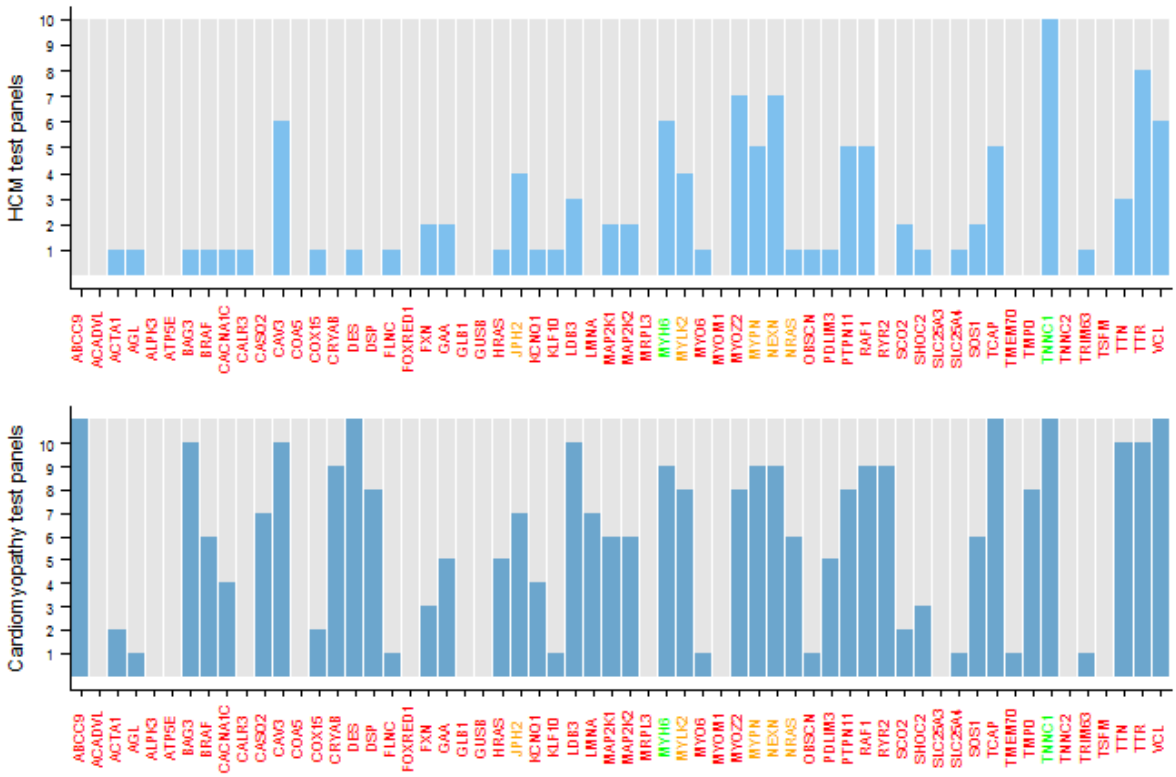


Figure 3.2-1 Frequency of inclusion of the 60 selected genes in current commercial test panels. Data from NCBI Gene Tests website (October 2016) and Genetic Test Registry (December 2017). Clinical HCM panels (n=10) and Clinical Cardiomyopathy panels (n=11). y-axis shows number of panels which included each gene. The gene labels are coloured according to the review status as annotated by Genomics England which gives an indication of the level of evidence supporting each gene:disease relationship: Green = high evidence, the gene is very likely be the cause of the disease and can be reported back to patients. Amber = moderate evidence, and should not yet be used for genome interpretation. Red = low evidence for a role in disease, or not suitable for clinical diagnosis at this time.

3.2.8 Coverage data

The mean depth of sequence coverage across coding regions in each gene was assessed to determine if there was sufficient sequencing data coverage to enable accurate variant calling. Using data from Ensembl, a BED file was created listing the chromosome number and start and end points of the canonical transcript for all genes. A second BED file was created that listed the chromosome number and start and end points for all the coding regions within the canonical transcript of each gene.

A Python script was written to extract the depth of sequence coverage across coding nucleotides in each gene across all the HCM sample BAM files. This coverage data was combined into a single file per gene.

An R script was created to import the coverage data for each gene, calculate the mean depth of coverage at each coding nucleotide across all 230 HCM samples, and generate a box plot to display data for each gene. This script also calculated coverage "bins" for each gene (e.g. <10x, 10-20x, 20-30x, >30x), to determine the percentage of coding bases covered across each range of read depths. Another R script was created to generate a plot showing the mean coverage, with 95% confidence intervals for every nucleotide across the entire transcript of a gene, annotated with coding regions. This plot provides a visual display of the coverage across a specified gene.

3.2.9 Analysis protocol

Chromosome co-ordinates for each gene and transcript were obtained from Ensembl Genome build GRCh37 and combined into a single BED file.

A range of open source bioinformatics tools were combined into a Python script to extract variants within regions of interest (R.O.I) from the BRRD project merged VCF files. This included the following key steps: (i)Tabix (http://www.htslib.org/doc/tabix.html) was used to extract variants from chromosome regions as specified in the BED file; (ii) variants within R.O.I were annotated using SNPEff (http://snpeff.sourceforge.net/); (iii) variant filtering was undertaken using SNPSift using the following criteria PASS filter, ExAC MAF <0.0001, ENSG ID, ENST ID and variants annotated by SNPEff as HIGH or MODERATE impact. Variants annotated by the SNPEff tool as MODERATE includes missense variants, in-frame insertions or

deletions, and other variants not predicted to cause protein truncation. SNPSift was also used to generate rare variant counts for cases and controls.

For some variants, discrepancies in MAF were noted between the BRRD controls and ExAC cohorts; this prompted inclusion of an additional variant filtering step. The Cardiodb Allele Frequency app was used to determine an appropriate conservative allele count threshold for filtering in the BRRD reference cohort (https://cardiodb.org/allelefrequencyapp/). Based on a disease prevalence of 1/500[10], a maximum pathogenic allele frequency of 3% (the upper 95% CI of the binomial distribution of *MYBPC3* Arg502Trp allele frequency in European HCM cohorts, LMM, unpublished data), penetrance of 30%, and a population/reference cohort size of 7253, a maximum credible population AF of 0.0001 was assigned. Appling this MAF threshold, the maximum tolerated allele count for pathogenic variants in the BRRD cohort was 4; as such, variants with an allele count ≥4 were excluded from analysis.

Student T test was used to compare age at genetic testing between cases with and without a family history of HCM. Comparison was also made for age at genetic testing and gender. Plots were generated to view the distribution of gender and family history by age. Analyses and plot generation were undertaken using R version 3.3.3 and R Studio version 1.0.136.

Fisher's exact test and odds ratios were calculated for each gene by comparing the proportion of cases and controls with a rare variant (defined for the purposes of these analyses as variants with a MAF ≤0.0001 in ExAC [138] and an allele count ≤4 in BRRD reference controls) within unrelated cases (n=230) and reference controls (n=7352). Analyses were undertaken combining rare variants across all genes and for each gene individually. Separate analyses were also undertaken focusing on missense, truncating

(frameshift, nonsense, splice donor/acceptors) and non-truncating (missense, in-frame insertions and deletions) variants only. As a control for technical factors, the prevalence of synonymous variants in each gene, the majority of which are not expected to be disease-causing, was also assessed.

Analysis of the 16 genes in the current Oxford Medical Genetics Laboratory (OMGL) diagnostic gene panel was also undertaken to confirm previous findings. Analyses were undertaken using R version 3.3.3 and R Studio version 1.0.136.

The mean proportion of individuals with a rare variant, with 95% CI calculated under the binomial distribution, were plotted for each gene in cases and reference controls. Analyses were also undertaken to explore the relationship between the number of rare variants detected, the number of coding nucleotides in gene, and estimates of ExAC missense constraint scores. Analyses and plot generation was undertaken using R version 3.3.3 and R Studio version 1.0.136.

3.2.10 Homology

Analysis was also undertaken to determine whether any of the selected genes had regions of high sequence homology to other genomic regions (e.g. within the same gene, a functionally related gene, or pseudogene). Data published by Mandelker *et al* [221], which lists regions of high homology across 4773 medically relevant genes, was used to identify potentially problematic regions.

In regions considered to be a high risk of misalignment, case-control analyses were repeated excluding variants in these regions to ensure false positive variant annotations did not impact on results conclusions.

3.2.11 Comparing variant distribution

Kernel Density Plots were generated to visualise the distribution of rare missense variants across amino-acid positions in each gene in cases and controls. Statistical comparison of variant distribution was undertaken using the Kolmogorov-Smirnov Test (K-ST). Genes with fewer than five missense variants were excluded from analysis. Analysis and plot generation was performed in R version 3.2.0.

3.2.12 Variant classification

Classification of individual rare variants detected in HCM cases was undertaken using American College of Medical Genetics and Genomics (ACMG) guidelines [128], adapted to incorporate disease specific knowledge relating to estimates of disease prevalence[10], penetrance, and rare variant frequency data from previous analyses [27,215,222].

3.3 Results

3.3.1 Characteristics of the cohort

These analyses included 230 unrelated HCM cases in which no putative pathogenic variants were detected following clinical genetic testing of confirmed genes. The mean age at specialist clinic evaluation was 53 years [SD 12 yrs, range 20-75yrs]. 75% of HCM cases were male, compared to 45% of BRRD reference controls (Figure 3.3-1). 32% of cases reported a family history of HCM. There was no difference in the mean age at genetic testing when comparing cases with and without a family history (p=0.14). The age distribution of HCM cases with and without a reported family history of HCM is shown in Figure 3.3-3 and distribution by gender in Figure 3.3-4. Comparison of age at genetic testing between individuals with and without a family history of HCM and by gender is shown in Figure 3.3-2.

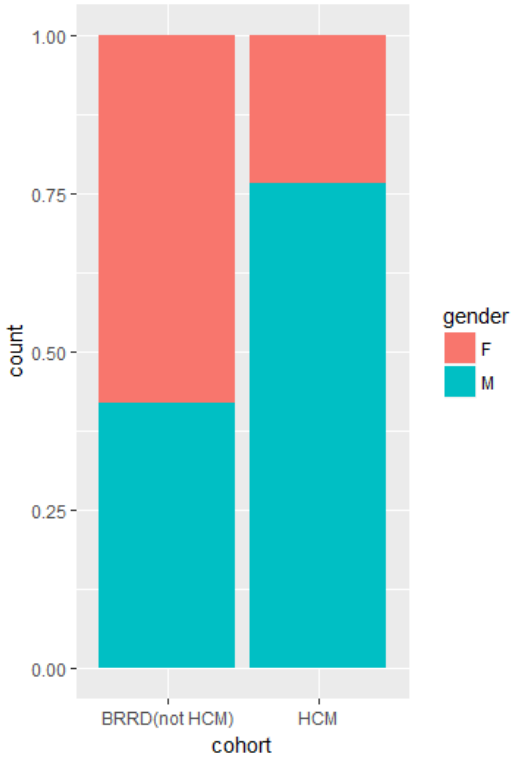


Figure 3.3-1 Gender BRRD (non-HCM) compared to HCM cohort. F=female, M=male.

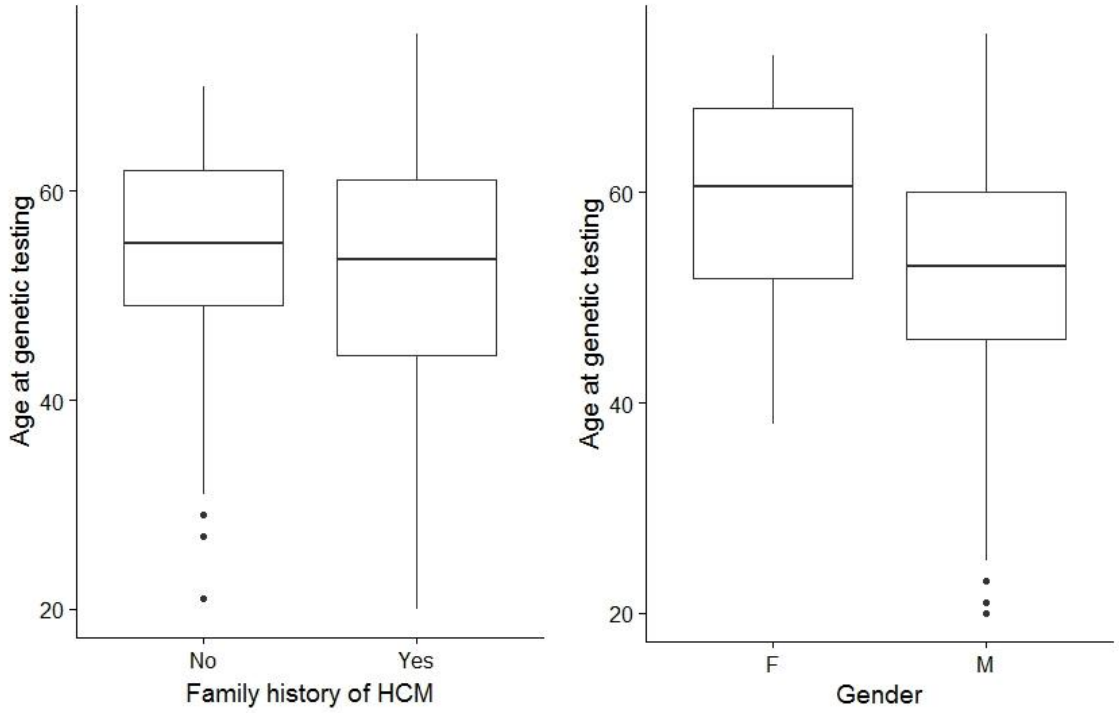


Figure 3.3-2 Boxplots comparing age at genetic testing (in years) in individuals with and without a reported family history of HCM and by gender F=female, M=male. (n=227, d.o.b not available for 3 individuals).

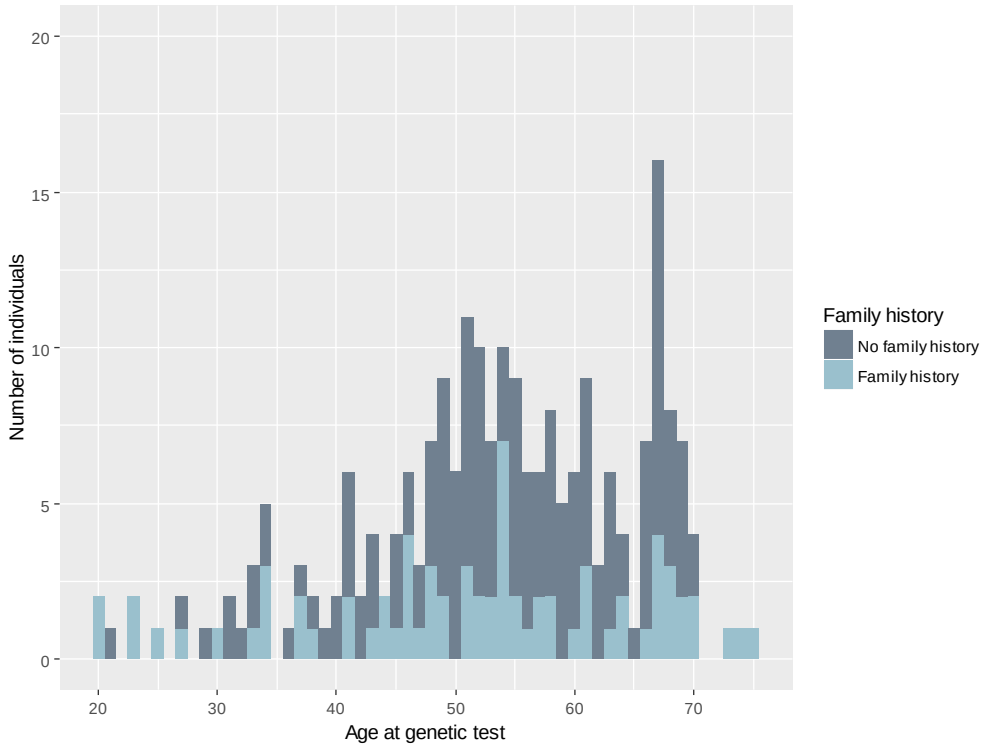


Figure 3.3-3 Histogram showing distribution of age (in years) of HCM cases by family history (n=227, date of birth not available for 3 individuals).

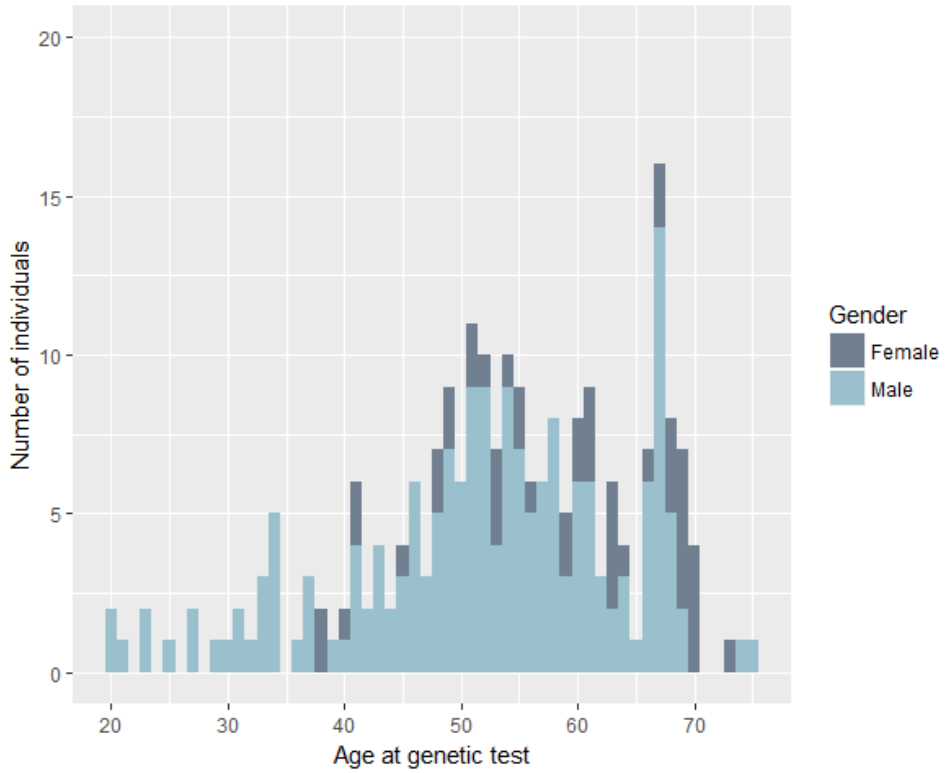


Figure 3.3-4 Histogram showing distribution of age (in years) of HCM cases by gender (n=227, date of birth not available for 3 individuals).

3.3.2 Ethnicity

The ethnicity of case and control cohorts, as predicted by PCA analysis, was broadly similar, with the majority (89.7% and 80% respectively) categorised as European (Table 3.3-1).

Cohort	African	East-Asian	European	Finnish	Other	South-Asian
HCM	2.5%	1.6%	89.7%	0%	3.7%	2.5%
Non-HCM	2.3%	0.8%	80%	0.2%	9.4%	7.2%

Table 3.3-1 . Ethnicity of HCM (n=243) and Non-HCM (n=8868) BRRD cohorts (including related individuals)

3.3.3 Gene coverage

Figure 3.3-5 shows the median coverage depth of coverage for each gene across 230 HCM samples. With the exception of one gene (*MYH6)*, the mean depth of sequence read coverage across coding nucleotides in HCM cases was at least 20x.

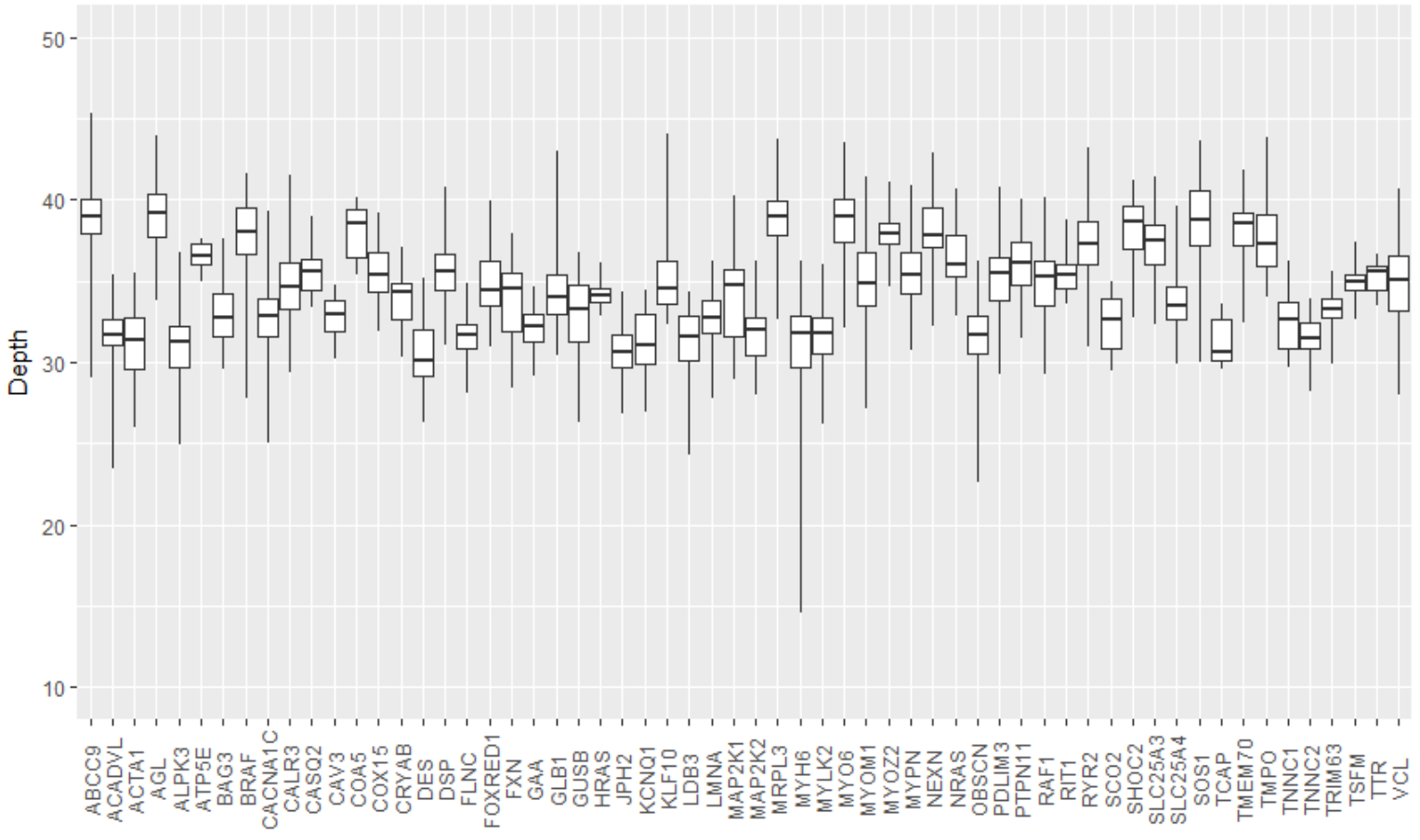


Figure 3.3-5 Boxplot showing coverage depth across 230 HCM samples for each gene. The horizontal line inside each rectangle indicates the median value. The rectangle represents the second and third quartiles. The lower and upper quartiles are shown as vertical lines on either side of the rectangle.

3.3.4 Regions of homology

Regions of high sequence homology were identified in 7 of the 60 genes selected for analysis (Table 3.3-2).

Gene	Region of homology	Number of exons affected		
		Dead Zone	High stringency	Low stringency
CACNA1C	Non-CDS	0	0	2
FLNC	Pseudogene	0	1	3
GUSB	Pseudogene	0	0	1
MAP2K1	Pseudogene	0	0	1
MAP2K2	Pseudogene	0	0	1
MYH6	Different gene (*MYH7*) *and* Pseudogene	0	1	5
TTN	Same gene	7	0	29

Table 3.3-2 Regions of homology in the 60 genes selected for these analyses which might impact on NGS read mapping.
Dead Zone= Short reads cannot be unambiguously mapped to any position in an exon or to a large stretch within a larger exon; therefore, there is a high probability of reads mapping to an identical sequence elsewhere in the genome.
High Stringency= A large fraction of the reads obtained for these regions likely to be at high risk of misalignment.
Low stringency= Contains at least some potentially difficult or unable to be analysed positions; therefore, there is a risk of read misalignment due to high homology at these sites.

3.3.4.1 TTN

Several exons within the *TTN* gene show high homology to other regions within the same gene, with at least seven exons in which short NGS reads cannot be aligned unambiguously.

3.3.4.2 FLNC

Four regions in the *FLNC* gene were highlighted as potentially problematic; exons 46-48 of *FLNC* show >98% sequence identity to the *FLNC* pseudogene (pseFLNC), which is located 53.6 kb downstream of the functional gene. Sequence coverage was relatively

uniform across the gene and no obvious issues were observed in these homologous regions (Figure 3.3-6).

A single "High Stringency" region was highlighted in this gene. This corresponds to *FLNC* exon 47, a 210bp coding region which shows 99.52% sequence identity to regions in pseFLNC. There is a single nucleotide difference between these regions, if pseFLNC sequence were misaligned to *FLNC*, this variant would be annotated as c.7877G>A p.(Ser2626Asn). This variant was observed in a total of 4 unrelated reference controls in the BRRD cohort, suggesting that misalignment of pseFLNC sequence reads to *FLNC* has occurred to some extent in this cohort.

In total 11 rare variants in 1 case and 18 reference controls were detected within homologous regions of the *FLNC* gene (Table 3.3-3). Four variants (detected in 1 case and 11 controls) were detected in Exon 47. Two variants (p.Arg2621Gln, p.Ser2626Asn) were excluded from case-control analyses due to their frequency in controls (AC ≥4). Removing the remaining variants from case-control analyses had no impact on the results conclusions (Table 3.3-4).

Chr7 g. (GRCh37)	AN	AC	Cases	Controls	c.	p.	FLNC Exon
128497175	18220	1	0,0,0	0,1,1	c.7565T>C	p.Val2522Ala	46
128497186	18220	2	0,0,0	0,1,1	c.7576T>A	p.Phe2526Ile	46
128497199	18220	1	0,0,0	0,1,1	c.7589C>A	p.Thr2530Asn	46
128497262	18220	1	0,0,0	0,1,1	c.7652A>G	p.Asp2551Gly	46
128497276	18220	1	0,0,0	0,1,1	c.7666C>G	p.Pro2556Ala	46
128497319	18220	2	0,0,0	0,1,1	c.7709A>G	p.Asn2570Ser	46
128497342	18220	1	0,0,0	0,1,1	c.7732G>A	p.Gly2578Ser	46
128498113	18220	2	0,0,0	0,1,1	c.7832T>G	p.Val2611Gly	47
128498143	18220	5	0,1,1	0,4,4	c.7862G>A	p.Arg2621Gln	47
128498158	18220	5	0,0,0	0,4,4	c.7877G>A	p.Ser2626Asn	47
128498190	18220	2	0,0,0	0,2,2	c.7909A>C	p.Lys2637Gln	47

Table 3.3-3 Variants detected in *FLNC* Exons 46-48 in BRRD cohort. Ensemble Gene ID:ENSG00000128591. Variants annotated against Ensembl transcript ID: ENST00000325888. In the columns Cases and Controls the numbers separated by commas refer to number of homozygous alleles, the number of heterozygous and total number of alleles respectively.

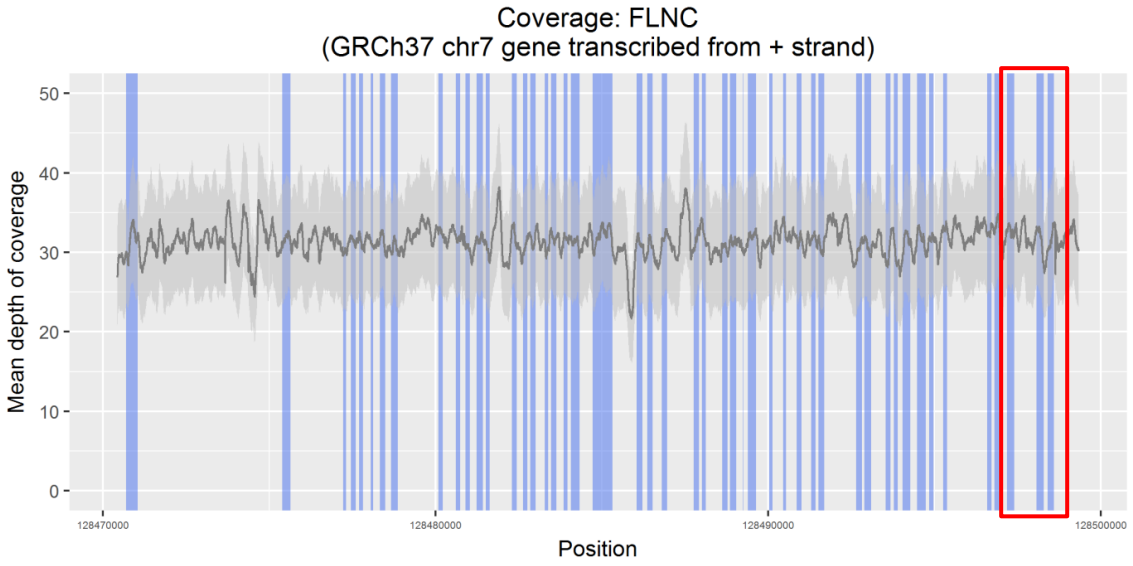


Figure 3.3-6 Mean per base depth of coverage across *FLNC* ENST00000325888. Dark grey= mean depth of coverage across 230 HCM samples. Light grey=standard deviation. Blue vertical bars=coding regions. Homologous regions (Exons46-48) marked by Red box.

Gene	Case with	Case without	Control with	Control Without	FET p value	OR	Lower_CI	Upper_CI
All (A)	14	216	268	6985	0.0752	1.689	0.896	2.947
All (B)	14	216	258	6995	0.0491	1.757	0.931	3.067
Missense (A)	12	218	263	6990	0.2096	1.462	0.735	2.652
Missense (B)	12	218	253	7000	0.1997	1.523	0.764	2.762

Table 3.3-4 Fisher's exact test results *FLNC* gene. Analysis of all rare variants and rare missense variants separately. (A) Analysis including rare variants detected in homologous regions (B) Analysis excluding rare variants detected in homologous regions.

3.3.4.3 MYH6

A total of 5/37 coding exons in *MYH6* showed strong sequence homology to another clinically relevant HCM gene (*MYH7*). One region (exon 26 in *MYH6* which corresponds to exon 27 in *MYH7*) was annotated as High Stringency. The nucleotide positions which differ between genes in these regions were noted (Table 3.3-5). The false positive variant annotations that would be detected if mis-mapping occurred within these

homologous regions were defined (Table 3.3-6). Several of these possible false positive variant annotations were detected in both the BRRD cohort and in the Genome Aggregation Database (gnomAD) (which includes exome and genome sequencing data from almost 150,000 individuals sequenced as part of various disease-specific and population genetic studies); a number had passed genotyping quality control filters (annotated as PASS filter) and therefore included in analyses (Table 3.3-6 and Table 3.3-7).

In the regions of *MYH6* which show homology to *MYH7*, 2 rare missense variants were detected in HCM cases and 27 missense and 2 truncating variants in controls (Table 3.3-7). Case-control analyses were undertaken with and without variants in regions of homology; in both analyses, no evidence was found to suggest there was a difference in rare variant frequency between cases and controls in this gene (Table 3.3-8).

Analysis of sequencing coverage detected a decrease in sequence read depth in *MYH6* exon 26, and a corresponding increase in coverage in the homologous exon 27 in *MYH7* (Figure 3.3-7 and Figure 3.3-8). No obvious discrepancies in sequence read coverage were noted in the other four homologous *MYH6/MYH7* coding regions.

MYH6			MYH7			No. nucleotide positions which differ /total no. nucleotides	% sequence identify
Exon	**Chr14(GRCh37)**	**Codon number**	**Exon**	**Chr14(GRCh37)**	**Codon number**		
14	23869635 to 23869465	471-527	15	23897879 to 23897707	470-526	2/171	98.83
19	23866026 to 23865903	723-764	20	23895027 to 23894904	721-762	3/124	97.58
26	23859655 to 23859266	1115-1244	27	23889442 to 23889054	1113-1242	0/390	100%
33	23855832 to 23855524	1551-1653	34	23885521 to 23885213	1507-1548	5/303	98.35
36	23853926 to 23853651	1764-1855	37	23884479 to 23884204	1720-1761	4/276	98.19

Table 3.3-5 Regions in *MYH6* and *MYH7* which show sequence homology. Codon numbers within homologous exons. Number of nucleotides that differ in homologous regions of these genes. Overall % of coding nucleotide identify within each homologous region.

MYH7 sequence mis-mapped to MYH6				MYH6 sequence mapped to MYH7			
Ex	**Variant**	**BBRD**	**gnomAD**	**Ex**	**Variant**	**BRRD**	**gnomAD**
14	c.1485C>T p.(Phe495Phe) c.1521T>C p.(Ile507Ile)	1/18848 2/18848	14-23869561-G-A (19/277218) 14-23869525-A-G (31/277196)	15	c.1482T>C p.(Phe494Phe) c.1518C>T p.(Ile506Ile)	ND ND	14-23897805-A-G (8/246270) 14-23897769-G-A (11/277216)
19	c.2189T>C p.(Val730Ala) c.2230A>G p.(Thr744Ala) c.2250T>C p.(Ser750Ser)	ND ND ND	14-23866006-A-G (4/276908) 14-23865965-T-C (3/245966) 14-23865945-A-G (11/245964)	20	c.2183C>Tp.(Ala728Val) c.2224G>A p.(Ala742Thr) c.2244C>T p.(Ser748Ser)	ND ND ND	14-23895007-G-A (40/276918) 14-23894966-C-T (1/245964) 14-23894946-G-A (1/30944)
26	n/a	-	-	27	n/a	-	-=
33	c.4695A>G p.(Leu1565Leu) c.4778A>T p.(Gln1593Leu) c.4812T>C p.(Asp1604Asp) c.4838T>C p.(Val1613Ala) c.4914T>C p.(Ala1638Ala)	1/18848 161/18848 2/18848 164/18848 9238/18848	14-23855788-T-C (9/276740) 14-23855705-T-A (3835/276204) 14-23855671-A-G (30/277172) 14-23855645-A-G (3770/276054) 14-23855695-A-G (127074/276690)	34	c.4689G>A p.(Leu1563Leu) c.4772T>A p.(Leu1591Gln) c.4806C>T p.(Asp1602Asp) c.4832C>T p.(Ala1611Val) c.4908C>T p.(Ala1636Ala)	1/198848 3/18848 (Not PASS filter OPR0.8) 6/18848 1/18848	14-23885477-C-T (2/245858-not PASS filter) 14-23885394-A-T (10/246256) 14-23885360-G-A (286/274438-Not PASS filter) 14-23885334-G-A (4/246252-didn't PASS filter. Found in Exomes only) 14-23885258-G-A (7/246212)
36	c.5367G>A p.(Glu1789Glu) c.5403G>A p.(Glu1801Glu) c.5427A>C p.(Gly1809Gly) c.5476_5477delinsAA p.(Gly1826Asn)	1/18848 1/18848 ND 8/18848 8/18848	14-23853849-C-T (14/277246) 14-23853813-C-T (2/246270) 14-23853789-T-G (8/277226) 14-23853739-C-T (47/277050) 14-23853740-C-T (47/277038) Reported as SNPs-review of BAMs confirms *in cis*	37	c.5361A>G p.(Glu1787Glu) c.5397A>G p.(Glu1799Glu) c.5421C>A p.(Gly1807Gly) c.5470_5471delinsGG p.(Asn1824Gly)	ND 1/18848 ND 1/18848 23884293 listed in VCF but AC 0.	14-23884402-T-C (218/277224) (208/34420 Latino) 14-23884366-T-C (17/277232) 14-23884342-G-T (2/246222) 14-23884292-T-C (5/277072 Of three BAM files available in all cases this variant was in cis with 23884295 14-23884293-T-C (3/277052) (two in cis with 23884292 and one single het)

Table 3.3-6 Variants which would arise if *MYH6* and MYH7 sequence reads were mis-mapped. Variants annotated against genome build GRCh37. *MYH6* Ensembl transcript ID:ENST00000356287, *MYH7* Ensembl transcript ID:ENST00000355349. Frequency of each variant (allele count/allele number) in BRRD cohort and Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/). ND=not detected.

Chr14 g. (GRCh37)	AN	AC	Cases	Controls	c.	p.	*MYH6* Exon
23853667	18220	1	0,0,0	0,1,1	c.5549A>G	p.Lys1850Arg	36
23853673	18220	2	0,0,0	0,1,1	c.5543G>A	p.Arg1848His	36
23853697	18220	2	0,0,0	0,1,1	c.5519A>G	p.Lys1840Arg	36
23853703	18220	1	0,0,0	0,1,1	c.5513C>T	p.Ser1838Leu	36
23853704	18220	1	0,0,0	0,1,1	c.5512T>G	p.Ser1838Ala	36
23853757	18220	1	0,0,0	0,1,1	c.5459G>A	p.Arg1820Gln	36
23853871	18220	1	0,0,0	0,1,1	c.5345A>G	p.Glu1782Gly	36
23853880	18220	1	0,0,0	0,1,1	c.5336C>A	p.Ala1779Asp	36
23853899	18220	1	0,0,0	0,1,1	c.5317A>G	p.Lys1773Glu	36
23853908	18220	1	0,0,0	0,1,1	c.5308G>A	p.Glu1770Lys	36
23853923	18220	7	0,0,0	0,6,6	c.5293G>A	p.Ala1765Thr	36
23855565	18220	1	0,0,0	0,1,1	c.4918G>A	p.Glu1640Lys	33
23855690	18220	1	0,1,1	0,0,0	c.4793C>T	p.Ser1598Leu	33
23855736	18220	1	0,0,0	0,1,1	c.4747G>A	p.Glu1583Lys	33
23855756	18220	1	0,0,0	0,1,1	c.4727G>A	p.Arg1576Gln	33
23855770	18220	1	0,0,0	0,0,0	c.4713G>T	p.Lys1571Asn	33
23855798	18220	1	0,1,1	0,0,0	c.4685G>T	p.Arg1562Leu	33
23855799	18220	2	0,0,0	0,2,2	c.4684C>T	p.Arg1562Trp	33
23855814	18220	1	0,0,0	0,1,1	c.4669G>A	p.Glu1557Lys	33
23855817	18220	1	0,0,0	0,0,0	c.4666G>A	p.Glu1556Lys	33
23855831	18220	1	0,0,0	0,1,1	c.4652C>A	p.Ala1551Asp	33
23859571	18176	2	0,0,0	0,2,2	c.3427C>T	p.Arg1143Trp	26
23859600	18220	1	0,0,0	0,1,1	c.3398A>G	p.Lys1133Arg	26
23859615	18220	1	0,0,0	0,1,1	c.3383G>A	p.Arg1128His	26
23859652	18220	3	0,0,0	0,2,2	c.3346C>A	p.Arg1116Ser	26
23859652	18220	1	0,0,0	0,1,1	c.3346C>T	p.Arg1116Cys	26
23865965	18220	1	0,0,0	0,1,1	c.2229delG	p.Thr744fs	19
23869470	18220	1	0,0,0	0,1,1	c.1576G>A	p.Glu526Lys	14
23869511	18220	2	0,0,0	0,1,1	c.1535T>C	p.Ile512Thr	14
23869545	18220	1	0,0,0	0,1,1	c.1501G>T	p.Glu501*	14
23869569	18220	1	0,0,0	0,1,1	c.1477C>T	p.His493Tyr	14

Table 3.3-7 Variants detected in BRRD cohort in *MYH6* regions showing homology to *MYH7* (Exons 14,19,26,22,36). g.= genome co-ordinates. AN=allele number. AC=allele count. Ensemble Gene ID:ENSG00000197616. Variants annotated against Ensembl transcript ID: ENST00000356287. In the columns "Cases" and "Controls" numbers refer to number of homozygous alleles, the number of heterozygous and total number of alleles.

MYH6	Case with	Case without	Control with	Control Without	FET p value	OR	Lower CI	Upper CI
All var (A)	5	225	168	7085	1	0.937	0.297	2.263
All var (B)	3	227	139	7114	0.8033	0.676	0.136	2.043
Missense (A)	5	225	149	7104	0.5121	0.937	0.297	2.263
Missense (B)	3	227	122	7131	1	0.772	0.155	2.341
Truncating (inc exon 26)	0	230	15	7238	1	-	-	8.833
Truncating (exc. exon 26)	0	230	13	7240	1	-	-	10.396
Non-truncating (inc. exon 26)	5	223	151	7102	0.8145	1.045	0.331	2.530
Non-truncating (exc. exon 26)	3	227	122	7131	1	0.772	0.155	2.341

Table 3.3-8 Fisher's exact test (FET). Results including (A) and excluding (B) variants detected in regions of *MYH6* homologous with *MYH7*.

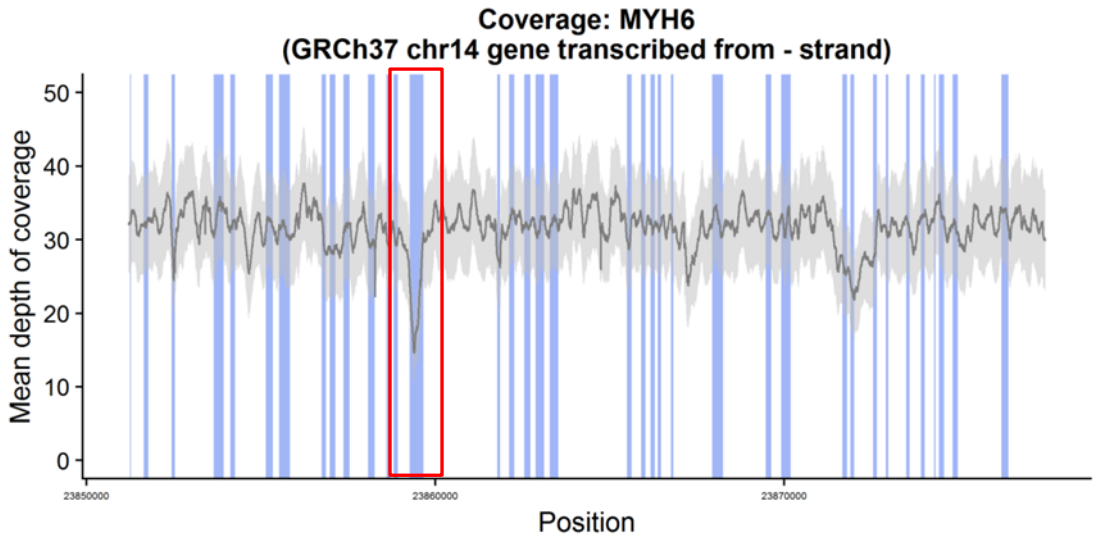


Figure 3.3-7 Mean per base depth of coverage across *MYH6*. Dark grey= mean depth of coverage across 230 HCM samples. Light grey=standard deviation. Blue vertical bars=coding regions. Homologous regions (Exon 26) marked by Red box. Ensembl transcript ID:ENST00000356287.

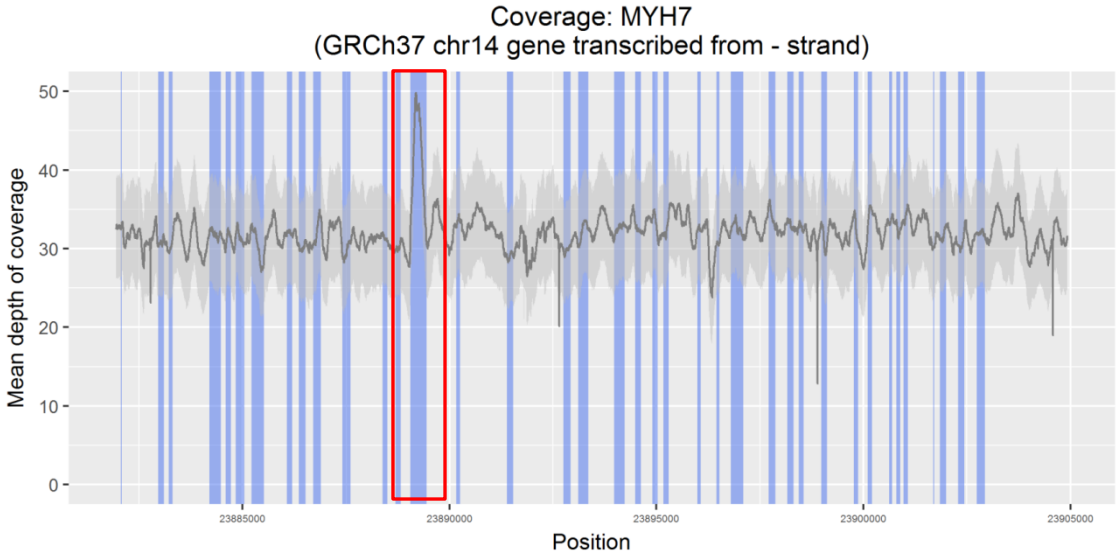


Figure 3.3-8 Mean per base depth of coverage across *MYH7*. Dark grey= mean depth of coverage across 230 HCM samples. Light grey=standard deviation. Blue vertical bars=coding regions. Homologous regions (Exon 27) marked by Red box. Ensembl transcript ID:ENST00000355349.

3.3.5 Gene level analyses

3.3.5.1 Comparing variant frequency

The frequency of rare variants in the 60 selected genes was compared in cases and controls. An overview of the different types of variants detected in each gene is provided in Table 3.3-10.

Figure 3.3-9 shows the number of coding nucleotides and the ExAC missense constraint score (Z score) of each gene. Figure 3.3-10 shows the distribution of the number of coding nucleotides and Z score distribution across genes. The number of rare variants detected in cases and controls showed a very strong positive correlation to the number of coding nucleotides in each gene (cases R=0.979, controls R=0.991, Figure 3.3-11). There was a moderate negative correlation between the number of rare variants and the ExAC missense constraint score (Z score), (cases=-0.463, controls =-0.436, Figure 3.3-11).

In 17/60 (28%) of genes, no rare variants were detected in HCM cases. These genes had a lower than average number of coding nucleotides (cds) (Figure 3.3-9). The *ABCC9* gene was a notable exception having a higher than average number of coding nucleotides (4650 cds) (Table 3.3-9), however this gene had one of the highest constraint scores for missense variants (ExAC Z score, 4.79), which suggests that it is likely to be intolerant to missense variation. A high constraint score was also assigned to the *ACTA1* gene (ExAC Z score 5.86), again suggesting that the lack of rare variants in this gene may be related to constraint rather than gene size per se (Figure 3.3-9).

Genes	n	Mean	SD	Median	Min	Max
All (including TTN)	60	4530	14164	1739	156	107,976
All (excluding TTN)	59	2776	4057	1728	156	26,772
No rare variants	17	1145	1068	897	156	4650

Table 3.3-9 Number of coding nucleotides in all 60 genes, and in the subset of genes in which no rare variants were detected in HCM cases. n (number of genes), SD (Standard deviation), Min (minimum) and Max (maximum). Due to its large size calculations were performed with and without TTN to reduce skewing.

At least one rare variant was detected in every gene in reference controls. In 13/60 (22%) of genes, no truncating variants were detected in either cases or controls. In one gene (*TCAP*) a truncating variant was detected in a single HCM case, however no truncating variants were detected in reference controls.

		HCM cases (n=230)		Reference controls (n=7253)	
		Rare missense	Rare truncating	Rare missense	Rare truncating
1	*ABCC9*	x	x	✓	✓
2	*ACADVL*	x	x	✓	✓
3	*ACTA1*	x	x	✓	✓
4	*AGL*	✓	✓	✓	✓
5	*ALPK3*	✓	✓	✓	✓
6	*ATP5E*	x	x	✓	✓
7	*BAG3*	✓	x	✓	x
8	*BRAF*	✓	x	✓	x
9	*CACNA1C*	✓	x	✓	✓
10	*CALR3*	✓	✓	✓	✓
11	*CASQ2*	✓	x	✓	✓
12	*CAV3*	x	x	✓	✓
13	*COA5*	x	x	✓	✓
14	*COX15*	✓	✓	✓	✓
15	*CRYAB*	✓	x	✓	✓
16	*DES*	✓	x	✓	✓
17	*DSP*	✓	x	✓	✓
18	*FLNC*	✓	✓	✓	✓
19	*FOXRED1*	✓	x	✓	✓
20	*FXN*	✓	✓	✓	✓
21	*GAA*	✓	✓	✓	✓
22	*GLB1*	✓	x	✓	✓
23	*GUSB*	✓	x	✓	✓
24	*HRAS*	x	x	✓	✓
25	*JPH2*	✓	x	✓	✓
26	*KCNQ1*	✓	✓	✓	✓
27	*KLF10*	✓	x	✓	✓
28	*LDB3*	✓	x	✓	✓
29	*LMNA*	x	x	✓	x
30	*MAP2K1*	x	x	✓	x
31	*MAP2K2*	✓	x	✓	✓
32	*MRPL3*	✓	x	✓	✓
33	*MYH6*	✓	x	✓	✓
34	*MYLK2*	✓	x	✓	✓
35	*MYO6*	✓	x	✓	✓
36	*MYOM1*	✓	x	✓	✓
37	*MYOZ2*	✓	x	✓	✓
38	*MYPN*	✓	x	✓	✓
39	*NEXN*	✓	x	✓	✓
40	*NRAS*	x	x	✓	x
41	*OBSCN*	✓	✓	✓	✓

		HCM cases (n=230)		Reference controls (n=7253)	
		Rare missense	Rare truncating	Rare missense	Rare truncating
42	*PDLIM3*	x	x	✓	✓
43	*PTPN11*	✓	x	✓	x
44	*RAF1*	✓	x	✓	x
45	*RYR2*	✓	x	✓	✓
46	*SCO2*	✓	✓	✓	✓
47	*SHOC2*	x	x	✓	x
48	*SLC25A3*	✓	x	✓	✓
49	*SLC25A4*	x	x	✓	x
50	*SOS1*	✓	x	✓	x
51	*TCAP*	x	✓	✓	x
52	*TMEM70*	x	x	✓	✓
53	*TMPO*	✓	x	✓	✓
54	*TNNC1*	✓	x	✓	x
55	*TNNC2*	x	x	✓	x
56	*TRIM63*	✓	x	✓	✓
57	*TSFM*	x	x	✓	✓
58	*TTN*	✓	✓	✓	✓
59	*TTR*	x	x	✓	x
60	*VCL*	✓	x	✓	✓

Table 3.3-10 Summary of findings in each gene.

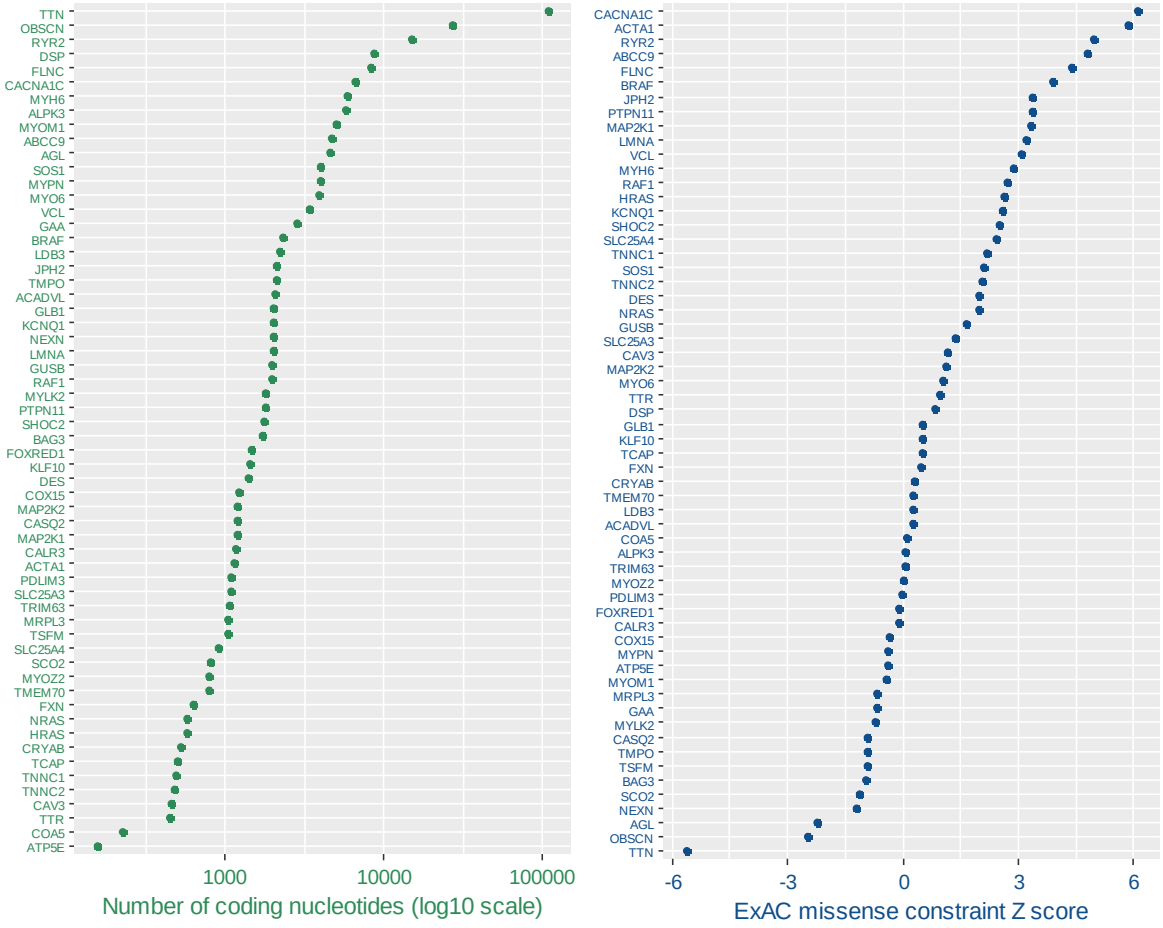


Figure 3.3-9 Left hand side (green): Genes sorted by increasing number of coding nucleotides (log10 scale). Right hand side (blue): Genes sorted by increasing Exome Aggregation Consortium (ExAC) missense constraint scores. Positive Z scores indicate increased constraint (intolerance to variation). Negative Z scores indicate that a gene is tolerant to variation. (see Lek et al 2016[138] or http://exac.broadinstitute.org/ for further information).

