

Genetic, genotype and imaging studies of hypertrophic cardiomyopathy

Submitted to Imperial College London
for the degree of Doctor of Philosophy

by

Roddy Walsh

M.Sc.

Cardiovascular Genetics and Genomics
Royal Brompton Cardiovascular Research Centre
Imperial College – National Heart and Lung Institute

December 2017

Declaration of originality

I, Roddy Walsh, declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Any contribution of others to the work described is duly acknowledged, and all else accordingly referenced.

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or distribution, researchers must make clear to others the licence terms of this work.

Abstract

Hypertrophic cardiomyopathy (HCM) is an inherited disorder of cardiac muscle whose genetic basis has been investigated for over 25 years. However, considerable uncertainty remains about the genetic aetiology of HCM, including the validity of implicated genes, the effective analysis of variants detected in patients and the interpretation and follow-up of genetic testing results.

Here, genetic data from large cardiomyopathy cohorts and 60,706 individuals of the Exome Aggregation Consortium (ExAC) was used to reassess the role of genes implicated in HCM and other cardiomyopathies. A significant excess of rare variation in cases over ExAC, along with supporting genetic evidence, was observed for only three non-sarcomeric genes, causative in <2% of cases, with moderate evidence of association for five other non-sarcomeric genes. This data suggests HCM genetic testing should be limited to 16 valid genes, including the eight core sarcomeric genes.

Variant analysis in sarcomeric genes has struggled to balance accuracy and sensitivity, with clinical genetics guidelines overly conservative in assigning pathogenicity. Here, by utilising stringent frequency thresholds and case-control analyses, classes of variants with high prior probabilities of pathogenicity ($\geq 95\%$) were identified in five sarcomeric genes. By adapting current variant interpretation guidelines, the yield of pathogenic variants in HCM cases was increased by 14%, corresponding to 4% of all tested HCM patients.

Building on this enhanced understanding of HCM genetics, genotype-phenotype analyses demonstrated that cases without causative variants follow a distinct clinical course, with later presentation, less pronounced fibrosis and markedly different hypertrophy patterns - lesser maximum wall thickness (19.85mm vs 18.51mm, $p=0.006$) but greater overall left ventricular mass (101.2g/m^2 vs 88.6g/m^2 , $p<0.001$). These results posit an informative interpretation of negative tests, indicating a benign prognosis and low likelihood of familial transmission.

Taken together, these data redefine the genetic architecture of cardiomyopathy and significantly enhance the interpretability of HCM genetic testing.

Table of contents

Declaration of originality	2
Copyright declaration	2
Abstract	3
Table of contents	4
List of Figures	8
List of Tables	10
Contribution of others to work presented in this thesis	10
Acknowledgements	13
Glossary of terms and abbreviations	14
Publications arising from this work	16
1 INTRODUCTION	17
1.1 <i>Cardiomyopathies</i>	17
1.2 <i>Overview of Hypertrophic Cardiomyopathy</i>	17
1.2.1 Definition of Hypertrophic Cardiomyopathy	17
1.2.2 Prevalence of HCM, issues in diagnosis and clinical features	19
1.2.3 Clinical course and management	20
1.2.4 Future treatment options	22
1.3 <i>Genetics of Hypertrophic Cardiomyopathy</i>	23
1.3.1 Current knowledge of the genetic basis of HCM	23
1.3.2 Role of genetic testing in HCM	25
1.3.3 Beyond a Mendelian model of disease	27
1.4 <i>Approaches to genetic sequencing and variant analysis</i>	28
1.4.1 Role of genetic sequencing in Mendelian diseases like HCM	28
1.4.2 Sequencing technologies and next generation sequencing approaches	30
1.4.3 Analysis of NGS data – variant detection	33
1.4.4 Analysis of NGS data – variant interpretation	33
1.5 <i>Aims of thesis</i>	36

2	DEFINING CARDIOMYOPATHY GENES	37
2.1	<i>Introduction</i>	37
2.1.1	Genes implicated in cardiomyopathies	38
2.1.2	Methods to associate genes with Mendelian disease	39
2.1.3	Impact on clinical genetic testing in cardiomyopathies	42
2.1.4	New insights from population databases	45
2.1.5	Exome Aggregation Consortium (ExAC)	45
2.1.6	Importance of validating the role of genes in cardiomyopathies	46
2.2	<i>Materials and Methods</i>	48
2.2.1	Sequencing cohorts	48
2.2.2	Next generation sequencing of prospective HCM cohort	56
2.2.3	Defining rare variants	57
2.2.4	Burden testing	58
2.2.5	Assessment of population-specific effects	60
2.2.6	Extended HCM gene analysis – data mining	60
2.2.7	Extended HCM gene analysis – classification	62
2.3	<i>Results</i>	64
2.3.1	Comparing rare variation between clinical cardiomyopathy cohorts and ExAC	64
2.3.2	Population-specific burden analysis	69
2.3.3	Assessing results of extended gene panel cardiomyopathy studies	72
2.3.4	Meta-analysis of genes implicated in HCM	75
2.4	<i>Discussion</i>	86
2.4.1	Why did the results from candidate gene studies fail to validate?	86
2.4.2	How reliable are the gene burden tests performed here?	87
2.4.3	What does lack of excess mean and how should this data inform clinical genetic testing?	89
2.4.4	The missing heritability of HCM	91
2.4.5	Comparison with ClinGen clinical validity framework	93
2.4.6	DCM gene burden data	95
2.4.7	Application to other Mendelian diseases	96
2.5	<i>Conclusion</i>	97
3	VARIANT INTERPRETATION IN HCM	98
3.1	<i>Introduction</i>	98
3.1.1	Early approaches to assessing variant pathogenicity	98
3.1.2	Insights from population genetics studies	100
3.1.3	ACMG Guidelines	101
3.1.4	Issues with current guidelines	103
3.1.5	Current and future challenges in variant interpretation	107
3.2	<i>Methods</i>	108
3.2.1	A more refined population allele frequency threshold for HCM	108
3.2.2	Identifying false positive disease associations in research literature	109
3.2.3	Defining Odds Ratios and Etiological Fractions	109
3.2.4	Refining EFs within genes	112
3.2.5	Adapting ACMG guidelines	115
3.2.6	Variant classification in clinical and prospective HCM cohorts	118

3.3	<i>Results</i>	122
3.3.1	Variant analysis in HCM confounded by both false positive and false negative associations	122
3.3.2	Odds Ratios and Etiological Fractions for HCM genes	124
3.3.3	Refining EFs within genes	127
3.3.4	Variant classification using adapted ACMG guidelines	135
3.3.5	Identification of non-rare variants enriched in case cohorts	138
3.4	<i>Discussion</i>	143
3.4.1	The limitations of generic approaches to variant analysis	143
3.4.2	Advantages of quantitative, EF-based approaches to variant interpretation	144
3.4.3	Limitations of EF-based analyses	146
3.4.4	Future challenges in variant interpretation	147
3.5	<i>Conclusion</i>	151
4	GENOTYPE-PHENOTYPE STUDIES IN HCM	152
4.1	<i>Introduction</i>	152
4.1.1	Assessing the correlation between genotype and phenotype	152
4.1.2	Genotype-phenotype studies in HCM	152
4.1.3	Genotype-positive versus genotype-negative	156
4.1.4	Effect of specific variant class	156
4.1.5	Effect of multiple pathogenic variants	159
4.1.6	Ongoing challenges in HCM genotype-phenotype analysis	159
4.2	<i>Materials and Methods</i>	161
4.2.1	Patient cohort	161
4.2.2	Defining genotype groups	161
4.2.3	Genotype-phenotype comparisons	162
4.2.4	Using genotype-phenotype studies to validate prior findings in this study	163
4.3	<i>Results</i>	165
4.3.1	Genotyping results and metrics	165
4.3.2	Genotype-positive vs genotype-negative	166
4.3.3	Distinguishing between familial and non-familial HCM	169
4.3.4	Genotype-positive subgroups	170
4.3.5	Cases with multiple pathogenic variants	172
4.3.6	Using genotype-phenotype data to validate conclusions about genetics of HCM	172
4.4	<i>Discussion</i>	178
4.4.1	Distinctive clinical profile of genotype-positive and negative patients	178
4.4.2	Impact on genetic testing in HCM	179
4.4.3	Phenotype data can help to validate hypotheses about the genetics of HCM	181
4.4.4	Limited phenotypic differences observed depending on class of pathogenic variant	182
4.5	<i>Conclusion</i>	184

5	Concluding Remarks	185
6	Bibliography	186
7	Appendices	212
7.1	<i>Summary of evidence for non-sarcomeric genes implicated in HCM</i>	212
7.2	<i>List of variants detected in prospective HCM cohort</i>	223
7.3	<i>Permissions to re-use third-party copyrighted works</i>	229

List of Figures

1 - Introduction

Figure 1.1: Comparison of HCM, DCM and ARVC to normal heart	18
Figure 1.2: Mortality rates for HCM and other chronic diseases.	21
Figure 1.3: Proteins and cellular components involved in cardiomyopathies	24
Figure 1.4: Comparison of WGS, WES and panel sequencing	32

2 - Defining Cardiomyopathy Genes

Figure 2.1: Increase in the number of genes implicated in HCM and DCM.	39
Figure 2.2: Outcome of clinical testing for DCM by LMM.	44
Figure 2.3: Comparison between ExAC, 1KG and ESP population databases	52
Figure 2.4: Sequencing metrics for Sanger-confirmed SOLiD variants.	57
Figure 2.5: Classification system to assess the evidence for genes implicated in HCM	62
Figure 2.6: Rare variant frequencies of HCM and DCM cohorts and ExAC	64
Figure 2.7: DCM rare variant excess for all samples and Caucasian only.	71
Figure 2.8: Comparison of the pathogenicity predictions of research studies with those predicted by excess over ExAC.	72
Figure 2.9: Evidence for genes that have been implicated in HCM	78
Figure 2.10: Expression levels of <i>MYH6</i> and <i>MYH7</i> from the GTEX portal.	83
Figure 2.11: Classes of evidence in the ClinGen clinical validity framework	94

3 - Variant interpretation in HCM

Figure 3.1: Evidence framework for the ACMG guidelines.	101
Figure 3.2: Combinations of ACMG rules required to achieve classifications of Pathogenic, Likely Pathogenic, Likely Benign and Benign.	102
Figure 3.3: LMM HCM genetic test results for white and non-white patients.	106

Figure 3.4: The use of EFs to evaluate methods for distinguishing pathogenic from benign variants.....	113
Figure 3.5: Proposed changes to the ACMG rules.....	117
Figure 3.6: Proportion of HCM cases with rare VUS compared to ExAC.....	124
Figure 3.7: Clustering of rare, non-truncating variants in HCM cohorts and ExAC for the five HCM genes with detected clustering of variants.....	130
Figure 3.8: Distribution of rare, non-truncating variants in HCM cohorts and ExAC for validated genes without significantly enriched clusters.....	133
Figure 3.9: Clinical HCM cases reclassified based on the new ACMG rules	136
Figure 3.10: Prospective HCM cases classified as actionable based on current and new ACMG rules.	137
Figure 3.11: Proportion of individuals in ExAC and gnomAD with <i>TNNT2</i> :p.R278C.	139

4 - Genotype-phenotype studies in HCM

Figure 4.1: Number of HCM cases in each of the defined genotype groups	165
Figure 4.2: Number of HCM cases in each of the genotype-positive subgroups.....	166
Figure 4.3: Phenotype comparisons for genotype-positive and negative cases.....	167
Figure 4.4: Correlation between maximum LVWT and indexed LVM.....	168
Figure 4.5: Phenotype comparisons for genotype-positive subgroups	171
Figure 4.6: Phenotype comparisons for genotype-positive cases, genotype-negative cases and cases with variants in putative, non-sarcomeric genes that have been shown to have no evidence for involvement in HCM	173

List of Tables

1 – Introduction

Table 1.1: Classes of evidence used in the interpretation of variants.	35
--	----

2 - Defining Cardiomyopathy Genes

Table 2.1: Studies associating sarcomeric genes with HCM between 1989 and 1999	38
---	----

Table 2.2: HCM gene test coverage from 17 clinical genetics laboratories and diagnostic companies as of July 2017	43
--	----

Table 2.3: Summary of the sample cohorts used in this study for gene burden testing.	48
--	----

Table 2.4: Transcripts analysed and number of samples sequenced in each cohort for each gene assessed in this study.	49
--	----

Table 2.5: Cohorts that comprise the ExAC database	51
---	----

Table 2.6: Classification system adopted in this study to assess the evidence for association for genes implicated in HCM.	63
--	----

Table 2.7: Frequencies of rare variation in clinical HCM cohorts compared to ExAC.....	66
---	----

Table 2.8: Frequencies of rare variation in clinical DCM cohorts compared to ExAC.....	69
---	----

Table 2.9: Comparison of the excess of rare variants in LMM DCM cases over ExAC between Caucasian only and all sample analysis.....	71
--	----

Table 2.10: Summary of genes associated with complex phenotypes that include left ventricular hypertrophy.....	76
---	----

Table 2.11: Evidence validating the role of eight validated sarcomeric genes in HCM.	77
--	----

Table 2.12: Evidence for association with HCM for 31 non-sarcomeric genes.....	79
---	----

3 - Variant interpretation in HCM

Table 3.1: ExAC status of variants implicated in HCM by HGMD.	123
---	-----

Table 3.2: Odds ratios and etiological fractions for HCM genes.	126
---	-----

Table 3.3: True positive rate (TPR) and true negative rate (TNR) values for missense functional prediction scores for eight core sarcomeric HCM genes.	128
Table 3.4: Comparison of variant clustering and consensus functional prediction scores in differentiating EF values for non-truncating/missense variants in 5 HCM genes.....	133
Table 3.5: Frequency of <i>TNNT2</i> :p.R278C in published HCM cohorts.....	139
Table 3.6: ORs and penetrance for <i>TNNT2</i> :p.R278C for HCM versus population cohorts....	140
Table 3.7: Summary of reports of <i>TNNT2</i> :p.R278C from Pubmed and ClinVar.	141

4 - Genotype-phenotype studies in HCM

Table 4.1: Summary of demographic, baseline clinical and outcome data for major published HCM studies comparing genotype-positive and genotype-negative patients.....	155
Table 4.2: Summary of demographic, baseline clinical and outcome data for major published HCM studies comparing MYBPC3 and MYH7 variant carriers.....	159
Table 4.3: Summary of demographic and clinical phenotype data for genotype-positive and genotype-negative HCM cases.	168
Table 4.4: Summary of demographic and clinical phenotype data for genotype-positive cases, genotype-negative cases with a FH of HCM or SCD and genotype-negative cases without a FH of HCM.....	169
Table 4.5: Summary of the demographic and clinical phenotype data for genotype-positive subgroups of HCM patients.	170
Table 4.6: Genotype-phenotype comparisons comparing carriers of rare variants in non-sarcomeric genes implicated in HCM with genotype-positive and negative cases	174
Table 4.7: Genotype-phenotype comparisons of patients with and without variants in putative modifiers genes.	175
Table 4.8: Summary of the demographic and phenotypic characteristics of two sub-groups of patients with VUS, compared to genotype-positive cases and negative cases	177

Contribution of others to work presented in this thesis

Stuart Cook and Sanjay Prasad assigned HCM patients from the prospective cohort described in this study to the Biobank database. Basic CMR measurements were in most cases performed by the reporting clinician. BRU research nurses administered the patient questionnaire. The HCM cohort in the database was curated by clinical Fellows and previous clinical PhD students. Demographics and CMR measurements were extracted from reports by Leanne Felkin.

Rachel Buchan, Alicja Wilk and William Midwinter extracted DNA, prepared the sequencing libraries and performed next generation sequencing for the prospective HCM cohort described in this study.

Sung Gong, Shibu John and Risha Govind established the sequence alignment, variant calling, and basic variant annotation pipelines used for the prospective HCM cohort.

The analysis between the clinical HCM genetic data and the ExAC database described in Chapter 2 and parts of Chapter 3 was done in collaboration with the group of Professor Hugh Watkins at the University of Oxford – Kate Thomson and I performed most of the data analysis with statistical support from Professor Martin Farrall.

The variant clustering algorithm described in Chapter 3 was developed by Francesco Mazzarotto of the Careggi University Hospital Florence.

Acknowledgements

Firstly I would like to thank my supervisors Professor Stuart Cook and Dr Paul Barton for their support and guidance throughout this project and for the opportunity to do research in such a wonderful lab. I have had the honour and privilege of working with a group of extraordinarily talented people for the last several years, doing world class genomics research and without whom none of what is described in this thesis would have been possible. I would particularly like to thank James Ware who provided guidance and sage advice throughout this project. I am also grateful for the support and assistance from fellow researchers in the Cardiovascular Genetics and Genomics lab – Francesco Mazzarotto, Paz Tayal, Leanne Felkin, Nicky Whiffin, Katie Francis and Gillian Rea (in particular thanks to Francesco for developing the variant clustering algorithm and to Leanne for her sterling efforts in assembling the HCM phenotype data). The sequencing data generated by Rachel Buchan, Alicja Wilk and Will Midwinter was not just of huge scale but also of consistently excellent quality. The bioinformatics infrastructure and variant analysis and annotation pipeline developed by Sung Gong, Shibu John, Risha Govind, Shun Liang, Mian Ahmed, Erica Mazaika and Liz Edwards enabled me (and many other researchers) to quickly and efficiently access variant data at great scale and accuracy. The insights we have gained from this into the genetics of cardiomyopathies highlights the enormous value to be gained from such an integrated sequencing and bioinformatics infrastructure. Work on the assessment of cardiomyopathy gene pathogenicity was done through a hugely productive and rewarding collaboration with Professor Hugh Watkins and his team, in particular Kate Thomson. Finally outside of the lab, huge thanks to friends and family for support, encouragement and laughs along the way, in good times and bad. Many fun evenings were had in the Sydney Arms (especially the glorious six “Chieftain” months!) – in particular thanks to my drinking buddies Francesco, Risha and Paul for long evenings in the pub but always making sure I made the last train home....

Glossary of terms and abbreviations

1KG	1000 Genomes project
ACMG	American College of Medical Genetics
ARP	Attributable Risk Percent among exposed
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
CI	Confidence Interval
CMRI	Cardiac Magnetic Resonance Imaging
CNV	Copy Number Variation
DCM	Dilated Cardiomyopathy
dHPLC	denaturing High Performance Liquid Chromatography
DNA	Deoxyribonucleic Acid
ECG	Electrocardiography
EF	Etiological Fraction
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
ExAC-NFE	Exome Aggregation Consortium - Non-Finnish European
FH	Family History
GATK	Genome Analysis Toolkit
GWAS	Genome Wide Association Study
HCM	Hypertrophic Cardiomyopathy
HGMD	Human Gene Mutation Database
HVOL	Healthy Volunteer
ICC	Inherited Cardiac Condition
ICD	Implantable Cardioverter Defibrillator
IHM	Interacting Heads Motif

LGE	Late Gadolinium Enhancement
LMM	Laboratory of Molecular Medicine
LOD	Logarithm Of Odds
LQTS	Long QT Syndrome
LV	Left Ventricular
LVEF	Left Ventricular Ejection Fraction
LVMI	Left Ventricular Mass (indexed)
LVNC	Left Ventricular Non-Compaction
LVOTO	Left Ventricular Outflow Tract Obstruction
LVWT	Left Ventricular Wall Thickness
MAF	Minor Allele Frequency
NGS	Next Generation Sequencing
NMC	Non-random Mutation Cluster
NYHA	New York Heart Association
OMGL	Oxford Molecular Genetic Laboratory
OMIM	Online Mendelian Inheritance in Man
OR	Odds Ratio
RAAS	Renin Angiotensin Aldosterone System
RCM	Restrictive Cardiomyopathy
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
SCD	Sudden Cardiac Death
VEP	Variant Effect Predictor
VUS	Variant of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Publications arising from this work

Published papers

1. **Walsh R***, Thomson KL*, Ware JS, Funke BH, Woodley J, McGuire KJ, Mazzarotto F, Blair E, Seller A, Taylor JC, Minikel EV, Exome Aggregation Consortium, MacArthur DG, Farrall M, Cook SA, Watkins H. ***Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples.*** Genetics in Medicine 2017 Feb;19(2):192-203.
2. **Walsh R**, Buchan R, Wilk A, John S, Felkin LE, Thomson KL, Chiaw TH, Loong CC, Pua CJ, Raphael C, Prasad S, Barton PJ, Funke B, Watkins H, Ware JS, Cook SA. ***Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes.*** European Heart Journal 2017 Dec;38(46):3461–3468.
3. **Walsh R** and Cook SA. ***Issues and Challenges in Diagnostic Sequencing for Inherited Cardiac Conditions.*** (REVIEW) Clinical Chemistry 2017 Jan;63(1):116–128.
4. Whiffin N, Minikel E, **Walsh R**, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, Ware JS. ***Using high-resolution variant frequencies to empower clinical genome interpretation.*** Genetics in Medicine 2017 Oct;19(10):1151-1158.

Conference oral presentations

1. Invited speaker at Cardiomyopathies: A look at the future, Florence, December 2016: *Re-evaluating the genetic architecture of cardiomyopathy.*
2. Invited speaker at BHF/EPSRC Joint Symposium: The Heart - Models, Measurements, Medicine, London, June 2017: *Identifying actionable variants in cardiomyopathy.*
3. Invited speaker at Slovenian Society of Medical Genetics, Ljubljana, October 2017: *Applying a quantitative approach to variant interpretation in inherited cardiac conditions.*
4. Oral presentation at International Cardiovascular Genomic Medicine Conference, Cardiff, October 2017: *Quantitative approaches to variant classification increase the yield, precision and equitability of genetic testing in HCM.*

1 Introduction

1.1 Cardiomyopathies

Cardiomyopathies are a range of disorders of the cardiac muscle leading to structural and functional abnormalities in the heart. Primary cardiomyopathies, i.e. those not the result of more general cardiac or systemic conditions, are often familial and tend to be inherited in a broadly Mendelian manner. The most common cardiomyopathies are hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM), both of which are estimated to affect at least 1 in 500 people and therefore represent some of the most common Mendelian genetic diseases. HCM is characterised by hypertrophy (i.e. thickening) of the walls of the left ventricle and is a leading cause of sudden cardiac death (SCD) in young people and athletes. DCM is characterised by enlargement of the left ventricle and systolic dysfunction - it is associated with poor outcomes and is the leading cause of heart transplantation. Less common cardiomyopathies include arrhythmogenic right ventricular cardiomyopathy (ARVC), left ventricular non-compaction (LVNC) and restrictive cardiomyopathy (RCM). The genetic basis of cardiomyopathies has been the source of intense investigation in recent years. Clinical genetic testing for these conditions is now regularly pursued and has been a clear example of the value of genomics in medical practice. The genetic architectures of each cardiomyopathy have distinctive features, though with overlap in some of the genes involved and in the challenges faced for the effective application of genomics in the clinic.

1.2 Overview of Hypertrophic Cardiomyopathy

1.2.1 *Definition of Hypertrophic Cardiomyopathy*

HCM is an inherited disorder of the cardiac muscle characterised by a complex pathophysiology and marked heterogeneity in morphology, clinical manifestations, genetic aetiology and outcomes¹. Typical features of HCM include asymmetric left ventricular (LV) hypertrophy that varies widely in extent and location, myocyte disarray (the spatial disorganisation of myocardial fibres) and increased replacement and interstitial fibrosis, usually in the mid-myocardium where wall thickness is at a maximum. Familial HCM is largely inherited in an autosomal dominant manner, though with variable disease expressivity and

often incomplete penetrance such that clinical diagnosis is often delayed until adolescence or adulthood. However, a substantial proportion of cases lack family history and affected relatives, which may be due to *de novo* variants in patients or the lack of any known causative genetic factor.

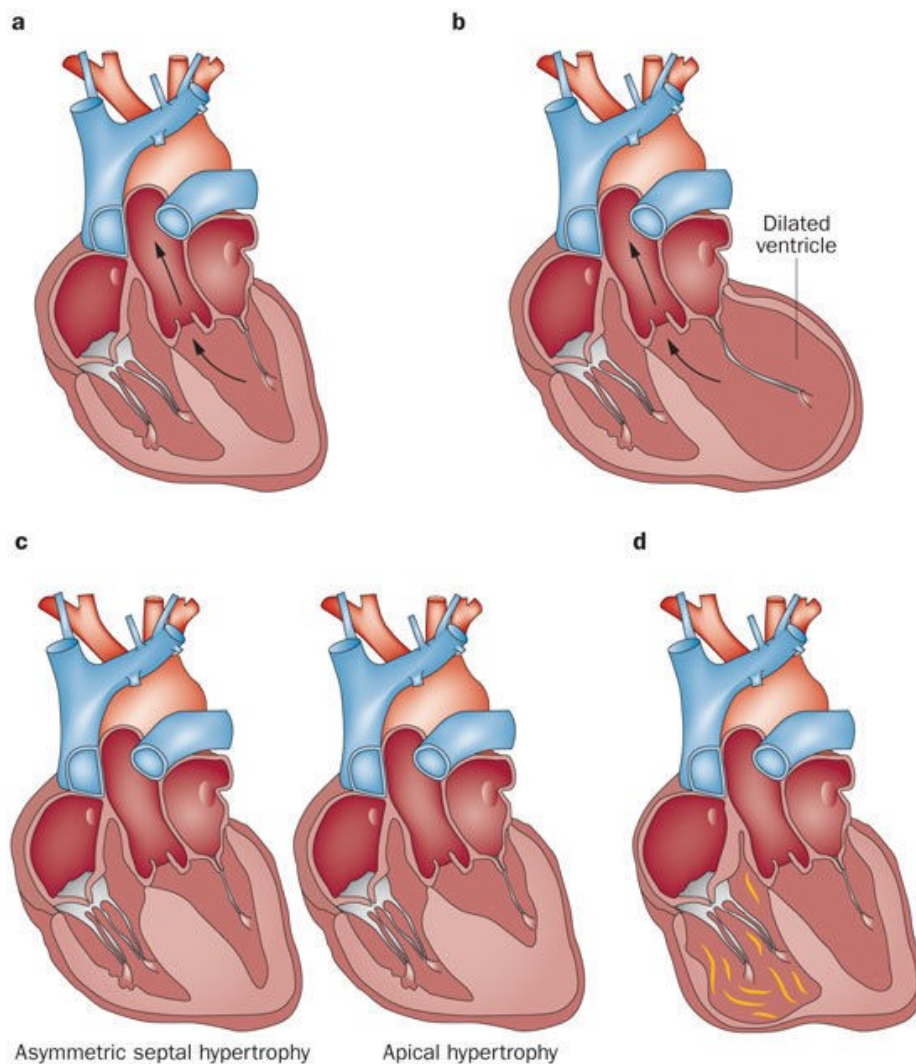


Figure 1.1: Examples of two common patterns of hypertrophy in HCM – asymmetric septal hypertrophy and apical hypertrophy (c) and comparison with normal heart (a), dilated cardiomyopathy (b) and arrhythmogenic right ventricular cardiomyopathy (d). Reproduced from Wilde and Behr² with permission.

HCM is defined by increased LV wall thickness (LVWT) with non-dilated LV chambers that is not solely explained by abnormal loading conditions such as hypertension and aortic stenosis³. A maximum LVWT of $\geq 15\text{mm}$, assessed by echocardiography or cardiac magnetic resonance imaging (CMRI), is typically the threshold used to diagnose HCM. Patients with LVWT between 13-14mm require additional evidence such as family history or other imaging abnormalities. For first degree relatives of HCM patients, an otherwise unexplained maximum LVWT of $\geq 13\text{mm}$ can be used for clinical diagnosis, while for children maximum LVWT should be more than two standard deviations from the expected mean (for age, sex and body size)³.

1.2.2 Prevalence of HCM, issues in diagnosis and clinical features

The prevalence of HCM in the population is generally estimated at approximately 1 in 500 people (0.2%)^{3,4} though, given the age-dependent onset of cardiac hypertrophy, this rate varies depending on the cohort studied. HCM was detected in 0.29% of 60 year olds undergoing echocardiography-based cardiac evaluation in the United States⁵ compared to an estimated annual incidence of 0.3-0.5 per 100,000 in paediatric studies^{6,7}. A recent study has suggested that the prevalence of HCM may have been underestimated, with more advanced imaging techniques and the frequency of sarcomeric gene variants leading to an estimate of 0.6%⁸, though these findings have been disputed⁹.

The use of maximum LVWT to both diagnose patients with HCM and to estimate the prevalence of the condition in the population is problematic for a number of reasons. The recommended threshold of 15mm is not adjusted for age, sex, race or body surface area and does not account for normal regional variation in LVWT. The threshold therefore can range between 2 and 6 standard deviations above the normal adult mean¹⁰. Additionally, although the guidelines stress that the detected hypertrophy should not be explained solely by loading conditions, distinguishing HCM from LV hypertrophy secondary to hypertension or athletic adaptation (“athlete’s heart”) can be challenging. This is particularly an issue with individuals of African descent who are more susceptible to LV hypertrophy and repolarization abnormalities in response to moderate cardiac stresses such as hypertension or exercise^{11,12}. As LV hypertrophy can be a late onset condition, it can also be challenging to distinguish it

from the relatively common condition of isolated basal septal hypertrophy, estimated to affect approximately 8% of those over the age of 70¹³. Finally, there are rare but important examples of patients with HCM but without LV hypertrophy – these are usually associated with variants in the *TNNT2* gene and can carry increased risk of SCD^{14,15}, as well as the increasing number of relatives of HCM patients who have a pathogenic variant but currently lack any overt phenotype.

For these reasons, factors other than maximum LVWT are taken into account when diagnosing HCM. LV hypertrophy in HCM patients is generally regional and asymmetric, in contrast to the more concentric hypertrophy characteristic of hypertension and physiological adaptation. Standard 12-lead electrocardiography (ECG) usually reveals anomalies, in particular pathological Q-waves and ST- and T-wave abnormalities. Family history for HCM, SCD and unexplained heart failure and genetic testing for sarcomeric gene variants (see below) can also be used to confirm diagnosis, although a significant proportion of HCM cases will lack family history and/or a causative variant. Left ventricular outflow tract obstruction (LVOTO) is present in approximately one-third of HCM patients and is associated with poorer prognosis and increased risk of SCD.

1.2.3 Clinical course and management

Perhaps one of the surprising aspects of HCM is the relatively favourable prognosis for patients, in contrast to the often negative perceptions they will have at diagnosis¹⁶. It is now estimated that the annual mortality rate for HCM is only 0.5%, equivalent to the general population and significantly lower than other chronic cardiac and non-cardiac conditions (Figure 1.2). The rate for children and young adults, who can often suffer a more aggressive clinical course, is not substantially higher than this^{16,17}. These low rates of mortality are in contrast to earlier reports from the 1970s and 1980s that estimated annual death rates at 2-4% in adults (and up to 6% in children)¹⁸, a decline that can be explained by both selection bias, as earlier reports were based on tertiary centres that were more likely to include more severe cases, and by improved treatment options for HCM patients, particularly implantable cardioverter defibrillators (ICDs) to prevent SCD.

While the majority of HCM patients may experience normal life expectancy without debilitating symptoms, certain subgroups are at risk of serious complications and premature death. The most severe of these is the risk of SCD which is particularly prevalent in those under 40 years of age. Given the grave and unpredictable nature of SCD, and the ability of ICDs to treat the ventricular tachyarrhythmias that cause it¹⁹, there have been substantial efforts to identify at-risk sub-groups for whom prophylactic ICD therapy will be most beneficial (complications with ICD devices, including lead defects and inappropriate shocks, ensure they need to be targeted to those at most risk). Current models for risk stratification account for the following factors – family history of HCM-related SCD, unexplained syncope, multiple repetitive non-sustained ventricular tachycardia, abnormal exercise blood pressure response, LV hypertrophy $\geq 30\text{mm}$ and late gadolinium enhancement (LGE, a CMRI-derived marker of myocardial fibrosis) of $\geq 15\%$ of LV mass. Two models have been developed – European²⁰ and American²¹ – that use different criteria to assess the risk of SCD.

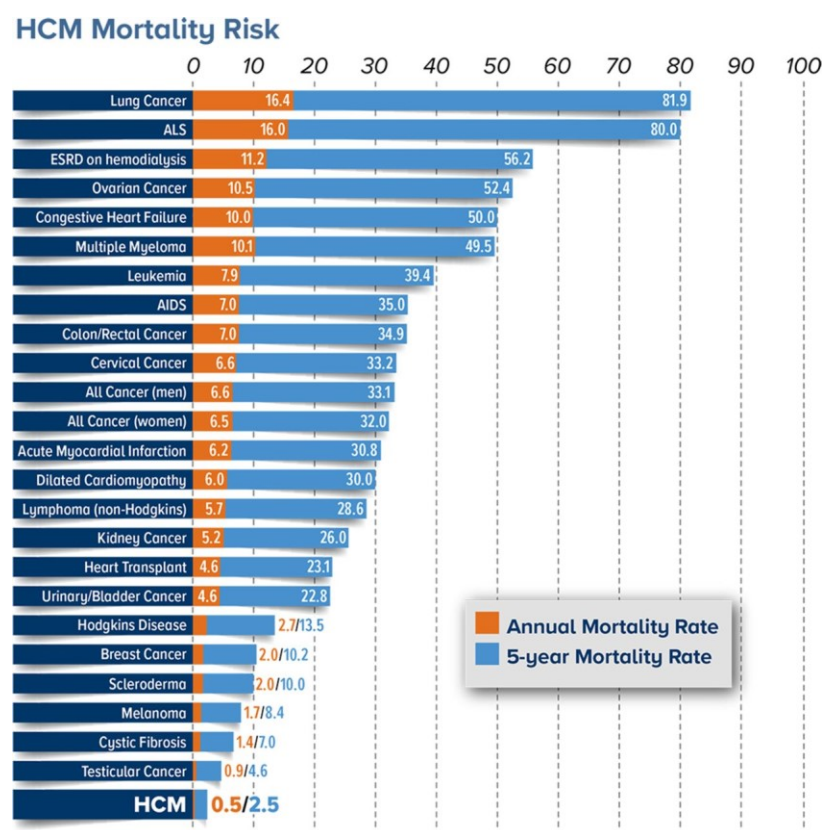


Figure 1.2: Comparison of the annual and 5-year mortality rates (with current treatments) for HCM and other chronic diseases. Reproduced from Maron et al.¹⁶ with permission.

Some degree of heart failure with preserved systolic function occurs in approximately half of HCM patients though in most this is readily treatable with drugs or surgery and does not progress to advanced heart failure. LVOTO is the most important cause of these symptoms, with β -adrenergic receptor blockers the first line of treatment for relieving ischaemic chest discomfort and attenuating exercise-induced LVOTO²². In patients with more advanced symptoms and who are refractory to pharmacological treatment, surgical septal myectomy is recommended and has been shown to be highly effective at reversing heart failure symptoms with minimal operative mortality²³. Percutaneous alcohol septal ablation has been developed as a non-surgical alternative to myectomy, particularly for those of an advanced age and for whom comorbidity may increase the risk of surgery. Patients with non-obstructive HCM will mostly experience a non-symptomatic and stable clinical course. However, a small minority (approximately 3%) will develop advanced, end-stage heart failure (referred to as “burned out HCM” that can resemble DCM) characterised by systolic dysfunction with left ventricular ejection fraction (LVEF) < 50%, adverse LV remodelling, regression of hypertrophy and cavity enlargement²⁴. As these patients tend to be relatively young with an average age of 43 years, heart transplantation is the only definitive treatment option.

1.2.4 Future treatment options

Although current treatment options are relatively successful for most HCM patients, they target symptoms rather than the underlying genetic defect or pathways involved in the pathogenesis of the disease. A number of novel approaches, at pre-clinical and clinical trial stage, are currently being investigated as potential new HCM therapies. Experiments to repair genetic defects in mice have utilised a range of techniques including gene therapy with adeno-associated virus²⁵, repair of *Mybpc3* mRNA with 5' trans-splicing²⁶, exon skipping mediated by antisense oligoribonucleotides inserted into modified U7 small nuclear RNA²⁷, and RNAi that silences mutant *Myh6* transcripts²⁸. In 2017, a ground-breaking study that used CRISPR-Cas9 to correct the pathogenic *MYBPC3* variant c.1420_1423delGAGT in human embryos made headlines across the world²⁹, suggesting that disease-causing variants could one day be repaired in embryos prior to implantation. However, this approach currently offers little practical advance on the current practice of pre-implantation genetic diagnosis, where embryos are genetically tested prior to implantation to select those lacking the identified

pathogenic variant. For treatments closer to potential clinical application, there were over 50 ongoing trials for HCM registered at <https://clinicaltrials.gov/> as of December 2017. Of most interest are the phase 2 VANISH trial assessing the effects of the angiotensin receptor blocker valsartan³⁰ and the phase 2 trial of the MYK-461 compound developed by Myokardia Inc, which inhibits myosin ATPase activity and is being evaluated in HCM patients with LVOTO³¹.

1.3 Genetics of Hypertrophic Cardiomyopathy

1.3.1 *Current knowledge of the genetic basis of HCM*

Several studies in the 1950s and 1960s described a myocardial disease, eventually defined as hypertrophic cardiomyopathy, that clustered in families. By the 1980s, the development of echocardiography, which could detect clinical features of HCM in affected family members, and linkage analysis, which used restriction fragment length polymorphisms (RFLPs) to identify chromosomal regions that segregated with disease, allowed researchers to identify the genes and mutations that were responsible for HCM in these families³². Studies by the Seidman lab and other research groups identified variants in the beta myosin heavy chain encoding *MYH7* gene, the myosin binding C protein encoding *MYBPC3* gene and six other sarcomeric genes as the causative variants in numerous large HCM pedigrees (Figure 1.3), providing definitive evidence for the role of these genes in the disease.

Subsequent sequencing of cohorts of HCM cases identified putative disease-causing variants in these 8 sarcomeric genes in 30-60% of patients, depending on the characteristics of the cohort analysed, in particular the proportion of cases with a family history. Over 1,500 distinct variants in these genes have since been identified in HCM patients – although there are examples of founder mutations in certain population groups and some relatively common recurrent pathogenic variants, many variants are private and unique to individual families. *MYH7* and *MYBPC3* account for over 70% of identified pathogenic variants in HCM, with variants in the other sarcomeric genes occurring in less than 5% of patients each^{33,34}. For most HCM genes, missense substitutions (leading to the alteration of a single amino acid) are the

predominant class of pathogenic variants and are believed to act in a dominant negative manner through incorporation of the mutated protein into the sarcomere. In contrast, almost half of the pathogenic variants in *MYBPC3* are truncating variants that would lead to radical changes in protein sequence, but whose transcripts are likely to be degraded before translation. They are believed to cause HCM through haploinsufficiency, i.e. a reduced amount of protein product available in cardiomyocytes^{35–37}.

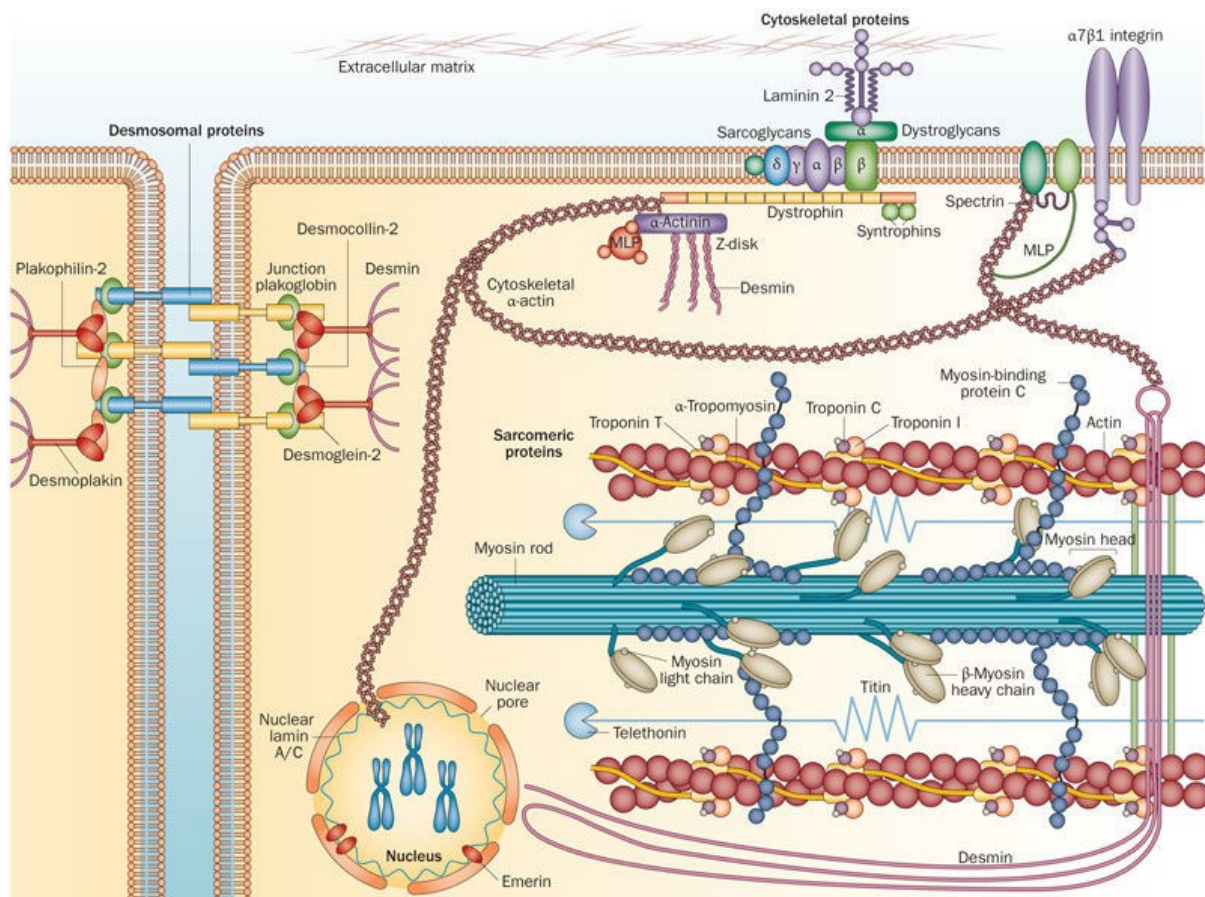


Figure 1.3: Major proteins and cellular components involved in cardiomyopathies. Most HCM cases with known genetic aetiology have variants in sarcomeric genes – thick filament (blue) and thin filament (red). Reproduced from Wilde and Behr² with permission.

The large proportion of cases that lacked a causative pathogenic variant in one of these sarcomeric genes led to an ongoing search for novel HCM-causing genes that could explain the apparent missing heritability for this condition. As we will see in Chapter 2, although numerous putative novel HCM genes have been proposed, most were assessed by candidate

gene studies rather than unbiased genomic analyses. As these studies lacked evidence from large family pedigrees (most of which yielded pathogenic variants in sarcomeric genes upon genetic analysis), there have been substantial doubts as to how prevalent pathogenic variants in these genes are, if at all, in causing HCM, particularly in light of the insights offered into human genetic variation by large scale population sequencing studies.

1.3.2 Role of genetic testing in HCM

The primary purpose of genetic testing in HCM patients is to facilitate cascade screening in relatives of the proband to identify family members both at risk of developing the disease and those free from risk. Current guidelines recommend clinical evaluation with imaging and 12-lead ECG for first degree relatives of all HCM patients – annually for children from 10-12 years until physical maturity and every 2-5 years for adults³. If a definitive pathogenic variant is identified in the proband, relatives who are carriers of the variant can receive regular targeted clinical surveillance and advice on reproduction and lifestyle choices (e.g. participation in competitive sports). Detailed clinical examination may detect early, pre-symptomatic signs of disease^{38–40} and enable treatment to be commenced prior to the development of serious symptoms. Recent studies have suggested that genotype-positive, phenotype-negative individuals are likely to have a benign clinical course^{41,42} although, as HCM is often a late-onset disease, ongoing evaluation is recommended with a proportion of such individuals expected to develop cardiomyopathy. Given the relatively benign prognosis for HCM and the variability in disease expression, pre-natal diagnosis is not recommended for the disease³, although preimplantation genetic diagnosis is now a feasible choice for family planning with variant carriers⁴³.

Equally importantly, genetic testing enables relatives who are non-carriers to be freed from ongoing clinical examination, assuming they have no clinical symptoms. In addition to the clear psychological benefit this provides for these individuals, it also has the potential to provide significant cost savings for healthcare providers. A US study on clinical genetic testing of 2,912 HCM probands estimated savings of \$700,000 from 691 genotype-negative family members³³, though this is likely to be a considerable underestimate as it was based on

Medicare rates for clinical evaluations. UK and Australian studies have also concluded that genetic cascade screening is more cost-effective for HCM than clinical diagnosis alone, with the incremental cost per life year saved/gained of €14,397 (UK, 2010⁴⁴) and €9,509 (Australia, 2012⁴⁵), well within the recommended thresholds of €35,000 (UK National Institute for Health and Care Excellence guidelines) and €31,400 (Australian Pharmaceutical Benefits Advisory Committee). Given the conservative estimates used in these models, the benefits associated with expanding cascade screening beyond first degree relatives and recent dramatic reductions in the costs of genetic sequencing, the cost-effectiveness of comprehensive genetic testing in HCM is likely to be substantially greater.

The benefits of genetic testing on patients who have already been diagnosed with HCM are limited in the majority of cases. While a positive genetic test can be used to support diagnosis in patients, particularly in borderline cases where the diagnosis may be uncertain or to definitively distinguish HCM from hypertensive cardiomyopathy or athlete's heart, it will not impact on clinical management or treatment for most patients. Although some early reports suggested that particular classes of variants might predict clinical outcome³⁴, subsequent studies have failed to identify specific and consistent correlations between genotype and phenotype, despite some broad associations that are described in Chapter 4. The exception to this is identifying patients who have conditions that mimic sarcomeric HCM upon cardiac imaging but are in fact different genetic conditions that have distinct clinical profiles, inheritance patterns and treatment options. Genetic testing can identify these "genocopies", often inaccurately referred to as phenocopies, when clinical screening has suggested a HCM phenotype, or can be used to confirm diagnosis of a suspected non-HCM phenotype. Patients with these genocopy conditions may account for 3-5% of those undergoing HCM genetic testing. Examples include Fabry disease, caused by variants in the *GLA* gene, which warrants enzyme replacement therapy⁴⁶ and Danon disease, caused by variants in the *LAMP2* gene, which has a particularly severe prognosis with survival past 25 years uncommon⁴⁷.

1.3.3 Beyond a Mendelian model of disease

The remarkable phenotypic heterogeneity observed in HCM patients with similar causative variants, even in family members with the same variant, belies a characterisation of HCM as a simple monogenic Mendelian disease. For many patients, HCM is of course still primarily a genetic disease in that a causative sarcomeric variant that has been identified in the family is necessary for expression of the disease (and thus enabling cascade screening). However, it is clear that other factors must influence whether HCM develops in variant carriers (penetrance of the variant) and the severity of the phenotype in those where disease occurs (expressivity of the variant). While definitive studies into these modifying factors have yet to be performed, initial reports suggest a variety of genetic, epigenetic, environmental and other factors may be involved.

In a minority of HCM cases, a more severe phenotype has been observed in patients carrying multiple pathogenic sarcomeric variants⁴⁸, although the prevalence of these has been recently called into question through re-analysis of variant pathogenicity⁴⁹ (see Chapter 4 for further details). Non-Mendelian genetic variants may also play a role in modifying disease expression. Rare variants in the muscle ring finger proteins MuRF1 and MuRF2 (coded for by the *TRIM63* and *TRIM55* genes) have been proposed to act as modifiers of HCM and were associated with an earlier age of onset and greater maximum LVWT in one study⁵⁰. Variants in troponin genes that are not rare enough to act as Mendelian variants (or would do so only with very low penetrance) may also influence disease expression in some patients (see Chapter 3, section 3.3.5, for details). There have been limited studies investigating a role for common variants in HCM – a small genome wide association study (GWAS) identified a polymorphism in *FHOD3* that was enriched in HCM cases, although carriers of this variant had similar phenotypes to non-carriers⁵¹. Other studies have noted that polymorphisms in the Renin Angiotensin Aldosterone System (RAAS) pathway, involving the *ACE*, *AGTR1*, *CMA*, *AGT* and *CYP11B2* genes, have been shown to influence the severity of LV hypertrophy^{52–54}.

Indirect genetic and epigenetic factors may also play a role in influencing LV hypertrophy and disease expression. Common variants may serve to regulate the expression of mutated

alleles, as has been observed in another genetic cardiac condition long QT syndrome (LQTS), where polymorphisms in the 3' untranslated region of *KCNQ1* were shown to alter the expression of the associated transcript and therefore the balance between normal and mutant protein⁵⁵ (though these findings have since been disputed^{56–58}). Micro RNAs (miRNA) are small, non-coding RNAs that regulate the expression of mRNA by repressing translation or decreasing mRNA stability. Numerous miRNAs have been shown to be differentially expressed with hypertrophy in humans and animal models^{59–61}. Variants in miRNA sequences themselves, or in the target regions of miRNAs, could conceivably play a role in modifying disease expression, although no such variants have thus far been identified for HCM.

Other environmental, physiological and medical factors are also likely to play a role in the modifying the expression of disease. As described above, physiological adaptation due to exercise or systemic hypertension are known to influence LV hypertrophy while a recent study has also associated obesity with increased LV mass⁶². Gender is also a recognised risk factor for HCM, with the condition more common in males than females (discussed further in Chapter 4), though the mechanisms behind this are unclear. Overall, while numerous genetic and non-genetic factors have been implicated as modifiers of HCM, much larger studies will be needed to definitively assess the role and extent of these associations, including large and sufficiently powered GWAS to identify common variants that may influence disease expression.

1.4 Approaches to genetic sequencing and variant analysis

1.4.1 Role of genetic sequencing in Mendelian diseases like HCM

Genetic sequencing can be utilised for a number of different reasons to investigate Mendelian diseases like HCM, both for clinical management of patients and their families and to research the genetic basis of the disease. Genetic testing can be divided into a) detecting a predefined variant in a patient or individual, b) identifying variants within a panel of genes that have

already been associated with the disease, and c) discovering variants within novel genes that have not been previously implicated in the condition.