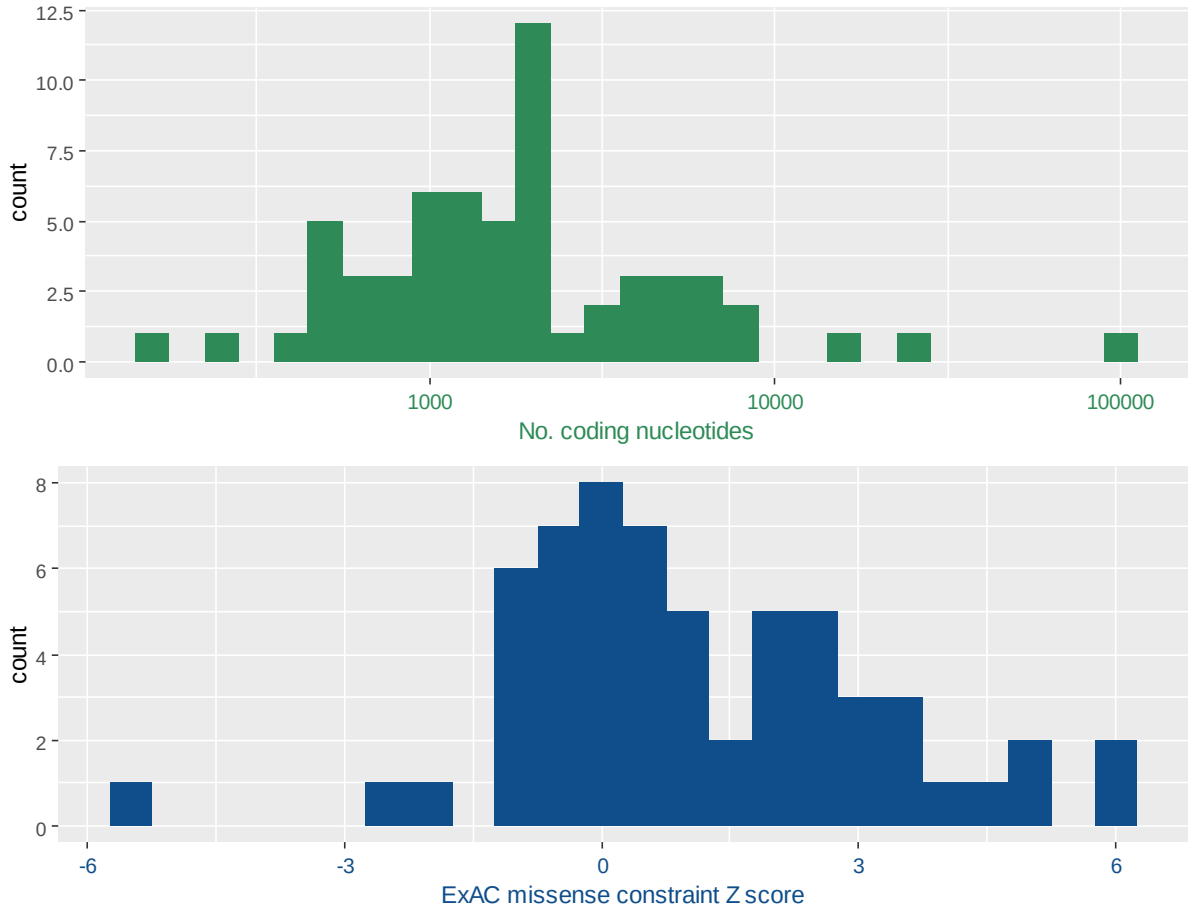


Figure 3.3-10 Distribution of the number of coding nucleotides (using log10 scale) and Exome Aggregation Consortium (ExAC) missense constraint scores (Z scores) across the 60 genes included in these analyses.

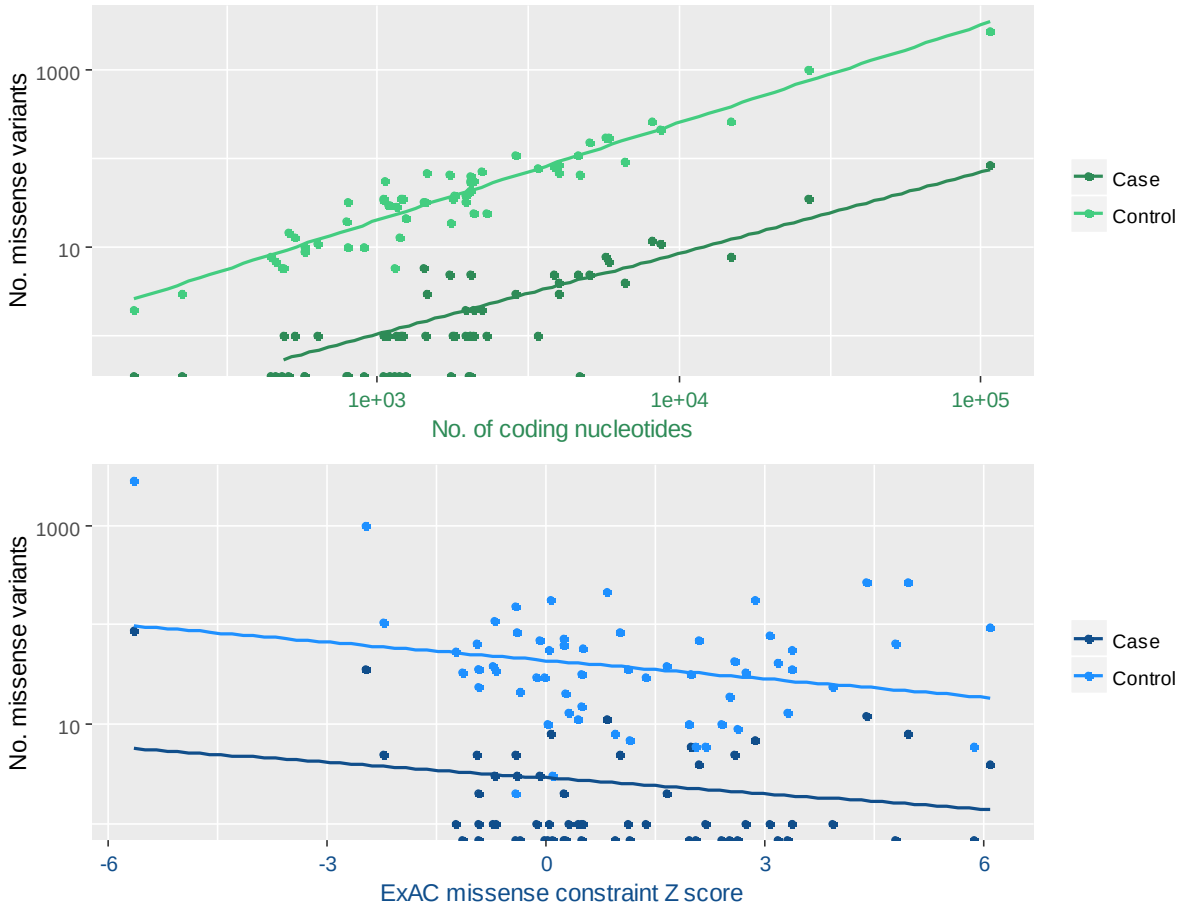


Figure 3.3-11 Scatterplot showing the number of missense variants detected in cases and controls and the number of coding nucleotides and Exome Aggregation Consortium (ExAC) missense constraint scores (Z scores). Number of missense variants plotted on log10 scale. Number of coding nucleotides plotted on log10 scale.

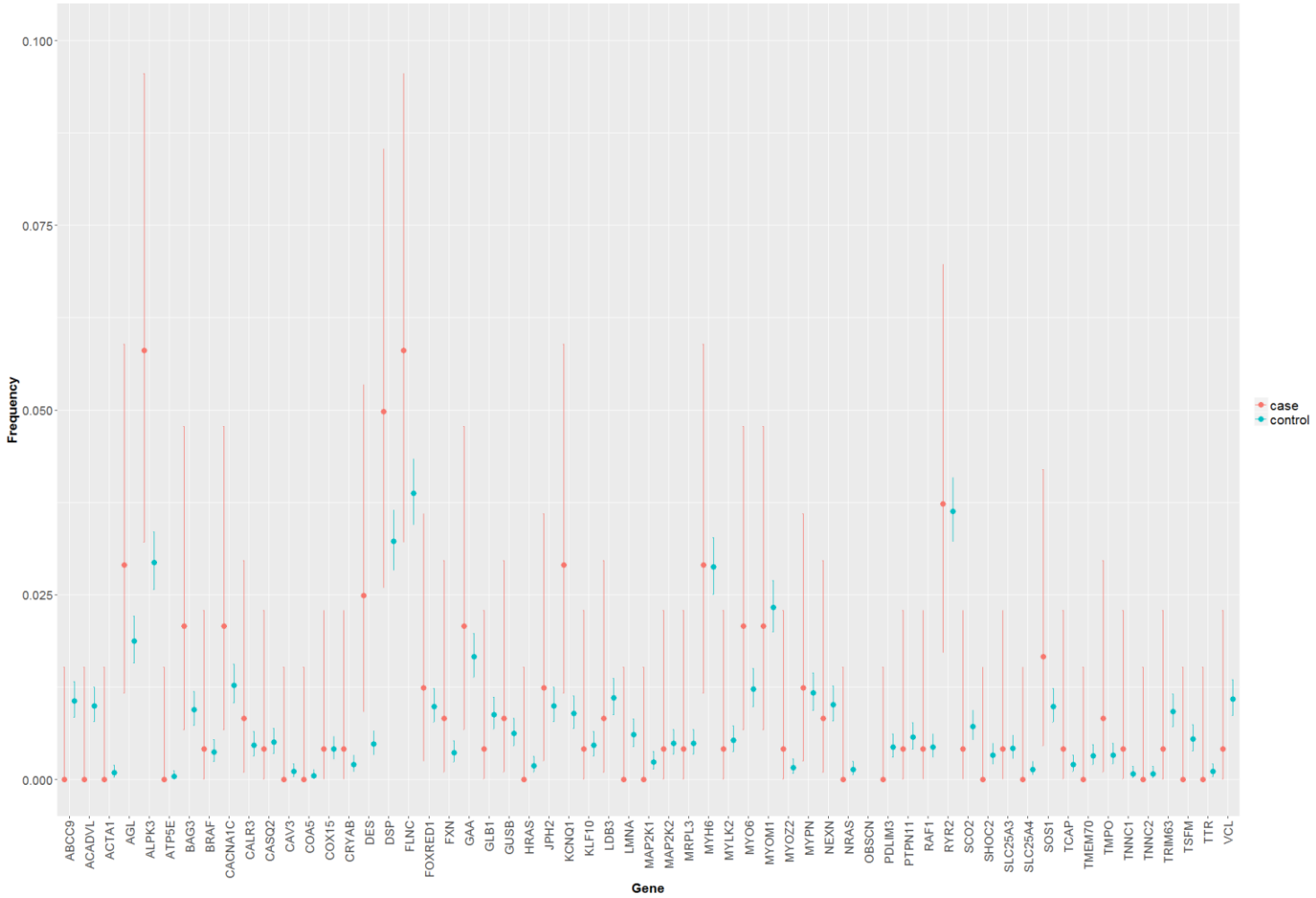


Figure 3.3-12 Proportion of cases (red) and controls (green) with a rare variant in each gene, with 95% confidence intervals under binomial distribution. Data from *TTN* and *OBSCN* genes was not plotted as the high frequency of rare variants in these genes skews the y axis and makes data from other genes difficult to differentiate. Results for these genes were as follows: *TTN* mean =0.473 95% CI [0.408-0.538], *OBSCN* mean=0.174, 95% CI [0.128-0.228].

The proportion of cases and controls with a rare variant in each gene with 95% confidence intervals calculated under the binomial distribution is shown in Figure 3.3-12.

Overall, combining data from all genes, there was no significant difference in the proportion of HCM cases and BRRD controls carrying one or more rare variant (OR 1.28 [95% CI 0.94-1.77] p=0.123). However, there was evidence supporting a small excess of cases with rare non-truncating variants (OR 1.35 [95% CI 1.02-1.80] p=0.03) and rare truncating variants (OR 1.74 [95% CI 1.01-2.83] p=0.03), (Table S9). There was no significant difference in the proportion of HCM cases with one or more rare synonymous variant compared to BRRD controls, indicating comparable sensitivity for rare variant detection (OR 0.91 [95% CI 0.67-1.22] p=0.56, Table S9).

The results of Fisher's exact test and odds ratio analyses are presented combining all rare variants (Table 3.3-11 and Figure 3.3-13), and separately for missense (Table 3.3-12 and Figure 3.3-14), truncating (Table 3.3-13 and Figure 3.3-15) and non-truncating variants (Table 3.3-14 and Figure 3.3-16).

All rare variants											
No.	**Gene**	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control Without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
1	ABCC9*	0	230	0	72	7181	0.0099	0.73	0.4259	0.0106	2.4727
2	ACADVL*	0	230	0	75	7178	0.0103	0.7339	0.409	0.0102	2.3712
3	ACTA1*	0	230	0	7	7246	0.0010	0.2468	3.9203	0.088	29.444
4	AGL	7	223	0.0304	128	7125	0.0176	0.1979	1.7471	0.6806	3.7654
5	ALPK3	13	217	0.0565	191	7062	0.0263	0.0116	2.2147	1.1394	3.9596
6	ATP5E*	0	230	0	3	7250	0.0004	0.1457	7.8428	0.1587	79.6199
7	BAG3	5	225	0.0217	66	7187	0.0091	0.0666	2.4194	0.7533	6.0189
8	BRAF	1	229	0.0043	28	7225	0.0039	0.5963	1.1268	0.0274	6.8778
9	CACNA1C	5	225	0.0217	96	7157	0.0132	0.2408	1.6566	0.5209	4.0593
10	CALR3	2	228	0.0087	35	7218	0.0048	0.3154	1.8088	0.2095	7.1158
11	CASQ2	1	229	0.0043	38	7215	0.0052	1	0.8291	0.0204	4.9537
12	CAV3*	0	230	0	8	7245	0.0011	0.2702	3.4842	0.0792	25.3271
13	COA5*	0	230	0	4	7249	0.0006	0.1722	6.2732	0.1321	56.3907
14	COX15	1	229	0.0043	25	7228	0.0034	0.5565	1.2625	0.0306	7.7819
15	CRYAB	1	229	0.0043	15	7238	0.0021	0.3935	2.1068	0.0498	13.8036
16	DES	6	224	0.0261	36	7217	0.0050	0.0016	5.367	1.8295	13.0371
17	DSP	12	218	0.0522	231	7022	0.0318	0.089	1.6731	0.839	3.0403
18	FLNC	14	216	0.0609	268	6985	0.0370	0.0752	1.6891	0.8958	2.947
19	FOXRED1	3	227	0.013	74	7179	0.0102	0.5128	1.2821	0.2566	3.9435
20	FXN	2	228	0.0087	13	7240	0.0018	0.076	4.8832	0.532	21.7616
21	GAA	4	226	0.0174	120	7133	0.0165	0.7929	1.0521	0.2796	2.8023
22	GLB1	1	229	0.0043	61	7192	0.0084	1	0.5149	0.0128	3.0078
23	GUSB	2	228	0.0087	47	7206	0.0065	0.6629	1.3448	0.1573	5.1914
24	HRAS*	0	230	0	14	7239	0.0019	0.396	2.0892	0.0494	13.687
25	JPH2	2	228	0.0087	63	7190	0.0087	1	1.0011	0.1179	3.8116
26	KCNQ1	6	224	0.0261	51	7202	0.0070	0.0077	3.7816	1.3122	8.9405
27	KLF10	1	229	0.0043	35	7218	0.0048	1	0.9006	0.0221	5.4075

All rare variants											
No.	**Gene**	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control Without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
28	LDB3	2	228	0.0087	83	7170	0.0114	1	0.7578	0.0897	2.8569
29	LMNA*	0	230	0	41	7212	0.0057	1	0.7435	0.0183	4.4169
30	MAP2K1*	0	230	0	13	7240	0.0018	0.3766	2.2387	0.0527	14.829
31	MAP2K2	1	229	0.0043	37	7216	0.0051	1	0.8517	0.0209	5.0964
32	MRPL3	1	229	0.0043	37	7216	0.0051	1	0.8517	0.0209	5.0964
33	MYH6	7	223	0.0304	195	7058	0.0269	0.6781	1.1361	0.4456	2.4253
34	MYLK2	1	229	0.0043	40	7213	0.0055	1	0.7875	0.0194	4.6908
35	MYO6	5	225	0.0217	86	7167	0.0119	0.2051	1.8518	0.5808	4.556
36	MYOM1	5	225	0.0217	169	7084	0.0233	1	0.9315	0.2957	2.25
37	MYOZ2	1	229	0.0043	12	7241	0.0017	0.3338	2.6345	0.0614	17.9433
38	MYPN	3	227	0.013	88	7165	0.0121	0.7589	1.076	0.2161	3.2898
39	NEXN	1	229	0.0043	63	7190	0.0087	0.7239	0.4984	0.0124	2.9081
40	NRAS*	0	230	0	10	7243	0.0014	0.3148	2.8502	0.066	19.7583
41	OBSCN	40	190	0.1739	1073	6180	0.1479	0.2998	1.2125	0.8345	1.7246
42	PDLIM3*	0	230	0	33	7220	0.0045	1	0.9194	0.0225	5.5322
43	PTPN11	1	229	0.0043	36	7217	0.0050	1	0.8754	0.0215	5.2475
44	RAF1	1	229	0.0043	33	7220	0.0045	1	0.9554	0.0234	5.7603
45	RYR2	9	221	0.0391	268	6985	0.0370	0.8582	1.0614	0.4737	2.0829
46	SCO2	1	229	0.0043	44	7209	0.0061	1	0.7155	0.0176	4.2401
47	SHOC2*	0	230	0	20	7233	0.0028	0.5002	1.4911	0.0359	9.35
48	SLC25A3	1	229	0.0043	32	7221	0.0044	1	0.9854	0.0241	5.954
49	SLC25A4*	0	230	0	10	7243	0.0014	0.3148	2.8502	0.066	19.7583
50	SOS1	4	226	0.0174	69	7184	0.0095	0.2874	1.8426	0.4841	5.0006
51	TCAP	1	229	0.0043	15	7238	0.0021	0.3935	2.1068	0.0498	13.8036
52	TMEM70*	0	230	0	24	7229	0.0033	0.5595	1.2519	0.0304	7.7161
53	TMPO	2	228	0.0087	25	7228	0.0034	0.2008	2.5358	0.2894	10.2719
54	TNNC1	1	229	0.0043	6	7247	0.0008	0.1964	5.2717	0.1142	43.6781

All rare variants											
No.	**Gene**	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control Without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
55	TNNC2*	0	230	0	6	7247	0.0008	0.2227	4.4801	0.099	35.0764
56	TRIM63	1	229	0.0043	63	7190	0.0087	0.7239	0.4984	0.0124	2.9081
57	TSFM*	0	230	0	41	7212	0.0057	1	0.7435	0.0183	4.4169
58	TTN	92	138	0.4	2905	4348	0.4005	1	0.9978	0.7547	1.3143
59	TTR*	0	230	0	8	7245	0.0011	0.2702	3.4842	0.0792	25.3271
60	VCL	1	229	0.0043	82	7171	0.0113	0.5217	0.3819	0.0095	2.21

Table 3.3-11 Case control analyses-All rare variants. HCM cases (n=230) compared to BRRD reference controls (n=7253). Bonferroni corrected p=≤0.0008. OR=odds ratio. CI=95% confidence interval. *For genes where the case allele count was zero, 0.5 (Haldane continuity adjustment [155])was added to all cells before calculating the odds ratio.

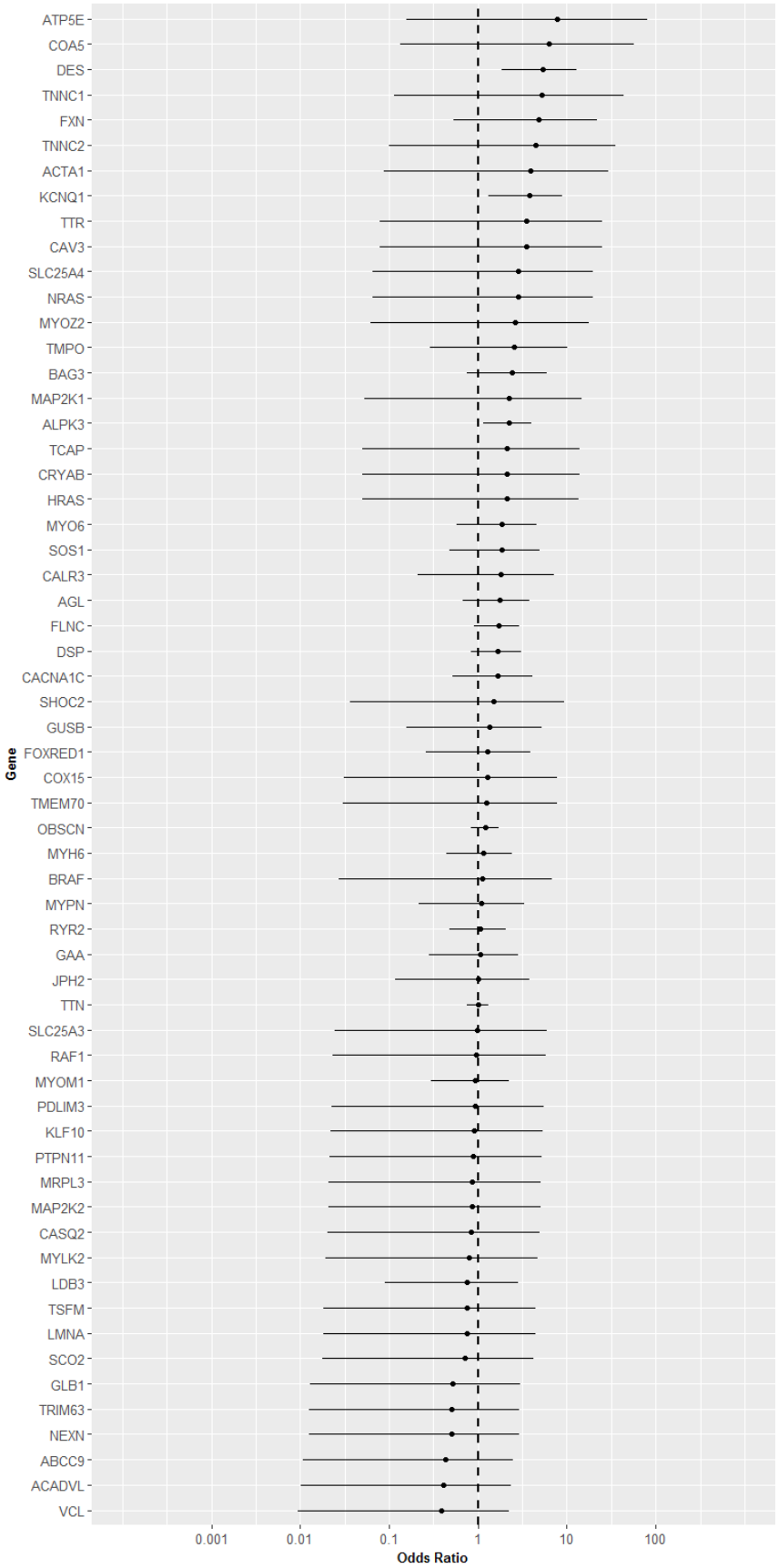


Figure 3.3-13 Odds ratio with 95% CI for all rare variants. HCM cases (n=230) compared to BRRD controls (n=7253). Data plotted on log10 scale.

Missense variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
ABCC9*	0	230	0	65	7188	0.0090	0.7244	0.47	0.01	2.75
ACADVL*	0	230	0	63	7190	0.0087	0.7239	0.49	0.01	2.84
ACTA1*	0	230	0	6	7247	0.0008	0.2227	4.48	0.10	35.08
AGL	5	225	0.0217	107	7146	0.0148	0.3972	1.48	0.47	3.62
ALPK3	8	222	0.0348	176	7077	0.0243	0.2788	1.45	0.61	2.97
ATP5E*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
BAG3	5	225	0.0217	65	7188	0.0090	0.0635	2.46	0.76	6.12
BRAF	1	229	0.0044	24	7229	0.0033	0.5424	1.32	0.03	8.14
CACNA1C	4	226	0.0174	93	7160	0.0128	0.5438	1.36	0.36	3.66
CALR3	1	229	0.0044	29	7224	0.0040	0.6088	1.09	0.03	6.62
CASQ2	1	229	0.0044	36	7217	0.0050	1.0000	0.88	0.02	5.25
CAV3*	0	230	0	7	7246	0.0010	0.2468	3.92	0.09	29.44
COA5*	0	230	0	3	7250	0.0004	0.1457	7.84	0.16	79.62
COX15*	0	230	0	21	7232	0.0029	0.5157	1.42	0.03	8.88
CRYAB	1	229	0.0044	13	7240	0.0018	0.3543	2.43	0.06	16.32
DES	6	224	0.0261	32	7221	0.0044	0.0010	6.04	2.04	14.84
DSP	11	219	0.0478	213	7040	0.0294	0.1129	1.66	0.80	3.09
FLNC	12	218	0.0522	263	6990	0.0363	0.2096	1.46	0.73	2.65
FOXRED1	3	227	0.0130	70	7183	0.0097	0.4915	1.36	0.27	4.18
FXN	1	229	0.0044	11	7242	0.0015	0.3126	2.87	0.07	19.93
GAA	3	227	0.0130	108	7145	0.0149	1.0000	0.87	0.18	2.66
GLB1	1	229	0.0044	57	7196	0.0079	1.0000	0.55	0.01	3.23
GUSB	2	228	0.0087	39	7214	0.0054	0.3607	1.62	0.19	6.34
HRAS*	0	230	0	9	7244	0.0012	0.2928	3.14	0.07	22.19
JPH2	1	229	0.0044	56	7197	0.0077	1.0000	0.56	0.01	3.29
KCNQ1	5	225	0.0217	43	7210	0.0059	0.0154	3.73	1.14	9.51
KLF10	1	229	0.0044	32	7221	0.0044	1.0000	0.99	0.02	5.95

Missense variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
LDB3	2	228	0.0087	71	7182	0.0098	1.0000	0.89	0.10	3.36
LMNA*	0	230	0	41	7212	0.0057	1.0000	0.74	0.02	4.42
MAP2K1*	0	230	0	13	7240	0.0018	0.3766	2.24	0.05	14.83
MAP2K2	1	229	0.0044	36	7217	0.0050	1.0000	0.88	0.02	5.25
MRPL3	1	229	0.0044	34	7219	0.0047	1.0000	0.93	0.02	5.58
MYH6	7	223	0.0304	176	7077	0.0243	0.5121	1.26	0.49	2.70
MYLK2	1	229	0.0044	38	7215	0.0052	1.0000	0.83	0.02	4.95
MYO6	5	225	0.0217	83	7170	0.0114	0.1967	1.92	0.60	4.73
MYOM1	5	225	0.0217	150	7103	0.0207	0.8135	1.05	0.33	2.55
MYOZ2*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
MYPN	3	227	0.0130	85	7168	0.0117	0.7524	1.11	0.22	3.41
NEXN	1	229	0.0044	54	7199	0.0075	1.0000	0.58	0.01	3.42
NRAS*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
OBSCN	35	195	0.1522	1000	6253	0.1379	0.5601	1.12	0.76	1.63
PDLIM3*	0	230	0	30	7223	0.0041	1.0000	1.01	0.02	6.11
PTPN11	1	229	0.0044	35	7218	0.0048	1.0000	0.90	0.02	5.41
RAF1	1	229	0.0044	33	7220	0.0046	1.0000	0.96	0.02	5.76
RYR2	8	222	0.0348	265	6988	0.0365	1.0000	0.95	0.40	1.93
SCO2*	0	230	0	33	7220	0.0046	1.0000	0.92	0.02	5.53
SHOC2*	0	230	0	19	7234	0.0026	0.4841	1.57	0.04	9.87
SLC25A3	1	229	0.0044	30	7223	0.0041	0.6208	1.05	0.03	6.38
SLC25A4*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
SOS1	4	226	0.0174	69	7184	0.0095	0.2874	1.84	0.48	5.00
TCAP*	0	230	0	15	7238	0.0021	0.4147	1.96	0.05	12.71
TMEM70*	0	230	0	20	7233	0.0028	0.5002	1.49	0.04	9.35
TMPO	2	228	0.0087	24	7229	0.0033	0.1896	2.64	0.30	10.74
TNNC1	1	229	0.0044	6	7247	0.0008	0.1964	5.27	0.11	43.68

Missense variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
TNNC2*	0	230	0	6	7247	0.0008	0.2227	4.48	0.10	35.08
TRIM63	1	229	0.0044	55	7198	0.0076	1.0000	0.57	0.01	3.35
TSFM*	0	230	0	35	7218	0.0048	1.0000	0.87	0.02	5.20
TTN	86	144	0.3739	2760	4493	0.3805	0.8903	0.97	0.73	1.28
TTR*	0	230	0	8	7245	0.0011	0.2702	3.48	0.08	25.33
VCL	1	229	0.0044	77	7176	0.0106	0.7344	0.41	0.01	2.36

Table 3.3-12 Case control analyses-Missense variants. HCM cases (n=230) compared to BRRD reference controls (n=7253). Bonferroni corrected p=≤0.0008. OR=odds ratio. CI=95% confidence interval. *For genes where the case allele count was zero, 0.5 (Haldane continuity adjustment [155])was added to all cells before calculating the odds ratio.

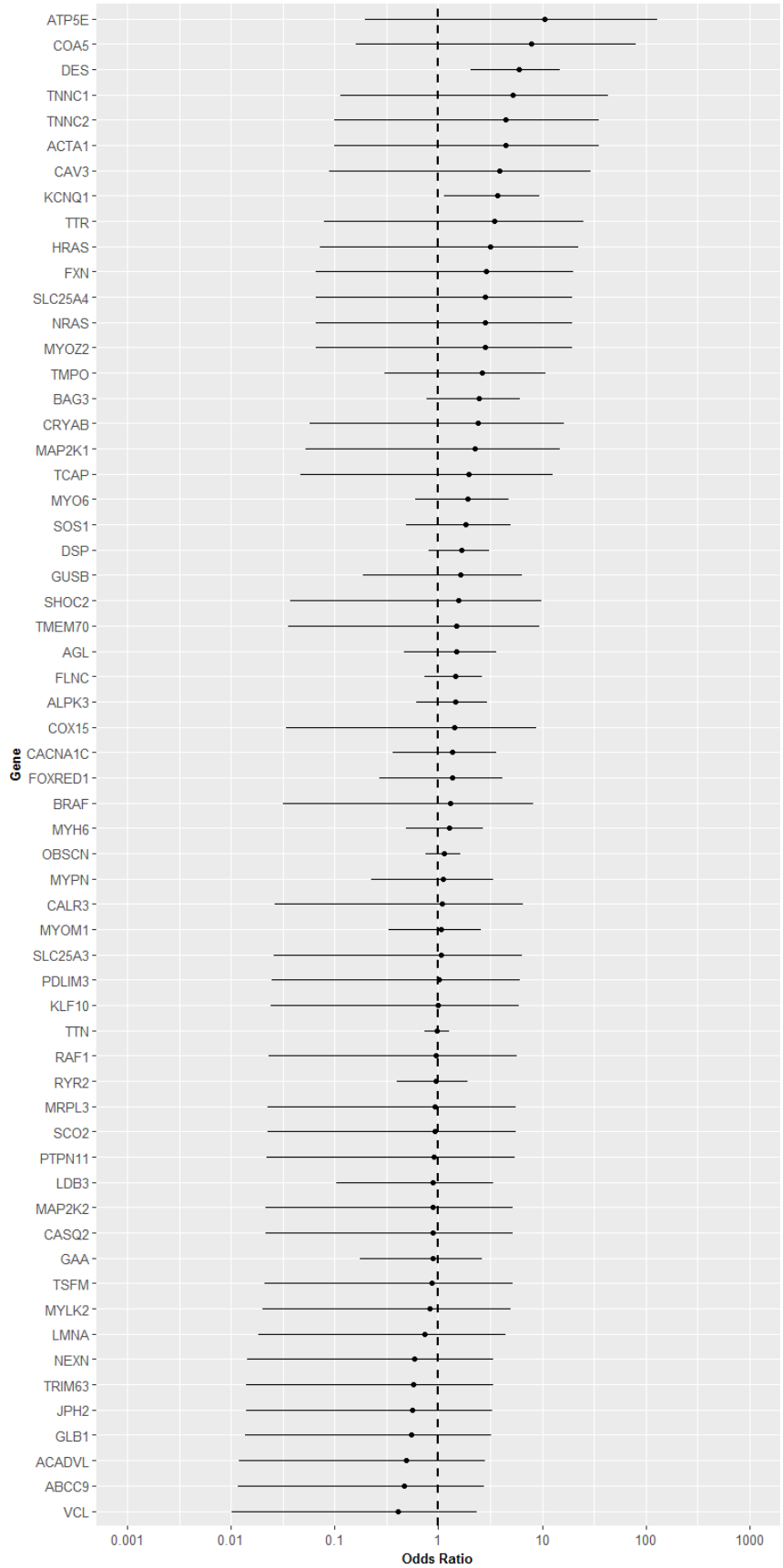


Figure 3.3-14 Odds ratio with 95% CI for missense variants. HCM cases (n=230) compared to BRRD controls (n=7253). Data plotted on log10 scale.

Truncating variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control Without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
ABCC9*	0	230	0	7	7246	0.0010	0.2468	3.92	0.09	29.44
ACADVL*	0	230	0	11	7242	0.0015	0.3360	2.61	0.06	17.79
ACTA1*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
AGL	2	228	0.0087	21	7232	0.0029	0.1564	3.02	0.34	12.48
ALPK3	5	225	0.0217	10	7243	0.0014	0.0001	16.07	4.27	52.05
ATP5E*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
BAG3	0	230	0	0	7253	0	-	-	-	-
BRAF	0	230	0	0	7253	0	-	-	-	-
CACNA1C*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
CALR3	1	229	0.0044	5	7248	0.0007	0.1709	6.33	0.13	56.91
CASQ2*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
CAV3*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
COA5*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
COX15	1	229	0.0044	4	7249	0.0006	0.1446	7.91	0.16	80.29
CRYAB*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
DES*	0	230	0	4	7249	0.0006	0.1722	6.27	0.13	56.39
DSP*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
FLNC	1	229	0.0044	2	7251	0.0003	0.0894	15.81	0.27	303.95
FOXRED1*	0	230	0	4	7249	0.0006	0.1722	6.27	0.13	56.39
FXN	1	229	0.0044	2	7251	0.0003	0.0894	15.81	0.27	303.95
GAA	1	229	0.0044	12	7241	0.0017	0.3338	2.63	0.06	17.94
GLB1*	0	230	0	3	7250	0.0004	0.1457	7.84	0.16	79.62
GUSB*	0	230	0	6	7247	0.0008	0.2227	4.48	0.10	35.08
HRAS*	0	230	0	4	7249	0.0006	0.1722	6.27	0.13	56.39
JPH2*	0	230	0	4	7249	0.0006	0.1722	6.27	0.13	56.39
KCNQ1	1	229	0.0044	7	7246	0.0010	0.2211	4.52	0.10	35.38
KLF10*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41

Truncating variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control Without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
LDB3*	0	230	0	4	7249	0.0006	0.1722	6.27	0.13	56.39
LMNA	0	230	0	0	7253	0	-	-	-	-
MAP2K2*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
MRPL3*	0	230	0	3	7250	0.0004	0.1457	7.84	0.16	79.62
MYH6*	0	230	0	17	7236	0.0023	0.4505	1.74	0.04	11.11
MYLK2*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
MYO6*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
MYOM1*	0	230	0	14	7239	0.0019	0.3960	2.09	0.05	13.69
MYOZ2*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
MYPN*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
NEXN*	0	230	0	7	7246	0.0010	0.2468	3.92	0.09	29.44
NRAS	0	230	0	0	7253	0	-	-	-	-
OBSCN	3	227	0.0130	60	7193	0.0083	0.4445	1.58	0.32	4.92
PDLIM3*	0	230	0	3	7250	0.0004	0.1457	7.84	0.16	79.62
PTPN11	0	230	0	0	7253	0	-	-	-	-
RAF1	0	230	0	0	7253	0	-	-	-	-
RYR2*	0	230	0	3	7250	0.0004	0.1457	7.84	0.16	79.62
SCO2	1	229	0.0044	10	7243	0.0014	0.2908	3.16	0.07	22.38
SHOC2	0	230	0	0	7253	0	-	-	-	-
SLC25A3*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
SLC25A4	0	230	0	0	7253	0	-	-	-	-
SOS1	0	230	0	0	7253	0	-	-	-	-
TCAP*	1	229	0.00435	0	7253	0.0000	0.0028	62.83	3.26	3610.86
TMEM70*	0	230	0	4	7249	0.0006	0.1722	6.27	0.13	56.39
TMPO*	0	230	0	1	7252	0.0001	0.0901	15.68	0.26	301.41
TNNC1	0	230	0	0	7253	0	-	-	-	-
TNNC2	0	230	0	0	7253	0	-	-	-	-

Truncating variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control Without**	**Control Freq.**	**FET p value**	**OR**	**Lower_CI**	**Upper_CI**
TRIM63*	0	230	0	8	7245	0.0011	0.2702	3.48	0.08	25.33
TSFM*	0	230	0	6	7247	0.0008	0.2227	4.48	0.10	35.08
TTN	3	227	0.0130	95	7158	0.0131	1.0000	1.00	0.20	3.04
TTR	0	230	0	0	7253	0	-	-	-	-
VCL*	0	230	0	5	7248	0.0007	0.1978	5.23	0.11	43.41

Table 3.3-13 Case control analyses-Truncating variants. HCM cases (n=230) compared to BRRD reference controls (n=7253). Bonferroni corrected p=≤0.0008. OR=odds ratio. CI=95% confidence interval. *For genes where the case allele count was zero, 0.5 (Haldane continuity adjustment [155])was added to all cells before calculating the odds ratio.

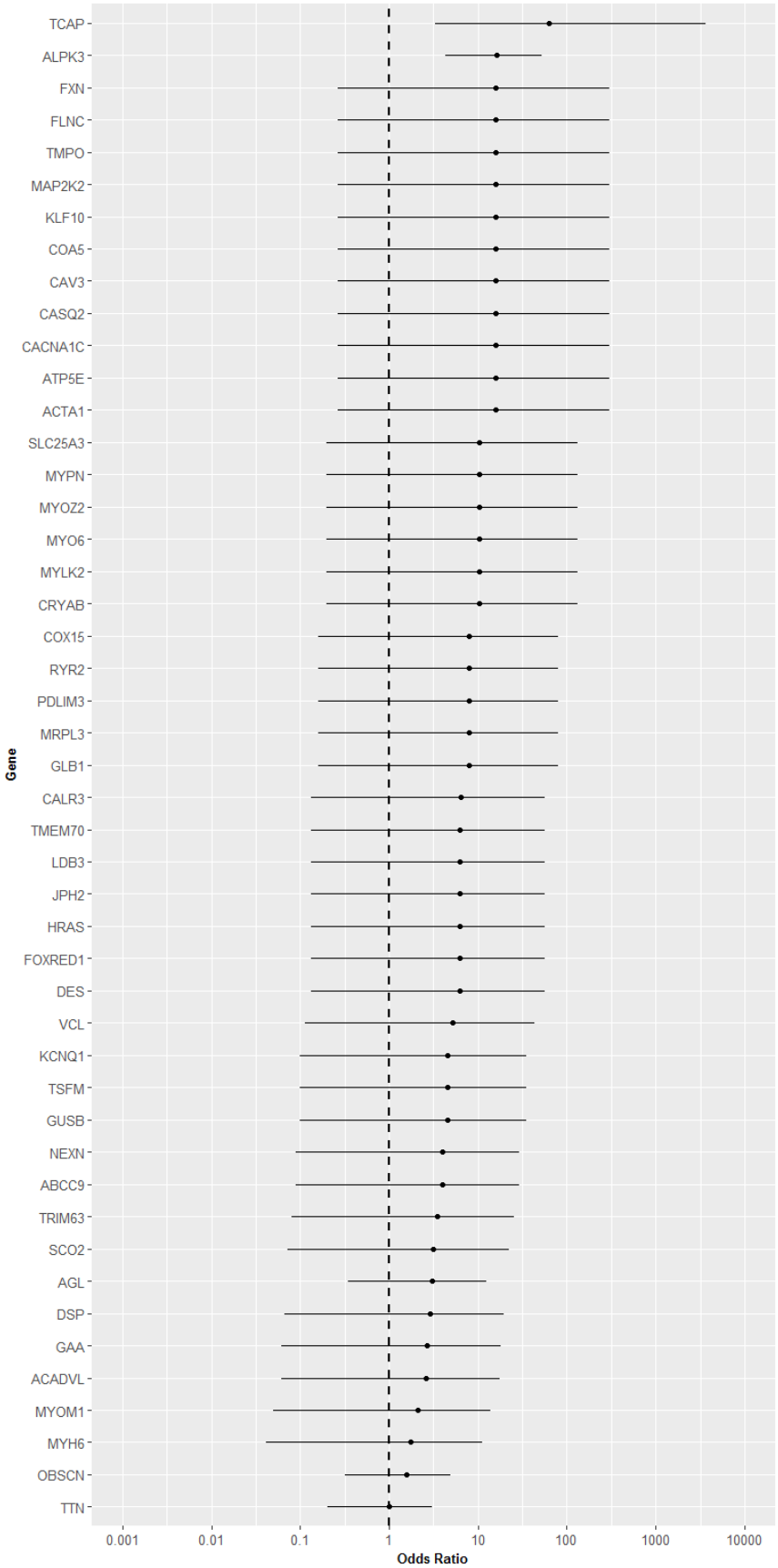


Figure 3.3-15 Odds ratio with 95% CI for truncating variants. HCM cases (n=230) compared to BRRD controls (n=7253). Data plotted on log10 scale.

Non-truncating variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control without**	**Control Freq.**	**FET p value**	**OR***	**Lower_CI**	**Upper_CI**
ABCC9*	0	230	0	65	7188	0.0090	0.7244	0.47	0.01	2.75
ACADVL*	0	230	0	64	7189	0.0088	0.7241	0.48	0.01	2.79
ACTA1*	0	230	0	6	7247	0.0008	0.2227	4.48	0.10	35.08
AGL	5	225	0.0217	107	7146	0.0148	0.3972	1.48	0.47	3.62
ALPK3	8	222	0.0348	181	7072	0.0250	0.3869	1.41	0.59	2.88
ATP5E*	0	230	0	2	7251	0.0003	0.1183	10.45	0.20	130.54
BAG3	5	225	0.0217	66	7187	0.0091	0.0666	2.42	0.75	6.02
BRAF	1	229	0.0044	28	7225	0.0039	0.5963	1.13	0.03	6.88
CACNA1C	5	225	0.0217	95	7158	0.0131	0.2368	1.67	0.53	4.10
CALR3	1	229	0.0044	30	7223	0.0041	0.6208	1.05	0.03	6.38
CASQ2	1	229	0.0044	37	7216	0.0051	1.0000	0.85	0.02	5.10
CAV3*	0	230	0	7	7246	0.0010	0.2468	3.92	0.09	29.44
COA5*	0	230	0	3	7250	0.0004	0.1457	7.84	0.16	79.62
COX15*	0	230	0	21	7232	0.0029	0.5157	1.42	0.03	8.88
CRYAB	1	229	0.0044	13	7240	0.0018	0.3543	2.43	0.06	16.32
DES	6	224	0.0261	32	7221	0.0044	0.0010	6.04	2.04	14.84
DSP	12	218	0.0522	221	7032	0.0305	0.0783	1.75	0.88	3.18
FLNC	13	217	0.0565	266	6987	0.0367	0.1527	1.57	0.81	2.80
FOXRED1	3	227	0.0130	70	7183	0.0097	0.4915	1.36	0.27	4.18
FXN	1	229	0.0044	11	7242	0.0015	0.3126	2.87	0.07	19.93
GAA	3	227	0.0130	108	7145	0.0149	1.0000	0.87	0.18	2.66
GLB1	1	229	0.0044	58	7195	0.0080	1.0000	0.54	0.01	3.17
GUSB	2	228	0.0087	41	7212	0.0057	0.3829	1.54	0.18	6.00
HRAS*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
JPH2	2	228	0.0087	59	7194	0.0081	0.7118	1.07	0.13	4.08
KCNQ1	5	225	0.0217	44	7209	0.0061	0.0167	3.64	1.12	9.28
KLF10	1	229	0.0044	34	7219	0.0047	1.0000	0.93	0.02	5.58

Non-truncating variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control without**	**Control Freq.**	**FET p value**	**OR***	**Lower_CI**	**Upper_CI**
LDB3	2	228	0.0087	79	7174	0.0109	1.0000	0.80	0.09	3.01
LMNA*	0	230	0	41	7212	0.0057	1.0000	0.74	0.02	4.42
MAP2K1*	0	230	0	13	7240	0.0018	0.3766	2.24	0.05	14.83
MAP2K2	1	229	0.0044	36	7217	0.0050	1.0000	0.88	0.02	5.25
MRPL3	1	229	0.0044	34	7219	0.0047	1.0000	0.93	0.02	5.58
MYH6	7	223	0.0304	178	7075	0.0245	0.5156	1.25	0.49	2.67
MYLK2	1	229	0.0044	38	7215	0.0052	1.0000	0.83	0.02	4.95
MYO6	5	225	0.0217	84	7169	0.0116	0.1993	1.90	0.59	4.67
MYOM1	5	225	0.0217	155	7098	0.0214	0.8187	1.02	0.32	2.46
MYOZ2	1	229	0.0044	10	7243	0.0014	0.2908	3.16	0.07	22.38
MYPN	3	227	0.0130	86	7167	0.0119	0.7545	1.10	0.22	3.37
NEXN	1	229	0.0044	56	7197	0.0077	1.0000	0.56	0.01	3.29
NRAS*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
OBSCN	37	193	0.1609	1013	6240	0.1397	0.3848	1.18	0.80	1.70
PDLIM3*	0	230	0	30	7223	0.0041	1.0000	1.01	0.02	6.11
PTPN11	1	229	0.0044	36	7217	0.0050	1.0000	0.88	0.02	5.25
RAF1	1	229	0.0044	33	7220	0.0046	1.0000	0.96	0.02	5.76
RYR2	9	221	0.0391	265	6988	0.0365	0.8575	1.07	0.48	2.11
SCO2*	0	230	0	34	7219	0.0047	1.0000	0.89	0.02	5.36
SHOC2*	0	230	0	20	7233	0.0028	0.5002	1.49	0.04	9.35
SLC25A3	1	229	0.0044	30	7223	0.0041	0.6208	1.05	0.03	6.38
SLC25A4*	0	230	0	10	7243	0.0014	0.3148	2.85	0.07	19.76
SOS1	4	226	0.0174	69	7184	0.0095	0.2874	1.84	0.48	5.00
TCAP*	0	230	0	15	7238	0.0021	0.4147	1.96	0.05	12.71
TMEM70*	0	230	0	20	7233	0.0028	0.5002	1.49	0.04	9.35
TMPO	2	228	0.0087	24	7229	0.0033	0.1896	2.64	0.30	10.74
TNNC1	1	229	0.0044	6	7247	0.0008	0.1964	5.27	0.11	43.68

Non-truncating variants										
Gene	**Case with**	**Case without**	**Case Freq.**	**Control with**	**Control without**	**Control Freq.**	**FET p value**	**OR***	**Lower_CI**	**Upper_CI**
TNNC2*	0	230	0	6	7247	0.0008	0.2227	4.48	0.10	35.08
TRIM63	1	229	0.0044	55	7198	0.0076	1.0000	0.57	0.01	3.35
TSFM*	0	230	0	35	7218	0.0048	1.0000	0.87	0.02	5.20
TTN	89	141	0.3870	2810	4443	0.3874	1.0000	1.00	0.75	1.32
TTR*	0	230	0	8	7245	0.0011	0.2702	3.48	0.08	25.33
VCL	1	229	0.0044	77	7176	0.0106	0.7344	0.41	0.01	2.36

Table 3.3-14 Case control analyses-Non-truncating variants. HCM cases (n=230) compared to BRRD reference controls (n=7253). Bonferroni corrected p≤0.0008 (non-truncating category includes missense, in frame ins/dels and other variants annotated as "MODERATE" impact by SNPEff) OR=odds ratio. CI=95% confidence interval. *For genes where the case allele count was zero, 0.5 (Haldane continuity adjustment [155]) was added to all cells before calculating the odds ratio.

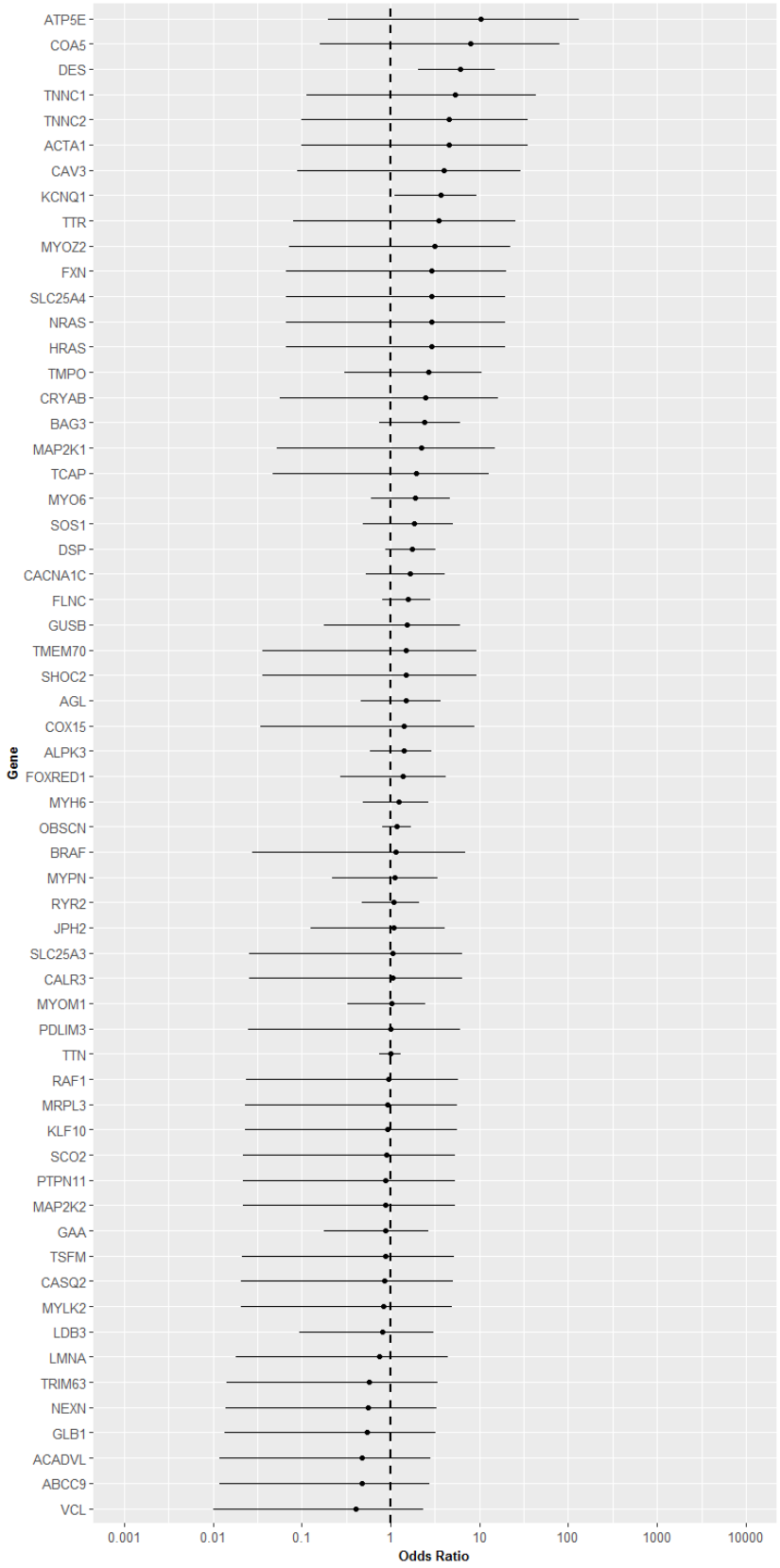


Figure 3.3-16 Odds ratios with 95% CI for non-truncating variants (all variants annotated as MODERATE by SNPEff, including missense and in-frame insertions and deletions). HCM cases (n=230) compared to BRRD controls (n=7253). Data plotted on log10 scale.

3.3.5.2 Comparing variant distribution

In addition to comparing the frequency of rare variants between cases and controls, differences in variant distribution were also assessed. The distribution of rare missense variants was compared between cases and controls in 13 of the 60 genes analysed (Table 3.3-15); there were too few missense variants in the other genes to perform these comparative analyses. While visualisation of the Kernel Density Plots indicated a possible difference in variant distribution between cases and controls in some of the analysed genes (Figure 3.3-17), analysis using the Kolmogorov-Smirnov Test (K-ST) indicated that the observed differences were not statistically significant (Table 3.3-15).

	Kolmogorov-Smirnov Test	
	D value	p -value
AGL	0.20741	0.9874
ALPK3	0.29397	0.5289
BAG3	0.5	0.2058
DES	0.4	0.5176
DSP	0.35574	0.1813
FLNC	0.19167	0.7999
KCNQ1	0.43889	0.3662
MYH6	0.42974	0.1206
MYO6	0.30149	0.7915
MYOM1	0.26364	0.8937
OBSCN	0.086433	0.9637
RYR2	0.21711	0.8608
TTN	0.063042	0.9047

Table 3.3-15 Kolmogorov-Smirnov Test results (genes with over 5 rare missense variant)

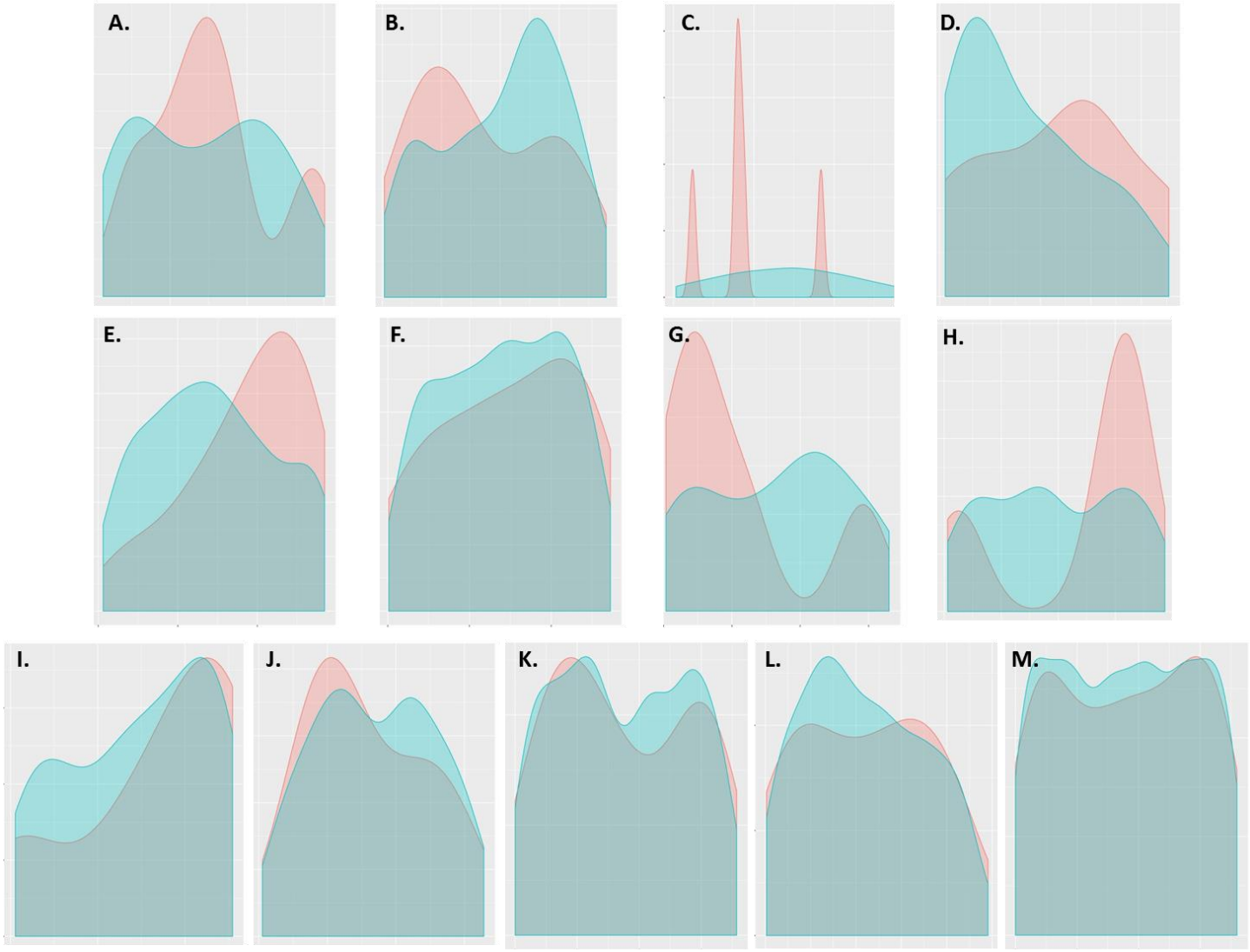


Figure 3.3-17 Density Plots. X axis=codon number. Y axis=density. Red = HCM cases. Blue=BRRD reference controls. A. *AGL*, B. *ALPK3*, C. *BAG3*, D. *DES*, E. *DSP*, F. *FLNC*, G. *KCNQ1*, H. *MYH6*, I. *MYO6*, J. *MYOM1*, K. *OBSCN*, L. *RYR2*, M. *TTN*.

3.3.5.3 ALPK3

Across all 60 genes and variant types, the strongest evidence supporting a difference in rare variant frequency between cases and controls was found for truncating variants in the *ALPK3* gene (p=0.0001, Bonferroni corrected significance threshold p=0.0008, OR: 16.97 [95% CI: 4.27-52.05], Table 3.3-16). Rare truncating variants were detected in 5/230 HCM cases and 10/7523 controls. One other truncating variant, p.Glu148fs, was detected in 1/230 HCM cases and 6/7253 controls and therefore excluded from case-control analyses. This variant was also detected at high frequency in the gnomAD database (Table 3.3-18).

ALPK3	Case Freq.	Control Freq.	FET p value	OR	Lower _CI	Upper CI
All variants	0.0565	0.0263	0.0116	2.2147	1.1394	3.9596
Missense	0.0348	0.0243	0.2788	1.45	0.61	2.97
Truncating	0.0217	0.0014	0.0001	16.07	4.27	52.05
Non-truncating	0.0348	0.0250	0.3869	1.41	0.59	2.88

Table 3.3-16 Summary of case-control analysis in *ALPK3*. FET=Fisher's exact test. OR=odds ratio. CI=95% confidence interval.

The evidence supporting the pathogenicity of heterozygous truncating variants detected in cases and controls was assessed (Table 3.3-18). No obvious difference in variant type or location was noted between cases and controls. All of the truncating variants detected in cases were classified as VUS, largely due to lack of evidence in the literature to support the pathogenicity of heterozygous loss of function variants in this gene.

The clinical features of individuals with heterozygous truncating variants in *ALPK3* are summarised in Table 3.3-17. One individual, who was heterozygous for the p.Gly223fs variant, had an atypical phenotype with features of both HCM and DCM; cardiac investigations undertaken in four first degree relatives were normal. None of the other

four cases in whom *ALPK3* truncating variants were detected had any unusual or atypical clinical features. In one family, the variant (c.2260-1G>A) was detected in the index case and an affected parent. None of the other cases had affected relatives, so segregation analyses were not possible.

ID	Variant	Age at first evaluation	LVH	FH of HCM or SCD	Other comments
013958	p.Lys403*	57	1.4	No	Apical/mild concentric hypertrophy.
009415	p.Gly223fs	68	2.9	Possible SCD	Severe concentric hypertrophy. Atypical, mixed HCM/DCM. RBBB. Possible hypertension.
007115	c.2260-1G>A	47	1.4	Yes	Mild HCM. Hypertension. Atypical type 1 diabetes
008833	p.Tyr1623*	47	2.0	No	Incidental diagnosis. Major hypertrophy.
013283	p.Ala604fs	47	1.7	No	No details

Table 3.3-17 Clinical details of individuals with truncating variants in *ALPK3*. Age at first evaluation=Age (in years)at first inherited cardiac clinic evaluation or first genetic test. LVH= Maximum left ventricular wall thickness (cm). FH =family history of HCM or sudden cardiac death (SCD)

c.	p.	Exon	Predicted effect	gnomAD	Case vs control	ClinVar
c.668delG	p.Gly223fs	1	This deletion creates a frame shift starting at codon Gly223. The new reading frame ends in a STOP codon 87	Not listed	Case (and affected first degree relative)	Not listed
c.1207A>T	p.Lys403*	5	Nonsense substitution.	Not listed	Case	Not listed
c.1810delG	p.Ala604fs	5	This deletion creates a frame shift starting at codon Ala604. The new reading frame ends in a STOP codon 13 positions downstream.	Not listed	Case	Not listed
c.2260-1G>A	.	Intron 5	This substitution is located in the acceptor splice site of intron 5. The consequence of this change is not predictable, but a skip of exon 6 is very likely.	Euro(NF):1/102714; Total:1/221606	Case	Not listed
c.4896C>G	p.Tyr1632*	10	Nonsense substitution.	Afr:1/24012; Euro(NF):1/126606; Total:2/277066	Case	Not listed
c.436_440dupCCAGG	p.Glu148fs	1	This duplication creates a frame shift starting at codon Glu148. The new reading frame ends in a STOP codon 7	Afr:169/9344 (2 homozygotes) ; Latino:7/8942; Other 1:2140; Euro(NF):1/29656; Total:178/67344. LoF Flag:Low coverage of region in ExAC, may indicate low-quality site.	1 Case : 6 Controls	Reported as Likely Pathogenic
c.334C>T	p.Gln112*	1	Nonsense substitution.	Not listed	Control	
c.1007delG	p.Arg336fs	4	This deletion creates a frame shift starting at codon Arg336. The new reading frame ends in a STOP codon 38 positions downstream.	Not listed	Control	Not listed
c.1699C>T	p.Gln567*	5	Nonsense substitution.	Euro(NF):6/110968; Total:6/244976	2 Controls	Not listed

c.	p.	Exon	Predicted effect	gnomAD	Case vs control	ClinVar
c.2650G>T	p.Glu884*	6	Nonsense substitution.	Euro(NF):4/111700; Total:4/246182	Control	Not listed
c.3061C>T	p.Arg1021*	6	Nonsense substitution.	Afr:1/15266; Euro(NF):1/111472; Total:2/245934	Control	Not listed
c.4819C>T	p.Arg1607*	10	Nonsense substitution.	SA:2/30778; Latino:1/33580; Total:3/246224	Control	Not listed
c.5714delG	p.Gly1905fs	14	This deletion creates a frame shift starting at codon Gly1905. The new reading frame ends in a STOP codon 52 positions downstream. Could escape NMD as canonical stop at p.1908	SA:1/30736; Total:1/239782	Control	Not listed
c.5721delG	p.Ter1908fs	14	This deletion creates a frame shift starting at codon *1908. The new reading frame ends in a STOP codon 49 positions downstream. Could escape NMD as canonical stop at p.1908	Euro(NF):3/124368; Total:3/267248	2 Controls	Not listed

Table 3.3-18 *ALPK3* putative loss of function variants detected in cases and controls. Variants annotated against genome build GRCh37, Ensembl transcript ID:ENST00000258888. c. =HGVS coding nucleotide, p. =HGVS protein nomenclature (http://www.hgvs.org/mutnomen/). Genome Aggregation database (gnomAD) https://gnomad.broadinstitute.org/. Euro(NF)=European (Non-Finnish) population. Afr=African population. SA=South Asian population. ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/).

3.3.5.4 DES

Investigations were also undertaken to explore a possible excess of missense variants in the *DES* gene in cases compared to controls (Table 3.3-19).

DES	Case with	Case without	Control with	Control Without	FET p-value	OR	Lower CI	Upper CI
All variants	6	224	36	7217	0.0016	5.367	1.8295	13.0371
Missense	6	224	32	7221	0.0010	6.04	2.04	14.84
Truncating	0	230	4	7249	0.1722	6.27	0.13	56.39
Non-truncating	6	224	32	7221	0.0010	6.04	2.04	14.84

Table 3.3-19 Summary of case-control analyses in *DES* gene. FET=Fisher's exact test. OR=odds ratio. CI=95% confidence interval. HCM cases n=230, reference controls n=7253.

A total of 6 rare missense variants were detected in the *DES* gene in 6/230 (2.6%) HCM cases. Four variants were classified as VUS (Table 3.3-20) and 2 variants (p.Ser298Leu and p.Arg429Gln) were classified as "Likely benign" on the basis of their frequency in gnomAD and BRRD non-HCM reference controls. One of these variants (p.Ser298Leu) had previously been reported in the ClinVar database as a VUS/Likely pathogenic variant. Of the four variants classified as VUS, two (p.Glu131Lys and p.Arg278Trp) were listed in the ClinVar database as VUS, however no information regarding phenotype was provided for these cases and segregation studies were not reported. Two variants (p.Gln131Lys and p.Arg278Trp) were located within the functionally important Coil1A and Coil 1B regions of the protein; however none were in the Coil 2B domain, where the majority of the previously published pathogenic variants have been reported.

The individual with the p.Asp455Val variant also had a VUS in the *BAG3* gene (p.Ala160Val) and the *MRPL3* (p.Ala342Val) gene. The presence of three apparently novel variants in this individual likely reflects their ethnicity. Principal Component Analysis (PCA) performed by the BRRD bioinformatics core team (see section 3.2.6), annotated this individual's ethnicity as "Other". Therefore, variants detected in

individuals of this ethnicity may be under-represented in reference cohorts and so be more likely to survive population MAF filtering.

Variants were distributed across all regions of the protein in cases and controls, however the majority of the rare variants detected in controls appear to lie in the head domain of the protein (Figure 3.3-18). Analysis of amino acid conservation across the protein indicates that the head domain is one of the least conserved regions and as such would be considered to be more tolerant of variation (Figure 3.3-19). Visualisation of the density plot indicates that rare variants in HCM cases and reference controls may be differentially distributed (Figure 3.3-17), however this was not supported by Kolmogorov-Smirnov analysis (Table 3.3-15). Unfortunately, the small size of the case cohort precludes these comparative analyses.

In the BRRD reference controls, 32 rare *DES* gene variants were detected in 36/7253 (0.49%) individuals (Table 3.3-21). This is comparable to the frequency found in ExAC (0.47%).

ID	Ethnicity	control n	c.	p.	Domain	ClinVar	gnomAD	*In silico* pathogenicity predictions	Class
014495	Euro	0	c.43C>A	p.Arg15Ser	Head	Not listed	Not detected	Mixed	VUS
011073	Euro	1	c.391C>A	p.Gln131Lys	Coil 1A	VUS	*NB. Variant not PASS filter.* Euro(NF):3/79992 (MAF 0.000037), Total: 3/187074 (MAF 0.000016).	Majority predict deleterious	VUS
011074	Euro	0	c.832C>T	p.Arg278Trp	Coil 2A	VUS	SA:1/30776(MAF 0.00003249) Total :1/245308 (MAF 0.000004)	Majority predict deleterious	VUS
008841	Euro	0	c.893C>T	p.Ser298Leu	Linker 2	Conflicting (LP x2, VUS x2)	Afr:6/23998 (MAF 0.00025) Other:1/6458 (MAF 0.00015) Euro(NF):14/126510 (MAF 0.0001) Latino:1/34416 (MAF 0.000029) Total:22/276948 (MAF 0.000079)	Mixed	LB
012462	Euro	3(unrelated)	c.1286G>A	p.Arg429Gln	Tail	Not listed	Other:1/5486 (MAF 0.00018) SA:3/30782 (MAF 0.000097) Euro(Fin):2/22294 (MAF 0.000089) Euro(NF):7/111712 (MAF 0.000062) EA:1/17248 (MAF 0.000057) Tot:14/246254 (MAF 0.000056)	Majority predict deleterious	LB
014508	Other	0	c.1364A>T	p.Asp455Val	Tail	Not listed	Not detected	Majority predict deleterious	VUS

Table 3.3-20 Rare variants found in *DES* gene in BRRD HCM cases. Variants annotated against genome build GRCh37, Ensembl transcript ID:ENST00000373960. Euro=European. c. =HGVS coding nucleotide, p. =HGVS protein nomenclature (http://www.hgvs.org/mutnomen/). Domain=protein domain. ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/). MAF=minor allele frequency. Euro(NF)=European (Non-Finnish) population. Euro(Fin)=Finnish population. EA=East Asian population. Afr=African population. SA=South Asian population. *In silico* pathogenicity predictions =range of tools used to predict consequences of variants. Class=assigned classification, VUS=variant of uncertain significance. LB=likely benign.

BRRD control n	Ethnicity	BRRD HCM n	c.	p.	Domain	ClinVar	gnomAD
1	Other	0	c.31G>T	p.Val11Leu	Head	Not listed	Not detected
2	EUR	0	c.40T>C	p.Tyr14His	Head	Not listed	Euro(NF): 4/114372 (MAF 0.00003497), Total:4/255182 (MAF 0.00001568)
1	EUR	0	c.50C>A	p.Thr17Asn	Head	Not listed	Not detected
1	EUR	0	c.68G>A	p.Gly23Asp	Head	Not listed	Not detected
1	AFR	0	c.155G>A	p.Arg52His	Head	Not listed	Not detected
1	EUR	0	c.162C>G	p.Tyr54*	Head	Not listed	Not detected
1	SAS	0	c.199G>A	p.Gly67Arg	Head	Not listed	Other:1/4468 (MAF 0.0002238), SA:2/26246 (MAF 0.00007620), Total:3/190606 (MAF 0.00001574)
1	EUR	0	c.206T>C	p.Leu69Pro	Head	Not listed	Euro(NF):2/79558 (MAF 0.00002514) , Total:2/18746 (MAF0.00001067)
1	AFR	0	c.218G>A	p.Arg73Gln	Head	VUS	Not detected
1	Other	0	c.221T>A	p.Leu74Gln	Head	Not listed	Not detected
1	EUR	0	c.226delA	p.Thr76fs	Head	Not listed	Euro(NF):1/71868 (MAF 0.00001391), Total:1/174330 (MAF 0.000005736)
1	EUR	0	c.250G>A	p.Gly84Ser	Head	VUS	Euro(NF):19/81384 (MAF 0.0002335) Other:1/5056 (MAF 0.0001978) Euro(Fin):2/19478 (MAF 0.0001027) Afr:1/17908 (MAF 0.00005584) Total:23/196058 (MAF 0.0001173)
1	EUR	0	c.298G>A	p.Glu100Lys	Head	Not listed	Not detected
1	EUR	0	c.299A>C	p.Glu100Ala	Head	VUS	Afr:1/17678 (MAF 0.00005657) Euro(NF):2/80692 (MAF 0.00002479) Total:3/194520 (MAF 0.00001542)
1	EUR	1	c.391C>A	p.Gln131Lys	Coil 1A	VUS	Not detected
1	SAS	0	c.418G>A	p.Val140Met	Coil 1A	Not listed	Not detected
1	EUR	0	c.566G>A	p.Arg189Gln	Coil 1B	VUS	Not detected
1	EAS	0	c.593T>C	p.Ile198Thr	Coil 1B	Not listed	EA:4/17248 (0.0002319) Total:4/246260 (MAF 0.00001624)
1	AFR	0	c.643G>A	p.Val215Met	Coil 1B	VUS	Afr:10/24036 (MAF 0.0004160) Latino:2/34420 (MAF 0.00005811) EA:1/188868 (MAF 0.000053) Euro (NF):2/126694 (MAF 0.00001579) Total:15/277214 (MAF 0.00005411)
1	Other	0	c.709G>A	p.Ala237Thr	Coil 1B	VUS	Euro(Fin):6/25734(0.0002332) Other:1/6466 (MAF 0.0001547) Euro(NF):12/126696 (MAF 0.00009471)

BRRD control n	Ethnicity	BRRD HCM n	c.	p.	Domain	ClinVar	gnomAD
							Total:19/277144 (MAF 0.00006856)
1	EUR	0	c.710C>T	p.Ala237Val	Coil 1B	Not listed	Euro(NF):1/111702 (MAF 0.000008952) Total:1/246192 (MAF 0.000004062)
1	EUR	0	c.823A>G	p.Arg275Gly	Coil 2A	Not listed	Not detected
1	EUR	0	c.944G>A	p.Arg315His	Coil 2B	VUS	EA:2/17244 (MAF 0.0001160) Euro(NF):8/111638 (MAF 0.00007166) Total:10/245908 (MAF 0.00004067)
1	EUR	0	c.1064G>A	p.Arg355Gln	Coil 2B	VUS	Other:1/6462 (MAF 0.0001548) Euro(NF):8/126702 (MAF 0.00006314) Total:9/277206 (MAF 0.00003247)
1	EUR	0	c.1099A>G	p.Ile367Val	Coil 2B	Not listed	Euro(NF):1/111706 (MAF 0.000008952), Total:1/246248 (MAF 0.000004061)
1	EUR	0	c.1142T>C	p.Met381Thr	Coil 2B	Not listed	EA:1/17246 (MAF 0.00005798), SA:1/30782 (MAF 0.0003249) Total:2/246260 (MAF 0.000008121)
1	Other	0	c.1235_1236dupAG	p.Glu413fs	Tail	Not listed	Not detected
3	EUR	0	c.1243C>T	p.Arg415Trp	Tail	VUS	EA:1/18862 (MAF 0.00005302) Afr:1/23826 (MAF 0.00004197) Euro (Fin):1/25732 (MAF 0.00003886), Euro(NonFin):3/126038 (MAF 0.00002380), Total:6/276182 (MAF 0.00002172)
3	SAS, EUR	1	c.1286G>A	p.Arg429Gln	Tail	VUS	Other:1/5486 (MAF 0.0001823) SA:3/30782 (MAF 0. 0.00009746) Euro(Fin):2/22294(MAF 0.00008971) Euro(NonFin):7/111712 (MAF 0.00006266) EA:1/17248 (MAF 0.00005798) Total:14/246254 (MAF 0.00005685)
1	EUR	0	c.1361G>A	p.Arg454Gln	Tail	VUS	Not detected
1	Other	0	c.1371+1G>A	c.1371+1G>A	Tail	LPx2(GenDX and Invitea), VUS(Stanford)	AshJewish:7/9454 (MAF 0.0007404) Other:1/5242 Total:8/230782 (MAF 0.00003466)

Table 3.3-21 Rare variants found in *DES* gene in BRRD reference cohorts. Variants annotated against genome build GRCh37, Ensembl transcript ID:ENST00000373960. Euro=European. c. =HGVS coding nucleotide. p. =HGVS protein nomenclature (http://www.hgvs.org/mutnomen/) . Domain=protein domain. ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/). Euro(NF)=European (Non-Finnish) population. Euro(Fin)=Finnish population. EA=East Asian population. Afr=African population. SA=South Asian population. AshJewish =Ashkenazi Jewish population. VUS=variant of uncertain significance. LP=likely pathogenic.

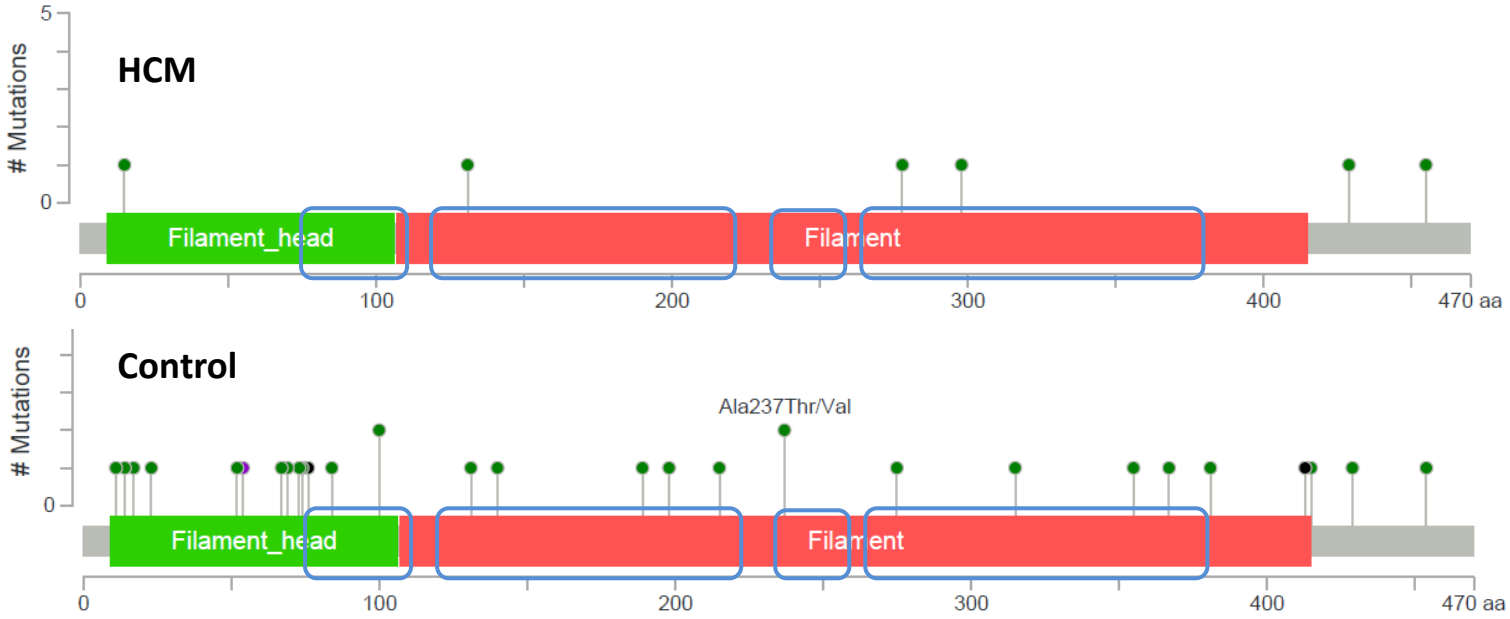


Figure 3.3-18 Distribution of BRRD HCM and control variants across DES gene. Blue boxes delineate key functional domains where the majority of pathogenic variants in other disorders have been reported previously. Plot generated at http://www.cbioportal.org/mutation_mapper.jsp.

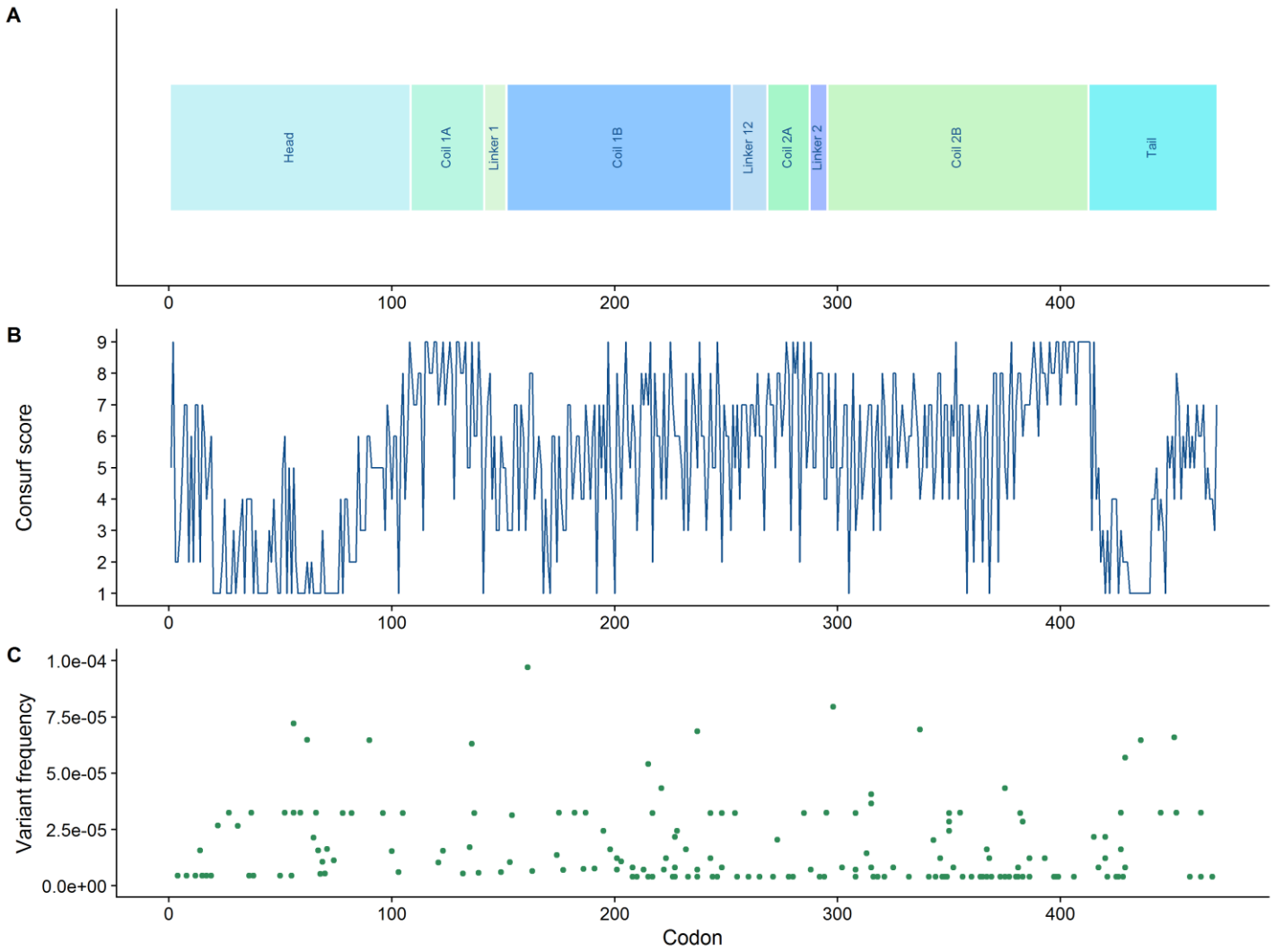


Figure 3.3-19 *DES* A. Protein domains (http://www.uniprot.org/uniprot/P17661). B. Consurf amino acid conservation scores (http://consurf.tau.ac.il/2016/overview.php). C. Genome Aggregation Database (gnomAD) missense variant frequency and distribution across protein (http://gnomad.broadinstitute.org/gene/ENSG00000175084).

ID	Variant	Age at first evaluation	LVH (cm)	FH	Clinical details
014495	p.Arg15Ser (Head)	54	1.5	No	Asymmetric septal hypertrophy Hypertension. No extra cardiac features noted.
011073	p.Gln131Lys (Coil 1A)	43	2.2	No	Asymmetric septal hypertrophy No extra cardiac features noted.
011074	p.Arg278Trp (Coil 2A)	58	1.6	Yes	Asymmetric septal hypertrophy No extra cardiac features noted.
014508	p.Asp455Val (Tail)	20	1.7	Yes	Asymmetric septal hypertrophy No extra cardiac features noted.

Table 3.3-22 Clinical details of individuals with VUS in *DES*. Age at first evaluation=Age (in years) at first inherited cardiac clinic evaluation or first genetic test. LVH= Maximum left ventricular wall thickness (cm). FH =family history of HCM or sudden cardiac death (SCD).

Review of clinical data in cases with *DES* missense variants did not reveal any evidence of skeletal muscle or cardiac conduction disease suggestive of desminopathy. In 2 out of 4 cases with VUS, there was a documented family history of HCM and/or SCD; however, no other affected family members were available for segregation analysis.

3.3.5.5 KCNQ1

The results of case-control analysis in the *KCNQ1* gene were also explored. Although higher than the conservative Bonferroni corrected p-value (p=0.0008), this was the second lowest p-value observed in these analyses (Table 3.3-23).

KCNQ1	Case with	Case without	Control with	Control without	FET p value	OR	Lower CI	Upper CI
All variants	6	224	51	7202	0.0077	3.7816	1.3122	8.9405
Missense	5	225	43	7210	0.0154	3.73	1.14	9.51
Truncating	1	229	7	7246	0.2211	4.52	0.10	35.38
Non-truncating	5	225	44	7209	0.0167	3.64	1.12	9.28

Table 3.3-23 Summary of case-control analyses in *KCNQ1* gene. FET=Fisher's exact test. OR=odds ratio. CI=95% confidence interval. HCM cases n=230, reference controls n=7253.

All six *KCNQ1* variants detected in the BRRD HCM cohort were classified as VUS. One variant (p.Val241Ile) lies within a functionally important transmembrane domain of the protein[223]. These regions are highly conserved and appear to be under greater constraint, and therefore less tolerant of variation than other domains[223,224] (Figure 3.3-20 and Figure 3.3-21).

Analysis of these six variants using a recently published tool, Q1 VarPred, which has been validated for pathogenicity predictions in *KCNQ1*[225], gave a high probability of pathogenicity for two variants p.Ser92Cys (probability of pathogencity=0.934) and p.Val241Ile (probability of pathogenicity=0.80).

No evidence was found to indicate that there was a difference in variant distribution between cases and controls (Table 3.3-15, Figure 3.3-17); however, limited conclusions can be drawn here given the small number of variants found in cases.

Review of the clinical records from the six HCM cases heterozygous for rare *KCNQ1* variants, revealed evidence of QT prolongation in one individual. There was no evidence of prolonged QT or Long QT syndrome in the other five cases (Table 3.3-25).

ID	Ethnicity	control n	c.	p.	Domain	ClinVar	gnomAD	*In silico* pathogenicity predictions	Q1 VarPred	Class
008467	Euro	1 ("Other")	c.41G>T	p.Arg14Leu	Cytoplasmic	Not listed	Not detected	Mixed	0.410	VUS
014519	Euro	0	c.275C>G	p.Ser92Cys	Cytoplasmic	Not listed	Euro(NF):1/99284 (MAF 0.00001) Total:1/220230 (MAF 0.0000045)	Mixed	0.934	VUS
011376	Euro	0	c.352A>G	p.Thr118Ala	Cytoplasmic	Not listed	Not detected	Mixed	0.564	VUS
006250	Euro	0	c.721G>A	p.Val241Ile	Transmembrane (S4)	VUS (single submission)	Other:1/5456 (MAF 0.00018), Latino:1/33530 (MAF 0.000029), Euro(NF):1/110846 (MAF 0.000009) Total:3/244640 (MAF 0.000001)	Mixed	0.800	VUS
006264	Euro	0	c.1252-3dupT	.	Cytoplasmic	Not listed	Not detected	Alamut-not predicted to affect splicing	n/a	VUS
008450	Euro	0	c.1747C>T	p.Arg583Cys	Cytoplasmic*	Pathogenic, risk factor (5 submissions, single submitter provided criteria)	SA 1/30782 (MAF 0.000032), Euro(NF) 3/111624 (MAF 0.000026) Total:4/245928 (MAF 0.000016)	Mixed	0.093	VUS

Table 3.3-24 *KCNQ1* rare variants found in BRRD HCM cases (n=230). Variants annotated against genome build GRCh37, Ensembl transcript ID:ENST00000155840. Euro=European. c. =HGVS coding nucleotide. p. =HGVS protein nomenclature (http://www.hgvs.org/mutnomen/). ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/). Euro(NF)=European (Non-Finnish) population. In silico pathogenicity predictions =range of tools used to predict consequences of variants. Class=assigned classification VUS=variant of uncertain significance.
*This variant does not lie within one of the highly conserved regions within C-terminus of protein (aa 349-391, 509-575, 585-607), which have been reported to strongly correspond with putative case mutations. [224]. Q1 VarPred Score=tool prediction score "Probability of deleterious", but note, the positive predictive value (PPV) of this score will be lower in individuals with low pre-test probability of LQTS as in these cases [225].

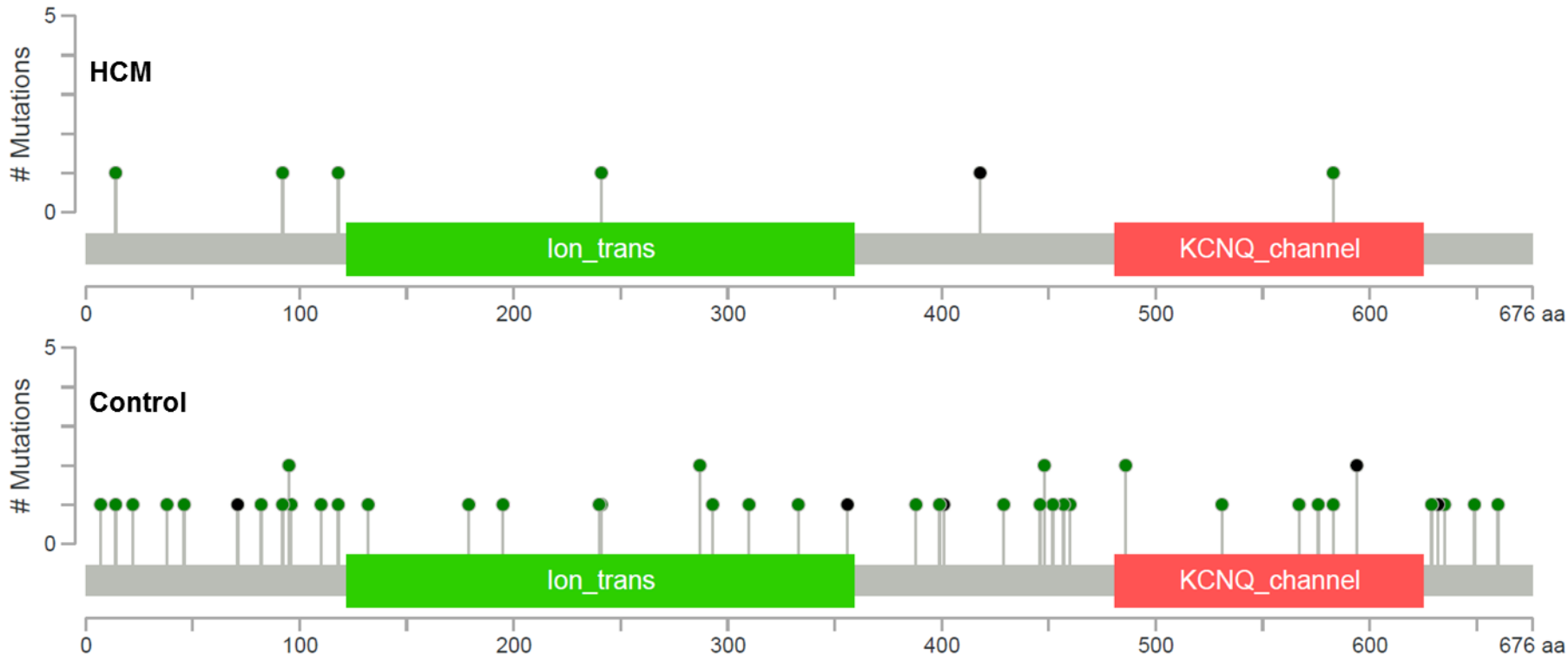


Figure 3.3-20 Distribution of variants in HCM cases and Non-HCM BRRD controls across the *KCNQ1* gene. Green=Missense, Black=Truncating. Plot generated at http://www.cbioportal.org/mutation_mapper.jsp.

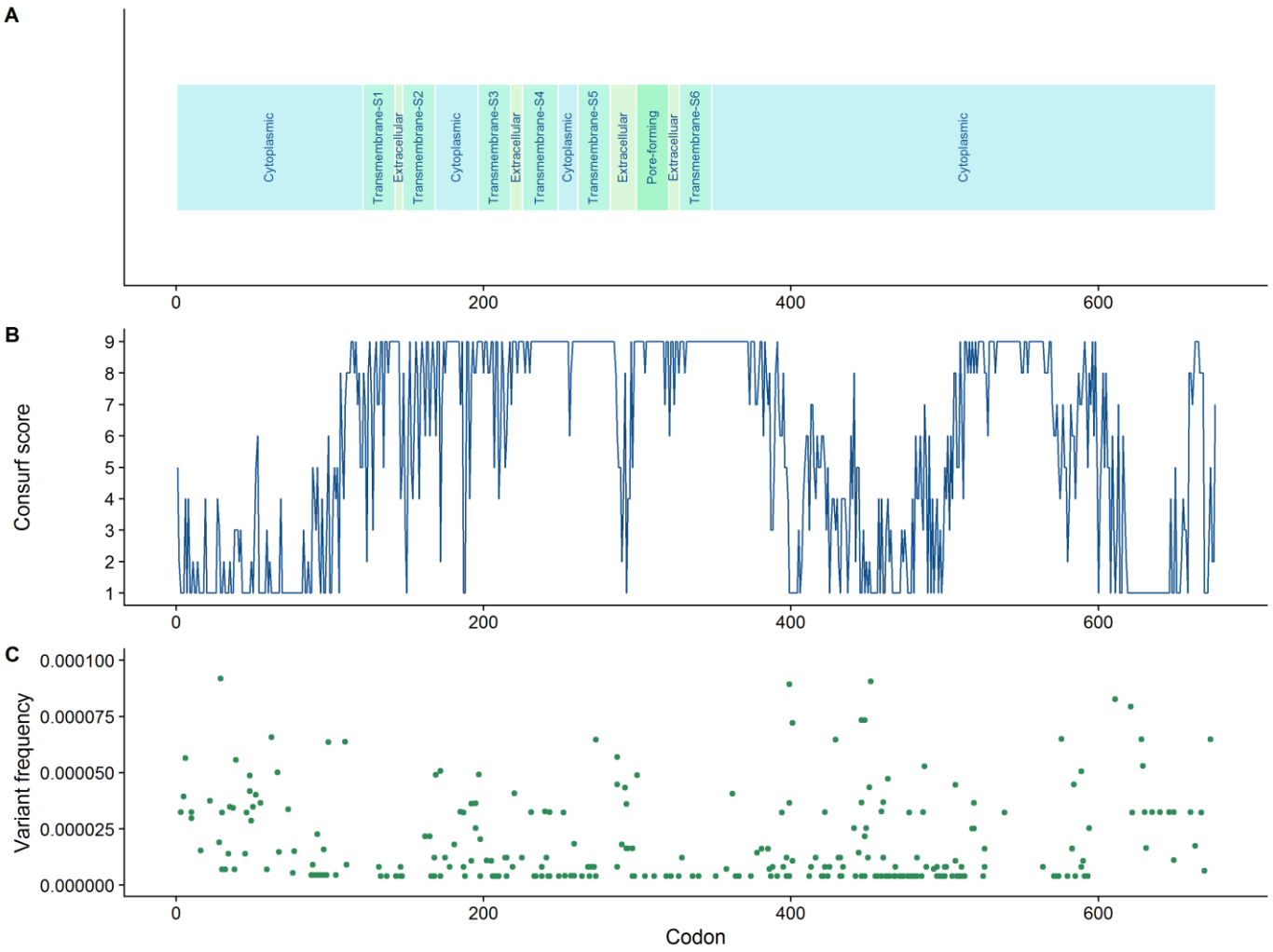


Figure 3.3-21 *KCNQ1* A. Protein domains (http://www.uniprot.org/uniprot/P51787). B. Consurf amino acid conservation scores (http://consurf.tau.ac.il/2016/overview.php). C. gnomAD missense variant frequency and distribution across protein (http://gnomad.broadinstitute.org/gene/ENSG0000005391).

ID	Variant	Age at first evaluation	LVH	FH of HCM or SCD	Clinical details
008467	p.Arg14Leu	56	2.3	No	Asymmetric septal hypertrophy Hypertension. "ECG unremarkable" (QTc not available)
014519	p.Ser92Cys	73	1.7	No	Asymmetric septal hypertrophy Left bundle branch block Anonymous 'legacy' sample-no further details available.
011376	p.Thr118Ala	46	1.5	Yes	Asymmetric septal hypertrophy Hypertension. Maternal uncle HCM. QTc=428
006250	p.Val241Ile	67	1.6	Yes	Asymmetric septal hypertrophy Hypertension. Atrial fibrillation Brother HCM. QTc=433
006264	c.1252-3dupT	67	2.4	No	Asymmetric septal hypertrophy Left ventricular outflow tract obstruction QTc=490
008450	p.Arg583Cys	70	1.6	No	Asymmetric septal hypertrophy Hypertension Qtc=443

Table 3.3-25 Clinical details of individuals with VUS in *KCNQ1*. Clinical details were obtained from existing clinical records. Age at first evaluation=Age (in years) at first inherited cardiac clinic evaluation or first genetic test. LVH= Maximum left ventricular wall thickness (cm). FH =family history of HCM or Sudden Death.

3.3.5.6 TCAP

A single truncating variant was found in this gene within the HCM cohort, c.66G>A (p.Trp22*), (Table 3.3-28) and within reference controls only 15 rare variants (all missense) were detected. Although the Fisher's exact test p value for truncating variants (p=0.0028) did not meet the Bonferroni corrected threshold for significance (p=0.0008) (Table 3.3-26), given the apparent rarity of truncating variants in this gene, this finding was considered worthy of further investigation.

TCAP is a small gene with only 2 exons encoding 167 amino acids (Figure 3.3-9). Data from ExAC indicates that this gene is not under constraint for missense or loss-of-function variants (Table 3.3-27).

TCAP	Case with	Case without	Control with	Control Without	FET p value	OR	Lower_ CI	Upper CI
All variants	1	229	15	7238	0.3935	2.1068	0.0498	13.8036
Missense	0	230	15	7238	0.4147	1.96	0.05	12.71
Truncating	1	229	0	7253	0.0028	62.83	3.26	3610.86
Non-truncating	0	230	15	7238	0.4147	1.96	0.05	12.71

Table 3.3-26 Summary of *TCAP* Fisher's exact test (FET) and odds ratios (OR) with 95% confidence intervals (CI).

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	26.7	27	z = -0.04
Missense	70.8	63	z = 0.45
LoF	4.0	2	pLI = 0.08
CNV	3.0	2	z = 0.21

Table 3.3-27 Exome aggregation consortium (ExAC) constraint scores for *TCAP* gene. (http://exac.broadinstitute.org/). Missense: For synonymous and missense, Z score represents the deviation of observed counts from the expected number. Positive Z scores indicate increased constraint (intolerance to variation) and therefore that the gene had fewer variants than expected. Negative Z scores are given to genes that had a more variants than expected. LoF (loss of function): The closer pLI is to one, the more LoF intolerant the gene appears to be. pLI >= 0.9 as an extremely LoF intolerant set of genes. [138]

Review of the clinical records from the HCM case in whom the truncating *TCAP* variant was detected revealed no atypical signs or symptoms and no family history of HCM or sudden death (SCD) was reported (Table 3.3-29).

ID	Ethnicity	c.	p.	ClinVar	gnomAD	Class
010414	Euro	c.66G>A	p.Trp22*	Likely pathogenic-Single submitter, no evidence.	Not detected	VUS

Table 3.3-28 Rare variant detected in *TCAP*. Variants annotated against genome build GRCh37, Ensembl transcript ID: ENST00000309889. Euro=European. c. =HGVS coding nucleotide. p. =HGVS protein nomenclature (http://www.hgvs.org/mutnomen/). ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/). Class=assigned classification. VUS=variant of uncertain significance.

ID	Variant	Age at first evaluation	LVH	FH	Clinical details
010414	**p.Trp22****	**61**	**2.1**	**No**	**Asymmetric septal hypertrophy**

Table 3.3-29 Clinical details Age at first evaluation=Age (in years) at first inherited cardiac clinic evaluation or first genetic test. LVH= Maximum left ventricular wall thickness (cm). FH =family history of HCM or Sudden Death.

3.3.6 Variant level analyses

To determine whether there was any evidence to support the pathogenicity of any individual variant, each was reviewed and classified according to clinical guidelines[128]. A total of 256 rare variants were detected in 43 of the 60 genes analysed. One variant was classified as likely pathogenic (LP) (0.4%) and 12 (5%) as benign or likely benign (B/LB). The vast majority of variants (243/256, 95%) were classified as variant of uncertain significance (VUS). Approximately one third of these VUS were in the *TTN* gene (86/243, 35%).

The variant classified as likely pathogenic was a splice site variant in the *FLNC* gene, c.6004+2T>C. This variant was absent from reference control samples (ExAC, gnomAD and BRRD) and *in silico* analyses predicted that this nucleotide substitution would disrupt the constitutive splice donor site, causing a frameshift of the amino acid sequence and premature termination of translation (Figure 3.3-22). Therefore, this variant was considered to meet the ACMG criteria for likely pathogenic classification (PVS1 and PM2)[5].

3.3.7 Case level analyses

To assess the clinical utility of extended gene analyses in this HCM cohort, the outcomes of testing were considered for each case. Across the 230 cases screened, a single clinically actionable variant, a splice donor site variant in *FLNC* (c.6004+2TC), was detected.

The *FLNC* c.6004+2TC variant was detected in an individual who was diagnosed with HCM at approximately 25 years of age; this patient remains asymptomatic with a maximum LV wall measurement of 1.7cm and good left ventricular systolic function at age 40. A raised CK level was reported at diagnosis; however subsequent CK measurements were normal. His father, who is now deceased, was diagnosed with restrictive cardiomyopathy (RCM) at 40 years of age and developed progressive heart failure requiring cardiac transplantation at age 49. The variant was present in DNA from the father, however, further segregation analyses were not possible as no additional relatives were known to be affected.

[5] **PVS1=** null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease.
In this case, this criteria was downgraded to "moderate" as a the time of reporting there was limited literature supporting this class of variant in this gene and no *in vitro* evidence to support consequences of splice site variant.
PM2=Absent from controls.

Overall, high levels of uncertain findings were found in this cohort, with 66% (151/230) of cases reported to have at least one VUS (Table 3.3-30 and Figure 3.3-23). In 30% of cases (70/230) more than one VUS was detected, with a maximum number of five VUS found in a single case. Excluding the *TTN* gene from analyses, the proportion of cases with at least one VUS was 51% (117/230), with 19% (44/230) having more than one VUS (Figure 3.3-23).

Analysis of the genes currently included on the OMGL HCM gene panel confirmed prior findings and did not detect any additional variants. Since initial clinical sequencing, one variant (*MYH7* p.Arg1712Gln) had been reclassified from VUS to likely pathogenic.



Figure 3.3-22 *FLNC* c.6004+2T>C *In silico* splice site analysis. Alamut Interactive Biosoftware. Details of splice site algorithms used available from http://www.interactive-biosoftware.com/doc/alamut-visual/2.7/splicing.html.

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
006226	55	No	European		OBSCN c.15267_15269dupGGC(p.Ala5090dup)[VUS] OBSCN c.15274G>A(p.Asp5092Asn)[VUS]
006227	48	No	European		
006228	64	Yes	European		
006229	41	No	European		FLNC: c.7205C>A(p.Ala2402Asp) [VUS] FXN: c.136_146dupATCGATGCGAC(p.Cys50fs) [VUS] TTN: c.88189_88191delATC(p.Ile29397del) [VUS]
006231	69	Yes	European		TTN: c.4532G>A(p.Gly1511Asp) [VUS] DSP c.4196C>T(p.Thr1399Ile)[VUS]
006232	63	No	European		TTN: c.14141G>A(p.Gly4714Asp) [VUS]
006233	68	No	European		
006234	52	Yes	European		TTN: c.54581G>A(p.Gly18194Asp) [VUS] DSP c.3146C>T(p.Ser1049Leu)[VUS]
006235	70	No	European	MYBPC3: c.3005G>A(p.Arg1002Gln) [VUS]	OBSCN c.22703C>T(p.Ala7568Val)[VUS]
006236	45	No	European	MYBPC3: c.3739G>A(p.Asp1247Asn) [VUS]	MYH6: c.4685G>T(p.Arg1562Leu) [VUS]
006237	45	No	European		
006238	52	No	European		LDB3: c.1633A>G(p.Ser545Gly) [VUS]
006239	48	No	European		TTN: c.23308G>A(p.Gly7770Arg) [VUS]
006240	64	Yes	European		BRAF: c.550G>T(p.Ala184Ser) [VUS] VCL: c.2285G>A(p.Arg762Gln) [VUS]
006241	51	No	European		GLB1: c.113G>A(p.Arg38Gln) [VUS]
006242	34	Yes	European		
006243	59	No	European		

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
006244	45	No	European		
006245	62	No	European		TNNC1: c.23C>T(p.Ala8Val) [VUS]
006246	31	No	European		
006247	66	No	European		TTN: c.101213G>A(p.Arg33738His) [VUS] OBSCN c.15064+2T>C(.)[VUS]
006248	37	No	European		
006249	53	No	European		OBSCN c.19631A>G(p.Glu6544Gly)[VUS]
006250	67	Yes	European	GLA: c.1028C>T(p.Pro343Leu) [VUS]	KCNQ1: c.721G>A(p.Val241Ile) [VUS] TTN: c.42236G>A(p.Arg14079Gln) [VUS]
006251	53	Yes	European		CALR3: c.403G>A(p.Asp135Asn) [VUS] MAP2K2: c.403G>A(p.Gly135Arg) [VUS]
006252	67	No	European		CRYAB: c.10G>A(p.Ala4Thr) [VUS]
006253	58	Yes	European		TTN: c.103216G>A(p.Gly34406Arg) [VUS]
006254	44	Yes	European		
006255	67	No	African		
006256	60	No	European		TTN: c.103724G>A(p.Arg34575His) [VUS]
006257	38	No	European		
006260	34	Yes	European		
006261	25	Yes	European		TTN: c.39343G>A(p.Val13115Ile) [VUS]
006263	66	Yes	European		OBSCN c.4685A>C(p.Gln1562Pro)[VUS] OBSCN c.4686G>T(p.Gln1562His)[VUS]
006264	67	No	European		KCNQ1: c.1252-3dupT [VUS] TTN: c.97442G>A(p.Gly32481Glu) [VUS]
006265	52	No	East-Asian	CSRP3: c.565_567delCAA(p.Gln189del) [VUS]	CACNA1C: c.5975G>T(p.Cys1992Phe) [VUS] TTN: c.5423A>G(p.Glu1808Gly) [VUS] TTN: c.102950A>G(p.Gln34317Arg) [VUS]
006267	51	No	European		TTN: c.19801C>T(p.Pro6601Ser) [VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
006270	49	Yes	European		TTN: c.45248G>A(p.Arg15083Gln) [VUS]
006275	61	No	European		TTN: c.29317G>A(p.Ala9773Thr) [VUS]
006276	48	No	European		TTN: c.10391A>C(p.Lys3464Thr) [VUS]
006279	40	No	European		
006280	45	Yes	European		
006424	-	-	European		
007105	67	No	European		
007107	33	Yes	European		**FLNC: c.6004+2T>C [Likely pathogenic]** TTN: c.66509G>T(p.Arg22170Leu) [VUS]
007108	56	No	Other		SCO2: c.157C>T(p.Gln53*) [Benign]
007109	60	Yes	European		
007110	51	Yes	European		
007111	-	-	European		TTN: c.25104C>A(p.Asp8368Glu) [VUS]
007114	46	Yes	European	ACTN2: c.1295C>T(p.Ala432Val) [VUS]	TTN: c.16010A>G(p.Asn5337Ser) [VUS] TTN: c.86120G>A(p.Gly28707Glu) [VUS]
007115	41	Yes	European		TTN: c.71912G>T(p.Trp23971Leu) [VUS] TTN: c.96172C>T(p.Arg32058Trp) [VUS] ALPK3 c.2260-1G>A(.)[VUS]
007116	63	No	European		SOS1: c.3709C>G(p.Pro1237Ala) [VUS]
007117	67	No	European	ANKRD1: c.300G>C(p.Arg100Ser) [VUS]	MYH6: c.431G>A(p.Gly144Asp) [VUS] MYOM1: c.1280G>A(p.Arg427His) [VUS]
007118	58	Yes	European		MYOM1: c.3362C>T(p.Ala1121Val) [VUS] TTN: c.2573C>G(p.Ser858Trp) [VUS]
007119	55	Yes	European		FLNC: c.6790G>A(p.Ala2264Thr) [VUS] SOS1: c.3286T>A(p.Ser1096Thr) [VUS]
007120	60	No	European		ALPK3 c.992C>T(p.Thr331Ile)[VUS]
007121	42	No	European		TTN: c.36364+2T>C [VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
007128	57	No	European		OBSCN c.12112G>A(p.Val4038Met)[VUS]
007129	43	No	European		DSP c.7481A>G(p.Lys2494Arg)[VUS]
007134	67	Yes	European		MYLK2: c.134C>T(p.Pro45Leu) [Benign]
007135	40	No	European		TTN: c.32858C>T(p.Ala10953Val) [VUS]
007136	66	No	European		CASQ2: c.479G>A(p.Arg160His) [VUS]
007139	59	No	European		
007144	32	No	African		CALR3: c.1013delG(p.Gly338fs) [VUS] FLNC: c.7862G>A(p.Arg2621Gln) [Benign] MYOM1: c.4243A>G(p.Ile1415Val) [VUS] TTN: c.95683A>G(p.Ser31895Gly) [VUS]
007146	68	No	European		TTN: c.8902C>A(p.Pro2968Thr) [VUS]
008447	23	Yes	European		
008449	69	No	European		
008450	69	No	European		KCNQ1: c.1747C>T(p.Arg583Cys) [VUS]
008451	60	No	European		FLNC: c.2650G>T(p.Val884Phe) [VUS] TTN: c.16555A>T(p.Thr5519Ser) [VUS] TTN: c.71635G>A(p.Ala23879Thr) [VUS] TTN: c.75527G>A(p.Arg25176His) [VUS]
008452	68	Yes	European		
008453	56	No	European		TTN: c.84766C>T(p.Pro28256Ser) [VUS]
008454	69	No	European		NEXN: c.1925C>G(p.Pro642Arg) [VUS]
008455	38	Yes	European		OBSCN c.7882C>A(p.Arg2628Ser)[VUS]
008461	39	No	African		OBSCN c.14632G>T(p.Val4878Phe)[VUS] TTN: c.12404A>G(p.Asn4135Ser) [Likely benign]
008466	37	Yes	European		