The simplest application of genetic sequencing is testing for the presence of a specific variant – given the high degree of genetic heterogeneity in HCM (and other inherited cardiac diseases), this is usually performed in the context of cascade screening of family members once a pathogenic variant has been identified in the index patient (as described above). The next level, and currently the most common use of genetics in HCM, is to sequence a panel of genes that have been implicated in HCM to discover the causative pathogenic variant in a patient or cohort of patients (either novel variants or variants that have been previously described in other patients or families). Such genetic testing has been available for HCM for approximately 15 years now and has become a standard and recommended part of the clinical management of this disease³. The key challenge in this approach remains determining if detected variants are disease-causing or if they simply represent rare benign variation (as discussed below and in detail in Chapter 3).

Finally, genetic sequencing can be used to identify variants in novel genes that have not been previously implicated in HCM, through either unbiased or targeted approaches. Unbiased techniques will scan the whole genome or exome to identify genetic variants or markers that are associated with the disease - such approaches were used to associate sarcomeric gene variation to HCM through linkage studies in the 1990s (as described above). Today, advances in sequencing technologies allow the whole genome or exome to be sequenced rapidly and cost effectively, enabling any source of genetic variation in a patient to be directly detected. However, the sheer number of rare variants in any one individual mean this is an impractical approach when applied to individual patients, given the difficulties in distinguishing the causative variant from the large numbers of rare benign variants. The only exceptions to this are if a large family pedigree, with defined clinical phenotypes, is available to test whether the candidate pathogenic variants segregate with disease, or if the mode of disease inheritance enables the list of rare variants to be rapidly triaged – most commonly this will involve recessive or *de novo* inheritance in an individual with unaffected parents (and sequencing data available from all three).

Targeted approaches to novel gene discovery take the form of candidate gene studies where a putative new disease gene (usually chosen on the basis of known function, involvement in related conditions or interaction with known disease genes) is sequenced in a cohort of disease probands, often followed by additional investigations to validate any novel association. This has been the most common method used until now to identify novel non-sarcomeric HCM genes but, as will be described in Chapter 2, this approach is prone to false positive findings if the studies are not rigorously designed and do not sufficiently control for rare benign variation.

1.4.2 Sequencing technologies and next generation sequencing approaches

When genetic testing was first applied to identify variants in sarcomeric genes in HCM patients, it depended on techniques such as denaturing high performance liquid chromatography (dHPLC), Sanger sequencing or array-based technologies. These methods were variously expensive (Sanger), time consuming (dHPLC, Sanger), lacking in analytical sensitivity (dHPLC) and difficult to scale up (all). Although diagnostic genetic testing using these techniques became established for HCM in a number of centres, these limitations prevented their wide-scale adoption as a routine part of clinical management. The development of next generation sequencing (NGS) technologies since 2005 has addressed many of these shortcomings and has hugely expanded the scope for genetic testing in HCM and many other conditions. NGS approaches have dramatically reduced the costs and timescales compared to traditional techniques, while offering analytically sensitive and accurate sequencing for variant detection, although many clinical laboratories still, likely unnecessarily, confirm findings with an alternative technique (usually Sanger sequencing). It also enables testing to be easily expanded to account for the increasing number of genes associated with genetic diseases, although this may be a double-edged sword given the number of dubious associations in the literature (Chapter 2).

Currently, NGS strategies for diagnostic testing include pre-defined gene panels, whole exome sequencing (WES) and whole genome sequencing (WGS), each of which offers distinct advantages and disadvantages. Gene panels can offer comprehensive sequencing of most or

all genes that have been associated with the disease in question, allowing both the sequencing and data analysis to focus on these validated genes. Panels such as Illumina's TruSight Cardio offer a single laboratory assay for a wide range of inherited cardiac conditions (ICCs)⁶³, with the possibility to restrict data analysis to just those genes or conditions in question. Gene panels have been demonstrated to obtain excellent coverage of target regions and analytical sensitivity in variant detection, outperforming both WES and WGS⁶³ (Figure 1.4) and enable more samples to be sequenced on single runs, thus reducing costs. However, panels do not allow for variant detection in novel genes, require periodic review to account for newly discovered gene associations and targeted sequencing is poor at detecting certain variants classes such as copy number variants (CNVs - large deletions or duplications of exons, genes and larger genomic regions).

WES involves the targeted sequencing of all coding regions of the genome, comprising approximately 30 million bases across more than 20,000 genes. WES enables comprehensive analysis of protein-altering variation in the genome (which accounts for the vast majority of known pathogenic variants in Mendelian disease), both for genes with a known role in the disease of interest and for the discovery of novel genetic associations. However as described above, in practice novel gene discovery is likely to be most effective in rare cases of recessive or *de novo* inheritance^{64,65}. WES can also be used as a comprehensive gene panel test, with analysis restricted to those genes associated with the disease and the possibility of re-analysis when novel genetic links are proposed. However, limitations in sequencing coverage that can affect important disease genes are a key limiting factor associated with WES and should be taken into account before pursuing this strategy.

WGS involves the complete sequencing of every base pair in the genome. It is the most comprehensive sequencing strategy available, provides more uniform coverage than WES, is more suited to detecting large structural variants and is increasingly an affordable option with ongoing reductions in sequencing costs. However, the practical advantages of WGS for diagnostic testing of patients with Mendelian diseases like HCM are currently questionable since prioritising the thousands of rare variants observed in each genome, and identifying actionable putative pathogenic variants, is an extremely difficult proposition, and even more

challenging than the analysis of novel coding variants identified by WES. There are also major cost implications relating to data processing and storage with WGS, as well as ethical issues around data ownership and the detection of incidental findings.

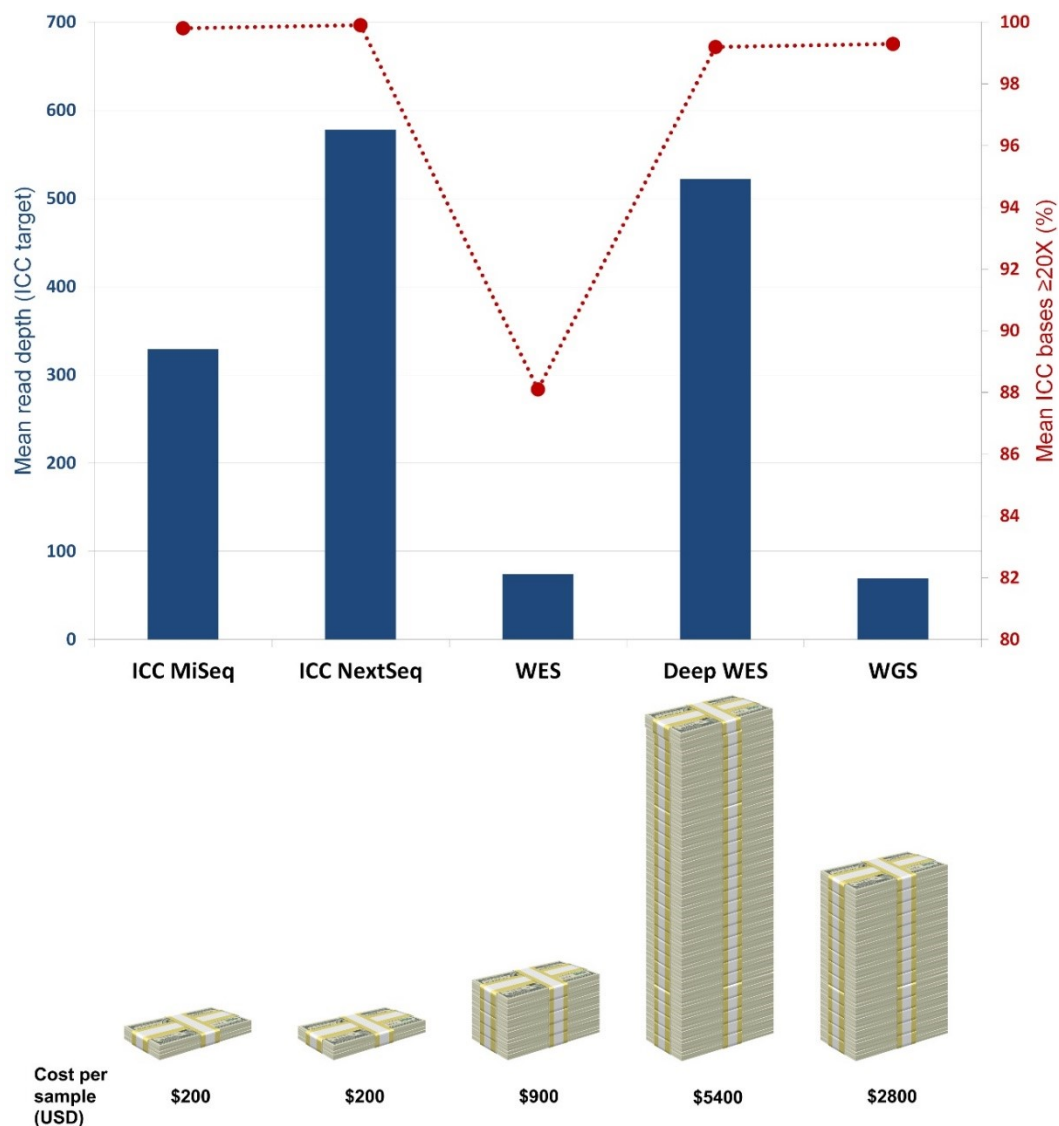


Figure 1.4: Sequencing coverage metrics for ICC genes on the TruSight Cardio panel using MiSeq and NextSeq and comparison with coverage for these genes using WES, Deep WES (i.e. 6 samples combined to obtain sequencing depth equivalent to TruSight Cardio panel) and WGS, as well as approximate targeted enrichment and sequencing cost per sample. Figure generated from data in Table 2 of Pua et al.⁶³

1.4.3 Analysis of NGS data – variant detection

The analysis of NGS data from germline DNA sequencing generally encompasses two distinct workflows – firstly the identification and annotation of genetic variants and secondly the interpretation of the effect the variant may have on the phenotype of a patient, i.e. whether the variant is pathogenic for the disease under investigation. The detection of variants from germline DNA sequencing is by now a well-established process, involving the mapping of sequencing reads to the reference genome template and the “calling” of variants (distinguishing the differences between the reads and reference genome that are genuine variants from sequencing errors or noise). Although the detection of CNVs and mosaic variants (where the variant is present in a subset of cell types and therefore detected in only a small proportion of NGS reads) remains challenging, the detection of standard heterozygous or homozygous substitutions and small insertions/deletions (<50 base pairs) has become consistent and standardised in recent years, especially for sequencing assays that produce a high depth of read coverage (such as targeted panels). The Genome Analysis Toolkit (GATK) software package, developed at the Broad Institute, is now the standard method for detecting variants of this type⁶⁶. The annotation of detected variants, i.e. translating the variant’s effect on the chromosome to its effect on transcript and protein sequences, is a relatively straightforward process through applications like the Variant Effect Predictor (VEP)⁶⁷, with inconsistencies only arising due to insertions/deletions that can be annotated at multiple positions or if different transcripts are used for annotation.

1.4.4 Analysis of NGS data – variant interpretation

In contrast, interpreting the likely clinical effect of variants found in patients remains one of the most challenging aspects in Mendelian disease genetics, with ongoing inconsistency between different reports and clinical genetics laboratories. A wide range of evidence types can be used to analyse variants (Table 1.1) but these vary in their availability for all variants, the strength of evidence they should confer and the consistency with which they are applied. An additional complication are the different approaches to variant interpretation employed in research and clinical genetics environments and the distinct motivations behind these. Many published research studies are seemingly motivated to maximise the yield of pathogenic variants and thus use relatively permissive approaches to assign pathogenicity

(such as non-stringent population frequency thresholds and non-specific computational algorithms). In contrast, clinical genetics laboratories strive to avoid false positive results, given the adverse medical consequences of such findings for patients and their families. Therefore, they require substantial evidence for the pathogenicity of detected variants but, given the genetic heterogeneity of diseases like HCM and the prevalence of private variants for which there is limited data, this may come at a cost of reduced test sensitivity and an overly conservative approach to variant interpretation (discussed in Chapter 3).

A number of resources are now available to assist in variant interpretation. Population genetic databases like the Exome Aggregation Consortium (ExAC) have transformed both our understanding of the extent of rare genetic variation in the human population and efforts to define whether variants detected in patients are rare enough to be plausibly pathogenic⁶⁸ (discussed in more detail in Chapters 2 and 3). They have also highlighted that many variants that had been previously associated with Mendelian diseases in the research literature are likely to be too common in the population to be disease-causing. Databases that aggregate variant-disease associations, such as OMIM (Online Mendelian Inheritance in Man, <https://www.omim.org>) and HGMD (Human Gene Mutation Database)⁶⁹ therefore have a substantial proportion of false positive entries that need to be taken into consideration when utilising them for variant interpretation. More recently, the ClinVar database⁷⁰ has attempted to provide a centralised resource for (theoretically) more stringent interpretation of variants, with clinical genetics laboratories encouraged to submit and share their assertions and associated evidence (assembled from published literature, databases and internal data). While this database often provides more rigorous and comprehensive variant analysis, the lack of consistency between laboratories (for both the assertion and the level of evidence provided), as well as the laborious nature of accessing its unstructured data, currently limit the effectiveness and usability of this resource.

Evidence	Pros	Cons
Segregation	- Can provide very strong evidence of pathogenicity if observed in sufficient affected family members, especially including more distant relatives. - Lack of segregation, i.e. absence of the variant in clearly affected family members, provides strong evidence against pathogenicity. - Data from other family pedigrees with the same variant (published and previous lab experience) can also be used as evidence.	- Large family pedigrees are rarely available for genetic testing in the clinic. - Incomplete penetrance of many ICCs means only affected family members can be used in segregation analysis. - Segregation may imply that the locus is linked to the disease rather than the specific variant. - Negates one of main benefits of genetic testing, i.e. cascade screening in family to assess disease status.
Known pathogenicity	- Established evidence that a variant is pathogenic for a disease provides a quick and clear result.	- Defining the amount and type of evidence that is sufficient to unambiguously classify a variant as pathogenic can be difficult and subjective. - Many variants in published studies have been incorrectly labelled as disease-causing. - Gathering evidence on known variants from the literature and online databases can be time-consuming.
Functional evidence	- Can provide powerful data on the functional effect of variants, particularly in well validated assays.	- Lack of standardisation of functional assays for many ICCs. - Translation of cellular and tissue assays and animal models to the clinical phenotype is often uncertain.
Population frequency	- Allows for quick filtering of most variants identified in broad genetic tests.	- Rarity in population databases does not imply that a variant is pathogenic. - As population databases can include affected individuals, determining a frequency that is compatible with pathogenicity can be difficult. - The patient's ethnicity should match the population studied, data for certain population groups can be limited or unavailable.
Computational prediction	- Data is easily calculated for all missense and splice region variants. - A multitude of algorithms are available, with consensus predictions offering increased accuracy.	- Low positive predictive values mean these can only be used as supporting evidence for pathogenicity. - Assess effect on protein function rather than pathogenicity in disease.
Pathogenic hotspots	- Variants occurring in regions enriched in disease cohorts and/or depleted in controls have increased likelihood of pathogenicity. - Large datasets can generate a prior probability of pathogenicity for variants found in enriched regions.	- Not relevant for all genes associated with Mendelian genetic diseases.

Table 1.1: Different classes of evidence used in the interpretation of variants and assessment of the pros and cons of each type.

In 2015, standards and guidelines for the interpretation of sequence variants were produced jointly by the American College of Medical Genetics (ACMG) and the Association of Molecular Pathology⁷¹ (described in detail in Chapter 3). These guidelines sought to provide a framework for variant interpretation, such that disease-specific approaches could be developed with standardised methods and terminology. By defining how different lines of evidence should be used and importantly the level of evidence required to denote a variant as actionable or disease-causing, they also aspired to introduce a stringent approach to variant analysis that would be more suitable for clinical genetics applications. While these guidelines have been widely adopted, they have highlighted how domain-specific knowledge is critical for the interpretation of variants in particular genes and diseases, and underlined the urgent need to adapt these guidelines to the distinct characteristics of each disease and gene. Such approaches will be required to develop a better balance between sensitivity and specificity than has been obtained thus far for clinical genetic testing in diseases like HCM.

1.5 Aims of thesis

1. To investigate the genetic architecture of HCM and DCM by:
 - a. Comparing rare variation between large clinical cohorts and ExAC reference samples.
 - b. Assessing evidence for involvement of non-sarcomeric genes in HCM and developing a framework to evaluate this evidence.
2. To improve the accuracy and sensitivity of sarcomeric variant interpretation in HCM and develop new approaches to variant analysis.
3. To utilise a large cohort of HCM cases screened with cardiac magnetic resonance imaging to explore genotype-phenotype correlations and enhance our understanding of genotype-negative HCM.

2 Defining cardiomyopathy genes

In this chapter, I undertake a comprehensive re-evaluation of the genes that have been implicated in cardiomyopathies, with particular emphasis on HCM. Taking advantage of the recent availability of large population and cardiomyopathy genetic datasets, I aim to use statistically robust approaches to identify genes that are enriched in cases and are therefore likely to yield informative and interpretable data when clinically tested. I intend to demonstrate how candidate gene studies and recent broad gene panel analyses of cardiomyopathies are insufficiently controlled for rare benign variation in the population and can lead to erroneous conclusions about the genetic architecture of these diseases. I also perform an in-depth reassessment of the published evidence associating genes with HCM and, integrating these findings with the rare variation burden analysis, define the current list of valid genes that should be included in clinical genetic testing panels, as well proposing a classification system to evaluate such evidence for other Mendelian diseases.

2.1 Introduction

Since the first gene responsible for HCM, the beta myosin heavy chain encoding *MYH7* gene, was discovered in 1989⁷², substantial efforts have gone into unravelling the genetics of cardiomyopathies. Clinical genetic testing is now well established for both HCM and DCM and has become an important part of the clinical management of patients with these conditions, particularly for identifying at risk relatives. After the genetic linkage studies in the 1990s associated 8 sarcomeric genes with HCM (Table 2.1), the first large cohort studies that used Sanger sequencing to detect variants in these genes found that approximately 40-50% of cases had causative variants^{73–75} (though not all of these variants are likely to still be regarded as pathogenic with our improved understanding of population variation). Subsequently, much research has focused on identifying additional genetic factors that could explain the apparent missing heritability in HCM cases.

Gene	Gene name	First report	Year
MYH7	myosin, heavy chain 7, cardiac muscle, beta	Jarcho et al. ⁷²	1989
MYBPC3	myosin binding protein C, cardiac	Solomon et al. ⁷⁶	1990
TNNT2	troponin T type 2 (cardiac)	Thierfelder et al. ⁷⁷	1994
TPM1	tropomyosin 1 (alpha)	Thierfelder et al. ⁷⁷	1994
TNNI3	troponin I type 3 (cardiac)	Kimura et al. ⁷⁸	1997
MYL3	myosin, light chain 3, alkali; ventricular, skeletal, slow	Poetter et al. ⁷⁹	1996
MYL2	myosin, light chain 2, regulatory, cardiac, slow	Poetter et al. ⁷⁹	1996
ACTC1	actin, alpha, cardiac muscle 1	Mogensen et al. ⁸⁰	1999

Table 2.1: Details of the studies associating variants in sarcomeric genes with HCM between 1989 and 1999.

2.1.1 Genes implicated in cardiomyopathies

Recent years have seen an exponential rise in the number of genes implicated in both HCM and DCM, driven by improvements in sequencing technology and in particular the development of NGS from 2007 onwards (Figure 2.1). Data from HGMD shows that as of the end of 2014, over 50 genes had been implicated in HCM and over 60 genes in DCM. However, given the study designs adopted for many of these reports (described below), doubts remain as to the strength of the evidence produced by these studies as well as the relevance the phenotypes analysed in certain studies have to standard clinical genetic testing in these conditions.

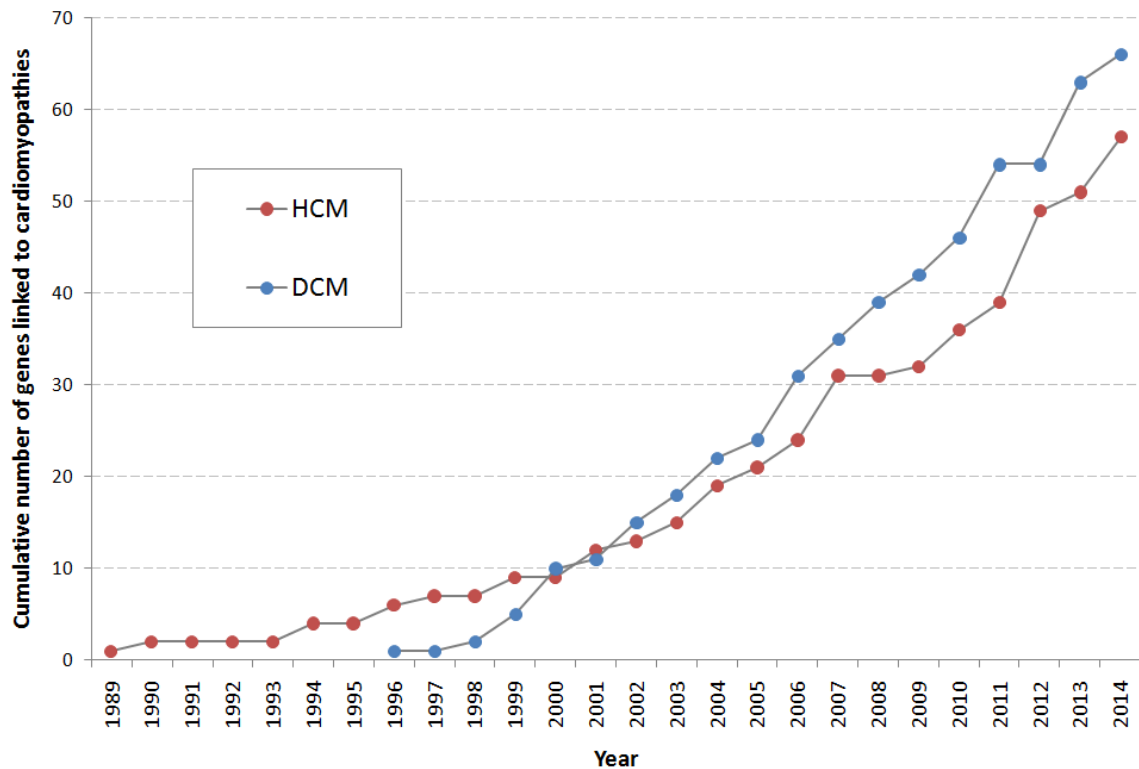


Figure 2.1: Cumulative increase in the number of genes with variants implicated in HCM and DCM from 1989 to 2014. Data from the HGMD, professional version 2015.3.

2.1.2 Methods to associate genes with Mendelian disease

2.1.2.1 Linkage and segregation studies

One of the most powerful ways to associate a gene to a Mendelian disease is to identify variants within that gene that fully segregate with the disease in large family pedigrees. As cardiomyopathies can display incomplete penetrance and late onset of disease, complete segregation is usually demonstrated by all affected members carrying the variant - individuals who are currently phenotype negative do not necessarily have to be non-carriers (though it is expected that a majority would be). Segregation analysis is most effective when performed in an unbiased, genome-wide manner, i.e. without pre-selecting a gene or region of interest. Before the era of NGS, this was accomplished through linkage studies identifying genetic markers that segregated with phenotype, which could eventually be narrowed down to a gene of interest, with the causative variant identified by Sanger sequencing. Such approaches

were used to associate the core sarcomeric genes with HCM. Today, whole exome and genome sequencing have become feasible options for the study of affected families that lack a pathogenic variant in a known gene, allowing causative variants that segregate with disease to be directly detected from unbiased scans of all genes. A limitation of this approach is that any identified variant may not in fact be the causative variant but rather be in linkage disequilibrium with the true pathogenic variant in another gene – demonstration of segregation with different variants in additional pedigrees can be used to unambiguously confirm the gene's role in disease.

2.1.2.2 Candidate gene studies

An alternative approach to discovering causative genes for rare Mendelian diseases is the candidate gene study. This has become more prevalent due to the increasing accessibility and lower costs of direct DNA sequencing technologies and the fact that it is not dependent on having large and clinically well-defined family pedigrees. These studies involve sequencing the gene of interest in a large cohort of probands to identify putative causative variants. The validity of the studies depends on the strength of the statistical tests applied to the variant data, how the study controls for benign variation and the amount of additional evidence gathered to support any association. If appropriately designed, candidate gene studies can be a powerful approach to identifying novel gene to disease associations, e.g. defining the role of truncating variants in *TTN* in DCM⁸¹.

Unlike the unbiased approach of linkage studies, candidate gene studies involve pre-selecting a plausible novel disease gene and applying various analyses to assess its role in disease. For cardiomyopathies, genes can be selected based on a number of factors – the known cardiac function of the gene that makes it a credible candidate, interaction with or similar function to established disease genes, or prior implication in other cardiac conditions. Once the gene has been selected for study, a number of experimental approaches can be adopted to assess its putative role in the disease. The success of the study, and the likely validity of the findings, depends on how many of these are used and how rigorously they are applied.

1. *Rarity of detected variants.* Variants detected in cases need to be absent in control groups or sufficiently rare in the population to be plausible candidates for rare and dominant genetic diseases. This has typically been assessed by confirming the absence of the variant in small control groups (usually less than 500) or below defined thresholds in early population databases (below). As we will see however, these approaches often failed to define rarity with sufficient stringency for pathogenicity.
2. *Statistical enrichment of variants in cases over controls.* One of the strongest methods of validating new gene to disease associations is to demonstrate a statistically significant excess of variation in cases cohorts compared to control or population cohorts, e.g. the *TTN* DCM study referenced above⁸¹. This can be assessed for distinct individual variants or, more commonly, all variants (or variants of a particular class) across a gene. The importance of this approach is that it adequately controls for rare, benign variation in the general population. Until now however, this method has only rarely been applied to candidate gene studies for cardiomyopathies, likely due to the cost of fully sequencing sizeable control cohorts in addition to cases or the general underappreciation until recently of the level of rare benign variation in the population.
3. *Genetic evidence in cases with variants.* For patients in whom variants in the novel gene were detected, supporting evidence can be obtained by examining the inheritance of the variant in the patient's family. If the variant fully segregates with disease in the family, the maximum logarithm of odds (LOD) score can be estimated – LOD scores greater than 2-3 (corresponding to at least 7 meioses between affected carriers) are generally regarded as providing statistically robust evidence of association between the variant and disease. However, such large pedigrees are rarely available for individual cases in such cohorts. Alternatively, demonstrating *de novo* inheritance of a variant (with confirmed maternity and paternity) also provides strong evidence of causality.
4. *Effect in functional assays.* The effect of a detected variant in functional assays can provide additional evidence of its role in disease. Such assays can include *in vitro* assays in cells, cell lines or tissues or *in vivo* animal models of disease. However, it is critical that any

functional assays used to support such associations are validated for the particular disease, with demonstrated and consistent effects with both known pathogenic and benign variants. In the absence of this (and such validation has been rarely applied to cardiomyopathies), it is questionable how well any observed effect in the laboratory would translate to the patient in the clinic. Therefore, considerable care should be taken to avoid relying on functional data to validate any novel gene to disease associations.

2.1.3 Impact on clinical genetic testing in cardiomyopathies

Many of the novel genes that have been recently implicated in cardiomyopathies through candidate gene studies have been incorporated into clinical genetics test panels. As we will see (in Methods), the latest DCM panel from the Laboratory of Molecular Medicine (LMM) of Partners Healthcare, Boston, tests 46 genes. In HCM, the current gene panel coverage for 17 major clinical laboratories is shown in Table 2.2, with the panels ranging from 19 to 47 genes. The impact of expanding test panels to include these novel gene associations has been assessed in recent publications from LMM, in which they describe their experience of genetic testing over 10 years for HCM³³ and 5.5 years for DCM⁸².

In DCM, the yield of genetic testing (i.e. cases with pathogenic or likely pathogenic variants) increased from 7.7% to 27% as the panel increased from 5 to 46 genes (Figure 2.2), largely driven by truncating variants in *TTN*. However, increasing panel sizes produced an even greater increase in the proportion of variants of uncertain significance (VUS) (6.5% to 61%), highlighting the challenges in interpreting variants in many of these novel genes (only 15 genes yielded actionable variants). Similarly, expanding HCM testing beyond the original 11 gene test (8 sarcomeric and 3 metabolic cardiomyopathy genes) to a larger 18 gene HCM NGS panel or a 46-51 gene pan-cardiomyopathy panel yielded only a single additional pathogenic variant (in the *PLN* gene). These findings highlight the current limited utility of expanding cardiomyopathy test panels to include novel associations from candidate gene studies, suggesting that variants in these genes are either very rarely causative, difficult to interpret and classify as actionable or have simply been wrongly associated with these conditions.

Gene	Amry Genetics	Blueprint Genetics	CeGaT GmbH	CENAGEN	CGC Genetics	Children's Hospital Eastern Ontario	Royal Brompton Hospital	DDC Clinic	EGL Genetic Diagnostics	GeneDX	Washington University	Invitae	La boratory of Molecular Medicine	Mayo Clinic	Prevention Genetics	Reference Laboratory Genetics	Sherbrooke Genomic Medicine
ACTA1																	
ACTC1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ACTN2	X	X	X														
AGL																	
ALPK3		X															
ANK2																	
ANKRD1	X		X														
BAG3																	
BRAF																	
CACNA1C																	
CALR3			X														
CASQ2																	
CAV3			X	X	X	X	X	X	X								
COX15																	
CRYAB																	
CSRP3	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
DES																	
FHL1			X														
FLNC																	
FXN	X																
GAA		X															
GBE1																	
GLA	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HRAS																	
JPH2	X	X	X	X													
KLF10																	
KRAS																	
LAMP2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LDB3			X														
MAP2K1																	
MAP2K2																	
MYBPC3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MYH6	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MYH7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MYL2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MYL3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MYLK2			X		X		X	X					X	X	X	X	X
MYO6																	
MYOZ2	X		X	X	X	X	X	X	X	X		X	X	X	X	X	X
MYPN	X		X	X			X	X	X								
NEXN	X		X		X	X	X	X									
NRAS																	
PDLIM3																	
PLEC																	
PLN	X		X		X	X	X	X	X	X	X	X	X	X	X	X	X
PRKAG2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
PRKAR1A																	
PTPN11	X																
RAF1	X	X															
RIT1																	
RYS2																	
SCO2																	
SHOC2																	
SLC22A5				X													
SLC25A4			X														
SOS1		X															
TCAP	X		X		X		X	X	X	X	X		X	X	X	X	X
TNNC1	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
TNNI3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
TNNT2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
TPM1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
TRIM63																	
TTN			X	X	X		X		X				X	X	X	X	X
TTR	X	X	X	X		X			X	X	X	X	X	X	X	X	X
VCL	X		X		X		X	X	X	X	X	X	X	X	X	X	X
Total	27	19	31	20	22	19	47	23	20	21	31	26	20	26	38	40	26

Table 2.2: HCM gene test coverage from 17 clinical genetics laboratories and diagnostic companies as of July 2017. Data from genetests.org and websites of ClinVar submitters with over 10,000 submissions.

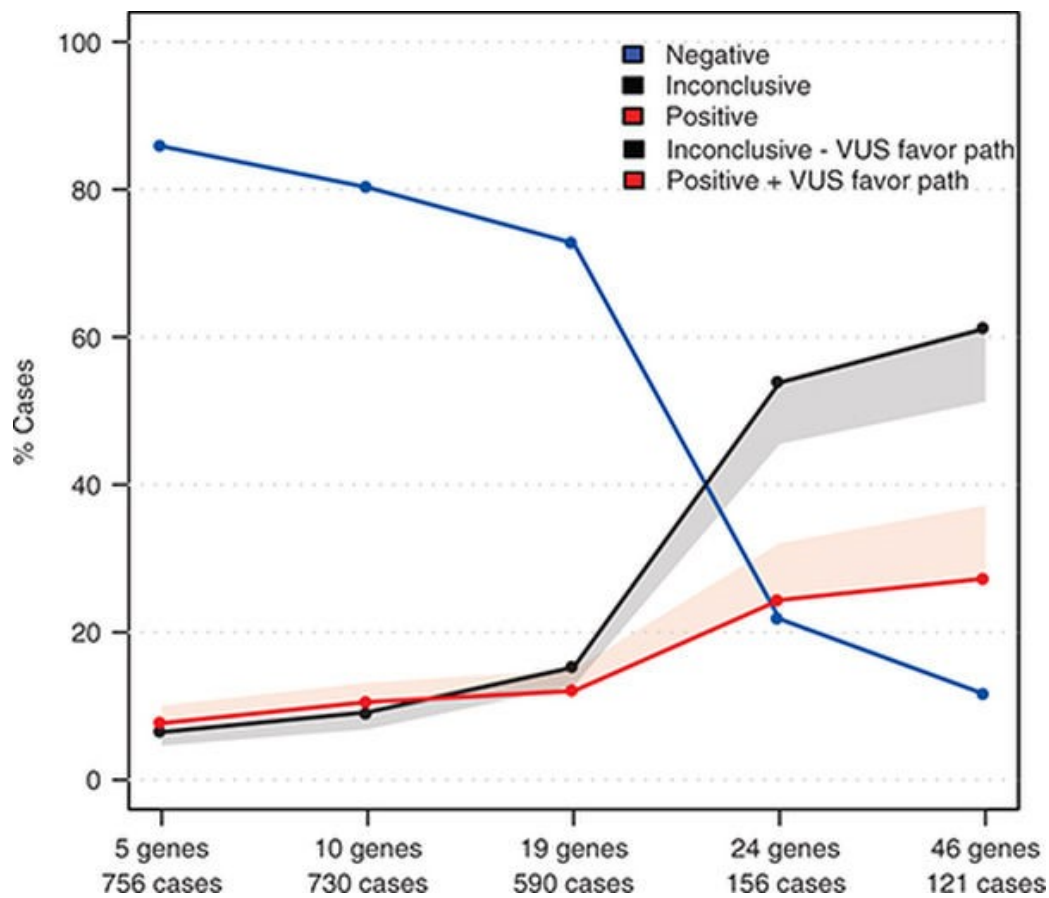


Figure 2.2: Outcome of clinical testing for DCM by LMM from 2007 to 2011 as the test panel increased from 5 to 46 genes. This highlights the modest increase in positive results (pathogenic and likely pathogenic) but a substantial increase in inconclusive results (VUS). The red shaded area denotes the upper bound of clinical sensitivity by including variants classified as “VUS favour pathogenic”. The grey shaded area denotes the lower bound of inconclusive cases by excluding variants classified as “VUS favour pathogenic”. Figure reproduced from Pugh et al⁸² with permission.

2.1.4 New insights from population databases

Projects involving WGS or WES of large control or common disease cohorts have offered significant new insights into the frequency of rare variation in the general population. The 1000 Genomes (1KG) Project ran between 2008 and 2015, sequencing 2,504 samples from a diverse range of population groups⁸³. It was the first project to sequence genomes at such a scale and provided a key resource to researchers investigating human genetic variation. Following on from 1KG, the Exome Sequencing Project (ESP) sequenced the exomes of 6,500 samples (with a 2:1 ratio of European-American and African-American individuals) from various well-phenotyped cohorts in the United States (<http://evs.gs.washington.edu/EVS/>).

The release of these datasets allowed researchers to re-evaluate variants that had been previously associated with diseases like cardiomyopathies. With the phase 1 of the 1KG release, Golbus et al. identified variants in *MYBPC3* and *MYH7* at a far higher frequency than would be expected based on the known prevalence of HCM and DCM⁸⁴. Similarly, in a study of the ESP database, Andreassen et al. discovered that 14% - 18% of variants that had been associated with HCM, DCM and ARVC were present in this population⁸⁵. If these associations were correct, it would correspond to a prevalence of 1:4 for HCM, 1:6 for DCM, and 1:5 for ARVC, far greater than the estimated incidence of 1:500 for HCM and DCM and 1:5000 for ARVC. Studies like these have highlighted that many variants in the research literature and databases like HGMD have been erroneously implicated in Mendelian diseases and more careful consideration of the population frequency of a variant is required before assigning pathogenicity. By underlining that the frequency of rare benign variation in the population is likely far greater than had previously been appreciated, they also imply that many of the novel gene to disease associations may have been based on inadequately controlled studies.

2.1.5 Exome Aggregation Consortium (ExAC)

The release of the Exome Aggregation Consortium or ExAC database in 2014 offered the clearest view yet of rare variation in the human exome⁶⁸. With over 60,000 samples from a series of common disease and population datasets (including subsets of 1KG and ESP), ExAC is the first such dataset powered to assess allele frequencies in the range of 1:1000 – 100,000.

Importantly, this resolution allows us to assess if the frequency of individual variants in the general population are consistent with penetrant Mendelian diseases of the prevalence of HCM and DCM (see Methods for how to define these frequency thresholds). It also allows us to accurately assess the level of rare benign variation across genes, even for small and constrained genes for which data in previous population datasets was limited. The insights into rare human variation afforded by ExAC thus enables us to reassess published claims of pathogenicity, both for individual variants and across genes.

2.1.6 Importance of validating the role of genes in cardiomyopathies

Clarification of which genes that have previously been implicated in cardiomyopathies are valid disease genes, and which are likely to be erroneous, is critical for a number of reasons. As described above, these associations can be used to inform the composition of the gene panels used in clinical genetic testing. Many of the more recent associations in HCM and DCM have already been shown to have limited utility for clinical testing – retaining incorrectly implicated genes in these panels increases the risk of false positive results and the uncertainty associated with growing numbers of VUS in clinical reports, while providing little or no chance of yielding actionable findings.

A critical appraisal of these gene to disease associations is also essential for increasing our understanding of the genetic architecture and aetiology of cardiomyopathies. Some recent studies have used sequencing of broad panels of cardiac genes in large HCM and DCM cohorts to explore the genetic variants underlying these diseases – 874 HCM patients were sequenced on a 41 gene panel⁸⁶ and 639 DCM patients were sequenced on an 84 gene panel⁸⁷. The variant analysis strategies employed by these studies included filtering by population frequencies (though at insufficiently stringent thresholds), annotation with disease associations in the literature (including the many invalid reports as described above) and generic missense functional prediction algorithms, with all genes analysed in a similar manner regardless of the strength of their association with HCM/DCM. Despite the sizable case cohorts involved, these studies did not attempt to assess any enrichment of rare variants in the genes tested in order to validate their role in these conditions.

As a result of implicating variants in numerous non-validated disease genes, these studies have made a number of erroneous conclusions about the genetics of these diseases. As well as incorrect inferences about individual gene to disease associations, they overestimate the genetic overlap between different inherited cardiac conditions and advocate multi-genic models for cardiomyopathies without properly controlling for rare variation in controls or the general population. Such assumptions can only serve to obscure our understanding of the genetic aetiology of cardiomyopathies, with potentially serious consequences for how genetic testing in these diseases is applied and interpreted clinically.

The availability of population genetic datasets of the scale of ExAC, along with increasingly large cardiomyopathy case cohorts, now affords us the opportunity to reassess the genetics of cardiomyopathies and in particular the pathogenicity of genes that have been implicated in these conditions. By focusing on variation that is rare enough in the population to be plausibly pathogenic, re-evaluating published evidence and applying statistically rigorous approaches to this analysis, we can define the genes and variant classes that cause these diseases and provide clarity to both our understanding of the genetics of these conditions and the application of clinical genetic testing in their management.

2.2 Materials and Methods

2.2.1 Sequencing cohorts

The cohorts used in this study to assess the role of genes implicated in cardiomyopathies are summarised in Table 2.3 and described in detail below. For details of the number of cases sequenced in each cohort for each gene, see Table 2.4. There are two major parts to the analysis in this chapter:

1. Burden testing of **clinical** HCM and DCM cohorts versus ExAC.
2. Evaluation of the evidence associating non-sarcomeric genes with HCM, including burden testing of **combined** case cohorts (clinical, prospective research and published data) versus ExAC.

Cohort	Cohort Type	Ref.	Sample size	Genes tested
<i>Hypertrophic Cardiomyopathy (HCM)</i>				
Oxford Molecular Genetics Laboratory	Clinical	⁸⁸	807 - 3267	16
Laboratory of Molecular Medicine	Clinical	^{33,88}	632 - 2912	18
Combined clinical	Clinical		632 - 6179	20
HCM prospective cohort	Research (prospective)	⁸⁹	804	35
HCM research (Lopes et al.)	Research (published)	⁸⁶	874	36
Combined HCM cohorts	All	⁸⁹	-	35
<i>Dilated Cardiomyopathy (DCM)</i>				
Oxford Molecular Genetics Laboratory	Clinical	⁸⁸	304 - 559	28
Laboratory of Molecular Medicine	Clinical	^{82,88}	121 - 756	46
Combined clinical	Clinical		121 - 1315	48
DCM research (Hershberger et al.)	Research (published)	⁹⁰⁻⁹²	312 - 324	12
<i>Population reference</i>				
Exome Aggregation Consortium (ExAC)	Reference database	⁶⁸	60706	All

Table 2.3: Summary of the sample cohorts used in this study for gene burden testing. For burden testing, the initial HCM and DCM analysis was performed with the “Combined Clinical” cohorts. The subsequent HCM analysis utilised the “Combined HCM cohorts” (clinical, published research and prospective research).

Gene	Ensembl transcript	Refseq transcript	HCM					DCM		
			Clinical		Research		Published Panel	Clinical		Published Panel
			OMGL	LMM	TruSight	SOLiD		OMGL	LMM	
ABCC9	ENST00000261201	NM_005691							590	
ACTA1	ENST00000366684	NM_001100			405					
ACTC1	ENST00000290378	NM_005159	1535	2650	405	399	874	355	748	
ACTN2	ENST00000366578	NM_001103	807	632	405	399		305	590	
ANK2	ENST00000264366	-					874			
ANKRD1	ENST00000371697	NM_014391	807		405	399		305	121	
CACNA1C	ENST00000399655	NM_000719			405	399				
CALR3	ENST00000269881	NM_145046			405	399				
CASQ2	ENST00000261448	NM_001232			405	399	874		121	
CAV3	ENST00000343849	NM_033337			405	399	874		121	
COX15	ENST00000370483	NM_004376			405					
CRYAB	ENST00000526180	NM_001885			405	399		304	121	
CSRP3	ENST00000533783	NM_003476	1535	632	405	399	874	355	590	313
CTF1	ENST00000279804	NM_001330							590	
DES	ENST00000373960	NM_001927					874	304	590	
DSC2	ENST00000280904	NM_024422					874	304	123	
DSG2	ENST00000261590	NM_001943					874	304	123	
DSP	ENST00000379802	NM_004415					874	304	123	
DTNA	ENST00000348997	NM_032978							121	
EMD	ENST00000369842	NM_000117							590	
FHL1	ENST00000370690	NM_001449	1535		405			355		
FHL2	ENST00000358129	NM_201555				399		304	121	
FXN	ENST00000377270	NM_000144			405	399				
GAA	ENST00000302262	NM_000152			405	399				
GLA	ENST00000218516	NM_000169	1535	2165	405	399		355	186	
JPH2	ENST00000372980	NM_020433			405					
JUP	ENST00000393930	NM_002230					874	304	121	
KCNE1	ENST00000399289	NM_000219					874			
KCNE2	ENST00000290310	NM_172201					874			
KCNH2	ENST00000262186	NM_000238					874			
KCNJ2	ENST00000535240	NM_000891					874			
KCNQ1	ENST00000155840	NM_000218			405	399	874			
KLF10	ENST00000285407	NM_005655			405					
LAMA4	ENST00000424408	NM_002290							121	
LAMP2	ENST00000200639	NM_002294	839	2451	405	399		309	223	

Table 2.4: Details of the transcripts analysed for all of the genes evaluated in this study, as well as the number of samples sequenced in the clinical (OMGL and LMM), prospective research (sequenced on Illumina with TruSight Cardio panel or SOLiD with custom Agilent panel) and published cardiomyopathy cohorts (HCM – Lopes et al.⁸⁶, DCM – Hershberger et al.^{90–92}). See text for full details of these cohorts. Table continued on the next page.

Gene	Ensembl transcript	Refseq transcript	HCM					DCM		
			Clinical		Research		Published Panel	Clinical		Published Panel
			OMGL	LMM	TruSight	SOLiD		OMGL	LMM	
LDB3	ENST00000361373	NM_007078			405		874		740	313
LMNA	ENST00000368300	NM_170707			405	399	874	304	740	324
MYBPC3	ENST00000545968	NM_000256	3267	2912	405	399	874	405	756	312
MYH6	ENST00000356287	NM_002471			405	399	874		121	312
MYH7	ENST00000355349	NM_000257	3200	2912	405	399	874	559	756	313
MYL2	ENST00000228841	NM_000432	1535	2650	405	399	874	355	188	
MYL3	ENST00000395869	NM_000258	1535	2650	405	399	874	355	188	
MYLK2	ENST00000375994	NM_033118			405	399			121	
MYO6	ENST00000369977	NM_004999			405					
MYOZ2	ENST00000307128	NM_016599		632	405	399			121	
MYPN	ENST00000358913	NM_032578			405	399				
NEXN	ENST00000334785	NM_144573		632	405	399			156	
PDLIM3	ENST00000284770	NM_014476			405	399	874			
PKP2	ENST00000070846	NM_004572					874	304	123	
PLN	ENST00000357525	NM_002667	1535	632	405	399	874	355	740	
PRKAG2	ENST00000287878	NM_016203	1535	2438	405	399		355	191	
RBM20	ENST00000369519	NM_001134363					874		156	
RYR2	ENST00000366574	NM_001035					874		121	
SCN5A	ENST00000333535	NM_198056					874	304		313
SGCD	ENST00000337851	NM_000337							590	
SLC25A4	ENST00000281456	NM_001151			405					
TAZ	ENST00000299328	NM_000116							740	
TCAP	ENST00000309889	NM_003673			405	399	874		590	313
TGFB3	ENST00000238682	NM_003239					874			
TMEM43	ENST00000306077	NM_024334					874	304	123	
TNNC1	ENST00000232975	NM_003280		632	405	399	874		156	312
TNNI3	ENST00000344887	NM_000363	3135	2912	405	399	874	483	756	312
TNNT2	ENST00000367318	NM_001001430	3191	2912	405	399	874	498	756	313
TPM1	ENST00000403994	NM_001018005	1535	2912	405	399	874	355	756	312
TRIM54	ENST00000296098	NM_032546				399				
TRIM55	ENST00000315962	NM_184085				399				
TRIM63	ENST00000374272	NM_032588			405	399				
TTN	ENST00000589042	NM_001267550						304	156	
TTR	ENST00000237014	NM_000371		632	405	399			121	
VCL	ENST00000211998	NM_014000			405	399	874		590	

2.2.1.1 Exome Aggregation Consortium (ExAC)

The Exome Aggregation Consortium (ExAC) is a database of whole exome sequences from 60,706 individuals, drawn from a series of common disease and population genetic studies⁶⁸ – see Table 2.5 for the details of each contributing study. Developed at the Broad Institute, ExAC takes the raw data from each component sequencing study and reprocesses it through the same pipeline, with variants identified using GATK Haplotype Caller.

Consortium/Cohort	Samples
1000 Genomes	1,851
Bulgarian Trios	461
GoT2D	2,502
Inflammatory Bowel Disease	1,675
Myocardial Infarction Genetics Consortium	14,622
NHLBI-GO Exome Sequencing Project (ESP)	3,936
National Institute of Mental Health (NIMH) Controls	364
SIGMA-T2D	3,845
Sequencing in Suomi (SISu)	948
Swedish Schizophrenia & Bipolar Studies	12,119
T2D-GENES	8,980
Schizophrenia Trios from Taiwan	1,505
The Cancer Genome Atlas (TCGA)	7,601
Tourette Syndrome Association International Consortium for Genomics (TSAICG)	297
Total	60,706

Table 2.5: Cohorts that comprise the ExAC database, including common disease and population genetic studies. Data from <http://exac.broadinstitute.org/faq>.

A comparison of the size and ethnic composition between ExAC and other population genetic databases is shown in Figure 2.3 – 1KG comprising 2,504 samples and ESP comprising 6,500 samples.