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
008467	58	No	European		FOXRED1: c.23A>G(p.His8Arg) [VUS] KCNQ1: c.41G>T(p.Arg14Leu) [VUS] TTN: c.65047C>A(p.Pro21683Thr) [VUS]
008468	48	Yes	European		
008469	61	Yes	European		
008470	75	Yes	European		MYO6: c.139C>A(p.Gln47Lys) [VUS]
					ALPK3 c.4546C>T(p.Arg1516Cys)[VUS]
008822	66	No	South-Asian	ANKRD1: c.196C>T(p.Arg66*) [VUS]	LDB3: c.2026G>A(p.Val676Met) [VUS]
008824	57	No	East-Asian	MYH7: c.5704G>C(p.Glu1902Gln) [VUS]	CACNA1C: c.5862_5873del(p.Pro1955_Thr1958del) [VUS] GAA: c.1087A>T(p.Met363Leu) [VUS] GUSB: c.961A>G(p.Thr321Ala) [VUS] TTN: c.93092G>A(p.Gly31031Glu) [VUS]
008827	60	No	European		
008833	62	No	European		SOS1: c.3772C>A(p.Pro1258Thr) [VUS] OBSCN c.15985_15990delACCTGG(p.Thr5329_Trp5330del)[VUS]
008834	56	Yes	European		
008841	66	No	European		DES: c.893C>T(p.Ser298Leu) [VUS]
008842	51	No	European		GUSB: c.313G>A(p.Val105Met) [VUS] TTN: c.35384A>G(p.Lys11795Arg) [VUS] TTN: c.75997G>A(p.Gly25333Ser) [VUS]
008843	65	No	European		
008846	54	Yes	Other	TNNT2: c.805A>G(p.Asn269Asp) [VUS]	TTN: c.9674A>G(p.Asn3225Ser) [VUS]
008854	70	Yes	European		
008855	48	No	European		
009319	54	Yes	South-Asian		GAA: c.794G>A(p.Ser265Asn) [VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
009325	31	No	South-Asian		FLNC: c.6682C>T(p.Arg2228Trp) [VUS] MYPN: c.3042T>G(p.Asp1014Glu) [VUS] DSP c.424C>T(p.Leu142Phe)[VUS]
009327	57	No	European	TNNT2: c.785A>G(p.Asn262Ser) [VUS]	OBSCN c.7427C>T(p.Ser2476Leu)[VUS]
009328	49	No	European		
009329	49	No	European		TTN: c.102155G>A(p.Arg34052Gln) [VUS] OBSCN c.3466C>T(p.Arg1156Cys)[VUS]
009336	51	No	European		AGL: c.3980G>A(p.Trp1327*) [VUS] OBSCN c.2351G>A(p.Arg784His)[VUS]
009338	56	No	European		
009415	64	No	European		ALPK3 c.668delG(p.Gly223fs)[VUS] ALPK3 c.668G>T(p.Gly223Val)[VUS] (in cis)
009420	46	No	European		FOXRED1: c.658C>T(p.Pro220Ser) [VUS] MYH6: c.1408G>A(p.Asp470Asn) [VUS] TTN: c.107290G>T(p.Ala35764Ser) [VUS] ALPK3 c.4091G>C(p.Gly1364Ala)[VUS]
009994	47	No	European		TTN: c.4247G>A(p.Arg1416His) [VUS] TTN: c.88285A>G(p.Ile29429Val) [VUS]
009995	67	No	European		PTPN11:c.173A>G(p.Asn58Ser) [VUS] OBSCN c.21107G>A(p.Gly7036Asp)[VUS]
009996	63	No	European		
010004	56	No	European		MYPN: c.1858A>G(p.Thr620Ala) [VUS] TTN: c.44588C>G(p.Thr14863Arg) [VUS] TTN: c.76538A>C(p.Asp25513Ala) [VUS] TTN: c.95243G>A(p.Arg31748His) [VUS] TMPO c.1003A>C(p.Thr335Pro)[VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
010005	67	No	European		
010016	53	No	European		
010017	70	No	European		
010089	54	Yes	European		JPH2: c.1837G>A(p.Glu613Lys) [VUS]
010103	63	Yes	European		MYOM1: c.1847G>A(p.Arg616His) [Likely benign]
010104	47	No	European		
010105	74	Yes	European		OBSCN c.8269G>T(p.Glu2757*)[VUS]
010106	55	No	European		KLF10: c.941A>G(p.Lys314Arg) [VUS] TTN: c.11183dupG(p.Leu3729fs) [VUS]
010107	52	No	European		TTN: c.82385C>T(p.Thr27462Met) [Likely benign]
010115	54	No	European		
010307	47	Yes	European		OBSCN c.17897G>A(p.Arg5966Gln)[VUS]
010313	53	No	European		MYH6: c.4208A>C(p.Glu1403Ala) [VUS]
010320	51	No	European		
010412	52	No	South-Asian		TTN: c.2283_2288del(p.Lys762_Ala763del) [VUS] TTN: c.21506A>G(p.Asn7169Ser) [VUS] OBSCN c.130G>A(p.Glu44Lys)[VUS]
010413	54	Yes	European		TTN: c.16517C>T(p.Ser5506Phe) [VUS]
010414	60	No	European		TCAP:c.66G>A(p.Trp22*) [VUS]
010415	20	Yes	European		FLNC: c.4061G>A(p.Arg1354Gln) [VUS] (phase unknown) FLNC: c.4277G>A(p.Arg1426Gln) [VUS] NEXN: c.1435C>T(p.Leu479Phe) [VUS] TTN: c.81852_81857dup (p.Val27286_Val27287insIleVal) [VUS]
010417	55	No	East-Asian		MYH6: c.317A>G(p.Lys106Arg) [VUS] DSP c.1027G>A(p.Ala343Thr)[VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
010427	60	Yes	European		TTN: c.35408T>C(p.Ile11803Thr) [VUS] TTN: c.82220T>C(p.Ile27407Thr) [VUS]
010592	61	No	European		TTN: c.105757G>A(p.Val35253Met) [VUS]
011072	55	No	European		
011073	43	No	European		DES: c.391C>A(p.Gln131Lys) [VUS]
011074	57	Yes	European		DES: c.832C>T(p.Arg278Trp) [VUS]
011077	56	No	European		TTN: c.10467_10498dup(p.His3500fs) [VUS] DSP c.7310A>G(p.Glu2437Gly)[VUS]
011078	61	Yes	European		MYPN: c.173A>T(p.Asp58Val) [VUS] OBSCN c.23594C>T(p.Ala7865Val)[VUS]
011092	61	No	European	ACTC1: c.435T>A(p.Tyr145*) [VUS]	
011103	33	No	African		TTN: c.8596G>A(p.Ala2866Thr) [VUS]
011106	46	Yes	European		DSP c.8372G>T(p.Arg2791Leu)[VUS] OBSCN c.22945G>A(p.Glu7649Lys)[VUS]
011115	52	No	European		TTN: c.14327A>G(p.Asn4776Ser) [VUS] OBSCN c.19736C>T(p.Thr6579Met)[VUS]
011266	27	No	European		
011267	69	Yes	European		GAA: c.131G>T(p.Gly44Val) [VUS]
011271	57	Yes	European		
011272	50	No	European	MYH7: c.964T>A(p.Ser322Thr) [VUS]	GAA: c.1504A>G(p.Met502Val) [VUS] TTN: c.55852T>G(p.Cys18618Gly) [VUS] DSP c.7523_7525delTCA(p.Ile2508del)[VUS]
011365	67	No	European		MYOZ2: c.213_230del(p.Glu71_Ser77delinsAsp) [VUS]
011376	46	Yes	European		KCNQ1: c.352A>G(p.Thr118Ala) [VUS]
011377	-	-	European		AGL: c.2632A>C(p.Ser878Arg) [VUS]
011461	60	No	European		DSP c.6332T>C(p.Ile2111Thr)[VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
011468	41	Yes	European		TTN: c.24535G>A(p.Asp8179Asn) [VUS] TTN: c.86947G>C(p.Glu28983Gln) [VUS] RYR2: c.2123G>T(p.Gly708Val) [VUS] TTN: c.24535G>A(p.Asp8179Asn) [VUS]
011962	58	No	European		MYO6: c.2020A>G(p.Lys674Glu) [VUS] DSP c.5833C>T(p.Arg1945Cys)[VUS]
011974	59	No	European		MYH6: c.5222C>T(p.Ser1741Leu) [VUS] MYO6: c.3608C>T(p.Ala1203Val) [VUS] TTN: c.39235G>A(p.Val13079Ile) [VUS]
011986	61	No	European		OBSCN c.25639G>A(p.Val8547Met)[VUS]
011998	55	No	European		TTN: c.1030C>G(p.Pro344Ala) [VUS]
012010	67	No	European		AGL: c.1679G>A(p.Ser560Asn) [VUS] TTN: c.83279A>T(p.Asn27760Ile) [VUS] TTN: c.90104G>A(p.Arg30035His) [VUS]
012220	63	No	African		
012225	66	No	European		OBSCN c.7966G>A(p.Asp2656Asn)[VUS] OBSCN c.25361G>A(p.Ser8454Asn)[VUS]
012284	67	No	European		TRIM63: c.434G>A(p.Cys145Tyr) [VUS]
012296	68	No	European		JPH2: c.1759_1779dup (p.Glu593_Ser594insProGluValSerGlySerGlu) [VUS]
012308	54	Yes	European		
012326	61	No	Other		TTN: c.50869A>G(p.Ile16957Val) [VUS] TTN: c.99154C>T(p.Arg33052Cys) [VUS]
012338	48	Yes	European	MYBPC3: c.3470C>T(p.Pro1157Leu) [VUS]	OBSCN c.3592_3596delACGTG(p.Thr1198fs)[VUS] TMPO c.1054G>A(p.Glu352Lys)[VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
012350	36	No	European	MYH7: c.5135G>A(p.Arg1712Gln) [Original: VUS, Revised: Likely Pathogenic]	ALPK3 c.1259C>T(p.Pro420Leu)[Likely benign]
012438	67	No	European		RAF1: c.66T>G(p.Phe22Leu) [VUS] TTN: c.47402C>T(p.Thr15801Ile) [VUS]
012450	55	No	African		KCNQ1: c.160_168dup(p.Ile54_Pro56dup) [Benign]
012462	49	Yes	European		DES: c.1286G>A(p.Arg429Gln) [VUS] OBSCN c.14470A>G(p.Thr4824Ala)[VUS]
012521	52	Yes	European	MYBPC3: c.655G>C(p.Val219Leu) [Likely Pathogenic]	GYS1: c.355G>A(p.Val119Met) [VUS] RYR2: c.1354G>A(p.Val452Ile) [VUS] GYS1: c.355G>A(p.Val119Met) [VUS] OBSCN c.4451C>G(p.Ser1484Cys)[VUS]
012522	68	Yes	European	MYBPC3: c.2429G>A(p.Arg810His) [VUS]	GYS1: c.1279A>G(p.Met427Val) [VUS] TTN: c.83600C>G(p.Pro27867Arg) [VUS]
012533	68	No	European	MYH7: c.5342G>A(p.Arg1781His) [VUS]	
012545	53	Yes	European		GAA: c.525delT(p.Glu176fs) [VUS]
012557	27	Yes	European	MYBPC3: c.3771C>A(p.Asn1257Lys) [VUS]	TTN: c.90616C>T(p.Leu30206Phe) [VUS]
012569	50	No	European	MYH7: c.2273T>C(p.Phe758Ser) [VUS]	
012581	58	No	European		AGL: c.3911delA(p.Asn1304fs) [VUS] TTN: c.2848A>G(p.Lys950Glu) [VUS] TTN: c.9851A>G(p.Lys3284Arg) [VUS]
012593	70	Yes	European	CSRP3: c.49G>A(p.Val17Ile) [VUS]	AGL: c.4271A>T(p.Tyr1424Phe) [VUS]
012605	69	No	European		TTN: c.62486T>G(p.Phe20829Cys) [VUS]
013283	43	No	European		FLNC: c.6310G>A(p.Glu2104Lys) [VUS] ALPK3 c.1810delG(p.Ala604fs)[VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
013285	30	Yes	European	MYH7: c.2631G>C(p.Met877Ile) [VUS]	MYOM1: c.1378T>A(p.Trp460Arg) [VUS] ALPK3 c.1846T>C(p.Ser616Pro)[VUS]
013295	23	Yes	Other		
013307	29	No	Other		MYO6: c.3791A>G(p.Asn1264Ser) [Likely benign]
013319	43	Yes	European		
013331	33	No	European		SOS1: c.890C>T(p.Ser297Leu) [VUS]
013343	49	No	European		
013354	37	Yes	European		FOXRED1: c.1190C>G(p.Ala397Gly) [VUS]
013355	67	No	European		
013366	46	Yes	European		
013674	68	No	European		CACNA1C: c.5347G>A(p.Gly1783Ser) [VUS] TTN: c.15430G>A(p.Glu5144Lys) [VUS] TTN: c.62867T>C(p.Ile20956Thr) [VUS]
013686	61	Yes	European		OBSCN c.7227G>C(p.Glu2409Asp)[VUS]
013698	34	No	South-Asian		MYH6: c.4793C>T(p.Ser1598Leu) [VUS] RYR2: c.10709G>A(p.Arg3570Gln) [VUS] OBSCN c.24427G>A(p.Ala8143Thr)[VUS]
013710	53	No	European		
013818	44	Yes	European		DSP c.7649G>A(p.Gly2550Asp)[VUS]
013958	55	No	Other		FLNC: c.1328C>T(p.Ala443Val) [VUS] FLNC: c.1606G>A(p.Glu536Lys) [Likely benign] ALPK3 c.1207A>T(p.Lys403*)[VUS]
013970	34	Yes	East-Asian	TPM1: c.82G>C(p.Asp28His) [VUS]	FXN: c.41C>A(p.Ala14Glu) [VUS] OBSCN c.8171C>T(p.Pro2724Leu)[VUS]
014005	68	Yes	European		CACNA1C: c.6388G>T(p.Asp2130Tyr) [VUS] TTN: c.6907G>C(p.Asp2303His)[VUS]

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
014017	52	No	European		TTN: c.55531A>T(p.Ile18511Phe) [VUS]
014029	67	Yes	European		
014041	51	No	European		ALPK3 c.4897G>A(p.Gly1633Arg)[VUS]
014358	34	No	European		
014369	53	No	European		OBSCN c.10790A>T(p.Glu3597Val)[VUS]
014370	49	No	European		
014381	62	No	European		TTN: c.22576G>T(p.Val7526Phe) [VUS] TTN: c.45088G>C(p.Glu15030Gln) [VUS] TTN: c.65515G>A(p.Ala21839Thr) [VUS]
014382	49	No	European		FLNC: c.4504A>G(p.Thr1502Ala) [VUS] TTN: c.64916G>A(p.Arg21639Gln) [VUS] OBSCN c.11728T>A(p.Trp3910Arg)[VUS]
014393	50	No	European		
014394	21	No	European		CACNA1C: c.1343A>C(p.Asp448Ala) [VUS]
014405	51	No	European		
014406	67	Yes	European		
014417	54	Yes	European		OBSCN c.3661C>T(p.Arg1221Trp)[VUS]
014418	61	No	European		FLNC: c.34C>G(p.Leu12Val) [VUS]
014429	48	Yes	European		
014430	52	No	European		
014434	63	No	European		
014435	59	No	European	MYH7: c.1405G>A(p.Asp469Asn) [VUS]	JPH2: c.692G>A(p.Arg231Gln) [Likely benign]
014436	50	No	European	FHL1: c.662G>A(p.Cys221Tyr) [VUS]	MYO6: c.3062C>G(p.Ala1021Gly) [VUS]
014437	42	No	European		
014446	59	No	European		
014447	57	No	European		TTN: c.85930A>G(p.Arg28644Gly) [VUS]
014448	46	No	European		

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
014449	58	No	European		
014458	58	No	Other		
014459	49	No	European		
014460	41	No	European		TTN: c.56942C>A(p.Ala18981Glu) [VUS]
014461	52	No	European		TTN: c.78947G>A(p.Gly26316Asp) [VUS] OBSCN c.9140G>C(p.Ser3047Thr)[VUS]
014470	51	Yes	European		AGL: c.766C>T(p.Arg256Cys) [VUS]
014471	54	No	European		
014472	54	Yes	European		
014482	55	Yes	European		TTN: c.60140G>A(p.Arg20047Lys) [VUS]
014483	58	No	European		
014484	49	No	European		TTN: c.39703C>G(p.Pro13235Ala) [VUS]
014494	41	No	South-Asian	MYH7: c.5771G>T(p.Ser1924Ile) [VUS]	ALPK3 c.2307C>A(p.Ser769Arg)[VUS]
014495	54	No	European		DES: c.43C>A(p.Arg15Ser) [VUS]
014496	51	Yes	European		
014506	41	No	Other		
014507	51	No	European		SLC25A3: c.659C>T(p.Ala220Val) [VUS] OBSCN c.4765C>T(p.Arg1589Trp)[VUS]
014508	20	Yes	Other	MYBPC3: c.1816G>A(p.Val606Ile) [VUS]	DES: c.1364A>T(p.Asp455Val) [VUS] MRPL3: c.1025C>T(p.Ala342Val) [VUS]
014518	50	No	European		FLNC: c.3650_3652del(p.Ser1217_Pro1218delinsThr) [VUS]
014519	69	No	European		KCNQ1: c.275C>G(p.Ser92Cys) [VUS]
014520	66	No	European		TTN: c.104305G>A(p.Glu34769Lys) [VUS]
014865	64	No	European	MYBPC3: c.3798C>G(p.Cys1266Trp) [VUS]	TTN: c.54704G>A(p.Arg18235His) [VUS]
014877	50	No	European		

ID	Age*	Family History HCM	Ethnicity	Variants detected in clinical test panel genes	Variants detected in the 60 selected genes
014889	73	Yes	European		TTN: c.6244G>A(p.Glu2082Lys) [VUS] DSP c.5363A>G(p.Gln1788Arg)[VUS]

Table 3.3-30 Variants detected in HCM clinical test panel genes and variants detected in the 60 genes selected for analyses. Clinical panel test genes include *ACTC1*, *ACTN2*, *ANKRD1*, *CSRP3*, *FHL1*, *GLA*, *LAMP2*, *MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *PLN*, *PRKAG2*, *TNNI3*, *TNNT2* and *TPM1*. Variant review and classification was undertaken using clinical guidelines. Variants identified in genes in clinical test panel were also assessed against current evidence, where classification changed since original clinical laboratory assessment, both original and revised classification listed. Age*=Age (in years) at first presentation to Inherited Cardiomyopathy Clinic (ICC). Ethnicity determined by PCA analysis (see Methods 3.2.6). Family history of HCM documented at ICC.

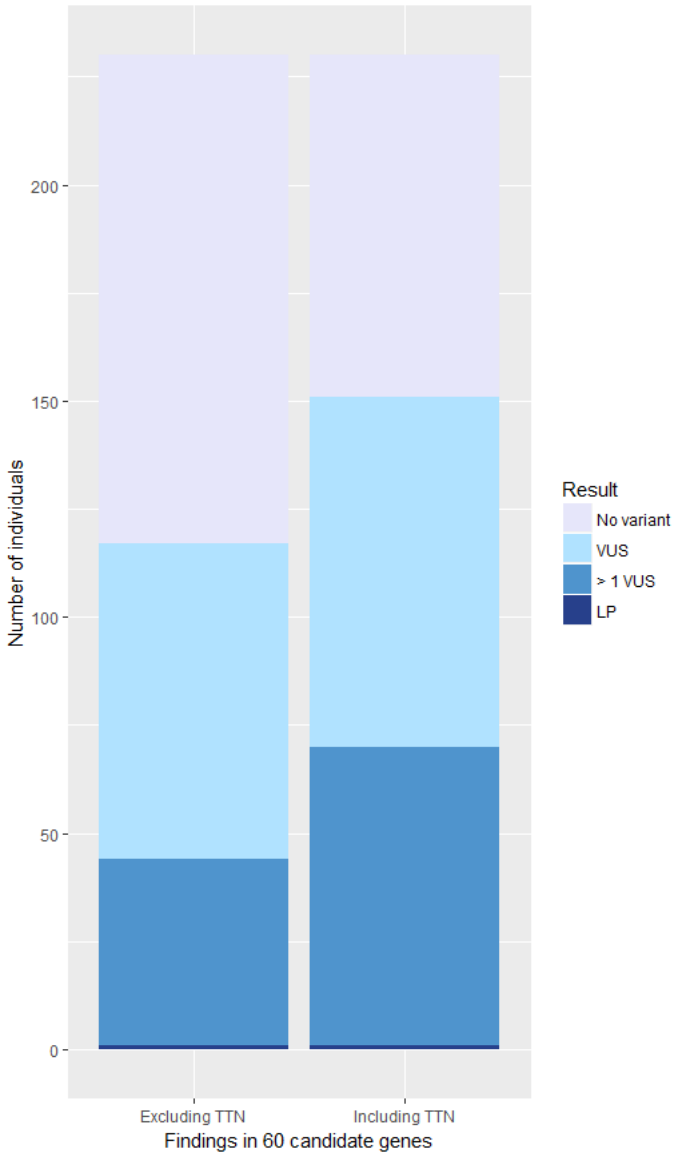


Figure 3.3-23 Summary of results of rare variant analyses in the 60 selected genes in 230 HCM cases. In the first column rare variants detected in the *TTN* gene were not included. The second column includes rare variants detected in the *TTN* gene. VUS=Variant of uncertain significance, >1 VUS =More than one variant of uncertain significance detected. LP=Likely pathogenic variant.

3.4 Discussion

3.4.1 Characteristics of the cohort

Cases were recruited from a cohort of patients with a clinical diagnosis of HCM in whom no pathogenic variants were found in key well-established HCM genes. This cohort should therefore be enriched for non-sarcomere genetic forms of HCM making it an ideal subgroup in which to explore the contribution of other putative causal genes. On review of the literature, this appears to be the largest HCM WGS cohort reported to date. However, this cohort was only sufficiently powered to allow coherent assessment of a small number of selected genes (Table 3.2-2).

The preponderance of males in this HCM cohort (75% of cases were male compared to 45% of controls) has been noted in previous HCM cohorts [226,227]. This is most likely to have a biological basis; however, it may reflect differences in behaviour (e.g. physical training) or ascertainment. None of the genes analysed are on the X-chromosome and case-control analyses were carried out on pooled genders as autosomal allele frequencies are not expected to show gender differences.

The availability of whole genome data from HCM cases along with extensive control data sequenced as part of the same study was integral to these analyses, as differences in rare variant frequency arising from variation in sequencing protocols and variant annotation are minimised [142].

3.4.2 Ethnicity

The cases and control cohorts were found to be of broadly similar ethnicity; this reduces potential confounding effects due to differences in rare variant frequency between populations of different ethnicity[228].

3.4.3 Gene coverage

The mean depth of sequence coverage across the coding regions of the selected genes was considered sufficient to enable accurate variant calling [229,230]. The *MYH6* gene was a notable outlier; further investigations indicated that the coverage issues in this gene are likely due to reads being discarded due to poor mapping quality of sequencing reads in regions that share high sequence homology with the *MYH7* gene.

3.4.4 Regions of Homology

Sequencing reads arising from highly homologous regions can difficult to map uniquely and mapping errors can arise, leading to inaccurate estimates of sequencing coverage and errors in variant annotation[221]. This is recognised as one of the most frequent causes of false-positive and variant calls in NGS data. Critically, the ability to distinguish between paralogous genes, or genes and non-functional pseudogenes is essential when predicting the consequence of variants.

Data from a publically available resource published by Mandelker *et al*[231] was used to identify genes which show high sequence homology to other genomic regions and could therefore be refractory to short read NGS analysis or prone to variant annotation errors.

While it is possible that the regions of homology highlighted in the *TTN* gene could result in false positive variant annotation, the risk of these being mis-interpreted and classified as pathogenic is considered to be minimal, as there is no evidence supporting a role for this gene in HCM. Regions of high sequence homology identified in two other genes, *FLNC* and *MYH6*, were considered to have the potential to result in clinically relevant false-positive and false-negative variant annotation and classification.

3.4.4.1 FLNC

FLNC is included in the Genomics England PanelApp HCM gene list and had been selected for review by the ClinGen Gene curation group. There are reports in the literature suggesting a possible role for missense variants in this gene in HCM[232–234]. Investigations were undertaken to explore the possibility of variant annotation errors in this gene arising due to misalignment of next generation sequencing reads from homologous regions of *FLNC* and the FLNC pseudogene (pseFLNC). Misalignment could lead to both false positive (pseFLNC reads mapping to *FLNC*) and false negative (*FLNC* reads mapping to pseFLNC) variant calls; notably, false positive variant calls in these regions have been reported previously[235].

Evidence of false-positive variant annotation arising due to mis-alignment of sequence reads from pseFLNC to *FLNC* was observed in this cohort. Exclusion of variants detected in regions that show high homology did not alter the results of case-control analysis. Although no rare variants were detected in homologous regions in HCM cases, these analyses highlight that variants detected in these homologous regions should be interpreted with caution and certainly should not be reported in a clinical setting without confirmatory analysis (i.e. Sanger sequencing, restriction enzyme digest or long read NGS sequencing).

3.4.4.2 MYH6

The *MYH6* gene is included in several clinical HCM test panels (Figure 3.2-1) and is listed by Genomics England Panel App as a definitive diagnostic-grade gene. *MYH6* encodes the alpha-myosin Heavy chain protein (α-MyHC), which is one of the two major myosin heavy chain (MHC) isoforms expressed in the human heart, the other being the beta-myosin heavy chain (β-MyHC), which is encoded by the *MYH7* gene[236]. α-MyHC is the

predominant isoform in the atria during cardiac muscle development and postnatally, with lower levels of expression seen in the ventricles, particularly in adult heart [236]. β-MyHC is the predominant isoform in ventricles in both the fetal and adult human heart[236,237].

Variants in the *MYH6* gene have been implicated in several cardiac conditions, including Atrial Septal Defects[238–240], Dilated Cardiomyopathy[172,184,241], Sick Sinus Syndrome[242,243] and HCM[168,244–246]. However, review of the literature indicates that there is little evidence to support the pathogenicity of the *MYH6* variants which have previously been reported in individuals with HCM (Table 3.4-2). Additionally, data from the ExAC consortium suggests that *MYH6* is not under the same levels of constraint for missense variants as *MYH7* (Table 3.4-1).

Investigations were undertaken to explore the likely impact of regions in the *MYH6* gene which show high homology to the *MYH7* gene. These genes, which are arranged in tandem approximately 4kb apart on chromosome 14 (Figure 3.4-1), show ~95% homology at the nucleotide level and 96% at the amino acid level. As pathogenic variants in *MYH7* are one of the most frequent causes of HCM, there is an increased risk that false negative and positive variant annotations arising from sequence read misalignment could have clinical consequences.

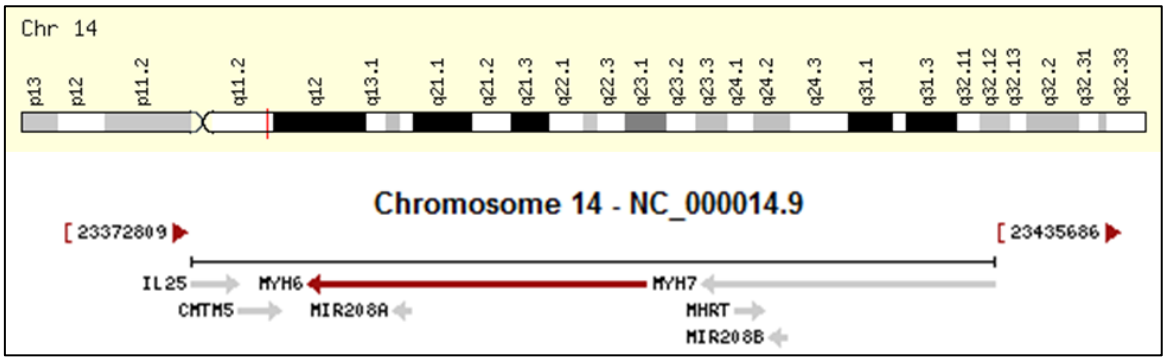


Figure 3.4-1 Location of *MYH6* and *MYH7* genes on Human Chromosome 14.

Review of coverage data highlighted a decrease in sequence coverage in exon 26 of *MYH6* and an increase in coverage in the corresponding homologous exon 27 in *MYH7*. These exons show 100% sequence identity and cover ~390 nucleotides. Given the size of the sequencing reads in this project (100-150bp), it is likely that a proportion of reads arising from these regions will not contain any unique sequence. Although, the expectation would be that non-unique reads would be discarded rather than misaligned.

In the other regions of *MYH6*/*MYH7* homology, a number of possible false positive variant annotations were identified (Table 3.3-6), suggesting that sequence mis-alignment may be occurring in these regions. Several of these possible false positive variant annotations were sufficiently rare that they were included in case-control analysis. Although no difference in rare variant frequency was found between cases and controls in this study, with or without inclusion of rare variants detected in homologous regions, two rare variants detected in these regions in cases had undergone pathogenicity investigations and were included in VUS counts.

These findings highlight uncertainty in the origin of sequence reads in homologous regions of *MYH6* and consequently a risk of false-positive results. Scrutiny of the associated BAM files might provide some evidence regarding the gene of origin of the sequence reads in which variants have been annotated. However, in a clinical setting, variants in these regions should be confirmed by a second method that allows sequence arising from homologous regions in these genes to be identified uniquely.

Constraint data from ExAC				
Gene	**Variant type**	**Expected no. variants**	**Observed no. variants**	**Constraint Metric**
MYH6	Synonymous	381.8	358	z = 0.76
	Missense	816.9	649	z = 2.87
	LoF	64.8	38	pLI = 0.00
	CNV	13.3	20	z = -0.50
MYH7	Synonymous	378.0	391	z = -0.41
	Missense	797.4	420	z = 6.54
	LoF	70.2	21	pLI = 0.00
	CNV	12.2	17	z = -0.39

Table 3.4-1 Exome aggregation consortium (ExAC) constraint scores for *MYH6* and *MYH7* genes. (http://exac.broadinstitute.org/). Missense: For synonymous and missense, Z score represents the deviation of observed counts from the expected number. Positive Z scores indicate increased constraint (intolerance to variation) and therefore that the gene had fewer variants than expected. Negative Z scores are given to genes that had a more variants than expected.
LoF (loss of function): The closer pLI is to one, the more LoF intolerant the gene appears to be. pLI >= 0.9 as an extremely LoF intolerant set of genes. [138]

Variant	Reference	Age of index case	Phenotype	Previous Testing	Comments	ClinVar	ExAC
c.2384G>A (p.Arg795Gln)	245	74yrs	"Elderly-onset" HCM	MYH7,MYBPC3, TNNT2,TNNI3, TPM1, MYH6.	No family history. No segregation. Not in 170 control chromosomes.	14147 (1 submission: 0 stars: Pathogenic)	ExAC(20/09/16): 4/121274 2/8650 East Asian 2/66648 Euro(NF)
c.3195G>C (p.Gln1065His)	244	27 years (at diagnosis) Died at 45 years.	HCM	None specified.	Not detected in offspring, 22 and 15 years at testing; asymptomatic and no evidence of HCM on clinical screening. Not in 300 "ethnically similar" control chromosomes.	14149 (4 submissions:1 Star: Conflicting)	ExAC(27/09/2016): 28/121410 2/908 Other 11/8654 East Asian 7/6614 Euro(Finnish), 2/16512 South Asian 6/66738 Euro(NF).
c.220G>A (p.Gly56Arg)	168	29 years	HCM	28 HCM gene panel	Sporadic HCM	44455 (3 submissions:2 stars: Benign)	ExAC(27/09/2016): 8472/121382 Detected in all population subgroups.
c.611G>A (p.Arg204His)	246	Not specified	"Late onset" HCM	17 gene HCM panel	No segregation	Not listed	ExAC(27/09/2016): 16/121392 14/66726 Euro(NF) 2/10406 African
c.2425C>T (p.Arg809Cys)		Not specified	"Late onset" HCM		No segregation	Not listed	ExAC(27/09/2016): 4/121388 4/16510 South Asian
c.5519A>G (p.Lys1840Arg)		Not specified	"Late onset" HCM		No segregation	Not listed	ExAC(27/09/2016): 10/121288 1/908 Other 9/66682 Euro(NF)

Table 3.4-2 Putative pathogenic variants reported in the *MYH6* gene in individuals with HCM. ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Variants annotated against genome build GRCh37, Ensembl transcript ID: ENST00000405093.7 Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/). Euro(NF)=European (Non-Finnish) population.

3.4.5 Gene level analyses

3.4.5.1 ALPK3

Of the 60 genes analysed, strong evidence for a case excess of rare variants was found in a single gene. Truncating variants in the *ALPK3* gene appeared to be more frequent in HCM cases compared to controls (p=0.0001, OR: 16.97 [95% CI: 4.27-52.05]). This gene had been selected for review by the ClinGen HCM gene curation working group on the basis of reports in the literature linking homozygous truncating variants to autosomal recessive infant-onset cardiomyopathy. However, it was not listed in OMIM, Orphanet, or any of the other key resources reviewed during gene selection, is not in the Genomics England Gene Panel App HCM list, and does not appear to be included in current clinical test panels (Figure 3.2-1). As such, this apparent excess of rare heterozygous truncating variants in HCM patients was considered a novel finding. The likely clinical significance of heterozygous truncating variants in *ALPK3* in the context of autosomal dominant HCM was investigated.

3.4.5.1.1 Gene structure and protein function

ALPK3 is located on the long arm of chromosome 15 (15q25.2, 15:84816680-84873482), it contains 14 exons, and is translated into Alpha-protein kinase 3, a 1907 amino acid 200kDa protein. Alpha-protein kinase 3 is a nuclear kinase protein belonging to the alpha-kinase family of proteins[247]. Protein kinases are catalytic proteins which regulate signalling cascades controlling a wide range of functions and physiological processes within the cell. The most common kinase proteins, referred to as the "conventional protein kinases" (CPKs), are the serine/threonine kinases and tyrosine kinases. These highly homologous proteins are encoded by over 500 genes in the human genome[247].

Alpha-protein kinases, which appear to be unique to eukaryotic organisms, were named in light of early studies showing their propensity to phosphorylate serine and threonine residues in the context of an alpha helix[248]. Alpha-kinase proteins are classified as "atypical protein kinases" (APKs); this family of kinases show little sequence homology with CPKs, and although there are similarities in the architecture of the catalytic domains, alpha-kinases harbour distinct protein motifs and employ different modes of substrate recognition[247–249].

Six genes encoding human alpha-kinase proteins have been identified to date (*eEF2K*, *ALPK1*, *ALPK2*, *ALPK3*, *TRPM6* and *TRPM7*). These proteins have been implicated in a diverse range of cellular processes including protein translation, magnesium homeostasis, intracellular transport, as well as cell migration, adhesion, growth and proliferation [247,249].

Defects in alpha-protein kinase genes have previously been implicated in cancer (*eEF2K*, *TRPM7*) and familial hypomagnesia (*TRPM6*). More recently, homozygous truncating mutations in *ALPK3* have been reported to cause cardiomyopathy[250–252].

Alpha-protein kinase 3 shows expression in several tissues, but is most abundantly expressed in fetal and adult heart, and in adult skeletal muscle[249]. It is thought to act as a transcriptional regulator implicated in early cardiac development[253]. It contains two Immunoglobulin-like (Ig-like) domains, which typically mediate protein-protein interactions, and an Alpha-type protein kinase domain.

3.4.5.1.2 Mouse studies

The role of *ALPK3* in cardiomyopathy was first reported in 2012. In a large scale study of mouse knockouts, aiming to identify genes critical for cardiac development and

function, Alpk3 null mice were found to develop cardiomyopathy. Mice deficient for Alpha-protein kinase 3 developed myocardial hypertrophy along with features of dilated cardiomyopathy (increase in end-diastolic and end-systolic LV volume, reduced cardiac output and stroke volume, and LV chamber dilation), however they did not go on to develop the ventricular wall thinning characteristic of DCM. Histological examination revealed mild myofibril disarray, a reduced number and abnormal appearance of intercalated discs, and abnormal nuclei [254].

The findings in this study suggest that Alpha-protein kinase 3 is important for the structural integrity of the intercalated discs (ICDs). These structures play a critical role in establishing the proper alignment of myofibrils and in maintaining mechanical coupling and chemical communication between cardiomyocytes (Figure 3.4-2).[255] The authors of this study proposed that loss of Alpha-protein kinase 3 results in defective intercalated disc formation, causing myofibril disarray which impairs cardiomyocyte contractility and precipitates cardiac insufficiency and cardiomyopathy[254].

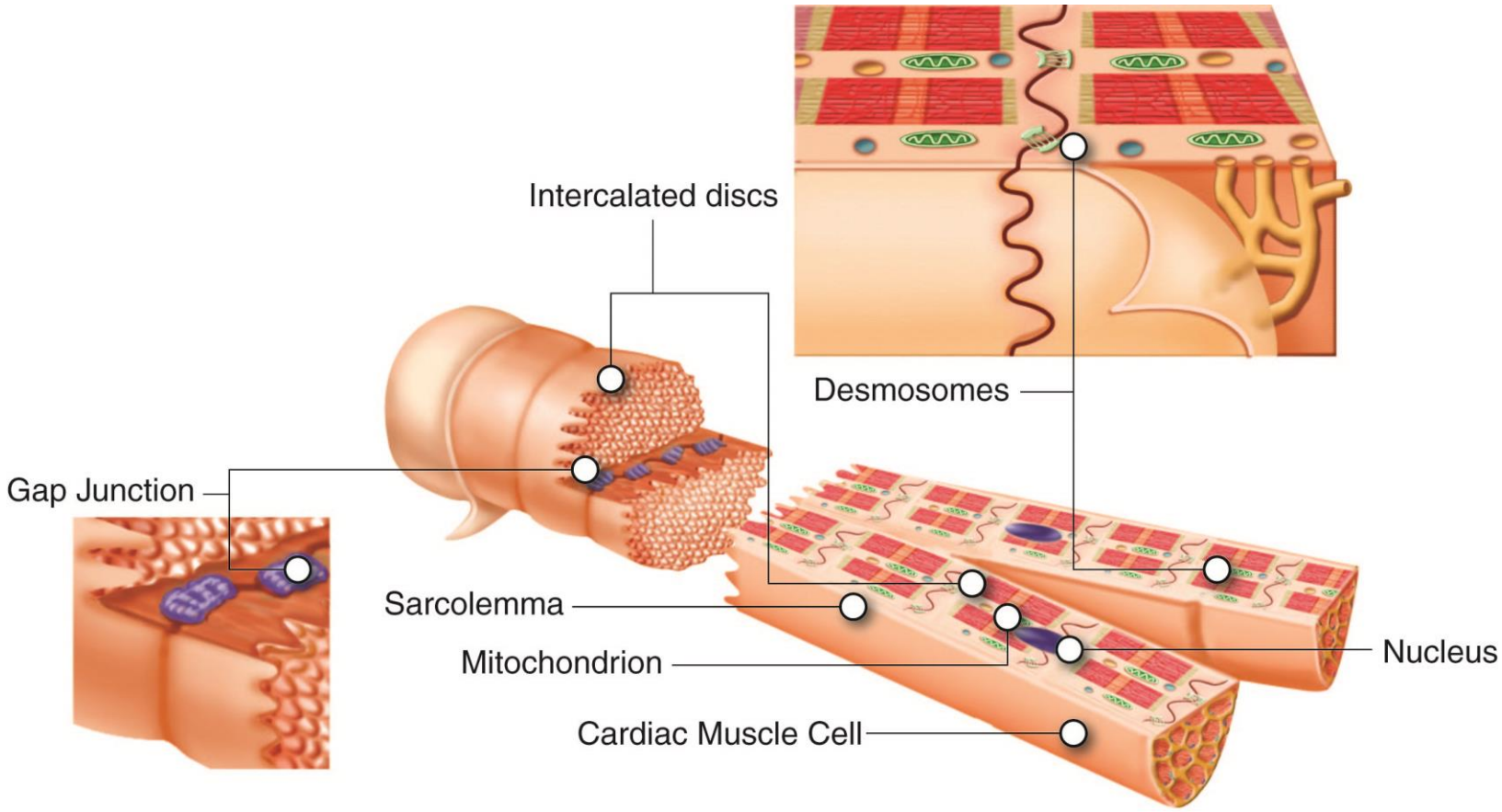


Figure 3.4-2 Anatomy of a cardiac muscle cell. Intercalated disc (IC) is a specialised intercellular junction between individual cardiac muscle cells. Gap Junctions lie within the IC and function to transmit electrical synapses. Structural connections are provided by the Desmosomes. Image from https://courses.lumenlearning.com/cuny-csi-ap-1/chapter/muscular-levels-of-organization/ Authored by: Open Learning Initiative. Provided by: Carnegie Mellon University. Located at: https://oli.cmu.edu/jcourse/workbook/activity/page?context=43488f3780020ca60140b159062e6e31. Project: Anatomy & Physiology. License: *CC BY-NC-SA: Attribution-NonCommercial-ShareAlike*

3.4.5.1.3 Role in disease

The first cases of *ALKP3*-related cardiomyopathy in humans were reported by Almomani *et al* in 2016[250]. In this study, using a combined approach of homozygosity mapping and exome sequencing, five individuals from three unrelated consanguineous families with early-onset cardiomyopathy were found to have homozygous putative truncating mutations in *ALPK3* (c.4736-1G>A, p.Val1579Glyfs*30; c.3781C>T, p.Arg1261*; and c.5294G>A, p.Trp1765*). All homozygous individuals presented with severe HCM and/or DCM *in utero*, at birth, or in early childhood. Three patients died from heart failure within the first week of life. Two siblings from one pedigree, who were heterozygous for the p.Trp1765* variant, had been diagnosed with an atypical form of hypertrophy at a young adult age. No evidence of cardiomyopathy was found in the other eight proven or obligate heterozygous carriers investigated (age range 2 months to 42 years).

Phelan *et al* [251] subsequently reported a consanguineous pedigree in which a homozygous truncating variant (c.3972G>A, p.W1264X) presented with infant onset cardiomyopathy (LV dilation and non-compaction), progressing to HCM with multiple pterygia and skeletal muscle abnormalities. They found no evidence of cardiomyopathy in the seven family members heterozygous for this variant.

More recently, a fifth pedigree was reported by Caglayan et al[252]. Here the index case, who was homozygous for the c.2018delC (p.Gln675SerfsX30) variant, was found to have a dilated cardiomyopathy (DCM) *in utero,* which progressed to HCM after birth. The index case was also reported to have dysmorphic features, with low set ears and high arched palate. Heterozygous family members were normal on echocardiogram. Although

this family was reported to be non-consanguineous, the parents were both from the same small village and as such are likely to be distantly related.

3.4.5.1.4 Histopathological, RNA and functional studies

In vitro studies have been used to explore the molecular mechanisms and cellular consequences of *ALPK3*-deficiency. Almomani *et al* performed post mortem examination of the hearts of two individuals homozygous for the *ALKP3* LoF variants c.4736-1G>A (p.Val1579Glyfs*30) and c.3781C>T (p.Arg1261*). Macroscopically, the hearts of these individuals showed severe cardiomegaly and biventricular dilation. Microscopically, they observed extensive fibroelastosis without apparent myocyte disarray. Patient material was not available for ultrastructural examination of the intercalated discs, however, immunohistochemical analysis in myocardial tissue from one patient revealed loss of signal from certain desmosomal proteins (plakoglobin and desmoplakin).

The consequence of the *ALPK3* c.4736-1G>A variant was also investigated. Analysis of cDNA from cultured fibroblasts confirmed abnormal splicing, with skipping of exon 10, leading to a deletion of 281 nucleotides, which is predicted to introduce a premature termination codon (p.Val1579Glyfs*30).

Phelan *et al* performed a series of analyses in induced pluripotent stem cell lines (iPSC) derived from a patient homozygous for the c.3792G.A (p.W1264X) variant, and in two independent CRISPR/Cas9 generated *ALPK3*-null human Embryonic Stem Cell lines (hESC).

Analysis of RNA from patient-derived fibroblasts, skeletal muscle and human pluripotent stem cells (hPSCs) demonstrated that the *ALPK3* transcript bearing the c.3792G.A (p.W1264X) variant underwent nonsense-mediated decay.

Transmission electron microscopy of *ALPK3*-mutant cardiomyocytes, from both iPSCs and hESC lines, demonstrated sarcomere disorganisation and abnormal intercalated disc morphology. Multi-electrode array (MEA) analysis in *ALPK3*-mutant hESC-derived cardiomyocytes indicated that loss of *ALPK3* disrupts membrane repolarisation. Also, irregularities in calcium flux were observed during contraction in *ALPK3*-mutant cardiomyocytes; this is indicative of abnormal calcium handling, as has been observed in other hPSC models of HCM[256]. Another key finding was that heterozygous correction of *ALPK3* in proband-derived iPSC cell lines using CRISPR/cas9 editing restored a wild type phenotype.

Taken together, the findings from these two studies provide strong support for the assertion that homozygous truncating variants in *ALPK3,* resulting in Alpha-protein kinase 3 deficiency, cause a distinct type of early onset cardiomyopathy.

3.4.5.1.5 Heterozygous ALPK3 truncating variants in HCM

Although evidence of an excess of rare truncating variants was found in our case cohort, the likely clinical consequences of heterozygous variants remain unclear, particularly as these types of variant are also found at low frequency in reference control cohorts (BRRD, ExAC and gnomAD).

The results from *in vitro* investigations undertaken to date suggest that the truncating variants reported in *ALPK3* lead to the introduction of a premature termination codon (PTC), and that these transcripts are degraded by Nonsense-mediated RNA decay (NMD)[250,251,255]. NMD is a surveillance pathway that exists in eukaryotic cells; it identifies mRNA transcripts that contain premature termination codons (PTCs) and targets them for degradation, preventing translation of potentially deleterious truncated proteins.

Phelan et al[251] speculated that the HCM phenotype in the two heterozygous cases reported by Almomani et al[250] could be the consequence of a dominant negative truncated protein, as opposed to protein haploinsufficiency. The variant in question, c.5294G>A (p.Trp1765*), lies within a C-terminal region of the protein and is predicted to introduce a PTC at position p.1765, 143 codons upstream of the canonical stop codon (p.1908). One of the main models of PTC driven NMD, the exon-junction complex (EJC) model, proposes that if an exon-exon junction (formed as a result of pre-mRNA splicing) is present more than 50-55 bases downstream of a stop codon, the NMD process is initiated [257,258]. The PTC introduced by the c.5294G>A (p.Trp1765*) variant is more than 50 nucleotides upstream of the last exon-exon junction, and as such would be expected to be a target for NMD; however, this has not been confirmed experimentally.

Of note, 12 of the 14 variants detected in cases and controls in this study were predicted to introduce a premature truncating codon more than 50 nucleotides upstream of the last exon-exon junction, and so would be expected to undergo nonsense mediated decay (Table 3.3-18). Two variants, both of which were detected in reference controls, were in the final exon of the protein and so could potentially produce transcripts which might evade NMD.

Alternatively, as has been observed in other genes, the presence of alternative translation start sites or differential splicing could "rescue" PTC bearing transcripts from NMD [259–262]. However, to date, a single splice isoform has been identified for *ALPK3* [6] and there is no *in vitro* or *in vivo* evidence to support generation of an alternatively translated or spliced protein product in individuals with truncating variants in this gene.

[6] https://www.proteinatlas.org/ENSG00000136383-ALPK3/tissue and http://www.ensembl.org/Homo_sapiens/Transcript/ProteinSummary?db=core;g=ENSG00000136383;r=15:84816680-84873482;t=ENST00000258888

3.4.5.1.6 Tissue/isoform specific loss of protein function

Another phenomenon, which has previously been reported to explain differences in penetrance of truncating variants in other cardiomyopathy genes (e.g. *TTN* [71] and *DSC2*[263]), relates to variants primarily affecting an isoform or exon that is not significantly expressed in cardiac tissue. However, as only one protein-encoding transcript has been reported for *ALPK3* to date[6], there is no evidence to support isoform/tissue specific loss of function.

3.4.5.1.7 Co-existing variants

Another possible mechanism of pathogenicity in cases heterozygous for truncating variants in *ALPK3*, is the existence of another functionally significant variant *in trans*. Whilst protein haploinsufficiency in itself might not be pathogenic, there may be pathogenic consequences if, due to an *in trans* missense variant, there is no "wild type" protein product being produced from the other allele. Complete analysis of the coding sequence in *ALPK3* in cases heterozygous for truncating variants did not reveal any other putative pathogenic variants. Additionally, no larger scale copy number variants (CNVs), encompassing all or part of *ALPK3* were detected.

Although pathogenic variants in the key well established HCM genes have been excluded in HCM cases heterozygous for truncating variants in *ALPK3* cohort, it is possible that there is a pathogenic variant in a more rarely associated or previously uncharacterised HCM gene. To explore this possibility, further analysis was undertaken in 1250 "biologically plausible" genes, selected on the basis of HCM Human Phenotype Ontology (HPO) terms. No putative pathogenic variants were detected in these genes in the selected cases. Further genome wide analysis may be considered; however, as these are singleton cases, the scope of these analyses is limited.

Alternatively, it is possible that the phenotype observed in our cases is more oligogenic in nature, arising from the combined effects of individually low penetrance variants in *ALPK3* and in genes encoding functionally related proteins. Although plausible, a significantly higher number of cases would be required to provide sufficient power to explore this oligogeneic model.

3.4.5.1.8 Variant penetrance

Another reasonable conclusion is that heterozygous *ALPK3* truncating variants can be pathogenic, but display reduced or incomplete penetrance. Incomplete and/or reduced penetrance is commonly observed in HCM, even with confirmed causal variants, and it is accepted that a combination of genetic and/or environmental factors can act in concert with the predominant causal variant to produce the phenotype.

Further clinical and genetic studies within HCM pedigrees with familial *ALPK3* truncating variants may clarify levels of penetrance. Clinical review of heterozygous individuals within reference control cohorts would also be valuable and could provide critical data regarding variant penetrance at the population level. This will be particularly important when interpreting the likely clinical significance of *ALPK3* truncating variants detected incidentally during genomic sequencing in individuals with no prior family history of cardiomyopathy.

3.4.5.1.9 False positive result/technical artefact

The possibility that the *ALPK3* findings in this cohort are a false positive result cannot be excluded, with technical problems relating to annotation of insertion/deletion variants as one possible source of error. However, as our case and control datasets were sequenced in a single centre and underwent combined analysis using a single analysis

pipeline, false positive results due to differences in variant annotation would not be expected.

Issues with sequencing coverage could also lead to inconsistencies in variant calling, skewing frequency estimates in our component cohorts; however, no sequence coverage issues were noted across coding regions of this gene.

3.4.5.1.10 Clinical review

Review of clinical notes from the five HCM cases heterozygous for *ALPK3* truncating variants was undertaken by Professor Hugh Watkins. An atypical phenotype, with features of both HCM and DCM was reported in one individual. No unusual or atypical clinical features were noted in the other four index cases. A family history of HCM was reported in only one proband, therefore at present there are minimal opportunities for segregation analyses in these cases. Further clinical investigations will be performed to examine the phenotype in these individuals.

3.4.5.1.11 Conclusions

While there is genetic evidence to support a causal role for heterozygous *ALPK3* truncating variants in HCM, further clinical, genetic and functional analyses will be required to investigate this finding. At present, the clinical significance of heterozygous truncating variants in the context of autosomal dominant HCM remains unclear.

3.4.5.2 DES

The results of case-control analyses in the *DES* gene were also examined. The *DES* gene is included in the Genomics England Gene Panel App HCM list and was selected for review by the ClinGen expert working group, however this gene is rarely included in clinical HCM test panels (Figure 3.2-1).

3.4.5.2.1 Gene structure and protein function

DES is located on the long arm of chromosome 2 (2q35, 2: 220,283,099-220,291,461), it contains 9 exons, and is translated into Desmin, a 470 amino acid 53kDa protein. Desmin is a class III intermediate filament protein (IF), highly expressed in skeletal, cardiac and smooth muscle cells[264]. This protein has a highly conserved 308 amino acid alpha-helical domain, comprising four helical segments (1A, 1B, 2A and 2B) linked by short non-helical segments, bound by globular N-terminal head and C-terminal tail domains[265].

In adult striated muscle, desmin homopolymers form a three-dimensional fibrous network, which surrounds the myofibrillar Z-disk connecting the myofibrils to the sarcolemma, intercalated disks, nuclei and other cytoplasmic organelles including the mitochondria [266–268]. This filament network plays a critical role in maintaining the structural and functional integrity of the myofibrils (Figure 3.4-3) [266–269].

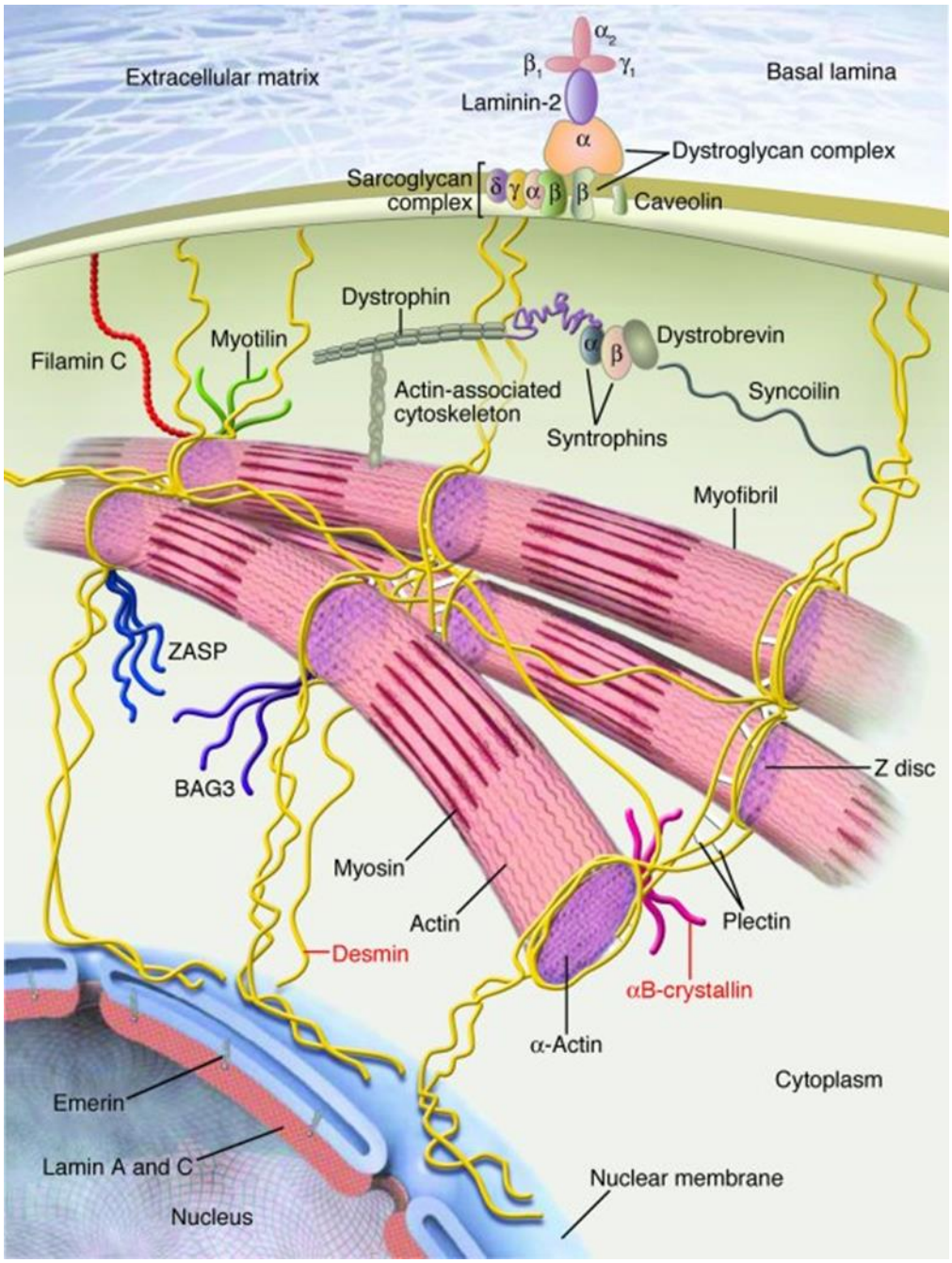


Figure 3.4-3 Molecular cytoarchitecture of a myocyte. Desmin is the main muscle intermediate filament protein. It interacts with other proteins to support myofibrils at the level of the Z disc and forms a continuous cytoskeletal IF network that maintains a spatial relationship between the contractile apparatus and other structural elements of the cell. Desmin provides maintenance of cellular integrity, force transmission, and mechanochemical signalling. Mutations in other sarcomeric and cytoskeletal proteins (plectin, myotilin, filamin C, αB-crystallin, Z band alternatively spliced PDZ-motif protein [ZASP], and BCL2- associated athanogene 3 [*BAG3*]) cause neuromuscular disorders. Reproduced with permission from [270], Copyright Massachusetts Medical Society.

3.4.5.2.2 Role in disease

Pathogenic variants in *DES* have been reported in a heterogeneous group of muscle disorders. These disorders are generally autosomal dominant, adult onset conditions; however, clinical presentation is variable and can encompass features of skeletal myopathy (myofibrillar myopathy, desmin-related myopathy, and distal myopathy), cardiomyopathy (including DCM, HCM, RCM and ARVC) and cardiac conduction disease. The cardiologic and neurologic features can appear in isolation or as a combined cardio-skeletal myopathy, and can present in any order. The histological hallmark of this disease is the intracytoplasmic accumulation of protein aggregate composed of desmin and other proteins[271]. Autosomal recessive cases have also been reported and appear to be associated with an earlier age of onset and more severe phenotype [272,273].

In a meta-analysis of published studies, Van Spaendonck-Zwarts *et al* reported that in approximately ~50% of individuals with desmin-related myopathy, both skeletal and cardiac signs were observed[274]. In 22% of cases, the cardiac phenotype was reported in isolation[274]. In individuals with a cardiac phenotype, approximately 50% reported to have cardiomyopathy. DCM was the most prevalent cardiomyopathy, reported in 17% of cases; around 13% of individuals had an unspecified cardiomyopathy, 12% were reported to have RCM, 6% HCM and 1% ARVC. Cardiac conduction disease (CCD) and/or arrhythmias were reported in 62% of cases with a cardiac phenotype, atrioventricular block (AVB) being the most frequent manifestation; two-thirds of individuals with CCD or arrhythmia also had cardiomyopathy.

The majority of reported pathogenic variants are missense changes, however small in-frame insertions/deletions [275,276] and splice site variants [270,277,278] have also been detected[279]. Most pathogenic missense variants have been reported within the 2B

domain (alpha helical rod domain), however variants have been detected in the other protein domains. Experimental studies indicate that many of these variants act in a dominant-negative fashion, disrupting protein assembly and filament network structure (Figure 3.4-4), ultimately leading to the cytoplasmic accumulation of abnormal desmin aggregates[270,272,275,280–284].

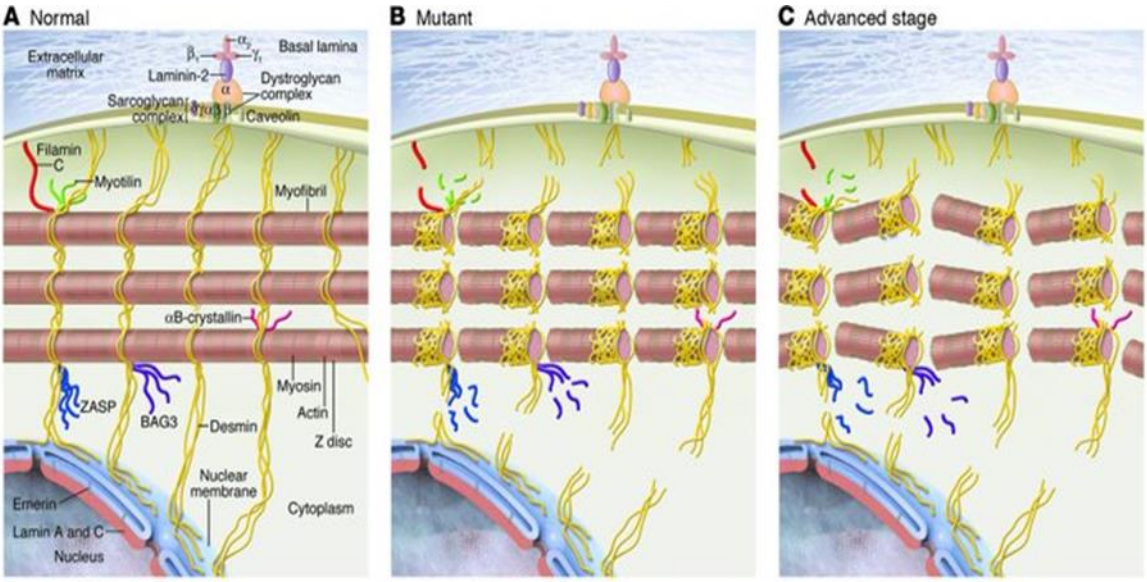


Figure 3.4-4 Schematic representation of the destructive effects of desmin mutations in muscles of patients with desminopathy.(A) Normal molecular cytoarchitecture of a myocyte (B) Mutant desmin causes disorganization of the Z discs and affects the integrity of the cellular IF network. (C) Fragmented desmin filaments form insoluble deposits that accumulate and eventually cause myopathy by disrupting the myofibrils, breaking the connections between them and upsetting the binding of the myofibrils to cellular membranes. Reproduced with permission from J Clin Invest 2009 Jul 1;119(7):1806-1813.

Loss-of-function variants (nonsense and frameshift) predicted to cause protein haploinsufficiency have been reported [273,285–287] and there is data to support homozygous or compound heterozygous loss-of-function variants in autosomal recessive desminopathy [273,285]; however, there is little evidence to support heterozygous loss-of-function variants being disease causing [285–287]. This is supported by data from

ExAC which indicates that the *DES* gene is not under constraint for loss-of-function variants (Table 3.4-3).

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	97.9	66	z = 2.00
Missense	202.0	134	z = 2.34
LoF	16.7	7	pLI = 0.00
CNV	4.1	4	z = 0.01

Table 3.4-3 Exome Aggregation Consortium (ExAC) constraint scores for *DES* gene. (http://exac.broadinstitute.org/). Missense: For synonymous and missense, Z score represents the deviation of observed counts from the expected number. Positive Z scores indicate increased constraint (intolerance to variation) and therefore that the gene had fewer variants than expected. Negative Z scores are given to genes that had a more variants than expected. LoF (loss of function): The closer pLI is to one, the more LoF intolerant the gene appears to be. pLI >= 0.9 as an extremely LoF intolerant set of genes. [138]

3.4.5.2.2.1 DCM

DES gene analysis has been performed in a number of DCM patient cohorts (Li et al. 1999 [n=44][288]; Tesson et al. 2000 [n=63][289]; Miyamoto et al. 2001 [n=265][290]; Taylor et al. 2007[n=425] [91]). From these analyses, it was estimated that ~1.5% of DCM patients may have a pathogenic variant in the *DES* gene [91,241,283,284,288,290,291]. However, for many of the variants detected in these cohorts, assertions of pathogenicity are not supported by robust segregation or functional data. Additionally, taking into account recent data from the gnomAD database, it seems that some of these variants may be benign polymorphisms (e.g. p.Ile451Met [270,288,290], p.Ser298Leu[91], p.Asp312Asn[91]. Also, in my more recent case-control analysis undertaken using data from the OMGL and LMM DCM cohorts (case n=894, ExAC control n=51128), there was no convincing evidence of an excess of rare variation in DCM cases (non-truncating variants OR 1.79 [95% CI 0.84-3.80] p=0.12, truncating variants OR 7.16 [95% CI 1.64-31.20] p=0.04).[215]. As such, it appears that current estimates of the prevalence of pathogenic *DES* variants in cohorts of DCM patients may be inflated.

Interestingly, one of these putative DCM causing variants, p.Asp312Asn, was analysed *in vitro* and reported to disrupt normal desmin cytoskeletal architecture and lead to accumulation of desmin-positive protein aggregates [91]. However, this variant has a MAF of 0.002 in the African population and at this frequency, it is highly unlikely that this is a penetrant pathogenic variant. This conflicting data highlights issues of specificity inherent to some of the functional assays used to investigate the consequences of variants in this gene.

3.4.5.2.2.2 ARVC

There is also published data relating to *DES* gene analysis in ARVC patient cohorts (Lorenzon et al. 2013 (n=91)[120]; Otten et al. 2010 (n=50)[292]). While no clearly pathogenic variants were found in these studies, case numbers were small. Data from previous case-control analysis (case n=93, ExAC control n=51128) using data from the OMGL was also limited by small case numbers (all rare variants OR 7.04 [95% CI 2.21-22.39], p=0.01).[215] Therefore, as yet, there is insufficient data to provide an accurate estimate about the contribution of *DES* variants to isolated ARVC and a larger case series will be required to examine the frequency and likely pathogenicity of *DES* gene variants in this setting.

3.4.5.2.3 DES variants in HCM

There are a number of reports in the literature indicating that HCM can be the predominant cardiac phenotype in individuals with pathogenic variants in *DES* (e.g. p.Ser13Phe [293], p.Pro419Ser [281], p.Leu392Pro [281] and in some cases the cardiac phenotype can precede the signs and symptoms of skeletal myopathy (e.g. p.Ser13Phe [293], p.Ile367Phe [281], p.Arg454Trp [282]. Therefore, *DES* would seem to be plausible HCM

gene, but to date, there do not appear to be any studies reporting its assessment in large HCM patient cohorts.

3.4.5.2.3.1 Clinical review

Review of clinical information from the individuals in this cohort heterozygous for rare *DES* variants did not reveal any evidence of skeletal myopathy, or other signs/symptoms suggestive of desminopathy. Further clinical review may be appropriate in these cases to explore the likely contribution of these *DES* variants to the phenotype in these individuals.

3.4.5.2.3.2 Analysis in larger HCM patient cohort

Further genetic analyses were undertaken to explore the putative relationship between isolated HCM and *DES*, highlighted through the BRRD case-control analyses. The *DES* gene is sequenced in all patients undergoing cardiomyopathy panel testing within the OMGL laboratory, but data from this gene is only analysed in individuals referred for DCM and ARVC. This meant that previously unanalysed *DES* gene data was available from 1287 patients referred for the HCM gene panel test; although it should be noted that a proportion of individuals in this cohort were included in the initial case-control analyses.

In total 6/1287 (0.47%) HCM cases in whom *DES* variant data was available, were found to have rare missense or in-frame deletion variants. Statistical analysis indicated that there was no difference in rare variant frequency between this OMGL HCM case cohort and the BRRD reference controls (6/1287 [0.47%] vs 36/7253 [0.47%], OR 0.93 [95% CI 0.3-2.24], p=1.00). Five of these variants were classified as VUS and one as likely benign (Table 3.4-4).

Therefore, extended analysis in a larger unselected HCM cohort (n=1287) found no evidence to suggest that *DES* is a key causal HCM gene. This raises the possibility that the result in the BRRD case-control cohort is a false positive finding. Alternatively, it is possible that the increased frequency of rare *DES* variants observed in the BRRD cohort reflects enrichment of this cohort for non-sarcomeric cardiomyopathy.

c.	p.	Domain	gnomAD	ClinVar	Classification
c.187G>T	p.(Ala63Ser)	Head	Not listed	Not listed	VUS
c.330_338delGGAGCTGCA	p.(110_113del)	Coil 1A	Not listed	Not listed	VUS
c.832C>T	p.(Arg278Trp)*	Coil 2A	SA: 1/30776 (MAF 0.00003) Total 1/245308 (MAF 4x10-6)	VUS	VUS
c.893C>T	p.(Ser298Leu)*	Linker 2	African:6/23998 (MAF 0.0002) Other: 1/6458 (MAF 0.00015) Euro(NF):14/126510 (MAF 0.0001) Latino:1/34416 (MAF 0.000029) Total: 22/276948(MAF 0.000079)	Conflicting (LPx2,VUSx2)	Likely benign
c.1243C>T	p.(Arg415Trp)	Tail	Total: 6/276182 (MAF 0.00002)	VUS	VUS
c.501G>T	p.(Glu167Asp)	Coil 1B	Not listed	Not listed	VUS

Table 3.4-4 OMGL HCM cohort *DES* gene findings. Variants annotated against genome build GRCh37, Ensembl transcript ID: ENST00000373960. * = same case described in BRRD case-control analyses. c. =HGVS coding nucleotide. p. =HGVS protein nomenclature (http://www.hgvs.org/mutnomen/). ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Genome Aggregation consortium (gnomAD) database (http://gnomad.broadinstitute.org/). Euro(NF)=European (Non-Finnish) population. SA=South Asian population. VUS=variant of uncertain significance. LP=Likely pathogenic.

3.4.5.2.4 Conclusions

At present, there is little genetic or clinical evidence to support inclusion of this gene in HCM diagnostic test panels.

3.4.5.3 KCNQ1

A possible case excess of rare missense variants in the *KCNQ1* gene was also investigated; while not achieving the Bonferroni corrected level of significance (p=0.0008), this gene was nonetheless a notable outlier. The *KCNQ1* gene is included in the Genomics England Gene Panel App HCM list and had been selected for review by the ClinGen HCM expert working (Table 3.2-4).

3.4.5.3.1 Gene structure and protein function

KCNQ1 encodes the alpha subunit of the voltage-gated cardiac potassium channel (KvLQT1, Kv7.1) which plays a critical role in the repolarization phase of the cardiac action potential. This gene encompasses ~400kb and is composed of 16 exons; it is located in an imprinted region of chromosome 11 and exhibits tissue-specific imprinting. Preferential expression from the maternal allele is observed in some tissues and bi-allelic expression in others [294]. The main protein isoform consists of 676 amino acid residues, which form six transmembrane domains, a pore loop with a typical potassium-channel pore motif (GYGD), bound by intracellular NH_2 and COOH terminal domains (Figure 3.4-5). Four KCNQ1 α-subunits combine with a number of KCNE1 β-subunits to form a functional channel [295].

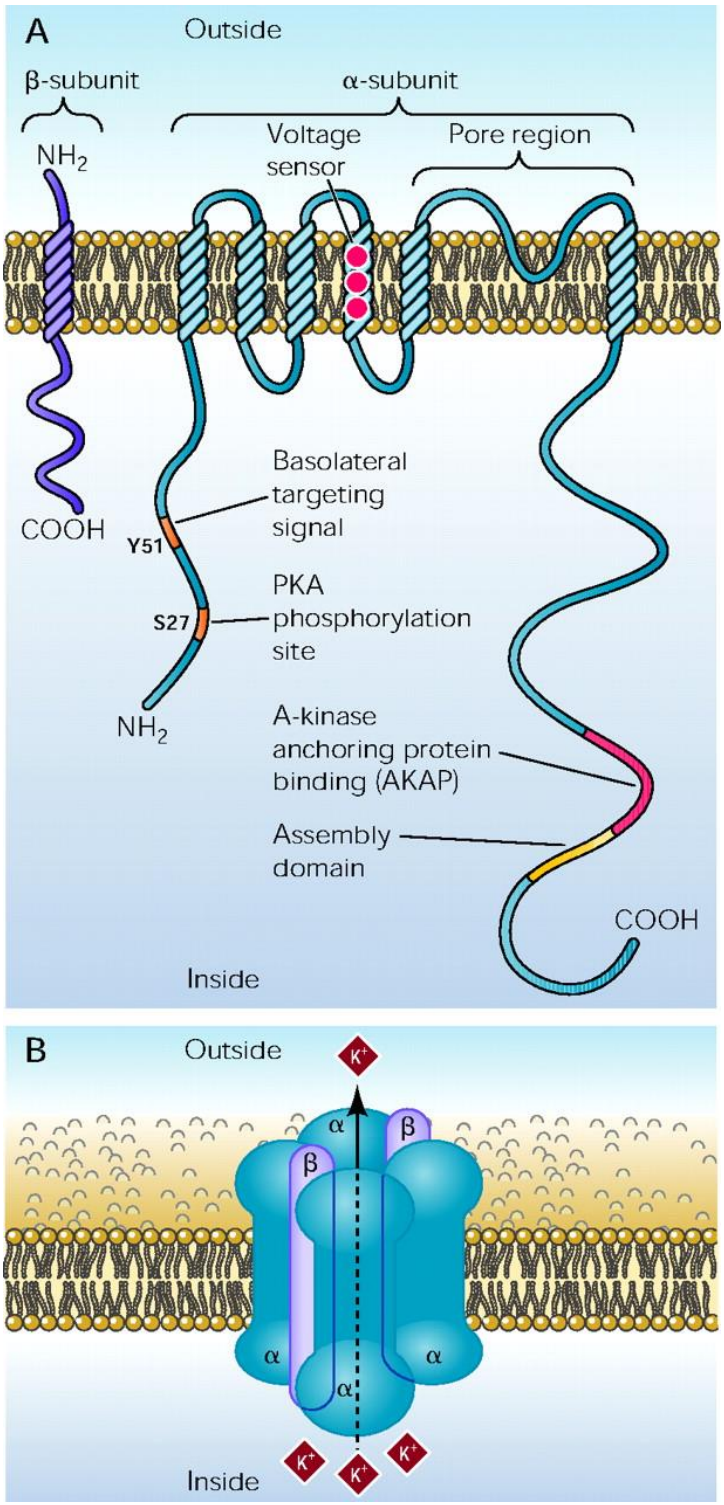


Figure 3.4-5 Topology and channel architecture of KCNE and KCNQ1 proteins.
A: topology of the KCNE proteins as well as the delayed rectifier potassium channel KCNQ1 with indications of some of the domains important for regulation of the channel. All five KCNE proteins hold one transmembrane domain but have varying lengths of NH_2 and COOH termini. B: KCNQ1/KCNE channel architecture. Four KCNQ1 α-subunits assemble to form the basic channel. Co-assembly with a number of KCNE β-subunits (probably between two and four KCNE proteins) leads to an α-subunit-specific change in the current characteristics. Image reproduced with permission from [295].

3.4.5.3.2 Role in disease

Pathogenic mutations in the *KCNQ1* gene are one of the most common causes of LongQT syndrome (LQTS), an inherited cardiac channelopathy characterised by QT prolongation (>440ms), syncope, seizures and SCD in the setting of a structurally normal heart [296]. Around 35-40% of LQTS patients will have a pathogenic variant in *KCNQ1* [297,298].

The majority of previously reported pathogenic mutations are missense variants [223] leading to loss of protein function via either a dominant negative mechanism or protein haploinsufficiency. Pathogenic variants have been reported to affect protein trafficking, assembly and regulation; the severity of phenotype being related to the degree of residual channel function [299].

Truncating variants are less common, but considered highly likely to be pathogenic [223]. Homozygous loss-of-function mutations in *KCNQ1* cause an autosomal recessive condition, Jervell and Lange-Neilsen syndrome (JLNS), which is associated with congenital profound bilateral sensorineural hearing loss and severe QT prolongation (>500ms).

More rarely, missense variants in *KCNQ1* have been reported to cause Short QT syndrome; here variants confer a gain-of-function effect on the protein, leading to shortened cardiac repolarisation, which can predispose to supraventricular arrhythmias and SCD [300,301]. Gain-of-function [302,303] and loss-of-function [303] missense variants have also been reported to cause familial atrial fibrillation.

Interpretation of missense variants in this gene is confounded by the high background rate of rare missense variation[223,224,304,305] and variable penetrance. Comparative case-control studies have refined the key functional regions and residues within the protein,

enabling better discrimination between benign and pathogenic variants[224,225,306–308]. Collectively, these data have enhanced variant interpretation in this gene, however some uncertainty remains and a high proportion of variants are classified as VUS[223].

3.4.5.3.3 KCNQ1 variants in HCM

Given current understanding of the pathophysiology of variants in *KCNQ1*, it is difficult to conceive how loss or gain of function variants in this gene could cause a structural cardiomyopathy mimicking HCM. Review of the literature found no evidence to suggest that *KCNQ1* is a causal gene in HCM. Indeed, in a study of 847 HCM patients undertaken by Lopes *et al* in 2014, rare *KCNQ1* gene variants were detected in 0.7% of cases[151]; this is the same as the frequency observed in the BRRD non-HCM controls and in ExAC.

No convincing evidence was found to suggest that any of the rare *KCNQ1* variants detected in this HCM cohort were pathogenic. Analysis using the Q1 VarPred prediction tool annotated two variants as having a high probability of pathogenicity (p.Ser92Cys [P=0.934] and p.Val241Ile [P=0.80]). This tool utilises data from 107 functionally characterised pathogenic *KCNQ1* variants and performs a quantitative analysis based on patterns of sequence conservation and variant distribution[225]. As this tool has been designed specifically for *KCNQ1*, it is reported to be more effective in predicting the functional impact of variants in this gene. However, the positive predictive value of this tool will be lower in individuals who do not have high pre-test probability of LQTS. Also, it is acknowledged that the functional impact of a single variant does not necessarily translate to a clinical phenotype, and that this evidence alone is not sufficient to confirm pathogenicity.

Although there is little genetic or biophysical evidence to support a causal role for rare *KCNQ1* variants in HCM, there is of course a possibility of co-inheritance of both *KCNQ1*-

related arrhythmia and HCM. One of the HCM cases in this study, who was heterozygous for the *KCNQ1* c.1252-3dupT variant, was reported to have a prolonged QT interval; however, this individual had severe hypertrophy and left ventricular outflow tract obstruction (LVOTO), factors which have previously been independently correlated with QT prolongation in a large HCM cohort[309]. Furthermore, on subsequent review of this variant by the OMGL, it was re-classified as 'likely benign'. Therefore, a dual HCM and LQTS diagnosis was considered unlikely in this case.

3.4.5.3.4 Conclusions

In the absence of any published data supporting a link between rare variants in *KCNQ1* and HCM, and with no strong genetic or clinical evidence to suggest that the *KCNQ1* variants detected in HCM patients in this cohort are pathogenic, the difference in rare variant frequency observed most likely reflects chance variation, rather any real difference between these populations.

3.4.5.4 TCAP

Very few rare variants were detected in the *TCAP* gene in cases and controls (1/230 and 15/7253 respectively); notably only one truncating variant was detected and this was found in an HCM case. Consequently, a possible casual role for *TCAP* truncating variants in HCM was investigated. The *TCAP* gene is included in several commercial clinical HCM gene panel tests, and both Genomics England and ClinGen include this gene in their HCM list.

3.4.5.4.1 Gene structure and protein function

TCAP is located on the long arm of chromosome 17 (17q12, 17:37821553-37821763), it is a small gene comprising only 2 coding exons and is highly expressed in skeletal and

cardiac muscle [264]. *TCAP* encodes Telethonin, a 167 amino acid, 19kDa sarcomeric Z-disk protein [310], which plays a critical role in titin filament assembly and Z-disk anchoring[311]. Telethonin protein binds to two titin proteins via their N-terminal Z1-Z2 domains, anchoring these molecules to the Z-disk (Figure 3.4-6); it has also been shown to interact with other Z-disc proteins such as muscle LIM protein (MLP, encoded by *CSRP3*) [312], skeletal muscle ankyrin repeat protein (sAnkyrin encoded by *ANKRD2*) [313], filamin-, actinin- and telethonin-Binding Protein (FATZ encoded by *MYOZ1*) [314], and potassium channel β-subunit (MinK encoded by *KCNE1)* [315], (Figure 3.4-6).

3.4.5.4.2 Role in disease

Missense variants in *TCAP* have been reported in a very small number HCM cases[54,167], however evidence supporting their pathogenicity is negligible. Homozygous truncating variants within *TCAP* have been reported in individuals with autosomal recessive limb-girdle muscular dystrophy type 2G (LGMD2G)[210,316–322] and congenital muscular dystrophy[323]. Of note, cardiomyopathy is not a reported phenotype of LGMD2G.

Complete loss of protein function is considered to be the main disease mechanism in these skeletal disorders[324,325]. Telethonin deficiency has been observed in muscle biopsies from the majority of reported cases[316,317,319,321–323]; however, sarcomeric localisation of truncated protein product has been noted in others [318]. I*n vivo* studies in mice [324] and zebrafish support a loss-of-function effect [325].

3.4.5.4.3 TCAP variants in HCM

In previous studies, individuals heterozygous for these truncating variants did not appear to show any signs or symptoms of skeletal or cardiac muscle disease[316–323].

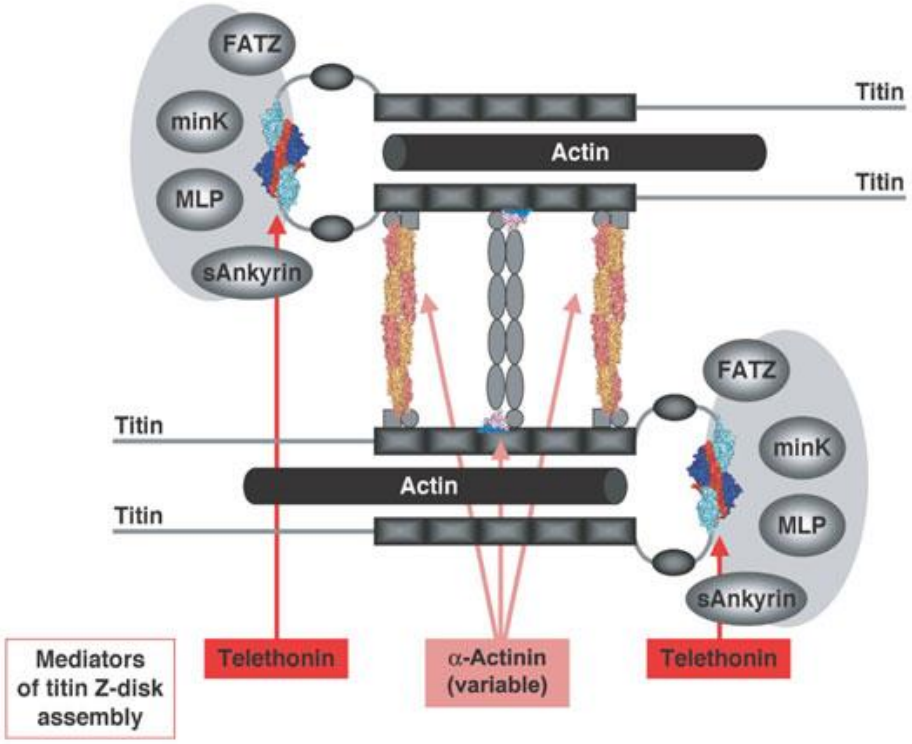


Figure 3.4-6 Model outlining the involvement of the titin–telethonin complex in the architecture of the sarcomeric Z-disk.
Titin filaments are assembled by a dual Z-disk bridging system, by α-actinin rods on a variable number of titin Z-repeats (three bridges are shown), and by telethonin by means of the N-terminal immunoglobulin domains Z1 and Z2. The titin N-terminus/telethonin complex forms a core that interacts with several ligands both inside and outside the sarcomeric Z-disk, including MLP, sAnkyrin, minK and FATZ. Image reproduced with permission from [311]

The scarcity of variants in this gene observed in this cohort, and in other large reference cohorts, likely reflects its small size rather than evidence of functional constraint. On the basis of published clinical case studies and results of *in vitro* and *in vivo* investigations, there is little evidence to suggest that a heterozygous truncating variant, which is predicted to result in protein haploinsufficiency, would cause hypertrophic cardiomyopathy. Therefore, at present, the c.66G>A (p.Trp22*) *TCAP* variant is classified as a VUS.

3.4.5.4.4 Conclusions

On the basis of current evidence there appears to be little indication for including the *TCAP* gene in HCM clinical test panels.

3.4.6 Variant level analyses

Although, in the vast majority of genes analysed, there was no evidence to suggest that there was a difference in rare variant frequency in HCM cases compared to reference controls, it is of course possible that some of the rare variants detected in these genes are disease-causing. Therefore, in addition to case-control analyses, the evidence supporting variant pathogenicity was reviewed and each classified according to clinical guidelines.

The results of these investigations illustrate the difficulties of variant interpretation in genes with limited evidence of causality. For the majority of variants, there was insufficient evidence to either confirm or exclude variant pathogenicity; as such, of the 256 rare variants detected in cases, 95% of variants were classified as VUS. As illustrated in the previous sections, in many cases the link between the gene and disease is a tentative one, and the mechanisms of pathogenicity ill-defined; therefore, there is no prior knowledge on which to base variant interpretation.

Unsurprisingly, given its large size (33,000 amino acids, coding sequence ~100kb), 35% of the VUS detected were in the *TTN* gene. Levels of rare truncating and non-truncating variants were the same in our HCM case series compared to reference controls (p=0.387, OR 0.9978 [95% CI 0.7547-1.13143]). Additionally, no evidence was found to suggest that any of these variants detected were in the regions of this gene that are considered to be enriched for pathogenic variants[68,69]. These findings support the results from

previous studies [67,210] and confirm that analysis of the *TTN* gene is not indicated for individuals with HCM.

Only one variant, detected in a single proband, was considered to have sufficient evidence to support pathogenicity. This splice donor site variant in *FLNC* gene variant (c.6004+2T>C), met the ACMG criteria for "Likely pathogenic" classification. *FLNC* splice site variants have previously been detected in families with dilated and restrictive cardiomyopathy characterised by a strong arrhythmogenic phentoype[7]. RNA studies have been initiated to examine the consequences of this variant *in vitro*. Further clinical, genetic and functional investigations will be required to explore the clinical significance of this finding.

3.4.7 Case level analyses

Excluding the *TTN* gene from analysis, 54% of index cases were reported to have at least one VUS and 19% had more than one VUS. Including these genes in clinical test panels would have a significant impact on patient management, as although not considered clinically actionable, variants of uncertain significance are generally included in the clinical report, and often elicit further clinical and genetic family studies. These can be costly both in terms of time and clinical resources, and often fail to allow reclassification, a particular problem in conditions such as HCM that can present in later life and show incomplete penetrance.

Based on information provided at referral, 68% of the individuals in this cohort do not have a family history of HCM; therefore segregation analysis cannot be performed in these cases. Even in the families where segregation study is possible, in most the value is limited by small number of affected relatives. While evidence of non-segregation is useful in excluding a variant being a fully-penetrant pathogenic variant, evidence of

segregation is unlikely to provide sufficient evidence to confirm pathogenicity (LOD ≥3) [326].

Of greater concern is the potential for erroneous interpretation and the impact of false positive results for at risk family members leading to misdiagnosis and inappropriate therapy, or false negative results, leading to inappropriate discharge from follow-up.

3.4.8 Limitations

3.4.8.1 Cohort size

A number of factors influence power calculations, including cohort size, the number of genes being analysed, rare variant frequency in those genes in the background population, the population prevalence of disease, variant penetrance, and the expected fold increase in rare variants in cases and controls.

Estimates of disease prevalence are considered to be relatively well established [10]. However, the frequency of very rare variants in each gene is influenced by several factors, including gene size, sequence context, and levels of constraint for different classes of variant, therefore no single estimate will be relevant for all genes. [204,327] In Chapter 2, the mean proportion of controls with a rare variant was 0.004, therefore this value was used in power calculations, along with other values (0.01 and 0.001), to provide a range of estimates (Table 3.2-2). The proportion of controls with a rare variant was higher in the genes included in these analyses (Table 3.4-5).

Genes	Variants	Mean	Median	SD	Min	Max
60 (inc. *TTN*)	All rare	0.02	0.01	0.05	0.0004	0.4
59 (ex. *TTN*)	All rare	0.01	0.01	0.02	0.0004	0.15
46 (inc. *TTN*)	Truncating	0.001	0.0006	0.003	0.0001	0.01
45(ex. *TTN*)	Truncating	0.0009	0.0006	0.001	0.0001	0.008

Table 3.4-5 Proportion of controls with rare variants in genes analysed. Mean =Mean proportion of cases with a rare variant in analysed genes. SD=standard deviation. Min=minimum. Max=maximum. Calculations were undertaken including (inc. *TTN*) and excluding (ex. *TT*N) *TTN* to ensure that the high levels of rare variation in this gene did not skew the results. Data shown for all rare variants and for truncating variants only.

Pre-study power analyses indicated that this cohort was sufficiently powered to validate genes where there was a 10-fold increased burden of rare variants in HCM cases compared to BRRD controls. Analysis on this scale would identify a significant excess of variants in most of the confirmed HCM genes. However, these analyses were not expected to have the power to detect a difference in the proportion of cases and controls with a rare variant in genes where only a very small fraction of the variants are pathogenic.

Post-hoc power calculations, based on the observed rare variant frequency in each individual gene in this study, indicate that these analyses were powered to detect a 10-fold increased burden in majority of the genes, and a 5-fold increased burden in around 50% of genes.[7] Data from a larger number of HCM cases would inevitably increase the power of these analyses and allow more definitive characterisation of the relationship between these genes and HCM (Table 3.4-6).

[7] Analyses performed by Prof Martin Farrall.

Control n	Significance level (60 genes)	Disease prevalence (1/500)	Proportion of controls with rare variant	5-fold No. cases required to achieve ~80% Power	10-fold No. cases required to achieve ~80%Power
7253	0.0008	0.002	0.01	~300	~130
7253	0.0008	0.002	0.004*	~750	~290
7253	0.0008	0.002	0.001	3000	1150

Table 3.4-6 Power Calculations: Size of case cohort required to achieve 80% Power
^Significance level p=0.05, Bonferroni correction for 60 genes p=0.0008.
Proportion of controls carrying a rare (ExAC MAF <0.0001) variant, 0.004 figure based on mean value from previous analysis of HCM (Chapter 2).
Mean proportion of controls with a rare variant in these analyses: Including *TTN* 0.02, Excluding *TTN* 0.01.

3.4.8.2 Gene selection

Extensive review and collation of data from a range of resources highlighted a large number of genes implicated in HCM. However, as detailed above, given the size of the cohort, allowing for multiple statistical testing of a very large number of genes would have reduced the power of these analyses. Therefore, taking into consideration these inherent limitations, rather than extend analysis to a large number of novel genes or perform genome wide analyses, a more targeted strategy was adopted, focusing on a smaller set of clinically relevant genes, in particular those included in current gene curation efforts.

3.4.8.3 Clinical heterogeneity

The majority of individuals recruited to this project had a clinical diagnosis of HCM made within an expert cardiac clinic. As such, the phenotypes in this cohort reflect the clinical heterogeneity seen in routine Inherited Cardiac Clinics (ICCs), rather than being

a highly selected cohort enriched for cases considered to have a high probability of autosomal dominant cardiomyopathy.

It is possible that some of the reference controls used in this study have HCM, or another inherited cardiomyopathy (e.g. DCM and ARVC). However, the frequency of these disorders in this control cohort is not expected to exceed the population prevalence of these disorders and therefore is not considered to be a significant limitation in these case-control analyses.

3.5 Conclusions

In this chapter, the results of case-control and rare variant analysis in 60 genes implicated in HCM, in a large sarcomere gene negative HCM patient cohort are reported. Through these analyses of 230 HCM cases, a single clinically actionable result was found (a splice site variant in the *FLNC* gene) and one gene was identified as a probable disease-causing gene (*ALPK3*).

In 59 of the 60 genes analysed, no robust evidence was found to suggest that rare variants were in excess in HCM cases compared to controls. This suggests that these are not prevalent causes of HCM. This does not exclude the possibility that a small proportion of rare variants in these genes could be disease causing; however, it does indicate that that the majority are not highly penetrant pathogenic variants. Without specific biophysical or functional evidence allowing discrimination between benign and pathogenic variants, extensive segregation data, or cases with additional clinical features (i.e. specifying a given "phenocopy"), rare variants in these genes are uninterpretable. This was evident in that the vast majority of rare variants detected in cases were classified as VUS. In the absence of evidence supporting a prominent role for these genes

in HCM, and taking into consideration issues of variant interpretability, these findings suggest that there is no clinical utility in including these genes in HCM clinical test panels.

For current clinical test panels focused on well-established HCM genes, approximately one-third of variants are classified as VUS and around 15% of patients are reported to have at least one VUS[27]. However, as shown in Chapter 2, even variants classified as VUS appear to be in excess in cases compared to controls, therefore it seems likely that the majority are pathogenic, so over time the clinical yield from these core genes is expected to increase and the proportion of VUS fall.

In contrast, the outcomes of testing large numbers of un-validated genes were apparent in this cohort, with over 99% of variants classified as VUS and a substantial increase in the proportion of patients reported to have one or more VUS. Importantly, in these 59 genes, an excess of rare variants was not found in cases, indicating that the vast majority of these VUS will not be disease causing.

There is increased awareness of the issues surrounding variant interpretation, particularly in the context of broad based gene analyses[328–330], and a drive towards more consistent and evidence based approaches to classification[128,142]. Thus, the results of these analyses provide much needed empirical data which will inform current gene curation initiatives, such as those led by the NIH Clinical Genome Resource (ClinGen)[331,332] and Genomics England. These efforts aim to confirm the definitive causal genes in each disorder and enable robust evidence based selection of genes for clinical testing. Evidence based approaches to genetic testing are critical in a clinical setting, and will become increasingly important as clinical exome and genome sequencing become more mainstream[333,334].

In efforts to define the causal genes in heterogeneous disorders such as HCM, whole genome approaches are likely to have limited utility when undertaken on a case-by-case basis. However, as illustrated in this study, consolidated analysis of genomic sequencing data from large case series can enable comprehensive assessment of genes and variants implicated in disease. This will counteract the recent trend of inclusion of putative disease-genes despite insufficient evidence and ensure that genetic testing strategies are optimally effective and clinically relevant.

Chapter 4. Exploring the population prevalence of pathogenic variants in inherited cardiac condition genes.

4.1 Introduction

Until recently, much of our understanding about inherited cardiac conditions (ICCs) has come from clinical and genetic studies in individuals and families presenting with disease.[190,335] However, as studies begin to explore ICC gene variants in population based cohorts, fundamental questions relating to the prevalence and penetrance of ostensibly pathogenic variants are emerging [140,141,336–338]. These studies highlight the need for further research to evolve our understanding of the natural history and impact of pathogenic ICC gene variants in the wider population.

Previous studies of cardiomyopathy and arrhythmia genes in population based cohorts have revealed an unexpectedly high prevalence of very rare variation, including a number of variants previously reported as pathogenic mutations[141,339–341]. While some prevalence estimates are likely over-inflated due to inclusion of variants which, on the basis of current evidence, are not considered to be pathogenic [141,336], even in studies focusing solely on variants for which there is stronger evidence supporting pathogenicity, when taking into consideration current estimates of disease prevalence, the frequencies are higher than expected. These findings indicate that, for some pathogenic ICC gene variants, estimates of penetrance derived from family-based segregation studies do not extrapolate to the wider population.[342,343]

A number of studies have utilised existing clinical data collected for research purposes[344,345] or electronic medical records (EMR)[341,346] to examine the penetrance of

putative pathogenic ICC variants detected in population-based cohorts. These investigations have reported penetrance estimates ranging from 0% to 50%. However, these estimates are based on evidence of related clinical features extracted from pre-existing records. While these datasets could confirm the presence of clinically overt disease, they have not been collected specifically for the purpose of diagnosing ICCs, and are expected to have low sensitivity in this setting. Therefore, while these studies contribute much needed evidence, further research is required to obtain more precise estimates of variant penetrance.

A more refined understanding of the population prevalence and penetrance of pathogenic ICC gene variants is particularly important, as many of these genes appear in the American College of Medical Genetics and Genomics (ACMG) recommendations as genes which should be considered for testing in all individuals referred for Clinical Exome and Genome Sequencing[143,347]. While these guidelines imply a duty for molecular laboratories and clinicians to seek and report pathogenic variants in specified ICC genes (referred to as secondary findings, SFs)[143,347], others argue that at present, there is insufficient knowledge to support variant interpretation in this setting [341,345,348–350]. Critically, without robust evidence relating to gene or allele-specific penetrance in the wider population, the clinical validity and utility of testing these genes in individuals with no known personal or family history of disease remains unclear[348,349].

In this study, data from two large cohorts, unselected for personal or family history of ICCs, were utilised to examine the population prevalence of stringently selected pathogenic ICC gene variants, and to examine the hypothesis that these variants are present in individuals without clinically overt disease at a higher than expected frequency considering current population estimates of disease prevalence.

The prevalence of previously reported putative pathogenic variants in two key cardiomyopathy genes (*MYH7* and *MYBPC3*) was explored using data from up to 150,000 participants genotyped on the UK Biobank array, with linked clinical data to ascertain the presence of ICC-related phenotypes.

In an alternative approach, analysis of genomic sequencing data from 7878 individuals recruited to the NIHR BioResource for Rare Disease (BRRD) project enabled detection of pathogenic variants in 16 inherited cardiomyopathy and arrhythmia genes. The results of these analyses were used to initiate follow-on clinical studies to directly examine variant penetrance.

The aim of these analyses was to provide estimates of the population prevalence and penetrance of pathogenic variants across key ICC genes, and contribute much needed evidence to inform variant interpretation in the context of secondary findings detected during clinical exome and genome sequencing.

4.2 Methods

4.2.1 UK BioBank cohort (UKBB)

Analyses were undertaken to explore the feasibility of using genotype array data to explore very rare variant frequency in population based datasets. The UK Biobank Axiom® Array includes 1,710 synthetic DNA markers designed to detect a series of previously reported rare variants within cardiac disease predisposition genes. This includes a subset of markers which enable genotyping of putative pathogenic variants in the *MYBPC3* and *MYH7* genes. Access to UKBB array genotype data, along with specified demographic and clinical datasets, was approved under UKBB application number: 11223 - Population prevalence of putative cardiomyopathy-causing genetic variants [351].

Minor allele frequency (MAF) data from up to 150,000 UKBB participants was obtained for variants annotated as "Rare variants in cardiac predisposition genes" within the *MYBPC3* and *MYH7* genes. Analysis was then restricted to known variants classified as either "Highly likely to be pathogenic" or "Likely to be pathogenic" by the Oxford Medical Genetics Laboratory, Oxford, UK (OMGL) and Partners Laboratory for Molecular Medicine, Boston, USA (LMM), clinical laboratories. MAF data for these variants was compared between the clinical cohorts, UKBB and the Exome Aggregation Consortium.[138] Genotype cluster plots were generated to enable visualisation and assessment of the quality of the rare variant calls; these plots were created using the ScatterShot application written by researchers within the Wellcome Centre for Human Genetics, University of Oxford (WCHG)[352]. Fisher's exact test and Odds Ratio (OR) analyses were undertaken to compare differences in MAF between cohorts. These analyses were undertaken in R version 3.2.0. Access to demographic information (e.g. gender, date of birth, age at sampling) and requested clinical datasets, including all cardiac related clinical datasets, were available for UKBB participants.

4.2.2 Bioresource rare disease cohort (BRRD)

Analysis was undertaken using de-identified whole genome sequencing data from the BRRD cohort (20161012 release). This included whole genome sequencing (WGS) data from 7878 individuals (6446 unrelated probands) from non-HCM cohorts (Table 4.2-1). With the exception of the Genomics England RD Pilot project, individuals in each of the component BRRD cohorts were recruited either on the basis of a specific rare disease diagnosis or were unaffected relatives. WGS was undertaken by Illumina on behalf of the NIHR Bioresource Rare Disease (BRRD) project; methods are detailed in Chapter 3.

Code	Project	Number of individuals (n=7878)
A	Bleeding and Platelet Disorders (BPD)	857
B	Ehlers Danlos and ED-like Syndromes	15
C	Genomics England RD Pilot	1965
D	UK Biobank - Red Blood Cell Extreme Traits	0
E	Pulmonary Arterial Hypertension (PAH)	765
F	Primary Immune Disorders	1277
G	Specialist Pathology: Evaluating Exomes in Diagnostics (SPEED)	1173
H	Thrombogenomics (BPD)	55
K	Bleeding and platelet disorders with Autism Spectrum Disorder	62
L	Stem Cell and Myeloid Disorders	139
M	Primary Membranoproliferative Glomerulonephritis	133
N	BPD Co-segregation	25
P	Controls for WGS protocol testing	5
R	Multiple Primary Malignant Tumours	376
S	Steroid Resistant Nephrotic Syndrome	161
T	Intrahepatic Cholestasis of Pregnancy	140
V	Neuropathic Pain Disorders	113
W	Pulmonary Arterial Hypertension (PAH)	373
Y	Cerebral Small Vessel Disease	184
Z	Leber Hereditary Optic Neuropathy	59

Table 4.2-1 BRRD component cohorts. Code relates to the BRRD project specific sample ID. Number of individuals= all sequenced individuals in component cohorts; this figure includes related individuals. In total there were 7877 probands within this data release, of which 6,446 were unrelated probands.

4.2.2.1 BRRD cohort genes analysed

Confirmed autosomal dominant ICC genes, which account for greater than 1% of cases referred for clinical genetic testing, and in which there is a well-established framework for investigating variant pathogenicity, were included in these analyses (Table 4.2-2). The *TTN* gene was not included as at the time of this study there was not considered to be a robust framework on which to base variant interpretation, and there were inconsistences in variant interpretation between clinical laboratories.

Inherited Cardiac Condition (ICC)	Genes selected for analysis
Hypertrophic cardiomyopathy (HCM)	*MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1*
Dilated cardiomyopathy (DCM)	*MYH7, TNNT2, TNNI3, LMNA, DSP*
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	*DSC2, DSG2, DSP, PKP2*
Long QT syndrome (LQTS)	*KCNQ1, KCNH2, SCN5A.*

Table 4.2-2 Genes selected for analysis

4.2.2.2 BRRD cohort variants selected for analysis

The aim of these analyses was to identify ICC gene variants detected in the BRRD non-HCM cohort for which there is strong evidence supporting pathogenicity. As such, analyses focused on variants classified as either "Highly likely to be pathogenic" or "Likely to be pathogenic", by two major diagnostic laboratories experienced in inherited cardiac disorders (OMGL and LMM). In this chapter, these variants are referred to as known pathogenic variants (KPV). Additionally, predicted truncating variants (nonsense, frameshift, splice acceptor/donor site variant) in genes where this class of variant is considered likely/highly likely to be pathogenic were also included (e.g. *MYBPC3*, *LMNA*, *PKP2*, *KCNQ1*, *KCNH2* and *SCN5A*). From here on, these variants are referred to as expected pathogenic variants (EPV).

4.2.2.3 BRRD cohort analysis protocol

A BED file was created with the co-ordinates for all genes undergoing analysis (Table 4.2-3). A range of open source tools were used to extract and filter for rare variants (Table 4.2-4), the commands for each of these steps were combined into a Bash script.

chr	start	end	gene	ENGS	ENST
15	35080297	35088340	*ACTC1*	ENSG00000159251	ENST00000290378
18	28646007	28682378	*DSC2*	ENSG00000134755	ENST00000280904
18	29078006	29128971	*DSG2*	ENSG00000046604	ENST00000261590
6	7541575	7586717	*DSP*	ENSG00000096696	ENST00000379802
7	150642049	150675403	*KCNH2*	ENSG00000055118	ENST00000262186
1	156052337	156109880	*LMNA*	ENSG00000160789	ENST00000368300
11	47350957	47376252	*MYBPC3*	ENSG00000134571	ENST00000545968
14	23879947	23906894	*MYH7*	ENSG00000092054	ENST00000355349
12	111348623	111358526	*MYL2*	ENSG00000111245	ENST00000228841
3	46899357	46923659	*MYL3*	ENSG00000160808	ENST00000395869
12	32943679	33049774	*PKP2*	ENSG00000057294	ENST00000340811
3	38589548	38691164	*SCN5A*	ENSG00000183873	ENST00000413689
19	55663135	55669141	*TNNI3*	ENSG00000129991	ENST00000344887
1	201328136	201346893	*TNNT2*	ENSG00000118194	ENST00000367318
15	63334831	63364114	*TPM1*	ENSG00000140416	ENST00000403994
11	2465914	2870339	*KCNQ1*	ENSG00000053918	ENST00000155840

Table 4.2-3 Gene co-ordinates (Genome build GRCh37) and Ensembl reference IDs[353].

Step	Tool used
Extract variant data from regions of interest (genes) from the BRRD merged-VCF files (release 20161012)	Tabix
Generate a list of the sample IDs that are listed in this VCF file	Bcftools
Generate a list of case and control Sample IDs	Grep
Annotate variants in VCF file	SNPEff
Separate the different annotations of each variant for the different transcripts onto a new line	SNPEff perl script
Separate the VCF into case and control VCF	SNPSift
Compress genotypes (remove individual sample genotype data from VCF and create a hom/het/missing count for each variant)	SNPSift
Filter variants (PASS filter, ExAC MAF < 0.0001, ENGS ID, Canonical ENST ID, impact predicted to be HIGH or MODERATE.	SNPSift
Extract fields of interest	SNPSift

Table 4.2-4 Steps in analysis

The output of these analysis steps was a list of rare (ExAC MAF <0.0001) protein coding variants. Variants were annotated using both Variant Effect Predictor [354] and SNPEff[355]. This list of variants was filtered to select KPVs and EPVs. For each variant the evidence supporting variant pathogenicity was independently reviewed by two investigators (Kate Thomson and Mike Mackley). The ClinGen Pathogenicity calculator tool [356] was used to access the range of resources used to explore variant pathogenicity, assign ACMG variant classification criteria[128] and record variant classifications. Where discrepancies in classification arose between either the original and revised classification, or between investigators, a third opinion was sought from the OMGL clinical laboratory to reach a consensus classification.

Information on age, gender, familial relationships and BRRD component cohort was examined to determine the age of individuals with a candidate variant, whether individuals with the same variant were related, and if there was any reported cardiac phenotype.

For KPV and EPV detected in both clinical cases and individuals in the BRRD cohort, Fishers exact test and OR analyses were undertaken to determine whether there were significant differences between cohorts. These analyses were undertaken in R version 3.2.0.

4.3 Results

4.3.1 UKBB cohort

A total of 825 variants annotated as "Rare variants in cardiac predisposition genes" within the *MYBPC3* and *MYH7* genes were included on the UKBB Axiom array. Of these, 139 had been reported as pathogenic or likely pathogenic by either the OMGL or LMM

clinical laboratories; UKBB MAF data was available for 120 of these pathogenic variants (see Table 4.3-1). ExAC MAF data was available for 41 of these 120 variants.

Gene	No. variants	Path/Likely Path
MYBPC3	376 (289)	60 (54)
MYH7	449 (349)	79 (66)
Total	825 (638)	139 (120)

Table 4.3-1 Variants annotated as "Rare variants in cardiac predisposition genes" detected in *MYBPC3* and *MYH7* genes in the UKBB cohort. Pathogenic variants classified as pathogenic by OMGL and/or LMM. Path=pathogenic. Number in brackets refers to the number of variants where UKBB MAF data was available.

On initial review, it was noted that for 11/41 variants, the UKBB MAF was higher than the conservative MAF threshold applied in previous analyses to define rare variants ($1x10^{-4}$). ExAC MAF data was available for 7 of these 11 variants; in 3 the UKBB MAF was more than twice the MAF observed in European individuals in ExAC (ExAC Euro Non-Finnish[NF]). The most extreme differences in MAF were found for the *MYBPC3* p.Arg502Trp variant (the most frequent pathogenic ICC variant in the Northern European population)[342,357], where the UKBB MAF was almost 5 fold higher than in ExAC Euro NF, and the *MYH7* p.Arg904Cys variant (a pathogenic variant found in 1/756 individuals in the LMM DCM cohort), where there was a 10 fold difference (Table 4.3-2). These MAF differences were less pronounced when data from the European population within the more recently available Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org), which is approximately twice the size of the ExAC cohort, was used for comparison. In gnomAD, there was a 2 fold difference in MAF for the *MYBPC3* p.Arg502Trp variant and the *MYH7* p.Arg904Cys variant was annotated as having been found in 1 of 9850 individuals of Ashkenazi Jewish ancestry, and not in 22,286 European individuals (Table 4.3-2).

For 9 of these 11 variants, on visual inspection of the cluster plots calls the heterozygous genotype clusters did not form a discrete cluster compared to the homozygous wild type genotypes and therefore the assigned rare variant genotype calls were considered unreliable (Figure 4.3-2 to Figure 4.3-7 and Figure 4.3-10 to Figure 4.3-11). Subsequent review of the cluster plots from other putative pathogenic variants (UKBB MAF < 0.0001), confirmed that the genotyping calls were not robust and therefore the associated MAF data unreliable.

The cluster plots for two variants (*MYBPC3* p.Arg502Trp, Figure 4.3-1 and *MYH7* p.Arg787Cys, Figure 4.3-8) showed slightly more discrimination between the homozygous wild type and the heterozygous clusters. There was no marked evidence of batch effects influencing these genotype calls (personal communication Dr Anuj Goel, WCHG). The UKBB individuals in whom these variants were detected are not closely related (1st, 2nd or 3rd degree relatives, or monozygotic twins), however they are self-reportedly of similar ethnicity ("British/Irish/other white"), (personal communication Dr Anuj Goel, WCHG). Analysis of flanking SNPs indicated that there was a single haplotype associated with the *MYBPC3* p.Arg502Trp variant (Table 4.3-4, analysis undertaken by Dr Giulia Poloni, WCHG). None of the individuals heterozygous for these variants were reported to have an ICD10 code (International Statistical Classification of Diseases and Related Health Problems 10th Revision) corresponding to HCM or any other related cardiac condition (personal communication, Dr Anuj Goel).

Fisher's exact test and OR ratio analyses indicated that there was a significant difference in the frequency of both variants when comparing the UKBB and gnomAD cohorts, *MYBPC3* p.Arg502Trp (OR 2.2 95% CI [1.2-4.3], p=0.005) and *MYH7* p.Arg787Cys (OR 2.52 95% CI [1.33-5.3], p=0.003), (Table 4.3-3). Further Fisher's exact test and OR ratio

analyses indicated that the *MYBPC3* p.Arg502Trp variant was in significant excess in HCM cases compared to both UKBB and gnomAD (Table 4.3-5 and Table 4.3-6). However, no evidence was found to suggest that the *MYH7* p.Arg787Cys variant, which had been classified as Likely pathogenic by the OMGL laboratory, was in excess in cases (Table 4.3-5 and Table 4.3-6).

Gene	Affymetrix ID	Variant	MAF UKBB	MAF ExAC (Global)	MAF ExAC (Euro)	MAF gnomAD (Euro)	MAF Clinical	Clinical classification (OMGL : LMM)	Figure
MYBPC3	80241798	p.Arg502Trp	0.0002	0.00002	0.00005	0.00010	0.008416	Path: Path	3-1
MYBPC3	86888786	p.Lys565*	0.0001	-	-	-	0.000172	n/a : Path	3-2
MYBPC3	91434825	p.Gly148Arg	0.0001	0.00004	0.00009	0.00016	0.000890	Likely path: VUS (FP)	3-3
MYBPC3	86885201	c.1928-2A>G	0.0001	-	-		0.002428	Path: Path	3-4
MYBPC3	86888809	p.Tyr237Ser	0.0001	-	-	-	0.000486	Likely path: Likely path	3-5
MYBPC3	89007238	p.Arg733Cys	0.0001	0.00009	0.00006	0.00007	0.000162	Likely path: VUS	3-6
MYH7	10272654	p.Lys146Asn	0.0005	-	-	-	0.000172	n/a: Likely path	3-7
MYH7	80248718	p.Arg787Cys	0.0002	0.00006	0.00009	0.00009	0.000313	Likely path: n/a	3-8
MYH7	80248725	p.Ala326Pro	0.0001	0.00007	0.00009	0.00015	0.000245	Likely path: VUS	3-9
MYH7	86888989	p.Arg904Cys	0.0001	0.00001	0.00001	-[$]	0.000661	n/a : Path	3-10
MYH7	86889047	p.Arg671Cys	0.0001	-	-	-	0.000245	Likely path: Likely path	3-11

Table 4.3-2 Variants annotated as pathogenic (Path) or likely pathogenic (Likely path) with UKBB MAF > 0.0001 (threshold applied to previous rare variant analyses)[357]. MAF=Minor allele frequency. ExAC=Exome Aggregation Consortium (http://exac.broadinstitute.org/), gnomAD=Genome Aggregation Database (https://gnomad.broadinstitute.org/). [$] Not in 22,286 European individuals, in 1/9850 Ashkenazi Jewish individuals (this ethnicity was not annotated in ExAC).

Gene	variant	UKBB AC	UKBB n	gnomAD AC	gnomAD n	FET p-value	OR	lower CI	upper CI
MYBPC3	p.Arg502Trp	69	152732	13	63,850	0.005	2.219	1.216	4.377
MYH7	p.Arg787Cys	67	152732	11	63,344	0.003	2.527	1.326	5.303

Table 4.3-3 Variants investigated as true positive results. Fisher's exact test (FET) p-value and Odds Ratio (OR) with 95% Confidence interval (CI), UKBB compared to Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/). AC=allele count. n=number of individuals.

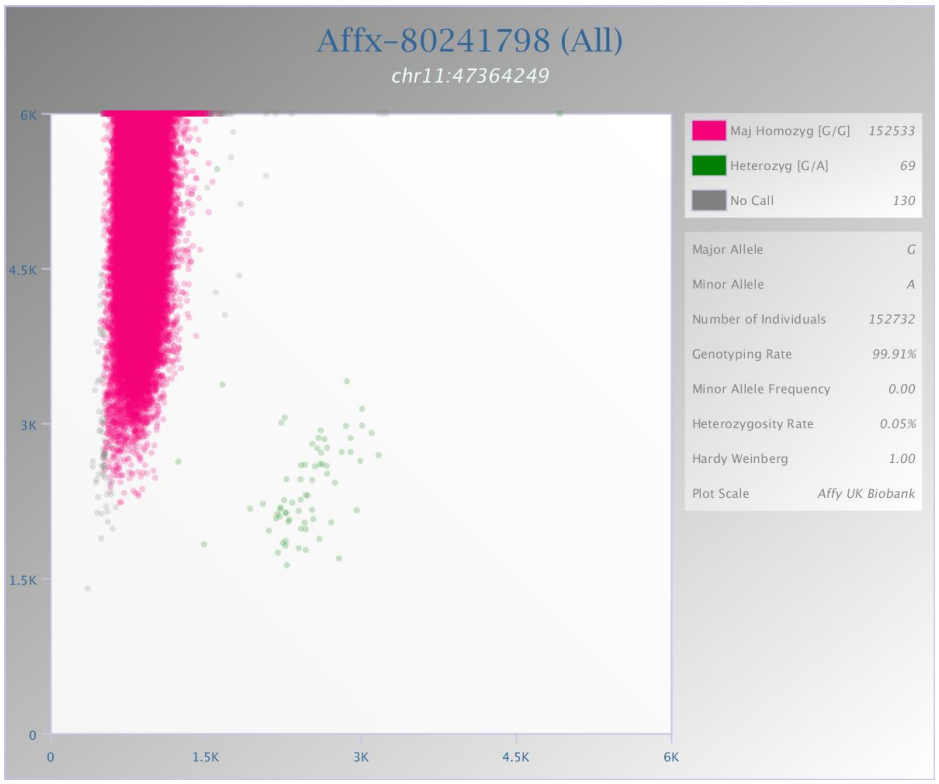


Figure 4.3-1 ScatterShot cluster plot *MYBPC3* p.Arg502Trp. Variant annotated as pathogenic by OMGL and LMM, detected in 104/6719 individuals in HCM clinical cohort.