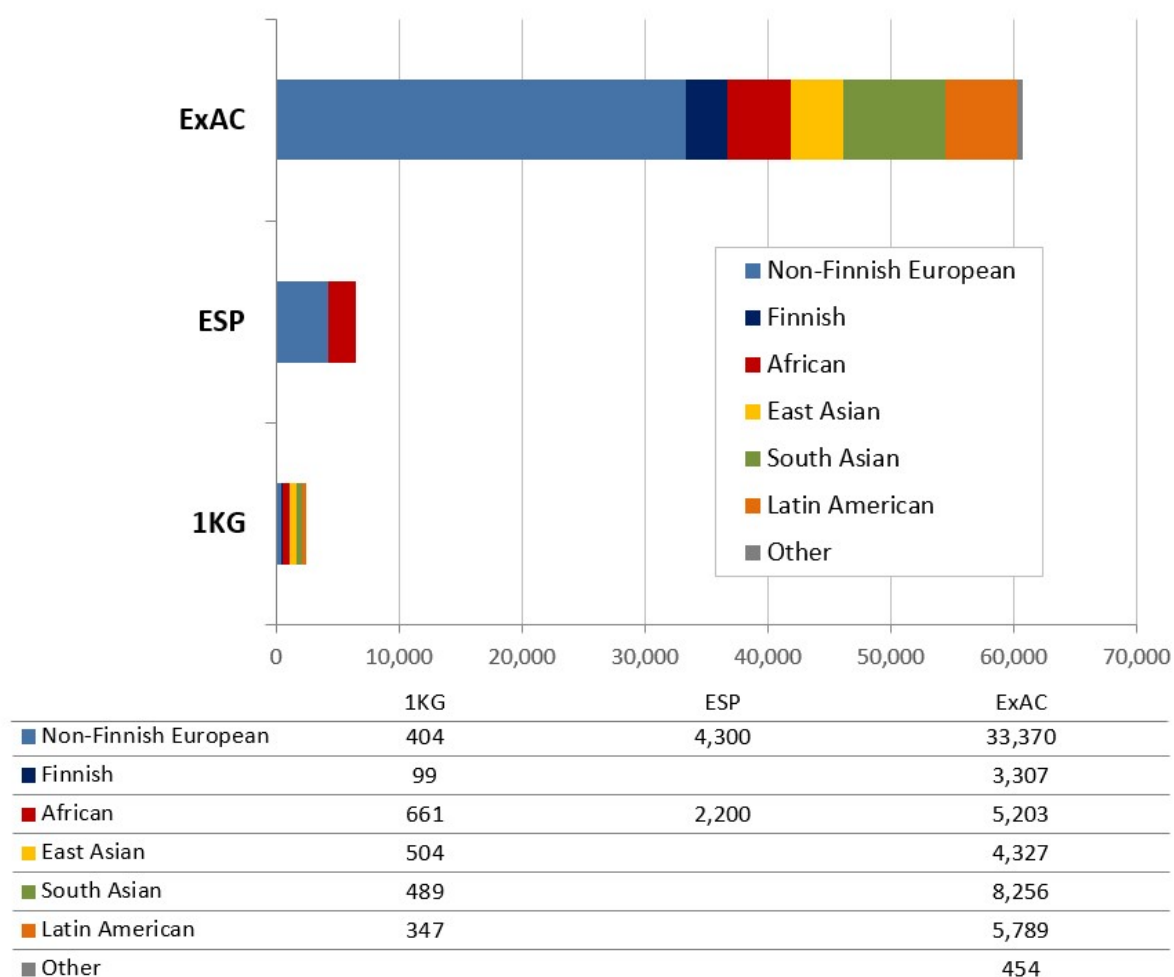


Figure 2.3: Number of individuals in the ExAC database per major ethnic group. Samples were clustered by principal component analysis, with “Other” groups consisting of samples that did not unambiguously cluster with major groups. The size of the Exome Sequencing Project (ESP) and 1000 Genomes (1KG) are shown for comparison. There is partial overlap between ExAC and 1KG (1,851 of the 2,504 samples in 1KG) and ESP (3,936 of the 6,500 samples in ESP).

2.2.1.2 Oxford Molecular Genetics Laboratory (OMGL)

The OMGL cohort comprises apparently unrelated index cases referred from Clinical Genetics centres across the UK, with diagnostic genetic testing undertaken in a routine clinical setting using clinical consent. 3267 individuals with HCM and 559 individuals with DCM were sequenced - the initial clinical diagnosis was made by a consultant cardiologist. The gene panels for both conditions have expanded over time such that not all patients were analysed for every gene. Up to 16 genes for HCM and 28 genes for DCM were sequenced for this cohort (see Table 2.4 for the exact number of cases for each gene). The variants in this cohort were classified according to national guidelines (<http://www.acgs.uk.com>) into three categories - highly likely to be pathogenic (Class 5), likely to be pathogenic (Class 4) and variant of unknown significance (VUS) (Class 3). Data on patient ethnicity was not available for this cohort but it is expected to be broadly representative of the UK population, i.e. predominantly Caucasian.

2.2.1.3 Laboratory of Molecular Medicine, Partners Healthcare, Boston (LMM)

The data from the LMM clinical cohort was downloaded from published studies on HCM³³ and DCM⁸² clinical genetic testing in this laboratory. The HCM cohort comprised 2912 unrelated probands referred for HCM clinical genetic testing between 2004 and 2014 and excludes 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 HCM³³. As with the OMGL cohort, the LMM HCM gene panel has expanded over this period from 5 to 18 genes.

The DCM cohort comprised 756 individual probands referred for DCM clinical genetic testing across a period of 5.5 years - all patients had DCM or clinical features consistent with DCM based on the medical and family history information. The cohort excludes any cases with confirmed diagnoses of other cardiomyopathies, structural heart disease, congenital heart disease or syndromic or environmental causes. Between 5 and 46 genes were sequenced in each patient (see Table 2.4 for the exact number of HCM and DCM cases analysed for each gene). Variants in these studies were classified as pathogenic, likely pathogenic, VUS or likely

benign according to the LMM's clinical grade variant classification criteria (the VUS were sub-classified into VUS favour pathogenic, VUS and VUS favour benign). Data on patient ethnicity was not available for the HCM cohort but it is expected to be broadly representative of the US population, i.e. majority Caucasian. The ethnicities of the DCM cohort were classified as: 464 white, 86 black or African-American, 43 Hispanic or Latino, 22 Asian, 7 Ashkenazi Jews, 22 mixed ancestry and 122 unspecified.

2.2.1.4 Clinical cohort data

As OMGL and LMM use similar clinical guidelines approaches for variant classification, it is expected that there will be broad though not total concordance between the two laboratories in the classification of specific variants⁹³. To standardise the different terms employed by OMGL and LMM, the following classifications were used in this study - Pathogenic (P) (OMGL Class 5 and LMM Pathogenic), Likely Pathogenic (LP) (OMGL Class 4 and LMM likely pathogenic) and VUS (OMGL Class 3 and LMM VUS, VUS favour pathogenic and VUS favour benign).

In these clinical cohorts, sequence data was generated using a range of mutation scanning and direct sequencing techniques of varying sensitivity - high-resolution DNA melting, dHPLC, DNA microarray, Sanger sequencing and targeted NGS. These targeted tests are designed to cover the coding regions and splice sites of the key genes of interest, with analytical sensitivity estimated to be in the region of 98-100%. Clinical genetic testing requires all putative pathogenic variants to be confirmed by Sanger sequencing, therefore the rate of false positive variant calls is expected to be negligible.

2.2.1.5 Published research studies

Data from two published cardiomyopathy genetic studies were downloaded to assess how such research studies interpret variants. The HCM research cohort, published in 2014, sequenced 874 patients across 35 genes⁸⁶ – 12 associated primarily with HCM (*ACTC1*, *CSRP3*, *MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *PLN*, *TCAP*, *TNNC1*, *TNNI3*, *TNNT2*, *TPM1*), 7 with DCM (*DES*, *LDB3*, *LMNA*, *MYH6*, *PDLIM3*, *RBM20*, *VCL*), 7 with ARVC (*DSC2*, *DSG2*, *DSP*, *JUP*, *PKP2*, *TGFB3*, *TMEM43*), and 9 with arrhythmias (*ANK2*, *CASQ2*, *KCNE1*, *KCNE2*, *KCNH2*, *KCNJ2*, *KCNQ1*, *RYR2*, *SCN5A*), as stated by Lopes et al. The DCM research cohort, published between 2008 and 2010, comprised 312-324 patients sequenced for 12 confirmed and putative DCM genes (though not including *TTN*)^{90–92}. In both of these studies, putative pathogenic variants were not identified by clinical-grade classification (such as ACMG guidelines) but rather by utilising information such as variant type, population frequency, and *in silico* functional algorithms (the details of the criteria used in each study are described in the published articles). As such, these studies are expected to be less stringent at assigning pathogenicity than clinical laboratories, particularly as they sequence broad panels of cardiac genes, regardless of whether these genes have been validated for the condition being studied.

2.2.1.6 Prospective HCM cohorts

Two prospective HCM cohorts were used to examine rare genetic variation across broad panels of cardiac genes. 684, mainly Caucasian, patients were recruited by the Cardiovascular Biomedical Research Unit at the Royal Brompton Hospital, London. 120, mainly Chinese, patients were recruited at the National Heart Centre, Singapore. All patients were confirmed with a diagnosis of HCM with reference to established cardiac MRI (UK) and/or echocardiographic (Singapore) diagnostic criteria. Of the 804 HCM probands analysed as part of this study, 405 samples were sequenced using the Illumina TruSight Cardio Sequencing Kit⁶³ (which includes 174 genes associated with ICCs) on the Illumina MiSeq and NextSeq platforms and 399 samples were sequenced using a custom Agilent SureSelect panel of genes associated with ICCs on the Life Technologies SOLiD 5500xl platform.

2.2.2 Next generation sequencing of prospective HCM cohort

Targeted DNA libraries were prepared according to manufacturers' protocols before performing paired end sequencing. For the Illumina sequenced samples, demultiplexing of sequence data were performed with MiSeq/NextSeq Control software or Bcl2FastQ conversion and the FastQ files were subjected to quality control with the FastQC v.0.10.1. After low quality reads (<20) were trimmed using PrinSeq v0.20.4, reads were aligned to the HG19 reference genome using BWA v0.7.10⁹⁴. Picard v1.115 and GATK v3.2-2⁶⁶ were used to mark duplicate reads and perform local realignment around indels and base quality score recalibration. Bases covered by at least 10 reads with a mapping quality ≥ 10 and base quality ≥ 20 were denoted as "callable", i.e. adequately covered for variant calling with recommended GATK parameters. Variant calling was performed with GATK HaplotypeCaller and UnifiedGenotyper. Variants called as "PASS" by either HaplotypeCaller or UnifiedGenotyper and with a quality of depth score ≥ 4 and an allelic balance ≥ 0.20 were considered to be valid variants.

For the SOLiD sequenced samples, reads were aligned in colour space using LifeScope™ v2.5.1 "Targeted re-sequencing" pipeline. The SOLiD Accuracy Enhancement Tool (SAET) was used to improve colour call accuracy prior to mapping. Variants were called by the diBayes and SmallIndel packages in LifeScope software as well as GATK UnifiedGenotyper. Bases covered by at least 10 reads with a mapping quality ≥ 10 and base quality ≥ 20 were denoted as "callable". All detected variants from both platforms were functionally annotated using the Ensembl API v75_37⁹⁵ with respect to the Ensembl transcripts defined in Table 2.4.

Based on prior experience in our laboratory with Sanger sequencing confirmation of variants detected on the SOLiD platform, some additional filtering was performed. For variants that were only called by the diBayes/SmallIndel packages, i.e. not independently confirmed by GATK UnifiedGenotyper, those with a coverage ≤ 100 and an allelic balance < 0.20 or a coverage < 50 and an allelic balance < 0.30 were filtered and removed from further analysis, given the likelihood that these were false positive sequencing errors (Figure 2.4).

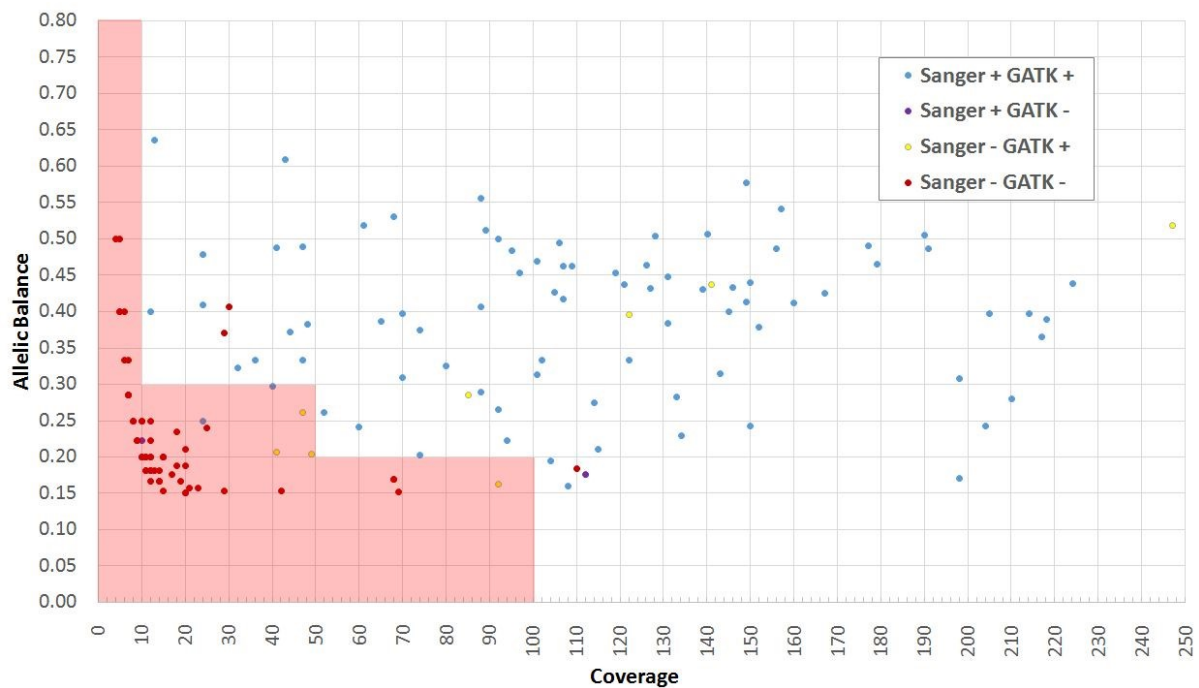


Figure 2.4: Comparison of coverage (x-axis) and allelic balance (y-axis) for variants identified from SOLiD-sequenced samples that were subject to Sanger confirmation. Note these variants are from samples from different cohorts sequenced by the Cardiovascular Genetics and Genomics group at the Royal Brompton Hospital and do not only include the HCM samples described in this study. Variants that were not independently called by GATK UnifiedGenotyper and with a coverage/allelic balance in the shaded sections were considered as poor quality variants that were likely to be sequencing artefacts due to the low probability they would be confirmed by Sanger sequencing.

2.2.3 Defining rare variants

A critical step in burden testing is defining the population frequency threshold (in this case in the ExAC database) for rare variants, i.e. a frequency above which a penetrant Mendelian variant could not be pathogenic. Arbitrary population minor allele frequency (MAF) cut-offs of 0.01 or 0.001 are often applied in published studies, despite the fact that such frequencies are far too common for diseases with a prevalence of 1:500 that exhibit high genetic heterogeneity (as with HCM and DCM). A variant should not be present in a population database more frequently than the disease being tested, or indeed than the proportion of

cases attributable to any single variant. Now, with the large case datasets available in this study, a more accurate threshold can be estimated.

In the combined clinical cohorts, the variant to which the largest proportion of cases is attributable is *MYBPC3*:c.1504C>T (p.R502W), found in 104/6179 or 1.7% (1.4-2.0%, 95% CI for binomial distribution) of HCM cases. The expected population frequency can therefore be calculated by:

$$104/6179 \text{ (case frequency)} \times 1/500 \text{ (disease prevalence)} \times 1/2 \text{ (diploid)} = 1.7 \times 10^{-5}$$

In ExAC, this variant is actually detected in 3/60557 individuals, yielding a MAF of 2.5×10^{-5} with a binomial distribution upper bound of the true population frequency of 5.3×10^{-5} . In the non-Finnish European subset of ExAC (ExAC-NFE), it is detected in 2/33329 individuals, MAF of 6.0×10^{-5} , upper bound of 7.2×10^{-5} .

To account for the variable penetrance observed in cardiomyopathy pathogenic variants, a threshold of 1×10^{-4} was applied for this study. While this is a conservative threshold based on the analysis above, it should be noted that it is still significantly more stringent than most studies published to date. This MAF threshold may not be appropriate for populations where there are more common founder mutations, e.g. *MYBPC3*:c.927-2A>G, present in up to 60% of Icelandic HCM patients⁹⁶, though this is not an issue for this study.

2.2.4 Burden testing

The initial burden testing analysis in this study compared variation in the combined OMGL and LMM clinical cohorts (separately for HCM and DCM) with ExAC. 20 genes sequenced in 632 – 6179 HCM patients and 46 genes sequenced in 121 – 1315 DCM patients were analysed (two of the DCM genes, *RBM20* and *CTF1*, were excluded due to poor coverage in ExAC). Only data from designated canonical transcripts (Table 2.4) was assessed.

For each gene, the overall frequency of rare variants (as defined above) was calculated for the case cohorts and for ExAC – this frequency is the proportion of individuals in the cohort

with a rare variant in the gene or variant class within that gene. For case cohorts, the frequency was calculated by dividing the sum of the rare variants detected by the number of samples sequenced for the particular gene. For ExAC, the frequency was calculated by dividing the sum of the adjusted allele count by the mean of the total adjusted alleles, i.e. the number of alleles for which there was sufficient coverage at that position for a sequencing call to be made. This allows us to adjust for potentially variable and inferior sequencing data from WES samples. Analysis was performed separately for truncating variants and non-truncating variants. Truncating variants were defined as comprising nonsense variants (single nucleotide substitutions that lead to a gain of a stop codon), frameshift variants (insertion or deletion of one or more nucleotides that leads to the disruption of the open reading frame) and variants affecting the splice donor and acceptor sites (first and last two bases of each intron). Non-truncating variants were defined as comprising missense variants (single nucleotide substitutions that lead to codons that code for a different amino acid) and in-frame insertions and deletions (of a multiple of 3 nucleotides in the coding sequence that doesn't lead to a disruption of the open reading frame).

This burden testing relies on a number of assumptions. We have assumed that variants for any given gene do not occur more than once in any individual in both case cohorts and ExAC – we do not have the individual sample data for either the clinical case cohorts or ExAC to confirm this but would expect any multiple hits to be very rare and therefore not affect this analysis. We also assume that the level of rare benign variation in cases and ExAC will be equivalent. Finally, while we expect that there will be cardiomyopathy cases, and therefore pathogenic variants, in ExAC, these should be sufficiently low so as not to affect these comparisons.

The case excess was calculated by subtracting the ExAC frequency from the case frequency. The significance of any observed case excess was tested with a Fisher's exact test (cases with/without and ExAC with/without the variant class being tested). Bonferroni correction was applied to account for multiple testing – $p < 0.0025$ for HCM (20 genes tested) and $p < 0.001$ for DCM (46 genes tested).

2.2.5 Assessment of population-specific effects

To ensure that there were no confounding effects from population-specific variants, the case-control analysis was repeated using specific population groups for the one cohort for which ethnicity data was available – the DCM cohort from LMM. Variant frequencies from self-declared Caucasian DCM samples were analysed against the ExAC-NFE subset, with the results compared to the all sample DCM analysis.

2.2.6 Extended HCM gene analysis – data mining

The HCM gene panels from the OMGL and LMM clinical laboratories were limited to a combined total of 20 genes, less than half of the number of genes implicated in HCM (Figure 2.1). To investigate all other genes associated with HCM, as well as clarifying uncertain results from the clinical sequencing data, a detailed analysis of all sequencing data (published, clinical and prospective cohorts) and published genetic evidence was performed. Genes with variants associated with HCM were identified by interrogating HGMD (professional version 2015.3) for entries with a disease-causing mutation tag (DM or DM?, the latter indicating some doubts about the association) and a list of manually curated disease descriptions related to HCM (see list of terms below). Genes associated with metabolic cardiomyopathies that mimic HCM, syndromic conditions that include cardiac hypertrophy and the eight core sarcomeric HCM genes were identified but not included in further analysis (see Results). Evidence for the remaining 31 genes was then collated.

HGMD disease terms for HCM

Cardiomyopathy, hypertrophic; Cardiomyopathy, hypertrophic with deafness; Cardiomyopathy, hypertrophic, with muscle hypertrophy; Cardiomyopathy, hypertrophic; Hypertrophic cardiomyopathy; Hypertrophic cardiomyopathy, early onset; Hypertrophy, left ventricular; Increased left ventricular wall thickness; Left ventricular hypertrophy; Mitochondrial myopathy & hypertrophic cardiomyopathy; Nemaline myopathy and hypertrophic cardiomyopathy; Perinatal problems & hypertrophic cardiomyopathy; Proximal muscle weakness & hypertrophic cardiomyopathy; X-linked myopathy with hypertrophic cardiomyopathy

For each gene, relevant published reports were identified in HGMD and by manual Pubmed searches. The following types of evidence were extracted from these reports:

- Where a cohort of unrelated HCM patients was sequenced for a gene, the frequency of rare variants identified (ExAC frequency < 0.01%) was calculated.
- The segregation of a variant within families was noted. Only the presence of a variant in affected individuals, i.e. those with a HCM or LV hypertrophy phenotype, was considered for variant segregation. Due to the incomplete penetrance and late disease onset characteristic of HCM, the presence of a variant in currently phenotypically-negative individuals was tolerated. In contrast, the absence of a variant in an affected individual was considered to indicate non-segregation of the variant. The maximum LOD scores for reported family pedigrees were noted if defined in published studies, otherwise these were estimated based on the number of meioses between affected variant carriers including obligate carriers (multiplying the number of meioses by 0.3 gives an approximate LOD score⁹⁷).
- The *de novo* occurrence of a variant with confirmed paternity and maternity.
- Functional validation assays of variants detected in HCM patients.

For each sequenced HCM cohort identified – from published reports, the combined clinical samples (OMGL and LMM) and the prospective HCM cohort – a burden test of rare variants in comparison to ExAC was performed as described above. This test was repeated for the

combined HCM cohort (clinical, prospective and published) for each gene where data was available. Genes were determined to have a significant case excess if the p-value for this combined cohort test was less than 0.0016 (0.05 with Bonferroni correction for 31 tests). For any specific variant occurring in ≥ 3 cases of the combined HCM cohort, a Fisher’s exact test was also performed comparing the variant’s frequency in cases and ExAC.

2.2.7 Extended HCM gene analysis – classification

The evidence described above, i.e. case excess at gene or variant level, highest LOD score, evidence of *de novo* inheritance and functional data, was amalgamated to classify the 31 genes according to the strength of their association with HCM. In addition to the 8 sarcomeric genes defined as “valid”, the genes were grouped into strong, moderate and weak evidence, as well as “functional evidence only” and “no evidence” according to the criteria described in Figure 2.5 and Table 2.6.

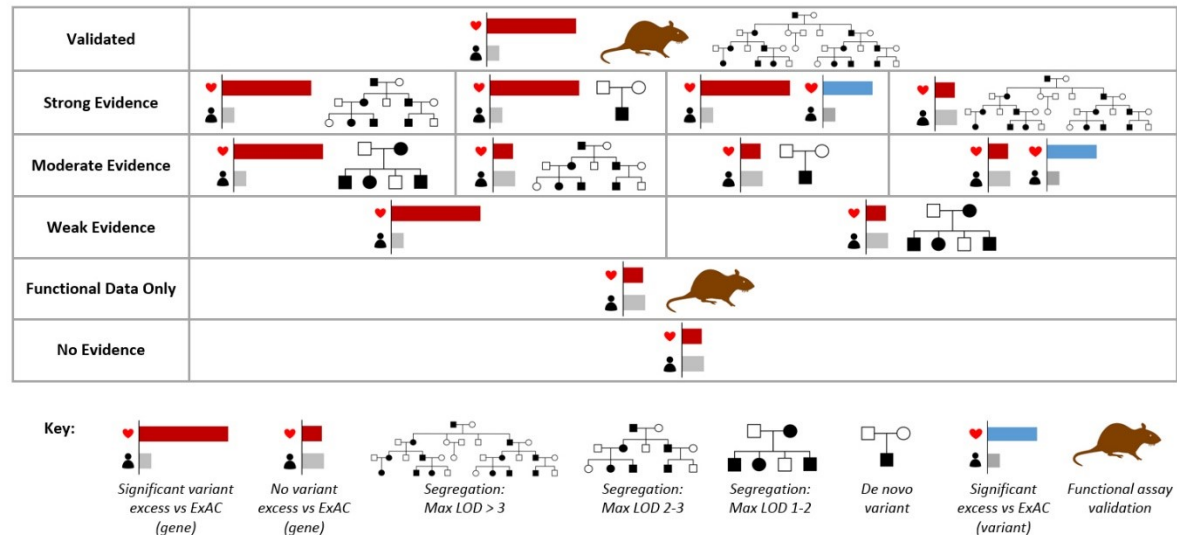


Figure 2.5: Classification system adopted in this study to assess the evidence for association for genes implicated in HCM⁸⁹. See text for details as to how these evidence strands were applied and Table 2.6 for more detailed definitions.

Validated Genes	1.	a) Significant excess in case cohorts over reference population AND b) Maximum LOD score >3 AND c) Functional validation of variants in <i>in vitro</i> or <i>in vivo</i> assays
Strong Evidence	1.	a) No significant excess in case cohorts over reference population AND b) Maximum LOD score >3.
	2.	a) Significant excess in case cohorts over reference population AND b) Maximum LOD score >2 and <3 OR <i>de novo</i> variant OR Significant excess of individual variant in cases over reference population.
Moderate Evidence	1.	a) No significant excess in case cohorts over reference population AND b) Maximum LOD score >2 and <3 OR <i>de novo</i> variant OR Significant excess of individual variant in cases over reference population.
	2.	a) Significant excess in case cohorts over reference population AND b) Maximum LOD score >1 and <2.
Weak Evidence	1.	a) Significant excess in case cohorts over reference population AND b) No maximum LOD score >1.
	2.	a) No significant excess in case cohorts over reference population AND b) Maximum LOD score >1 and <2.
Functional data only	1.	Variants demonstrated to have a functional effect in the absence of segregation data and without a significant excess in case cohorts over population/controls.
No Evidence	1.	a) Reported variants are present at a frequency (MAF>0.0001 in ExAC) which is incompatible with being a causative HCM mutation AND b) No evidence for family segregation, functional effects of variants or excess of variation in case cohorts.
	2.	a) No significant excess of rare variants in case cohorts over reference population AND b) No evidence for family segregation or functional effects of variants.

Table 2.6: Classification system adopted in this study to assess the evidence for association for genes implicated in HCM⁸⁹. See text for details as to how these evidence strands were applied.

2.3 Results

2.3.1 Comparing rare variation between clinical cardiomyopathy cohorts and ExAC

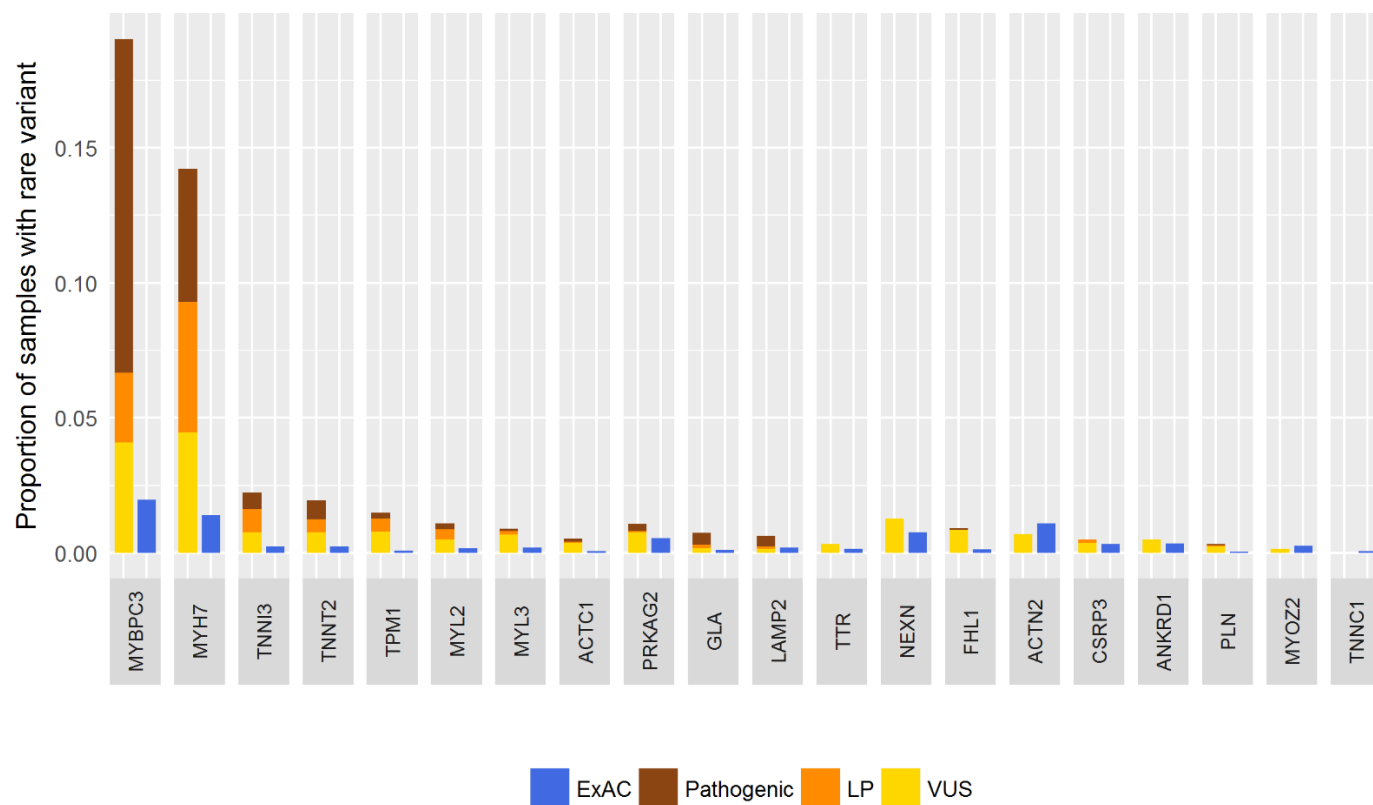
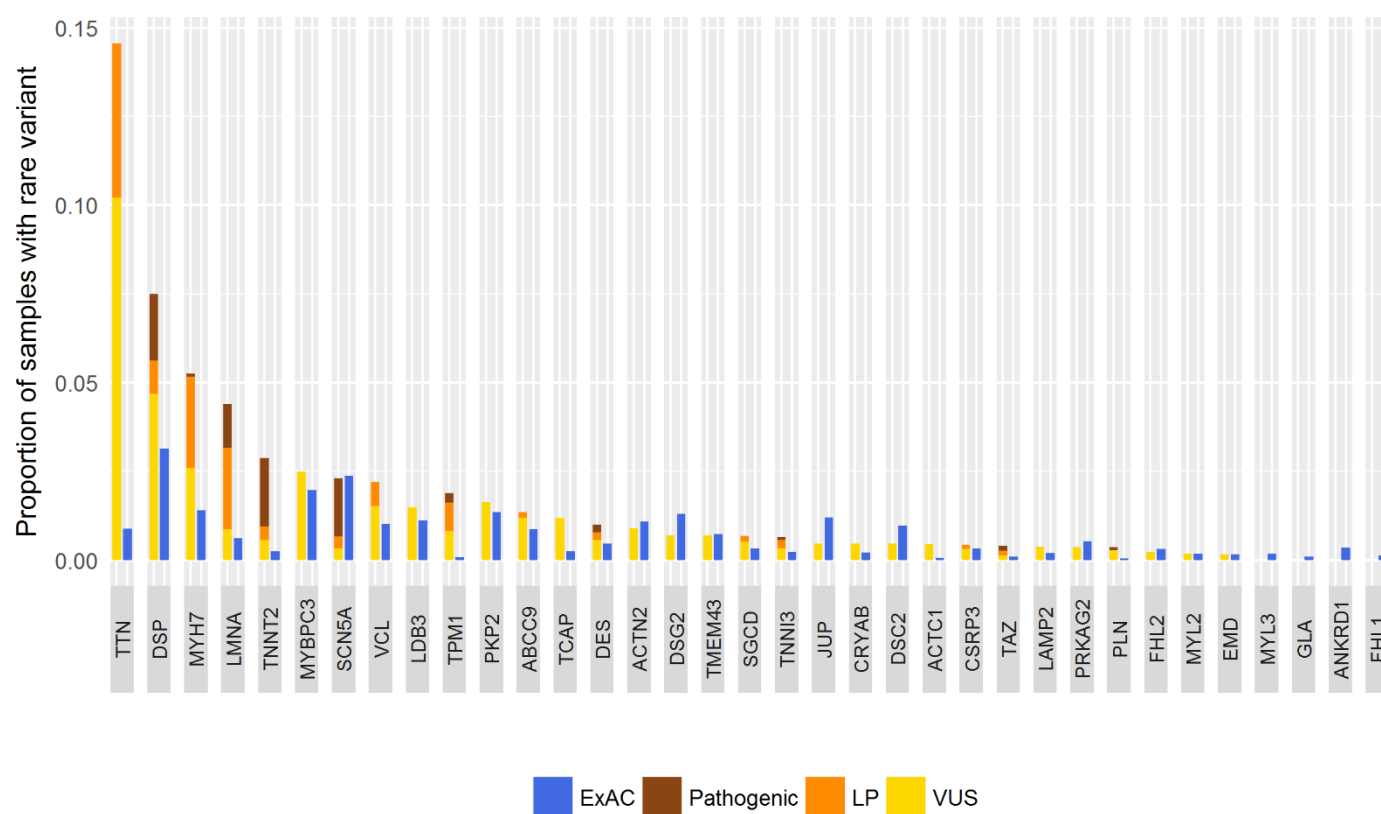
The frequency of rare variation in the HCM and DCM clinical cohorts and comparison to ExAC reference population samples are shown in Figure 2.6 and Table 2.7 and Table 2.8.

2.3.1.1 Hypertrophic cardiomyopathy (HCM)

A significant excess of rare variants was observed in the HCM clinical cohorts compared to ExAC for the 8 core sarcomeric genes associated with HCM – *MYBPC3*, *MYH7*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3* and *ACTC1*. As observed in other published cohorts, *MYBPC3* and *MYH7* account for the majority of rare variation in HCM, with case excesses of 17.1% and 12.8% respectively. Variants in the other genes are less common, with cases excesses ranging from 2.0% (*TNNI3*) to 0.5% (*ACTC1*).

A significant excess was also observed in three genes associated with conditions that mimic HCM – *GLA* (Fabry disease), *LAMP2* (Danon disease) and *PRKAG2* (PRKAG2-related HCM). This indicates a proportion of these cases actually had these alternate conditions and were referred for HCM genetic testing to either discover or confirm such diagnoses. While LMM claimed to exclude patients with these metabolic cardiomyopathies, the nature of these conditions is such that they often present as HCM and are only diagnosed with a positive genetic test. It is notable that the frequency of rare variants in these genes is lower in the LMM cohort than the OMGL cohort, which did not proactively exclude such cases (0.46% vs 1.10%, 0.41% vs 1.30% and 0.94% vs 1.24% for *GLA*, *LAMP2* and *PRKAG2* in LMM and OMGL respectively).

Figure 2.6 (next page): The proportion of individuals with rare protein-altering variants for (A) combined HCM clinical cohorts (OMGL and LMM) and ExAC and (B) for combined DCM clinical cohorts (OMGL and LMM) and ExAC. The HCM and DCM gene frequencies are sub-classified by the variant classifications by the clinical laboratories. Only genes with a minimum of 300 sequenced cardiomyopathy cases are shown. See Table 2.7 and Table 2.8 for full details.

A**B**

In contrast, for genes more recently implicated in HCM, there is either limited (*PLN*, *FHL1*), non-significant (*ANKRD1*, *CSRP3*, *NEXN*) or no (*ACTN2*, *MYOZ2*, *TNNC1*) excess of rare variants. It should be noted however that fewer cases were sequenced for many of these genes than for the sarcomeric genes (Table 2.4). It is therefore possible that a larger sequenced cohort would reveal a significant case excess for some of these genes, though this data would suggest that at best they would be very rare causes of HCM.

Gene	Cases	ExAC samples	Non-truncating variants				Truncating variants			
			ExAC	Cases	Excess	p-value	ExAC	Cases	Excess	p-value
ACTC1	4185	60198	0.062%	0.526%	0.464%	<0.0001	0.004%	0.000%	-0.004%	1.0000
ACTN2	1439	60388	1.076%	0.625%	-0.451%	0.1172	0.010%	0.069%	0.059%	0.1520
ANKRD1	807	60300	0.304%	0.372%	0.068%	0.7398	0.052%	0.124%	0.072%	0.3466
CSRP3	2167	60647	0.288%	0.369%	0.081%	0.4189	0.036%	0.138%	0.102%	0.0536
FHL1	1535	60268	0.123%	0.847%	0.724%	<0.0001	0.000%	0.065%	0.065%	-
GLA	3700	59404	0.100%	0.730%	0.630%	<0.0001	0.000%	0.000%	0.000%	-
LAMP2	3290	59987	0.198%	0.243%	0.045%	0.5454	0.000%	0.395%	0.395%	-
MYBPC3	6179	47155	1.884%	9.872%	7.988%	<0.0001	0.086%	9.160%	9.074%	<0.0001
MYH7	6112	60471	1.350%	14.136%	12.786%	<0.0001	0.048%	0.082%	0.034%	0.2356
MYL2	4185	60516	0.154%	1.027%	0.873%	<0.0001	0.024%	0.072%	0.048%	0.0932
MYL3	4185	60601	0.178%	0.884%	0.706%	<0.0001	0.004%	0.000%	-0.004%	1.0000
MYOZ2	632	60436	0.236%	0.158%	-0.078%	1.0000	0.020%	0.000%	-0.020%	1.0000
NEXN	632	59277	0.692%	1.108%	0.416%	0.2192	0.064%	0.158%	0.094%	0.3388
PLN	2167	60475	0.042%	0.138%	0.096%	0.0713	0.006%	0.185%	0.179%	0.0001
PRKAG2	3973	57747	0.520%	1.007%	0.487%	0.0002	0.012%	0.050%	0.038%	0.1103
TNNC1	632	59156	0.054%	0.000%	-0.054%	1.0000	0.004%	0.000%	-0.004%	1.0000
TNNI3	6047	53856	0.186%	2.133%	1.947%	<0.0001	0.042%	0.099%	0.057%	0.0571
TNNT2	6103	56924	0.214%	1.770%	1.556%	<0.0001	0.030%	0.180%	0.150%	<0.0001
TPM1	4447	58727	0.084%	1.484%	1.400%	<0.0001	0.004%	0.000%	-0.004%	1.0000
TTR	632	60635	0.136%	0.316%	0.180%	0.2150	0.004%	0.000%	-0.004%	1.0000

Table 2.7: Frequencies of rare variation (non-truncating and truncating variants) in combined clinical HCM cohorts (OMGL and LMM) compared to ExAC, including case excess and Fisher's exact test p-value. Genes with a significant excess ($p < 0.0025$ after Bonferroni correction) are highlighted in bold.

For the valid sarcomeric genes, an excess of rare variants was observed in cases even when only assessing variants classified as VUS by the clinical labs (Figure 2.6). This suggests that clinical laboratories are being overly conservative in assigning pathogenicity to variants in these genes and that improved methods for discriminating between benign and pathogenic variation could substantially increase the yield from HCM genetic testing (see Chapter 3).

2.3.1.2 Dilated Cardiomyopathy (DCM)

DCM is a highly genetically heterogeneous disease with over 60 genes implicated in the condition. Consequently, clinical testing gene panels have expanded to incorporate these findings, with the OMGL panel now comprising 28 genes and the LMM panel comprising 46 genes (Table 2.4). Despite this, a significant case excess was observed in only a minority of these genes.

Truncating variants in *TTN* were the most common variant class detected in this cohort with a case excess of 13.7%, in line with findings from other large research cohorts^{81,98,99}. Interestingly, the clinical laboratories differed in their interpretation of this variant class, with LMM classifying them as likely pathogenic (or VUS favour pathogenic) and OMGL classifying them as VUS. Truncating variants in several other genes were also significantly enriched in the DCM cohort, including *DSP* (2.7% excess), *LMNA* (1.5%), *SCN5A* (0.9%) and *VCL* (0.6%).

For non-truncating variants, a significant though modest excess was observed in four well characterised DCM genes, *MYH7* (3.9%), *TNNT2* (2.7%), *LMNA* (2.2%) and *TPM1* (1.6%), as well as *TCAP* (0.97%). A significant excess was also observed in two other established DCM genes (*RBM20* and *TNNC1*) although this data should be viewed with caution due to the limited number of DCM cases sequenced (156) and the poor coverage of the *RBM20* gene in ExAC (the mean coverage for *RBM20* variants is 20,094 alleles from a total of 121,412).

Gene	Cases	ExAC samples	Non-truncating variants				Truncating variants			
			ExAC	Cases	Excess	p-value	ExAC	Cases	Excess	p-value
ABCC9	590	60453	0.800%	1.186%	0.386%	0.2487	0.074%	0.169%	0.095%	0.3542
ACTC1	1103	60198	0.062%	0.363%	0.301%	0.0061	0.004%	0.091%	0.087%	0.0534
ACTN2	895	60388	1.076%	0.894%	-0.182%	0.7436	0.010%	0.000%	-0.010%	1.0000
ANKRD1	426	60300	0.304%	0.000%	-0.304%	0.6435	0.052%	0.000%	-0.052%	1.0000
CASQ2	121	59954	0.462%	0.826%	0.364%	0.4298	0.046%	0.826%	0.780%	0.0549
CAV3	121	60262	0.124%	0.000%	-0.124%	1.0000	0.014%	0.000%	-0.014%	1.0000
CRYAB	425	50492	0.198%	0.471%	0.273%	0.2105	0.010%	0.000%	-0.010%	1.0000
CSRP3	945	60647	0.288%	0.317%	0.029%	0.7558	0.036%	0.106%	0.070%	0.2993
DES	894	51128	0.436%	0.671%	0.235%	0.2998	0.036%	0.224%	0.188%	0.0376
DSC2	427	60064	0.938%	0.468%	-0.470%	0.4491	0.026%	0.000%	-0.026%	1.0000
DSG2	427	59667	1.226%	0.703%	-0.523%	0.5021	0.072%	0.000%	-0.072%	1.0000
DSP	427	59620	3.078%	4.684%	1.606%	0.0660	0.070%	2.810%	2.740%	<0.0001
DTNA	121	60219	0.426%	0.826%	0.400%	0.4100	0.010%	0.000%	-0.010%	1.0000
EMD	590	59063	0.166%	0.169%	0.003%	1.0000	0.000%	0.000%	0.000%	-
FHL1	355	60268	0.123%	0.000%	-0.123%	1.0000	0.000%	0.000%	0.000%	-
FHL2	425	59228	0.300%	0.235%	-0.065%	1.0000	0.010%	0.000%	-0.010%	1.0000
GLA	541	59404	0.100%	0.000%	-0.100%	1.0000	0.000%	0.000%	0.000%	1.0000
JUP	425	55896	1.180%	0.471%	-0.709%	0.2538	0.018%	0.000%	-0.018%	1.0000
LAMA4	121	59741	2.180%	4.959%	2.779%	0.0535	0.102%	0.000%	-0.102%	1.0000
LAMP2	532	59987	0.198%	0.376%	0.178%	0.2383	0.000%	0.000%	0.000%	1.0000
LDB3	740	58166	1.074%	1.892%	0.818%	0.0458	0.048%	0.000%	-0.048%	1.0000
LMNA	1044	51266	0.608%	2.874%	2.266%	<0.0001	0.014%	1.533%	1.519%	<0.0001
MYBPC3	1161	47155	1.884%	2.498%	0.614%	0.1278	0.086%	0.000%	-0.086%	1.0000
MYH6	121	60328	2.336%	2.479%	0.143%	0.7655	0.128%	0.000%	-0.128%	1.0000
MYH7	1315	60471	1.350%	5.247%	3.897%	<0.0001	0.048%	0.076%	0.028%	0.4756
MYL2	543	60516	0.154%	0.184%	0.030%	0.5224	0.024%	0.000%	-0.024%	1.0000
MYL3	543	60601	0.178%	0.000%	-0.178%	1.0000	0.004%	0.000%	-0.004%	1.0000
MYLK2	121	57226	0.768%	1.653%	0.885%	0.2447	0.014%	0.000%	-0.014%	1.0000
MYOZ2	121	60436	0.236%	0.000%	-0.236%	1.0000	0.020%	0.000%	-0.020%	1.0000
NEXN	156	59277	0.692%	2.564%	1.872%	0.0241	0.064%	0.641%	0.577%	0.0975
PKP2	427	58406	1.280%	1.639%	0.359%	0.5117	0.076%	0.000%	-0.076%	1.0000
PLN	1095	60475	0.042%	0.274%	0.232%	0.0132	0.006%	0.091%	0.085%	0.0858
PRKAG2	546	57747	0.520%	0.366%	-0.154%	1.0000	0.012%	0.000%	-0.012%	1.0000
RYR2	121	53692	3.370%	4.959%	1.589%	0.3136	0.114%	0.000%	-0.114%	1.0000
SCN5A	304	56420	2.318%	1.316%	-1.002%	0.3361	0.064%	0.987%	0.923%	0.0010
SGCD	590	50725	0.314%	0.508%	0.194%	0.4418	0.022%	0.169%	0.147%	0.1195
TAZ	740	51321	0.091%	0.270%	0.179%	0.1538	0.001%	0.135%	0.134%	0.0282
TCAP	590	55233	0.218%	1.186%	0.968%	0.0004	0.024%	0.000%	-0.024%	1.0000
TMEM43	427	60220	0.690%	0.703%	0.013%	0.7712	0.040%	0.000%	-0.040%	1.0000
TNNC1	156	59156	0.054%	1.923%	1.869%	0.0001	0.004%	0.000%	-0.004%	1.0000
TNNI3	1239	53856	0.186%	0.646%	0.460%	0.0033	0.042%	0.000%	-0.042%	1.0000
TNNT2	1254	56924	0.214%	2.871%	2.657%	<0.0001	0.030%	0.000%	-0.030%	1.0000
TPM1	1111	58727	0.084%	1.710%	1.626%	<0.0001	0.004%	0.180%	0.176%	0.0020
TTN	156	58613	-	-	-	-	0.876%	14.565%	13.689%	<0.0001
TTR	121	60635	0.136%	0.000%	-0.136%	1.0000	0.004%	0.000%	-0.004%	1.0000
VCL	590	59393	0.984%	1.525%	0.541%	0.2024	0.032%	0.678%	0.646%	0.0001

Table 2.8 (previous page): Frequencies of rare variation (non-truncating and truncating variants) in combined clinical DCM cohorts (OMGL and LMM) compared to ExAC, including case excess and Fisher's exact test p-value. Genes with a significant excess ($p < 0.001$ after Bonferroni correction) are highlighted in bold.

2.3.2 Population-specific burden analysis

Individual patient ethnicity data was available for only one subset of the clinical cardiomyopathy cohorts – the DCM cohort from LMM. To test if population-specific variants were having a confounding effect on the burden testing of the full cardiomyopathy and ExAC cohorts, I compared the results from the “all DCM versus all ExAC” burden test to the “reported Caucasian DCMs versus ExAC-NFE” subset. While there were marginal differences in case and ExAC frequencies between all and Caucasian/ExAC-NFE, there was complete correlation between the genes with a significant excess in both comparisons (Table 2.9). The Pearson correlation coefficient for the case excess values between the two comparisons was 0.97 (Figure 2.7). These results indicate that population stratification effects are not having a significant effect on the burden testing performed in this study.

Gene	1. Caucasian DCM cases vs ExAC-NFE				2. All DCM cases vs ExAC			
	ExAC	Case	Excess	p-value	ExAC	Case	Excess	p-value
ABCC9	0.74%	1.64%	0.90%	0.0580	0.87%	1.36%	0.48%	0.1847
ACTC1	0.07%	0.88%	0.81%	0.0004	0.06%	0.67%	0.60%	0.0002
ACTN2	0.78%	1.10%	0.32%	0.3751	1.09%	0.85%	-0.24%	0.8399
ANKRD1	0.27%	0.00%	-0.27%	1.0000	0.35%	0.00%	-0.35%	1.0000
CASQ2	0.46%	2.74%	2.28%	0.0452	0.51%	1.65%	1.15%	0.1269
CAV3	0.10%	0.00%	-0.10%	1.0000	0.14%	0.00%	-0.14%	1.0000
CRYAB	0.15%	0.00%	-0.15%	1.0000	0.21%	0.83%	0.62%	0.2243
CSRP3	0.25%	0.82%	0.58%	0.0651	0.32%	0.68%	0.35%	0.1304
CTF1	0.02%	0.27%	0.26%	0.0747	0.05%	0.17%	0.12%	0.2456
DES	0.35%	1.10%	0.75%	0.0431	0.47%	1.02%	0.55%	0.0651
DSC2	0.83%	0.00%	-0.83%	1.0000	0.96%	0.81%	-0.15%	1.0000
DSG2	1.01%	1.35%	0.35%	0.5272	1.30%	1.63%	0.33%	0.6755
DSP	2.27%	8.11%	5.84%	0.0069	3.15%	7.32%	4.17%	0.0162
DTNA	0.33%	1.37%	1.04%	0.2137	0.44%	0.83%	0.39%	0.4120
EMD	0.12%	0.27%	0.15%	0.3704	0.23%	0.17%	-0.06%	0.6265
FHL2	0.28%	0.00%	-0.28%	1.0000	0.31%	0.83%	0.52%	0.3149
GLA	0.07%	0.00%	-0.07%	1.0000	0.14%	0.00%	-0.14%	1.0000
JUP	0.82%	0.00%	-0.82%	1.0000	1.20%	0.00%	-1.20%	0.4098
LAMA4	1.59%	1.37%	-0.22%	1.0000	2.28%	4.96%	2.68%	0.0599
LAMP2	0.14%	0.77%	0.63%	0.1645	0.27%	0.90%	0.62%	0.0745
LDB3	0.77%	1.11%	0.34%	0.4064	1.12%	1.89%	0.77%	0.3760
LMNA	0.53%	6.22%	5.70%	<0.0001	0.62%	4.73%	4.11%	<0.0001
MYBPC3	1.50%	2.40%	0.90%	0.1199	1.97%	2.65%	0.68%	0.1866
MYH6	1.79%	2.74%	0.95%	0.3783	2.47%	2.48%	0.01%	1.0000
MYH7	1.34%	6.10%	4.76%	<0.0001	1.40%	5.95%	4.55%	<0.0001
MYL2	0.14%	0.93%	0.78%	0.1469	0.18%	0.53%	0.36%	0.2849
MYL3	0.20%	0.00%	-0.20%	1.0000	0.18%	0.00%	-0.18%	1.0000
MYLK2	0.48%	0.00%	-0.48%	1.0000	0.78%	1.65%	0.87%	0.2456
MYOZ2	0.15%	0.00%	-0.15%	1.0000	0.26%	0.00%	-0.26%	1.0000
NEXN	0.69%	3.16%	2.47%	0.0291	0.76%	3.21%	2.45%	0.0071
PKP2	0.83%	5.41%	4.57%	0.0036	1.36%	4.07%	2.71%	0.0271
PLN	0.02%	0.89%	0.87%	<0.0001	0.05%	0.54%	0.49%	0.0007
PRKAG2	0.46%	0.93%	0.47%	0.3930	0.53%	1.05%	0.52%	0.2711
RBM20	0.00%	6.32%	6.32%	<0.0001	1.41%	5.77%	4.36%	0.0005
RYR2	2.90%	5.48%	2.58%	0.1625	3.48%	4.96%	1.48%	0.3208
SGCD	0.26%	0.27%	0.02%	0.6125	0.34%	0.68%	0.34%	0.1437
TAZ	0.10%	0.22%	0.12%	0.3778	0.13%	0.41%	0.28%	0.0360
TCAP	0.23%	1.64%	1.42%	0.0003	0.24%	1.19%	0.94%	0.0008
TMEM43	0.50%	2.70%	2.20%	0.0537	0.73%	1.63%	0.90%	0.2270
TNNC1	0.04%	3.16%	3.12%	<0.0001	0.06%	1.92%	1.86%	0.0001
TNNI3	0.19%	0.65%	0.47%	0.0593	0.23%	0.79%	0.57%	0.0094
TNNT2	0.16%	3.49%	3.33%	<0.0001	0.24%	3.31%	3.07%	<0.0001
TPM1	0.07%	2.40%	2.33%	<0.0001	0.09%	2.12%	2.03%	<0.0001
TTR	0.13%	0.00%	-0.13%	1.0000	0.14%	0.00%	-0.14%	1.0000
VCL	0.67%	2.47%	1.79%	0.0011	1.02%	2.20%	1.19%	0.0111

Table 2.9 (previous page): Comparison of the excess of rare variants in LMM DCM cases over ExAC between Caucasian only (Caucasian DCM cases versus non-Finnish European subset of ExAC) and all sample analysis (all cases versus full ExAC dataset). Significant p-values (Fisher's exact test $p < 0.001$ with Bonferroni correction for 48 tests) for burden testing are shaded in green. Note that as this analysis only include LMM samples, there are some genes with a significant excess here (ACTC1, PLN) that were not significant with the combined DCM cohort.

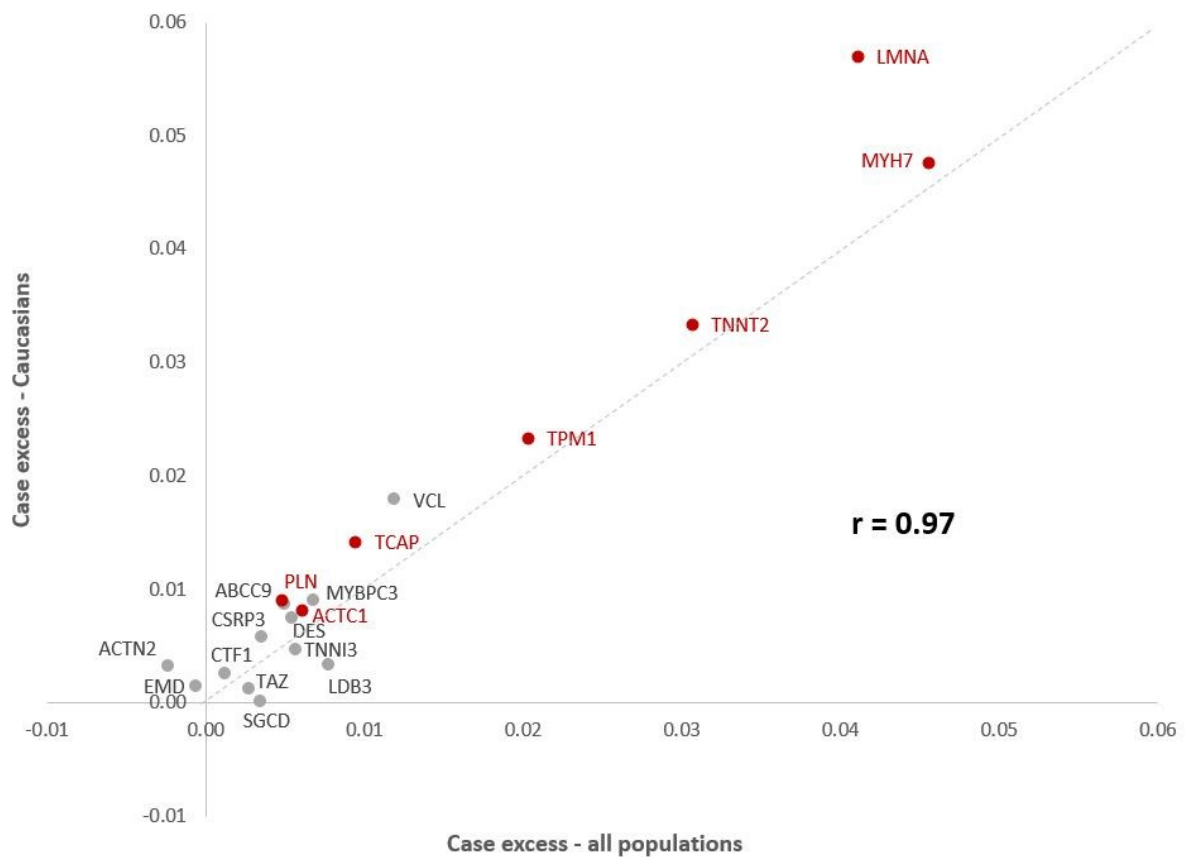
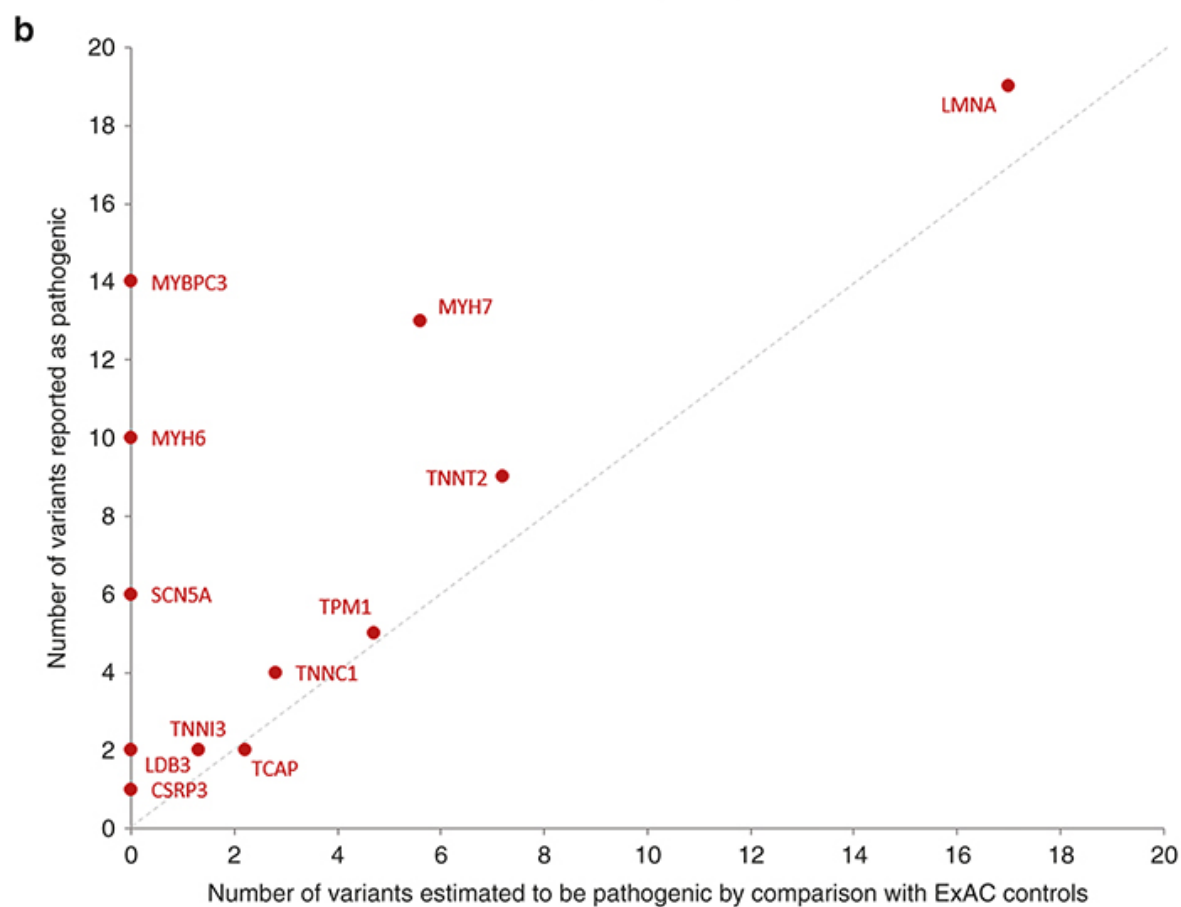
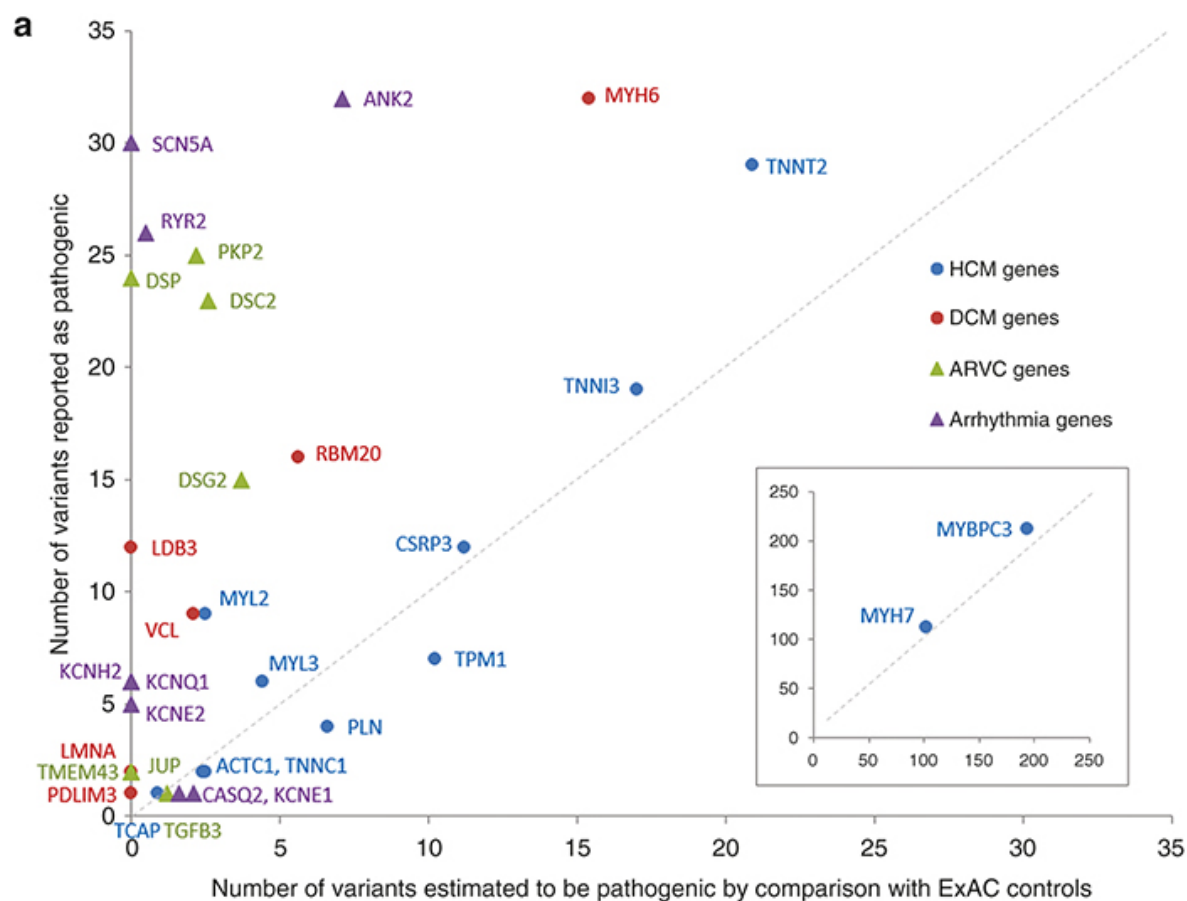


Figure 2.7: Comparison of the excess of rare variants in LMM DCM cases over ExAC between all sample analysis (all cases versus full ExAC dataset) and Caucasian only (Caucasian DCM cases versus non-Finnish European subset of ExAC). Genes with a significant excess in cases (Fisher's exact test $p < 0.001$ with Bonferroni correction for 48 tests) are shown in red. The Pearson correlation coefficient was calculated at 0.97. Only genes with >300 DCM samples sequenced are shown. See Table 2.9 for full details.

2.3.3 Assessing results of extended gene panel cardiomyopathy studies

Advances in NGS technology, i.e. increasing capacity and decreasing costs, have recently enabled the direct and simultaneous sequencing of large disease cohorts across expanding panels of genes. These panels can include genes with little or no evidence for involvement in the disease in question that, allied to insufficiently stringent interpretation of detected variants, can lead to erroneous associations of both variants and genes with diseases like cardiomyopathies, as well as misleading claims about the genetic architecture of these conditions^{86,87}. The data from two such studies were reassessed by comparing the frequency of reported pathogenic variants with the frequency of rare variants in excess over ExAC for each gene.

Figure 2.8 (next page): Comparison of the number of variants reported as putatively pathogenic in HCM (A) and DCM (B) research studies (using generic analysis criteria such as variant class, missense functional prediction algorithms and variant population frequency in ESP) with those predicted as pathogenic by the excess of rare variants in cases over ExAC in each gene. For the HCM study (A), genes are coloured by the cardiac disease for which they are primarily associated, as defined by Lopes et al.⁸⁶ While there is good concordance between the research findings and the ExAC predictions for established HCM genes, for genes primarily associated with DCM, ARVC and arrhythmias, the variation in cases is similar to that in controls. While the DCM study (B) only assessed genes implicated in DCM^{90–92}, variation burden in MYBPC3, SCN5A and MYH6 is similar between cases and ExAC, suggesting most variants in these genes are unlikely to be causing DCM.



The HCM study sequenced 36 genes in 874 cases⁸⁶, including genes that are primarily associated with DCM, ARVC and arrhythmias, and variants from all genes were similarly analysed (regardless of the evidence, or lack of it, implicating these genes with HCM). Figure 2.8A compares the number of variants reported in this study as pathogenic with the number predicted to be pathogenic based on the excess of rare variation over ExAC. For bona fide HCM genes (blue), there is broad correlation between the reported and predicted numbers, indicating that significant excesses of rare variants are observed in these genes as expected, with the majority of these classified as pathogenic by the methods used in the published study. In contrast, there is little excess variation in the DCM (red) and especially ARVC (green) and arrhythmia (purple) genes, suggesting that most of the variants detected in these genes in the HCM cases are benign background variants that do not contribute to disease burden. However, substantial numbers of variants in these genes are reported as pathogenic using the non-specific methods employed in the report, highlighting the dangers of expanding genetic testing to genes not proven to be associated with the disease in question.

For the DCM study, data from three reports from the same study published between 2008 and 2010^{90–92} was similarly reassessed (a broader DCM study sequencing 84 genes in 639 bases could not be analysed due to the fact that the data was not published in an accessible and reproducible manner⁸⁷). While this study sequenced 12 genes that had been implicated in DCM, there was no excess of rare variants for several of these genes, including *MYBPC3*, *MYH6* and *SCN5A*, indicating they may not in fact contribute to DCM (Figure 2.8B). While a significant excess of *SCN5A* truncating variants was identified in the OMGL clinical cohort, only missense variants were detected in this cohort, suggesting this variant class does not contribute to DCM. Although *MYH7* is a valid DCM gene, over twice as many variants in this gene were reported as pathogenic than the excess over ExAC would suggest, highlighting the necessity of considering the background level of rare, benign variation when interpreting variants.

2.3.4 Meta-analysis of genes implicated in HCM

To clarify the HCM burden testing results described, and extend the analysis to all other genes that have been implicated in the condition, a comprehensive approach was employed that included both expanded burden testing and the evaluation of published evidence. Variants in 51 genes were associated with HCM in HGMD at the time of searching (version 2015.3). Those genes implicated in classical, isolated HCM were distinguished from genes associated with complex phenotypes including LV hypertrophy.

2.3.4.1 Genes associated with complex and syndromic conditions

The genes associated with complex phenotypes that include LV hypertrophy are summarised in Table 2.10. These include 5 genes robustly associated with metabolic cardiomyopathies that can mimic, and initially present as, HCM – *GAA* (Pompe disease), *GLA* (Fabry disease), *LAMP2* (Danon disease), *PRKAG2* (PRKAG2-related HCM) and *TTR* (Hereditary transthyretin-related amyloidosis). Some or all of these genes are regularly included in HCM clinical genetic testing to identify or confirm these diagnoses.

Variants in two ion channel genes (*CACNA1C* and *KCNQ1*) were identified in patients with mixed HCM and arrhythmia phenotypes but it is thought that these variants were responsible for the arrhythmogenic component of these phenotypes^{100–102}. Five other genes were associated with rare and complex disorders that included both HCM and other non-cardiac symptoms, e.g. nemaline myopathy (*ACTA1*), psychomotor retardation (*MRPL3*) and hearing loss (*MYO6*). These conditions are often early onset and mostly exhibit autosomal recessive inheritance (Table 2.10). While it is important for clinical geneticists to be aware of these genetic associations when investigating patients with complex phenotypes, the inclusion of such genes in gene panels for standard HCM genetic testing is unwarranted.

Category	Gene	Condition	Inheritance	Refs
Rare conditions incl. LV hypertrophy	ACTA1	Infantile-onset nemaline myopathy and HCM	<i>de novo</i>	103,104
	COX15	Fatal infantile HCM and COX deficiency	Autosomal recessive	105
	MRPL3	HCM with psychomotor retardation and multiple respiratory chain deficiency	Autosomal recessive	106
	MYO6	HCM and sensorineural hearing loss	Autosomal dominant	107
	SLC25A4	Complex phenotype including HCM, mild myopathy, lactic acidosis, ophthalmoplegia and exercise intolerance	Autosomal recessive	108–111
Dual HCM / arrhythmia	CACNA1C	HCM and Timothy/Brugada syndromes	-	100,102
	KCNQ1	HCM and Long QT syndrome	-	101,102
Metabolic cardiomyopathy	LAMP2	Danon disease / Glycogen storage disease 2b	-	-
	GAA	Pompe disease / Glycogen storage disease 2	-	-
	GLA	Fabry disease	-	-
	PRKAG2	PRKAG2-related HCM	-	-
	TTR	Hereditary transthyretin-related amyloidosis	-	-

Table 2.10: Summary of genes associated with HCM where the conditions are actually complex phenotypes that include left ventricular hypertrophy.

2.3.4.2 Valid sarcomeric genes

The 8 core sarcomeric HCM genes – *MYBPC3*, *MYH7*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, *ACTC1* – were first associated with the disease through detailed family linkage studies in the 1990s (see Table 2.1 for first references) that provided convincing evidence for the involvement in HCM of both the detected variants and therefore these genes. Since then, there have been hundreds of further reports describing additional variants in these genes in HCM patients and families, many accompanied by segregation data and validation by functional assays. As shown above, a significant excess of rare variants in the combined clinical cohort over ExAC has also been demonstrated for these genes, with over 500 distinct variants in these genes given a clinical grade classification of pathogenic or likely pathogenic in these cohorts alone (Table 2.11). Taking all of this evidence together, these genes can be considered as validated HCM genes.

Gene	HGMD variants	HGMD references	ExAC frequency	HCM frequency	Case excess	Pathogenic variants
MYH7	424	168	1.40%	14.22%	12.82%*	157
MYBPC3	505	155	1.97%	19.03%	17.06%*	276
TNNT2	65	75	0.24%	1.95%	1.71%*	20
TNNI3	52	48	0.23%	2.23%	2.01%*	27
TPM1	22	36	0.09%	1.48%	1.40%*	11
MYL2	25	29	0.18%	1.10%	0.92%*	11
MYL3	23	20	0.18%	0.88%	0.70%*	4
ACTC1	17	16	0.06%	0.53%	0.46%*	2

Table 2.11: Overview of the evidence validating the role of eight validated sarcomeric genes in HCM. Displayed are the distinct number of variants and references in HGMD (version 2015.3) that associate the gene with HCM, the frequency of rare variants in ExAC and clinical genetics cohorts (using combined data from OMGL and LMM laboratories), the case excess of rare variants in the clinical cohorts over ExAC (= Fisher's exact test p-value < 0.0001) and the number of distinct variants with clinical-grade pathogenic and likely pathogenic classification from either or both of the OMGL and LMM cohorts.*

2.3.4.3 Putative HCM genes

The remaining 31 genes from the HGMD search are largely non-sarcomeric genes that have been implicated in HCM by one or more published studies. Table 2.12 summarises the frequencies of rare variation in the prospective, clinical and published cohorts, gene burden test versus ExAC and published genetic evidence (see Appendix 7.1 for full details).

2.3.4.4 Genes with strong, moderate or weak evidence for a primary pathogenic role in HCM

11 genes were classified as having strong (*CSRP3*, *FHL1*, *PLN*), moderate (*ACTN2*, *CRYAB*, *FLNC*, *MYOZ2*) or weak (*MYH6*, *TNNC1*, *TRIM55*, *TRIM63*) evidence according to the criteria outlined in Figure 2.5 and Table 2.6. The evidence for these associations is summarised in Figure 2.9 and Table 2.12 and described in detail in the Appendix (7.1).

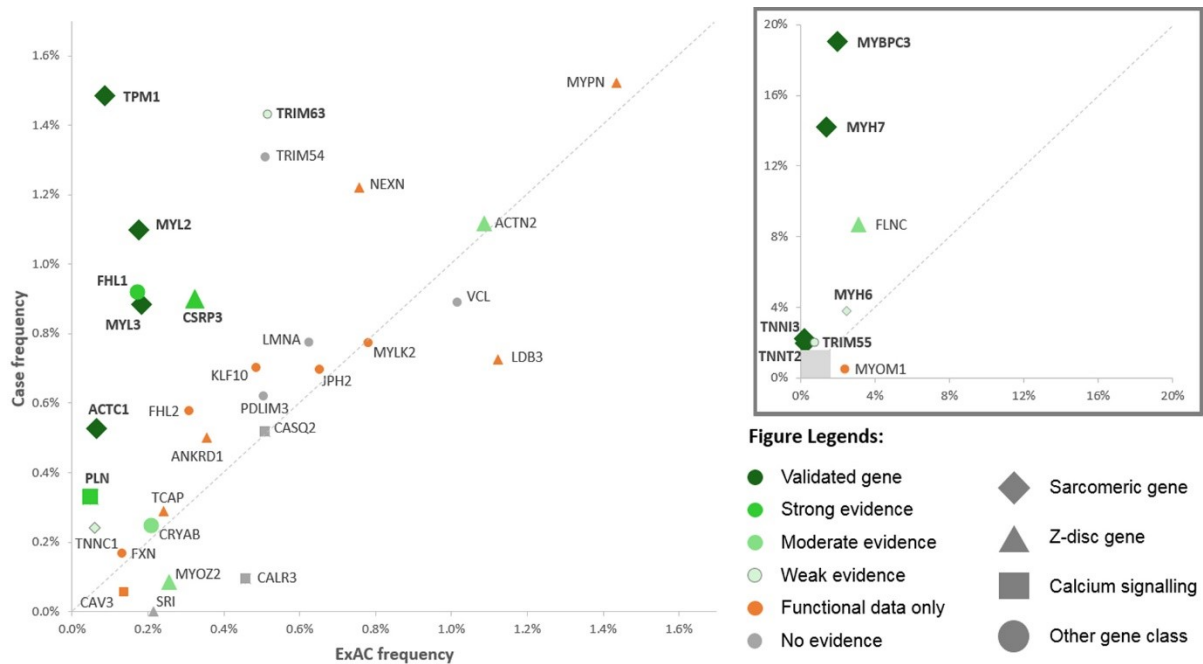


Figure 2.9: Summary of the evidence for genes that have been implicated in HCM from the meta-analysis. The scatter plot compares the frequency of rare variation in cases (y-axis) and ExAC (x-axis). Genes with a case excess will therefore be plotted near the y-axis (as with the sarcomeric genes in dark green) while those with no excess are plotted near the diagonal. Data points are coloured by evidence class, shaped by type of gene and sized by strength of published genetic evidence (as indicated).

The three genes classified as having strong evidence each demonstrated a significant excess of rare variants over ExAC as well as additional human genetic evidence from published reports.

- *CSRP3* (a Z-disc gene) had an excess of 0.58% based on a combined cohort of 4866 cases. The variant p.C58G was observed to segregate in a family pedigree, present in 8 affected and 3 unaffected and absent in 20 unaffected individuals, for an estimated maximum LOD score of 3.6¹¹². Additionally the variant p.L44P was significantly enriched in cases (11/4866) versus ExAC (0/60706) ($p < 1 \times 10^{-4}$).

Gene	ExAC freq. (%)	HCM frequency (%)				Fisher's exact test	Max LOD score	<i>de novo</i>	Single variant excess
		This study	OMGL / LMM	Other data	combined (total cases)				
Strong evidence									
CSRP3	0.32	0.75	0.51	1.42	0.90 (4866)	***	3.6	-	√
FHL1	0.12	0.49	0.91	2.48	0.92 (2061)	***	-	√	√
PLN	0.05	0.25	0.32	0.36	0.33 (5440)	***	-	-	√
Moderate Evidence									
ACTN2	1.09	1.99	0.69	0.93	1.12 (2779)	n/s	2.82	-	-
CRYAB	0.21	0.25	-	-	0.25 (804)	n/s	-	√	-
FLNC	3.21	-	-	4.91	4.91 (448)	n/s	2.33	-	-
MYOZ2	0.26	0.12	0.16	0	0.08 (2390)	no excess	2.03	-	-
Weak Evidence									
MYH6	2.47	3.36	-	4.2	3.80 (1709)	**	-	-	-
TNNC1	0.06	0.25	0	0.32	0.24 (3335)	*	1.2	-	-
TRIM55	0.73	1.5	-	2.36	2.01 (993)	***	-	-	-
TRIM63	0.51	1	-	2.02	1.43 (1398)	***	-	-	-
Functional data only (no genetic evidence)									
ANKRD1	0.35	0.37	0.5	0.78	0.50 (1995)	n/s	-	-	-
CAV3	0.14	0	-	0.1	0.06 (1824)	no excess	0.6	-	-
FHL2	0.31	0.5	-	0.83	0.58 (520)	n/s	-	-	-
FXN	0.13	0.12	-	0.26	0.17 (1193)	n/s	-	-	-
JPH2	0.65	0.99	-	0.56	0.70 (1292)	n/s	-	-	-
KLF10	0.49	0.99	-	0.54	0.68 (1328)	n/s	-	-	-
LDB3	1.12	1.48	-	0.45	0.73 (1518)	no excess	-	-	-
MYLK2	0.78	1.24	-	0	0.77 (1294)	no excess	-	-	-
MYOM1	2.39	-	-	0.53	0.53 (188)	no excess	0.6	-	-
MYPN	1.43	1.49	-	1.56	1.52 (1381)	n/s	-	-	-
NEXN	0.76	1.24	1.27	0.83	1.22 (1557)	n/s	0.9	-	-
TCAP	0.24	0.12	-	0.37	0.29 (2413)	n/s	-	-	-
No Evidence									
CALR3	0.46	0	-	0.4	0.10 (1056)	no excess	-	-	-
CASQ2	0.51	0.37	-	0.62	0.52 (1930)	n/s	-	-	-
LMNA	0.62	1	-	0.57	0.78 (1678)	n/s	-	-	-
OBSCN	11.07	-	-	-	0.00 (0)	-	-	-	-
PDLIM3	0.51	0.87	-	0.41	0.62 (1771)	n/s	-	-	-
SRI	0.21	-	-	0	0.00 (252)	no excess	-	-	-
TRIM54	0.51	0.25	-	2.02	1.31 (993)	*	-	-	-
VCL	1.02	1	-	0.84	0.89 (2358)	no excess	-	-	-

Table 2.12: Summary of the evidence for 31 non-sarcomeric genes implicated in HCM. Frequencies of rare variation in ExAC and HCM cohorts (prospective, clinical, other published and combined) are shown, along with Fisher's exact p-value for combined cohorts versus ExAC (n/s = not significant, * = $P < 0.05$, ** = $P < 0.0016$ correction for multiple testing threshold, *** = $P < 0.0001$). Also highlighted are the maximum LOD scores for published pedigrees and

whether de novo inheritance or the excess of individual variants in cases over controls has been observed.

- The X-linked *FHL1* gene (desmosomal gene) had an excess of 0.80% based on a combined cohort of 2061 cases. A confirmed *de novo* truncating variant, c.134delA, was also detected in a male patient diagnosed as a young adult¹¹³. Additionally the variant p.N38S was significantly enriched in cases (4/2061) versus ExAC (0/60706) ($p < 1 \times 10^{-4}$).
- *PLN* (calcium signalling) had an excess of 0.28% based on a combined cohort of 5440 cases. The nonsense variant p.L39X was also significantly enriched in cases (4/5440) versus ExAC (1/60706) ($p = 2 \times 10^{-4}$).

Of the four genes classified as having moderate evidence, variants in two Z-disc genes have been reported to segregate in relatively large pedigrees even though no overall excess of rare variants across the genes was observed.

- In *ACTN2*, the variant p.A119T was reported to be present in 23 members of a 3-generation family, including 7 affected, with a reported LOD score of 2.82¹¹⁴. In another report, the variant p.M228T was detected in 18 members of a 4-generation family, including 11 affected (5 with LV hypertrophy), with an estimated LOD score of 2.1¹¹⁵. However, the overall case frequency of 1.12% (total of 2779 cases) was similar to the ExAC frequency of 1.09% ($p = 0.851$), indicating that most *ACTN2* variants detected in HCM cases will represent background, benign variation.
- In *MYOZ2*, the variant p.S48P was detected in 6 affected and absent in 4 unaffected members of a 3 generation family (LOD=2.03)¹¹⁶, with the gene initially identified through a linkage study and prioritised for sequencing from 30 genes at the locus based on cardiac expression and function. A follow-up study assessing the effect of this variant in mouse models demonstrated that the mutant mice showed molecular,

cellular, and gross cardiac hypertrophy, preserved systolic function, interstitial fibrosis and mal-aligned Z disks¹¹⁷. However this effect was also observed with a second variant, p.I246M, identified through screening of 516 HCM patients. As this variant is present in 230 individuals in ExAC, and at a frequency of 0.025 in Finns, it is highly unlikely to be disease-causing, casting doubt on how well the mouse model data translates to human clinical effect. Only two rare *MYOZ2* variants were observed in a total of 2390 cases, with the frequency of 0.08% not in excess of the ExAC frequency of 0.26%.

The other genes with moderate evidence were:

- *CRYAB*: Although no excess of rare variation is observed in the *CRYAB* gene (0.25% in cases versus 0.21% in ExAC), a truncating variant, p.Q26X, was reported as a *de novo* variant in a paediatric HCM patient who went on to die of SCD at 17⁴². A different nonsense mutation, p.E99X, was identified in a HCM patient in our prospective cohort which was also present in his affected son. Truncating variants in *CRYAB* have previously been associated with myofibrillar myopathy and muscular dystrophy (with infantile onset in recessive cases) and posterior polar cataract, while a read-through variant has been linked to a family with DCM and congenital posterior pole cataract. They are rare in the general population, with only one nonsense/frameshift variant in ExAC. Although evidence is currently limited, this data suggests truncating variants in *CRYAB* may be a rare cause of HCM.
- The Z-disc gene *FLNC* is one of the most recently implicated HCM genes, with a report in 2014 identifying a variant, p.A1539T, through whole exome sequencing that segregated in a family - present in 4 affected (and 3 obligate carriers) and absent in 5 affected members, with a LOD score of 2.33¹¹⁸. This study also sequenced a small HCM cohort of 92 patients and detected 8 further rare *FLNC* variants for a frequency of 8.7% – this compared to a frequency in ExAC of 3.2% ($p = 0.008$, though not significant after Bonferroni correction for multiple testing). However a follow-up study from the same group extended their analysis to 448 HCM patients (including the 92 sequenced previously) and detected only 22 rare variants for a frequency of 4.9% ($p = 0.059$)¹¹⁹.

While the current evidence suggests some *FLNC* variants may be causative, the high background rate of rare, benign variants will make variant interpretation in this gene challenging (see Chapter 3).

Four genes have been classified as having weak evidence, three of which are limited to case excess data only.

- Rare variants in *MYH6* were detected in 3.8% of cases versus 2.5% in ExAC ($p = 0.001$), though the frequency was lower in our prospective cohort (3.4%) than the extended gene panel study described above⁸⁶ (4.2%). Given the high background rate (2.5%), clinical interpretation of *MYH6* variants is likely to be extremely challenging. As *MYH6* is more highly expressed in atria than ventricles in human heart (Figure 2.10), its role as a causative HCM gene is questionable, even if the case excess observed here is replicated in further cohorts.
- The *TRIM63* and *TRIM55* genes, encoding muscle ring finger proteins MuRF1 and MuRF2 (ubiquitin E3 ligases that modulate cardiac mass), have been proposed to act as modifiers of HCM⁵⁰. An excess of variants in these genes was observed in a Chinese HCM cohort, with presence of such variants associated with an earlier age of onset and greater maximum LVWT. As 56% of patients with *TRIM63/TRIM55* variants also had causative sarcomeric mutations, the authors speculated that these variants may act to modify the expression of the HCM phenotype in patients. The frequency of rare variants in our prospective cohort was substantially lower than the published report (1.0% vs 2.0% and 1.5% vs 2.4%, respectively) and not significantly in excess over ExAC, though a significant excess was maintained for the combined data. The differences between the published and prospective cohorts may be due to the different populations investigated (Chinese/East Asian versus predominantly Caucasian). Further studies will be needed to definitively ascertain any role of these genes in HCM.

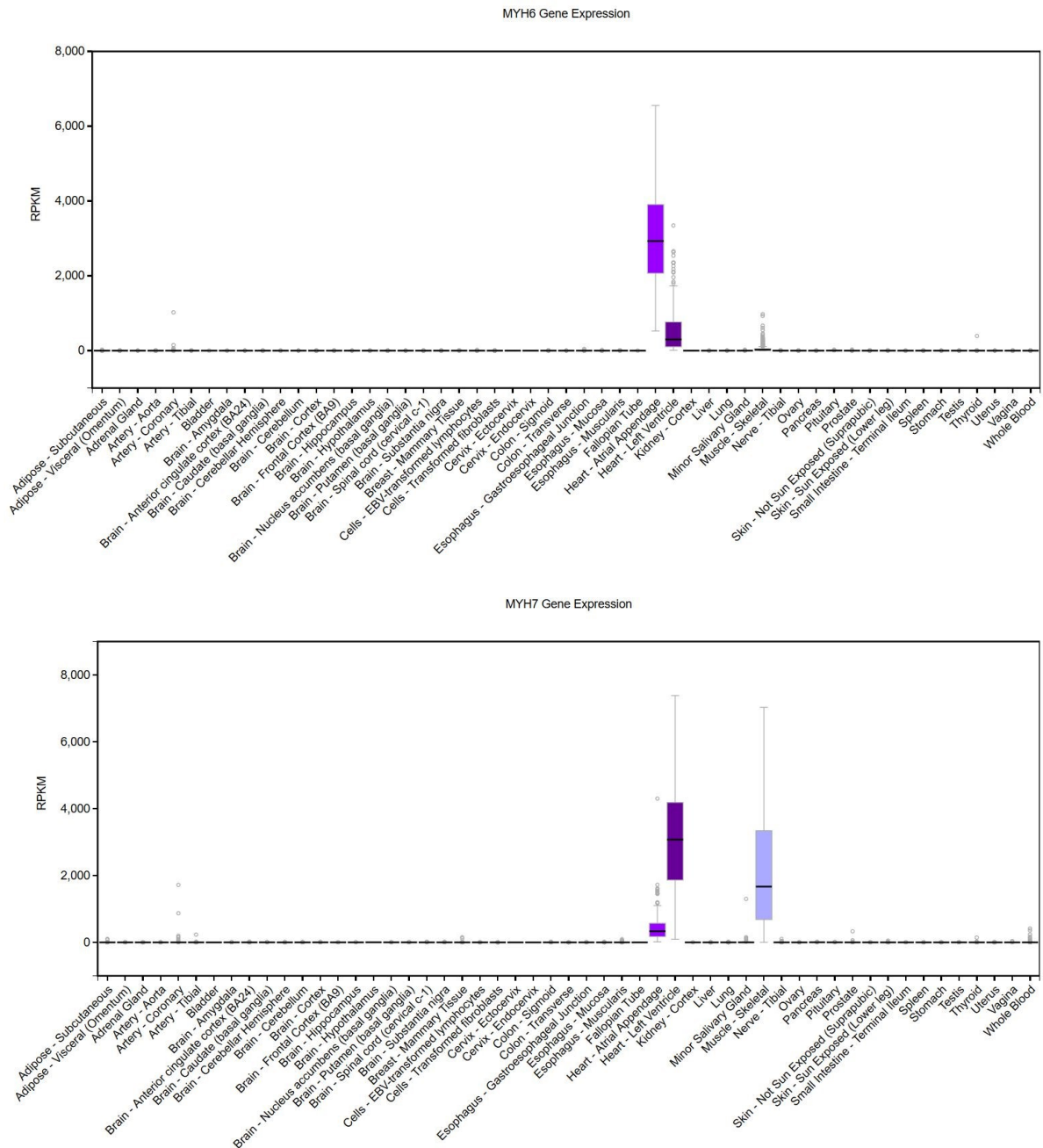


Figure 2.10: Comparison of the expression levels of MYH6 and MYH7. In the human heart MYH6 is predominantly expressed in atrial muscle (detected here using atrial appendage) whereas MYH7 is expressed in ventricular muscle (and shown here for left ventricle). Data from the GTEX portal - <https://gtexportal.org>.

- *TNNC1* is a sarcomeric gene whose role in HCM has been investigated in several studies and which is often included in HCM gene panels (Table 2.2). However the evidence as a causative gene in HCM remains limited. Variants in *TNNC1* are extremely rare, occurring in only 0.24% of patients in a combined cohort of 3335 cases. This is still in excess over ExAC (0.06%), although this does not reach statistical significance with multiple testing correction ($p=0.002$). Data from individual patients and families is very limited with a maximum LOD score of only 1.2¹²⁰. Given the paucity of *TNNC1* variants in the general population and the fact that it encodes a troponin protein that is a key part of the sarcomere, it is conceivable that *TNNC1* does play a role in HCM. The sequencing of additional large HCM cohorts and detailed investigation of *TNNC1* carriers and their families will be need to confirm this role.

2.3.4.5 Genes with functional data only

Several studies that have implicated genes in HCM have demonstrated functional effects of variants detected in HCM patients in various cellular and animal model assays. For 12 genes (*ANKRD1*, *CAV3*, *FHL2*, *FXN*, *JPH2*, *KLF10*, *LDB3*, *MYLK2*, *MYOM1*, *MYPN*, *NEXN*, *TCAP*), this functional data was not accompanied by any clear human genetic evidence associating these genes to HCM, i.e. no significant excess of variation (across genes or for single variants) and no evidence of segregation (LOD > 1) or *de novo* inheritance (Table 2.12).

As an example, rare variants in the *MYPN* gene (encoding the myopalladin protein) have been detected in 1.5-2.0% of HCM patients. The primary study assessing a role for this gene in HCM developed cardiac-restricted transgenic mice expressing the p.Y20C variant, which had been detected in one of the 484 HCM patients sequenced in this report¹²¹. These mice developed a HCM phenotype suggesting this variant may indeed be causative for HCM. However this variant occurs in 111 individuals in ExAC (MAF = 9×10^{-4}) which is too common to be a primary pathogenic variant, indicating that data from animal models such as these may not be fully translatable to human genetics. Without a significant overall excess of rare variants and other human genetic evidence, *MYPN* cannot currently be classified as a causative HCM gene.

2.3.4.6 *Genes with no convincing evidence for association with HCM*

For 8 genes, there was currently no evidence (human genetic or functional) to support any role in HCM (Table 2.12). These included genes where the reported variants were too common in ExAC ($MAF > 1 \times 10^{-4}$) to be pathogenic (*LMNA*, *OBSCN*, *SRI*) or genes where the overall frequency of rare variants was similar in cases and ExAC (*CALR3*, *CASQ2*, *PDLIM3*). Reports into the role of *VCL* (encoding the vinculin protein) in HCM found that patients with *VCL* variants had a marked reduction in vinculin protein levels in intercalated discs of their myectomy samples^{122,123}. However, a subsequent report indicated the reduced expression of vinculin may be a general feature of HCM, regardless of the underlying genotype¹²⁴.

2.4 Discussion

Advances in NGS technology that have reduced the cost and time (while increasing yield) for DNA sequencing have transformed the application of genetic testing in Mendelian diseases like HCM. While clinical testing is now well established for HCM, enabling the identification of disease-causing variants in a substantial subset of cases (and cascade screening in the patients' families), these advances have also empowered researchers to explore novel gene to disease associations that may explain the apparent missing heritability in HCM. These candidate gene studies discovered numerous plausible novel genes for HCM but, with limited replication studies and questionable clinical impact³³, the exact role of these genes in HCM was uncertain. Now, by combining large-scale case-control burden analysis with a systematic review of published evidence, I can conclude that the non-sarcomeric genes that have been implicated in HCM are either very rarely causative or, in most cases, are likely to have been wrongly associated with the disease.

2.4.1 *Why did the results from candidate gene studies fail to validate?*

In retrospect, we can identify why these candidate gene studies have largely failed to discover valid disease genes. The studies were often poorly controlled for variation in healthy or population cohorts – variants detected in cases were generally assessed for absence in several hundred control subjects. As described in this chapter's Methods on defining a rare variant threshold for cardiomyopathy burden analysis, and has been illuminated by the ExAC database, significantly larger control or population databases are required to prove that a variant is rare enough to be plausibly pathogenic for a Mendelian disease like HCM. Even when truly rare variants were detected in cases, the studies seldom controlled for rare variation across the entire gene in control cohorts, i.e. they did not prove that rare variation in a gene was enriched in cases. As databases like 1KG, ESP and in particular ExAC have shown, rare variation in the general population is much more prevalent than previously recognised and is collectively quite common for many genes. Therefore, the detection of a rare variant in a HCM patient, even in a gene with a plausible role in the disease (or one that has been implicated in other cardiac conditions) is insufficient evidence alone to prove a role for that gene.

Secondly, most of these studies have not provided sufficient additional evidence to establish a causal role in HCM for the detected variants. With few exceptions, segregation analysis in the affected families was either not reported or was performed in a pedigree of an insufficient size to generate a LOD score that would verify a role for the variant in disease (generally greater than 2 – 3). Instead, these studies often attempted to validate findings with various functional assays, ranging from cell lines to *in vivo* models. However, there is significant doubt as to how an effect seen in a functional assay will translate into a disease-causing effect in a patient. Additionally, these assays were generally not controlled with known benign or common variants to demonstrate that the observed effect was specific to the putative pathogenic variant. Indeed in a number of studies, clear functional effects were observed with variants that are too common to be disease-causing, e.g. *MYOZ2*:p.I246M (ExAC MAF=0.0019). This highlights how the uncertain translatability of functional assays, particularly those not well established or validated for specific diseases, makes them unsuitable for use as key evidence to demonstrate a role for novel genes in disease.

For the genes with no excess in cases, particularly those with reports demonstrating some functional effect of variants detected in HCM patients, it could be argued they may act as potential modifiers of HCM disease expression (as claimed for variants in the *TRIM63* and *TRIM55* genes⁵⁰). For any such role to be verified, the effect of such variants would need to be assessed either at the cohort level (see Chapter 4) or in large family pedigrees, by evaluating whether carriers of putative modifier variants have more severe phenotypes (or less severe in the case of putative protective modifier variants) than non-carriers. However, given the lack of enrichment of such variants in cases and limited evidence to currently support such a role, the more parsimonious explanation is that such genes do not contribute to phenotype in HCM.

2.4.2 How reliable are the gene burden tests performed here?

A clear limitation of the analyses described in this chapter are the potential confounding factors when comparing genetic data between cohorts from differing settings, studies and sequencing platforms. An ideal study design would involve clinically confirmed

cardiomyopathy and healthy control individuals, sequenced on the same platforms and panels yielding accurate, high coverage data, controlled for population stratification (e.g. all confirmed Caucasian samples) with sufficient samples to accurately gauge the level of rare variation even in highly constrained genes. As such datasets are clearly not currently available and are unlikely to be available for the foreseeable future, we have had to utilise the datasets that are available for this study.

As ethnicity data was not available for many of the cardiomyopathy cohorts used in this study, the main analysis was performed using all samples from both cases and ExAC. With differing ethnic composition of case and control cohorts, this is a potential confounding factor for burden analysis, as population-specific or enriched variants may affect overall gene frequencies. However, a number of factors suggest that population stratification is not a significant issue for the results of the analyses presented here. The fact that we have used a more stringent threshold for the definition of rare variants ($< 0.01\%$ in ExAC) than most other studies means that the effect of any population-specific variants will be limited when assessing rare variant frequencies across genes. Additionally, the ethnic composition of the ExAC dataset (60% European) is expected to be broadly similar to that of the clinical cardiomyopathy cohorts - UK and US populations that are majority Caucasian but with substantial minority of other ethnicities (the LMM DCM cohort was 61% white). To confirm that the ethnic composition of the cohorts was not a confounding factor in this analysis, we compared results from the burden analysis of all samples with Caucasian cases versus ExAC-NFE for the LMM DCM cohort (the only clinical cohort for which we had self-declared ethnicity data) and found no significant differences (Figure 2.7 and Table 2.9).

The other potential confounding factor is differing sequencing coverage between the ExAC and case cohorts. It is expected that the coverage from the WES data that comprises the ExAC database will be poorer than the gene panel data from our prospective cohort or the clinical laboratories (as diagnostic genetic services require full coverage of tested genes)⁶³. To control for this, the gene frequencies for the ExAC data were adjusted based on the mean coverage across each gene. It should also be noted that for most of the genes assessed in this study, no excess was observed in cases over ExAC – if there is some residual effect from the lower WES

coverage in ExAC, it would mean we were actually underestimating ExAC gene frequencies and therefore would not affect the results for these genes.

2.4.3 What does lack of excess mean and how should this data inform clinical genetic testing?

For most of the HCM genes in which no excess of rare variation was observed in cases versus ExAC, there was very limited or no other human genetic evidence to support a role in the disease. This does not imply that these genes are definitively not HCM disease genes – as with *ACTN2* and *MYOZ2*, it is possible that particular variants in these genes can cause disease and would segregate in families, even if there is no overall excess and most detected rare variants would be benign. However, we can make a number of assumptions about such genes and the role they could have in HCM clinical genetic testing.

With no excess in large case cohorts, variants in these genes are at best likely to be very rare causes of HCM as most detected variants will be the background, benign variation that would be observed in any disease or control cohort. Given that there are no known variants in these genes that would be classified as pathogenic by clinical genetic guidelines, and the lack of excess means there is no prior probability that a rare variant in these genes will be pathogenic (as described in Chapter 3), we can also conclude that variants in these genes are essentially uninterpretable. Applying the ACMG guidelines to any variant identified in a HCM patient can only yield a classification of VUS at best, regardless of the variant's rarity in population databases or its predicted effect by computational algorithms. Therefore, the recommendations of this study are to currently restrict standard genetic testing for HCM to those genes validated for disease and likely to yield interpretable data – the core sarcomeric genes (*MYBPC3*, *MYH7*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, *ACTC1*), the novel though rarely associated genes (*CSRP3*, *FHL1*, *FLNC*, *PLN*, *TNNC1*), the genes for which a subset of variants may be actionable (*ACTN2*, *CRYAB*, *MYOZ2*) though variants should be assessed cautiously and genes associated with mimics of HCM (*GAA*, *GLA*, *LAMP2*, *PRKAG2*, *TTR* and possibly genes associated with Noonan syndrome¹²⁵).

The only exception to this point is if segregation analysis is available for the family of the patient being tested, with a large enough pedigree (and well defined clinical status for several family members) to generate a statistically significant LOD score, assuming the variant segregates with disease. However, such pedigrees are rarely available in routine clinical genetic testing for diseases like HCM, particularly with large enough families to generate LOD scores that will upgrade a variant from VUS to Likely Pathogenic. This approach would also negate the primary purpose of HCM genetic testing - to facilitate predictive testing of family members. Nevertheless if such families were available, a tiered approach to genetic testing would offer a more sensible solution to detecting the causative variant. If the standard HCM test (with the validated genes described here) was negative, then proceeding to a whole exome or whole genome scan would offer a far greater likelihood of success than extending the analysis to a few dozen genes with limited or no evidence of involvement in HCM. For most genotype-negative HCM patients however, extended pedigrees with affected family members will not be available – as described in Chapter 4, such cases are much more likely to have no family history of disease and indeed lack any dominant, causative Mendelian variant. For rare recessive cases of HCM, i.e. an affected child with two unaffected parents, WES or WGS is the strategy of choice for detecting the causative variant as results can be effectively triaged to identify homozygous, compound heterozygous or *de novo* variants, e.g. the recent discovery of homozygous truncating *ALPK3* variants in paediatric cardiomyopathy¹²⁶.