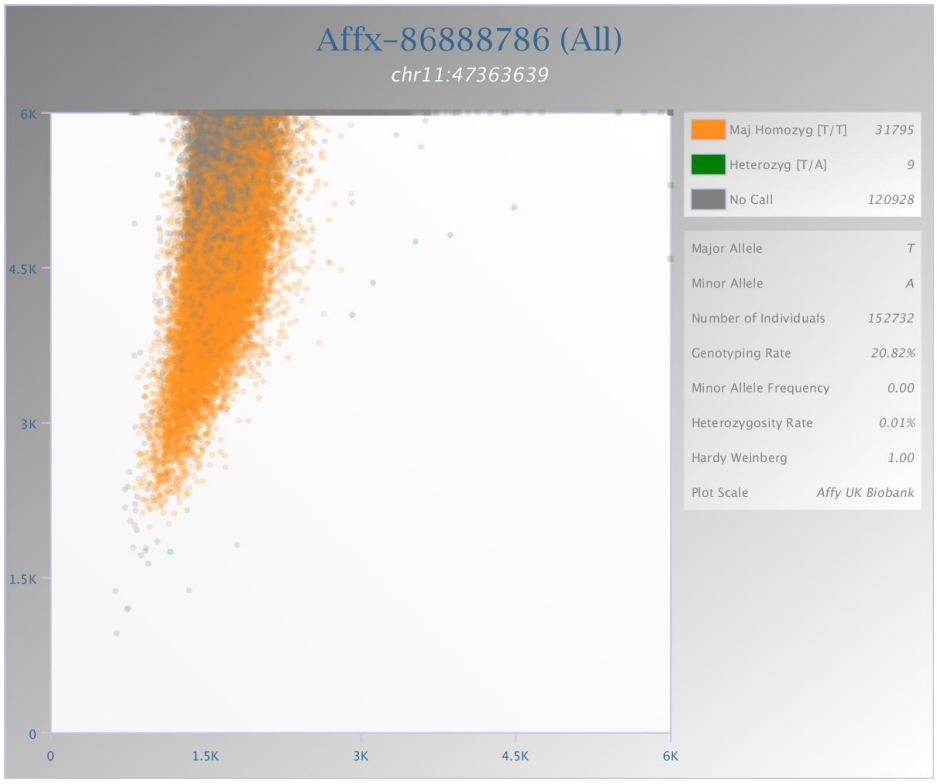


Figure 4.3-2 ScatterShot cluster plot *MYBPC3* p.Lys565*. Variant annotated as pathogenic by LMM, detected in 1/2912 individuals in LMM and 0/3200 in OMGL HCM clinical cohorts.

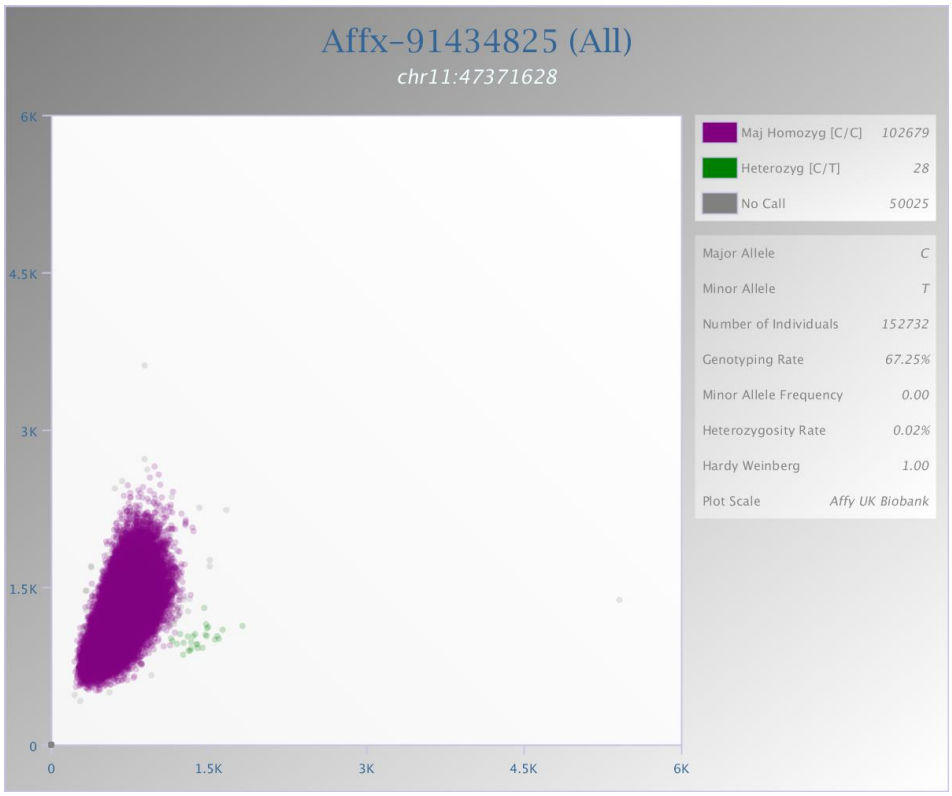


Figure 4.3-3 ScatterShot cluster plot *MYBPC3* p.Gly148Arg. Variant annotated as likely pathogenic by OMGL and VUS favour pathogenic by LMM, detected in 11/6719 individuals in HCM clinical cohort.

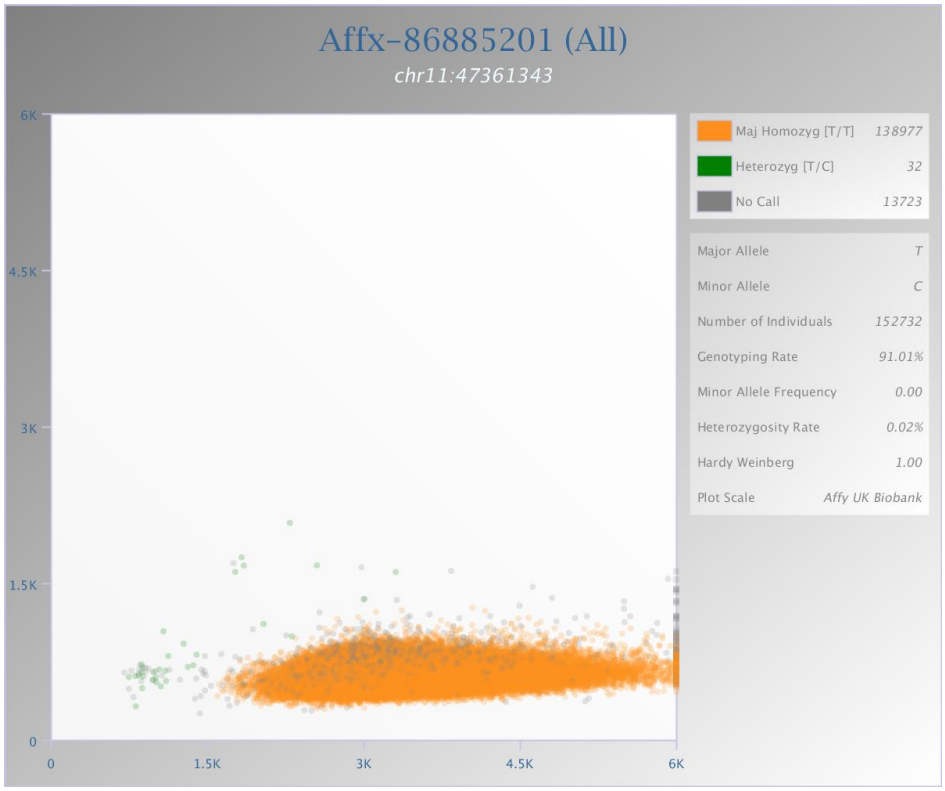


Figure 4.3-4 ScatterShot cluster plot *MYBPC3* c.1928-2A>G. Variant annotated as pathogenic by OMGL and LMM, detected in 30/6719 individuals in HCM clinical cohort.

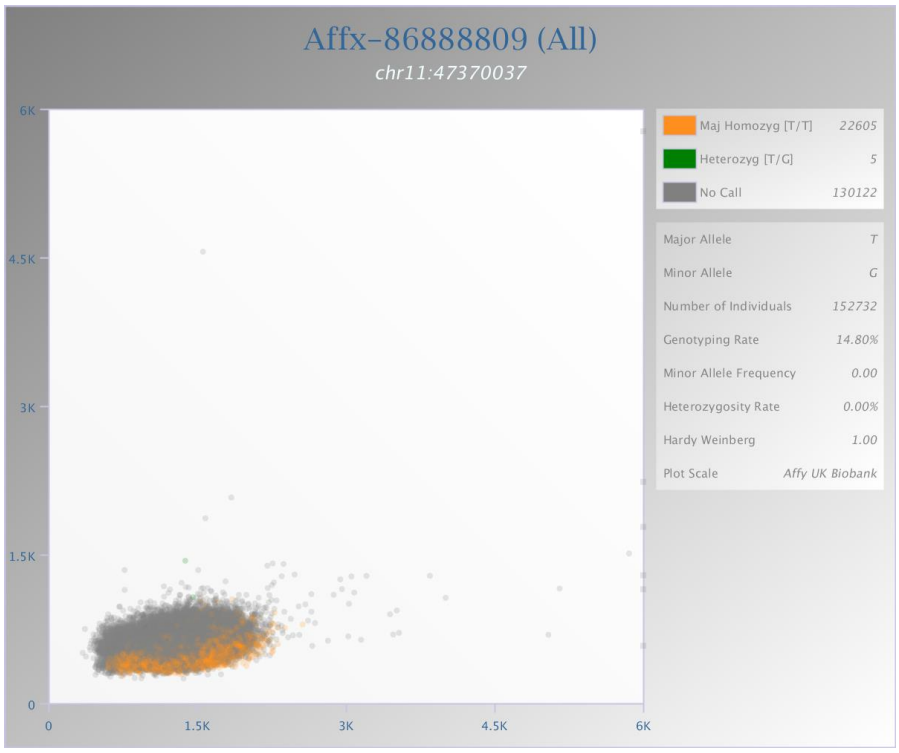


Figure 4.3-5 ScatterShot cluster plot *MYBPC3* p.Tyr237Ser. Variant annotated as likely pathogenic by OMGL and LMM, detected in 6/6719 individuals in HCM clinical cohort.

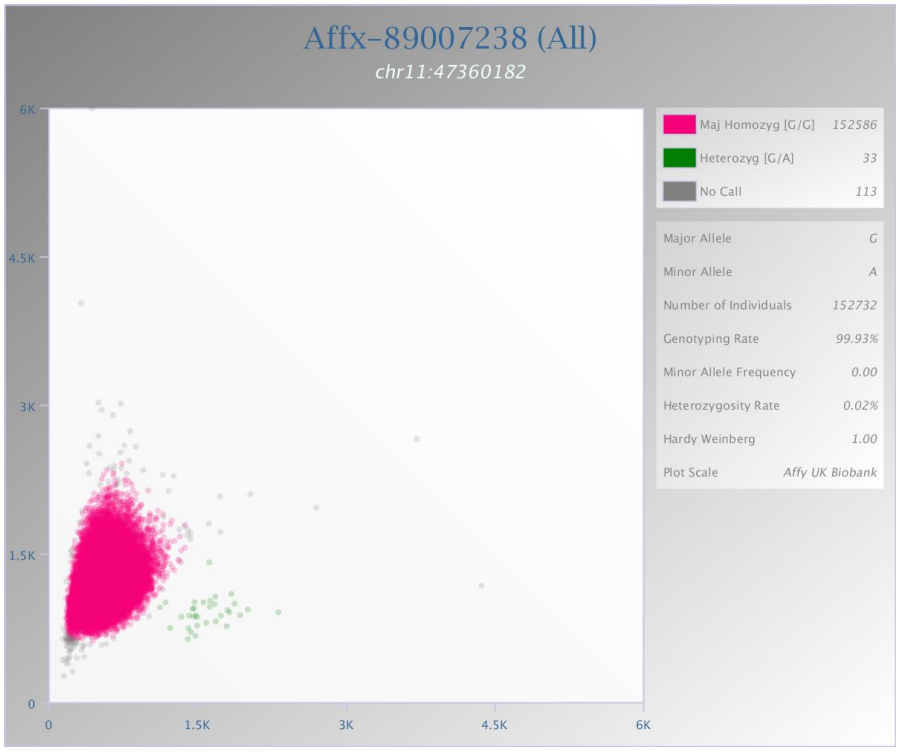


Figure 4.3-6 ScatterShot cluster plot *MYBPC3* p.Arg723Cys. Variant annotated as likely pathogenic by OMGL and VUS by LMM, detected in 2/6719 individuals in HCM clinical cohort.

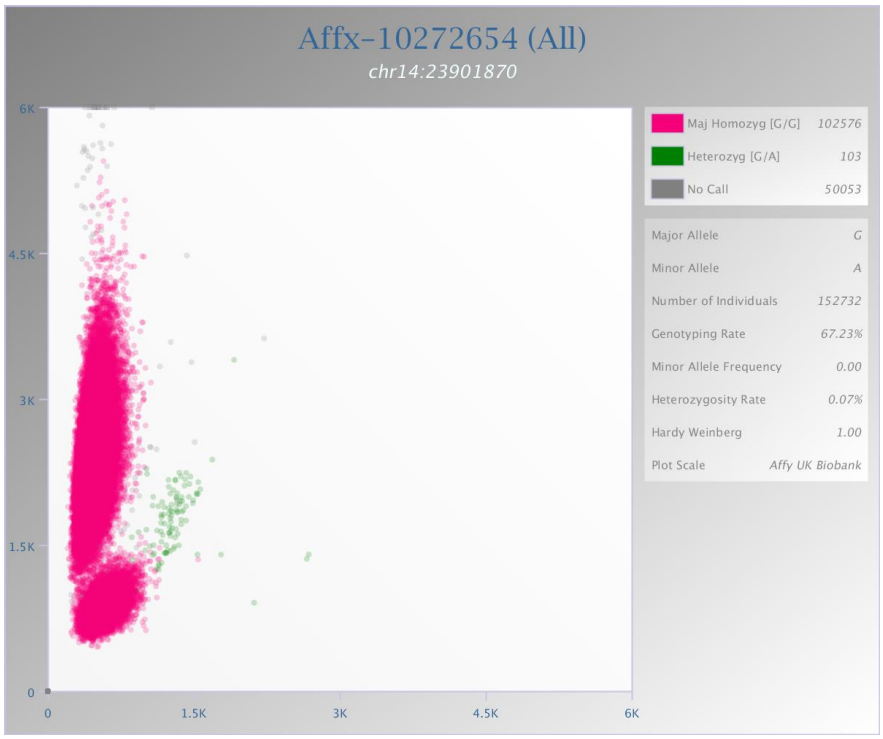


Figure 4.3-7 ScatterShot cluster plot *MYH7* p.Lys146Asn. Variant annotated as likely pathogenic by LMM, detected in 1/2912 individuals in LMM cohort and 0/3200 OMGL HCM cohort.

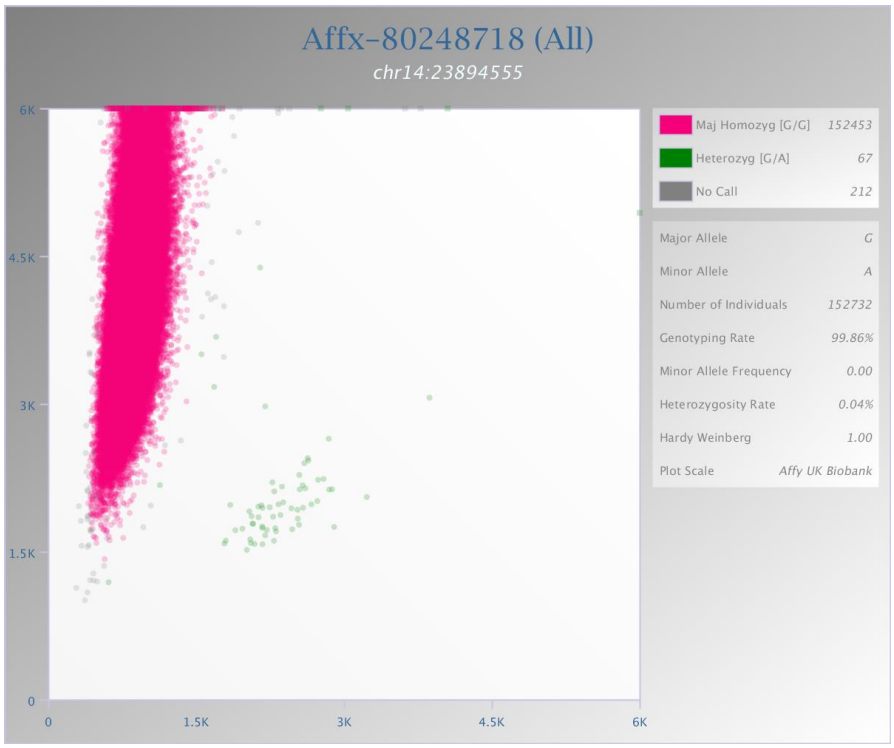


Figure 4.3-8 ScatterShot cluster plot *MYH7* p.Arg787Cys. Variant annotated as likely pathogenic by OMGL, detected in 2/3200 individuals in OMGL HCM clinical cohort and 0/2912 LMM cohort.

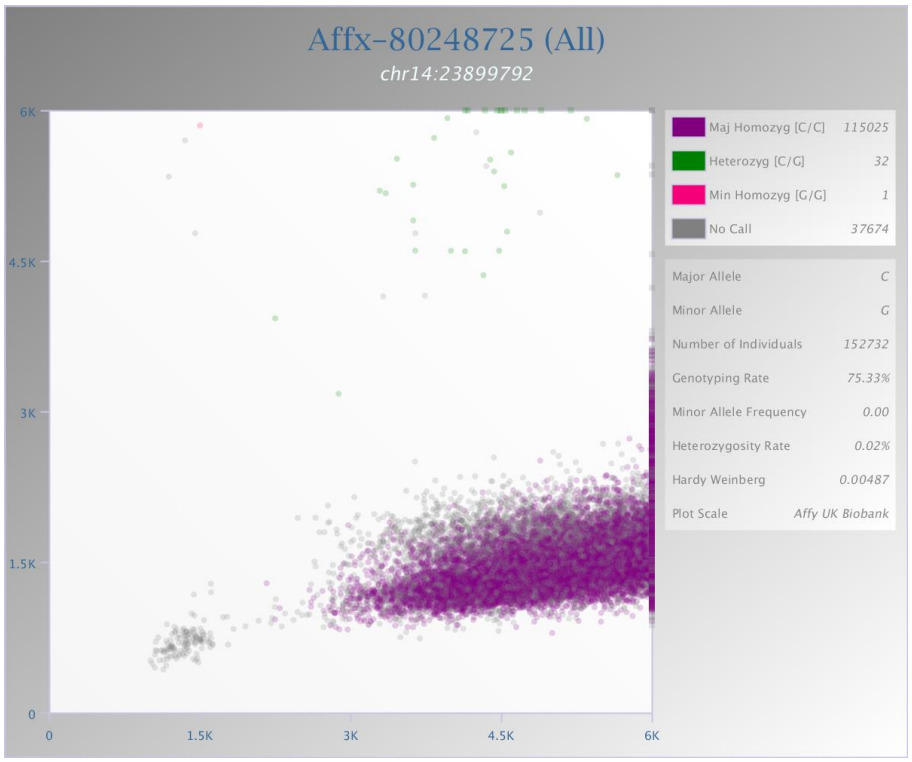


Figure 4.3-10 ScatterShot cluster plot *MYH7* p.Ala326Pro Variant annotated as likely pathogenic by OMGL and VUS by LMM, detected in 2/3200 individuals in OMGL HCM clinical cohort and 1/2912 LMM cohort.

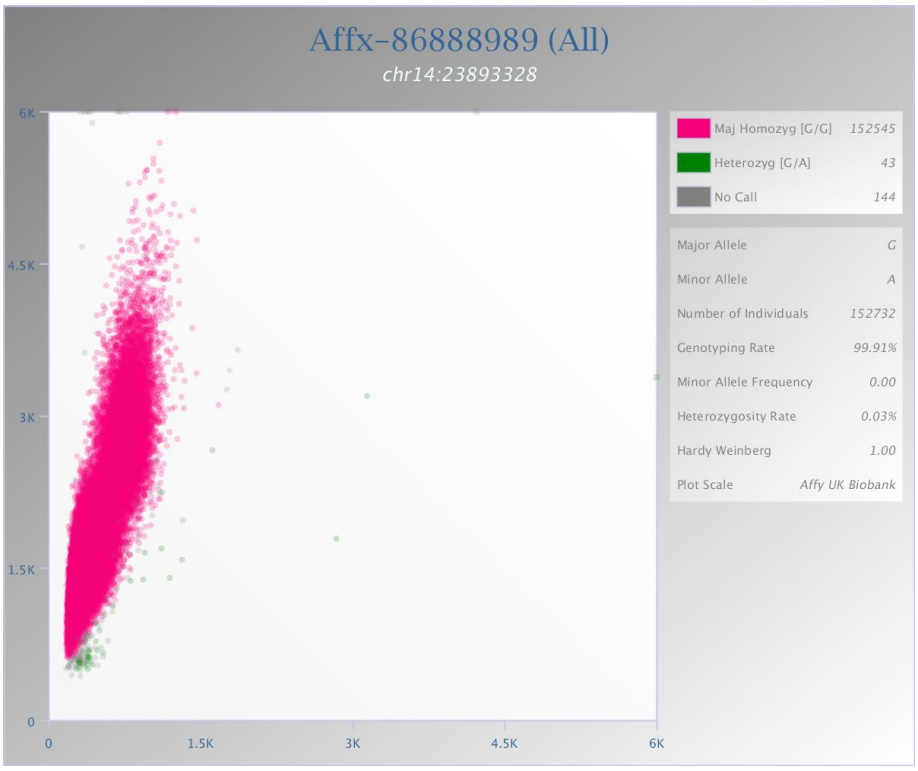


Figure 4.3-9 ScatterShot cluster plot *MYH7* p.Arg904Cys. Variant annotated pathogenic by LMM, detected in 1/756 individuals in DCM clinical cohort, not detected in 559 individuals in OMGL DCM cohort.

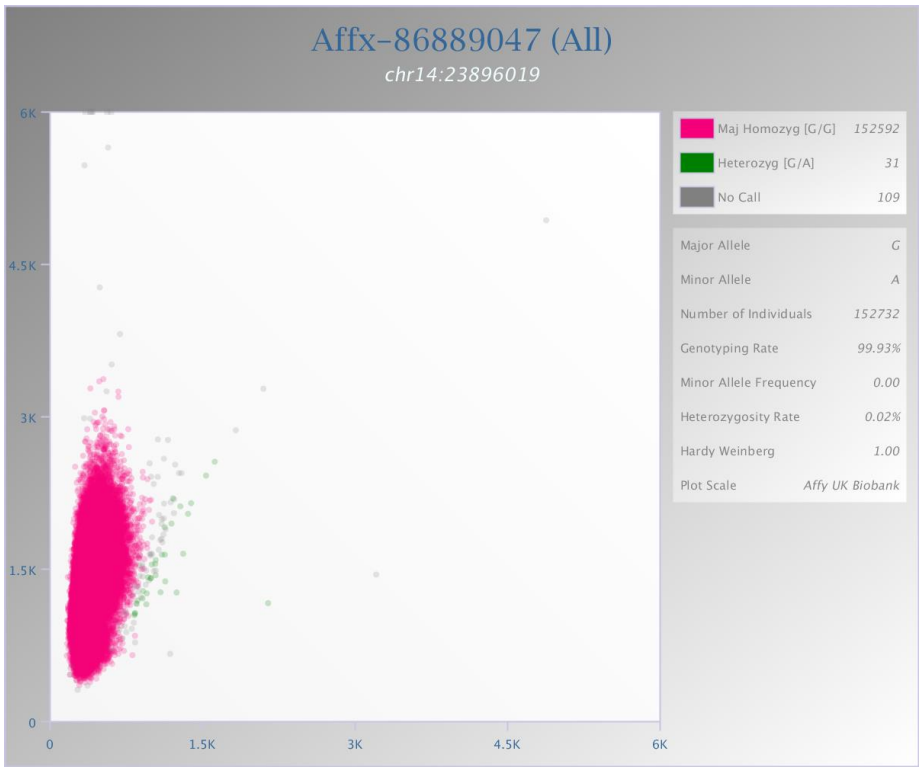


Figure 4.3-11 ScatterShot cluster plot *MYH7* p.Arg671Cys Variant annotated as likely pathogenic by OMGL and LMM, detected in 2/3200 individuals in OMGL HCM clinical cohort and 1/2912 LMM.

Affx-5247979	Affx-5248024	Affx-5248028	Affx-5248110	Affx-5249966	Affx-5250067	Affx-5250133	Affx-5250292	***Affx-80241798***	Affx-5250529	Affx-5250546	Affx-5250572	Affx-5250861	Affx-5250964	Pop. Freq.
C	G	C	T	T	C	C	T	G	T	C	C	A	C	0.211
T	A	C	G	T	T	T	C	G	T	T	C	G	T	0.137
T	G	C	T	C	C	C	T	G	T	T	C	A	T	0.066
T	G	C	T	C	C	C	T	G	T	C	C	A	T	0.065
T	G	C	T	T	C	T	C	G	C	T	C	G	T	0.063
T	***G***	***T***	***T***	***C***	***C***	***C***	***T***	***A***	***T***	***C***	***C***	***A***	***T***	***0.056***
T	G	C	T	T	C	C	T	G	T	C	T	G	T	0.046
T	G	T	T	C	C	C	T	G	T	C	C	A	T	0.044
T	G	C	T	T	C	T	C	G	C	C	C	G	T	0.042
T	G	C	T	T	C	C	T	G	T	C	C	A	C	0.025
T	G	C	T	C	C	C	T	G	T	T	T	G	T	0.025
T	G	C	G	T	C	T	C	G	T	C	C	A	T	0.023
T	A	C	G	T	T	T	C	G	T	T	C	A	T	0.018
C	G	C	T	T	C	C	T	G	T	C	C	G	T	0.015
C	G	C	T	T	C	C	C	G	C	C	C	G	T	0.013
C	G	C	T	T	C	C	T	G	T	C	C	A	T	0.012
T	G	C	T	T	C	C	T	G	T	C	T	A	C	0.012
T	A	C	G	T	T	T	C	G	T	T	C	A	C	0.012
T	A	C	G	T	T	T	C	G	T	C	C	G	T	0.011

Table 4.3-4 Graphical representation of haplotype analysis around *MYBPC3* p. Arg502Trp mutation spanning approximately83 kb from chr11: 47303275 (Affx-5247979) to chr11: 47387599 (Affx-5250964). The conserved haplotype, with the mutant allele and the Affymetrix ID corresponding to the variant is highlighted in bold and italics. Data generated by Dr Giulia Poloni, WCHG.

Gene	variant	HCM AC	HCM n	UKBB AC	UKBB n	FET p-value	OR	lower CI	upper CI
MYBPC3	p.Arg502Trp	104	6,179	69	152,732	2.2x10-16	38.99	28.38	53.84
MYH7	p.Arg787Cys	2	6,112	67	152,732	1	0.75	0.09	2.80

Table 4.3-5 Fisher's Exact Test (FET) and odds ratio (OR) with 95% confidence intervals (CI) comparing the proportion of HCM cases (OMGL and LMM combined) with either the *MYBPC3* p.Arg502Trp or *MYH7* p.Arg787Cys variants (in which the quality of the cluster plots was considered acceptable) with the proportion in the UKBB cohort. AC=allele count. n=number of individuals.

Gene	variant	HCM AC	HCM n	gnomAD AC	gnomAD n	FET p-value	OR	lower CI	upper CI
MYBPC3	p.Arg502Trp	104	6179	13	63,850	2.2x10-16	84.18	47.06	162.65
MYH7	p.Arg787Cys	2	6112	11	63,344	0.3193	1.88	0.20	8.64

Table 4.3-6 Fisher's Exact Test (FET) and odds ratio (OR) with 95% confidence intervals (CI) comparing the proportion of HCM cases (OMGL and LMM combined) with either the *MYBPC3* p.Arg502Trp or *MYH7* p.Arg787Cys variants (in which the quality of the cluster plots was considered acceptable) with the proportion in the Genome Aggregation Database (gnomAD) cohort (https://gnomad.broadinstitute.org/). AC=allele count. n=number of individuals.

4.3.2 BRRD cohort

A total of 955 rare (ExAC MAF <0.0001) variants were detected in coding regions and canonical splice sites within the 16 selected ICC genes in 1,412/7,878 (17.9%) of individuals from the BRRD non-HCM cohorts. Of these 955 variants, 39 (4%) were KPV, and 33 (3.4%) were EPV (of which 6 had been previously reported by OMGL and/or LMM). In total, 66 KPV and EPV variants (66/955, 6.9%), detected in 100/7,878 (1.3%) individuals were selected for further review (See Table 4.3-7).

Gene	No. distinct rare variants in BRRD non-HCM cohort	Known pathogenic variants [KPV]	Expected pathogenic variants [EPV] (number KPV)	No. variants selected for 2nd review (% of all rare variants)
ACTC1	2	0	0	0
DSC2	64	0	5 (0)	5 (7.8%)
DSG2	85	4	5 (1)	8 (9.4%)
DSP	184	1	8(1)	8 (4.3%)
KCNH2	81	4	0	4 (4.9%)
KCNQ1	44	4	2 (2)	4 (9%)
LMNA	41	2	0	2 (4.9%)
MYBPC3	120	8	4 (0)	12 (10%)
MYH7	91	7	n/a	7 (7.7%)
MYL2	11	0	0	0
MYL3	12	2	n/a	2 (16.6%)
PKP2	70	2	8 (2)	8 (11.4%)
SCN5A	111	3	1 (0)	4 (3.6%)
TNNI3	19	2	n/a	2 (10.5%)
TNNT2	16	0	n/a	0
TPM1	4	0	n/a	0
TOTAL	955	39	33(6)	66

Table 4.3-7 Total number of rare (ExAC MAF <0.0001) unique variants detected in BRRD non-HCM cohorts (n=7,878). Known pathogenic variants [KPV] =number of variants that had been reported as "highly likely pathogenic" or "likely pathogenic" by either the OMGL or LMM clinical laboratories. Expected pathogenic variants [EPV] =Number of putative truncating variants detected in genes in which truncating variants are generally considered pathogenic. n/a=putative loss-of-function variants in these genes are not considered pathogenic.

Following review, 43 variants (43/955, 4.5%) detected in 64 individuals (64/7,878, 0.8 %,) were classified as "Highly likely" or "Likely" to be pathogenic (Table 4.3-9). When considering unrelated individuals only, the overall prevalence was 0.7% [95% CI 0.53-0.89], (45/6,466). The remaining 23 variants were classified as VUS. The total number of "Highly likely" and "Likely" pathogenic variants detected according to the related ICC, along with current estimates of disease prevalence, is shown in Table 4.3-8.

ICC genes	Variants in ICC genes classified as Highly likely/Likely to be pathogenic			Estimated population prevalence of disease
	All individuals (n=7,878)	Unrelated individuals (n=6,466)	% unrelated individuals [95% CI]	
ARVC : *DSC2* *DSG2* *PKP2*	28*	18* <*15*>	0.28%* [0.18-0.41] <*0.23%*>	~0.05% [358,359]
ARVC/DCM: *DSP*	12*	8* <*6*>	0.12%* [0.06-0.22] <*0.09%*>	ARVC: ~0.05%[61] DCM: ~0.4%[61]
LQTS: KCNH2 KCNQ1 SCN5A	6	5	0.08% [0.03-0.16]	~0.02-0.04%[360–362]
HCM: *MYBPC3* *MYH7* *MYL3* *TNNI3*	18	14	0.22% [0.13-0.34]	~0.2-0.5% [10,363]

Table 4.3-8 Summary of findings by ICC. This table summarises the number of Highly Likely/Likely pathogenic variants. These comprise candidate gene variants (variants previously detected in disease cohorts by OMGL and/or LMM clinical laboratories and considered Likely/Highly likely to be pathogenic and/or putative truncating variants in genes where this class of variant is considered to be pathogenic) which on review were considered Highly Likely/Likely pathogenic variants. Individuals=Total number of individuals from non-HCM BRRD cohorts, this includes related individuals. Unrelated=probands only. % unrelated individuals=% of unrelated individuals with a Highly Likely/Likely pathogenic variants with 95% Binomial confidence intervals.
* On review of BRRD records, some of the individuals with variants in these genes were reported to have ARVC or DCM. <>=count or % with ICC cases removed

ICC	Gene	c.	p.	AC	OMGL	LMM	Reviewer 1	Reviewer 2	Final
ARVC	*DSC2*	c.1123C>T	p.Arg375*	1	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC	*DSC2*	c.959delA	p.Gln320fs	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC	*DSC2*	c.939_942delAGAG	p.Arg313fs	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC	*DSC2*	c.379G>T	p.Glu127*	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC	*DSC2*	c.2533G>T	p.Glu845*	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC	*DSG2*	c.1880-2A>G	.	1	Pathogenic	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC	*DSG2*	c.355C>T	p.Arg119*	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC	*DSG2*	c.375dupT	p.Leu126fs	2	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC	*DSG2*	c.603_604delAT	p.Leu203fs	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC	*DSG2*	c.2894_2897delAGAG	p.Glu965fs	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC/DCM	*DSP*	c.3G>C	p.Met1?	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
ARVC/DCM	*DSP*	c.1288G>T	p.Glu430*	3	Pathogenic[$]	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC/DCM	*DSP*	c.2436+2T>C	.	1	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC/DCM	*DSP*	c.3672dupA	p.Ser1225fs	1	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC/DCM	*DSP*	c.4423delA	p.Thr1475fs	1	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC/DCM	*DSP*	c.5887dupA	p.Thr1963fs	2	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC/DCM	*DSP*	c.6273delA	p.Ala2092fs	2	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC/DCM	*DSP*	c.7968_7972delCACAG	p.Cys2656fs	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
LQTS	*KCNH2*	c.1744C>T	p.Arg582Cys	1	Pathogenic	Not reported	Likely Path	Likely Path	Likely Path
LQTS	*KCNQ1*	c.1066C>T	p.Gln356*	2	Pathogenic	Not reported	Pathogenic	Pathogenic	Pathogenic
LQTS	*KCNQ1*	c.1780C>T	p.Arg594*	1	Pathogenic	Not reported	Pathogenic	Pathogenic	Pathogenic
LQTS	*KCNQ1*	c.1781G>A	p.Arg594Gln	1	Likely Path	Not reported	Likely Path	Likely Path	Likely Path
HCM	*MYBPC3*	c.1504C>T	p.Arg502Trp	3	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic
HCM	*MYBPC3*	c.1624G>C	p.Glu542Gln	1	Pathogenic	Likely Path	Pathogenic	Likely Path	Pathogenic
HCM	*MYBPC3*	c.2373dupG	p.Trp792fs	1	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic
HCM	*MYBPC3*	c.3192dupC	p.Lys1065fs	1	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
HCM	*MYBPC3*	c.772G>A	p.Glu258Lys	2	Pathogenic	Pathogenic	Pathogenic	Likely Path	Pathogenic

ICC	Gene	c.	p.	AC	OMGL	LMM	Reviewer 1	Reviewer 2	Final
HCM	*MYBPC3*	c.3592dupG	p.Ala1198fs	2	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
HCM	*MYBPC3*	c.994G>T	p.Glu332*	1	Not reported	Not reported	Pathogenic	Likely Path	Pathogenic
HCM/DCM	*MYH7*	c.4066G>A	p.Glu1356Lys	1	Likely Path ^	Likely Path ^	Likely Path	Likely Path	Likely Path
HCM/DCM	*MYH7*	c.1142C>A	p.Ala381Asp	1	Likely Path ^	Not reported	Likely Path	Likely Path	Likely Path
HCM	*MYL3*	c.170C>G	p.Ala57Gly	2	Not reported	Likely Path	Likely Path	Likely Path	Likely Path
HCM	*MYL3*	c.281G>A	p.Arg94His	1	Not reported	Likely Path	Likely Path	VUS	Likely Path
ARVC	*PKP2*	c.1237C>T	p.Arg413*	1	Pathogenic	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC	*PKP2*	c.1799delA	p.Asp600fs	3	Not reported	Not reported	Pathogenic	Likely Path	Pathogenic
ARVC	*PKP2*	c.2146-1G>C		6	Pathogenic	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC	*PKP2*	c.2489+1G>A		3	Not reported	Not reported	Pathogenic	Pathogenic	Pathogenic
ARVC	*PKP2*	c.2478delT	p.Ser826fs	1	Not reported	Not reported	Pathogenic	Likely Path	Pathogenic
ARVC	*PKP2*	c.2378_2379dupCC	p.Ser794fs	1	Not reported	Not reported	Pathogenic	Likely Path	Pathogenic
ARVC	*PKP2*	c.2194C>T	p.Gln732*	1	Not reported	Not reported	Pathogenic	Likely Path	Pathogenic
ARVC	*PKP2*	c.1035-1G>C		1	Not reported	Not reported	Pathogenic	Likely Path	Pathogenic
LQTS	*SCN5A*	c.6017dupC	p.Ser2007fs	1	Not reported	Not reported	Likely Path	Likely Path	Likely Path
HCM	*TNNI3*	c.433C>T	p.Arg145Trp	2	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic

Table 4.3-9 Candidate gene variants considered "Highly likely" or "Likely" Pathogenic following review. AC=allele count in BRRD cohort (n=7,878), OMGL=variant classification made by Oxford clinical laboratory (data accessed Dec 2014). LMM= variant classification from LMM clinical laboratory [77,124] Reviewer 1 (Kate Thomson), Reviewer 2 (Mike Mackley). Variants reviewed Jan-March 2017. ^=variant detected in a patient with HCM [$]=variant detected in a patient with ARVC.

Fisher's Exact Test and OR analyses confirmed that the well-established recurrent pathogenic variants detected in the *MYBPC3*, *TNNI3* and *PKP2* genes were in significant excess in clinical cases compared to unrelated individuals from non-HCM BRRD cohorts (Table 4.3-10).

For the 64 individuals with one of the 43 pathogenic or likely pathogenic variants, the BRRD database was interrogated to retrieve demographic data (age at sampling, gender, related samples, and ethnicity), details of referral cohort and phenotype data. For several individuals, date of birth or age at sampling was not available. Additionally, for many individuals, clinical/phenotype data was not provided. For five variants, detected in a total of 17 individuals in the Genomics England RD Pilot cohort, the records indicated that a familial ICC or other cardiac phenotype was present (Table 4.3-10).

Gene	c.	p.	Class	BRRD	BRRD unrelated	ICC recorded	Clinical Labs	FET p-value	OR [95 % CI]
DSC2	c.1123C>T	p.Arg375*	P	1/7878	1/6446	No	-		
DSC2	c.959delA	p.Gln320fs	LP	1/7878	1/6446	No	-		
DSC2	c.939_942delAGAG	p.Arg313fs	LP	1/7878	1/6446	No	-		
DSC2	c.379G>T	p.Glu127*	LP	1/7878	1/6446	No	-		
DSC2	c.2533G>T	p.Glu845*	LP	1/7878	1/6446	No	-		
DSG2	c.1880-2A>G	.	P	1/7878	1/6446	No	1/354	0.014	18.23 [0.23 - 1414.76]
DSG2	c.355C>T	p.Arg119*	LP	1/7878	1/6446	No	-		
DSG2	c.375dupT	p.Leu126fs	P	2/7878	1/6446	No	-		
DSG2	c.603_604delAT	p.Leu203fs	LP	1/7878	1/6446	No	-		
DSG2	c.2894_2897delAGAG	p.Glu965fs	LP	1/7878	1/6446	No	-		
DSP	c.3G>C	p.Met1?	LP	1/7878	1/6446	No	-		
DSP	**c.1288G>T**	**p.Glu430***	**P**	**3/7878**	**1/6446**	**DCM**	*2/352*	*0.008*	*36.76 [1.91 -2137]*
DSP	c.2436+2T>C	.	P	1/7878	1/6446	**RVF, PAH**	-		
DSP	c.3672dupA	p.Ser1225fs	P	1/7878	1/6446	No	-		
DSP	c.4423delA	p.Thr1475fs	P	1/7878	1/6446	No	-		
DSP	c.5887dupA	p.Thr1963fs	P	2/7878	1/6446	No	-		
DSP	c.6273delA	p.Ala2092fs	P	2/7878	**1/6446**	**ARVC**	-		
DSP	**c.7968_7972delCACAG**	**p.Cys2656fs**	**LP**	**2/7878**	**1/6446**	**RVF, PAH**	-		
KCNH2	c.1744C>T	p.Arg582Cys	LP	1/7878	1/6446	No	-		
KCNQ1	c.1066C>T	p.Gln356*	P	2/7878	1/6446	No	-		
KCNQ1	c.1780C>T	p.Arg594*	P	1/7878	1/6446	No	-		
KCNQ1	c.1781G>A	p.Arg594Gln	LP	1/7878	1/6446	No	-		
MYBPC3	c.1504C>T	p.Arg502Trp	**<P>**	3/7878	2/6446	No	104/6179	2.2x10-16	55 [14.87 - 461.06]
MYBPC3	c.1624G>C	p.Glu542Gln	**<P>**	1/7878	1/6446	No	41/6197	4.4 x10-12	43.04 [7.3 -1727.29]
MYBPC3	c.2373dupG	p.Trp792fs	**<P>**	1/7878	1/6446	No	66/6197	2.2x10-16	69.55 [12.04 - 2750.60]
MYBPC3	c.3192dupC	p.Lys1065fs	P	1/7878	1/6446	No	-		

Gene	c.	p.	Class	BRRD	BRRD unrelated	ICC recorded	Clinical Labs	FET p-value	OR [95 % CI]
MYBPC3	c.772G>A	p.Glu258Lys	**<P>**	2/7878	1/6446	No	69/6197	2.2x10-16	72.75 [12.63 -2873]
MYBPC3	c.3592dupG	p.Ala1198fs	P	2/7878	1/6446	No	-		
MYBPC3	c.994G>T	p.Glu332*	P	1/7878	1/6446	No	-		
MYH7	c.4066G>A	p.Glu1356Lys	LP	1/7878	1/6446	No	6/6112	0.06	6.33 [0.77 - 290.96]
MYH7	c.1142C>A	p.Ala381Asp	LP	1/7878	1/6446	No	5/3200	0.02	10.08 [1.13 - 475.95]
MYL3	c.170C>G	p.Ala57Gly	LP	2/7878	1/6446	No	2/2650	0.2	4.867 [0.25 - 286.70]
MYL3	c.281G>A	p.Arg94His	LP	1/7878	1/6446	No	4/2650	0.03	9.74 [0.963 - 478.64]
PKP2	c.1237C>T	p.Arg413*	P	1/7878	1/6446	No	2/361	0.008	35.84 [1.86 - 2084.43]
PKP2	**c.1799delA**	**p.Asp600fs**	**P**	**3/7878**	**1/6446**	**ARVC**	-		
PKP2	**c.2146-1G>C**	**.**	**<P>**	**6/7878**	**1/6446**	**ARVC**	24/361	*2.2x10-16*	*457.71 [74.18 - 16,384]*
PKP2	**c.2489+1G>A**	**.**	**P**	**3/7878**	**1/6446**	**ARVC**	-		
PKP2	c.2478delT	p.Ser826fs	P	1/7878	1/6446	No	-		
PKP2	c.2378_2379dupCC	p.Ser794fs	P	1/7878	1/6446	No	-		
PKP2	c.2194C>T	p.Gln732*	P	1/7878	1/6446	No	-		
PKP2	c.1035-1G>C	.	P	1/7878	1/6446	No	-		
SCN5A	c.6017dupC	p.Ser2007fs	LP	1/7878	1/6446	No	-		
TNNI3	c.433C>T	p.Arg145Trp	**<P>**	2/7878	2/6446	No	13/6047	0.003	6.94 [1.57 - 63.43]

Table 4.3-10 Comparison of variant frequency between the BRRD cohort and OMGL and LMM clinical cohorts. Class column lists final (post review) variant classification, P=Pathogenic, LP=Likely pathogenic. <P>=well established recurrent pathogenic variant. ICC recorded=whether any cardiac related phenotype was recorded in the BRRD summary data table. Where an ICC diagnosis or cardiac phenotype was recorded, the row has been highlighted in BOLD. Clinical labs=combined total frequency of variant in OMGL and LMM. FET=Fisher's exact test, OR=odds ratio. Results in grey italics relate to variants where the BRRD case was reported to have an ICC/cardiac related phenotype. RVF=Right ventricular failure. PAH=Pulmonary hypertension.

4.4 Discussion

Analysis of array data from 150,000 UKBB participants failed to provide reliable data about the prevalence of known rare variants within the *MYBPC3* and *MYH7* genes. The findings in this study indicate that this methodology is not suitable for assessing the prevalence of very rare variants in the wider population.

Analysis of genome sequencing data from 6,466 unrelated individuals from the BRRD cohort was undertaken to explore the prevalence of known and expected pathogenic variants (KPV and EPV) in 16 definitive ICC genes. Overall, the prevalence of variants in ARVC and LQTS genes was higher than current population estimates of disease prevalence[275,358–360,362,364,365]. The prevalence in HCM genes was similar to population estimates of disease prevalence[10]. Although lack of specific phenotyping and clinical data precluded comprehensive assessment of variant penetrance, in some individuals review of summary level data confirmed a personal or family history of ICC. Follow-on clinical studies have been initiated to explore variant penetrance in this unselected cohort.

4.4.1 UKBB cohort

Following visualisation of the cluster plots, the vast majority of genotype calls in this dataset appear to be false positive results, implying that this methodology was unreliable for analysis of the very low frequency variants which were the focus of this study (ExAC MAF $<1x10^{-4}$). This was not entirely unexpected as genotyping arrays are not routinely used to assess the frequency of very rare variants. In general, genotyping calls made within array datasets are based on quantitative measures of allelic abundance; inevitably, for very rare variants it can be difficult to differentiate between

signal and background noise. Also, the genotype clustering algorithms have not been optimised in this setting and there are no reliable metrics or quality control parameters to assess the confidence of variant calls (e.g. for very rare variants it is not possible to assess deviation from Hardy Weinberg equilibrium $p<1x10^{-6}$) [366,367].

However, in two of the variants being investigated because of an unexpectedly high MAF in the UKBB cohort, *MYBPC3* p.Arg502Trp and *MYH7* p.Arg787Cys, the cluster plots were of acceptable quality. Therefore, these results were investigated as possible true positive results.

The *MYBPC3* (p.Arg502Trp) variant, which is the most frequently reported pathogenic variant in HCM cohorts of Northern European decent [27,215,342], was detected in 0.045% [95% CI 0.036-0.055%] individuals in the UKBB cohort and 0.02% [95% CI 0.01-0.03] of European individuals in the gnomAD cohort. Despite the difference in MAF between the UKBB and gnomAD European cohorts, which likely reflects differences in population substructure [368,369], this variant was significantly enriched in the OMGL and LMM HCM cases series, compared to these cohorts (Table 4.3-5 and Table 4.3-6). While a common haplotype has been reported in HCM cases with this variant, it has been detected on more than one haplotype, suggesting that it has arisen recurrently[342]. In the UKBB dataset, there appears to be a common haplotype for this variant, suggesting a founder origin rather than multiple unrelated occurrences in this cohort. Family studies suggest that the penetrance of this variant is ~50% by 45 years of age [342]. All UKBB participants are >45 years of age. On the basis of ICD10 codes, there was no reported evidence of cardiomyopathy in the UKBB participants heterozygous for this variant.

The *MYH7* p.Arg787Cys variant was detected in 0.04% [95% CI 0.03-0.05%] individuals in the UKBB cohort and 0.02% [95% CI 0.01-0.03] of European individuals in the

gnomAD cohort. This variant was reported in two HCM cases within the OMGL cohort and classified as "Likely pathogenic". While this variant has been reported a number of times in the literature and in clinical databases in individuals with HCM, in the majority of cases it has been reported as a variant of uncertain clinical significance [370–372]. This variant does not appear to be in excess in the OMGL and LMM HCM case series when compared to UKBB or gnomAD (Table 4.3-5 and Table 4.3-6). On review of the ICD 10 codes, there was no indication that any of the UKBB participants heterozygous for this variant had a clinical diagnosis of HCM.

Although no clear evidence was found to indicate that the genotype calls relating to these two variants were false positive results, given the overall high error rate observed for rare variant genotype calls using this methodology, these prevalence estimates may well be unreliable. A follow-on application has been submitted to request DNA from the UKBB participants reported to be heterozygous for these variants, to allow genotype confirmation by Sanger sequencing. If these genotypes are confirmed, a more thorough clinical review of cardiac related datasets, ideally including electrocardiographic and echocardiographic results, would be undertaken to explore the variant penetrance. If not confirmed, this would indicate that, for very rare variants, erroneous genotype calls cannot be easily detected even by careful data inspection.

4.4.2 BRRD cohort

4.4.2.1 Variant prevalence

Overall, of 6,446 unrelated individuals in the BRRD cohort, 0.7% [95% CI 0.53-0.89] had a known or expected pathogenic variant, considered "Highly likely" or "Likely" to be pathogenic, in one of 16 definitive ICC genes. The results of these analyses support the hypothesis that pathogenic gene variants in ARVC and LQTS genes are at a higher than

expected frequency in unselected cohorts; however, the prevalence of KPV and EVP in HCM genes was found to be similar to estimates of the population prevalence of disease [10,363].

Given the comparatively large number of individuals screened, and range of ICC genes analysed, the results of this study provide relatively robust estimates relating to the prevalence of secondary findings in unselected cohorts. While broadly consistent with findings in previous studies (Table 4-4.4-1), which have generally reported higher than expected prevalences of ICC gene SFs (0.06-3%) in unselected cohorts, there are notable differences in the methodologies employed in each study. These include the source, size and ethnicity of cohorts; the genes selected for analyses; and the criteria used for variant classification. As such, the prevalence estimates derived from this study, and from previous published studies, are not directly comparable and may not accurately reflect the prevalence of pathogenic ICC gene variants in the wider population.

4.4.2.2 Variant penetrance

For these exploratory rare variant analyses, clinical and demographic data were not available for BRRD participants, so investigations into variant penetrance were not possible. However, following post-analysis review of de-identified summary referral data, which was available for a small proportion of individuals, an ICC diagnosis was noted in some cases. In 3 of the 8 index cases with a truncating variant in the *PKP2* gene, a personal or family history of ARVC was reported. Similarly, in 2 of 8 index cases with truncating variants in *DSP*, diagnoses of DCM and ARVC were recorded. All these cases had been recruited to a cohort that could have included ICC disorders (Genomics England RD Pilot).

Another potentially interesting observation arising from review of summary data, was that in two individuals with truncating variants in *DSP,* a phenotype of pulmonary arterial hypertension (PAH) with right ventricular failure was recorded. Further clinical investigations are required to determine whether the cardiac phenotype in these cases is related to these expected pathogenic *DSP* variants [373], or is coincidental.

Previous studies using electronic medical records to explore variant penetrance in unselected cohorts have provided some preliminary data. However, in general, case numbers have been small, and due to variation in methodology, it is difficult to draw any firm conclusions regarding the variant penetrance from these studies (Table 4-4.4-1)[141,341,344–346]. Critically, while these record-based approaches may confirm the presence of clinically overt disease, they lack the sensitivity needed to obtain accurate data on penetrance.

A critical advantage of the BRRD study is that consent is in place to approach participants to take part in follow-on studies, allowing specific phenotyping for the cardiac disorder implied by the detected ICC gene variant. Eligible individuals, along with age, gender and cohort matched controls, have been invited to participate in a follow-on clinical study, which aims to assess and evaluate the clinical significance, psychosocial and behavioural impact of pathogenic ICC variants detected as secondary findings (Secondary Cardiac Findings Evaluation [SCARFE]). In contrast to previous studies using EMR data, individuals with ICC SFs will be invited for clinical assessment within a specialist ICC unit, allowing a clinically appropriate and comprehensive assessment of variant penetrance to be undertaken. Additionally, this strategy allows systematic collection and evaluation of family history, and where relevant, clinical assessment of relatives, which is not possible through EMR approaches.

The results of this follow-on clinical study will provide much needed evidence regarding variant interpretation in the context of secondary findings and inform the ongoing debate surrounding the clinical validity and utility of genetic testing in this setting. [347–350]

4.4.2.3 Study limitations

There are a number of limitations to consider when assessing the accuracy of the prevalence estimates obtained in this study. These analyses focused on known and expected pathogenic variants so the prevalence estimates obtained are likely to be conservative. For example, only a small proportion (~7%) of the 955 rare ICC gene variants detected in the BRRD cohort were considered for analyses. While this stringent approach to variant selection was necessary to ensure that the follow-on clinical studies focused on legitimate pathogenic variants for which there is an existing evidence base, inevitably this approach excludes a proportion of pathogenic variants from consideration. In future studies, to gain a more comprehensive understanding about the ICC variants in the wider population, it will be necessary to expand analyses to include less well-established and novel ICC gene variants.

Despite the stringent approach used to select variants for these analyses, it is possible that some of these variants are not pathogenic. The inherent challenges of ICC variant interpretation and classification have been described in recent studies, which show lack of concordance in ICC variant interpretation, even amongst specialist ICC clinical laboratories[339,374–376]. Additionally, despite efforts to improve the stringency and reliability of variant classification[128,212], historical errors perpetuate in the literature and mutation databases, and can confound variant interpretation[374,377]. This highlights the critical importance of reviewing the evidence supporting variant pathogenicity, even

when considering variants previously reported to be pathogenic. As an illustration, following review, around one third of the "known" ICC variants (23/66, 35%) detected in this study, were considered to have insufficient evidence to allow classification as "Highly likely" or "Likely" to be pathogenic.

Although none of the component cohorts in the BRRD dataset were expected to be enriched for ICCs (Table 4.2-1), in the absence of clinical data, it was not possible to systematically exclude individuals with ICC disorders prior to analyses. The discovery that some of the individuals within the BRRD cohort have pre-existing clinical ICC diagnoses highlights another potential source of error for the prevalence estimates in this study.

While it is anticipated that the follow-on SCARFE study will provide much needed empirical evidence regarding variant penetrance in the context of SFs, it is acknowledged that the number of participants remains relatively small (n=~60) and penetrance estimates may only be relevant to a subset of ICC variants (n=43). However, the findings of this feasibility study are expected to inform similar studies in larger cohorts (such as UK Biobank or the 100,000 Genomes Project). Cumulative data from these studies will enable more thorough assessment of the clinical validity and utility of ICC gene testing in unselected cohorts and inform policy in this area.

Study	ICC	Population size	Genes tested	Criteria for variant selection	Prevalence	Penetrance
This study	Mixed	6,466 (unrelated) 7,878 (including related) Mean age not available Clinical data not available cohort wide ~80% European	16 genes ARVC: DSC2, DSG2, DSP, PKP2 DCM: MYH7, TNNT2, TNNI3, LMNA, DSP LQT:KCNH2, KCNQ1, SCN5A HCM: MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1	Known (OGML and LMM clinical laboratories) and expected pathogenic Reviewed and classified as Path/Likely Path	Path & Likely Path Total: 45/6,466 (0.69%) ARVC: 18/6,466 (0.28%) ARVC/DCM: 8/6466(0.12%) LQTS: 5/6,466 (0.08%) HCM: 14/6,466 (0.22%)	In progress
Haggerty et al 2017	ARVC	30,716 individuals recruited to MyCode repository Predominantly European ancestry No exclusions made for relatedness or ancestry	7 genes PKP2, DSG2, DSC2, DSP, JUP, TMEM43, TGFβ3	Database listed as Pathogenic or Likely pathogenic identified and evidence reviewed. Post evidence review 9 Putative LOF variants (pLOF) classified as P/LP. All missense classified as VUS or likely benign.	pLOF: 18/30,716 (0.06%)	Penetrance assessed using electronic health records: 0%
Van Driest et al 2016	Arrhythmia	2022 individuals recruited for non-antiarrhythmic drug exposure phenotypes	2 genes KCNH2 and SCN5A	Variants assessed by three laboratories with ion channel expertise and the ClinVar database.	Path or Likely Path (classified by at least one laboratory or ClinVar) Total: 63/2022 (3%)	Penetrance assessed using electronic medical records: 35% (22/63) Had ECG or arrhythmia phenotype, only 2 had QTC >500 milliseconds

Study	ICC	Population size	Genes tested	Criteria for variant selection	Prevalence	Penetrance
Natarajan et al 2016	Mixed	3,685 462 European American Framingham Heart Study (FHS) n=290 Offspring cohort (mean age 36.8yrs) n=172 Generation 3 cohort (mean age at enrolment, 44.5 yrs.) Mean age at enrolment 36.8 3223 African-American Jackson Heart Study (JHS) Mean age at enrolment 55.6 years	(Selected from ACMG 56) 11 HCM: MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL3, ACTC1, PRKAG2, GLA, MYL2, LMNA 1 CPVT: RYR2 5 ARVC: PKP2, DSP, DSC2, TMEM43, DSG2 3 LQTS: KCNQ1, KCNH2, SCN5A	ACMG classification Path and Likely path.	Total P/LP: FHD (Combined ICC):3/462 (0.65%) JHS (Combined ICC):14/3,233 (0.43%)	FHS clinical phenotypes from the most recent clinical examination. Participants followed for an average of 20.4 (SD, 14.3) years. 2/3 (66.7%) JHS participant clinical phenotypes from the baseline examination. 4/14 (28.6%)
Ng et al 2013	Mixed	870 ClinSeq participants Mean age 57 years (range 45-65) 45-65 years Phenotyping (ECG, ECHO, MDCT coronary Ca Not selected for FH or ICC or SCD 89% Caucasian	63 genes WES/WGS Targeted analysis of, 22 Arrhythmia (AF, BS, CPVT, LQTS, SQTS) 41 Cardiomyopathy (ARVC, DCM, HCM, LVNC)	Path and Likely path study specific criteria ICC specific frequency filter applied to remove variants considered too common to be disease causing (>1/2000 BS, >1/10,000 CPVT, >1/25000 LQTS, >1/1000 ARVC, >1/500 HCM, >1/500 DCM) HGMD and LSDB reviewed	Pathogenic Total: 6/870 (0.68%) DCM: 1/870 (0.1%) HCM: 2/870(0.2%) LVNC: 1/870 (0.1%) LQT: 2/870 (0.2%)	Phenotypes assess based on clinical data and family history data collected at recruitment. Total: 3/6 (50%)

Study	ICC	Population size	Genes tested	Criteria for variant selection	Prevalence	Penetrance
Bick et al 2012	HCM	3600 1,637 unrelated individuals form Offspring cohort of Framingham Heart Study 1.963 unrelated individuals from Jackson Heart Study	8 genes MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1	Rare known or likely pathogenic sarcomere variants based on the Laboratory for Molecular Medicine criteria. (NB: Analyses undertaken pre ExAC)	Path & Likely Path Total: 22/3600 (0.6%)	Clinical data collected at study recruitment. 4/22 (18%)

Table 4-4.4-1 Summary of previous published studies investigating variant prevalence and penetrance in ICC disorders.