It could be argued that retaining these genes in HCM clinical gene panels will actually do no harm. If ACMG guidelines are correctly applied, then no variant in these genes should be incorrectly classified as actionable (Pathogenic or Likely Pathogenic) and therefore there is no risk of a false positive result that could make incorrect inference about the clinical status of family members. However, as described in Chapter 3, the ACMG guidelines can be vague and open to interpretation, with no guarantee that all clinical genetics laboratories, especially those less experienced in cardiomyopathy genetics, will utilise them correctly and avoid erroneous interpretation of variants. As an example, a recent study applying WES to ICCs classified truncating variants in *TRPM4* and *ABCC1* as Pathogenic in HCM patients¹²⁷, citing the rule ACMG rule PVS1 (*predicted null variant in a gene where loss of function is a known mechanism of disease*) despite the fact that truncating variants in these genes are not a

known mechanism for HCM. In addition, the mere presence of VUS on clinical genetics reports may be confusing to both clinicians and patients, creating further uncertainty and the belief or expectation that these may actually be causative variants for which insufficient evidence of pathogenicity exists to be currently classified as causative¹²⁸ (despite the fact that the ACMG guidelines state that “*a variant of uncertain significance should not be used in clinical decision making*”).

2.4.4 The missing heritability of HCM

Studies to identify novel non-sarcomeric genes that are causative in HCM were driven by the attempt to explain the apparent missing heritability in HCM, as observed by the large proportion of cases without pathogenic variants in sarcomeric genes. As described above, it is expected that the limited number of likely valid non-sarcomeric HCM genes that have been identified will contribute to only a very small proportion of cases, estimated as approximately 2%. I will argue in Chapter 4 that most of the remaining sarcomeric-negative cases are likely in fact to represent non-genetic forms of the disease, lacking dominant causative variants and Mendelian inheritance in families. However, for the minority of cases that lack sarcomeric variants but have a family history of HCM or SCD, it is possible that as yet undiscovered genetic factors may have primary causative roles in HCM.

The identification of novel genetic causes of HCM will require more robustly designed studies than many of the previously published candidate gene approaches. Unless additional large family pedigrees are available for linkage analysis or whole genome sequencing, these will involve case-control analyses and are likely to require very large case cohorts to demonstrate statistically significant effects, as novel factors will probably only contribute to a small proportion of cases. Collaborations such as the SHaRe Cardiomyopathy Registry¹²⁹ will be essential for generating such cohorts. While different classes of genetic variation are potential causes of HCM, each present substantial challenges for the validation of any novel associations.

Protein-coding variants in novel genes are perhaps still the most likely candidate for new HCM genetic factors, given the fact that this variant class is responsible for the vast majority of known variation underlying Mendelian disease. However identifying novel genes through unbiased case-control analysis using WES or WGS will prove to be extremely challenging – with the likely low prevalence of any novel genes and the need to correct for multiple testing in over 20,000 genes, very large case cohorts will be required to detect a statistically significant association. More candidate-based approaches, that integrate family pedigree analysis with cohort sequencing (such as the initial study implicating *FLNC* in HCM¹¹⁸), may increase the likelihood of discovering novel disease genes, though care will need to be taken to avoid the pitfalls of the candidate gene studies described in this chapter.

It is possible that different classes of variants affecting known (or indeed novel) HCM genes may also contribute to disease burden. CNVs that lead to the deletion or duplication of genes (whole genes or large regions of genes) have been implicated in many genetic diseases, though they are hampered by the difficulties in detecting, validating and interpreting such data. There have been limited studies thus far assessing CNVs in HCM, with results suggesting they may play only a very minor role in the disease. 4 validated CNVs were found in 505 cases (0.8%) by Lopes et al.¹³⁰, though two of these affected genes with limited evidence for involvement in HCM (*LMNA* and *PDLIM3*). 4 CNVs were also detected in a cohort of 303 cases (1.3%) by Mademont-Soler et al.¹³¹ and these did involve deletions in genes where loss of function is a validated mechanism for HCM (*MYBPC3* and *PLN*). Ceyhan-Birsoy et al. detected 4 CNVs in 708 HCM cases (0.6%)¹³², involving an intragenic deletion in *MYBPC3*, a whole gene duplication in *NEXN*, an intragenic duplication in *MYOZ2* and a trisomy X potentially affecting *EMD*, *GLA*, *LAMP2* and *TAZ*. These studies indicate that CNVs are likely to explain only a very limited proportion of cases with negative results from standard genetic testing. As HCM is a disease for which known loss of function or haploinsufficiency mechanisms account for only a small proportion of cases, it is perhaps unsurprising that CNVs appear to occur in so few cases. The detection of CNVs may be more relevant to cardiomyopathies like DCM and ARVC where these mechanisms play a larger role. Indeed, a recent report detected CNVs in 5 genes in 6.9% of 160 genotype-negative ARVC cases¹³³. Interpreting the effect of CNVs remains

challenging for gene or exon duplications and for CNVs affecting genes not fully validated in the disease.

It remains to be determined whether other variant classes (such as those involving regulatory or RNA sequences, or variants in exonic or deep intronic sequence that may affect splicing and can be detected through transcriptome sequencing¹³⁴) have a role to play in explaining some of the missing heritability of HCM. The difficulties in interpreting such variants, that have rarely thus far been implicated in Mendelian diseases, means they will require innovative experimental design and functional validation studies to identify and define their role in HCM.

2.4.5 Comparison with ClinGen clinical validity framework

The Clinical Genome Resource (ClinGen) is an NIH-funded initiative that aims to “*define the clinical relevance of genes and variants for use in precision medicine and research*”¹³⁵. A key aspect of this is to establish frameworks to assess the evidence of gene-disease relationships that will enable an informed approach towards the composition of genetic testing panels for Mendelian diseases. The proposed framework was published in June 2017 with the classes of evidence and relative weights summarised in Figure 2.11⁹⁷. In contrast to the classification system described here (Figure 2.5), the ClinGen framework adopts a scoring system allowing for 0 – 18 points to be assigned for a gene-disease relationship, yielding classifications of Limited (1-6 points), Moderate (7-11 points), Strong (12-18 points) and Definitive (12-18 points and replicated over time). A range of points is defined for each evidence class, allowing adjustment based on the strength of evidence, though perhaps at the risk of inconsistency between different classifications and curators.

Unlike the classification system I have proposed here, a maximum of 6 points is allowed for experimental evidence in the ClinGen framework, including up to 2 points for basic gene/protein function, i.e. relevant biochemical function, interaction with other proteins involved in the disease and expression in the appropriate tissue. Most of the genes implicated in HCM that I have shown to have little or no human genetic evidence would score points

Case-Level Data ^A	Evidence Type		Case Information			Suggested Points/Case		Points Given	Max Score
						Default	Range		
	Variant Evidence	Autosomal Dominant OR X-Linked Disorder ^B	Variant is <i>de novo</i> ^C			2	0-3		12
			Proband with predicted or proven null variant ^D			1.5	0-2		10
			Proband with other variant type with some evidence of gene impact ^E			0.5	0-1.5		7
		Autosomal Recessive	Two variants in <i>trans</i> and at least one <i>de novo</i> ^C or a predicted/proven null variant ^D			2	0-3		12
			Two variants (not predicted/proven null) with some evidence of gene impact ^E in <i>trans</i>			1	0-1.5		
	Segregation ^F Evidence	Evidence of segregation in one or more families	LOD Score Examples	3	5	0-7		7	
				2	4				
				1.5	3				
1				1.5					
Case-Control Data ^G	Case-Control Study Type ^H	Case-Control Quality Criteria ^I			Suggested Points/Study		Points Given	Max Score	
	Single Variant Analysis ^{Ha}	• Variant Detection Methodology ^{Ia} • Power ^{Ib} • Bias and Confounding Factors ^{Ic} • Statistical Significance ^{Id}			0-6			12	
	Aggregate Variant Analysis ^{Hb}				0-6				
TOTAL ALLOWABLE POINTS for Genetic Evidence									12

Evidence Category	Evidence Type	Suggested Points		Points Given	Max Score
		Default	Range		
Function	Biochemical Function	0.5	0-2		2
	Protein Interaction		0-2		
	Expression		0-2		
Functional Alteration	Cells from affected individual	1	0-2		2
	Engineered cells	0.5	0-1		
Models & Rescue	Animal model	2	0-4		4
	Cell culture model system	1	0-2		
	Rescue in animal model	2	0-4		
	Rescue in engineered equivalent	1	0-2		
Total Allowable Points for Experimental Evidence					6

Figure 2.11: Classes of genetic evidence (top) and experimental evidence (bottom) in the ClinGen clinical validity framework. Figures from Strande et al.⁹⁷ (reused with permission).

based on these gene functional criteria – indeed these were the very reasons they were proposed for the candidate gene studies. The ClinGen guidelines do state that experimental evidence alone cannot yield a classification of Limited or above and that “*at least one human genetic variant with a plausible causal association must be present*”. However, the evidence required to demonstrate this “plausible” association lacks clarity and stringency (defined as an impact on gene function, recurrence in affected individuals and not too common in population databases), particularly as segregation evidence is considered separately (Figure 2.11). As a consequence of this scoring system, it is likely that many of the HCM genes shown here to have no (or only functional) evidence will be classified as “Limited evidence” by the ClinGen framework, despite having no clear genetic proof (case excess or segregation data) for disease association. Genetic labs seeking to design comprehensive test panels may choose to retain or include genes with this “Limited evidence” classification, with implications for the accuracy of genetic testing and the generation of ambiguous or erroneous results.

2.4.6 DCM gene burden data

Although the primary focus of this project is the genetics of HCM, I have included the DCM gene burden analysis as this was performed as part of the same study. DCM is regarded as a more genetically heterogeneous than HCM, in terms of the number of genes that have been implicated in the condition and the size of the gene panels used for clinical genetic testing (as demonstrated in the OMGL and LMM data). However, only 10 genes displayed a significant excess of rare variants in DCM cases over ExAC (truncating variants in *DSP*, *LMNA*, *SCN5A*, *TTN* and *VCL* and non-truncating variants in *LMNA*, *MYH7*, *TCAP*, *TNNC1*, *TNNT2* and *TPM1*).

It should be noted however that the number of cases sequenced for many of the genes in these cohorts was much more limited than for the HCM genes (including 12 genes with data for only 121 – 156 cases), due to the more recent development of clinical genetic testing for DCM. Larger case cohorts will be needed to more accurately assess whether there is a significant case excess for such genes – we are currently investigating such a dataset comprising an extended clinical cohort from LMM and a research cohort of 1,040 DCM cases

compiled at the Royal Brompton Hospital London and National Heart Centre Singapore and sequenced on the TruSight Cardio panel.

Some of the DCM genes without an observed case excess may be relevant to specific population groups due to the presence of founder mutations, e.g. the *PLN*:p.Arg14del variant is present in over 10% of DCM cases in the Netherlands¹³⁶. For other genes, a detailed re-assessment of the evidence implicating them in DCM, similar to the analysis performed here for HCM, will be required to definitively prove whether these are valid disease genes or have been erroneously associated with the condition. Given the number of reports to evaluate, this task is best suited to collaborative efforts like ClinGen. However, based on our findings from the HCM analysis, it is likely that many of these reports will have been based on inadequately controlled study designs, with incidental findings of rare variants in DCM patients not supported by convincing genetic evidence.

2.4.7 Application to other Mendelian diseases

The issues described here for cardiomyopathies – implicating genes in these conditions based on poorly designed and controlled studies – are likely to be just as prevalent across other Mendelian diseases. As described above, large panels of putative disease genes have recently been sequenced in HCM⁸⁶ and DCM⁸⁷ cohorts, implicating variants in numerous genes but without testing for a statistical enrichment of variants in these genes. A search of the HGMD database can quickly identify similar studies across a range of other Mendelian diseases – including deafness¹³⁷, retinal dystrophy¹³⁸ and bicuspid aortic valve¹³⁹. This implies that erroneous gene to disease associations, and the repercussions for clinical genetic testing and understanding of the genetic architecture of these conditions, is a serious and widespread issue across the spectrum of genetic disease. Evaluating these associations with techniques like those described in this chapter, and initiatives such as ClinGen, will be essential to clarify the aetiology and genetics of these conditions. A crucial aspect of this will involve publishing and sharing genetic cohort data like that of OMGL and LMM, particularly for rarer and less well studied genetic diseases.

2.5 Conclusion

The failure to validate most of the non-sarcomeric genes that have been implicated in HCM emphasises the need for a rigorous statistical approach when asserting associations between genes and Mendelian diseases. This will be ever more critical as genetic sequencing becomes a standardised part of the clinical management of genetic diseases and WES or WGS are increasingly used as the primary sequencing assay. Given the prevalence of rare variation in the population for most genes, and the non-specific nature of many variant assessment techniques (from computational algorithms to functional assays), there is a strong temptation to over-interpret plausible sounding associations like those described here for both researchers and clinical geneticists. It is therefore essential that knowledge and appreciation of the statistical evidence required to validate such associations becomes entrenched in those working in this field, including journal editors and reviewers, to prevent further publication of spurious genetic associations.

3 Variant interpretation in HCM

In this chapter, I assess the challenges of variant interpretation in HCM and explore methods to enhance both the specificity and sensitivity of genetic testing for this disease. In this context, variant interpretation refers to the identification of disease-causing variants in validated HCM genes and distinguishing these from rare benign variants. This is distinct from validating variants in novel and putative HCM genes as described in Chapter 2, though clearly there will be much overlap in the techniques used to assess these variants. Pathogenic variants for diseases like HCM are assumed to be primary Mendelian variants necessary for phenotypic expression in families being studied (though allowing for variable penetrance that is characteristic of HCM), and do not refer to putative modifier variants that may alter the expression of disease in the presence of a primary mutation (there has been very little research on such putative variants in HCM to date). The accurate identification of pathogenic variants is one of the key challenges for genetic testing in HCM and one of the limiting factors towards the comprehensive adoption of genomics as part of the clinical management for such diseases.

3.1 Introduction

3.1.1 Early approaches to assessing variant pathogenicity

In the decade following the discovery that variants in sarcomeric genes were causative for HCM, these genes began to be routinely sequenced in patients and small research cohorts, initially by dHPLC and Sanger sequencing and subsequently with NGS. The following HCM cohorts (with a minimum of 100 patients sequenced) were published between 2000 and 2010 – a 389 patient cohort from USA (2003)^{73–75}, a 197 patient cohort from France (2003)¹⁴⁰, a 108 patient cohort from Germany (2003)¹⁴¹, a 100 patient cohort from China (2005)¹⁴², a 203 patient cohort from Italy (2008)¹⁴³, a 120 patient cohort from Spain (2009)¹⁴⁴ and a 192 patient cohort from France (2010)¹⁴⁵.

The interpretation of variants detected in these studies was hampered by the lack of the following:

- Large-scale population datasets to assess the frequency of individual variants and the overall level of rare variation in the population.
- Extensive catalogues of well characterised pathogenic variants, particularly with clinical grade classification.
- Standardised methods and protocols for the assessment of variants.

In the absence of these, such studies used some or all of the following techniques to assess the likely effect of variants detected in sarcomeric genes:

- Variants found in cases were generally tested for absence in small, ethnically-matched control cohorts (typically between 100 and 300 individuals). However, while this approach could detect common polymorphisms in the populations being studied, they were far too small to confirm that case variants were rare enough to be plausibly pathogenic (see Methods below for how such thresholds should be calculated).
- The predicted effect of the variants was assessed by noting the variant class (i.e. whether the variant was synonymous or altered the protein product with missense or truncating variation), exploring the conservation of affected amino acids of missense variants and for later studies the use of computational functional prediction algorithms such as SIFT and Polyphen. While these techniques may identify benign variants that are unlikely to be pathogenic, their specificity is limited.
- The segregation of detected variants with disease in family members was assessed where possible. However, details of these family pedigrees were rarely provided and pedigrees of sufficient size to definitively associate variants with HCM were not available for the vast majority of cases studied.

While the majority of the associations in these studies would not meet clinical grade classification of disease-causing variants, the extent of false positives remained uncertain until the publication of large population genetic datasets.

3.1.2 *Insights from population genetics studies*

As described in Chapter 2, the release of data from population genetics studies such as the 1KG project (2,504 genomes from a diverse range of population groups) and ESP (6,500 exomes from well-phenotyped cohorts of European-American and African-American ancestry) demonstrated that rare variation in the general population was collectively more common than previously recognised. Subsequent reports highlighted that many variants previously implicated in Mendelian disease were not only present in these databases, but were often at frequencies incompatible with being a penetrant pathogenic variant for rare genetic diseases like cardiomyopathies^{84,85}. A number of conclusions could be drawn from these findings:

- The traditional technique of assessing variant rarity by demonstrating absence in healthy control cohorts of several hundred individuals is inadequate for most genetic diseases.
- Unlike defined small cohorts of healthy controls, population databases may include individuals with genetic diseases and therefore the presence of a particular variant at low frequencies in these databases does not necessarily preclude a variant from being disease-causing.
- The frequency of rare benign variation for many genes is much higher than previously understood, suggesting that new datasets and methods would be required to definitively ascertain whether variants detected in patients are sufficiently rare to be plausibly pathogenic.
- With the amount of rare benign variants likely to be detected in patients, especially as gene panels increase in size, new and more stringent approaches to variant interpretation for clinical genetic testing were required to prevent or minimise false positive results.

3.1.3 ACMG Guidelines

Standards and guidelines for the interpretation of sequence variants were produced jointly by the American College of Medical Genetics (ACMG) and the Association of Molecular Pathology in 2015 (henceforth known as the ACMG guidelines)⁷¹. The aims of these guidelines were to standardise methods for variant interpretation, improve the consistency of analysis between laboratories, describe how different lines of evidence should be used and define the level of evidence required to denote a variant as actionable or disease-causing. The guidelines describe a series of rules for pathogenic and benign evidence with different strengths (Figure 3.1) and five levels of variant classification – Benign (B), Likely Benign (LB), Uncertain Significance (VUS), Likely Pathogenic (LP) and Pathogenic (P) – depending on the number and type of rules activated for the specific variant (Figure 3.2).

	Benign			Pathogenic		
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 3.1: Evidence framework for the ACMG guidelines highlighting the different evidence classes and strengths. Reproduced from Richards et al.⁷¹, with permission.

Pathogenic					
VS S	VS M/P M/P	S S	S M M M	S M M P P	S M P P P P
Likely Pathogenic					
VS M	S M	S P P	M M M	M M P P	M P P P P
Likely Benign					
S P	P P				
Benign					
BA1	S S				

VS

 Very Strong (PVS1)

S

 Strong (PS1-PS4)

M

 Moderate (PM1-PM6)

P

 Supporting (PP1-PP5)

BA1

 Stand-alone (BA1)

S

 Strong (BS1-BS4)

P

 Supporting (BP1-BP7)

Figure 3.2: Combinations of ACMG rules required to achieve classifications of Pathogenic, Likely Pathogenic, Likely Benign and Benign (minimum number of rules required are shown). A variant is classified as of Uncertain Significance if the criteria shown here are not met or if both benign and pathogenic criteria are met. The combinations shaded in grey may be incorrectly classified based on Bayesian classifier analysis that assumed the four levels of evidence displayed exponentially scaled odds of pathogenicity (abstract by Leslie Biesecker, Curating the Clinical Genome meeting, June 2017).

There are a number of important aspects to note about the ACMG guidelines. Firstly, they are designed to be a broad framework for use across all Mendelian disease (dominant and recessive), with the intention that disease-specific adaptations will be developed within this framework by domain specialists. Thus, the guidelines and rules are intentionally described and defined in a vague manner. Secondly, given the importance of avoiding false positive results in a clinical setting (especially where such results inform cascade screening in families and subsequently treatment options of family members) and the preponderance of false positive associations in the literature as described above, the guidelines are stringent in nature and require a substantial level of evidence before classifying a variant as disease-causing. Specifically, and in contrast to variant classification methods described above for earlier research studies, the guidelines propose stricter population frequency thresholds than previously applied, give limited weight to computational prediction algorithms, define truncating variants as actionable only when affecting a gene where loss of function is a known mechanism for disease and stress the importance of family pedigree evidence like co-segregation of variants with disease and *de novo* inheritance. Finally, and importantly for this study and variant classification in HCM, the guidelines place most weight on variant-specific evidence, particularly for non-truncating variants. In practice, this means that segregation of the variant in the tested family (which is rarely available or pursued in standard clinical genetic testing), prior functional or genetic characterisation of the variant or the common recurrence of the variant in cases will be required to classify a variant as actionable.

3.1.4 Issues with current guidelines

3.1.4.1 Consistency between laboratories

One of the key goals of the ACMG guidelines is to improve the concordance amongst different clinical genetics laboratories in classifying the same variant. Initial reports have indicated that this has been at least partially successful, although the nature of the ACMG guidelines has meant that their consistent application between clinical laboratories remains an ongoing challenge. A pilot of these guidelines amongst 9 laboratories found that concordance increased from 34% to 71%, but only after consensus discussions and detailed review of how the criteria were applied⁹³. A recent study of 4 laboratories found initial concordance of

88.3%, with 87.2% of a subset of the discordant variants able to be resolved through data sharing and reassessment of how the criteria were applied¹⁴⁶. These studies have highlighted how the deliberately broad and vague nature of the ACMG guidelines have necessitated expert interpretation and application to specific disease areas, as well as consensus building between different clinical genetics groups and data sharing to facilitate equivalent interpretation for individual variants. Improved techniques to more efficiently apply these guidelines to specific diseases like HCM and produce more unbiased approaches to variant interpretation are clearly required.

3.1.4.2 Stringent approach underestimates pathogenicity and limits yield of genetic testing

As was shown in Chapter 2 (section 2.3.1.1 and Figure 2.6), current guidelines for variant classification in clinical laboratories are overly conservative and lead to the under-calling of pathogenic variants in HCM. This was demonstrated by the significant excess of VUS alone over the background rate of rare variation in ExAC in each of the 8 core sarcomeric HCM genes. As all rare truncating variants in *MYBPC3* are likely to be classified as at least likely pathogenic (with rule PVS1 activated), these excess VUS will mostly be non-truncating / missense variants that have not been previously demonstrated as disease-causing. Clearly, better data sharing between clinical laboratories or improved variant analysis techniques are required to distinguish between pathogenic and benign VUS and increase the yield of genetic testing in HCM. Recently published clinical outcome data from the SHaRe registry of HCM patients has provided additional evidence that many VUS in sarcomeric genes are likely to be causative. SHaRe patients with VUS had outcomes intermediate between genotype-positive and negative patients (the risk of developing the overall composite outcome by age 50 was 31.5% for sarcomeric-positive, 15.1% for sarcomeric negative and 26.2% for sarcomeric VUS), indicating a substantial proportion are likely to be pathogenic [publication in press].

3.1.4.3 Does genetic testing have an ethnic bias?

One of the consequences of the stringent approach encompassed by the ACMG guidelines is that it is very difficult for novel or relatively uncharacterised non-truncating variants to achieve a P/LP classification and therefore an actionable test result. In the absence of co-

segregation data, *de novo* inheritance or functional characterisation of a variant, variants would need to activate the following rules to become actionable - PM2 (rarity), PP3 (predicted to have a deleterious effect by computational algorithms), PM5 (affects an amino acid where a different change was known to be pathogenic) and one of the PP2 or PM1 rules on the relative frequencies of variants between cases and controls. As this is likely to apply to only a minority of novel non-truncating variants, the guidelines inevitably lead to an under-calling of such variants as demonstrated above.

As a result of this approach, genetic testing can become somewhat of a lottery for individual patients and families. If the causative mutation in the family is one that has been previously observed and characterised, e.g. it has been shown to segregate with disease (PP1) or is statistically enriched in cases over controls (PS4), then a positive test result will be obtained and the family can avail of cascade screening to identify members both at risk and free from risk of disease. However, if the causative variant is non-truncating and has not been previously characterised, the family is likely to obtain an uncertain result. Without an opportunity to genetically identify at risk family members, all phenotype-negative relatives of the proband are instead obliged to receive regular clinical screening.

This bias in the application of genetic testing does not just apply to individual patients and families but to different populations and ethnic groups. As much of the genetic research for cardiomyopathies (like many other conditions) has been carried out in the countries of Europe and North America with Caucasian-majority populations, it is conceivable that founder and recurrent variants in these populations will be more thoroughly evaluated than recurrent variants in other ethnic groups. Indeed, this is what has been observed in clinical genetic testing for HCM by LMM³³. 34.7% of white HCM patients tested by LMM received a positive result as compared to 24.2% of non-white patients, while 13.9% of white patients had a VUS compared to 20.6% of non-white patients (Fisher's exact test, $p < 1 \times 10^{-4}$) (Figure 3.3). While the lower proportion of pathogenic variants in non-whites may be partially explained by a different genetic architecture of disease in these populations, the higher proportion of VUS in these groups indicates that most of this effect is likely due to increasing numbers of uncharacterised variants in these populations.

These findings highlight the necessity for efforts to make genetic testing in diseases like HCM more unbiased and equitable, both for individuals and different population groups. More research into the genetic architecture of Mendelian diseases in non-Caucasians, and the characterisation of causative mutations in different population groups, is essential, but these are likely to be long term efforts. For now, novel techniques to improve variant interpretation and make it less dependent on the prior observation and characterisation of individual variants are required. Such analyses should ideally fit within the broad framework of the ACMG guidelines, but with adaptations and customisation to specific rules if required.

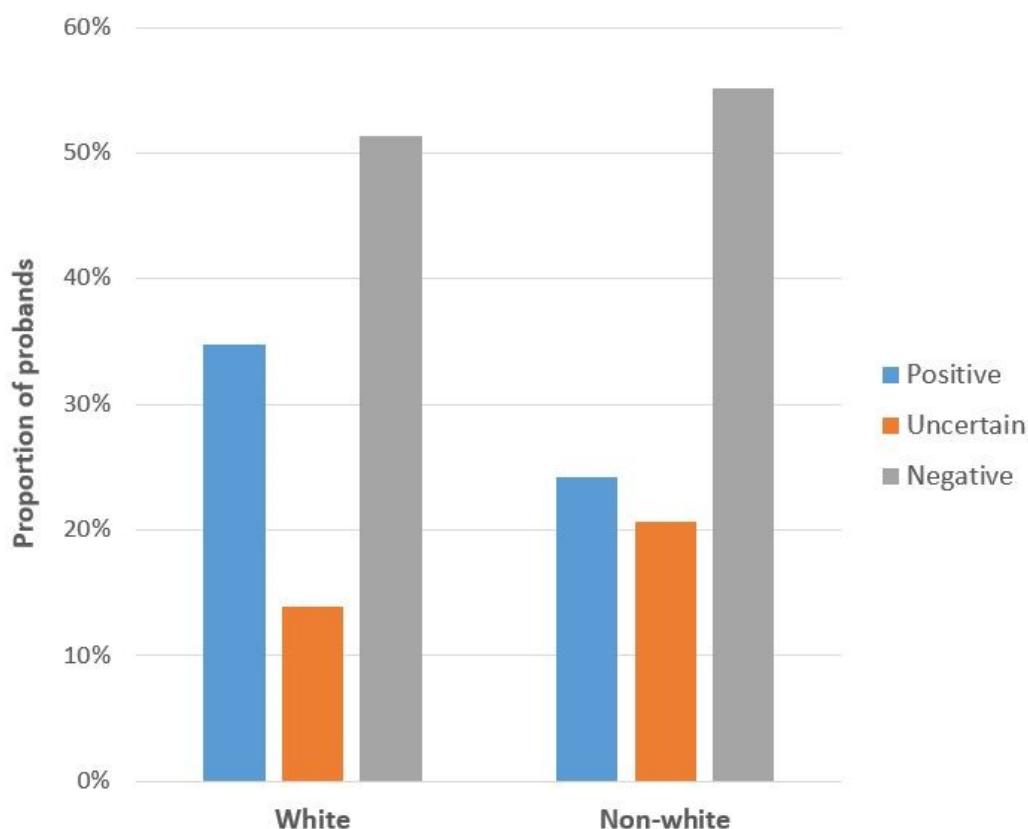


Figure 3.3: Proportion of HCM probands from LMM with positive (P/LP), uncertain (VUS) and negative (LB/B/no variant) genetic tests. Data from 1806 white patients and 558 non-white patients (212 Black or African-American, 184 Asian, 71 Hispanic, 15 American-Indian, 7 native Hawaiian and 69 mixed/other) are shown. Data for Ashkenazi Jewish and unspecified ethnicities not included³³.

3.1.5 Current and future challenges in variant interpretation

The interpretation of variants in HCM genes has undergone broadly two phases or approaches since linkage studies first implicated sarcomeric genes in the disease. Research studies that utilised lenient population frequency thresholds and other generic variant analysis techniques produced a rate of false positive findings that would be unacceptable in clinical genetic testing. The stringent criteria adopted by clinical genetics laboratories, now codified in the ACMG guidelines, sought to minimise false positive results but, as is now clear, this has come at the cost of test sensitivity. New approaches to variant interpretation in HCM that improve the yield of genetic testing in this disease, while maintaining rigorous standards in deeming a variant as actionable, are clearly required to maximise the potential of genetic testing in HCM. The availability of large genetic datasets from both population studies and disease cases now offers the opportunity to develop new disease and gene-specific methods to interpret variants and better differentiate classes of variants highly likely to be pathogenic from those for which strong, variant-specific evidence will be required to demonstrate pathogenicity.

3.2 Methods

3.2.1 A more refined population allele frequency threshold for HCM

In Chapter 2, I defined the methods for estimating the population allele frequency threshold used to define rare variants for assessing burden in putative cardiomyopathy genes, which utilised the disease prevalence and genetic heterogeneity of HCM. These methods have since been further developed and refined in a joint effort between our laboratory and the MacArthur group at the Broad Institute¹⁴⁷. The new method incorporates estimates of disease prevalence, genetic heterogeneity and penetrance, with the following formula used to calculate the maximum credible population allele frequency for dominant Mendelian diseases.

$$\text{Max population AF} = \frac{\text{disease prevalence} * \text{genetic heterogeneity}}{\text{penetrance}}$$

For HCM, disease prevalence was estimated at 1 in 500 people (or 1 in 1000 alleles or 0.001), genetic heterogeneity was estimated as 0.02 (as we have seen the most common variant in the HCM cohort, *MYBPC3*:p.R502W, was present in 1.7% of cases, with 95% CI for binomial distribution of 1.4%-2.0%), and penetrance was conservatively stated as 0.5 (the lowest of variants assessed for this study). The maximum credible population allele frequency for HCM was therefore calculated as 4×10^{-5} .

Instead of comparing this with the overall ExAC frequency as previously, this new approach defines the “filtering allele frequency” for each variant in ExAC, which accounts for sampling variance of ExAC counts (by calculating the CIs of Poisson distribution) and uses the highest frequency in any one of the major population groups in ExAC (non-Finnish European, African, East Asian, South Asian, Latino), on the assumption that a variant too common in any one ethnicity is too common to cause disease in any patient.

3.2.2 Identifying false positive disease associations in research literature

To assess the extent of putative false positive disease associations in the research literature, variants implicated in HCM in HGMD professional version 2016.3 were analysed with respect to their presence and population frequency in ExAC. For the 19 genes with some evidence for association with HCM (described in Chapter 2), the number of variants implicated in HCM that occur in ExAC above two frequency thresholds was calculated – 0.01% for overall ExAC frequency (as described in Chapter 2) and the frequency filter of 0.004%¹⁴⁷ (as defined above). The total number of individuals in ExAC with a variant implicated in HCM was calculated, as well as the number of individuals with variants below the threshold frequencies described above, i.e. variants that are plausibly pathogenic based on rarity in the population.

3.2.3 Defining Odds Ratios and Etiological Fractions

3.2.3.1 Odds Ratios (OR) and Etiological Fractions (EF)

The odds ratio (OR) is a measure of association between an exposure and an outcome and represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure¹⁴⁸. For its use in Mendelian genetics, exposure refers to the presence of a particular variant or class of variant while the outcome refers to the presence or absence of the disease in question. ORs can be calculated for individual variants or for classes of variants, e.g. all rare variants in a gene, by comparing the frequencies of carriers in case and control or population cohorts. The OR measures the enrichment of specified variants in cases and can be used to assess the likely pathogenicity or interpretability of the variant or variant class.

Attributable risk percent among exposed (ARP), provides an estimate of the proportion of the risk in an exposed population that can be attributed to the exposure which, for Mendelian disease, corresponds to the proportion of the risk of disease in variant carriers that can be attributable to the variant. ARP is defined as follows¹⁴⁹:

$$ARP(\%) = \frac{R_e - R_0}{R_e} \times 100\%$$

where

R_e = risk in the exposed (i.e. carriers of rare variant in gene of interest)

R_o = risk in the unexposed (i.e. individuals with no rare variant in gene of interest)

The equation can be conveniently rewritten as:

$$ARP(\%) = \frac{RR - 1}{RR} \times 100\%$$

where

RR = relative risk (ratio of risk among exposed to risk among unexposed)

For cross-sectional (i.e. case-control) data, ORs provide accurate estimates of the underlying relative risk, leading to:

$$ARP(\%) = \frac{OR - 1}{OR} \times 100\%$$

When expressed as a fraction, the ARP has been referred to as the etiological fraction (EF)¹⁵⁰. For the interpretation of variants in Mendelian diseases like HCM, the EF can describe the proportion of variant-carrying probands in which the variant is causal or the probability that an individual rare variant, found in a proband, is responsible for the disease.

3.2.3.2 Calculation of ORs and EFs for HCM genes

ORs are calculated by the following formula:

$$OR = (a/b)/(c/d)$$

where:

a = number of disease cases in the exposed group, i.e. with the variant of interest.

b = number of control or reference individuals in the exposed group.

c = number of disease cases in the non-exposed group, i.e. without the variant of interest.

d = number of control or reference individuals in the non-exposed group.

The 95% CIs for OR values was calculated as described by Altman (1991)¹⁵¹:

$$95\% \text{ CI} = \exp(\ln(OR) - 1.96 * SE\{\ln(OR)\}) \text{ to } \exp(\ln(OR) + 1.96 * SE\{\ln(OR)\})$$

where the standard error of the log OR was given by:

$$SE\{\ln(OR)\} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

An OR greater than 1 indicates an enrichment of the variant(s) in cases – this enrichment is deemed significant if the lower CI value is also greater than 1.

The EF was derived from the OR:

$$EF = (OR - 1)/OR$$

95% CIs for EF values were calculated as described by Hildebrandt et al.¹⁵²:

$$\left[\widehat{EF} - z_{1-\frac{\alpha}{2}} \sqrt{\widehat{VAR}(\widehat{\Phi})}, \min \left\{ \widehat{EF} + z_{1-\frac{\alpha}{2}} \sqrt{\widehat{VAR}(\widehat{\Phi})}, 1 \right\} \right]$$

where the asymptomatic variance estimator of $\widehat{\Phi}$ is given by:

$$\widehat{VAR}(\widehat{\Phi}) = \widehat{\Phi}^2 \left[\frac{1 - \widehat{\pi}_0}{N_0 \widehat{\pi}_0} + \frac{1 - \widehat{\pi}_1}{N_1 \widehat{\pi}_1} \right]$$

and:

$$\widehat{\pi}_1 = a/b + a \text{ (estimator for the probability of disease for an exposed person)}$$

$$\widehat{\pi}_0 = c/d + c \text{ (estimator for the probability of disease for an unexposed person)}$$

$$N_0 = c + d \text{ (total number of unexposed persons in the cohort)}$$

$$N_1 = a + b \text{ (total number of exposed persons in the cohort)}$$

$$\widehat{\Phi}^2 = \widehat{\pi}_0 / \widehat{\pi}_1$$

3.2.4 Refining EFs within genes

3.2.4.1 Using EFs as a means of quantifying the performance of variant classifiers

The EF is dependent on the relative frequencies of variants in cases and population controls. Given the relatively high frequency of non-truncating variants in the population for many genes, the overall gene EF will usually not be high enough to adequately distinguish between benign and pathogenic variation in a clinical genetic setting (the ACMG guidelines refer to a 90% likelihood of pathogenicity to be actionable). Therefore additional methods are required to discriminate between causative and background variants. This is illustrated by the hypothetical example in Figure 3.4 – there is a case frequency of 9.5% for all rare variants in this gene but as this includes a background level of 1.5% for rare benign variants (ExAC), the EF is 0.85 (1). A perfect discriminator of pathogenic and benign variants will identify the proportion of causative variants that is equal to the case excess (8.0%), yielding an EF of 1.0 (2), with the proportion of benign variants equal to the population reference frequency of ExAC (1.5%) and an EF of 0 (3).

In practice, while it is unlikely that full discrimination like this will be achieved, methods to differentiate between pathogenic and benign variants should aim to maximise the EF while capturing the maximum number of variants up to the case excess. This EF-based approach also allows us to evaluate such methods without requiring predefined lists of irrefutable pathogenic and benign variants (as used by standard assessments of specificity and sensitivity), which can be limited when performing analysis on specific genes.

3.2.4.2 Assessment of missense functional prediction algorithms for HCM genes

Functional prediction scores from the dbNSFP database¹⁵³ (version 3.2) were downloaded for all missense variants in the 13 HCM genes for which missense variants have been shown to be potentially causative from Chapter 2 (*MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, *ACTC1*, *CSRP3*, *FHL1*, *FLNC*, *PLN*, *TNNC1*). Eight scores that provide binary predictions, i.e. damaging vs benign/neutral were assessed – fathmm-MKL coding, FATHMM, LRT, Mutation assessor, MutationTaster, Polyphen2-HDIV, PROVEAN and SIFT, as well as the CADD algorithm (damaging variants were defined with a CADD phred score ≥ 15). A consensus prediction

between the 9 scores was defined as being damaging if greater than 50% of the scores predicted a damaging effect. Additionally two consensus algorithms, MetaLR and MetaSVM, were also evaluated¹⁵⁴.

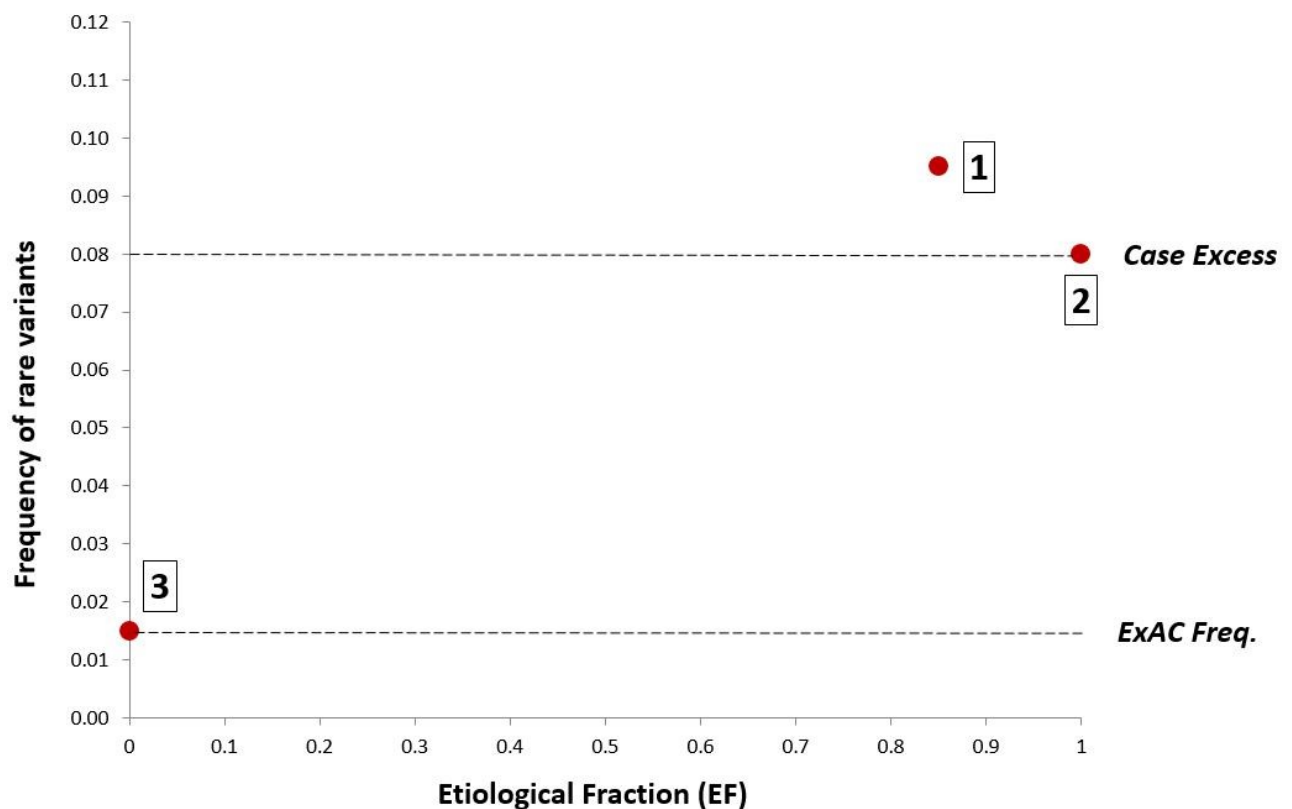


Figure 3.4: Illustration of how EFs can be used to evaluate methods for distinguishing pathogenic from benign variants. For the hypothetical gene shown here, the overall EF of 0.85 [1] is based on a case frequency of 9.5% and a reference frequency of 1.5%. The aim of variant classification methods is to fully distinguish between pathogenic variants (producing an EF of 1.0 with frequency equal to case excess [2]) and benign variants (producing an EF of 0 with frequency equal to ExAC reference [3]).

The functional prediction algorithms were assessed in three ways. Firstly, the proportion of available predictions for each score for all potential missense variants in each gene was calculated to identify algorithms that do not provide sufficient predictions for specific genes. The gene : predictor combinations for which less than 20% of predictions are available in dbNSFP were - LRT (*MYH7* and *MYBPC3*), Mutation assessor (*FHL1*, *PLN*, *TPM1* and *TNNT2*) and SIFT (*ACTC1*). Secondly, the effectiveness of these prediction scores for individual HCM genes was evaluated with lists of known pathogenic and benign missense variants. Pathogenic variants were defined by rarity in ExAC as described above and:

- 1) classified as pathogenic or likely pathogenic in HCM patients by one or more clinical laboratories with no conflicting classifications (VUS or benign) by other laboratories. Clinical laboratories refer to OMGL, LMM and other ClinVar submitters excluding OMIM.
- 2) significantly enriched in the OMGL/LMM cohorts compared to ExAC (Fisher's exact test).

Benign variants were defined as:

- 1) presence in more than one individual in ExAC and not associated with any disease in ClinVar (P/LP/VUS) or HGMD.
- 2) associated with disease in ClinVar (though not classified as pathogenic or likely pathogenic) or HGMD but at a frequency >0.001 in ExAC.

The sensitivity (true positive rate) and specificity (true negative rate) was calculated for the 8 functional prediction scores and 3 consensus scores for each of the 8 core sarcomeric genes (there were insufficient known pathogenic variants for the minor genes).

Finally, the ability of the three consensus scores to differentiate pathogenic from benign variation was assessed by calculating EFs for the subset of damaging variants in each HCM gene, using the OMGL/LMM clinical cohorts and ExAC. The results were then compared to EFs calculated by variant clustering (below).

3.2.4.3 Identification of variant clustering in HCM genes

To detect if rare variants in cases are enriched in particular regions or domains of sarcomeric genes, a bespoke clustering algorithm was developed in collaboration with Dr Francesco Mazzarotto of the University of Florence. In our initial publication on cardiomyopathy genetics⁸⁸, the non-random mutation cluster algorithm (implemented in the iPAC Bioconductor R package)¹⁵⁵ was used to identify clusters of variants in *MYH7*. However, this method struggled to detect concise clusters in other genes. Therefore, a new regional enrichment detection algorithm was developed.

The algorithm compares the total count of observed variants in a randomly chosen window of 5% of the protein size to the expected count obtained from an equally-sized virtual cohort, in which the same number of variants are randomly distributed. Given that random assignment of variant locations could result in stochastic clustering, we initialized 100 virtual cohorts per gene, and performed 1000 random window tests against each, for a total of 100,000 tests per gene. Clusters were defined as the union of windows in which the enrichment was significant (after multiple testing correction). Consequently, boundaries were adjusted to prevent clusters from excluding neighbouring variant residues or from including variant-free flanking regions.

3.2.5 Adapting ACMG guidelines

3.2.5.1 Current rules focusing on relative variant frequencies between cases and controls

There are two rules in the ACMG guidelines that relate to the relative frequencies of classes of variants between cases and controls (although they are classified as “functional data” rules) – PP2 (*missense in gene with a low rate of benign missense variants and pathogenic missense variants are common*) and PM1 (*mutational hot spot or well-studied functional domain without benign variation*) (Figure 3.1). These are specifically designed to be applied to non-truncating or missense variants and, unlike many of the ACMG rules, can be activated by novel or relatively uncharacterised variants. As with other ACMG criteria, these rules are intentionally vaguely defined (referring to “low rate” or “common” variation) and non-quantitative, with PM1 also requiring a complete absence of benign variants from key

domains (this condition will likely be too stringent for most genes in the age of ExAC and gnomAD). It is expected that only one of these two rules would be activated by any variant and, as such (and as described in the introduction), adopting either of these rules would not lift any novel variant beyond VUS, regardless of how enriched such variant classes are in cases.

3.2.5.2 Proposed changes to these rules

The proposed changes to the ACMG guidelines for this evidence class are shown in Figure 3.5. For conditions such as HCM where we have large case datasets, the current PP2 and PM1 rules would be superseded – although they refer to genes and domains respectively, both rules essentially concern the enrichment of a particular class of variants in cases, without adequately defining the level of enrichment necessary to achieve either supporting or moderate evidence. In their place, the PM1 rule is defined with three levels of evidence, denoted as PM1_supporting, PM1_moderate and PM1_strong. The specific rule and level of evidence activated depends on the pre-calculated EF value for the class of the variant in question – this may refer to rare non-truncating variants across an entire gene or within a specified region shown to have enrichment or depletion of variants in case or control cohorts. An EF between 0.8 and 0.9 would constitute supporting evidence (PM1_supporting), between 0.9 and 0.95 would be moderate (PM1_moderate) and greater than 0.95 would provide strong evidence (PM1_strong).

The new rules offer the following advantages:

- Although they introduce a semi-quantitative approach to variant assessment, the rules fit within the existing ACMG framework. They are structured similarly to the segregation rule (PP1) which, although defined as supporting evidence, can be progressed to moderate and strong with increasing numbers of segregations or LOD scores (Figure 3.1).
- The EF thresholds applied correspond to a minimum OR of 5 for supporting, 10 for moderate and 20 for strong evidence, equivalent to a two-fold increase in odds between evidence categories.

Current ACMG guidelines:

Pathogenic			
Supporting	Moderate	Strong	Very strong
Missense in gene with low rate of benign, missense variants and path. missenses common PP2	Mutational hot spot or well-defined functional domain without benign variation PM1		

Proposed adapted ACMG guidelines:

Pathogenic			
Supporting	Moderate	Strong	Very strong
Non-truncating variant in gene or protein region with $0.8 \leq EF < 0.9$ PM1_supporting	Non-truncating variant in gene or protein region with $0.9 \leq EF < 0.95$ PM1_moderate	Non-truncating variant in gene or protein region with $EF \geq 0.95$ PM1_strong	

Figure 3.5: Proposed changes to the ACMG rules relating to the relative frequencies of rare variants between cases and controls.

- These changes address the discrepancy between the current rules for truncating and non-truncating variants. All truncating variants in genes where loss of function is a known mechanism for the disease in question will achieve a classification of at least likely pathogenic, courtesy of the (very strong) PVS1 rule (assuming sufficient rarity of the detected variant). While the weight of this rule partly derives from the fact that a non-functional protein is likely to be produced by the truncating variant (albeit with the caveats described by Richards et al.⁷¹), it also reflects the rarity of such variants in the population and consequently the high odds of a variant detected in a patient being pathogenic, e.g. *MYBPC3* truncating variants in HCM, with an $EF > 0.99$ and an OR of 115 (see Results below). Non-truncating variant classes that are similarly highly enriched in case cohorts should also have this evidence reflected when evaluating variants.

- Critically, these modified rules enable novel or limited evidence missense variants with an EF ≥ 0.95 to be classified as likely pathogenic, assuming they are of sufficient rarity (PM2), with one strong and one moderate rule activated (Figure 3.2). For rare and novel inframe insertions and deletions, an EF ≥ 0.9 would be sufficient for a likely pathogenic classification, yielding 3 lines of moderate evidence – PM2, PM4 (*protein length changing variant*) and PM1_moderate.
- Importantly, even though these changes aim to classify certain non-truncating variants as actionable in the absence of segregation or functional evidence, and even if the variant is completely novel, they maintain the rigorous ACMG approach to clinical variant assessment. The ACMG guidelines define likely pathogenic as a “*greater than 90% certainty of a variant being disease-causing*”. The approach described here actually takes a more conservative stance than these recommendations in requiring 95% likelihood of pathogenicity for such variants to activate the strong rule and yield a classification of likely pathogenic for a novel variant.

3.2.6 Variant classification in clinical and prospective HCM cohorts

3.2.6.1 Reclassifying VUS from clinical cohorts based on adapted ACMG guidelines

To assess the effect of the adapted ACMG guidelines on the yield of pathogenic variants in HCM, rare VUS in sarcomeric genes in the OMGL/LMM clinical cohort were re-evaluated. The proportion of patients with variants that would be upgraded from VUS to Likely Pathogenic with the additional activation of the new rules was calculated. These variants correspond to those previously classified as VUS and either in a variant class with an EF ≥ 0.95 and therefore activating, as a minimum, the PM1_strong and PM2 (rarity in the population) rules or an inframe indel with an EF ≥ 0.90 and therefore activating, as a minimum, the PM1_moderate, PM2 and PM4 (protein length changing variant) rules.

3.2.6.2 Classification of sarcomeric variants in prospective HCM cohort

To further assess the impact of these adapted ACMG guidelines on variant classification in HCM genes, and to define a high confidence set of samples with pathogenic variants for

subsequent genotype-phenotype analysis (Chapter 4), data from the prospective HCM cohort described in Chapter 2 was analysed using a subset of key rules from the ACMG guidelines. Additionally, we sought to assess how the new EF-based rules (that can be automatically applied) compare with manual curation of individual variant evidence (such as segregation data), both in terms of increasing the yield of actionable variants and reducing the time required for variant analysis.

The following rules were applied for the initial automated analysis (without applying the EF rules or manual curation):

- PM2 – filtering allele frequency in ExAC $< 4 \times 10^{-5}$. This rule must be activated to denote a causative variant¹⁴⁷.
- PVS1 – truncating variants in *MYBPC3*, *TNNT2*, *TNNI3*, *CSRP3*, *FHL1*, *PLN* (genes with truncating variants statistically enriched in HCM cohorts versus ExAC) and *CRYAB*.
- PS4 – individual variant statistically enriched in cases over controls, based on LMM/OMGL cohort versus ExAC with the rule activated if the case count was >2 and the Fisher's exact test p-value $< 1.79 \times 10^{-6}$ (Bonferroni correction), detailed in our paper describing the CardioClassifier resource¹⁵⁶.
- PM4 – protein length changing variant, i.e. an inframe indel or stop lost variant.
- PP3 – missense variant with multiple lines of computational evidence suggesting a deleterious effect. The following eight algorithms were used - SIFT, PolyPhen2 Var, LRT, Mutation Taster, Mutation Assessor, FATHMM, CADD and Grantham scores, with the rule activated if only one tool predicts benign and >3 have unknown classifications or if ≥ 3 tools have unknown outcomes and all other tools predict deleterious (as applied by CardioClassifier¹⁵⁶).
- PM5/PS1 – novel missense change at an amino acid where a different missense variant is pathogenic (PM5) or novel missense variant with same amino acid change as an established pathogenic variant (PS1). Pathogenicity here is defined as a pathogenic classification in ClinVar by multiple submitters with no conflicting evidence (as applied by CardioClassifier¹⁵⁶).

For rare variants (activating the PM2 rule), the following rules assessing published variant-specific evidence were applied by curation of ClinVar entries and published reports described in HGMD.

- PP1 – co-segregation with disease. This rule was defined as supporting for ≥ 3 observed meioses, moderate for ≥ 5 meioses and strong for ≥ 7 meioses. These criteria were those used for both CardioClassifier¹⁵⁶ and recommendations for *MYH7* variant classification developed by the ClinGen Inherited Cardiomyopathy Expert Panel (in press). Where multiple pedigrees have been described, these were used to increase the evidence strength, e.g. two “supporting” pedigrees (3-4 meioses) combined to produce moderate evidence.
- PS2/PM6 – *de novo* inheritance (with / without confirmed paternity and maternity).
- Rules relating to effects in functional studies (PS3/BS3) were not applied due to the lack of standardisation and validation in functional assays for HCM variants.

The new EF-based rules for non-truncating variants with EF ≥ 0.8 were then applied as described above to assess the additional yield provided by this approach.

3.2.6.3 Detection and characterisation of non-rare variants enriched in cases

To detect non-rare variants that may contribute to disease with reduced penetrance, all variants with a population frequency above the threshold applied (i.e. not activating PM2) that occurred in more than one case in the prospective HCM cohort were assessed. Enrichment of these variants was tested with Fisher’s exact tests against ExAC and a healthy volunteer (HVOL) cohort of 912 unrelated individuals sequenced on the TruSight panel. As more common variants can have varying frequencies in different population groups, this analysis was repeated for Caucasian/European samples, i.e. self-declared Caucasian HCM and HVOL samples and the non-Finnish European subset of ExAC.

For non-rare variants that were enriched in HCM cases, their estimated penetrance was calculated as follows:

$$\text{penetrance} = \text{disease prevalence} * \frac{\text{allele frequency in cases}}{\text{allele frequency in population}}$$

as described by Minikel et al.¹⁵⁷, with 95% CIs calculated by the method of Kirov et al.¹⁵⁸, i.e. the binomial distribution for case and population allele frequencies were calculated, with the upper bound of 95% CI calculated with the upper bound for case allele frequency and the lower bound for population allele frequency (and vice versa for the lower bound of 95% CI).

3.3 Results

3.3.1 Variant analysis in HCM confounded by both false positive and false negative associations

3.3.1.1 Extent of false positive variant : disease associations in the research literature

The number of variants implicated in HCM (as identified by HGMD, version 2016.3) above defined frequency thresholds (overall ExAC frequency of 0.01% and filtering allele frequency of 0.004%) is shown in Table 3.1, for 19 genes with evidence of association to HCM (as defined in Chapter 2). Of the 1,408 variants linked to HCM, 79 (5.6%) have an ExAC frequency > 0.01% and 119 (8.5%) have a filtering allele frequency > 0.004%. In total, variants implicated in HCM are observed 23,610 times in ExAC, corresponding to a maximum of 38.9% of individuals in ExAC (assuming no individual had more than one variant). Although ExAC is not a control database of healthy individuals and is likely to include HCM patients, it is not expected to be enriched beyond the normal disease prevalence of approximately 1 in 500 or 0.2%. Therefore this analysis indicates there are significant numbers of false positive reports for HCM in the literature.

Focusing solely on rare variants, i.e. those below the frequency thresholds described, there are 1,299 (2.14%) ExAC individuals (ExAC MAF < 0.01%) or 626 (1.03%) ExAC individuals (filtering allele frequency < 0.004%) with HCM implicated variants. With an expected disease prevalence of HCM in ExAC of 0.2% (121 individuals), and approximately half of these unlikely to have a causative variant in these genes, this indicates that a substantial proportion of implicated variants that are rare enough to be plausibly pathogenic will still be false positive associations. It should also be noted that this assessment actually understates the number of false positive associations for HCM as it does not include variants in genes that were deemed to have insufficient evidence to be putative HCM genes, as described in Chapter 2.

Gene		HGMD variants	ExAC MAF 0.01%		ExAC FAF 0.004%		Total ExAC individuals with HCM variant
			# variants >0.01%	Individuals with variant <0.01%	# variants >0.004%	Individuals with variant < 0.004%	
Validated genes	MYH7	520	8	468	20	309	5,680
	MYBPC3	578	46	498	67	215	13,151
	TNNT2	70	5	45	6	8	771
	TNNI3	62	2	71	2	44	1,746
	TPM1	26	0	7	0	3	7
	MYL2	32	3	27	3	3	532
	MYL3	23	1	108	2	26	122
	ACTC1	19	0	10	0	3	10
Strong	CSRP3	11	1	9	1	7	183
	FHL1	12	0	0	0	0	0
	PLN	1	0	2	0	0	2
Moderate	ACTN2	6	1	5	1	4	57
	CRYAB	2	1	0	1	0	45
	FLNC	8	0	11	1	0	11
	MYOZ2	2	1	0	1	0	448
Weak	MYH6	10	6	14	9	0	365
	TNNC1	7	1	0	1	0	112
	TRIM55	8	1	5	0	0	21
	TRIM63	11	2	19	4	4	347
Total		1,408	79	1,299	119	626	23,610

Table 3.1: Variants implicated in HCM as described by HGMD (version 2016.3). Displayed are the number of variants above two defined thresholds – an ExAC MAF of 0.01% and ExAC filtering allele frequency of 0.004%, as well as the number of individuals in ExAC with implicated variants below these thresholds. Also shown are the total number of individuals in ExAC with implicated HCM variants regardless of ExAC frequency but excluding three MYBPC3 polymorphisms, c.706G>A (ExAC MAF of 0.108), c.2686G> (0.013) and c.2737+12C>T (0.101).

3.3.1.2 Proportion of VUS in clinical HCM cohorts likely to be false negatives

In contrast to many research studies, clinical laboratories adopt stringent approaches to variant classification to minimise false positive findings. As demonstrated in Chapter 2 (section 2.3.1.1 and Figure 2.6) and discussed in the Introduction, these guidelines are likely to be overly conservative and lead to false negative findings. Focusing solely on VUSs in the HCM clinical cohorts, a significant excess over the background rate of rare variation in ExAC

was observed in each of the 8 core sarcomeric HCM genes (Figure 3.6). Rare VUS in these genes occurred in 12.36% of HCM patients in this cohort, compared to 4.35% of individuals in ExAC (the expected background rate of rare benign variation). Therefore, although there is not sufficient evidence to classify these variants individually as causative based on current guidelines, we can deduce that almost two-thirds of these variants are likely to be disease-causing, corresponding to 8.0% of all patients referred for genetic testing.

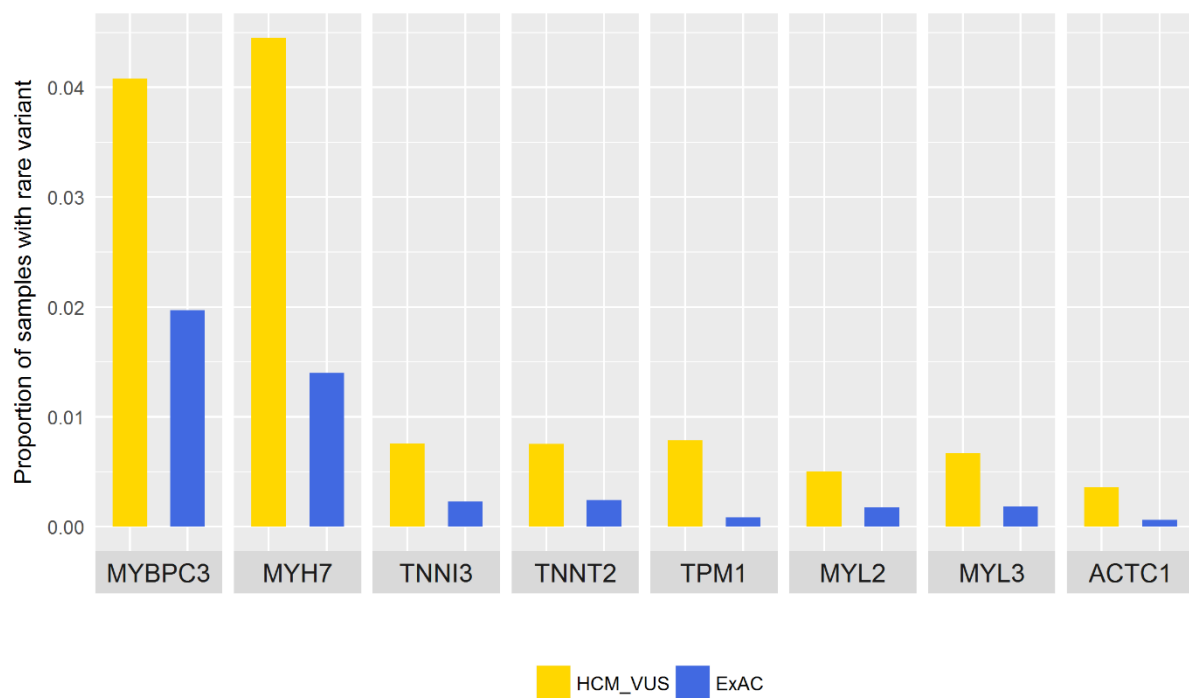


Figure 3.6: Proportion of cases in joint LMM/OMGL clinical cohort with rare variants classified as VUS by the clinical laboratories compared to the proportion of individuals in ExAC with rare variants (corresponding to the background rate of rare benign variants in these genes).

3.3.2 Odds Ratios and Etiological Fractions for HCM genes

The etiological fraction (EF) for a given gene or variant class describes the proportion of cases with such variants where that variant is causative, which corresponds to the prior probability that a rare variant, without considering additional evidence, is pathogenic. It can be used to

identify variant classes highly enriched in cases that, by incorporating the new ACMG rules described in Methods, could improve the yield of pathogenic variants.

Odds ratios (OR) and EFs were calculated separately for truncating and non-truncating variants for each HCM gene. Only gene variant classes with a significant excess of rare variants in case cohorts over ExAC were assessed. A more stringent definition of rare variation (filtering allele frequency in ExAC $< 4 \times 10^{-5}$) than the analysis in Chapter 2 was used here, as described in Methods (3.2.1)¹⁴⁷. The LMM/OMGL clinical cohort was used for the 8 core sarcomeric genes and the combined cohorts described in Chapter 2 were used for the additional HCM genes (*CSRP3*, *FHL1*, *FLNC*, *PLN*, and *TNNC1*). Although non-truncating variants in *FLNC* and *TNNC1* were not significantly enriched after Bonferroni correction in the analysis in Chapter 2, they were included in this analysis given their likely valid role in HCM.

The data from this analysis is shown in Table 3.2. As expected, high OR (115.3) and EF (0.99) values were observed for truncating variants in *MYBPC3*, given the prominent role of this variant class in HCM (occurring in over 9% of cases) and the low rate in the general population (0.09%). Values for truncating variants in other genes are more modest due to their low case frequency (occurring in $< 0.2\%$ of cases in each gene). It should be noted however that all rare truncating variants in these genes will achieve at least a Likely Pathogenic classification by ACMG guidelines (courtesy of the PVS1 and PM2 rules, Figure 3.2), assuming that a loss of function mechanism for these genes in HCM is acknowledged.

For non-truncating variants, the EFs are generally more modest due to the higher frequency of such variants in the population and vary widely according to gene. Only non-truncating variants in *TPM1* have an EF ≥ 0.95 , yielding strong evidence (PM1_strong) with the adapted guidelines described above. For the other sarcomeric genes, these gene-level scores would only produce supporting or moderate evidence from the adapted rules, such that variants in these genes would continue to be VUS in the absence of additional variant-specific evidence. For this reason, further methods to differentiate pathogenic from benign variant classes are required to improve EF values.

Gene	Frequency Cases (variants/total)	Frequency ExAC (variants/total)	Fisher's exact p-value	Odds Ratio (OR)	Etiological Fraction (EF)
<i>Non-truncating variants</i>					
MYH7	13.89% (849/6112)	1.11% (672/60469)	<0.0001	14.4 (12.9-15.9)	0.930 (0.923-0.938)
MYBPC3	9.35% (578/6179)	1.21% (555/45794)	<0.0001	8.4 (7.5-9.5)	0.881 (0.868-0.895)
TNNT2	1.69% (103/6103)	0.15% (86/57018)	<0.0001	11.4 (8.5-15.2)	0.912 (0.889-0.935)
TNNI3	2.10% (127/6047)	0.15% (79/52607)	<0.0001	14.3 (10.8-18.9)	0.930 (0.912-0.948)
TPM1	1.44% (64/4447)	0.07% (42/58642)	<0.0001	20.4 (13.8-30.1)	0.951 (0.933-0.969)
MYL2	1.03% (43/4185)	0.11% (69/60521)	<0.0001	9.1 (6.2-13.3)	0.890 (0.851-0.930)
MYL3	0.84% (35/4185)	0.14% (85/60605)	<0.0001	6.0 (4.0-8.9)	0.833 (0.772-0.895)
ACTC1	0.53% (22/4185)	0.06% (37/60198)	<0.0001	8.6 (5.1-14.6)	0.884 (0.826-0.941)
PLN	0.17% (9/5440)	0.02% (15/60475)	<0.0001	6.7 (2.9-15.3)	0.850 (0.737-0.964)
CSRP3	0.62% (30/4866)	0.19% (115/60647)	<0.0001	3.3 (2.2-4.9)	0.694 (0.579-0.808)
FHL1	0.78% (16/2061)	0.09% (53/60278)	<0.0001	8.9 (5.1-15.6)	0.888 (0.826-0.949)
TNNC1	0.24% (8/3335)	0.06% (33/59192)	0.0013	4.3 (2.0-9.3)	0.768 (0.598-0.938)
FLNC	3.79% (17/448)	2.15% (1225/56897)	0.0314	1.8 (1.1-2.9)	0.442 (0.172-0.712)
<i>Truncating variants</i>					
MYBPC3	9.16% (566/6179)	0.09% (40/45794)	<0.0001	115.3 (83.6-159.1)	0.991 (0.988-0.995)
TNNT2	0.18% (11/6103)	0.03% (17/57018)	<0.0001	6.1 (2.8-12.9)	0.835 (0.722-0.948)
TNNI3	0.08% (5/6047)	0.01% (5/52607)	0.0019	8.7 (2.5-30.1)	0.885 (0.757-1.013)
PLN	0.17% (9/5440)	0.01% (4/60475)	<0.0001	25.1 (7.7-81.4)	0.960 (0.917-1.003)
CSRP3	0.14% (7/4866)	0.02% (14/60647)	0.0006	6.2 (2.5-15.5)	0.840 (0.705-0.974)
FHL1	0.15% (3/2061)	0.00% (0/60278)	<0.0001	205.0 (10.6-3969.8)	0.995 (0.981-1.009)

Table 3.2: Odds ratios and etiological fractions for non-truncating and truncating variants (with 95% CIs) in HCM genes with a significant excess of rare variation over ExAC. Variant counts and overall frequencies in case cohorts and ExAC are also displayed, along with p-values from Fisher's exact tests.

3.3.3 Refining EFs within genes

3.3.3.1 Evaluation of missense functional prediction algorithms for HCM genes

In order to evaluate the performance of computational functional prediction algorithms on differentiating EFs, consensus functional prediction scores taken from the dbNSFP database were initially assessed for their gene-specific performance using known pathogenic and benign variants (Table 3.3). As well as observing the expected and well recognised levels of high sensitivity but limited specificity that characterise these algorithms, it was notable that the FATHMM predictor (and the MetaLR and MetaSVM consensus scores that incorporate FATHMM) substantially under-calls pathogenic variants for *MYBPC3*, *MYL2* and potentially *MYL3* (though data is limited for this gene). This is despite the fact that FATHMM has generally been shown to be one of the better performing predictors¹⁵⁴. This data highlights how such generic techniques may be unsuitable for interpreting variants in specific genes and the importance of developing disease and gene-specific approaches for variant classification.

3.3.3.2 Identification of clusters of case variants in HCM genes

Clusters of case variants were detected in 5 HCM genes – *MYH7* (residues 167 - 969), *MYBPC3* (residues 178 - 311, 432 - 597, 750 - 858, 1194 - 1266), *TNNI3* (residues 136 - 171, 186 - 209), *TNNT2* (residues 79 - 104, 151 - 179, 278 - 286) and *MYL3* (residues 149 - 156) (Figure 3.7). The EFs for rare truncating variants in these clusters are shown in Table 3.4, with an EF ≥ 0.95 identified for the clusters in *MYH7*, *TNNI3*, *TNNT2* and *MYL3*. Rare variants detected in these regions would activate the PM1_strong rule (in addition to the PM2 rule on variant rarity) and thus have at least a likely pathogenic classification, regardless of any other variant-specific evidence. The EF for the *MYBPC3* clusters is 0.937, equivalent to the PM1_moderate rule. Novel inframe indel variants in *MYBPC3* would therefore be classified as likely pathogenic (PM2, PM1_moderate and PM4 rules activated) while missense variants would still be VUS and require further, variant-specific evidence to become actionable.

Predictor	MYBPC3	MYH7	TNNT2	TNNI3	TPM1	MYL2	MYL3	ACTC1
No. of Variants (Path / Benign)	42 / 119	173 / 83	17 / 15	25 / 12	15 / 6	11 / 15	5 / 13	3 / 6
CADD	0.93 / 0.22	1.00 / 0.17	0.94 / 0.27	0.96 / 0.17	1.00 / 0.17	1.00 / 0.20	1.00 / 0.23	1.00 / 0.17
FATHMM	0.19 / 0.97	0.92 / 0.25	0.82 / 0.00	1.00 / 0.08	0.93 / 0.33	0.18 / 0.93	0.40 / 0.08	1.00 / 0.00
fMKL	0.93 / 0.26	0.94 / 0.19	0.94 / 0.47	0.88 / 0.33	1.00 / 0.00	1.00 / 0.07	1.00 / 0.23	1.00 / 0.00
LRT	- / -	- / -	1.00 / 0.73	0.72 / 0.58	0.93 / 0.00	0.91 / 0.13	0.80 / 0.69	0.00 / 0.00
MutationAssessor	0.74 / 0.54	0.79 / 0.29	- / -	0.72 / 0.67	- / -	0.64 / 0.53	0.80 / 0.38	1.00 / 0.67
MutationTaster	0.90 / 0.29	0.97 / 0.13	0.94 / 0.40	0.92 / 0.50	1.00 / 0.00	1.00 / 0.07	1.00 / 0.23	1.00 / 0.00
Polyphen2-DIV	0.83 / 0.43	0.83 / 0.39	0.94 / 0.47	0.96 / 0.42	0.47 / 0.50	1.00 / 0.33	0.60 / 0.38	0.33 / 0.83
PROVEAN	0.71 / 0.60	0.88 / 0.42	0.82 / 0.67	0.84 / 0.67	0.80 / 0.67	0.91 / 0.53	0.80 / 0.62	0.33 / 0.67
SIFT	0.79 / 0.39	0.96 / 0.22	0.82 / 0.73	0.88 / 0.75	0.87 / 0.33	0.73 / 0.60	0.80 / 0.31	- / -
Consensus	0.88 / 0.38	0.98 / 0.24	0.94 / 0.53	0.96 / 0.42	0.93 / 0.17	1.00 / 0.20	1.00 / 0.46	1.00 / 0.17
MetaSVM	0.43 / 0.90	0.99 / 0.30	0.88 / 0.27	0.92 / 0.25	1.00 / 0.17	0.64 / 0.53	0.40 / 0.54	1.00 / 0.17
MetaLR	0.33 / 0.91	1.00 / 0.28	0.88 / 0.20	0.96 / 0.17	1.00 / 0.17	0.73 / 0.60	0.60 / 0.38	1.00 / 0.00

Table 3.3: True positive rate (TPR) and true negative rate (TNR) values for missense functional prediction scores for eight core sarcomeric HCM genes, based on defined pathogenic and benign variants as described in Methods. Consensus indicates >50% of individual predictors with a deleterious score. Notably for MYBPC3, MYL2 and MYL3, the TPR of the FATHMM predictor (and therefore the MetaLR and MetaSVM consensus scores that incorporate FATHMM, all three of which display high correlation - Dong et al, 2015¹⁵⁴) is substantially lower compared to other predictors, despite the fact that FATHMM is a generally high performing predictor (Dong et al, 2015¹⁵⁴).

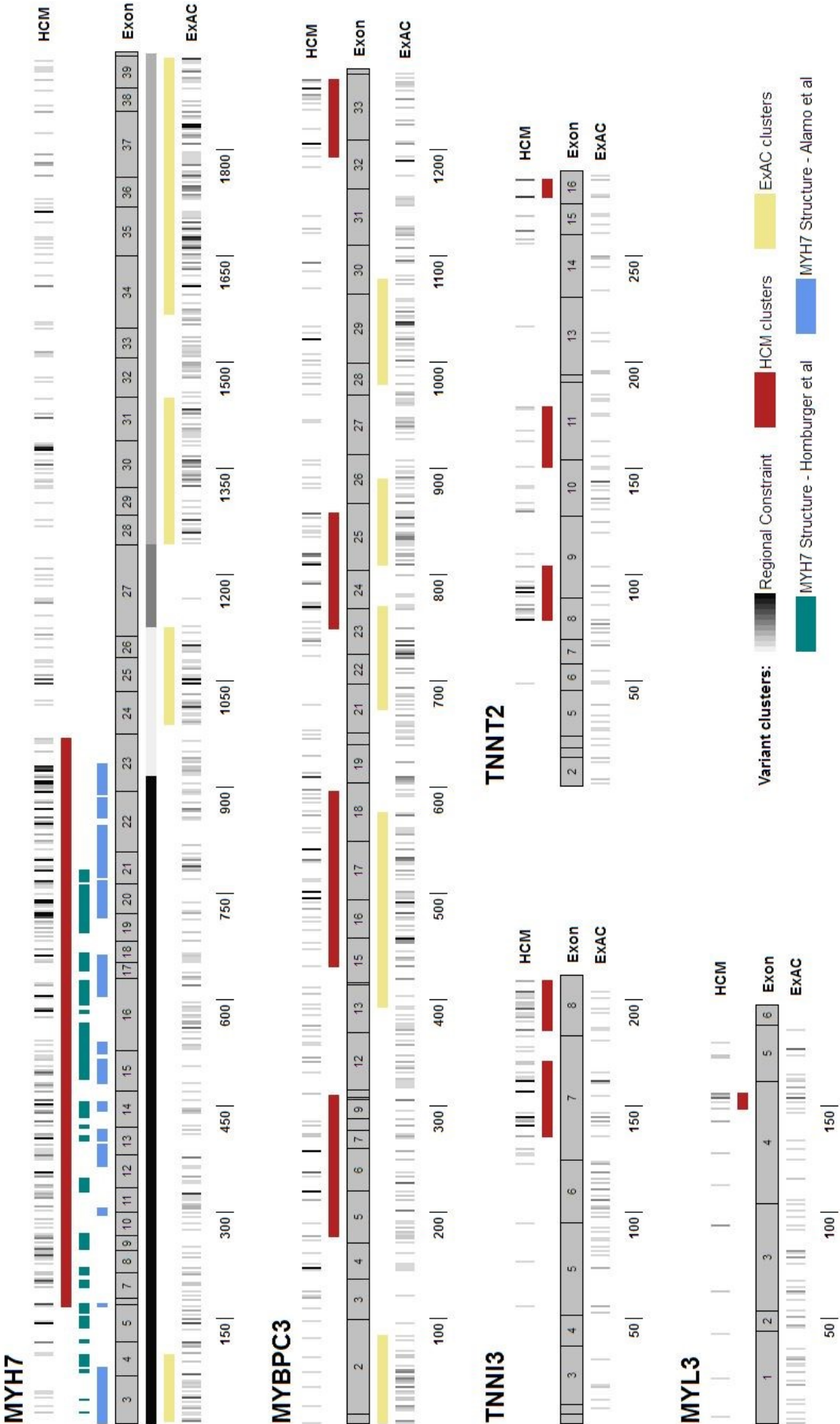


Figure 3.7 (previous page): Distribution of rare, non-truncating variants in HCM cohorts and ExAC for the five HCM genes with detected clustering of variants (variant density increases with darker shades of grey). The HCM clusters detected were: MYH7 (residues 167-969), MYBPC3 (178-311, 432-597, 750-858, 1194-1266), TNNI3 (136-171, 186-209), TNNT2 (79-104, 151-179, 278-286) and MYL3 (149-156). For MYH7, key structure-function residues (Alamo et al.¹⁵⁹ and Homburger et al.¹⁶⁰) and constraint regions (Samocha et al.¹⁶¹) are also displayed.

3.3.3.3 Overlap between case clusters and regions from other studies

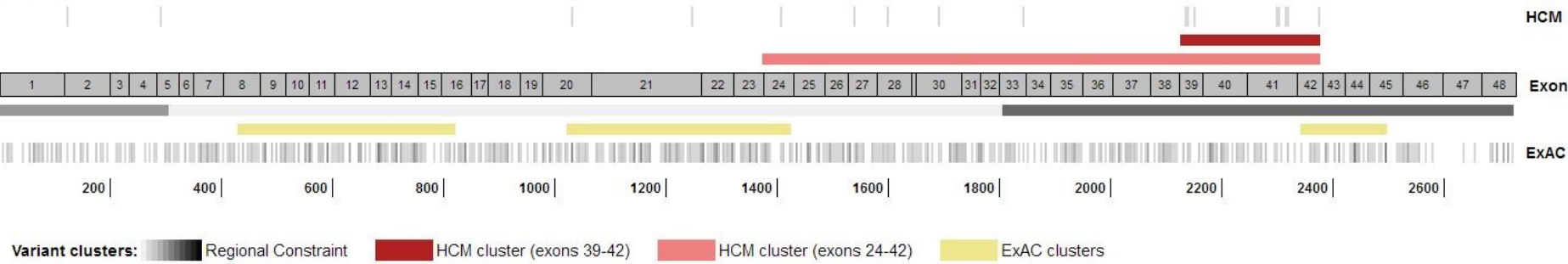
A different clustering approach, non-random mutation cluster (NMC)¹⁵⁵ implemented in the iPAC Bioconductor R package, was used to identify enriched regions in *MYH7* in our published study from the analysis presented in Chapter 2 (Walsh and Thomson et al., 2017⁸⁸). The HCM case cluster identified by NMC (using the same dataset but including variants with an ExAC frequency < 0.01% rather than a filtering allele frequency < 0.004% in this study) was similar, though slightly smaller than the cluster identified here – residues 181-837 compared to residues 167-969. An ExAC cluster in the C-terminal myosin rod region identified by NMC (residues 1271-1903) broadly overlapped with two clusters detected here (residues 1244-1450 and 1569-1929) though two other ExAC clusters were also identified here (residues 4-100 and 990-1126) (Figure 3.7).

The differential pathogenicity of regions within genes has also been observed in a recent study assessing missense variant constraint in the ExAC database¹⁶¹. For the HCM genes analysed in this study, sub-genic regions of constraint were only identified in *MYH7*, *FLNC* and *TNNC1*. For *MYH7*, there is broad correlation between regions of relative constraint and the HCM and ExAC clusters identified in this study (Figure 3.7). However, utilising both case and control data enables us to quantify the likelihood of pathogenicity in specific regions, as well as detecting pathogenic clusters in other genes for which no regional constraint data exists.

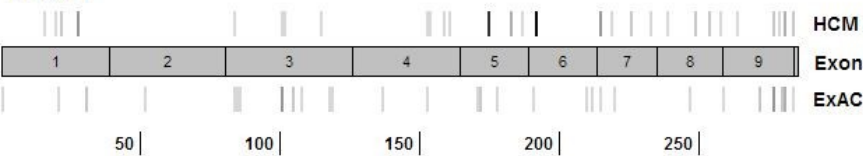
Two recent studies exploring structure-function models in β -cardiac myosin (*MYH7*) have attempted to identify residues that are key to protein function (and therefore susceptible to pathogenic variation), with variants affecting these residues enriched in case over population reference cohorts. Homburger et al. modelled β -cardiac myosin before and after the myosin power stroke and identified the converter kinetic domain and myosin mesa surface area as regions enriched in disease-associated variants¹⁶⁰. Alamo et al. used an interacting-heads motif (IHM) model to define interaction sites between paired myosin heads and other proteins, noting that variants in HCM cases disproportionately alter IHM residues¹⁵⁹. The *MYH7* residues identified by these studies largely overlap with the HCM cluster we have identified by one-dimensional clustering (Figure 3.7), though we do not detect enrichment in N terminal section of the myosin head (and indeed detect a cluster of ExAC variants here). However, particular groups of residues detected by these analyses are highly enriched in disease-associated variants (yielding higher EFs than our cluster), with 7 IHM groups yielding an EF>0.99 and accounting for 44% of variants found in HCM cases¹⁵⁹. EF-based variant analysis thus requires a balance between specificity (strength of EF) and sensitivity (the proportion of case variants encompassed by the identified clusters), or a tiered approach with different confidence levels for pathogenicity.

FLNC has recently been proposed as a relatively frequent cause of HCM¹¹⁸. However the high rate of rare variation in ExAC indicates a substantial proportion of *FLNC* variants detected in HCM patients will not be pathogenic, with only a modest overall EF of 0.44. Although enrichment of case variants towards the C-terminus has been noted in the published studies^{118,119}, no clusters were detected in this study, likely due to the limited cohort size available (448 cases¹¹⁹). Based on exon boundaries alone, the EF for non-truncating *FLNC* variants rises to 0.74 in a broad HCM cluster (exons 24 – 42) and 0.89 in a narrower HCM cluster (exons 39 – 42) (Figure 3.8). However much larger case datasets will be required to confirm and delineate any such clusters in this gene. No clusters were detected in the other HCM genes analysed here (Figure 3.8), although the overall EF for *TPM1* is 0.951 and therefore actionable for novel non-truncating variants per the new rules.

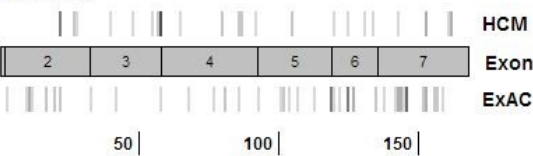
FLNC



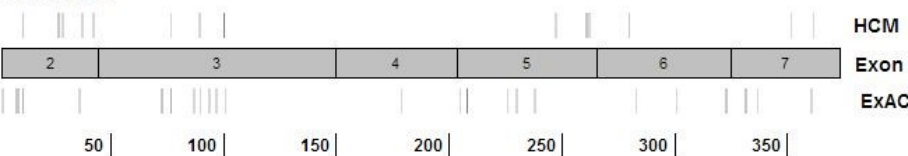
TPM1



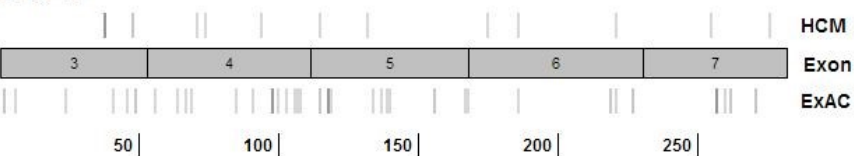
MYL2



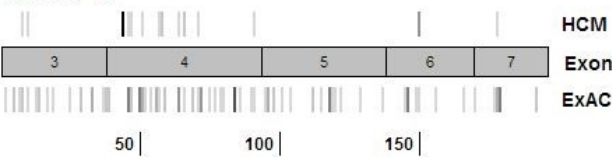
ACTC1



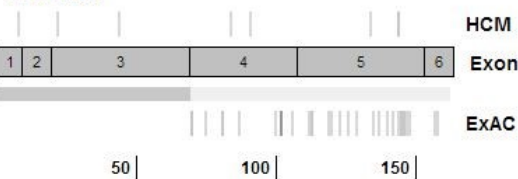
FHL1



CSRP3



TNNC1



PLN

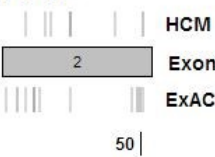


Figure 3.8 (previous page): Distribution of rare, non-truncating variants in HCM cohorts and ExAC for validated genes without significantly enriched clusters using the methods applied in this study (variant density increases with darker shades of grey). For FLNC, putative suggested HCM clusters between exons 39 and 42 and between exons 24 and 42 are highlighted. The overall EF for non-truncating variants in FLNC is 0.44 which rises to 0.74 in the broad HCM cluster (exons 24-42) and 0.89 in the narrow HCM cluster (exons 39-42). Regional constraint boundaries described by Samocha et al.¹⁶¹ are also highlighted.

3.3.3.4 Comparison of clustering and missense functional prediction algorithms

The ability of variant clustering and missense functional prediction algorithms to distinguish between pathogenic and benign variants, and therefore differentiate EFs, were compared for the five HCM genes where clustering was detected (Table 3.4). For effective predictors, the pathogenic prediction (i.e. a variant is within a cluster or classified as deleterious by consensus algorithms) will have a case frequency approaching the case excess and an EF approaching 1 (or at least ≥ 0.95) while the non-pathogenic prediction (i.e. outside of the cluster or classified as tolerated by consensus algorithms) will have a case frequency close to the ExAC background rate and therefore a low or non-significant EF. The results in Table 3.4 show that variant clustering, specific to gene and disease, substantially outperforms the more generic functional prediction algorithms in distinguishing between rare variants in HCM genes.

Table 3.4 (next page): Comparison of the performance of variant clustering and consensus functional prediction scores in differentiating EF values for non-truncating/missense variants in 5 HCM genes where the clustering of case variants was detected. Clustering identified subclasses of variants in 4 genes with an $EF \geq 0.95$ (bold) and generally outperformed consensus functional prediction scores. Fisher's exact p-values for comparison of rare variation in cases and ExAC reference samples were <0.0001 unless otherwise noted. Note – for MYBPC3 (italics), the FATHMM predictor was not included in the Consensus scores due to its poor performance for this gene, which also affected the MetaSVM and MetaLR consensus scores.

Gene	Case Excess	EF (whole gene)	Predictor method	Pathogenic Prediction		Non-Pathogenic Prediction	
				Case freq.	EF	Case freq.	EF
MYH7	12.78%	0.930 (0.923-0.938)	HCM cluster	10.80%	0.975 (0.971-0.980)	3.09%	0.743 (0.703-0.783)
			Consensus	12.55%	0.940 (0.933-0.947)	1.34%	0.783 (0.728-0.839)
			MetaSVM	12.53%	0.944 (0.937-0.951)	1.36%	0.739 (0.675-0.804)
			MetaLR	13.29%	0.944 (0.938-0.951)	0.60% (p=0.0155)	0.406 (0.185-0.627)
MYBPC3	8.14%	0.881 (0.868-0.895)	HCM cluster	7.36%	0.937 (0.927-0.947)	1.99%	0.647 (0.582-0.712)
			Consensus	8.42%	0.904 (0.892-0.916)	0.93%	0.524 (0.379-0.670)
			MetaSVM	4.27%	0.945 (0.934-0.957)	5.08%	0.811 (0.786-0.837)
			MetaLR	1.78%	0.900 (0.874-0.925)	7.57%	0.871 (0.855-0.887)
TNNT2	1.54%	0.912 (0.889-0.935)	HCM cluster	1.31%	0.976 (0.965-0.988)	0.38%	0.684 (0.549-0.820)
			Consensus	1.20%	0.909 (0.880-0.937)	0.49%	0.832 (0.730-0.934)
			MetaSVM	1.11%	0.894 (0.861-0.927)	0.58%	0.905 (0.848-0.961)
			MetaLR	1.11%	0.889 (0.856-0.923)	0.58%	0.921 (0.872-0.971)
TNNI3	1.95%	0.930 (0.912-0.948)	HCM cluster	1.90%	0.974 (0.963-0.984)	0.20% (p=0.0370)	0.502 (0.222-0.783)
			Consensus	1.93%	0.957 (0.943-0.970)	0.17% (p=0.0383)	0.566 (0.280-0.852)
			MetaSVM	1.77%	0.939 (0.921-0.957)	0.33%	0.873 (0.803-0.944)
			MetaLR	1.87%	0.932 (0.913-0.951)	0.23%	0.903 (0.833-0.973)
MYL3	0.70%	0.833 (0.772-0.895)	HCM cluster	0.36%	0.954 (0.920-0.989)	0.48%	0.742 (0.623-0.861)
			Consensus	0.79%	0.869 (0.817-0.921)	0.05% (p=0.6503)	0.310 (-0.629-1.249)
			MetaSVM	0.50%	0.840 (0.763-0.917)	0.34%	0.833 (0.735-0.930)
			MetaLR	0.53%	0.809 (0.722-0.897)	0.31%	0.883 (0.809-0.958)

3.3.4 Variant classification using adapted ACMG guidelines

3.3.4.1 Effect of rule changes on yield of genetic testing in clinical HCM cohorts

To evaluate how the EF-based modified ACMG guidelines could improve the yield of genetic testing in HCM, the proportion of VUS in the OMGL/LMM cohorts with an EF ≥ 0.95 (i.e. non-truncating variants in all of *TPM1* and case-enriched clusters of *MYH7*, *TNNI3*, *TNNT2* and *MYL3*) or inframe indels with $0.90 \leq \text{EF} \leq 0.95$ (case enriched clusters in *MYBPC3*) was calculated. 93 variants in 127 cases for *MYH7*, 17 variants in 33 cases for *TNNI3*, 15 variants in 20 cases for *TNNT2*, 22 variants in 33 cases for *TPM1*, 7 variants in 12 cases for *MYL3* and 1 variant in 2 cases for *MYBPC3* would be upgraded based on this analysis. In all, variants in 4.0% of cases could be upgraded to Likely Pathogenic by activating this strong evidence rule. The yield of pathogenic and likely pathogenic variants in the eight sarcomeric genes would therefore increase by 13.8% in this cohort, from a total of 29.0% of HCM cases with a positive result to 33.0% of cases (Figure 3.9). It should be noted this is a conservative estimate as additional VUS in variant classes with EFs < 0.95 may also be upgraded when the moderate and supporting EF rules are added to existing evidence.

3.3.4.2 Identifying pathogenic variants in prospective HCM cohort

For the prospective cohort of 684 HCM patients described in Chapter 2, variants in the valid HCM genes were classified using the amended ACMG guidelines to assess their pathogenicity. Initially, automatically applied ACMG rules (PM2, PVS1, PS4, PM4, PP3, PM5, PS1) yielded 131 pathogenic / likely pathogenic variants in 131 cases. Manual curation and EF-based rules were then applied to increase the yield of actionable variants (Figure 3.10).

- ClinVar entries and published reports were analysed to identify evidence from family pedigrees that could activate segregation (PP1) or *de novo* inheritance (PM6/PS2) rules. Only four additional variants in four cases were upgraded from VUS by this additional evidence – *MYH7*:p.I263T, *TNNT2*:p.R92Q, *TNNI3*:p.R186Q and *MYL3*:p.M149V. However, the number of samples with variants classified as Pathogenic increased from 9 to 51 when incorporating this evidence, highlighting that while previously published pedigree data only marginally increases the yield of

actionable variants, it can increase the confidence in pathogenicity for a substantial subset of cases.

- The EF-based rules (PM1_supporting, PM1_moderate, PM1_strong) were then applied based on the gene and variant cluster values defined above. 33 variants in 32 cases were upgraded from VUS by this additional evidence over the initial automated rules (i.e. not including curated evidence), including the four variants upgraded by the curated segregation data.

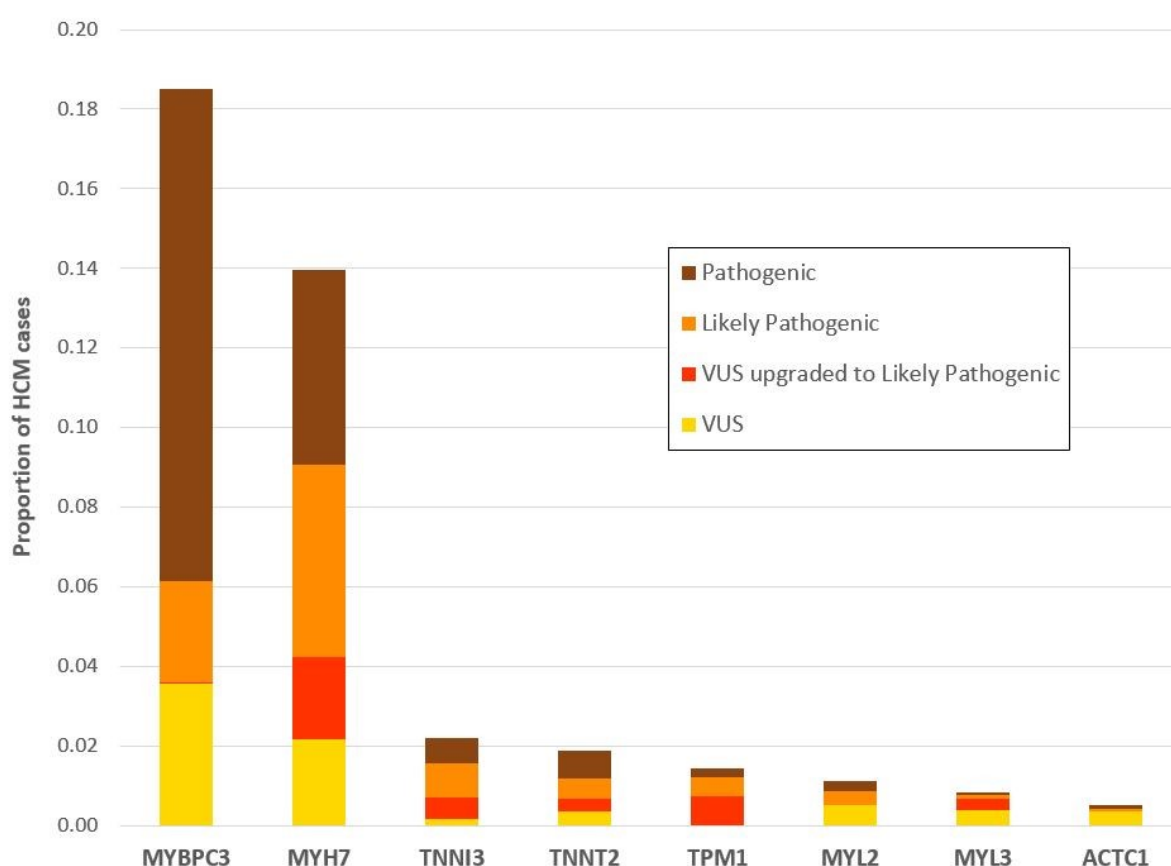


Figure 3.9: Proportion of cases from the OMGL/LMM HCM cohorts with variants in 8 sarcomeric genes (only rare variants, ExAC filtering frequency $< 4 \times 10^{-5}$, are shown, excluding non-essential splice site variants). Variant proportions are coloured according to clinical classification by OMGL and LMM, as well as variants initially classified as VUS that could be reclassified as Likely Pathogenic based on the new ACMG rules.

One additional variant, *MYBPC3*:p.R820Q (present in 2 cases) was upgraded by a combination of segregation and EF-based rules. In total therefore, variants in 30 cases were upgraded based on these adapted guidelines, over and above the standard ACMG rules (automated and curated segregation and *de novo* data). This corresponds to 4.39% of cases tested and an increase of 22% in the number of cases with a positive result (from 19.7% of cases to 24.1%). As well as increasing the yield of actionable variants, the number of cases with variants classified as Pathogenic increased from 51 to 77 by incorporating the EF-based rules. See Appendix (7.2) for the full list of rare variants, including their classification, ACMG rules activated and associated evidence.

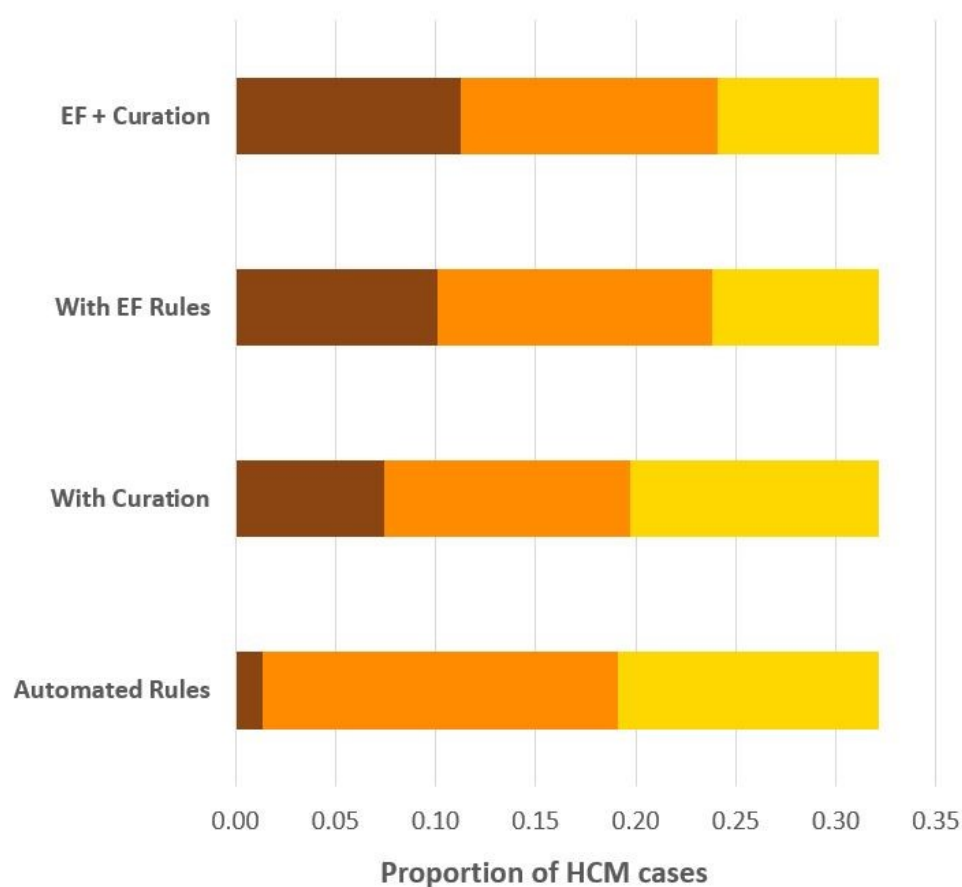


Figure 3.10: Proportion of cases in the prospective HCM cohort classified as actionable based on automated ACMG rules and the addition of manual curation of published evidence and new EF-based rules. 32 extra cases (4.7%) are upgraded with EF-based rules compared to just 4 (0.6%) with manual curation. MYBPC3:p.R820Q (in two cases) required both manual curation and the EF rule to be classified as likely pathogenic.

3.3.5 Identification of non-rare variants enriched in case cohorts

Non-rare variants in the prospective HCM cohort, i.e. those with a filtering allele frequency > 0.004% (and therefore not activating rule PM2) were analysed to see if any was enriched in cases compared to ExAC and HVOL controls, the latter to account for any platform-specific issues around variant calling. 26 non-rare protein-altering variants in the valid HCM genes occurred in more than one HCM patient in our cohort. One variant was significantly enriched in HCM cases with both the all sample and Caucasian only comparisons (after Bonferroni correction for multiple testing, $p < 0.0019$) – *TNNT2*:p.R278C present in 0.73% (0.96%) of cases compared to 0.09% (0.12%) in ExAC and 0% in HVOLs (percentages in parentheses denote Caucasian/ExAC-NFE). Interestingly, one variant, *MYBPC3*:p.V896M, seemed significantly depleted in HCM compared to ExAC (0.44% vs 2.54%). However, the HVOL frequency was lower than ExAC (1.10%), as was the frequency in ESP (0.96%), which indicates that exome coverage issues in ExAC (only 23292 alleles are called at this position) may explain this discrepancy.

The *TNNT2*:p.R278C variant was analysed in depth to investigate a possible role in HCM. The variant frequency in published HCM cohorts was noted (Table 3.5). Given the variant's absence from all Asian cohorts in ExAC (Figure 3.11), the HCM cohorts were split into European or majority Caucasian countries (though a minority of cases in these cohorts are likely to be non-Caucasian) and Asian countries. R278C is present in 0.7% of European/Caucasian HCM cohort samples (52/7181 cases) and in 0.0% of Asian cases (0/668 cases). The variant is present in 0.12% of non-Finnish European individuals in both ExAC and gnomAD exomes and at 0.08% in gnomAD genomes (Figure 3.11) – gnomAD is the updated version of ExAC comprising 123,136 exome sequences and 15,496 whole-genome sequences (<http://gnomad.broadinstitute.org/>). The frequency of R278C in the combined European/Caucasian HCM cohort (Table 3.5) and the non-Finnish European subsets of ExAC and gnomAD were used to assess the significance of enrichment (Fisher's exact test), odds ratios and penetrance. An OR of 6 and a penetrance of 1.2% was calculated when comparing to both ExAC-NFE and gnomAD-NFE (Table 3.6).