4.5 Conclusions

Analyses of array data from the UKBB cohort were mostly unsuccessful and did not provide reliable data on variant prevalence; the findings in this study highlight the limitations of array methodology for genotyping very rare variation.

Analysis of genomic sequencing data from the BRRD cohort provides relatively reliable estimates of the prevalence of pathogenic ICC gene variants in unselected cohorts. These findings, along with the results of ongoing clinical studies being undertaken to investigate variant penetrance, contribute much needed evidence to inform the ongoing debate about the clinical validity and utility of testing ICC genes in individuals with no personal or family history of disease [347–350].

Importantly, this knowledge enhances understanding about the natural history and impact of pathogenic variants in ICC genes in the wider population.

Chapter 5. Conclusions

5.1 Aims

The aims of this study were to refine understanding of the causal genes and variants in inherited cardiomyopathies, to improve the variant interpretation in causal genes, and to identify genes not currently appropriate for clinical genetic testing.

5.2 Summary of findings

Initial analyses were undertaken to reassess the genes in current clinical test panels. This involved in-depth comparison of rare variation from two large clinical case cohorts and a large reference cohort. These analyses highlighted the genes, regions of genes, and classes of variant that can be most reliably interpreted in a clinical setting. Additionally, they demonstrated that in confirmed genes, there is scope to further refine variant interpretation and increase the yield of clinically actionable results. Importantly, the findings in this study supported the hypothesis that some of the genes implicated in these disorders may not be disease causing, and therefore may not be appropriate for testing in a clinical setting.

In a follow-on study, employing a similar case-control approach, the contribution of 60 genes implicated in HCM was examined in a selected cohort of 230 sarcomere gene-negative HCM cases. These analyses highlighted a possible casual role for truncating variants in the *ALPK3* gene in HCM. However, in the vast majority of genes analysed, the gene-disease relationship could not be validated, and lack of evidence precluded variant interpretation leading to a high number of uncertain results. Notably, the diagnostic

yield of these analyses was extremely low, with just a single clinically actionable result (a splice site variant in the *FLNC* gene). The results of this study indicated that extended panel tests have limited clinical utility in HCM and highlighted the need for rigorous, evidence-based selection of genes for clinical testing.

Analyses were also undertaken to explore the prevalence of pathogenic variants in key inherited cardiac condition (ICC) genes in large reference cohorts. The results of these analyses supported the hypothesis that the population prevalence of disease-causing variants in some ICC genes is higher than expected. They also established prevalence estimates for pathogenic ICC gene variants in a large reference cohort, and provided the foundation for follow-on studies into variant penetrance. These analyses demonstrated the technical challenges of array based analyses in the context of very rare variation. Also, they highlighted the various sources of bias that limit the accuracy of current population prevalence and penetrance estimates.

5.3 Implications of findings

The results of this study demonstrate that consolidated analysis of genomic sequencing data from large case and reference cohorts can enable comprehensive and robust statistical assessment of causal genes and putative causal variants[378–381].

Collectively, the analyses undertaken in this study refine understanding of the genetic architecture of inherited cardiomyopathies; confirming the key causal genes and classes of variant that contribute to these disorders. Importantly, the results of this study provide much needed empirical data which will inform current gene and variant curation initiatives[212–214]. This will facilitate evidence-based selection of genes for

clinical test panels, and ensure that genetic testing strategies are effective and clinically beneficial.

The findings in the sarcomere-negative HCM cohort indicate that testing beyond the confirmed well-established genes has limited clinical utility. Certainly, at present, there is insufficient evidence to support the use of large gene panel tests, or exome and genome sequencing approaches, in a clinical diagnostic setting.

The results of this study also contribute data regarding the potential impact of ICC gene analyses in individuals with no personal or family history of disease. This will inform the emerging evidence base relating to the analysis and reporting of variants in ICC genes in the context of secondary findings[143].

5.4 Future work

5.4.1 Case-control analyses

Case-control analyses are being expanded to incorporate more recent data from both cardiomyopathy and arrhythmia clinical cohorts. This will increase the power of these analyses and provide stronger statistical support for results conclusions; this will be particularly valuable for genes where only small case numbers were available for the analyses undertaken in this study.

Efforts are also being made to access data from additional sarcomere-negative HCM cases from other whole genomic sequencing (WGS) studies (e.g. the 100,000 Genomes Project). This would increase the power of these analyses and enable more robust conclusions to be drawn regarding the contribution of the 60 genes considered in this study. Also, this would enable exploration of a broader range of candidate genes, both

those already implicated in the literature, and others that may follow. This cohort could also prove to be a valuable resource for disease-gene discovery; however, a significantly larger case series would be required to achieve the sufficient power for genome-wide analyses[228].

5.4.2 Additional genetic analyses

Additional genetic analyses are being undertaken in the sarcomere-screen negative HCM cases recruited to this study. These include screening for copy number variants in HCM genes (although on the basis of previous studies, the yield of these analyses is expected to be low[208–210]) and exploring the contribution of mitochondrial DNA variants, which have been detected previously in individuals with HCM[16,382,383]. Also, WGS analysis will be performed on a case-by-case basis. The success rate of WGS in single cases with presumed autosomal dominant disorders is typically low[384–387], and without segregation studies and/or functional evidence it will be difficult to assign pathogenicity to any candidate variants[384]. Nevertheless, these analyses will provide empirical data about the diagnostic yield of WGS analysis in sarcomere-negative HCM, and the potential utility of this approach in a clinical diagnostic setting.

5.4.3 Assessing the penetrance of pathogenic variants detected in BRRD non-HCM cohorts

As discussed in Chapter 4, individuals from the Bioresource rare disease cohort (BRRD) in whom pathogenic ICC variants were detected, have been invited to participate in a follow-on clinical study, Secondary Cardiac Findings Evaluation (SCARFE). This study aims to assess and evaluate the clinical significance and psychosocial and behavioural impact of pathogenic ICC variants detected as secondary findings. This study is ongoing,

however it is hoped that the results of these analyses will contribute to the debate about whether secondary findings in ICC genes should be sought and returned.

5.4.4 Clinical studies in sarcomere-negative HCM cohort

Further clinical analyses will also be undertaken in the sarcomere-negative HCM cases analysed within this project. Lack of a pathogenic variant despite extensive genetic analyses, and absence of family history, increases the probability that their cardiomyopathy has a non-Mendelian basis. In a recent study, approximately 40% of HCM probands were reported to have a "non-familial" subtype of HCM, defined by negative family history and negative genetic test. On average, individuals within this "non-familial" subtype had a later disease onset and less severe clinical course than individuals with sarcomeric-HCM[191]. More detailed exploration of the clinical characteristics of individuals within our sarcomere-negative cohort could provide an opportunity to validate previous findings[191] and more clearly delineate the proposed "non-familial" clinical subgroup. These clinical observations could have implications for clinical management strategies in HCM; a combination of clinical, family history, and genetic analyses could potentially be used to change management in some of these individuals, for example, reduced clinical follow-up[191].

5.5 Conclusion

In conclusion, consolidated analysis of sequence data from large case series and unselected reference cohorts has provided valuable insight into the causal genes and variants in inherited cardiomyopathies. Importantly, the methods described could be applied to other dominant Mendelian disease genes, and could enhance understanding and interpretation of rare variation across a range of rare diseases. The results of these

analyses will enable more effective evidence-based selection of genes for clinical test panels and reduce uncertainty in results interpretation. It is hoped that this will increase the yield of clinically actionable variants, and consequently, the clinical utility of genetic testing in these medically important conditions.

Bibliography

1. Watkins, H., Ashrafian, H. & Redwood, C. Inherited Cardiomyopathies. *N. Engl. J. Med.* **364,** 1643–1656 (2011).

2. Cecchi, F., Tomberli, B. & Olivotto, I. Clinical and molecular classification of cardiomyopathies. *Glob. Cardiol. Sci. Pract.* **2012,** 4 (2012).

3. Wilcox, J. E. & Hershberger, R. E. Genetic cardiomyopathies. *Curr. Opin. Cardiol.* 1 (2018). doi:10.1097/HCO.0000000000000512

4. Elliott, P. M. *et al.* 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur. Heart J.* **35,** 2733–2779 (2014).

5. Gersh, B. J. *et al.* 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **124,** 2761–96 (2011).

6. Fatkin, D. Guidelines for the Diagnosis and Management of Familial Dilated Cardiomyopathy. *Hear. Lung Circ.* **20,** 691–693 (2011).

7. Smith, W. Guidelines for the Diagnosis and Management of Arrhythmogenic Right Ventricular Cardiomyopathy. *Hear. Lung Circ.* **20,** 757–760 (2011).

8. Charron, P. *et al.* Genetic counselling and testing in cardiomyopathies: A position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. *Eur. Heart J.* **31,** 2715–2728 (2010).

9. Hershberger, R. E. *et al.* Genetic Evaluation of Cardiomyopathy-A Heart Failure Society of America Practice Guideline. *J. Card. Fail.* **15,** 83–97 (2009).

10. Maron, B. J. *et al.* 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,** 785–789 (1995).

11. Seidman, J. G. & Seidman, C. The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* **104,** 557–567 (2001).

12. Varnava, A. M., Elliott, P. M., Mahon, N., Davies, M. J. & McKenna, W. J. Relation between myocyte disarray and outcome in hypertrophic cardiomyopathy. *Am. J. Cardiol.* **88,** 275–279 (2001).

13. Maron, B. J. Hypertrophic cardiomyopathy: a systematic review. *JAMA* **287,** 1308–1320 (2002).

14. Arad, M. *et al.* Glycogen Storage Diseases Presenting as Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **352,** 362–372 (2005).

15. Merante, F., Tein, I., Benson, L. & Robinson, B. H. Maternally inherited hypertrophic cardiomyopathy due to a novel T-to-C transition at nucleotide 9997 in the mitochondrial

tRNA(glycine) gene. *Am. J. Hum. Genet.* **55,** 437–446 (1994).

16. Liu, Z. *et al.* The novel mitochondrial 16S rRNA 2336T>C mutation is associated with hypertrophic cardiomyopathy. *J. Med. Genet.* **51,** 176–184 (2014).

17. Silvestri, G. *et al.* A new mtDNA mutation in the tRNA(Leu(UUR)) gene associated with maternally inherited cardiomyopathy. *Hum. Mutat.* **3,** 37–43 (1994).

18. Thierfelder, L. *et al.* Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. *Cell* **77,** 701–712 (1994).

19. Geisterfer-Lowrance, A. A. T. *et al.* A molecular basis for familial hypertrophic cardiomyopathy: A β cardiac myosin heavy chain gene missense mutation. *Cell* **62,** 999–1006 (1990).

20. Solomon, S. D. *et al.* A locus for familial hypertrophic cardiomyopathy is closely linked to the cardiac myosin heavy chain genes, CRI-L436, and CRI-L329 on chromosome 14 at q11-q12. *Am. J. Hum. Genet.* **47,** 389–394 (1990).

21. Watkins, H. *et al.* Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. *Nat. Genet.* **11,** 434–7 (1995).

22. Bonne, G. *et al.* Cardiac myosin binding protein-C gene splice acceptor site mutation is associated with familial hypertrophic cardiomyopathy. *Nat. Genet.* **11,** 438–40 (1995).

23. Alders, M. *et al.* The 2373insG mutation in the MYBPC3 gene is a founder mutation, which accounts for nearly one-fourth of the HCM cases in the Netherlands. *Eur. Heart J.* **24,** 1848–1853 (2003).

24. van der Linde, I. H. M. *et al.* A Dutch MYH7 founder mutation, p.(Asn1918Lys), is associated with early onset cardiomyopathy and congenital heart defects. *Netherlands Hear. J.* 675–681 (2017). doi:10.1007/s12471-017-1037-5

25. Kuusisto, J., Sipola, P., Jaaskelainen, P. & Naukkarinen, A. Current perspectives in hypertrophic cardiomyopathy with the focus on patients in the Finnish population: a review. *Ann. Med.* **48,** 496–508 (2016).

26. Saltzman, A. J. *et al.* Short communication: the cardiac myosin binding protein C Arg502Trp mutation: a common cause of hypertrophic cardiomyopathy. *Circ. Res.* **106,** 1549–1552 (2010).

27. Alfares, A. a. *et al.* Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet. Med.* **17,** 880–888 (2015).

28. Maron, B. J., Maron, M. S. & Semsarian, C. Double or compound sarcomere mutations in hypertrophic cardiomyopathy : A potential link to sudden death in the absence of conventional risk factors. *HRTHM* **9,** 57–63 (2012).

29. Girolami, F. *et al.* Clinical features and outcome of hypertrophic cardiomyopathy associated with triple sarcomere protein gene mutations. *J. Am. Coll. Cardiol.* **55,** 1444–1453 (2010).

30. Ingles, J. *et al.* Compound and double mutations in patients with hypertrophic cardiomyopathy: implications for genetic testing and counselling. *J Med Genet* **42,** e59 (2005).

31. García-castro, M., Coto, E., Reguero, J. R., Berrazueta, J. R. & Álvarez, V. Mutations in

Sarcomeric Genes MYH7 , MYBPC3 , TNNT2 , TNNI3 , and TPM1 in Patients With Hypertrophic Cardiomyopathy. **62,** 48–56 (2009).

32. Fourey, D. *et al.* Prevalence and Clinical Implication of Double Mutations in Hypertrophic CardiomyopathyCLINICAL PERSPECTIVE. *Circ. Cardiovasc. Genet.* **10,** e001685 (2017).

33. Burns, C., Bagnall, R. D., Lam, L., Semsarian, C. & Ingles, J. Multiple Gene Variants in Hypertrophic Cardiomyopathy in the Era of Next-Generation Sequencing. *Circ. Cardiovasc. Genet.* **10,** (2017).

34. Wessels, M. W. *et al.* Compound heterozygous or homozygous truncating MYBPC3 mutations cause lethal cardiomyopathy with features of noncompaction and septal defects. *Eur. J. Hum. Genet.* **23,** 922–928 (2014).

35. Marziliano, N. *et al.* A case of compound mutations in the MYBPC3 gene associated with biventricular hypertrophy and neonatal death. *Neonatology* **102,** 254–258 (2012).

36. Ortiz, M. F. *et al.* A homozygous MYBPC3 gene mutation associated with a severe phenotype and a high risk of sudden death in a family with hypertrophic cardiomyopathy. *Rev. Esp. Cardiol.* **62,** 572–575 (2009).

37. Zahka, K. *et al.* Homozygous mutation of MYBPC3 associated with severe infantile hypertrophic cardiomyopathy at high frequency among the Amish. *Heart* **94,** 1326–1330 (2008).

38. Xin, B., Puffenberger, E., Tumbush, J., Bockoven, J. R. & Wang, H. Homozygosity for a novel splice site mutation in the cardiac myosin-binding protein C gene causes severe neonatal hypertrophic cardiomyopathy. *Am. J. Med. Genet. A* **143A,** 2662–2667 (2007).

39. Lekanne Deprez, R. H. *et al.* Two cases of severe neonatal hypertrophic cardiomyopathy caused by compound heterozygous mutations in the MYBPC3 gene. *Journal of medical genetics* **43,** 829–832 (2006).

40. Frey, N., Luedde, M. & Katus, H. A. Mechanisms of disease: Hypertrophic cardiomyopathy. *Nat. Rev. Cardiol.* **9,** 91–100 (2012).

41. Redwood, C. S., Moolman-Smook, J. C. & Watkins, H. Properties of mutant contractile proteins that cause hypertrophic cardiomyopathy. *Cardiovasc. Res.* **44,** 20–36 (1999).

42. Palmer, B. M. *et al.* Differential cross-bridge kinetics of FHC myosin mutations R403Q and R453C in heterozygous mouse myocardium. *Am. J. Physiol. Hear. Circ.* **287,** H91-9 (2004).

43. Witjas-Paalberends, E. R. *et al.* Faster cross-bridge detachment and increased tension cost in human hypertrophic cardiomyopathy with the R403Q MYH7 mutation. *J. Physiol.* **592,** 3257–3272 (2014).

44. Robinson, P. *et al.* Alterations in thin filament regulation induced by a human cardiac troponin T mutant that causes dilated cardiomyopathy are distinct from those induced by troponin T mutants that cause hypertrophic cardiomyopathy. *J. Biol. Chem.* **277,** 40710–40716 (2002).

45. Ormerod, J. O. M., Frenneaux, M. P. & Sherrid, M. V. Myocardial energy depletion and dynamic systolic dysfunction in hypertrophic cardiomyopathy. *Nat. Rev. Cardiol.* **13,** 677–687 (2016).

46. Crilley, J. G. *et al.* Hypertrophic cardiomyopathy due to sarcomeric gene mutations is

characterized by impaired energy metabolism irrespective of the degree of hypertrophy. *J. Am. Coll. Cardiol.* **41,** 1776–1782 (2003).

47. Pinto, Y. M. *et al.* Clinical utility gene card for: hypertrophic cardiomyopathy (type 1–14). *Eur. J. Hum. Genet.* **19,** 1–4 (2011).

48. Geier, C. *et al.* Mutations in the human muscle LIM protein gene in families with hypertrophic cardiomyopathy. *Circulation* **107,** 1390–1395 (2003).

49. Chiu, C. *et al.* Mutations in Alpha-Actinin-2 Cause Hypertrophic Cardiomyopathy. A Genome-Wide Analysis. *J. Am. Coll. Cardiol.* **55,** 1127–1135 (2010).

50. Landstrom, A. P., Adekola, B. A., Bos, J. M., Ommen, S. R. & Ackerman, M. J. PLN-encoded phospholamban mutation in a large cohort of hypertrophic cardiomyopathy cases: summary of the literature and implications for genetic testing. *Am. Heart J.* **161,** 165–171 (2011).

51. Arimura, T. *et al.* Cardiac Ankyrin Repeat Protein Gene (ANKRD1) Mutations in Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* **54,** 334–342 (2009).

52. Osio, A. *et al.* Myozenin 2 Is a Novel Gene for Human Hypertrophic Cardiomyopathy. *Circ. Res.* **100,** 766–768 (2007).

53. Tanigawa, G. *et al.* A molecular basis for familial hypertrophic cardiomyopathy: An alpha/beta cardiac myosin heavy chain hybrid gene. *Cell* **62,** 991–998 (1990).

54. Hayashi, T. *et al.* Tcap gene mutations in hypertrophic cardiomyopathy and dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **44,** 2192–2201 (2004).

55. Teekakirikul, P., Kelly, M. a., Rehm, H. L., Lakdawala, N. K. & Funke, B. H. Inherited cardiomyopathies: Molecular genetics and clinical genetic testing in the postgenomic era. *J. Mol. Diagnostics* **15,** 158–170 (2013).

56. Waldmüller, S. *et al.* Targeted 46-gene and clinical exome sequencing for mutations causing cardiomyopathies. *Mol. Cell. Probes* 1–7 (2015). doi:10.1016/j.mcp.2015.05.004

57. Mook, O. R. F. *et al.* Targeted sequence capture and GS-FLX Titanium sequencing of 23 hypertrophic and dilated cardiomyopathy genes: implementation into diagnostics. *J. Med. Genet.* **50,** 614–26 (2013).

58. Authors/Task Force members *et al.* 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur. Heart J.* **35,** 2733–79 (2014).

59. Ackerman, M. J. *et al.* HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies. *Europace* **13,** 1077–1109 (2011).

60. Wilde, A. A. M. & Behr, E. R. Genetic testing for inherited cardiac disease. *Nat. Rev. Cardiol.* **10,** 571–583 (2013).

61. Hershberger, R. E., Hedges, D. J. & Morales, A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* **10,** 531–47 (2013).

62. Petretta, M., Pirozzi, F., Sasso, L., Paglia, A. & Bonaduce, D. Review and metaanalysis of the frequency of familial dilated cardiomyopathy. *Am. J. Cardiol.* **108,** 1171–6 (2011).

63. Codd, M. B., Sugrue, D. D., Gersh, B. J. & Melton, L. J. 3rd. Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted County, Minnesota, 1975-1984. *Circulation* **80,** 564–572 (1989).

64. Fatkin, D. *et al.* 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,** 1715–1724 (1999).

65. Kamisago, M. *et al.* Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N. Engl. J. Med.* **343,** 1688–1696 (2000).

66. McNally, E. M. & Mestroni, L. Dilated cardiomyopathy: Genetic determinants and mechanisms. *Circ. Res.* **121,** 731–748 (2017).

67. Herman, D. S. *et al.* Truncations of Titin Causing Dilated Cardiomyopathy. *New England Journal of Medicine* **366,** 619–628 (2012).

68. Roberts, A. M. *et al.* Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* **7,** 270ra6 (2015).

69. Ware, J. S. & Cook, S. A. Role of titin in cardiomyopathy: from DNA variants to patient stratification. *Nat. Rev. Cardiol.* **15,** 241–252 (2018).

70. Akinrinade, O., Koskenvuo, J. W. & Alastalo, T. P. Prevalence of titin truncating variants in general population. *PLoS One* **10,** 1–14 (2015).

71. Roberts, A. M. *et al.* Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* (2015). doi:10.1126/scitranslmed.3010134

72. Schafer, S. *et al.* Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat. Genet.* **49,** 46–53 (2017).

73. Deo, R. C. Alternative Splicing, Internal Promoter, Nonsense-Mediated Decay, or All Three: Explaining the Distribution of Truncation Variants in Titin. Circulation: Cardiovascular Genetics **9,** (2016).

74. Tabish, A. M., Azzimato, V., Alexiadis, A., Buyandelger, B. & Knöll, R. Genetic epidemiology of titin-truncating variants in the etiology of dilated cardiomyopathy. *Biophys. Rev.* **9,** 207–223 (2017).

75. Arbustini, E. *et al.* Autosomal dominant dilated cardiomyopathy with atrioventricular block: A lamin A/C defect-related disease. *J. Am. Coll. Cardiol.* **39,** 981–990 (2002).

76. Captur, G. *et al.* Lamin and the heart. *Heart* **104,** 468–479 (2018).

77. Pugh, T. J. *et al.* The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet. Med.* **16,** 601–8 (2014).

78. Hershberger, R. E. & Siegfried, J. D. Update 2011: Clinical and Genetic Issues in Familial Dilated Cardiomyopathy. *J. Am. Coll. Cardiol.* **57,** 1641–1649 (2011).

79. Ackerman, M. J. *et al.* 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). *Hear. Rhythm* **8,** 1308–1339 (2011).

80. Mogensen, J. *et al.* Severe disease expression of cardiac troponin C and T mutations in patients with idiopathic dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **44,** 2033–2040 (2004).

81. Dyer, E. C. *et al.* Functional analysis of a unique troponin c mutation, gly159asp, that causes Familial dilated cardiomyopathy, studied in explanted heart muscle. *Circ. Hear. Fail.* **2,** 456–464 (2009).

82. Olson, T. M., Kishimoto, N. Y., Whitby, F. G. & Michels, V. V. Mutations that alter the surface charge of α-tropomyosin are associated with dilated cardiomyopathy. *J. Mol. Cell Cardiol.* **33,** 723–732. (2001).

83. Li, D. *et al.* Novel cardiac troponin T mutation as a cause of familial dilated cardiomyopathy. *Circulation* **104,** 2188–2193 (2001).

84. Robinson, P., Griffiths, P. J., Watkins, H. & Redwood, C. S. Dilated and hypertrophic cardiomyopathy mutations in troponin and α-tropomyosin have opposing effects on the calcium affinity of cardiac thin filaments. *Circ. Res.* **101,** 1266–1273 (2007).

85. Mirza, M. *et al.* Dilated cardiomyopathy mutations in three thin filament regulatory proteins result in a common functional phenotype. *J. Biol. Chem.* **280,** 28498–28506 (2005).

86. Schmitt, J. P. *et al.* Cardiac myosin missense mutations cause dilated cardiomyopathy in mouse models and depress molecular motor function. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 14525–14530 (2006).

87. Debold, E. P. *et al.* Hypertrophic and dilated cardiomyopathy mutations differentially affect the molecular force generation of mouse alpha-cardiac myosin in the laser trap assay. *Am. J. Physiol. Heart Circ. Physiol.* **293,** H284-91 (2007).

88. Mogensen, J. *et al.* Severe disease expression of cardiac troponin C and T mutations in patients with idiopathic dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **44,** 2033–2040 (2004).

89. Duboscq-Bidot, L. *et al.* Mutations in the Z-band protein myopalladin gene and idiopathic dilated cardiomyopathy. *Cardiovasc. Res.* **77,** 118–125 (2008).

90. Olson, T. M. *et al.* Metavinculin mutations alter actin interaction in dilated cardiomyopathy. *Circulation* **105,** 431–437 (2002).

91. Taylor, M. R. G. *et al.* Prevalence of desmin mutations in dilated cardiomyopathy. *Circulation* **115,** 1244–1251 (2007).

92. Ortiz-Genga, M. F. *et al.* Truncating FLNC Mutations Are Associated With High-Risk Dilated and Arrhythmogenic Cardiomyopathies. *J. Am. Coll. Cardiol.* **68,** 2440–2451 (2016).

93. Garcia-Pavia, P. *et al.* Desmosomal protein gene mutations in patients with idiopathic dilated cardiomyopathy undergoing cardiac transplantation: A clinicopathological study. *Heart* **97,** 1744–1752 (2011).

94. Refaat, M. M. *et al.* Genetic variation in the alternative splicing regulator RBM20 is associated with dilated cardiomyopathy. *Hear. Rhythm* **9,** 390–396 (2012).

95. Haas, J. *et al.* Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur. Heart J.* **36,** 1123–35a (2015).

96. McNally, E. M., Golbus, J. R. & Puckelwartz, M. J. Genetic mutations and mechanisms in dilated cardiomyopathy. *J. Clin. Invest.* **123,** 19–26 (2013).

97. Hoorntje, E. T. *et al.* Arrhythmogenic cardiomyopathy: Pathology, genetics, and concepts in pathogenesis. *Cardiovasc. Res.* **113,** 1521–1531 (2017).

98. Sen-Chowdhry, S., Morgan, R. D., Chambers, J. C. & McKenna, W. J. Arrhythmogenic Cardiomyopathy: Etiology, Diagnosis, and Treatment. *Annu. Rev. Med.* **61,** 233–253 (2010).

99. Sen-Chowdhry, S. *et al.* Left-Dominant Arrhythmogenic Cardiomyopathy. An Under-Recognized Clinical Entity. *J. Am. Coll. Cardiol.* **52,** 2175–2187 (2008).

100. Sen-Chowdhry, S. *et al.* Clinical and genetic characterization of families with arrhythmogenic right ventricular dysplasia/cardiomyopathy provides novel insights into patterns of disease expression. *Circulation* **115,** 1710–1720 (2007).

101. Towbin, J. A. Arrhythmogenic right ventricular cardiomyopathy: A paradigm of overlapping disorders. *Ann. Noninvasive Electrocardiol.* **13,** 325–326 (2008).

102. Sen-Chowdhry, S., Syrris, P. & McKenna, W. J. Genetics of right ventricular cardiomyopathy. *J. Cardiovasc. Electrophysiol.* **16,** 927–935 (2005).

103. Garrod, D. & Chidgey, M. Desmosome structure, composition and function. *Biochim. Biophys. Acta - Biomembr.* **1778,** 572–587 (2008).

104. Van Der Zwaag, P. A. *et al.* A genetic variants database for arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Hum. Mutat.* **30,** 1278–1283 (2009).

105. Kapplinger, J. D. *et al.* Distinguishing arrhythmogenic right ventricular cardiomyopathy/dysplasia-associated mutations from background genetic noise. *J. Am. Coll. Cardiol.* **57,** 2317–27 (2011).

106. Rigato, I. *et al.* Compound and digenic heterozygosity predicts lifetime arrhythmic outcome and sudden cardiac death in desmosomal gene-related arrhythmogenic right ventricular cardiomyopathy. *Circ. Cardiovasc. Genet.* **6,** 533–542 (2013).

107. Xu, T. *et al.* Compound and Digenic Heterozygosity Contributes to Arrhythmogenic Right Ventricular Cardiomyopathy. *J. Am. Coll. Cardiol.* **55,** 587–597 (2010).

108. Kirchhof, P. *et al.* Age- and training-dependent development of arrhythmogenic right ventricular cardiomyopathy in heterozygous plakoglobin-deficient mice. *Circulation* **114,** 1799–1806 (2006).

109. Remme, C. A. & Bezzina, C. R. Cardiac desmosomal (dys)function and myocyte viability. *Cell Cycle* **9,** 1246–1252 (2010).

110. Kaplan, S. R. *et al.* Remodeling of myocyte gap junctions in arrhythmogenic right ventricular cardiomyopathy due to a deletion in plakoglobin (Naxos disease). *Hear. Rhythm* **1,** 3–11 (2004).

111. Garcia-gras, E. *et al.* Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhytmogenic right ventricular cardiomyopathy. *J. Clin. Invest.* **116,** 2012–2021 (2006).

112. Lombardi, R. & Marian, A. J. Molecular genetics and pathogenesis of arrhythmogenic right ventricular cardiomyopathy: A disease of cardiac stem cells. *Pediatr. Cardiol.* **32,** 360–365 (2011).

113. Haywood, A. F. M. *et al.* Recurrent missense mutations in TMEM43 (ARVD5) due to founder

effects cause arrhythmogenic cardiomyopathies in the UK and Canada. *Eur. Heart J.* **34,** 1002–1011 (2013).

114. Liang, J. J., Goodsell, K., Grogan, M. & Ackerman, M. J. LMNA-Mediated Arrhythmogenic Right Ventricular Cardiomyopathy and Charcot-Marie-Tooth Type 2B1: A Patient-Discovered Unifying Diagnosis. *J. Cardiovasc. Electrophysiol.* **27,** 868–871 (2016).

115. Kato, K. *et al.* LMNA cardiomyopathy detected in Japanese arrhythmogenic right ventricular cardiomyopathy cohort. *J. Cardiol.* **68,** 346–351 (2016).

116. Mayosi, B. M. *et al.* Identification of Cadherin 2 (CDH2) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circ. Cardiovasc. Genet.* **10,** (2017).

117. Turkowski, K. L., Tester, D. J., Bos, J. M., Haugaa, K. H. & Ackerman, M. J. Whole exome sequencing with genomic triangulation implicates CDH2-encoded N-cadherin as a novel pathogenic substrate for arrhythmogenic cardiomyopathy. *Congenit. Heart Dis.* **12,** 226–235 (2017).

118. Taylor, M. *et al.* Genetic Variation in Titin in ARVC-Overlap Syndromes. *Circulation* **124,** 876–885 (2011).

119. Roux-Buisson, N. *et al.* Prevalence and significance of rare RYR2 variants in arrhythmogenic right ventricular cardiomyopathy/dysplasia: Results of a systematic screening. *Hear. Rhythm* **11,** 1999–2009 (2014).

120. Lorenzon, A. *et al.* Desmin Mutations and Arrhythmogenic Right Ventricular Cardiomyopathy. *Am. J. Cardiol.* **111,** 400–405 (2013).

121. Hedberg, C., Melberg, A., Kuhl, A., Jenne, D. & Oldfors, A. Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy 7 is caused by a DES mutation. *Eur. J. Hum. Genet.* **20,** 984–985 (2012).

122. Karmouch, J., Protonotarios, A. & Syrris, P. Genetic basis of arrhythmogenic cardiomyopathy. *Curr. Opin. Cardiol.* 1 (2018). doi:10.1097/HCO.0000000000000509

123. Pugh, T. J. *et al.* The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet. Med.* **16,** 601–8 (2014).

124. Alfares, A. a. *et al.* Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet. Med.* (2015). doi:10.1038/gim.2014.205

125. Magnusson, P., Palm, A., Branden, E. & Mörner, S. Misclassification of hypertrophic cardiomyopathy: Validation of diagnostic codes. *Clin. Epidemiol.* **9,** 403–410 (2017).

126. Haas, J. *et al.* Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur. Heart J.* (2014). doi:10.1093/eurheartj/ehu301

127. Wordsworth, S. *et al.* DNA testing for hypertrophic cardiomyopathy: A cost-effectiveness model. *Eur. Heart J.* **31,** 926–935 (2010).

128. Richards, S., Aziz, N., Bale, S., Bick, D. & Das, S. ACMG Standards and Guidelines Standards and guidelines for the interpretation of sequence variants : a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. **17,** 1–20 (2015).

129. Jarvik, G. P. & Browning, B. L. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am. J. Hum. Genet.* **98,** 1077–1081 (2016).

130. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7,** (2010).

131. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31,** (2003).

132. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nat. Methods* **11,** 361–362 (2014).

133. Kircher, M. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. g* **46,** 310–315 (2014).

134. Flanagan, S. E., Patch, A.-M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet. Test. Mol. Biomarkers* **14,** 533–537 (2010).

135. Ernst, C. *et al.* Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med. Genomics 2018 111* **11,** 35 (2018).

136. Grimm, D. G. *et al.* The Evaluation of Tools Used to Predict the Impact of Missense Variants Is Hindered by Two Types of Circularity. *Hum. Mutat.* **36,** 513–523 (2015).

137. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337,** 64–69 (2012).

138. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).

139. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).

140. Andreasen, C. *et al.* New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur. J. Hum. Genet.* **21,** 918–28 (2013).

141. Bick, A. G. *et al.* Burden of rare sarcomere gene variants in the framingham and jackson heart study cohorts. *Am. J. Hum. Genet.* **91,** 513–519 (2012).

142. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508,** 469–76 (2014).

143. Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15,** 565–74 (2013).

144. Xue, Y., Ankala, A., Wilcox, W. R. & Hegde, M. R. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet. Med.* **17,** (2014).

145. NCBI. NCBI Genetic Test Registry. at <https://www.ncbi.nlm.nih.gov/gtr/>

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

147. Wooderchak-Donahue, W. *et al.* Clinical utility of a next generation sequencing panel assay

for Marfan and Marfan-like syndromes featuring aortopathy. *Am. J. Med. Genet. Part A* **167,** 1747–1757 (2015).

148. Nijman, I. J. *et al.* Targeted next-generation sequencing: A novel diagnostic tool for primary immunodeficiencies. *J. Allergy Clin. Immunol.* **133,** 529–534.e1 (2015).

149. The 1000 Genomes Project Consortium. An integrated map of genetic variation. *Nature* **135,** 0–9 (2012).

150. NHLBI Exome Sequencing Project (ESP). at <http://evs.gs.washington.edu/EVS/>

151. Lopes, L. R. *et al.* Novel genotype-phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy. *Heart* **101,** 294–301 (2014).

152. A map of human genome variation from population scale sequencing. *Nature* **467,** 1061–1073 (2010).

153. Norton, N. *et al.* Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ. Cardiovasc. Genet.* **5,** 167–174 (2012).

154. Wallis, Y. *et al. Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics.o Title*. (2013). at <http://www.acgs.uk.com/media/774853/evaluation_and_reporting_of_sequence_variants_bpgs_june_2013_-_finalpdf.pdf>

155. Haldane, J. B. S. The estimation and significance of the logarithim of a ratio of frequencies. *Ann. Hum. Genet.* 309–311 (1955).

156. Ye, J., Pavlicek, A., Lunney, E. a, Rejto, P. a & Teng, C.-H. Statistical method on nonrandom clustering with application to somatic mutations in cancer. *BMC Bioinformatics* **11,** 11 (2010).

157. Ryslik, G. & Zhao, H. Identification of Protein Amino acid clustering (iPAC). (2012). at <https://bioconductor.org/packages/release/bioc/html/iPAC.html>

158. Kamisago, M. *et al.* Mutations in Sarcomere Protein Genes as a Cause of Dilated Cardiomyopathy. *N. Engl. J. Med.* **343,** 1688–1696 (2000).

159. Rampazzo, A. *et al.* Mutation in Human Desmoplakin Domain Binding to Plakoglobin Causes a Dominant Form of Arrhythmogenic Right Ventricular Cardiomyopathy. *Am. J. Hum. Genet.* **71,** 1200–1206 (2002).

160. Gerull, B. *et al.* Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat. Genet.* **36,** 1162–1164 (2004).

161. Yoskovitz, G. *et al.* A novel titin mutation in adult-onset familial dilated cardiomyopathy. *Am. J. Cardiol.* **109,** 1644–50 (2012).

162. Roberts, A. M. *et al.* HHS Public Access. **7,** (2015).

163. Kapplinger, J. D. *et al.* Distinguishing Hypertrophic Cardiomyopathy-Associated Mutations from Background Genetic Noise. *J. Cardiovasc. Transl. Res.* 347–361 (2014). doi:10.1007/s12265-014-9542-z

164. NCBI-ClinVar. ClinVar [CSRP3]. at <https://www.ncbi.nlm.nih.gov/clinvar/?term=CSRP3%5Bgene%5D>

165. Geier, C. *et al.* Beyond the sarcomere: CSRP3 mutations cause hypertrophic cardiomyopathy. *Hum. Mol. Genet.* **17,** 2753–2765 (2008).

166. Andersen, P. S. *et al.* Diagnostic yield, interpretation, and clinical utility of mutation screening of sarcomere encoding genes in Danish hypertrophic cardiomyopathy patients and relatives. *Hum. Mutat.* **30,** 363–370 (2009).

167. Bos, J. M. *et al.* Genotype-phenotype relationships involving hypertrophic cardiomyopathy-associated mutations in titin, muscle LIM protein, and telethonin. *Mol. Genet. Metab.* **88,** 78–85 (2006).

168. Santos, S. *et al.* High resolution melting: improvements in the genetic diagnosis of hypertrophic cardiomyopathy in a Portuguese cohort. *BMC Med. Genet.* **13,** 17 (2012).

169. Bagnall, R. D., Molloy, L. K., Kalman, J. M. & Semsarian, C. Exome sequencing identifies a mutation in the ACTN2 gene in a family with idiopathic ventricular fibrillation, left ventricular noncompaction, and sudden death. *BMC Med. Genet.* **15,** 99 (2014).

170. Girolami, F. *et al.* Novel alpha-actinin 2 variant associated with familial hypertrophic cardiomyopathy and juvenile atrial arrhythmias: a massively parallel sequencing study. *Circ. Cardiovasc. Genet.* **7,** 741–750 (2014).

171. NCBI-ClinVar. ClinVar [ACTN2]. at <https://www.ncbi.nlm.nih.gov/clinvar/?term=ACTN2%5Bgene%5D>

172. Hershberger, R. E. *et al.* Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* **3,** 155–161 (2010).

173. Daehmlow, S. *et al.* Novel mutations in sarcomeric protein genes in dilated cardiomyopathy. *Biochem. Biophys. Res. Commun.* **298,** 116–120 (2002).

174. Van Driest, S. L., Ommen, S. R., Tajik, A. J., Gersh, B. J. & Ackerman, M. J. Sarcomeric genotyping in hypertrophic cardiomyopathy. *Mayo Clin. Proc.* **80,** 463–469 (2005).

175. Marston, S. *et al.* Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency*. *Circ. Res.* **105,** 219–222 (2009).

176. Barefield, D. *et al.* Haploinsufficiency of MYBPC3 exacerbates the development of hypertrophic cardiomyopathy in heterozygous mice. *J. Mol. Cell. Cardiol.* **79,** 234–243 (2015).

177. Rottbauer, W. *et al.* Myosin Binding Protein-C Mutation Novel Splice Donor Site Mutation in the Cardiac Myosin-binding Protein-C Gene in Familial Hypertrophic Cardiomyopathy Characterization of Cardiac Transcript and Protein. *J. Clin. Invest* **100,** 475–482 (1997).

178. Marston, S. *et al.* How do MYBPC3 mutations cause hypertrophic cardiomyopathy? *J. Muscle Res. Cell Motil.* **33,** 75–80 (2012).

179. Van Dijk, S. J. *et al.* Cardiac myosin-binding protein C mutations and hypertrophic ardiomyopathy haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction. *Circulation* **119,** 1473–1483 (2009).

180. Harris, S. P. *et al.* Hypertrophic cardiomyopathy in cardiac myosin binding protein-C knockout mice. *Circ. Res.* **90,** 594–601 (2002).

181. Vignier, N. *et al.* Nonsense-mediated mrna decay and ubiquitin-proteasome system regulate

cardiac myosin-binding protein c mutant levels in cardiomyopathic mice. *Circ. Res.* **105,** 239–248 (2009).

182. McConnell, B. K. *et al.* Dilated cardiomyopathy in homozygous myosin-binding protein-C mutant mice. *J. Clin. Invest.* **104,** 1235–1244 (1999).

183. Møller, D. V. *et al.* The role of sarcomere gene mutations in patients with idiopathic dilated cardiomyopathy. *Eur. J. Hum. Genet.* **17,** 1241–1249 (2009).

184. Merlo, M. *et al.* Poor Prognosis of Rare Sarcomeric Gene Variants in Patients with Dilated Cardiomyopathy. *Clin. Transl. Sci.* **6,** 424–428 (2013).

185. Wells, Q. S. *et al.* Familial dilated cardiomyopathy associated with congenital defects in the setting of a novel VCL mutation (Lys815Arg) in conjunction with a known MYPBC3 variant. *Cardiogenetics* **1,** e10 (2011).

186. Shimizu, M. *et al.* Gene mutations in adult Japanese patients with dilated cardiomyopathy. *Circ. J.* **69,** 150–153 (2005).

187. Ehlermann, P. *et al.* Adverse events in families with hypertrophic or dilated cardiomyopathy and mutations in the MYBPC3 gene. *BMC Med. Genet.* **9,** 1–11 (2008).

188. Merlo, M. *et al.* Poor prognosis of rare sarcomeric gene variants in patients with dilated cardiomyopathy. *Clin. Transl. Sci.* **6,** 424–428 (2013).

189. Maron, B. J. *et al.* Development of left ventricular hypertrophy in adults with hypertrophic cardiomyopathy caused by cardiac myosin-binding protein C gene mutations. *J. Am. Coll. Cardiol.* **38,** 315–321 (2001).

190. Richard, P. *et al.* Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* **107,** 2227–32 (2003).

191. Ingles, J. *et al.* Clinical predictors of genetic testing outcomes in hypertrophic cardiomyopathy. *Genet Med* **15,** 972–977 (2013).

192. Van Tintelen, J. P. *et al.* Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circulation* **113,** 1650–1658 (2006).

193. Van Driest, S. L., Ommen, S. R., Tajik, A. J., Gersh, B. J. & Ackerman, M. J. Yield of genetic testing in hypertrophic cardiomyopathy. *Mayo Clin. Proc.* **80,** 739–744 (2005).

194. Richard, P. *et al.* Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* **107,** 2227–32 (2003).

195. Kimura, A. Molecular genetics and pathogenesis of cardiomyopathy. *J. Hum. Genet.* **61,** 41–50 (2016).

196. Redwood, C. *et al.* Investigation of a Truncated Cardiac Troponin T That Causes Familial Hypertrophic Cardiomyopathy. *Circ. Res.* **86,** 1146–1152 (2000).

197. Eilbeck, K., Quinlan, A. & Yandell, M. Settling the score: variant prioritization and Mendelian disease. *Nat. Rev. Genet.* **18,** 599–612 (2017).

198. Elliott, K., Watkins, H. & Redwood, C. S. Altered regulatory properties of human cardiac troponin I mutants that cause hypertrophic cardiomyopathy. *J. Biol. Chem.* **275,** 22069–22074 (2000).

199. Cheng, Y. & Regnier, M. Cardiac troponin structure-function and the influence of hypertrophic cardiomyopathy associated mutations on modulation of contractility. *Arch. Biochem. Biophys.* **601,** 11–21 (2016).

200. Van Den Wijngaard, A. *et al.* Recurrent and founder mutations in the Netherlands: Cardiac Troponin i (TNNI3) gene mutations as a cause of severe forms of hypertrophic and restrictive cardiomyopathy. *Found. Mutat. Inherit. Card. Dis. Netherlands* 43–50 (2014). doi:10.1007/978-90-368-0705-0_7

201. Rayment, I., Holden, H. M., Sellers, J. R., Fananapazir, L. & Epstein, N. D. Structural interpretation of the mutations in the beta-cardiac myosin that have been implicated in familial hypertrophic cardiomyopathy. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3864–3868 (1995).

202. Buvoli, M., Hamady, M., Leinwand, L. A. & Knight, R. Bioinformatics assessment of beta-myosin mutations reveals myosin's high sensitivity to mutations. *Trends Cardiovasc. Med.* **18,** 141–149 (2008).

203. Ryslik, G. A., Cheng, Y., Cheung, K.-H., Modis, Y. & Zhao, H. Utilizing protein structure to identify non-random somatic mutations. *BMC Bioinformatics* **14,** 190 (2013).

204. Samocha, K. E. *et al.* Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* (2017). doi:10.1101/148353

205. Chang, Y.-S., Huang, H.-D., Yeh, K.-T. & Chang, J.-G. Evaluation of whole exome sequencing by targeted gene sequencing and Sanger sequencing. *Clin. Chim. Acta.* **471,** 222–232 (2017).

206. Goldfeder, R. L. *et al.* Medical implications of technical accuracy in genome sequencing. *Genome Med.* **8,** 1–12 (2016).

207. Chen, Y. C., Liu, T., Yu, C. H., Chiang, T. Y. & Hwang, C. C. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLoS One* **8,** (2013).

208. Lopes, L. R. *et al.* Use of high-throughput targeted exome-sequencing to screen for copy number variation in hypertrophic cardiomyopathy. *Eur. J. Med. Genet.* **58,** 611–616 (2015).

209. Ceyhan-Birsoy, O. *et al.* Next generation sequencing-based copy number analysis reveals low prevalence of deletions and duplications in 46 genes associated with genetic cardiomyopathies. *Mol. Genet. Genomic Med.* **4,** 143–151 (2016).

210. Mademont-Soler, I. *et al.* Additional value of screening for minor genes and copy number variants in hypertrophic cardiomyopathy. *PLoS One* **12,** e0181465 (2017).

211. Bezzina, C. R. *et al.* Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat. Genet.* **45,** 1044–9 (2013).

212. Kelly, M. A. *et al.* Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet. Med.* **00,** 1–9 (2018).

213. ClinGen. Hypertrophic Cardiomyopathy Gene Curation. (2018). at

<https://www.clinicalgenome.org/working-groups/clinical-domain/sub-groups/cardiovascular/hypertrophic-cardiomyopathy-gene-curation/>

214. Genomics England. Genomcis England Panel App. (2018). at <https://panelapp.genomicsengland.co.uk/>

215. Walsh, R. *et al.* Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet. Med.* (2016). doi:10.1038/gim.2016.90

216. Pua, C. J. *et al.* Development of a Comprehensive Sequencing Assay for Inherited Cardiac Condition Genes. *J. Cardiovasc. Transl. Res.* **9,** 3–11 (2016).

217. Walsh, R. *et al.* Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes. *Eur. Heart J.* ehw603 (2017). doi:10.1093/eurheartj/ehw603

218. Wallis, Y. *et al.* Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics . (2013).

219. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

220. Conomos MP and Thornton T. GENESIS: GENetic EStimation and Inference in Structured samples (GENESIS). (2016).

221. Mandelker, D. *et al.* Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet. Med.* 1–8 (2016). doi:10.1038/gim.2016.58

222. Whiffin, N. *et al.* Using & high (resolution & variant & frequencies & to & empower & clinical & genome & interpretation &. (2016).

223. Kapa, S. *et al.* Genetic testing for long-qt syndrome: Distinguishing pathogenic mutations from benign variants. *Circulation* **120,** 1752–1760 (2009).

224. Kapplinger, J. D. *et al.* Enhancing the Predictive Power of Mutations in the C-Terminus of the KCNQ1-Encoded Kv7.1 Voltage-Gated Potassium Channel. *J. Cardiovasc. Transl. Res.* 187–197 (2015). doi:10.1007/s12265-015-9622-8

225. Li, B. *et al.* Predicting the Functional Impact of KCNQ1 Variants of Unknown SignificanceCLINICAL PERSPECTIVE. *Circ. Cardiovasc. Genet.* **10,** e001754 (2017).

226. Kubo, T. *et al.* Gender-specific differences in the clinical features of hypertrophic cardiomyopathy in a community-based Japanese population: Results from Kochi RYOMA study. *J. Cardiol.* **56,** 314–319 (2010).

227. Olivotto, I. *et al.* Gender-related differences in the clinical presentation and outcome of hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **46,** 480–487 (2005).

228. Cirulli, E. T. The Increasing Importance of Gene-Based Analyses. *PLOS Genet.* **12,** e1005852 (2016).

229. Cheng, A. Y., Teo, Y.-Y. & Ong, R. T.-H. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. *Bioinformatics* **30,** 1707–1713 (2014).

230. Meynert, A. M., Ansari, M., Fitzpatrick, D. R. & Taylor, M. S. Variant detection sensitivity and

biases in whole genome and exome sequencing. *BMC Bioinformatics* **15,** (2014).

231. Mandelker, D. Towards comprehensive clinical exome sequencing : Inclusive testing of disease genes with high homology NGS Testing Landscape 2014.

232. Jaafar, N. *et al.* Spectrum of Mutations in Hypertrophic Cardiomyopathy Genes Among Tunisian Patients. *Genet. Test. Mol. Biomarkers* **20,** 674–679 (2016).

233. Valdés-Mas, R. *et al.* Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. *Nat. Commun.* **5,** 5326 (2014).

234. Gómez, J. *et al.* Screening of the Filamin C Gene in a Large Cohort of Hypertrophic Cardiomyopathy Patients. *Circ. Cardiovasc. Genet.* **10,** (2017).

235. Odgerel, Z., van der Ven, P. F. M., Fürst, D. O. & Goldfarb, L. G. DNA sequencing errors in molecular diagnostics of filamin myopathy. *Clin. Chem. Lab. Med.* **48,** 1409–14 (2010).

236. England, J. & Loughna, S. Heavy and light roles: Myosin in the morphogenesis of the heart. *Cell. Mol. Life Sci.* **70,** 1221–1239 (2013).

237. Reiser, P. J., Portman, M. a, Ning, X. H. & Schomisch Moravec, C. Human cardiac myosin heavy chain isoforms in fetal and failing adult atria and ventricles. *Am. J. Physiol. Heart Circ. Physiol.* **280,** H1814–H1820 (2001).

238. Ching, Y.-H. *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat. Genet.* **37,** 423–428 (2005).

239. Granados-Riveron, J. T. *et al.* Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. *Hum. Mol. Genet.* **19,** 4007–4016 (2010).

240. Posch, M. G. *et al.* Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One* **6,** e28872 (2011).

241. Zhao, Y. *et al.* Targeted next-generation sequencing of candidate genes reveals novel mutations in patients with dilated cardiomyopathy. *Int. J. Mol. Med.* (2015). doi:10.3892/ijmm.2015.2361

242. Holm, H. *et al.* Syndrome. **43,** 316–320 (2011).

243. Ishikawa, T. *et al.* Novel mutation in the α-myosin heavy chain gene is associated with sick sinus syndrome. *Circ. Arrhythmia Electrophysiol.* **8,** 400–408 (2015).

244. Carniel, E. *et al.* Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation* **112,** 54–59 (2005).

245. Niimura, H. *et al.* Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation* **105,** 446–451 (2002).

246. Rubattu, S. *et al.* A Next-Generation Sequencing Approach to Identify Gene Mutations in Early- and Late-Onset Hypertrophic Cardiomyopathy Patients of an Italian Cohort. *Int. J. Mol. Sci.* **17,** (2016).

247. Drennan, D. & Ryazanov, A. G. Alpha-kinases: Analysis of the family and comparison with conventional protein kinases. *Prog. Biophys. Mol. Biol.* **85,** 1–32 (2004).

248. Ryazanov, A. G. *et al.* Alpha-kinases: a new class of protein kinases with a novel catalytic domain. *Curr. Biol.* **9,** R43-5 (1999).

249. Middelbeek, J., Clark, K., Venselaar, H., Huynen, M. A. & Van Leeuwen, F. N. The alpha-kinase family: An exceptional branch on the protein kinase tree. *Cell. Mol. Life Sci.* **67,** 875–890 (2010).

250. Almomani, R. *et al.* Biallelic Truncating Mutations in ALPK3 Cause Severe Pediatric Cardiomyopathy. *J. Am. Coll. Cardiol.* **67,** 515–525 (2016).

251. Phelan, D. G. *et al.* ALPK3-deficient cardiomyocytes generated from patient-derived induced pluripotent stem cells and mutant human embryonic stem cells display abnormal calcium handling and establish that ALPK3 deficiency underlies familial cardiomyopathy. *Eur. Heart J.* **37,** 2586–2590 (2016).

252. Caglayan, A. O. *et al.* ALPK3 GENE MUTATION IN A PATIENT WITH CONGENITAL CARDIOMYOPATHY AND DYSMORPHIC FEATURES. *Cold Spring Harb. Mol. case Stud.* (2017). doi:10.1101/mcs.a001859

253. Hosoda, T. *et al.* A novel myocyte-specific gene Midori promotes the differentiation of P19CL6 cells into cardiomyocytes. *J. Biol. Chem.* **276,** 35978–89 (2001).

254. Van Sligtenhorst, I. *et al.* Cardiomyopathy in α-Kinase 3 (ALPK3)–Deficient Mice. *Vet. Pathol.* **49,** 131–141 (2012).

255. Arber, S. *et al.* MLP-Deficient Mice Exhibit a Disruption of Cardiac Cytoarchitectural Organization, Dilated Cardiomyopathy, and Heart Failure. *Cell* **88,** 393–403 (2017).

256. Lan, F. *et al.* NIH Public Access. **12,** 101–113 (2014).

257. Zhang, J., Sun, X., Qian, Y., LaDuca, J. P. & Maquat, L. E. At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. *Mol. Cell. Biol.* **18,** 5272–83 (1998).

258. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem. Sci.* **23,** 198–199 (1998).

259. Zou, J. *et al.* An internal promoter underlies the difference in disease severity between N- and C-terminal truncation mutations of Titin in zebrafish. *Elife* **4,** e09406 (2015).

260. De Roeck, A. *et al.* Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. *Acta Neuropathol.* **134,** 475–487 (2017).

261. Nishiyama, A. *et al.* Dystrophin nonsense mutations can generate alternative rescue transcripts in lymphocytes. *Ann. Hum. Genet.* **72,** 717–724 (2008).