Group	Year	Population	Sequenced	R278C	Frequency
EUROPE / CAUCASIAN	2017 ⁸⁹	UK	685	5	0.7%
	2017 *	Italy	1198	8	0.7%
	2015 ³³	USA	2912	16	0.5%
	2015 ⁸⁶	UK	874	8	0.9%
	2012 ¹⁶²	Portugal	77	1	1.3%
	2010 ¹⁴⁵	France	192	4	2.1%
	2009 ¹⁶³	UK	79	1	1.3%
	2009 ¹⁴⁴	Spain	120	1	0.8%
	2009 ¹⁶⁴	Spain	127	2	1.6%
	2005 ¹⁶⁵	Greece	143	2	1.4%
	2005 ¹⁶⁶	Australia	80	1	1.3%
	2003 ⁷³⁻⁷⁵	USA	389	3	0.8%
	2003 ¹⁴⁰	France	197	0	0.0%
	2003 ¹⁴¹	Germany	108	0	0.0%
	Total		7181	52	0.7%
EAST ASIA	2017 ⁸⁹	Singapore	120	0	0.0%
	2013 ¹⁶⁷	China	200	0	0.0%
	2013 ¹⁶⁸	China	136	0	0.0%
	2005 ¹⁴²	China	100	0	0.0%
	2012 ¹⁶⁹	Japan	112	0	0.0%
	Total		668	0	0.0%

Table 3.5: Frequency of TNNT2:p.R278C in published HCM cohorts (* From Olivotto group, manuscript in press).

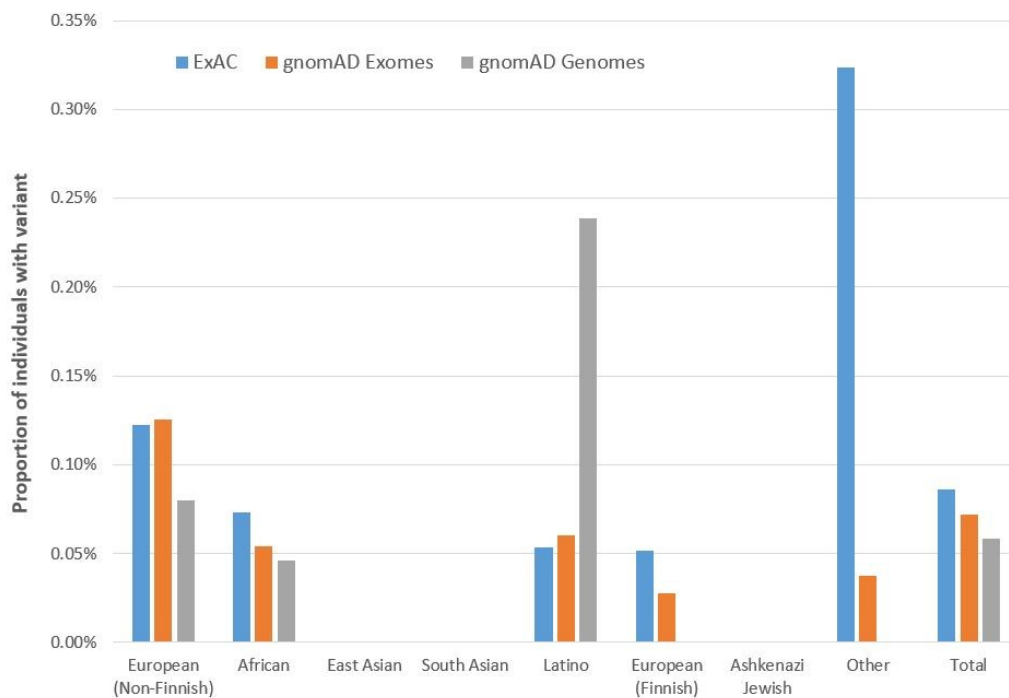


Figure 3.11: Proportion of individuals in subgroups of ExAC and gnomAD with TNNT2:p.R278C.

	HCM	ExAC	gnomAD exomes	gnomAD genomes	gnomAD total
Alleles/individuals	52/7181	33/26897	68/54180	6/7482	74/61662
Sample frequency	0.00724	0.00123	0.00126	0.00080	0.00120
p-value	-	<0.0001	<0.0001	<0.0001	<0.0001
OR (95% CI)	-	5.94 (3.84-9.19)	5.80 (4.04-8.34)	9.09 (3.90-21.17)	6.07 (4.25-8.66)
Penetrance (95% CI)	-	0.012 (0.006-0.022)	0.012 (0.007-0.019)	0.018 (0.006-0.052)	0.012 (0.007-0.020)

Table 3.6: Frequency of TNNT2:p.R278C in HCM, ExAC and gnomAD cohorts, with calculated Fisher's exact p-value, OR and penetrance for HCM versus population cohorts.

These low OR and penetrance values highlight the uncertainty about the clinical significance of R278C and whether such variants would be actionable for cascade screening in families. Published reports and ClinVar evidence summaries about this variant reveal variable expressivity and clinical presentation, reduced penetrance (as expected) and a high proportion of cases with second sarcomeric variants – see details in Table 3.7. It is important to note however that although the population penetrance of R278C is very low, this does not necessarily reflect the risk that carrier relatives of R278C HCM patients have of developing disease, given that they may share additional genetic and environmental influences.

The OR and penetrance of R278C is far lower than would traditionally be considered as actionable by clinical geneticists, given that such a small proportion of disease risk is accounted for by this variant. Nevertheless, if the presence of this variant is necessary for disease expression in these families, and family members are already undergoing regular clinical surveillance, then identifying those relatives who are not variant carriers could release them from ongoing screening. However, this would require the clinical geneticist to be confident enough that R278C was the main driver of disease in the affected family, and that non-carrier relatives were not at risk from additional, and as yet undetected, variants.

Published Reports		
Author	Year	Summary
Ripoll-Vera ¹⁷⁰	2016	Family 12 - proband diagnosed at 18, father at 57, cousin at 37. Family 13 - proband diagnosed at 37, brother at 42, mother (75) unaffected. Family 14 - proband diagnosed at 55, daughter (27) unaffected. Two of the probands had second variants - <i>MYBPC3</i> :p.R502Q and <i>MYH7</i> :p.R723C.
Brito ¹⁶²	2012	The R278C missense mutation was identified in a small family with a mother and daughter both affected. There was a family history of premature (aged <50 years) sudden death in first-degree relatives.
Millat ¹⁴⁵	2010	Second variant, <i>MYBPC3</i> :p.D610H, was detected in 1/4 R278C cases.
Garcia-Castro ¹⁴⁴	2009	R278C patient had second variant - <i>MYBPC3</i> :p.R733H.
Gimeno ¹⁶⁴	2009	Family M92 had 6 R278C carriers of which 3 were affected - all 3 had second variant, <i>MYH7</i> :p.D928N (one <i>MYH7</i> only carrier was also affected).
Miliou ¹⁶⁵	2005	Proband A (AIII:5) was diagnosed with HCM at young age and died suddenly at the age of 16 years (no family history of SCD). Proband B (BIV:1) was diagnosed with HCM at the age of 32 years (family history of SCD at advanced age). Four relatives were R278C carriers, one affected. Cases not tested for additional variants.
Theopistou ¹⁷¹	2004	R278C present in proband (male, diagnosed at 33) and aunt (diagnosed at 64), mother is obligate carrier (diagnosed with HCM at 50, sudden death at 68). Two sons (41, 38) and granddaughter (14) were carriers but unaffected. History of SCD in family (aged 62, 64, 70). Second proband was male diagnosed at 13, SCD at 15, parents and sister (17) unaffected (family not tested genetically). Cases not tested for additional variants.
Garcia-Castro ¹⁷²	2003	The R278C mutation was found in a 60-year-old woman with severe hypertrophy, whereas her sister and her daughter were carriers and showed normal echocardiographic values.
Watkins ¹⁷³	1995	Identified (by screening 26 cases) in 1 proband and 2 relatives (LOD=0.6).
ClinVar Comments		
Laboratory	Summary	
Centenary Institute SCV000212627.1	Variant detected in 3 unrelated families, though 2 families carry an additional pathogenic/likely pathogenic variant in another HCM gene.	
GeneDx SCV000209268.8	Variant has been identified in multiple unrelated individuals referred for cardiomyopathy genetic testing at GeneDx, some of whom also harbour pathogenic or likely pathogenic variants in other cardiomyopathy genes.	
LMM SCV000060277.4	Variant identified by laboratory in many individuals with cardiomyopathy (mainly Caucasian), though approximately half carried a second clinically significant variant in another gene. Many affected individuals carrying R278C present at age 50 or older, suggesting a milder effect.	
Stanford SCV000280532.1	14 of 47 cases (not internal cases) who had sequencing of at least <i>MYH7</i> and <i>TNNT2</i> had another variant - 5 are likely pathogenic/pathogenic, 4 are VUS, other variant details not available to confirm classification.	

Table 3.7: Summary of reports of *TNNT2*:p.R278C from Pubmed and ClinVar.

Conversely, given the high proportion of R278C carriers who have other sarcomeric variants, it may act more like a modifier variant – influencing the expression and severity of disease but not itself the critical determinant of whether family members are at risk. In families with additional likely pathogenic sarcomeric variants (with presumably a far higher estimated penetrance than R278C), this second variant is likely to be the main focus for cascade screening, with R278C providing some additional indication of likely expression or severity. If R278C is the only sarcomeric variant identified, additional genetic and non-genetic factors may determine whether family members will develop HCM. These additional genetic factors may be Mendelian variants of relatively high penetrance in as yet undiscovered HCM genes or other rare or common variants with low effect sizes that can also modify disease expression. Clearly much more research into the effect of R278C, and detailed investigation into carrier families, is needed before it can be confidently used to inform clinical management.

A common variant enriched in cardiomyopathy cases has also been described in South Asian populations¹⁷⁴. A 25 base pair deletion in intron 32 of *MYBPC3* was detected in 10.9% of cases compared to 1.7% of controls, yielding an OR of 7 (broadly similar to *TNNT2*:p.R278C). The variant was shown to lead to exon skipping, with substantial variability in disease severity in cases and their families – symptoms were mild in most carriers but in some, particularly homozygotes, a severe phenotype was observed with several instances of SCD. However, as approximately 4% of the South Asian population are carriers of this variant, it clearly exhibits very low penetrance and may act more as a risk factor for cardiomyopathy and heart failure. Interestingly, the authors speculate that the variant may contribute to the wide disparity in cardiac mortality between southern and northern Indian states, as the frequency of the variant is significantly higher in southern and western states¹⁷⁴.

3.4 Discussion

In this chapter, I have focused on the issues affecting variant interpretation in validated HCM genes and the challenges in developing methods that balance specificity and sensitivity for clinical genetic testing. By utilising large genetic datasets and developing statistically robust quantitative approaches to variant analysis, I have been able to increase the yield of variants that can be confidently classified as disease-causing while keeping within the framework and recommendations of current clinical guidelines. These results demonstrate the importance of taking a disease and gene-specific approach to variant interpretation in order to maximise the yield of pathogenic variants and enable a more consistent, unbiased and equitable approach to genetic testing in HCM.

3.4.1 The limitations of generic approaches to variant analysis

Following the discovery that variants in sarcomeric genes are the major cause of HCM, largely through linkage studies of large affected pedigrees, variant analysis in these genes proceeded through two main phases (similar to many other Mendelian genes and conditions). Novel sarcomeric variants identified through the sequencing of research cohorts were deemed as putatively pathogenic based on permissive criteria – generally the absence in small control cohorts, conservation of affected amino acids in selected species and a predicted deleterious effect by computational algorithms. This was based on a misunderstanding of the rate and frequency of variation in the population for many genes (due to the then absence of resources like ExAC), which assumed that non-polymorphisms (i.e. variants with a MAF < 1%) are likely to be both rarely present in the population and individually rare enough to be plausibly pathogenic (the non-specific nature of computational algorithms was also likely under recognised). Nevertheless, the majority of variants detected in these studies are still likely to be disease-causing, given the relative variant frequencies between case and population cohorts defined here.

As described above, the release of population databases like ExAC has highlighted that many variants associated with Mendelian diseases like HCM are too common to be pathogenic for penetrant conditions and that even very rare variants can be collectively common, which

needs to be taken into account when interpreting variants. As a result of this, the ACMG guidelines were developed in 2015 to implement more stringent approaches to clinical genetic testing and have become widely adopted by clinical genetics laboratories. While successful in reducing false positive results and increasing concordance between laboratories, it is clear that these approaches, with an emphasis on variant-specific evidence from prior studies, have led to overly conservative interpretation and reduced test sensitivity, as well as potentially producing an ethnic bias in genetic testing. While the ACMG guidelines themselves are a sensible and popular framework for variant analysis, quantitative and disease-specific implementation of these guidelines is essential.

3.4.2 Advantages of quantitative, EF-based approaches to variant interpretation

The approach to variant interpretation outlined in this chapter offers a number of advantages compared to other methods:

- It is compatible with the existing ACMG framework, requiring merely the modification of existing rules to be implemented. As these guidelines have become widely adopted since being published, it is critical that novel methods can be seamlessly integrated into existing workflows. Although the rules/points based ACMG approach seems initially at odds with the more quantitative and probabilistic methods developed here, conversion of EF values into a semi-quantitative series of rules ensures compatibility, in a similar manner to the incorporation of segregation data. Other more fully quantitative and statistical approaches to variant classification, based on Bayesian or machine learning methodologies, have been proposed¹⁷⁵ and are continuing to be developed. However, beyond the challenge of producing models of sufficient specificity and sensitivity, deploying such methods in standard clinical practice will be fraught with difficulties in terms of replacing existing guidelines, offering transparency as to how classifications are generated and gaining the trust of end users.
- Critically this method maintains the stringent approach towards variant classification that characterises the ACMG guidelines. These state that likely pathogenic should be defined as a “*greater than 90% certainty of a variant being disease-causing*”⁷¹ but, as we have

seen, the current guidelines prevent uncharacterised non-truncating variants from being classified beyond VUS, regardless of how high the prior probability for pathogenicity is for the variant in question. Our adapted rules correct this discrepancy but actually introduce more stringency for such variants – a 95% probability is required for variants to become actionable in the absence of further supporting evidence, as I believe a 1 in 10 false positive rate is too high for clinical genetic testing and a 95% threshold is more in line with current clinical practice. It is also worth noting that genetic testing can have other uses beyond its primary purpose for family cascade screening, for which the risk of false positives is less critical than reducing false negatives, e.g. eligibility for certain therapies.

- The adapted ACMG guidelines described here produce an increase in the yield of high confidence actionable variants of 14% and 22% respectively in the clinical and research HCM cohorts assessed in this study. This corresponds to a large proportion of the VUS case excess observed in such cohorts, i.e. sarcomeric gene variants that individually could not be deemed disease-causing until now but which collectively could be predicted to include a substantial percentage of pathogenic variants. By identifying variant classes with high prior probabilities of pathogenicity, this approach will reduce the “genetic lottery” of current clinical testing, where the likelihood of a positive result can depend on whether the causative mutation in a family has previously been characterised. It may also begin to address the ethnic bias associated with genetic testing in HCM and other diseases, although in-depth investigation of the genetic architecture of HCM in non-Caucasian populations is still critically needed. Interestingly, the increased yield of actionable variants in sarcomeric genes described here is likely to have a greater impact on HCM genetic testing than all of the previous efforts to identify novel, non-sarcomeric genetic causes in this condition, as described in Chapter 2. This highlights how research resources can be poorly allocated due to the misguided understanding of the genetics of such conditions, where the majority of genotype-negative cases were presumed to have an unknown genetic cause rather than likely to represent non-Mendelian HCM as now seems the case (Chapter 4). It also reflects how research, and in particular scientific publishing, can be overly driven by the pursuit of “novelty” rather than the more detailed and painstaking work of understanding variation in known disease genes.

3.4.3 Limitations of EF-based analyses

Despite this, there are a number of issues to take into account when implementing this approach to variant analysis for Mendelian diseases like HCM.

- EF analysis requires access to sequencing data from large case cohorts which may not be available for rarer conditions or those where genetic sequencing is less established. It also highlights the importance of research groups and clinical laboratories publishing genetic data in an accessible and reproducible format¹⁷⁶. Much of the analysis done here used published data from the OMGL and LMM clinical laboratories, as well as other published datasets. Sharing and re-analysis of such data is critical in gaining further insights into the genetic aetiologies of Mendelian disease.
- Although we have identified highly pathogenic variant classes in a number of HCM genes, for others it is more challenging to sufficiently differentiate between benign and pathogenic variation, either due to limited case data (e.g. *FLNC*) or the relatively high rate of rare population variation (e.g. *MYBPC3*). Non-truncating variants in *MYBPC3*, which are causative in up to 10% of HCM cases, will still be classified as VUS in the absence of additional variant-specific evidence, despite the identification of some case-enriched clusters. Research into other methods to discriminate between pathogenic and benign variation will be required for such genes, which could include protein structure analysis or the development of computational prediction techniques that are specific to (and validated in) key disease genes, given the limitations of the generic and consensus scores that we have observed in this study.
- EFs are dependent on the case and control or population cohorts used to generate them, both in terms of their composition and the sequencing technology and analysis applied to generate variant data. For cases, the use of clinical referral cohorts (as we have largely employed in this study) will produce more conservative EF values than cohorts with definitively diagnosed patients or higher proportions of familial cases, but I believe these more cautious, referral EFs will be more relevant to their use in a clinical genetics setting. While ExAC is not strictly a control cohort, I believe the advantages of its scale outweigh issues of exome sequencing coverage and the likely inclusion of a small number of cases

(the variable coverage of ExAC was also adjusted for in this study). This is particularly important for smaller genes and those more constrained of rare genetic variation which require very sizable control cohorts to accurately estimate their frequency of rare variation.

- The EF-based rules described here can only be applied to assessing variants from patients with the disease in question, as EFs correspond to the probability of pathogenicity given the presence of disease. Variants detected as incidental findings in other individuals, or through any future population screening studies, will need to be assessed in a more stringent manner, given the low prior probability of disease in the individuals tested.
- Although I have taken care to ensure that this approach remains within the framework and recommendations of current guidelines, it represents a new way of interpreting variation in genetic disease. Classifying a variant as likely pathogenic despite limited evidence about the specific variant may prove disconcerting to clinical geneticists who are more accustomed to relying on evidence such as segregation in affected families and will require careful explanation of the benefits and drawbacks of this approach. The EF values also explicitly state the level of uncertainty and proportion of false positives associated with this type of interpretation. However, it is important to recognise that such uncertainty also exists with current classifications, even if the lack of quantification may conceal the extent of this. Communicating the level of confidence and uncertainty associated with the results of genetic testing to both clinicians and patients can be one of the most challenging aspects of genomic medicine. The EF-based analysis provides the basis of a more quantitative way of defining variant pathogenicity that I believe can increase the transparency around variant interpretation.

3.4.4 Future challenges in variant interpretation

Although the methods outlined in this chapter lead to a substantial increase in the yield of actionable variants in HCM, and bring a more quantitative and coherent approach to the ACMG guidelines, it is important to note that these are designed to be complementary to current variant interpretation techniques rather than a replacement for them. Variant-

specific evidence, in particular prior observations of segregation in affected families, will continue to be integral to variant interpretation, both for variant classes with an EF<0.95 and to provide further supporting evidence for high EF variants that can elevate classifications from Likely Pathogenic to Pathogenic. Such data not only provides strong and direct evidence for the effect of a variant, but has been the basis of variant interpretation in clinical genetics for many years, such that clinical geneticists and genetic counsellors are experienced and comfortable with applying it. On the other hand, as well as impacting on test sensitivity as described, this evidence relies on trusting that published evidence is accurately reported (both phenotype and genetics data) and can be difficult and time-consuming to access. It should also be noted that, at least for the prospective HCM cohort in this study, such evidence was decisive in the classification of only 4 variants/cases. All of the other segregation data relevant to this cohort related to variants that were statistically enriched in cases and therefore were already classified as actionable by activation of the PS4 rule. The limited added value of this evidence class further highlights the necessity for developing approaches like those described here.

Although some progress has been made in communicating this variant-specific evidence between clinical genetics laboratories, further efforts are needed to expand this data sharing, as well as seeking improvements in databases like ClinVar that allow such data to be disseminated¹⁷⁶. Data submission to ClinVar has been cumbersome and likely off-putting for clinical genetics laboratories that lack extensive bioinformatics support. Accessing this database is also hampered by the poor design of its web interface, which can make it time consuming to find key evidence amongst the repetitive and often irrelevant data in the evidence summaries submitted. Additionally, disease-specific initiatives to curate the evidence for recurrent pathogenic variants, such as the *MYH7* curation effort by ClinGen (manuscript in press) or the knowledge base developed by our laboratory to complement the CardioClassifier resource¹⁵⁶, will play an important role in unambiguously defining the status of important subsets of variants.

The EF-based approach described here could also be the basis for a more comprehensive quantitative framework for variant interpretation. With the EF as the prior for particular

variant classes, additional evidence could be quantitatively assessed to increase or decrease this probability of pathogenicity in a Bayesian manner. Our group has been involved in a pilot study for such an approach, mainly focusing on variants in Long QT syndrome, though that study was based on limited population variant data and did not incorporate ACMG rules and guidelines¹⁷⁵. An approach which is complementary to the ACMG guidelines, and produces a quantitative score from the evidence in activated rules, may provide more granularity to the overall variant classification and increase the confidence and transparency associated with the output of this analysis. It could also potentially highlight where certain rules or evidence types are applied with inappropriate weighting. By integrating with these existing and widely adopted guidelines and classifications, I believe such a methodology has a greater chance of being implemented in clinical genetics environments than completely bespoke quantitative approaches based on Bayesian or machine learning methods.

There has been substantial progress made in recent years in defining sufficiently stringent population frequency thresholds that are compatible with pathogenicity for variants in Mendelian diseases. Methods to define disease-specific thresholds that account for penetrance and the genetic heterogeneity of diseases, utilising the power of resources like ExAC and gnomAD, have been developed by our group^{88,147}. These stringent thresholds have been shown to be consistent with the vast majority of known pathogenic variants in HCM¹⁴⁷. However, as described above, certain variants such as *TNNT2*:p.R278C, with higher population frequencies but nevertheless significantly enriched in cases, may have a role to play in HCM. Although these variants are calculated to have a very low penetrance, and are therefore consistent with these frequency threshold approaches (<http://cardiodb.org/allelefrequencyapp/>), they introduce new challenges to variant interpretation and clinical management. Identifying such variants requires careful and statistically robust analysis of large case cohorts, ideally restricted to individuals of specific populations with matched control cohorts. Once identified and defined, these low penetrance variants will need cautious assessment as to their potential utility in cascade screening of the families of affected variant carriers. While variable penetrance is frequently observed in HCM and similar conditions, dealing with such low penetrance variants will require new approaches and guidelines – should unaffected variant carriers be offered routine clinical assessment and

are non-carrier family members at such low risk that they can be released from ongoing surveillance? Research into identifying genetic and non-genetic modifiers that may influence disease expressivity in such cases may illuminate the role of such variants in HCM and how they should be clinically managed.

As highlighted in the introduction to this chapter, much of the research into the genetics of HCM and other cardiomyopathies has been done in western countries with Caucasian-majority populations. This is likely to be one of the reasons for a lower yield of actionable variants in non-Caucasians which was one of the drivers for developing the EF-based approach described here. However, there may also be distinct features in the genetic architecture of HCM in other populations, as observed in the unique aetiologies of countries like the Netherlands¹⁷⁷ and Iceland⁹⁶. Regions with high rates of consanguinity, such as North Africa and the Middle East, may exhibit recessive forms of the disease, either affecting known sarcomeric genes or novel genes where dominant mutations are not disease-causing. Population-specific founder, recurrent or low penetrant variants may also be prevalent in different regions of the world which will require the collection and sequencing of large non-Caucasian case and control cohorts to identify and characterise such factors.

The techniques described here for variant classification in HCM are also potentially applicable to other cardiomyopathies (as well as other Mendelian diseases in general). As shown in Chapter 2 and our associated publication⁸⁸, there is also an excess of VUS over ExAC for several DCM and ARVC genes, indicating an under-calling of pathogenic variants similar to HCM. However, truncating variants are more prevalent in these conditions than in HCM, e.g. *TTN*, *DSP*, *LMNA*, *VCL* in DCM and *PKP2*, *DSP*, *DSG2*, *DSC2* in ARVC⁸⁸. Given the relative frequencies of truncating variants between case and population cohorts, and the activation of the PVS1 rule where appropriate for the particular gene and disease, the classification of these variants is less challenging than for non-truncating variants. Conversely, the lower case frequencies of non-truncating variants in DCM/ARVC genes (compared to HCM) may make it more difficult to identify variant classes with high EF values, at either the whole gene or regional cluster level. Recent studies exploring the genetic architecture of LVNC, involving broad panel or exome sequencing of case cohorts (of admittedly still limited sizes)^{178,179}, also

offer the potential to also expand this analysis to this genetically understudied cardiomyopathy. As emphasised in this chapter, it is likely that specific approaches to variant interpretation will be needed for the individual characteristics of each gene and disease, albeit utilising broadly similar techniques such as the EF analysis presented here.

3.5 Conclusion

Variant interpretation in HCM and other Mendelian diseases has long struggled to balance specificity and sensitivity when applied to genetic testing. Although current guidelines understandably place a premium on avoiding false positive results, it is incumbent on the research community to develop methods that increase the sensitivity of genetic testing to allow us to more fully and equitably exploit the benefits of genomics in conditions like HCM. The results described in this chapter emphasise the importance of applying disease and gene-specific techniques, datasets and expertise to variant interpretation, while maintaining rigorous statistical approaches to avoid the erroneous associations that have previously afflicted HCM genetics. While we have made significant progress in improving variant classification in HCM, efforts to enhance our understanding of sarcomeric gene variation are ongoing and will be dependent on the continued generation and sharing of sequencing data from clinical cohorts and the further development of new analysis techniques.

4 Genotype-phenotype studies in HCM

In this chapter, I utilise a large cohort of HCM patients with gene panel sequencing and cardiac magnetic resonance imaging (CMRI) to investigate the association between the genotype of a patient (as defined by the presence or absence of disease-causing variants in known HCM genes) and their demographic and clinical characteristics. Building on the insights into HCM genetics from Chapters 2 and 3 and benefitting from the largest CMRI cohort yet described, I aim to thoroughly assess the effect of different classes of disease-causing variants on intermediate clinical phenotypes, integrate these findings with published studies and evaluate how the insights into HCM gained from this project may impact on genetic testing and clinical management for HCM patients and their families.

4.1 Introduction

4.1.1 Assessing the correlation between genotype and phenotype

Genotype-phenotype studies aim to assess the correlation between the genotype of patients (those with a variant or class of variants that have been associated with a disease) and the phenotypic expression of that disease, i.e. particular clinical characteristics or the severity of the condition and its outcomes. The genetic variants assessed in these studies may be common variants known to increase the risk of developing a disease, or rare variants that have been shown to be the primary cause – for Mendelian diseases like HCM, most genotype-phenotype studies explore the latter. The ultimate aim of these studies is to enable the knowledge of a patient's genetic status to influence clinical management of their condition – by impacting on the choice of treatment options or identifying patients either at risk of developing a severe form of the condition or at risk of particular clinical outcomes such as SCD in HCM.

4.1.2 Genotype-phenotype studies in HCM

There have been substantial efforts over the last 20 years or more to demonstrate a correlation between the genotype of HCM patients and their demographic characteristics,

disease severity and outcomes. Some early studies suggested distinctive phenotypic features were associated with particular variants or genes, e.g. *TNNT2* variants with a high degree of SCD¹⁷³ or *MYBPC3* variants with delayed disease onset¹⁸⁰. However, such findings were not always replicated in subsequent studies, with specific and consistent correlations between genotype and phenotype rarely observed. In addition to the intrinsic heterogeneity that characterises HCM, this may have been due to limited cohort sizes analysed, differing cohort composition (with respect to disease severity or the proportion of familial cases), inconsistent gene selection or variant analysis and discrepancies in the measurement of phenotype markers. A meta-analysis of 13 genotype-phenotype studies by Lopes et al. (ranging in size from 46 to 471 patients for a total of 2459) was performed in 2013 to identify statistically robust genotype-phenotype associations¹⁸¹ (Table 4.1). Together with several larger studies that have since been published^{86,182–185} (Table 4.1), a more consistent picture has emerged of the distinctive phenotypic features associated with different HCM genotypes.

For such studies, phenotype can refer to a variety of patient characteristics including demographics, baseline clinical attributes or clinical outcomes from longitudinal studies. For HCM, the demographic details regularly assessed include gender, age (either at diagnosis or inclusion in study) and family history (FH) of HCM or SCD. Traditionally, echocardiography has been used to measure clinical features of left ventricular dimensions and morphology. The most widely assessed marker is the maximum left ventricular wall thickness (LVWT) which is both diagnostic for HCM and a measure of disease severity. CMRI has become more widely used in recent years in both research and clinical settings and confers several advantages over echocardiography. As well as affording more accurate measurement of volumes and LVWT, CMRI enables the quantification of left ventricular mass indexed to body surface area (LVMI) and late gadolinium enhancement (LGE), presumed to be a marker of intramyocardial fibrosis which may be a predictor of SCD in HCM patients¹⁸⁶. For outcome studies, a number of end-points are typically used for HCM – mortality (all, cardiovascular or heart failure related), SCD or aborted SCD and combined heart failure events.

Centre	Mayo Clinic, USA	UCL, UK	Toronto General Hospital, Canada	Erasmus Medical Centre, Netherlands	Centenary Institute, Australia	Meta-analysis (13 cohorts)
Study	Bos et al. 2014 ¹⁸²	Lopes et al. 2014 ⁸⁶	Li et al. 2014 ¹⁸³	van Velzen et al. 2016 ¹⁸⁴	Ingles et al. 2013 ¹⁸⁵	Lopes et al. 2013 ¹⁸¹
Cohort Size	1053	874	558	512	265	2459
Genes tested	9	8	8	Large panel	10	-
Genotype positive (%)	34.1%	43.8%	35.5%	45.7%	52.1%	-
Demographics						
Age at inclusion		45.8±14.7 vs 53.1±14.9 p<0.001		46±15 vs 55±15 p<0.001	47±16 vs 56±17 p<0.001	
Age at diagnosis	36.4±17 vs 48.5±18 p<0.001		39.5±15.2 vs 48.5±14.8 p<0.001		34±17 vs 44±18 p<0.001	38.4±10.3 vs 46.0±10.4 p<0.001
Gender (% male)	58.5% vs 60.4% p=0.6		56.1% vs 70.3% p=0.001	67.9% vs 61.5% p=0.13	53.6% vs 70.9% p=0.005	57.5% vs 61.5% p=0.422
FH HCM (%)	50.4% vs 22.9% p<0.001	39.8% vs 15.6% p<0.001	52.5% vs 20.0% p<0.001		73.9% vs 30.7% p<0.001	50.6% vs 23.1% p<0.001
FH SCD (%)	27.0% vs 15.0% p<0.001	28.5% vs 15.2% p<0.001	16.7% vs 7.2% p=0.002	19.7% vs 5.4% p<0.001	41.3% vs 6.3% p<0.001	27.0% vs 14.9% p<0.001
Baseline characteristics						
Max LVWT (mm)	22.6±6 vs 20.1±5 p<0.001	18.8±4.4 vs 18.1±4.1 p=0.004	20.8±4.8 vs 19.6±4.9 p=0.16 (adjusted)	20±5 vs 18±4 p<0.001	22±6 vs 21±5 p=0.03	21.0±4.1 vs 19.3±3.5 p=0.03
Hypertension (%)	19.2% vs 43.2% p<0.001		25.3% vs 46.9% p<0.001			22.8% vs 41.7% p=0.004
Clinical outcomes - hazard ratio (log rank p-value)						
Mean follow-up years			6.6±6.3 / 6.2±5.6	12±9		
CV mortality		3.99 (p=0.001)		2.82 (p=0.002)		
HF-related mortality				6.33 (p=0.004)		
SCD / aborted SCD		3.44 (p=0.028)		2.88 (p=0.015)		
Combined HF events			4.51 (p<0.001)			

Table 4.1 (previous page): Summary of demographic, baseline clinical and outcome data for major published HCM studies comparing genotype-positive and genotype-negative patients. Genotype-positive status is defined by the presence of a putatively pathogenic variant – the vast majority of these occur in one of the 8 core sarcomeric genes (MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1). The meta-analysis by Lopes et al. involved 13 previously published cohorts - this included subsets of the cohorts from the Mayo Clinic and Toronto General Hospital that had been published prior to the larger versions of those cohorts described here.

Two types of HCM genotype-phenotype studies are generally performed depending on the definition of genotype - genotype-positive versus genotype-negative studies (referring to the presence or absence of causative, usually sarcomeric, variants) or studies comparing the effects of variants in different genes or classes of genes. In the former, genotype-positive cases are those with a putative disease-causing variant in a HCM gene - the vast majority of these variants occur in the eight core sarcomeric HCM genes (*MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3, ACTC1*), though some studies may have included small numbers of additional variants in other putative HCM genes. Genotype-negative cases are those without a putative disease-causing variant in the sequenced genes. The absence of an identified pathogenic variant does not necessarily indicate a non-genetic form of HCM - the disease in these cases could be due to protein-coding variants in as yet unknown or recently described HCM genes (e.g. *FLNC*^{118,119}), variants classes such as CNVs that are not detectable with standard genetic sequencing or non-protein coding variants (such as those affecting regulatory sequences) which may not be sequenced or interpretable in standard testing. Nevertheless, given the low proportion of cases with FH of HCM, and the distinctive phenotype that will be described, it is expected that a significant proportion of genotype-negative HCM cases will not have a Mendelian genetic aetiology.

4.1.3 Genotype-positive versus genotype-negative

Based on the recent analysis of several large cohorts (and the meta-analysis of earlier studies) (Table 4.1), genotype-positive HCM cases have been shown to be diagnosed significantly younger (on average between 8 and 10 years) and have a greater FH of both HCM and SCD than genotype-negative cases. They are also associated with increased disease severity as measured by a greater maximum LVWT, with mean differences ranging from 0.7mm to 2.5mm. The studies that assessed the presence of hypertension found it was significantly more prevalent in genotype-negative cases. Three studies assessed clinical outcomes between the two groups and observed that genotype-positive cases were associated with more adverse outcomes, including mortality^{86,184}, SCD or aborted SCD^{86,184} and combined heart failure events¹⁸³. A CMRI study of 245 HCM patients also found that LGE was detected in significantly more genotype-positive (75%) than genotype-negative (53%) patients¹⁸⁷.

4.1.4 Effect of specific variant class

While genotype-positive cases present with more severe characteristics overall compared to genotype-negative cases, the effect of variants in different genes or classes of genes can be highly variable. This is likely due to the limited cohort sizes available for gene-level analysis (only 30-60% of HCM cases will have an identifiable causative variant), exacerbating the other potential confounding factors described above. Inconsistencies in study design, e.g. some studies compared carriers of variants in one gene with all other cases rather than carriers of variants in other genes, adds to the difficulty in obtaining comparable findings. Some of the large cohort studies (and meta-analysis) described above^{181–183,185} have attempted to clarify these associations (Table 4.2). Additionally, some recent studies have utilised CMRI to distinguish between the effects of *MYBPC3* and *MYH7* variants¹⁸⁸ and between carriers of variants in thin filament genes (*TNNT2*, *TNNI3*, *TPM1*, *ACTC1*) and thick filament genes (*MYBPC3* and *MYH7*)¹⁸⁹.

Comparing *MYBPC3* and *MYH7* variant carriers, no consistent differences have been observed for age of presentation, FH of HCM or SCD or maximum LVWT. Some studies have suggested that patients with *MYBPC3* variants tend to present later and have a less severe disease

course^{180,190}, consistent with the fact that most HCM founder mutations occur in *MYBPC3* (the maintenance of which in the population suggests limited impact during the child bearing years of early adulthood)^{96,174}. However, evidence from larger cohort studies is mixed, with only one observing a significant later onset for *MYBPC3* carriers (Table 4.2). In contrast, although earlier studies have reported no difference in the proportion of males between *MYBPC3* and *MYH7* variant carriers, recent studies have shown a consistently lower proportion of males in *MYH7* cases compared to *MYBPC3* (Table 4.2). A detailed study on CMRI phenotypes between *MYBPC3* and *MYH7* carriers revealed no significant differences in LV volumes (end diastolic/systolic volumes, ejection fraction), LVMI, left atrial volume, morphology, mitral valve leaflet lengths and LGE metrics¹⁸⁸. The study also found no difference in short-term outcomes (SCD, septal reduction, LVEF < 50%, atrial fibrillation, mean follow-up of 23.5 months), in agreement with an earlier longer-term study that assessed heart failure-free survival over a mean follow-up of 6.3 years¹⁸³. However, none of the studies comparing the effects of variants in *MYBPC3* and *MYH7* distinguished between truncating and non-truncating variants in *MYBPC3*.

A recent study by Coppini et al. compared the clinical features and outcomes of patients with variants in thin filament genes (80) and thick filament genes (150), a subset of whom were assessed by CMRI¹⁸⁹. Thin filament variants were associated with lesser maximum LVWT and more atypically distributed hypertrophy, a lower prevalence of LV outflow tract obstruction (LVOTO), smaller LVMI, a lower proportion of LGE and a lower LVEF. Although mortality and SCD rates did not differ between the cohorts, patients with thin filament variants were more likely to progress to NYHA class III and IV and to develop advanced LV dysfunction. In a CMRI comparison of 31 patients with thin filament variants and 128 patients with thick filament variants, Weissler-Snir et al. also observed a lower prevalence of LVOTO and a lower LVEF in thin filament patients, although no statistically significant differences in LVWT, LGE, LV mass and morphology were found¹⁸⁸. Both studies demonstrated a significant difference in FH of SCD between the two groups, though in opposite directions (36% vs 18% and 10% vs 33% for thin and thick filament groups respectively). No differences in age, gender of FH of HCM were observed in either study.

Centre	University Health Network, Toronto	Mayo Clinic, USA	Toronto General Hospital, Canada	Centenary Institute, Australia	Meta-analysis (13 cohorts)
Study	Weissler-Snir et al. 2017 ¹⁸⁸	Bos et al. 2014 ¹⁸²	Li et al. 2014 ¹⁸³	Ingles et al. 2013 ¹⁸⁵	Lopes et al. 2013 ¹⁸¹
Cohort Size	358	1053	558	265	2459
MYBPC3 / MYH7	75 / 53	182 / 121	88 / 74	53 / 44	-
Demographics					
Age at inclusion	43.9±14.3 vs 41.8±16.4 p=0.44			52±16 vs 42±14 p=0.0035	
Age at diagnosis		37.6±15 vs 36.4±19 p=0.5		40±17 vs 27±15 p=0.0036	37.6±10.5 vs 35.3±13.2 p=0.316
Gender (% male)	65% vs 40% p=0.004	64% vs 51% p=0.03		63% vs 47% ns	
FH HCM (%)		50% vs 49% p=0.8		74% vs 77% ns	46.5% vs 50.9% p=0.284
FH SCD (%)	33% vs 32% p=0.88	25% vs 28% p=0.7		44% vs 45% ns	31.6% vs 31.0% p=0.651
Baseline characteristics					
Max LVWT (mm)	20.6±5.4 vs 20.6±6.1 p=0.99	22.9±6 vs 22.3±6 p=0.4		23±7 vs 23±5 ns	21.0±3.8 vs 19.2±3.7 p=0.574
LVMi (g/m ²)	76.7±27.2 vs 71.6±26.2 p=0.28				
LGE mass (g)	4.1 (0-13.2) vs 2.53 (0-7.2) p=0.41				
Clinical outcomes - hazard ratio (log rank p-value)					
Mean follow-up years	1.96		6.6±6.3 / 6.2±5.6		
SCD / aborted SCD	ns				
Septal reduction	ns				
Atrial fibrillation	ns				
LVEF <50%	ns				
Combined HF events			p=0.949		

Table 4.2 (previous page): Summary of demographic, baseline clinical and outcome data for major published HCM studies comparing MYBPC3 and MYH7 variant carriers. The meta-analysis by Lopes et al. involved 13 previously published cohorts - this included subsets of the cohorts from the Mayo Clinic and Toronto General Hospital that had been published prior to the larger versions of those cohorts described here. ns = non-significant.

4.1.5 Effect of multiple pathogenic variants

One of the most consistent findings on the association between genotype and phenotype in HCM relates to the effect of multiple pathogenic variants, i.e. homozygosity (two copies of the same pathogenic variant), compound heterozygosity (two different pathogenic variants in the same gene) or double mutations (pathogenic variants in different genes). Several studies have shown that patients with more than one pathogenic variant are more likely to have a severe phenotype – earlier onset, more severe LV hypertrophy and a higher incidence of heart failure and SCD^{48,73,191}. An analysis of published cohorts by Fourey et al.⁴⁹ observed that the prevalence of double mutations ranged between 3% and 19% of all genotype-positive patients. However, this study showed that when these published cohorts were re-analysed according to the recommended stricter guidelines for ascertaining pathogenicity⁷¹, the prevalence of double pathogenic variants dropped from 8% to 0.4%, the same proportion observed in their retrospective cohort of 1,411 cases⁴⁹. This analysis suggests that multiple pathogenic variants are less common than previously believed, with most “second hits” actually corresponding to VUS. However, published cases where two variants were correctly classified as pathogenic or likely pathogenic were indeed associated with more severe phenotypes, in particular those with homozygous or compound heterozygous truncating variants in *MYBPC3* (effectively “human knockouts” for this gene).

4.1.6 Ongoing challenges in HCM genotype-phenotype analysis

While numerous studies have investigated genotype-phenotype associations in HCM, additional efforts are needed to further clarify the effects of causative genetic variants on

disease expression. In particular, only a small number of studies thus far have utilised CMRI, limiting findings into the effects of genotype on LGE and LVMI. With the largest cohort yet assembled of HCM probands with both genetic sequencing and CMRI (656 patients), we can thoroughly assess the effect of genotype on these key factors, in addition to more accurate quantification of maximum LVWT values. Recent advances in understanding the genetics of HCM^{88,89,147}, and the population frequencies of variants in HCM genes⁶⁸, also allows us to more accurately classify genotype-positive cases (Chapters 2 and 3). Additionally the large cohort size will enable us to better distinguish between the effects of specific classes of variants, including between truncating and non-truncating variants in *MYBPC3* which have not been assessed in any prior study, and attempt to clarify some of the discordant findings between published studies.

4.2 Materials and Methods

4.2.1 Patient cohort

The cohort in this study consisted of 656 unrelated HCM patients who have undergone both gene panel sequencing and CMRI screening, i.e. a subset of the cohort described in Chapters 2 and 3. The cohort was 73.2% male, with an average age (at inclusion in the study) of 55.4 ± 15.7 years. 26.0% had a FH of HCM and 26.1% had a FH of SCD (percentages exclude those cases for which information on FH was unavailable).

4.2.2 Defining genotype groups

The modified ACMG guidelines used in Chapter 3 to define pathogenic and likely pathogenic variants were also applied here to define the genotype-positive group and included variants in the following genes - *MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, *ACTC1*, *CSRP3*, *FHL1*, *PLN*, *CRYAB* and *TNNC1*. Additional variants were classified as VUS if they were rare in ExAC (activating the PM2 rule with a filtering allele frequency $< 4 \times 10^{-5}$) and were of the following variant type, i.e. classes that have been shown to be pathogenic in HCM:

- Non-truncating variants (missense, in-frame insertions and deletions) in *MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, *ACTC1*, *CSRP3*, *FHL1* (hemizygous only), *PLN* and *TNNC1*.
- Truncating variants (frameshift, nonsense, essential splice site) in *MYBPC3*, *TNNT2*, *CRYAB*, *CSRP3*, *FHL1*, *PLN*, and *TNNC1*.
- Splice region variants in *MYBPC3*.

Genotype-negative samples were defined as those lacking variants described above as well as lacking variants of the following classes:

- Rare variants in HCM genes of the variant types not included above (variant classes not currently considered as potentially pathogenic in HCM).
- Non-rare variants in HCM genes, i.e. filtering allele frequency $> 4 \times 10^{-5}$ but an ExAC frequency < 0.001 .

- Rare variants in minor HCM genes (*ACNT2*, *MYH6*, *MYOZ2*, *TRIM55*, *TRIM63*) – as described in Chapter 2, these genes are likely to yield variants that are uninterpretable but very unlikely to be pathogenic.
- Variants in metabolic cardiomyopathy genes (*GAA*, *GLA*, *LAMP2*, *PRKAG2*, *TTR*) with an ExAC frequency $< 1 \times 10^{-4}$ that may be pathogenic for these genocopy conditions.

Although these variant classes are, with the exception of those in the metabolic cardiomyopathy genes, unlikely to be disease-causing, their removal from the genotype-negative group allows us to generate a “cleaner” set of non-Mendelian HCM samples for comparison with genotype-positive cases. For a separate analysis, this genotype-negative group was sub-classified into familial cases, i.e. those with a FH of HCM or SCD and non-familial cases, i.e. those with no FH of HCM (cases where this data was unavailable were removed from this analysis).

Patients in the genotype-positive group, i.e. those with pathogenic or likely pathogenic variants were then sub-classified into the following groups to assess for phenotype differences based on variant class:

- *MYH7*.
- *MYBPC3* truncating.
- *MYBPC3* non-truncating.
- thin filament genes (*TNNT2*, *TNNI3*, *TPM1*, *ACTC1*, *TNNC1*).
- other HCM genes (*MYL2*, *MYL3*, *CSRP3*, *PLN*, *FHL1*, *CRYAB*).
- multiple variants if more than one pathogenic variant was detected.

4.2.3 Genotype-phenotype comparisons

Based on the genotype groups described above, the following comparisons were performed:

1. Genotype-positive vs genotype-negative cases.
2. Familial genotype-negative vs non-familial genotype-negative cases.

3. Genotype-positive subclasses – a) thick filament genes vs thin filament genes, b) *MYBPC3* vs *MYH7* and c) a comparison between the major genotype-positive subgroups (*MYBPC3* truncating, *MYBPC3* non-truncating, *MYH7*, thin filament genes).
4. The effects of multiple sarcomeric variants on phenotype.

The following phenotype characteristics were assessed – gender, age (at enrolment in study), FH of HCM and SCD, maximum LVWT (as a continuous variable and the proportion of cases with extreme LVWT, defined as $\geq 30\text{mm}$), LVMi (LVM mass in grams indexed to body surface area), LGE (presence/absence, a continuous variable of the proportion of LVM and the proportion of cases with extreme LGE, defined as $>15\%$ of LVM) and the proportion of cases with LVOTO.

Continuous variables were expressed as mean $\pm$ standard deviation. Significance was assessed by student t-test (comparing two groups) or ANOVA (comparing >2 groups) for normally distributed data (maximum LVWT) and by Mann-Whitney U test (comparing two groups) or Kruskal-Wallis test (comparing >2 groups) for non-normally distributed data (age, LVMi and LGE). Categorical variables were expressed as the number of cases and percentage of the total for which the measurement was available, with statistical significance assessed by Fisher's Exact test (comparing two groups) or chi-squared test (comparing >2 groups). All statistical analysis and data visualisation was performed in R.

4.2.4 Using genotype-phenotype studies to validate prior findings in this study

In Chapter 2, I showed that there is currently no genetic evidence supporting a role in HCM for the majority of non-sarcomeric genes that have been implicated in the disease. Genotype-phenotype analysis can provide supporting evidence for these findings. Firstly, if these non-sarcomeric genes are valid HCM genes, we would expect patients with rare variants in these genes to be phenotypically more similar to patients with sarcomeric gene variants than to genotype-negative patients. To test this, we compared the demographic and CMRI characteristics between three sub-groups – the genotype-positive cohort described above, cases with rare (ExAC frequency $< 0.01\%$) variants in the non-sarcomeric genes for which we

found no genetic evidence for association with HCM (*ANKRD1*, *CALR3*, *CASQ2*, *CAV3*, *FHL2*, *FXN*, *JPH2*, *KLF10*, *LDB3*, *LMNA*, *MYLK2*, *MYPN*, *NEXN*, *PDLIM3*, *TCAP*, *TRIM54*, *VCL*) or where variants are unlikely to be pathogenic (*ACTN2*, *MYH6*, *MYOZ2*, *TRIM55*, *TRIM63*) and a genotype-negative cohort that did not include any of these variants (or non-rare variants in HCM genes / rare variants in metabolic cardiomyopathy genes as described above).

Secondly, the analysis in Chapter 2 identified a group of genes with rare variants found in HCM patients that were demonstrated to have functional effects. To test if such variants may act to modify disease expression, cases with protein-altering variants (ExAC frequency < 0.1%) in the following genes were identified – *ANKRD1*, *CAV3*, *FHL2*, *FXN*, *JPH2*, *KLF10*, *LDB3*, *MYLK2*, *MYOM1*, *MYPN*, *NEXN*, *TCAP*. A higher ExAC threshold was used as these variants are not being tested as putative disease-causing Mendelian variants but rather as potential modifiers of disease and therefore can be expected to be tolerated at a higher frequency in the population. Four genotype sub-groups were defined to test this hypothesis with two comparisons – genotype-positive samples (as described above) with and without additional variants in these genes and genotype-negative samples (as described above) with and without putative modifier variants.

Finally, the analysis in Chapter 3 identified a group of variants that would be classified as VUS by current ACMG guidelines but were upgraded to Likely Pathogenic due to an EF ≥ 0.95 for the particular variant class. Cases with these variants were compared to those with variants still classified as VUS, as well as the genotype-positive cases (using current ACMG guidelines only) and the genotype-negative samples described above. The aim of this analysis was to determine if cases with the high EF variants were phenotypically more similar to genotype-positive cases, providing additional evidence of their likely pathogenicity.