262. Cardiero, G. *et al.* Role of nonsense-mediated decay and nonsense-associated altered splicing in the mRNA pattern of two new alpha-thalassemia mutants. *Int. J. Biochem. Cell Biol.* **91,** 212–222 (2017).

263. De Bortoli, M. *et al.* The p.A897KfsX4 frameshift variation in desmocollin-2 is not a causative mutation in arrhythmogenic right ventricular cardiomyopathy. *Eur. J. Hum. Genet.* **18,** 776–82 (2010).

264. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).

265. Fuchs, E. & Weber, K. INTERMEDIA TE FILAMENTS : Structure, Dynamics, Function and Disease. *Annu. Rev. Biochem.* **63,** 345–82 (1994).

266. Konieczny, P. *et al.* Myofiber integrity depends on desmin network targeting to Z-disks and costameres via distinct plectin isoforms. *J. Cell Biol.* **181,** 667–681 (2008).

267. Kouloumenta, A., Mavroidis, M. & Capetanaki, Y. Proper perinuclear localization of the TRIM-like protein myospryn requires its binding partner desmin. *J. Biol. Chem.* **282,** 35211–35221 (2007).

268. Capetanaki, Y., Bloch, R. J., Kouloumenta, A., Mavroidis, M. & Psarras, S. Muscle intermediate filaments and their links to membranes and membranous organelles. *Exp. Cell Res.* **313,** 2063–2076 (2007).

269. Paulin, D. & Li, Z. Desmin: A major intermediate filament protein essential for the structural integrity and function of muscle. *Exp. Cell Res.* **301,** 1–7 (2004).

270. Dalakas, M. C. *et al.* Desmin myopathy, a skeletal myopathy with cardiomyopathy caused by mutations in the desmin gene. *N. Engl. J. Med.* **342,** 770–780 (2000).

271. Goldfarb, L. G., Olivé, M., Vicart, P. & Goebel, H. H. Intermediate filament diseases: desminopathy. *Adv. Exp. Med. Biol.* **642,** 131–64 (2008).

272. Goudeau, B. *et al.* Variable pathogenic potentials of mutations located in the desmin alpha-helical domain. *Hum. Mutat.* **27,** 906–913 (2006).

273. McLaughlin, H. M. *et al.* Compound heterozygosity of predicted loss-of-function DES variants in a family with recessive desminopathy. *BMC Med. Genet.* **14,** 68 (2013).

274. van Spaendonck-Zwarts, K. Y. *et al.* Desmin-related myopathy. *Clin. Genet.* **80,** 354–66 (2011).

275. Kaminska, A. *et al.* Small deletions disturb desmin architecture leading to breakdown of muscle cells and development of skeletal or cardioskeletal myopathy. *Hum. Genet.* **114,** 306–313 (2004).

276. Cao, L. *et al.* A novel heterozygous deletion-insertion mutation in the desmin gene causes complete atrioventricular block and mild myopathy. *Clin. Neuropathol.* **32,** 9–15 (2013).

277. Dunand, M. *et al.* Confirmation that abnormal desmin accumulation and migration are due to a desmin gene mutation in a familial cardiomyopathy and distal myopathy. *Neuromuscul. Disord.* **19,** 802 (2009).

278. Park, K. Y. *et al.* Desmin splice variants causing cardiac and skeletal myopathy. *J. Med. Genet.* **37,** 851–7 (2000).

279. Goldfarb, L. G. & Dalakas, M. C. Tragedy in a heartbeat: Malfunctioning desmin causes skeletal and cardiac muscle disease. *J. Clin. Invest.* **119,** 1806–1813 (2009).

280. Raats, J. M. *et al.* Muscle-specific expression of a dominant negative desmin mutant in transgenic mice. *Eur. J. Cell Biol.* **71,** 221–236 (1996).

281. Olive, M. *et al.* Phenotypic patterns of desminopathy associated with three novel mutations in the desmin gene. *Neuromuscul. Disord.* **17,** 443–450 (2007).

282. Bar, H. *et al.* Conspicuous involvement of desmin tail mutations in diverse cardiac and skeletal myopathies. *Hum. Mutat.* **28,** 374–386 (2007).

283. Brodehl, A. *et al.* Functional characterization of the novel DES mutation p.L136P associated with dilated cardiomyopathy reveals a dominant filament assembly defect. *J. Mol. Cell. Cardiol.* **91,** 207–214 (2016).

284. Brodehl, A. *et al.* The novel desmin mutant p.A120D impairs filament formation, prevents intercalated disk localization, and causes sudden cardiac death. *Circ. Cardiovasc. Genet.* **6,** 615–623 (2013).

285. NCBI-ClinVar. NCBI-ClinVar. *ClinVar* (2017). at <)>

286. Hong, D. *et al.* A series of Chinese patients with desminopathy associated with six novel and one reported mutations in the desmin gene. *Neuropathol. Appl. Neurobiol.* **37,** 257–70 (2011).

287. Schröder, R. *et al.* On noxious desmin: functional effects of a novel heterozygous desmin insertion mutation on the extrasarcomeric desmin cytoskeleton and mitochondria. *Hum. Mol. Genet.* **12,** 657–69 (2003).

288. Li, D. *et al.* Desmin mutation responsible for idiopathic dilated cardiomyopathy. *Circulation* **100,** 461–464 (1999).

289. Tesson, F. *et al.* Epidemiology of desmin and cardiac actin gene mutations in a european population of dilated cardiomyopathy. *Eur. Heart J.* **21,** 1872–6 (2000).

290. Miyamoto, Y. *et al.* Frequency and clinical characteristics of dilated cardiomyopathy caused by desmin gene mutation in a Japanese population. *Eur. Heart J.* **22,** 2284–2289 (2001).

291. Tse, H. F. *et al.* Patient-specific induced-pluripotent stem cells-derived cardiomyocytes recapitulate the pathogenic phenotypes of dilated cardiomyopathy due to a novel DES mutation identified by whole exome sequencing. *Hum. Mol. Genet.* **22,** 1395–1403 (2013).

292. Otten, E. *et al.* Desmin mutations as a cause of right ventricular heart failure affect the intercalated disks. *Heart Rhythm* **7,** 1058–1064 (2010).

293. van Tintelen, J. P. *et al.* Severe cardiac phenotype with right ventricular predominance in a large cohort of patients with a single missense mutation in the DES gene. *Hear. Rhythm* **6,** 1574–1583 (2009).

294. Lee MP, Hu RJ, Johnson LA, F. A. Human KvLQT1 gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. *Nat Genet* **15,** 181–185 (1997).

295. Jespersen, T., Grunnet, M. & Olesen, S.-P. The KCNQ1 potassium channel: From gene to physiological function. *Physiology* **20,** 408–416 (2005).

296. Ackerman, M. J. OUR TENTH YEAR Cardiac channelopathies : it ' s in the genes. *Nat. Med.* **10,** 463–464 (2004).

297. Kapplinger, J. D. *et al.* NIH Public Access. *Heart* **6,** 1297–1303 (2011).

298. Bezzina, C. R., Lahrouchi, N. & Priori, S. G. Genetics of Sudden Cardiac Death. *Circ. Res.* **116,** 1919–1936 (2015).

299. Moss, A. J. *et al.* Clinical aspects of type-1 long-QT syndrome by location, coding type, and biophysical function of mutations involving the KCNQ1 gene. *Circulation* **115,** 2481–2489 (2007).

300. Mazzanti, A., Underwood, K., Nevelev, D., Kofman, S. & Priori, S. G. The new kids on the block of arrhythmogenic disorders: Short QT syndrome and early repolarization. *J. Cardiovasc. Electrophysiol.* (2017). doi:10.1111/jce.13265

301. Bellocq, C. *et al.* Mutation in the KCNQ1 gene leading to the short QT-interval syndrome. *Circulation* **109,** 2394–2397 (2004).

302. Chen, Y. *et al.* Mutation in Familial Atrial. *Science (80-.).* **299,** 251–254 (2003).

303. Steffensen, A. B. *et al.* IKs Gain- and Loss-of-Function in Early-Onset Lone Atrial Fibrillation. *J. Cardiovasc. Electrophysiol.* **26,** 715–723 (2015).

304. Driest, S. L. Van *et al.* Association of Arrhythmia-Related Genetic Variants With Phenotypes Documented in Electronic Medical Records. **0575,** 47–57 (2016).

305. Refsgaard, L. *et al.* High prevalence of genetic variants previously associated with LQT syndrome in new exome data. *Eur. J. Hum. Genet.* **20,** 905–908 (2012).

306. Giudicessi, J. R. *et al.* Phylogenetic and physicochemical analyses enhance the classification of rare nonsynonymous single nucleotide variants in type 1 and 2 long-QT syndrome. *Circ. Cardiovasc. Genet.* **5,** 519–528 (2012).

307. Ware, J. S., Walsh, R., Cunningham, F., Birney, E. & Cook, S. A. Paralogous annotation of disease-causing variants in long QT syndrome genes. *Hum. Mutat.* **33,** 1188–1191 (2012).

308. Kapa, S. *et al.* Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. *Circulation* **120,** 1752–1760 (2009).

309. Johnson, J. N. *et al.* Prevalence and clinical correlates of QT prolongation in patients with hypertrophic cardiomyopathy. *Eur. Heart J.* **32,** 1114–1120 (2011).

310. Valle, G. *et al.* Telethonin, a novel sarcomeric protein of heart and skeletal muscle. *FEBS Lett.* **415,** 163–168 (1997).

311. Zou, P. *et al.* Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. *Nature* **439,** 229–233 (2006).

312. Knöll, R. *et al.* 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 (2002).

313. Kojic, S. *et al.* The Ankrd2 protein, a link between the sarcomere and the nucleus in skeletal muscle. *J. Mol. Biol.* **339,** 313–325 (2004).

314. Faulkner, G. *et al.* FATZ, a filamin-, actinin-, and telethonin-binding protein of the Z-disc of skeletal muscle. *J. Biol. Chem.* **275,** 41234–41242 (2000).

315. Furukawa, T. *et al.* 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 (2001).

316. Ikenberg, E. *et al.* Rare diagnosis of telethoninopathy (LGMD2G) in a Turkish patient. *Neuromuscul. Disord.* **27,** 856–860 (2017).

317. de Fuenmayor-Fernández de la Hoz, C. P., Hernández-Laín, A., Olivé, M., Fernández-Marmiesse, A. & Domínguez-González, C. Novel mutation in TCAP manifesting with asymmetric calves and early-onset joint retractions. *Neuromuscul. Disord.* **26,** 749–753 (2016).

318. Barresi, R. *et al.* Conserved expression of truncated telethonin in a patient with limb-girdle muscular dystrophy 2G. *Neuromuscul. Disord.* **25,** 349–352 (2015).

319. Francis, A. *et al.* Novel TCAP mutation c.32C>A causing limb girdle muscular dystrophy 2G. *PLoS One* **9,** 1–11 (2014).

320. Negrão, L., Matos, A., Geraldo, A. & Rebelo, O. Limb-girdle muscular dystrophy in a Portuguese patient caused by a mutation in the telethonin gene. *Acta Myol.* **29,** 21–24 (2010).

321. Olive, M. *et al.* Transcription-terminating mutation in telethonin causing autosomal recessive muscular dystrophy type 2G in a European patient. *Neuromuscul. Disord.* **18,** 929–933 (2008).

322. Moreira, E. S. Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. *Nat. Genet.* **24,** 163–166 (2000).

323. Ferreiro, A. *et al.* Telethonin-deficiency initially presenting as a congenital muscular dystrophy. *Neuromuscul. Disord.* **21,** 433–438 (2011).

324. Markert, C. D. *et al.* Functional muscle analysis of the Tcap knockout mouse. *Hum. Mol. Genet.* **19,** 2268–2283 (2010).

325. Zhang, R., Yang, J., Zhu, J. & Xu, X. Depletion of zebrafish Tcap leads to muscular dystrophy via disrupting sarcomere-membrane interaction, not sarcomere assembly. *Hum. Mol. Genet.* **18,** 4130–4140 (2009).

326. Jarvik, G. P. & Browning, B. L. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am. J. Hum. Genet.* **98,** 1077–1081 (2016).

327. Guo, M. H. *et al.* Determinants of Power in Gene-Based Burden Testing for Monogenic Disorders. *Am. J. Hum. Genet.* **99,** 527–539 (2016).

328. Amendola, L. M. *et al.* Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res.* **25,** 305–315 (2015).

329. Green, R. C. *et al.* Clinical Sequencing Exploratory Research Consortium: Accelerating Evidence-Based Practice of Genomic Medicine. *Am. J. Hum. Genet.* **98,** 1051–1066 (2016).

330. Rehm, H. L. *et al.* Spe ci a l R e p or t ClinGen — The Clinical Genome Resource. 1–8 (2015).

331. Hunter, J. E. *et al.* A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet. Med.* **18,** 1258–1268 (2016).

332. Strande, N. T. *et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100,** 895–906 (2017).

333. Manolio, T. A. *et al.* Implementing genomic medicine in the clinic: the future is here. *Genet. Med.* **15,** 258–267 (2013).

334. Bowdin, S. *et al.* Recommendations for the Integration of Genomics into Clinical Practice. **18,** 1075–1084 (2017).

335. Van Driest, S. L. *et al.* Comprehensive Analysis of the Beta-Myosin Heavy Chain Gene in 389 Unrelated Patients With Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* **44,** 602–610 (2004).

336. Golbus, J. R. *et al.* Population-based variation in cardiomyopathy genes. *Circ. Cardiovasc. Genet.* **5,** 391–9 (2012).

337. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132,** 1077–1130 (2013).

338. Kruszka, P., Weiss, K. & Hadley, D. W. *KCNQ1* Gene Variants in Large Asymptomatic Populations: Considerations for Genomic Screening of Military Cohorts. *Mil. Med.* **182,** e1795–e1800 (2017).

339. Andreasen, C. *et al.* New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur. J. Hum. Genet.* **21,** 918–28 (2013).

340. Golbus, J. R. *et al.* Targeted Analysis of Whole Genome Sequence Data to Diagnose Genetic Cardiomyopathy. *Circ. Cardiovasc. Genet.* **7,** 751–759 (2014).

341. Van Driest, S. L. *et al.* Association of Arrhythmia-Related Genetic Variants With Phenotypes Documented in Electronic Medical Records. *JAMA* **315,** 47–57 (2016).

342. Saltzman, A. J. *et al.* The Cardiac Myosin Binding Protein C Arg502Trp Mutation: A Common Cause of Hypertrophic Cardiomyopathy. *Circ Res* **106,** 1549–1552 (2010).

343. Christiaans, I. *et al.* Manifest disease, risk factors for sudden cardiac death, and cardiac events in a large nationwide cohort of predictively tested hypertrophic cardiomyopathy mutation carriers: Determining the best cardiological screening strategy. *Eur. Heart J.* **32,** 1161–1170 (2011).

344. Ng, D. *et al.* Interpreting Secondary Cardiac Disease Variants in an Exome Cohort. *Circ. Cardiovasc. Genet.* **6,** 10.1161/CIRCGENETICS.113.000039 (2013).

345. Natarajan, P. *et al.* Aggregate penetrance of genomic variants for actionable disorders in European and African Americans. **151,** 1–14 (2016).

346. Haggerty, C. M. *et al.* Electronic health record phenotype in subjects with genetic variants associated with arrhythmogenic right ventricular cardiomyopathy: a study of 30,716 subjects with exome sequencing. *Genet. Med.* **00,** 1–8 (2017).

347. ACMG. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* **17,** 68–69 (2015).

348. Feero, W. G. Establishing the Clinical Validity of Arrhythmia-Related Genetic Variations Using the Electronic Medical Record A Valid Take on Precision Medicine ? *Jama* **315,** 33–35 (2016).

349. Burke, W. *et al.* Recommendations for returning genomic incidental findings? We need to talk! *Genet. Med.* **15,** 854–859 (2013).

350. Murray, M. F. Your DNA is not your diagnosis : getting diagnoses right following secondary genomic findings. **18,** 2015–2017 (2016).

351. Biobank, U. UK Biobank. *11223* (2016). at <http://www.ukbiobank.ac.uk/2016/08/professor-hugh-watkins-university-of-oxford/>

352. Robertson, N. (University of O. Scatter Shot. (2018). at

<http://mccarthy.well.ox.ac.uk/static/scattershot/>

353. Aken, B. L. *et al.* The Ensembl gene annotation system. *Database* **2016,** baw093 (2016).

354. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 1–14 (2016).

355. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6,** 80–92 (2012).

356. Patel, R. Y. *et al.* ClinGen Pathogenicity Calculator: a configurable system for assessing pathogenicity of genetic variants. *Genome Med.* **9,** 3 (2017).

357. Walsh, R. *et al.* Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *bioRxiv* (2016). at <http://biorxiv.org/content/early/2016/02/24/041111.abstract>

358. Azaouagh, A., Churzidse, S., Konorza, T. & Erbel, R. Arrhythmogenic right ventricular cardiomyopathy/dysplasia: A review and update. *Clin. Res. Cardiol.* **100,** 383–394 (2011).

359. Philips, B. & Cheng, A. 2015 Update on the diagnosis and management of arrhythmogenic right ventricular cardiomyopathy. *Curr. Opin. Cardiol.* **31,** 1 (2015).

360. Crotti, L., Celano, G., Dagradi, F. & Schwartz, P. J. Congenital long QT syndrome. *Orphanet J. Rare Dis.* **3,** 1–16 (2008).

361. Schwartz, P. J. *et al.* Prevalence of the Congenital Long QT Syndrome. *Circulation* **120,** 1761–1767 (2009).

362. Goldenberg, I., Zareba, W. & Moss, A. J. Long QT Syndrome. *Curr. Probl. Cardiol.* **33,** 629–694 (2008).

363. Maron, B. J. New Perspectives on the Prevalence of Hypertrophic Cardiomyopathy. **65,** 0–5 (2015).

364. Lazzarini, E. *et al.* The ARVD/C genetic variants database: 2014 update. *Hum. Mutat.* **36,** 403–410 (2015).

365. Schwartz, P. J. *et al.* Prevalence of the congenital long-qt syndrome. *Circulation* **120,** 1761–1767 (2009).

366. Laurie, C., Doheny, K. & Mirel, D. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34,** 591–602 (2010).

367. UKBiobank. *UK Biobank Axiom Array Datasheet*. (2014).

368. Zawistowski, M. *et al.* Analysis of rare variant population structure in Europeans explains differential stratification of gene-based tests. *Eur. J. Hum. Genet.* **22,** 1137–44 (2014).

369. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

370. Morita, H. *et al.* Shared genetic causes of cardiac hypertrophy in children and adults. *N. Engl. J. Med.* **358,** 1899–1908 (2008).

371. Coto, E. *et al.* Resequencing the whole MYH7 gene (including the intronic, promoter, and 3'

UTR sequences) in hypertrophic cardiomyopathy. *J. Mol. Diagnostics* **14,** 518–524 (2012).

372. ClinVar. ClinVar ID 161327. (2018). at <https://www.ncbi.nlm.nih.gov/clinvar/variation/161327/>

373. Castelletti, S. *et al.* Desmoplakin missense and non-missense mutations in arrhythmogenic right ventricular cardiomyopathy: Genotype-phenotype correlation. *Int. J. Cardiol.* **249,** 268–273 (2017).

374. Manrai, A. K. *et al.* Genetic Misdiagnoses and the Potential for Health Disparities. *N. Engl. J. Med.* **375,** 655–665 (2016).

375. Driest, S. L. Van *et al.* Association of Arrhythmia-Related Genetic Variants With Phenotypes Documented in Electronic Medical Records. **0575,** 47–57 (2016).

376. Amendola, L. M. *et al.* Actionable exomic incidental findings in 6503 participants : challenges of variant classification. 305–315 (2015). doi:10.1101/gr.183483.114.Freely

377. Rehm, H. L. *et al.* Spe ci a l R e p or t ClinGen — The Clinical Genome Resource. 2235–2242 (2015).

378. de Andrade, K. C. *et al.* Higher-than-expected population prevalence of potentially pathogenic germline TP53 variants in individuals unselected for cancer history. *Hum. Mutat.* **38,** 1723–1730 (2017).

379. Kim, J., Field, A., Schultz, K. A. P., Hill, D. A. & Stewart, D. R. The prevalence of DICER1 pathogenic variation in population databases. *Int. J. Cancer* **141,** 2030–2036 (2017).

380. Newey, P. J., Berg, J. N., Zhou, K., Palmer, C. N. A. & Thakker, R. V. Utility of Population-Level DNA Sequence Data in the Diagnosis of Hereditary Endocrine Disease. *J. Endocr. Soc.* **1,** 1507–1526 (2017).

381. Tarailo-Graovac, M., Zhu, J. Y. A., Matthews, A., van Karnebeek, C. D. M. & Wasserman, W. W. Assessment of the ExAC data set for the presence of individuals with pathogenic genotypes implicated in severe Mendelian pediatric disorders. *Genet. Med.* **19,** 1300–1308 (2017).

382. Casali, C. *et al.* 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,** 1584–1589 (1999).

383. Ito, T. *et al.* Mitochondrial DNA mutations in cardiomyopathy. *Jpn. Circ. J.* **56,** 1045–1053 (1992).

384. Taylor, J. C. *et al.* Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat. Genet.* **47,** 717–726 (2015).

385. Mendes De Almeida, R. *et al.* Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy. *PLoS One* **12,** 1–19 (2017).

386. Xu, J. *et al.* Investigation of Pathogenic Genes in Chinese sporadic Hypertrophic Cardiomyopathy Patients by Whole Exome Sequencing. *Nat. Publ. Gr.* 1–13 (2015). doi:10.1038/srep16609

387. Cirino, A. L. *et al.* A Comparison of Whole Genome Sequencing to Multigene Panel Testing in Hypertrophic Cardiomyopathy Patients. *Circ. Cardiovasc. Genet.* **10,** e001768 (2017).

Appendix I

Gene	Protein	Summary of protein function
ABCC9	ATP Binding Cassette Subfamily C Member 9	The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member of the MRP subfamily which is involved in multi-drug resistance. This protein is thought to form ATP-sensitive potassium channels in cardiac, skeletal, and vascular and non-vascular smooth muscle. Protein structure suggests a role as the drug-binding channel-modulating subunit of the extra-pancreatic ATP-sensitive potassium channels. Mutations in this gene are associated with cardiomyopathy dilated type 1O. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Apr 2011]
ACADVL	Acyl-CoA Dehydrogenase, Very Long Chain	The protein encoded by this gene is targeted to the inner mitochondrial membrane where it catalyzes the first step of the mitochondrial fatty acid beta-oxidation pathway. This acyl-Coenzyme A dehydrogenase is specific to long-chain and very-long-chain fatty acids. A deficiency in this gene product reduces myocardial fatty acid beta-oxidation and is associated with cardiomyopathy. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Jul 2008]
ACTA1	Actin, Alpha 1, Skeletal Muscle	The product encoded by this gene belongs to the actin family of proteins, which are highly conserved proteins that play a role in cell motility, structure and integrity. Alpha, beta and gamma actin isoforms have been identified, with alpha actins being a major constituent of the contractile apparatus, while beta and gamma actins are involved in the regulation of cell motility. This actin is an alpha actin that is found in skeletal muscle. Mutations in this gene cause nemaline myopathy type 3, congenital myopathy with excess of thin myofilaments, congenital myopathy with cores, and congenital myopathy with fiber-type disproportion, diseases that lead to muscle fiber defects. [provided by RefSeq, Jul 2008]
AGL	Amylo-Alpha-1, 6-Glucosidase, 4-Alpha-Glucanotransferase	This gene encodes the glycogen debrancher enzyme which is involved in glycogen degradation. This enzyme has two independent catalytic activities which occur at different sites on the protein: a 4-alpha-glucotransferase activity and a amylo-1,6-glucosidase activity. Mutations in this gene are associated with glycogen storage disease although a wide range of enzymatic and clinical variability occurs which may be due to tissue-specific alternative splicing. Alternatively spliced transcripts encoding different isoforms have been described. [provided by RefSeq, Jul 2008]
ALPK3	Alpha-Protein Kinase 3	Kinase that recognizes phosphorylation sites in which the surrounding peptides have an alpha-helical conformation. Plays a role in cardiomyocyte differentiation (By similarity). [No details on NCBI gene, these details taken from UniProt 2016]

Gene	Protein	Summary of protein function
ATP5E	ATP Synthase, H+ Transporting, Mitochondrial F1 Complex, Epsilon Subunit	This gene encodes a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. ATP synthase is composed of two linked multi-subunit complexes: the soluble catalytic core, F1, and the membrane-spanning component, Fo, comprising the proton channel. The catalytic portion of mitochondrial ATP synthase consists of 5 different subunits (alpha, beta, gamma, delta, and epsilon) assembled with a stoichiometry of 3 alpha, 3 beta, and a single representative of the other 3. The proton channel consists of three main subunits (a, b, c). This gene encodes the epsilon subunit of the catalytic core. Two pseudogenes of this gene are located on chromosomes 4 and 13. Read-through transcripts that include exons from this gene are expressed from the upstream gene SLMO2.[provided by RefSeq, Mar 2011]
BAG3	BCL2 Associated Athanogene 3	BAG proteins compete with Hip for binding to the Hsc70/Hsp70 ATPase domain and promote substrate release. All the BAG proteins have an approximately 45-amino acid BAG domain near the C terminus but differ markedly in their N-terminal regions. The protein encoded by this gene contains a WW domain in the N-terminal region and a BAG domain in the C-terminal region. The BAG domains of BAG1, BAG2, and BAG3 interact specifically with the Hsc70 ATPase domain in vitro and in mammalian cells. All 3 proteins bind with high affinity to the ATPase domain of Hsc70 and inhibit its chaperone activity in a Hip-repressible manner. [provided by RefSeq, Jul 2008]
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase	This gene encodes a protein belonging to the raf/mil family of serine/threonine protein kinases. This protein plays a role in regulating the MAP kinase/ERKs signaling pathway, which affects cell division, differentiation, and secretion. Mutations in this gene are associated with cardiofaciocutaneous syndrome, a disease characterized by heart defects, mental retardation and a distinctive facial appearance. Mutations in this gene have also been associated with various cancers, including non-Hodgkin lymphoma, colorectal cancer, malignant melanoma, thyroid carcinoma, non-small cell lung carcinoma, and adenocarcinoma of lung. A pseudogene, which is located on chromosome X, has been identified for this gene. [provided by RefSeq, Jul 2008]
CACNA1C	Calcium Voltage-Gated Channel Subunit Alpha1 C	This gene encodes an alpha-1 subunit of a voltage-dependent calcium channel. Calcium channels mediate the influx of calcium ions into the cell upon membrane polarization. The alpha-1 subunit consists of 24 transmembrane segments and forms the pore through which ions pass into the cell. The calcium channel consists of a complex of alpha-1, alpha-2/delta, beta, and gamma subunits in a 1:1:1:1 ratio. There are multiple isoforms of each of these proteins, either encoded by different genes or the result of alternative splicing of transcripts. The protein encoded by this gene binds to and is inhibited by dihydropyridine. Alternative splicing results in many transcript variants encoding different proteins. Some of the predicted proteins may not produce functional ion channel subunits. [provided by RefSeq, Oct 2012]

Gene	Protein	Summary of protein function
CALR3	Calreticulin 3	The protein encoded by this gene belongs to the calreticulin family, members of which are calcium-binding chaperones localized mainly in the endoplasmic reticulum. This protein is also localized to the endoplasmic reticulum lumen, however, its capacity for calcium-binding may be absent or much lower than other family members. This gene is specifically expressed in the testis, and may be required for sperm fertility. Mutation in this gene has been associated with familial hypertrophic cardiomyopathy. [provided by RefSeq, Dec 2011]
CASQ2	Calsequestrin 2	The protein encoded by this gene specifies the cardiac muscle family member of the calsequestrin family. Calsequestrin is localized to the sarcoplasmic reticulum in cardiac and slow skeletal muscle cells. The protein is a calcium binding protein that stores calcium for muscle function. Mutations in this gene cause stress-induced polymorphic ventricular tachycardia, also referred to as catecholaminergic polymorphic ventricular tachycardia 2 (CPVT2), a disease characterized by bidirectional ventricular tachycardia that may lead to cardiac arrest. [provided by RefSeq, Jul 2008]
CAV3	Caveolin 3	This gene encodes a caveolin family member, which functions as a component of the caveolae plasma membranes found in most cell types. Caveolin proteins are proposed to be scaffolding proteins for organizing and concentrating certain caveolin-interacting molecules. Mutations identified in this gene lead to interference with protein oligomerization or intra-cellular routing, disrupting caveolae formation and resulting in Limb-Girdle muscular dystrophy type-1C (LGMD-1C), hyperCKemia or rippling muscle disease (RMD). Alternative splicing has been identified for this locus, with inclusion or exclusion of a differentially spliced intron. In addition, transcripts utilize multiple polyA sites and contain two potential translation initiation sites. [provided by RefSeq, Jul 2008]
COA5	Cytochrome C Oxidase Assembly Factor 5	This gene encodes an ortholog of yeast Pet191, which in yeast is a subunit of a large oligomeric complex associated with the mitochondrial inner membrane, and required for the assembly of the cytochrome c oxidase complex. Mutations in this gene are associated with mitochondrial complex IV deficiency, a disorder of the mitochondrial respiratory chain with heterogeneous clinical manifestations, ranging from isolated myopathy to a severe disease affecting several tissues and organs. [provided by RefSeq, Dec 2011]
COX15	COX15 Cytochrome C Oxidase Assembly Homolog	Cytochrome c oxidase (COX), the terminal component of the mitochondrial respiratory chain, catalyzes the electron transfer from reduced cytochrome c to oxygen. This component is a heteromeric complex consisting of 3 catalytic subunits encoded by mitochondrial genes and multiple structural subunits encoded by nuclear genes. The mitochondrially-encoded subunits function in electron transfer, and the nuclear-encoded subunits may function in the regulation and assembly of the complex. This nuclear gene encodes a protein which is not a structural subunit, but may be essential for the biogenesis of COX formation and may function in the hydroxylation of heme O, according to the yeast mutant studies. This protein is predicted to contain 5 transmembrane domains localized in the mitochondrial inner membrane. Alternative splicing of this gene generates two transcript variants diverging in the 3' region. [provided by RefSeq, Jul 2008]

Gene	Protein	Summary of protein function
CRYAB	Crystallin Alpha B	Mammalian lens crystallins are divided into alpha, beta, and gamma families. Alpha crystallins are composed of two gene products: alpha-A and alpha-B, for acidic and basic, respectively. Alpha crystallins can be induced by heat shock and are members of the small heat shock protein (HSP20) family. They act as molecular chaperones although they do not renature proteins and release them in the fashion of a true chaperone; instead they hold them in large soluble aggregates. Post-translational modifications decrease the ability to chaperone. These heterogeneous aggregates consist of 30-40 subunits; the alpha-A and alpha-B subunits have a 3:1 ratio, respectively. Two additional functions of alpha crystallins are an autokinase activity and participation in the intracellular architecture. The encoded protein has been identified as a moonlighting protein based on its ability to perform mechanistically distinct functions. Alpha-A and alpha-B gene products are differentially expressed; alpha-A is preferentially restricted to the lens and alpha-B is expressed widely in many tissues and organs. Elevated expression of alpha-B crystallin occurs in many neurological diseases; a missense mutation cosegregated in a family with a desmin-related myopathy. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jan 2014]
DES	Desmin	This gene encodes a muscle-specific class III intermediate filament. Homopolymers of this protein form a stable intracytoplasmic filamentous network connecting myofibrils to each other and to the plasma membrane. Mutations in this gene are associated with desmin-related myopathy, a familial cardiac and skeletal myopathy (CSM), and with distal myopathies. [provided by RefSeq, Jul 2008]
DSP	Desmoplakin	This gene encodes a protein that anchors intermediate filaments to desmosomal plaques and forms an obligate component of functional desmosomes. Mutations in this gene are the cause of several cardiomyopathies and keratodermas, including skin fragility-woolly hair syndrome. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jan 2016]
FLNC	Filamin C	This gene encodes one of three related filamin genes, specifically gamma filamin. These filamin proteins crosslink actin filaments into orthogonal networks in cortical cytoplasm and participate in the anchoring of membrane proteins for the actin cytoskeleton. Three functional domains exist in filamin: an N-terminal filamentous actin-binding domain, a C-terminal self-association domain, and a membrane glycoprotein-binding domain. Two transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]
FOXRED1	FAD-Dependent Oxidoreductase Domain Containing 1	This gene encodes a protein that contains a FAD-dependent oxidoreductase domain. The encoded protein is localized to the mitochondria and may function as a chaperone protein required for the function of mitochondrial complex I. Mutations in this gene are associated with mitochondrial complex I deficiency. Alternatively spliced transcript variants have been observed for this gene. [provided by RefSeq, Dec 2010]
FXN	Frataxin	This nuclear gene encodes a mitochondrial protein which belongs to FRATAXIN family. The protein functions in regulating mitochondrial iron transport and respiration. The expansion of intronic trinucleotide repeat GAA results in Friedreich ataxia. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jun 2009]

Gene	Protein	Summary of protein function
GAA	Glucosidase Alpha, Acid	This gene encodes lysosomal alpha-glucosidase, which is essential for the degradation of glycogen to glucose in lysosomes. The encoded preproprotein is proteolytically processed to generate multiple intermediate forms and the mature form of the enzyme. Defects in this gene are the cause of glycogen storage disease II, also known as Pompe's disease, which is an autosomal recessive disorder with a broad clinical spectrum. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jan 2016]
GLB1	Galactosidase Beta 1	This gene encodes a member of the glycosyl hydrolase 35 family of proteins. Alternative splicing results in multiple transcript variants, at least one of which encodes a preproprotein that is proteolytically processed to generate the mature lysosomal enzyme. This enzyme catalyzes the hydrolysis of a terminal beta-linked galactose residue from ganglioside substrates and other glycoconjugates. Mutations in this gene may result in GM1-gangliosidosis and Morquio B syndrome. [provided by RefSeq, Nov 2015]
GUSB	Glucuronidase Beta	This gene encodes a hydrolase that degrades glycosaminoglycans, including heparan sulfate, dermatan sulfate, and chondroitin-4,6-sulfate. The enzyme forms a homotetramer that is localized to the lysosome. Mutations in this gene result in mucopolysaccharidosis type VII. Alternative splicing results in multiple transcript variants. There are many pseudogenes of this locus in the human genome.[provided by RefSeq, May 2014]
HRAS	Harvey Rat Sarcoma Viral Oncogene Homolog	This gene belongs to the Ras oncogene family, whose members are related to the transforming genes of mammalian sarcoma retroviruses. The products encoded by these genes function in signal transduction pathways. These proteins can bind GTP and GDP, and they have intrinsic GTPase activity. This protein undergoes a continuous cycle of de- and re-palmitoylation, which regulates its rapid exchange between the plasma membrane and the Golgi apparatus. Mutations in this gene cause Costello syndrome, a disease characterized by increased growth at the prenatal stage, growth deficiency at the postnatal stage, predisposition to tumor formation, mental retardation, skin and musculoskeletal abnormalities, distinctive facial appearance and cardiovascular abnormalities. Defects in this gene are implicated in a variety of cancers, including bladder cancer, follicular thyroid cancer, and oral squamous cell carcinoma. Multiple transcript variants, which encode different isoforms, have been identified for this gene. [provided by RefSeq, Jul 2008]
JPH2	Junctophilin 2	Junctional complexes between the plasma membrane and endoplasmic/sarcoplasmic reticulum are a common feature of all excitable cell types and mediate cross talk between cell surface and intracellular ion channels. The protein encoded by this gene is a component of junctional complexes and is composed of a C-terminal hydrophobic segment spanning the endoplasmic/sarcoplasmic reticulum membrane and a remaining cytoplasmic domain that shows specific affinity for the plasma membrane. This gene is a member of the junctophilin gene family. Alternative splicing has been observed at this locus and two variants encoding distinct isoforms are described. [provided by RefSeq, Jul 2008]

Gene	Protein	Summary of protein function
KCNQ1	Potassium Voltage-Gated Channel Subfamily Q Member 1	This gene encodes a voltage-gated potassium channel required for repolarization phase of the cardiac action potential. This protein can form heteromultimers with two other potassium channel proteins, KCNE1 and KCNE3. Mutations in this gene are associated with hereditary long QT syndrome 1 (also known as Romano-Ward syndrome), Jervell and Lange-Nielsen syndrome, and familial atrial fibrillation. This gene exhibits tissue-specific imprinting, with preferential expression from the maternal allele in some tissues, and biallelic expression in others. This gene is located in a region of chromosome 11 amongst other imprinted genes that are associated with Beckwith-Wiedemann syndrome (BWS), and itself has been shown to be disrupted by chromosomal rearrangements in patients with BWS. Alternatively spliced transcript variants have been found for this gene. [provided by RefSeq, Aug 2011]
KLF10	Kruppel-Like Factor 10	This gene encodes a member of a family of proteins that feature C2H2-type zinc finger domains. The encoded protein is a transcriptional repressor that acts as an effector of transforming growth factor beta signaling. Activity of this protein may inhibit the growth of cancers, particularly pancreatic cancer. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jun 2013]
LDB3	LIM Domain Binding 3	This gene encodes a PDZ domain-containing protein. PDZ motifs are modular protein-protein interaction domains consisting of 80-120 amino acid residues. PDZ domain-containing proteins interact with each other in cytoskeletal assembly or with other proteins involved in targeting and clustering of membrane proteins. The protein encoded by this gene interacts with alpha-actinin-2 through its N-terminal PDZ domain and with protein kinase C via its C-terminal LIM domains. The LIM domain is a cysteine-rich motif defined by 50-60 amino acids containing two zinc-binding modules. This protein also interacts with all three members of the myozenin family. Mutations in this gene have been associated with myofibrillar myopathy and dilated cardiomyopathy. Alternatively spliced transcript variants encoding different isoforms have been identified; all isoforms have N-terminal PDZ domains while only longer isoforms (1, 2 and 5) have C-terminal LIM domains. [provided by RefSeq, Jan 2010]
LMNA	Lamin A/C	The nuclear lamina consists of a two-dimensional matrix of proteins located next to the inner nuclear membrane. The lamin family of proteins make up the matrix and are highly conserved in evolution. During mitosis, the lamina matrix is reversibly disassembled as the lamin proteins are phosphorylated. Lamin proteins are thought to be involved in nuclear stability, chromatin structure and gene expression. Vertebrate lamins consist of two types, A and B. Alternative splicing results in multiple transcript variants. Mutations in this gene lead to several diseases: Emery-Dreifuss muscular dystrophy, familial partial lipodystrophy, limb girdle muscular dystrophy, dilated cardiomyopathy, Charcot-Marie-Tooth disease, and Hutchinson-Gilford progeria syndrome. [provided by RefSeq, Apr 2012]

Gene	Protein	Summary of protein function
MAP2K1	Mitogen-Activated Protein Kinase Kinase 1	The protein encoded by this gene is a member of the dual specificity protein kinase family, which acts as a mitogen-activated protein (MAP) kinase kinase. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act as an integration point for multiple biochemical signals. This protein kinase lies upstream of MAP kinases and stimulates the enzymatic activity of MAP kinases upon wide variety of extra- and intracellular signals. As an essential component of MAP kinase signal transduction pathway, this kinase is involved in many cellular processes such as proliferation, differentiation, transcription regulation and development. [provided by RefSeq, Jul 2008]
MAP2K2	Mitogen-Activated Protein Kinase Kinase 2	The protein encoded by this gene is a dual specificity protein kinase that belongs to the MAP kinase kinase family. This kinase is known to play a critical role in mitogen growth factor signal transduction. It phosphorylates and thus activates MAPK1/ERK2 and MAPK2/ERK3. The activation of this kinase itself is dependent on the Ser/Thr phosphorylation by MAP kinase kinase kinases. Mutations in this gene cause cardiofaciocutaneous syndrome (CFC syndrome), a disease characterized by heart defects, mental retardation, and distinctive facial features similar to those found in Noonan syndrome. The inhibition or degradation of this kinase is also found to be involved in the pathogenesis of Yersinia and anthrax. A pseudogene, which is located on chromosome 7, has been identified for this gene. [provided by RefSeq, Jul 2008]
MRPL3	Mitochondrial Ribosomal Protein L3	Mammalian mitochondrial ribosomal proteins are encoded by nuclear genes and help in protein synthesis within the mitochondrion. Mitochondrial ribosomes (mitoribosomes) consist of a small 28S subunit and a large 39S subunit. They have an estimated 75% protein to rRNA composition compared to prokaryotic ribosomes, where this ratio is reversed. Another difference between mammalian mitoribosomes and prokaryotic ribosomes is that the latter contain a 5S rRNA. Among different species, the proteins comprising the mitoribosome differ greatly in sequence, and sometimes in biochemical properties, which prevents easy recognition by sequence homology. This gene encodes a 39S subunit protein that belongs to the L3P ribosomal protein family. A pseudogene corresponding to this gene is found on chromosome 13q. [provided by RefSeq, Jul 2008]
MYH6	Myosin, Heavy Chain 6, Cardiac Muscle, Alpha	Cardiac muscle myosin is a hexamer consisting of two heavy chain subunits, two light chain subunits, and two regulatory subunits. This gene encodes the alpha heavy chain subunit of cardiac myosin. The gene is located ~4kb downstream of the gene encoding the beta heavy chain subunit of cardiac myosin. Mutations in this gene cause familial hypertrophic cardiomyopathy and atrial septal defect 3. [provided by RefSeq, Mar 2010]
MYLK2	Myosin Light Chain Kinase 2	This gene encodes a myosin light chain kinase, a calcium/calmodulin dependent enzyme, that is exclusively expressed in adult skeletal muscle. [provided by RefSeq, Jul 2008]
MYO6	Myosin VI	This gene encodes a reverse-direction motor protein that moves toward the minus end of actin filaments and plays a role in intracellular vesicle and organelle transport. The protein consists of a motor domain containing an ATP- and an actin-binding site and a globular tail which interacts with other proteins. This protein maintains the structural integrity of inner ear hair cells and mutations in this gene cause non-syndromic autosomal dominant and recessive hearing loss. Alternative splicing results in multiple transcript variants encoding distinct isoforms. [provided by RefSeq, Jul 2014]

Gene	Protein	Summary of protein function
MYOM1	Myomesin 1	The giant protein titin, together with its associated proteins, interconnects the major structure of sarcomeres, the M bands and Z discs. The C-terminal end of the titin string extends into the M line, where it binds tightly to M-band constituents of apparent molecular masses of 190 kD (myomesin 1) and 165 kD (myomesin 2). This protein, myomesin 1, like myomesin 2, titin, and other myofibrillar proteins contains structural modules with strong homology to either fibronectin type III (motif I) or immunoglobulin C2 (motif II) domains. Myomesin 1 and myomesin 2 each have a unique N-terminal region followed by 12 modules of motif I or motif II, in the arrangement II-II-I-I-I-I-I-II-II-II-II-II. The two proteins share 50% sequence identity in this repeat-containing region. The head structure formed by these 2 proteins on one end of the titin string extends into the center of the M band. The integrating structure of the sarcomere arises from muscle-specific members of the superfamily of immunoglobulin-like proteins. Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jul 2008]
MYOZ2	Myozenin 2	The protein encoded by this gene belongs to a family of sarcomeric proteins that bind to calcineurin, a phosphatase involved in calcium-dependent signal transduction in diverse cell types. These family members tether calcineurin to alpha-actinin at the z-line of the sarcomere of cardiac and skeletal muscle cells, and thus they are important for calcineurin signaling. Mutations in this gene cause cardiomyopathy familial hypertrophic type 16, a hereditary heart disorder. [provided by RefSeq, Aug 2011]
MYPN	Myopalladin	Striated muscle in vertebrates comprises large proteins which must be organized properly to contract efficiently. Z-lines in striated muscle are a sign of this organization, representing the ends of actin thin filaments, titin, nebulin or nebulette and accessory proteins required for structure and function. This gene encodes a protein which interacts with nebulin in skeletal muscle or nebulette in cardiac muscle and alpha-actinin. In addition, this gene product can interact with a protein with the I-band indicating it has a regulatory as well as structural function. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2011]
NEXN	Nexilin F-Actin Binding Protein	This gene encodes a filamentous actin-binding protein that may function in cell adhesion and migration. Mutations in this gene have been associated with dilated cardiomyopathy, also known as CMD1CC. Alternatively spliced transcript variants have been described.[provided by RefSeq, Feb 2010]
NRAS	Neuroblastoma RAS Viral Oncogene Homolog	This is an N-ras oncogene encoding a membrane protein that shuttles between the Golgi apparatus and the plasma membrane. This shuttling is regulated through palmitoylation and depalmitoylation by the ZDHHC9-GOLGA7 complex. The encoded protein, which has intrinsic GTPase activity, is activated by a guanine nucleotide-exchange factor and inactivated by a GTPase activating protein. Mutations in this gene have been associated with somatic rectal cancer, follicular thyroid cancer, autoimmune lymphoproliferative syndrome, Noonan syndrome, and juvenile myelomonocytic leukemia. [provided by RefSeq, Jun 2011]

Gene	Protein	Summary of protein function
OBSCN	Obscurin-Myosin Light Chain Kinase	The obscurin gene spans more than 150 kb, contains over 80 exons and encodes a protein of approximately 720 kDa. The encoded protein contains 68 Ig domains, 2 fibronectin domains, 1 calcium/calmodulin-binding domain, 1 RhoGEF domain with an associated PH domain, and 2 serine-threonine kinase domains. This protein belongs to the family of giant sacromeric signaling proteins that includes titin and nebulin, and may have a role in the organization of myofibrils during assembly and may mediate interactions between the sarcoplasmic reticulum and myofibrils. Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jul 2008]
PDLIM3	PDZ And LIM Domain 3	The protein encoded by this gene contains a PDZ domain and a LIM domain, indicating that it may be involved in cytoskeletal assembly. In support of this, the encoded protein has been shown to bind the spectrin-like repeats of alpha-actinin-2 and to colocalize with alpha-actinin-2 at the Z lines of skeletal muscle. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. Aberrant alternative splicing of this gene may play a role in myotonic dystrophy. [provided by RefSeq, Apr 2012]
PTPN11	Protein Tyrosine Phosphatase, Non-Receptor Type 11	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP contains two tandem Src homology-2 domains, which function as phospho-tyrosine binding domains and mediate the interaction of this PTP with its substrates. This PTP is widely expressed in most tissues and plays a regulatory role in various cell signaling events that are important for a diversity of cell functions, such as mitogenic activation, metabolic control, transcription regulation, and cell migration. Mutations in this gene are a cause of Noonan syndrome as well as acute myeloid leukemia. Two transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, May 2012]
RAF1	Raf-1 Proto-Oncogene, Serine/Threonine Kinase	This gene is the cellular homolog of viral raf gene (v-raf). The encoded protein is a MAP kinase kinase kinase (MAP3K), which functions downstream of the Ras family of membrane associated GTPases to which it binds directly. Once activated, the cellular RAF1 protein can phosphorylate to activate the dual specificity protein kinases MEK1 and MEK2, which in turn phosphorylate to activate the serine/threonine specific protein kinases, ERK1 and ERK2. Activated ERKs are pleiotropic effectors of cell physiology and play an important role in the control of gene expression involved in the cell division cycle, apoptosis, cell differentiation and cell migration. Mutations in this gene are associated with Noonan syndrome 5 and LEOPARD syndrome 2. [provided by RefSeq, Jul 2008]
RYR2	Ryanodine Receptor 2	This gene encodes a ryanodine receptor found in cardiac muscle sarcoplasmic reticulum. The encoded protein is one of the components of a calcium channel, composed of a tetramer of the ryanodine receptor proteins and a tetramer of FK506 binding protein 1B proteins, that supplies calcium to cardiac muscle. Mutations in this gene are associated with stress-induced polymorphic ventricular tachycardia and arrhythmogenic right ventricular dysplasia. [provided by RefSeq, Jul 2008]

Gene	Protein	Summary of protein function
SCO2	SCO2 Cytochrome C Oxidase Assembly Protein	Cytochrome c oxidase (COX) catalyzes the transfer of electrons from cytochrome c to molecular oxygen, which helps to maintain the proton gradient across the inner mitochondrial membrane that is necessary for aerobic ATP production. Human COX is a multimeric protein complex that requires several assembly factors; this gene encodes one of the COX assembly factors. The encoded protein is a metallochaperone that is involved in the biogenesis of cytochrome c oxidase subunit II. Mutations in this gene are associated with fatal infantile encephalocardiomyopathy and myopia 6. [provided by RefSeq, Oct 2014]
SHOC2	SHOC2 Leucine-Rich Repeat Scaffold Protein	This gene encodes a protein that consists almost entirely of leucine-rich repeats, a domain implicated in protein-protein interactions. The protein may function as a scaffold linking RAS to downstream signal transducers in the RAS/ERK MAP kinase signaling cascade. Mutations in this gene have been associated with Noonan-like syndrome with loose anagen hair. [provided by RefSeq, May 2010]
SLC25A3	Solute Carrier Family 25 Member 3	The protein encoded by this gene catalyzes the transport of phosphate into the mitochondrial matrix, either by proton cotransport or in exchange for hydroxyl ions. The protein contains three related segments arranged in tandem which are related to those found in other characterized members of the mitochondrial carrier family. Both the N-terminal and C-terminal regions of this protein protrude toward the cytosol. Multiple alternatively spliced transcript variants have been isolated. [provided by RefSeq, Jul 2008]
SLC25A4	Solute Carrier Family 25 Member 4	This gene is a member of the mitochondrial carrier subfamily of solute carrier protein genes. The product of this gene functions as a gated pore that translocates ADP from the cytoplasm into the mitochondrial matrix and ATP from the mitochondrial matrix into the cytoplasm. The protein forms a homodimer embedded in the inner mitochondria membrane. Mutations in this gene have been shown to result in autosomal dominant progressive external opthalmoplegia and familial hypertrophic cardiomyopathy. [provided by RefSeq, Jun 2013]
SOS1	SOS Ras/Rac Guanine Nucleotide Exchange Factor 1	This gene encodes a protein that is a guanine nucleotide exchange factor for RAS proteins, membrane proteins that bind guanine nucleotides and participate in signal transduction pathways. GTP binding activates and GTP hydrolysis inactivates RAS proteins. The product of this gene may regulate RAS proteins by facilitating the exchange of GTP for GDP. Mutations in this gene are associated with gingival fibromatosis 1 and Noonan syndrome type 4. [provided by RefSeq, Jul 2008]
TCAP	Titin-Cap	Sarcomere assembly is regulated by the muscle protein titin. Titin is a giant elastic protein with kinase activity that extends half the length of a sarcomere. It serves as a scaffold to which myofibrils and other muscle related proteins are attached. This gene encodes a protein found in striated and cardiac muscle that binds to the titin Z1-Z2 domains and is a substrate of titin kinase, interactions thought to be critical to sarcomere assembly. Mutations in this gene are associated with limb-girdle muscular dystrophy type 2G. [provided by RefSeq, Jul 2008]
TMEM70	Transmembrane Protein 70	This gene likely encodes a mitochondrial membrane protein. The encoded protein may play a role in biogenesis of mitochondrial ATP synthase. Mutations in this gene have been associated with neonatal mitochondrial encephalocardiomyopathy due to ATP synthase deficiency. Alternatively spliced transcript variants have been described. [provided by RefSeq, Feb 2010]

Gene	Protein	Summary of protein function
TMPO	Thymopoietin	The protein encoded by this gene resides in the nucleus and may play a role in the assembly of the nuclear lamina, and thus help maintain the structural organization of the nuclear envelope. It may function as a receptor for the attachment of lamin filaments to the inner nuclear membrane. Mutations in this gene are associated with dilated cardiomyopathy. Alternatively spliced transcript variants encoding different isoforms have been noted for this gene. [provided by RefSeq, May 2010]
TNNC1	Troponin C1, Slow Skeletal And Cardiac Type	Troponin is a central regulatory protein of striated muscle contraction, and together with tropomyosin, is located on the actin filament. Troponin consists of 3 subunits: TnI, which is the inhibitor of actomyosin ATPase; TnT, which contains the binding site for tropomyosin; and TnC, the protein encoded by this gene. The binding of calcium to TnC abolishes the inhibitory action of TnI, thus allowing the interaction of actin with myosin, the hydrolysis of ATP, and the generation of tension. Mutations in this gene are associated with cardiomyopathy dilated type 1Z. [provided by RefSeq, Oct 2008]
TNNC2	Troponin C Type 2	Troponin (Tn), a key protein complex in the regulation of striated muscle contraction, is composed of 3 subunits. The Tn-I subunit inhibits actomyosin ATPase, the Tn-T subunit binds tropomyosin and Tn-C, while the Tn-C subunit binds calcium and overcomes the inhibitory action of the troponin complex on actin filaments. The protein encoded by this gene is the Tn-C subunit. [provided by RefSeq, Jul 2008]
TRIM63	Tripartite Motif Containing 63	This gene encodes a member of the RING zinc finger protein family found in striated muscle and iris. The product of this gene is an E3 ubiquitin ligase that localizes to the Z-line and M-line lattices of myofibrils. This protein plays an important role in the atrophy of skeletal and cardiac muscle and is required for the degradation of myosin heavy chain proteins, myosin light chain, myosin binding protein, and for muscle-type creatine kinase. [provided by RefSeq, Feb 2012]
TSFM	Ts Translation Elongation Factor, Mitochondrial	This gene encodes a mitochondrial translation elongation factor. The encoded protein is an enzyme that catalyzes the exchange of guanine nucleotides on the translation elongation factor Tu during the elongation step of mitchondrial protein translation. Mutations in this gene are associated with combined oxidative phosphorylation deficiency-3 syndrome. Alternate splicing results in multiple transcript variants.[provided by RefSeq, Mar 2010]

Gene	Protein	Summary of protein function
TTN	Titin	This gene encodes a large abundant protein of striated muscle. The product of this gene is divided into two regions, a N-terminal I-band and a C-terminal A-band. The I-band, which is the elastic part of the molecule, contains two regions of tandem immunoglobulin domains on either side of a PEVK region that is rich in proline, glutamate, valine and lysine. The A-band, which is thought to act as a protein-ruler, contains a mixture of immunoglobulin and fibronectin repeats, and possesses kinase activity. An N-terminal Z-disc region and a C-terminal M-line region bind to the Z-line and M-line of the sarcomere, respectively, so that a single titin molecule spans half the length of a sarcomere. Titin also contains binding sites for muscle associated proteins so it serves as an adhesion template for the assembly of contractile machinery in muscle cells. It has also been identified as a structural protein for chromosomes. Alternative splicing of this gene results in multiple transcript variants. Considerable variability exists in the I-band, the M-line and the Z-disc regions of titin. Variability in the I-band region contributes to the differences in elasticity of different titin isoforms and, therefore, to the differences in elasticity of different muscle types. Mutations in this gene are associated with familial hypertrophic cardiomyopathy 9, and autoantibodies to titin are produced in patients with the autoimmune disease scleroderma. [provided by RefSeq, Feb 2012]
TTR	Transthyretin	This gene encodes transthyretin, one of the three prealbumins including alpha-1-antitrypsin, transthyretin and orosomucoid. Transthyretin is a carrier protein; it transports thyroid hormones in the plasma and cerebrospinal fluid, and also transports retinol (vitamin A) in the plasma. The protein consists of a tetramer of identical subunits. More than 80 different mutations in this gene have been reported; most mutations are related to amyloid deposition, affecting predominantly peripheral nerve and/or the heart, and a small portion of the gene mutations is non-amyloidogenic. The diseases caused by mutations include amyloidotic polyneuropathy, euthyroid hyperthyroxinaemia, amyloidotic vitreous opacities, cardiomyopathy, oculoleptomeningeal amyloidosis, meningocerebrovascular amyloidosis, carpal tunnel syndrome, etc. [provided by RefSeq, Jan 2009]
VCL	Vinculin	Vinculin is a cytoskeletal protein associated with cell-cell and cell-matrix junctions, where it is thought to function as one of several interacting proteins involved in anchoring F-actin to the membrane. Defects in VCL are the cause of cardiomyopathy dilated type 1W. Dilated cardiomyopathy is a disorder characterized by ventricular dilation and impaired systolic function, resulting in congestive heart failure and arrhythmia. Multiple alternatively spliced transcript variants have been found for this gene, but the biological validity of some variants has not been determined. [provided by RefSeq, Jul 2008]