4.3 Results

4.3.1 Genotyping results and metrics

The number of cases in each of the defined genotype groups are shown in Figure 4.1 and Figure 4.2. 154 cases (23.5%) were classified as genotype-positive (i.e. had pathogenic or likely pathogenic variants with the adapted ACMG guidelines). There were 52 patients with rare VUS in HCM genes, 75 patients with non-rare variants in HCM genes or rare variants in uninterpretable HCM genes and 17 with rare variants in genes associated with metabolic cardiomyopathies – these three groups were not included in main analyses below. The genotype-negative sub-group comprised 358 HCM probands.

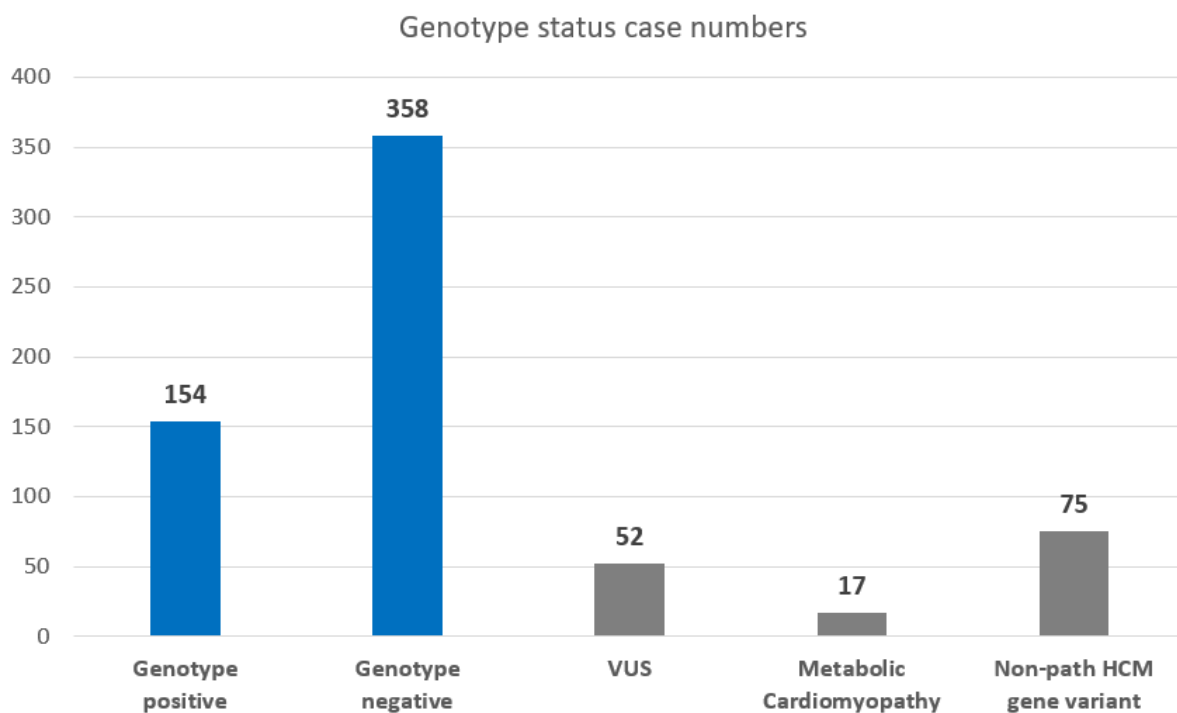


Figure 4.1: Number of HCM cases in each of the defined genotype groups. See text for details on the composition of these groups.

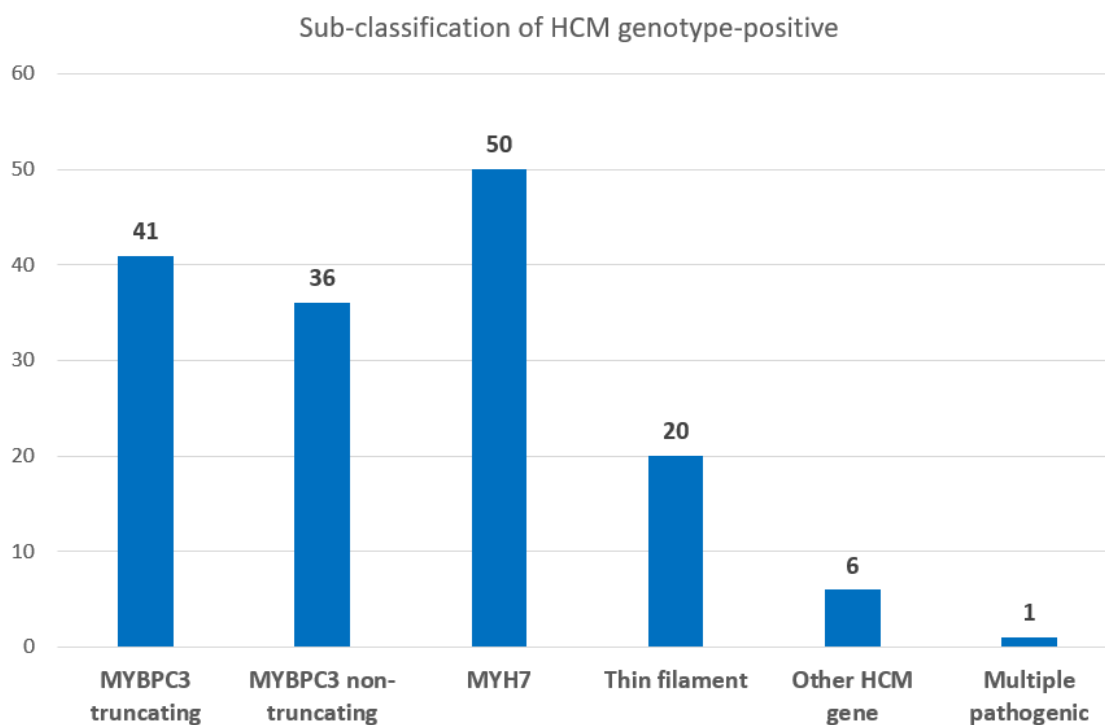


Figure 4.2: Number of HCM cases in each of the genotype-positive subgroups. See Methods for details on the composition of these groups.

4.3.2 Genotype-positive vs genotype-negative

Table 4.3 and Figure 4.3 summarise the results of the genotype-positive versus genotype-negative comparisons. In agreement with published studies, genotype-positive cases were significantly younger (48.1 ± 15.7 years vs 58.9 ± 14.7 years, $p < 0.001$) and had a greater FH of both HCM (50.8% vs 15.0% , $p < 0.001$) and SCD (43.1% vs 20.3% , $p < 0.001$) than genotype-negative patients. They also had both a greater maximum LVWT (19.85 ± 5.34 mm vs 18.51 ± 4.19 mm, $p = 0.006$) and a larger proportion of cases with extreme LVWT (≥ 30 mm) (5.9% vs 1.1% , $p = 0.004$). Although maximum LVWT and LVMI levels were correlated across the whole cohort ($r = 0.56$, $p < 0.001$, Figure 4.4), LVMI was significantly lower in genotype-positive patients (88.6 ± 31.0 vs 101.2 ± 31.8 , $p < 0.001$).

No difference was observed in the proportion of patients with LGE (92.2% vs 90.7% , $p = 0.701$) but the quantity of LGE as a percentage of LV mass was greater in genotype-positive cases ($10.8 \pm 10.6\%$ vs $8.0 \pm 8.5\%$, $p = 0.011$). More genotype-positive cases had extensive LGE, defined

as $\geq 15\%$ of LVM which has been suggested as a risk factor for SCD¹⁹², although this did not achieve statistical significance (25.9% vs 17.9%, $p=0.098$). There was no significant difference in the proportion of cases with LVOTO (32.4% vs 36.3%, $p=0.471$).

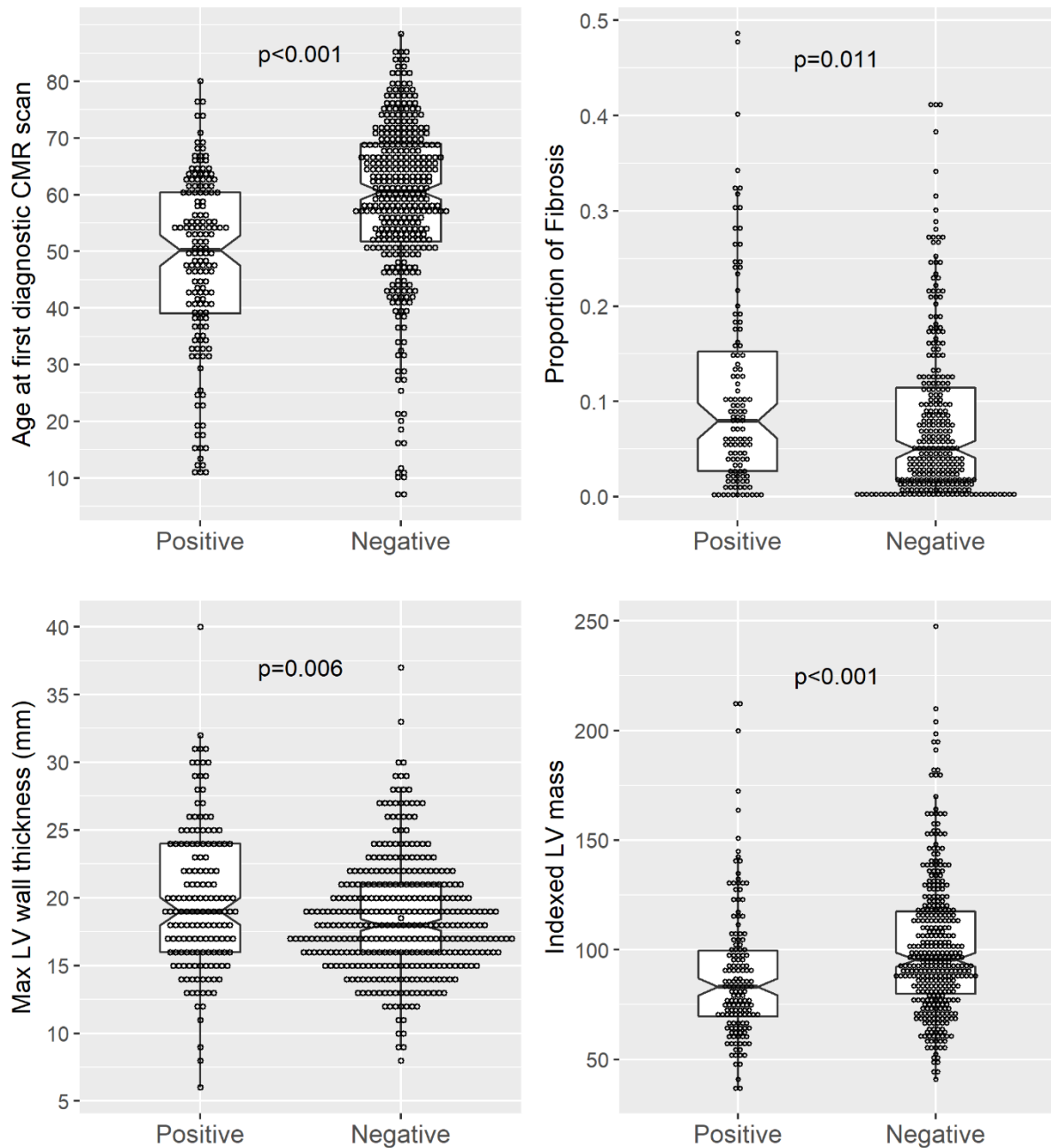


Figure 4.3: Box plots comparing genotype-positive and genotype-negative cases for age, proportion of LV LGE, maximum LVWT and indexed LV mass.

Category	Total	Genotype positive	Genotype negative	p value
Gender (% male)	372 (72.7%)	103 (66.9%)	269 (75.1%)	0.066
FH HCM	107 (26.5%)	66 (50.8%)	41 (15%)	<0.001
FH SCD	99 (27.1%)	47 (43.1%)	52 (20.3%)	<0.001
Age (years)	55.6 ± 15.8	48.1 ± 15.7	58.9 ± 14.7	<0.001
Max LVWT (mm)	18.91 ± 4.6	19.85 ± 5.34	18.51 ± 4.19	0.006
Extreme LVWT (>30mm)	13 (2.6%)	9 (5.9%)	4 (1.1%)	0.004
LVM (g/m ²)	97.4 ± 32	88.6 ± 31.0	101.2 ± 31.8	<0.001
LGE presence	360 (91.1%)	107 (92.2%)	253 (90.7%)	0.701
LGE proportion	0.088 ± 0.093	0.108 ± 0.106	0.08 ± 0.085	0.011
Extensive LGE (>15%)	80 (20.3%)	30 (25.9%)	50 (17.9%)	0.098
LVOTO	172 (35.1%)	48 (32.4%)	124 (36.3%)	0.471

Table 4.3: Summary of demographic and clinical phenotype data for genotype-positive and genotype-negative HCM cases.

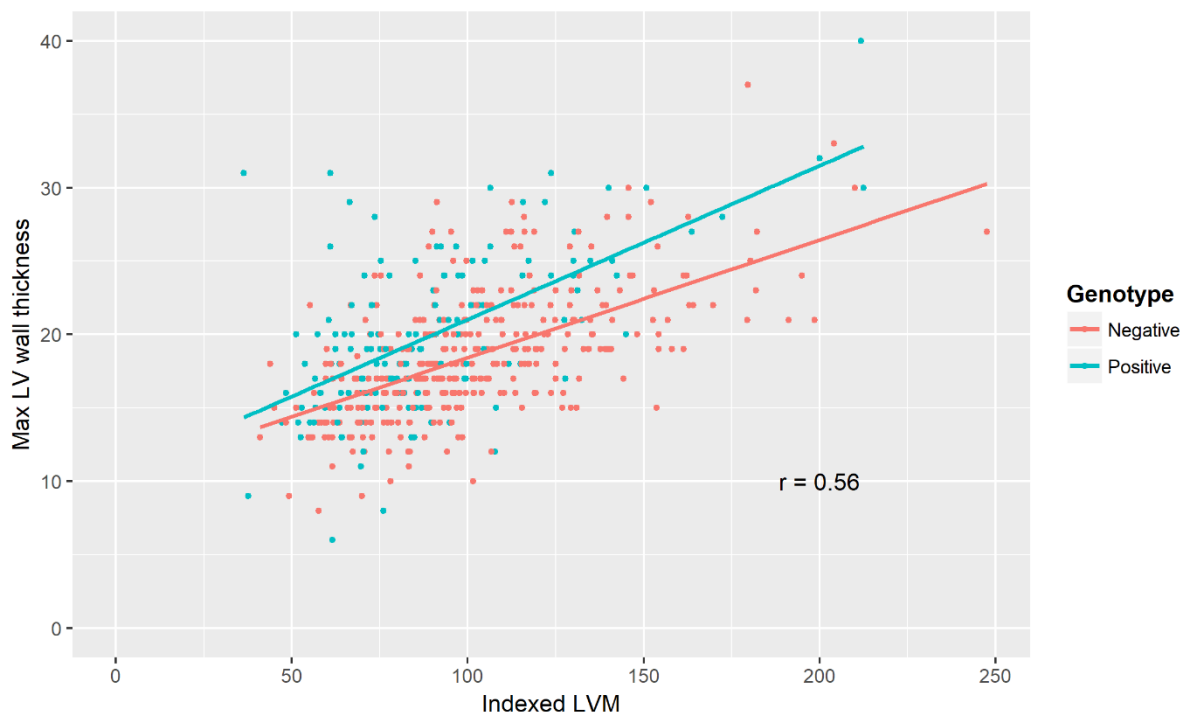


Figure 4.4: Correlation between maximum LV wall thickness and indexed LVM for genotype-positive and genotype-negative cases.

4.3.3 Distinguishing between familial and non-familial HCM

To distinguish between genotype-negative cases that may still potentially have unknown causative genetic variation (e.g. in a novel gene, non-coding variant or CNV as discussed in Chapter 2) and those that are more likely to be non-Mendelian, I compared genotype-negative cases with a FH of HCM or SCD to those without a FH of HCM. As shown in Table 4.4, there are no significant differences in the major demographic or phenotypic characteristics between the two genotype-negative sub-groups. However, there is a trend towards a more genotype-positive phenotype for the familial genotype-negative sub-group for age (lower), maximum LVWT (thicker) and LVMI (reduced mass). This suggests that at least some of these cases may have as yet undiscovered Mendelian variants, either in novel genes or alternate variant classes, which makes them more phenotypically similar to cases with pathogenic sarcomeric variants. However, it should be noted that self-declared FH may not accurately define all cases of familial HCM and at least some of those cases with FH may have a non-Mendelian genetic basis for the disease.

Category	Positive (154)	Negative Familial (83)	Negative NonFamilial (201)	P values		
				Pos / Neg-Fam	Pos / Neg-NonFam	Neg-Fam / Neg-NonFam
Gender (% male)	103 (66.9%)	62 (74.7%)	151 (75.1%)	0.357	0.292	1
Age (years)	48.1 ± 15.7	55.8 ± 14.9	59 ± 15.1	<0.001	<0.001	0.074
Max LVWT (mm)	19.8 ± 5.3	18.9 ± 4.7	18.5 ± 4.2	0.275	0.024	0.853
LVMI (g/m ²)	88.6 ± 31	97.9 ± 34.7	102.6 ± 32.5	0.038	<0.001	0.156
LGE proportion	0.108 ± 0.106	0.067 ± 0.087	0.082 ± 0.085	0.003	0.081	0.06

Table 4.4: Summary of demographic and clinical phenotype data for genotype-positive cases, genotype-negative cases with a FH of HCM or SCD and genotype-negative cases without a FH of HCM.

4.3.4 Genotype-positive subgroups

Table 4.5 and Figure 4.5 summarise the results of the genotype-positive subgroup comparisons. While a smaller (66.9% vs 75.1%) though non-significant ($p=0.066$) proportion of genotype-positive cases were male, a breakdown of the genotype-positive cohort by gene and variant class revealed that 51% of *MYH7* variant carriers were male compared to 70-80% for all other groups (including genotype-negative) ($p=0.045$). No significant differences in age or FH of HCM were observed between the subgroups, although a lower proportion of cases with *MYBPC3* non-truncating variants had a family history of SCD ($p=0.021$).

Category	Samples	Gender (% male)	FH HCM	FH SCD	Age (years)	Max LVWT (mm)	LVM (g/m ²)	LGE proportion
Total	512	372 (72.7%)	107 (26.5%)	99 (27.1%)	55.6±15.8	18.91±4.6	97.4±32	0.088±0.093
Geno. pos.	154	103 (66.9%)	66 (50.8%)	47 (43.1%)	48.1±15.7	19.85±5.34	88.6±31	0.108±0.106
Geno. neg.	358	269 (75.1%)	41 (15%)	52 (20.3%)	58.9±14.7	18.51±4.19	101.2±31.8	0.08±0.085
p value		0.066	<0.001	<0.001	<0.001	0.006	<0.001	0.011
Thick filament	128	83 (64.8%)	56 (53.3%)	39 (43.8%)	47.5±15.7	20.1±5.4	88.2±30.5	0.1±0.1
Thin filament	20	14 (70%)	9 (47.4%)	5 (33.3%)	50.7±16.5	17.5±3.6	79.4±19.4	0.116±0.122
p value		0.802	0.804	0.576	0.311	0.013	0.234	0.983
MYBPC3	77	57 (74%)	31 (50%)	21 (39.6%)	47.9±13.6	19.9±5.4	88.6±26.5	0.098±0.097
MYH7	51	26 (51%)	25 (58.1%)	18 (50%)	46.8±18.5	20.4±5.5	87.6±35.9	0.12±0.116
p value		0.009	0.433	0.387	0.71	0.625	0.235	0.641
MYBPC3tv	41	32 (78%)	20 (58.8%)	17 (56.7%)	46.2±13.3	21±6.2	92.2±32.7	0.095±0.093
MYBPC3non-tv	36	25 (69.4%)	11 (39.3%)	4 (17.4%)	49.9±13.9	18.6±4	84.6±16.8	0.101±0.103
MYH7	51	26 (51%)	25 (58.1%)	18 (50%)	46.8±18.5	20.4±5.5	87.6±35.9	0.12±0.116
Thin filament	20	14 (70%)	9 (47.4%)	5 (33.3%)	50.7±16.5	17.5±3.6	79.4±19.4	0.116±0.122
p value		0.045	0.354	0.021	0.517	0.048	0.362	0.953

Table 4.5: Summary of the demographic and clinical phenotype data for genotype-positive subgroups of HCM patients.

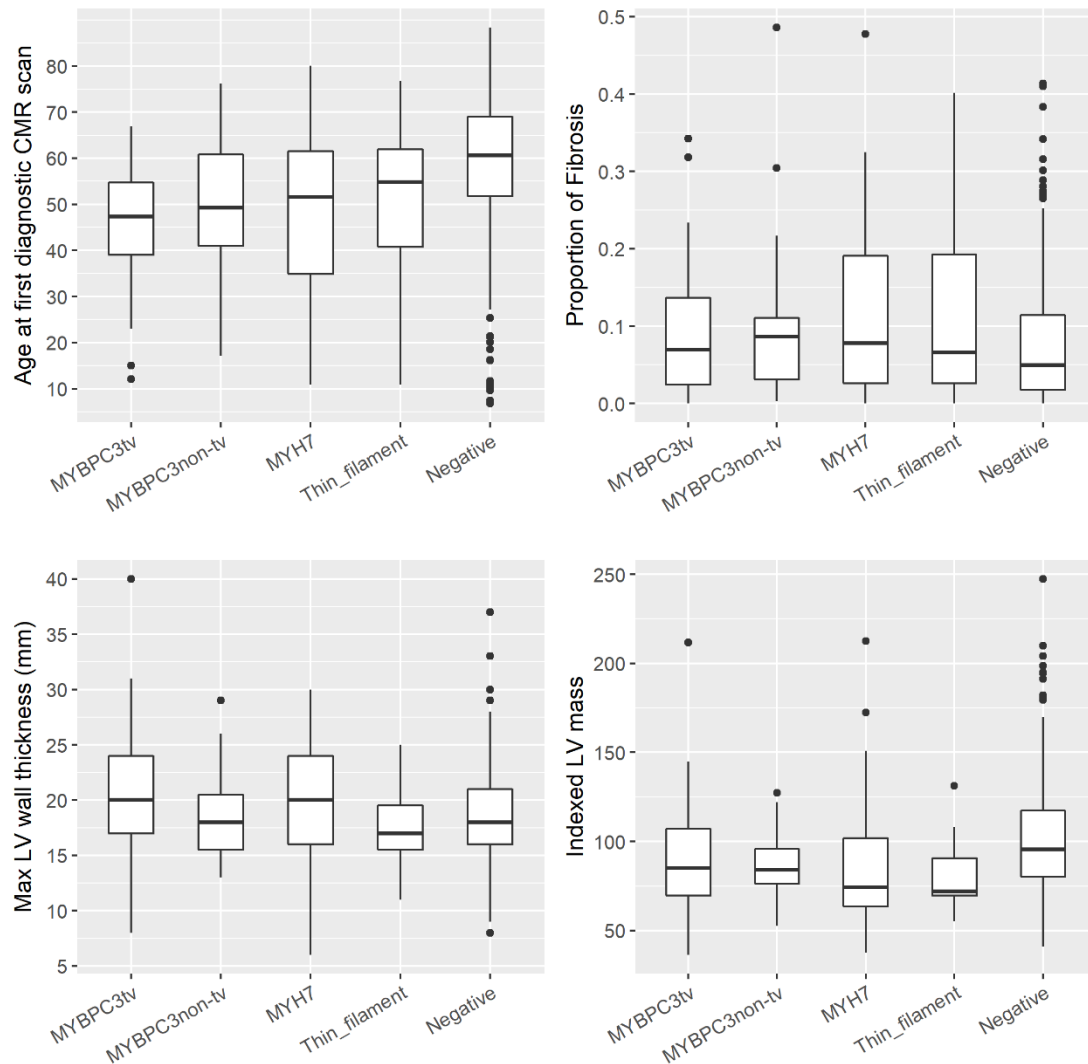


Figure 4.5: Box plots comparing genotype-positive subgroups (MYBPC3 truncating and non-truncating variants, MYH7 variants and thin filament gene variants) and genotype-negative cases for age, proportion of LV LGE, maximum LV wall thickness and indexed LV mass.

Maximum LVWT was significantly greater among carriers of thick filament gene variants than thin filament ($20.1 \pm 5.4\text{mm}$ vs $17.5 \pm 3.6\text{mm}$, $p=0.013$). Although not statistically significant, LVWT trended lower in MYBPC3 non-truncating carriers ($18.6 \pm 4.0\text{mm}$) compared to MYBPC3 truncating variant carriers ($21.0 \pm 6.2\text{mm}$) ($p=0.205$, Tukey post-hoc test of ANOVA). Although there was a trend for greater LVMI in thick filament samples compared to thin filament samples (88.2 ± 30.5 vs 79.4 ± 19.4), this was not statistically significant ($p=0.234$). No differences in the proportion of LGE was observed for any of the subgroups.

4.3.5 Cases with multiple pathogenic variants

With the stringent, ACMG guideline-based approach to variant interpretation adopted for this study, only one patient had more than one actionable pathogenic variant. This patient had two *MYH7* variants that unusually affected the same residue, p.S748F and p.S748T, both of which were novel (as well as a non-rare *MYBPC3* variant - p.P608S, ExAC frequency 0.00014). They presented with a relatively severe phenotype – inclusion at an early age (34 years) with a maximum LVWT of 24mm. Although the frequency of patients with multiple pathogenic variants here is much lower than some earlier reports, it is consistent with a recent study which observed that the prevalence of double mutations in HCM patients has been overestimated, with many of these reported variants now not being classified as pathogenic according to current guidelines⁴⁹. While we cannot make any conclusions about the effect of multiple pathogenic variants from this study, in conjunction with recent reports it seems likely that the presence of multiple variants can only be very rarely used to predict deleterious outcomes in HCM patients, with only *MYBPC3* “knockouts” (i.e. homozygous or compound biallelic/*in trans* truncating variants) definitively predicting a more severe phenotype⁴⁹.

4.3.6 Using genotype-phenotype data to validate conclusions about genetics of HCM

4.3.6.1 Phenotype characteristics of cases with variants in non-sarcomeric genes

In Chapter 2, I concluded that for most of the non-sarcomeric genes that have been implicated in HCM, there is insufficient human genetic evidence to support a role for variants in these genes to be causative in this condition. To further validate these findings, the phenotypic characteristics of cases with rare variants in these genes were compared to genotype-positive and negative sub-groups. As expected, this non-valid gene cohort were phenotypically similar to the genotype-negative cohort, with no statistically significant differences in the demographic and CMRI metrics that have been shown to differentiate genotype-positive and genotype-negative cases (Table 4.6 and Figure 4.6). This provides further evidence that these variants are not causative for HCM in these patients and instead are simply rare, benign variants occurring at a frequency we would expect to see in any population cohort or group. HCM patients with variants in these genes therefore are currently considered to be genotype-negative.

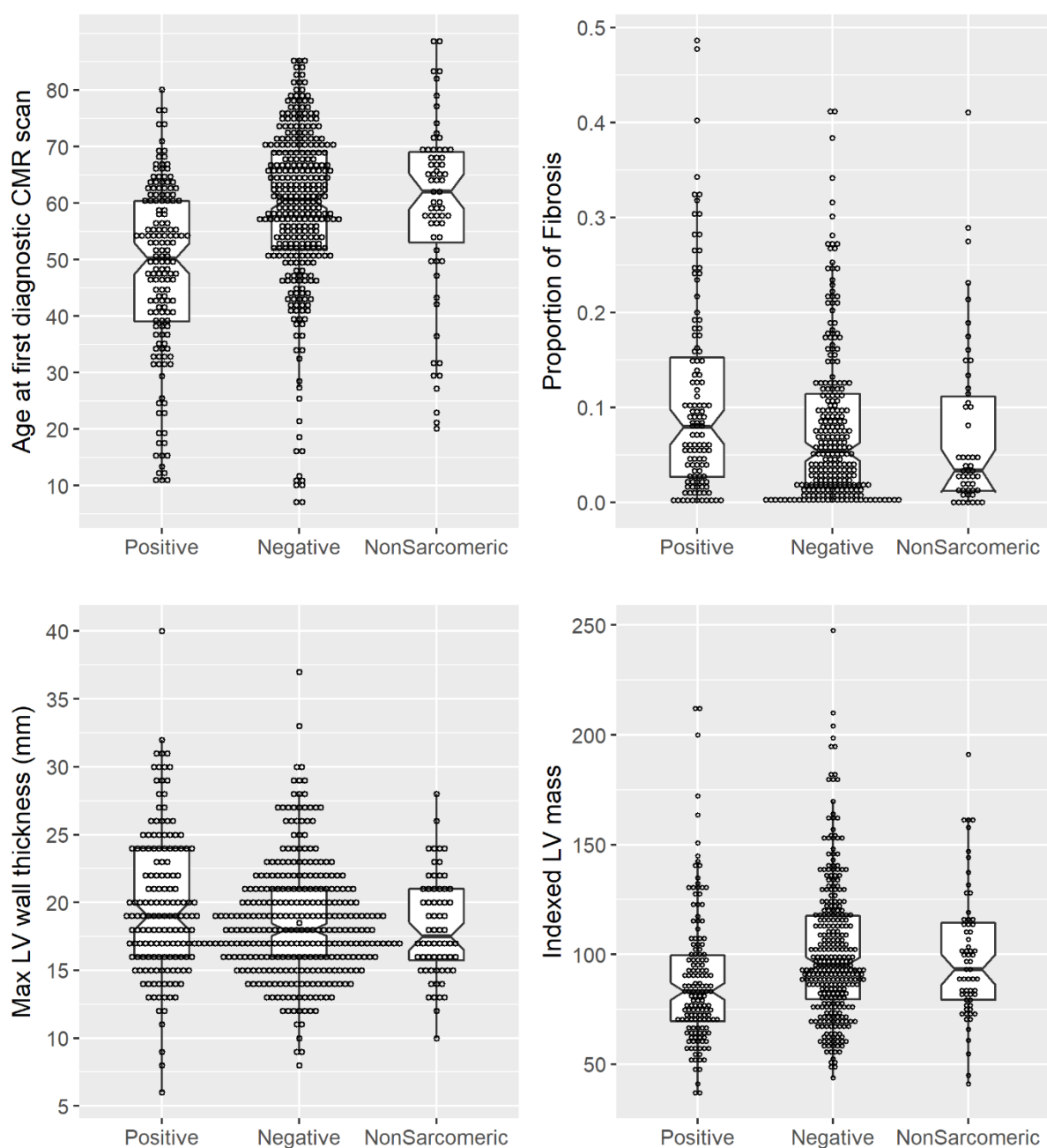


Figure 4.6: Box plots comparing genotype-positive cases, genotype-negative cases and cases with variants in putative, non-sarcomeric genes that have been shown to have no evidence for involvement in HCM in Chapter 2 or for which variants are unlikely to be pathogenic. Comparisons are shown for age, proportion of LV LGE, maximum LV wall thickness and indexed LV mass.

Category	Positive (154)	Negative (313)	Other (64)	P values		
				Pos/Neg	Pos/Other	Neg/Other
Gender (% male)	103 (66.9%)	236 (75.4%)	47 (73.4%)	0.182	0.634	0.752
FH HCM	66 (42.9%)	36 (11.5%)	7 (10.9%)	<0.001	<0.001	1
FH SCD	47 (30.5%)	46 (14.7%)	7 (10.9%)	<0.001	0.002	0.542
Age	48.1 ± 15.7	59 ± 14.5	59.1 ± 16.1	<0.001	<0.001	0.775
Max LVWT (mm)	19.8 ± 5.3	18.6 ± 4.3	18.2 ± 3.7	0.015	0.036	0.781
LVM (g/m ²)	88.6 ± 31	101.2 ± 31.8	99.1 ± 29.9	0	0.008	0.726
LGE	0.108 ± 0.106	0.08 ± 0.083	0.074 ± 0.09	0.026	0.036	0.317

Table 4.6: Genotype-phenotype comparisons comparing carriers of rare variants in non-sarcomeric genes implicated in HCM (“Other”) with genotype-positive and genotype-negative cases. Patients with variants in these other genes are more phenotypically similar to genotype-negative cases, indicating these variants are unlikely to be disease-causing.

4.3.6.2 Effect of putative modifier genes

A subset of the non-sarcomeric genes identified in Chapter 2 that lacked convincing human genetic evidence for association with HCM did have published reports demonstrating the functional effects of variants (detected in HCM patients) using *in vivo* and *in vitro* assays. While variants in these genes are unlikely to be disease-causing Mendelian variants in HCM patients, it is possible that they may act to modify expression of disease given their observed effects. However, as shown in Table 4.7, no significant differences were observed based on the presence or absence of these putative functional gene variants in either the genotype-positive or genotype-negative sub-groups, indicating that variants in these genes, or at least this group of genes as a whole, do not act as phenotype modifiers.

Category	Genotype positive			
	Total	With func. gene variant	Without func. gene variant	p value
Gender (% male)	102 (66.7%)	12 (66.7%)	90 (66.7%)	1
FH HCM	66 (51.2%)	11 (64.7%)	55 (49.1%)	0.3
FH SCD	46 (42.6%)	7 (50%)	39 (41.5%)	0.574
Age	48.1 ± 15.7	46.5 ± 16	48.2 ± 15.7	0.484
Max LVWT (mm)	19.85 ± 5.34	20.28 ± 4.48	19.76 ± 5.46	0.658
LVM (g/m ²)	88.6 ± 31	83.2 ± 22.7	89.5 ± 32.1	0.534
LGE proportion	0.108 ± 0.106	0.118 ± 0.099	0.106 ± 0.107	0.429
Category	Genotype negative			
	Total	With func. gene variant	Without func. gene variant	p value
Gender (% male)	269 (75.1%)	39 (69.6%)	230 (76.2%)	0.314
FH HCM	41 (15%)	6 (15%)	35 (15%)	1
FH SCD	52 (20.3%)	6 (14.6%)	46 (21.4%)	0.401
Age	58.9 ± 14.7	58.9 ± 14.5	58.9 ± 14.7	0.739
Max LVWT (mm)	18.51 ± 4.19	18.25 ± 3.88	18.55 ± 4.25	0.608
LVM (g/m ²)	101.2 ± 31.8	103.8 ± 30.3	100.7 ± 32.1	0.341
LGE proportion	0.08 ± 0.085	0.071 ± 0.087	0.082 ± 0.085	0.33

Table 4.7: Genotype-phenotype comparisons comparing patients with and without variants in genes that have been implicated in HCM with demonstrated effects of variants in functional assays. These comparisons test whether these variants may have a modifying effect on genotype-positive or genotype-negative cases.

4.3.6.3 Phenotypes of cases with VUS of differing EFs

Finally, the distinctive characteristics of genotype-positive and genotype-negative HCM cases offers an opportunity to assess variant interpretation in sarcomeric genes, in particular to observe whether patients with VUS are more phenotypically similar to genotype-positive cases (suggesting that a large proportion of the VUS may actually be pathogenic) or genotype-negative (indicating that most are likely to be rare benign variants). For this study, I compared cases with VUS that were upgraded to likely pathogenic by the new EF-based rules described in Chapter 3 with cases harbouring variants still classified as VUS by these adapted ACMG

rules, as well as to genotype-positive (based on current ACMG guidelines) and genotype-negative cases.

A summary of the demographic and phenotypic metrics for each of these groups is shown in Table 4.8, along with results from statistical comparisons between the four groups. The evidence for pathogenicity of the two VUS sub-groups is mixed, perhaps reflecting the limited number of cases with these variant classes (29 cases with upgraded VUS and 52 cases with standard VUS). The sub-group with upgraded variants had a low proportion of males (reflecting the large number of *MYH7* variant carriers in this group) and a similar proportion of patients with a FH of HCM as the genotype-positive cases (37.9% and 44.0% respectively). The mean age of both VUS groups was intermediate between genotype-positive and genotype-negative cases, though closer on average to the genotype-positive cases. In contrast, the mean maximum LVWT for the VUS groups is more similar to genotype-negative cases – as this metric is not indexed to age, sex or body mass, it may be susceptible to inconsistent demographic compositions between the groups, especially given the limited cohort sizes of the VUS groups. However, distinctive values for LVMi data are observed between the two VUS groups, with the upgraded cases displaying similar means to the genotype-positive group (85.2 ± 23.4 and 89.2 ± 32.5) and the VUS group similar to genotype-negative cases (101.9 ± 24.2 and 101.2 ± 31.8).

Category	Positive (125)	Negative (358)	Positive-EF (29)	VUS (52)	P values					
					Pos / Neg	Pos / VUS	Pos / PosEF	Neg / VUS	Neg / PosEF	VUS / PosEF
Gender (% male)	89 (71.2%)	269 (75.1%)	14 (48.3%)	40 (76.9%)	0.558	0.558	0.055	0.864	0.022	0.041
FH HCM	55 (44%)	41 (11.5%)	11 (37.9%)	15 (28.8%)	<0.001	0.024	0.82	0.024	0.001	0.231
FH SCD	40 (32%)	52 (14.5%)	7 (24.1%)	12 (23.1%)	<0.001	0.26	0.556	0.471	0.307	0.571
Age	47.3±15.2	58.9±14.7	51.3±17.7	52±15.2	<0.001	0.103	0.12	0.004	0.047	0.869
Max LVWT (mm)	20.2±5.4	18.5±4.2	18.4±4.9	19±4.1	0.002	0.39	0.235	0.899	0.999	0.946
LVM (g/m ²)	89.2±32.5	101.2±31.8	85.6±23.4	101.9±24.2	<0.001	0.001	0.986	0.479	0.027	0.024
LGE	0.112 ±0.105	0.08 ±0.085	0.093 ±0.112	0.094 ±0.074	0.024	0.886	0.341	0.276	0.907	0.487

Table 4.8: Summary of the demographic and phenotypic characteristics of two sub-groups of patients with VUS – Positive-EF (PosEF) indicates those cases with a variant upgraded to likely pathogenic based on the adapted ACMG guidelines described in Chapter 3 ($EF \geq 0.95$), while VUS indicates those cases with a variant of a pathogenic class for HCM but which are still classified as VUS by the adapted ACMG guidelines. These are compared to genotype-positive cases (as defined by the current ACMG guidelines, not including the new EF-based rules) and genotype-negative cases.

4.4 Discussion

Initial research into the correlation between genotype and phenotype in HCM patients produced inconsistent and even contradictory findings. This was likely due to both the intrinsic phenotypic heterogeneity associated with HCM and to problems relating to study design and the accuracy in assessing both phenotype and genotype. Issues around cohort size and the reliance on echocardiography for phenotypic measurements have long been recognised and are beginning to be addressed with the publication of data from larger cohorts and the increasing use of CMRI in research studies. The accurate identification of truly pathogenic variants in these disease cohorts remains a critical and challenging issue, reflecting both the inclusion of genes that may not be actually involved in HCM (Chapter 2) and the under or over calling of pathogenic variants in valid HCM genes (Chapter 3). This study aims to overcome these limitations by analysing the largest CMRI-assessed HCM cohort yet described. Building on our improved understanding of the genetics of HCM (Chapters 2 and 3), likely causative variants can be more confidently identified, thereby yielding more accurate genotype subsets. The results have clarified the effect of genotype on demographics and intermediate clinical phenotypes, and evaluated the effect of genotype on the LVMI and LGE metrics obtained by CMRI.

4.4.1 Distinctive clinical profile of genotype-positive and genotype-negative patients

There is a growing body of evidence that genotype-positive and genotype-negative HCM patients follow a distinct clinical course¹⁹³. Genotype-positive cases present approximately a decade earlier and are associated with more adverse outcomes^{86,183,184} (Table 4.1). While confirming previous reports that maximum LVWT is greater in genotype-positive cases (on average by over 1mm in our study), I have also demonstrated that overall hypertrophy, as measured by LVMI, is actually significantly greater in genotype-negative samples, despite a broad correlation between maximum LVWT and LVMI across all samples. These findings imply that a distinctive pattern of hypertrophy will present in patients depending on the presence or absence of sarcomeric variants – more severe and localised hypertrophy in genotype-positive samples and less severe but more widespread hypertrophy in genotype-negative samples.

Additionally, this study observed that the amount of LGE as a proportion of LV mass was significantly greater in the genotype-positive subset, implying the presence of more widespread fibrosis in these cases. As intramyocardial fibrosis is thought to be a predictor of SCD in HCM patients¹⁸⁶, this finding is consistent with the observed higher incidence^{86,184} and greater FH (Table 4.1) of SCD in genotype-positive patients. Overall these findings, in conjunction with published outcome data, indicate that the presence of sarcomeric variants in HCM patients is associated with a more severe form of the disease.

These findings call into question how HCM is defined and diagnosed and its nature as a Mendelian genetic disease. The diagnosis of the disease is based on fairly arbitrary diagnostic thresholds of continuous biological parameters – LVWT, LVOTO, myocyte disarray etc. Distinguishing between primary LV hypertrophy and secondary physiological and pathological adaptation in cases without causative sarcomeric variants can be difficult, even though by definition a diagnosis of HCM requires the observed LV hypertrophy to not be fully explained by abnormal loading conditions. It is notable that in two large published cohorts (Table 4.1), hypertension was observed in substantial proportions of genotype-negative cases, raising doubts as to how effectively we can differentiate between primary and secondary hypertrophy. The divergent characteristics that distinguish genotype-negative from genotype-positive HCM observed in this and other studies suggest this may be viewed as a distinctive pathological condition, and one that lies on a spectrum of physiological to pathological adaptation presumably driven by a range of environmental and non-Mendelian genetic risk factors.

4.4.2 *Impact on genetic testing in HCM*

The main role of genetic testing in HCM has been to facilitate the identification of at-risk relatives through cascade screening for a pathogenic variant (assuming one is identified in the proband). For first-degree relatives of HCM patients without a detected causative variant, standard (and ongoing) clinical screening is recommended³. In these genotype-negative cases, it has been uncertain until now whether Mendelian inheritance of HCM still occurs in these families (caused by undetected or uninterpretable genetic variation) or whether the

proband has a non-familial form of the disease (with a combination of genetic and non-genetic aetiology that would make co-occurrence in families sporadic at best). However, there is accumulating evidence that the majority of genotype-negative cases may in fact have a non-Mendelian and non-familial form of HCM.

Since the original linkage studies of the 1990s that identified variants in the eight core sarcomeric genes as causative for HCM, only a handful of additional genes have been confidently implicated in HCM, and these probably account for only a small percentage of cases (Chapter 2). Given that most large family pedigrees have probably been genetically characterised by now (at least in Europe and North America), it seems logical to assume that rather than being caused by as yet undiscovered genes, most genotype-negative cases constitute a non-Mendelian form of the disease. The distinctive clinical characteristics of genotype-positive and genotype-negative cases described here further support the theory that these two subtypes have different disease aetiologies. Finally, the significantly lower proportion of genotype-negative cases with a FH of HCM (between 15% and 25%) indicates that many, if not most, of these cases will not be inherited in a Mendelian manner (although the minority with FH may have as yet unidentified causative variants).

Building on these observations, Ingles et al. have described a non-familial sub-group of HCM, i.e. those without both a causative sarcomeric variant and FH of the disease¹⁹⁴. Compared to genotype-positive cases, this non-familial sub-group was characterised by later onset and less adverse clinical outcomes. While this is similar to what is observed between genotype-positive and genotype-negative sub-groups, the likelihood of family members of the non-familial sub-group being at risk of developing HCM is believed to be low, given the lack of both FH of the condition and any identified Mendelian causative variant in the proband. The authors suggest a revised clinical screening pathway in which adult relatives of patients with non-familial HCM would not require ongoing clinical screening given their reduced risk of developing disease. In this study, I have directly compared genotype-negative cases with FH to those without. While there are no statistically significant differences between these sub-groups, there was a trend towards a more “genotype-positive” phenotype in those cases with FH, further supporting the idea that non-familial cases will follow a distinct clinical course.

As further evidence of this hypothesis, a recent study sought to assess the risk of developing disease for relatives of genotype-negative HCM patients with no prior family history and found that only 2 of 64 (3%) of those relatives clinically screened were diagnosed with HCM (both over 40 and with no HCM-related adverse events)¹⁹⁵. In contrast, HCM was diagnosed in 17% of relatives of patients with sarcomeric variants or prior family history, with 65% of these relatives experiencing HCM-related adverse events. While further long-term studies are needed to definitively ascertain the risk to such relatives, these findings have implications on how the results of genetic testing in HCM are interpreted.

Until now, the failure to detect a pathogenic variant in a HCM proband was indecisive in determining whether family members were at risk of developing the disease. First-degree relatives of such patients are still recommended to undergo regular clinical screening to detect signs of developing left ventricular hypertrophy, which leads to both a psychological burden for family members and a significant financial burden for healthcare providers. However these insights now suggest that far from being a “negative” result as has been previously believed, a genetic test without an identified pathogenic variant may in fact be an informative and indeed encouraging outcome. For patients without prior FH of HCM, a negative test suggests the disease is unlikely to be familial and relatives may be spared from ongoing clinical surveillance. It also implies that the proband may experience a more benign prognosis than would be the case if a variant had been detected.

4.4.3 Phenotype data can help to validate hypotheses about the genetics of HCM

The distinctive demographic and phenotypic characteristics of genotype-positive and genotype-negative HCM cases allows us to use this data to assess the likely pathogenicity of particular variant classes. If patients with these variants are broadly more phenotypically similar to genotype-positive cases, it implies that these variants are more likely to represent pathogenic, Mendelian variants, while if they are more similar to genotype-negative cases, then these variants may just be rare benign variants without any causative role. I have used this approach to test some of the conclusions from the analysis of HCM genetics described in Chapters 2 and 3. Cases with variants in non-sarcomeric genes for which there was very

limited evidence for involvement in HCM (Chapter 2) had a profile similar to genotype-negative cases (Table 4.6 and Figure 4.6), providing further supporting evidence that at least the majority of variants in these genes are unlikely to be pathogenic. Similarly, there was no evidence at the cohort level that variants in genes in which the functional effects of variants have been reported have a modifier effect on disease expression (Table 4.7), though further studies that are specific to particular genes, assess larger cohorts or investigate family pedigrees may be needed to more fully explore this hypothesis. Finally, although the two subgroups of cases with VUS (those upgraded by the adapted ACMG guidelines and those still VUS) weren't consistently similar to either the genotype-positive or genotype-negative groups (perhaps reflecting the limited number of cases in these groups), the age and LVMI data in particular did indicate that they are likely to include a substantial number of actual pathogenic variants. However, data from much larger cohorts will be needed to more definitively utilise phenotype data to support the pathogenicity of such variant classes.

4.4.4 Limited phenotypic differences observed depending on class of pathogenic variant

While early HCM genotype-phenotype studies suggested associations between certain variants (and variant classes) and disease expression and severity, most of these have failed to be replicated in subsequent larger studies. Therefore initial hopes that a patient's genotype could be prognostic for disease severity and characteristics have been unfulfilled. Currently, knowledge of the specific causative variant does not impact on the clinical management of disease for HCM patients, with the notable and important exception of the identification of variants in metabolic cardiomyopathy genes that may mimic the HCM phenotype and mandate specific therapy. Nevertheless, broad distinctions between carriers of different classes are now being recognised which yield insights into how genetic variation translates into clinical phenotype and outcomes.

Numerous studies have attempted to assess the effect on clinical phenotype of variants in the two major HCM genes (*MYBPC3* and *MYH7*), with some early reports associating *MYH7* variants with a more severe phenotype. Consistent with more recent research in larger cohorts¹⁸⁸, this study observed no significant difference in CMRI phenotypes or demographics

between *MYBPC3* and *MYH7* carriers, with the exception of a lower proportion of males amongst *MYH7* cases. The gender distribution of HCM cases is an intriguing and still largely unexplained phenomenon, given that it is an autosomal dominant disease and yet up to 75% of HCM patients are male. It has been hypothesised that gender may act as an environmental factor that modifies expression of the disease or that there may be a bias towards males when diagnosing HCM. However, the data presented here imply that this effect is not observed in carriers of *MYH7* variants (51% male), indicating more complete penetrance of disease-causing variants in this gene.

It is notable that some differences were observed between patients with truncating and non-truncating variants in *MYBPC3* – a lower proportion with FH of SCD and less severe (though not statistically significant) LV hypertrophy in cases with non-truncating variants. The cohort size for these sub-groups was limited (41 and 36 for truncating and non-truncating respectively), but this data suggests that future genotype-phenotype studies in HCM would benefit from distinguishing between *MYBPC3* variant classes. This study also confirms a previous report¹⁸⁹ that a distinctive LV morphology exists between patients with variants in thin filament genes as compared to thick filament genes – hypertrophy is less severe in thin filament patients (defined by both maximum LVMT and overall mass). The effect this has on long term outcome for these two sub-groups is unclear – the one study that assessed outcomes between these groups showed that while thin filament variants were more likely to lead to advanced LV dysfunction, this did not correspond to increased rates of mortality or SCD¹⁸⁹.

Long-term studies with very large cohorts of HCM patients will be needed to definitively assess the effect of variant class on disease progression and outcome. Given the need to sub-classify the patient cohort according to gene and variant class, and the relatively small proportion of cases with pathogenic variants in these sub-groups, consortia such as the SHaRe Registry (<http://theshareregistry.org>) that amalgamate data from several leading cardiovascular centres will be best placed to face this challenge. However, although some differences are observed between carriers of different variant classes at the cohort level, there are still large overlaps between sub-groups when assessing individual cases (Figure 4.5).

Therefore it is unlikely that such findings will translate into informing or altering the clinical management of individual patients depending on their particular causative variant.

It is conceivable that the phenotypic variability observed between and within carriers of different variant classes may be due additional genetic variants, either rare variants or more common polymorphisms, that exert smaller effects which may modify disease expression. If such variants were identified, they could be used as a genetic risk score in conjunction with the primary Mendelian pathogenic variant to produce a more complete assessment of the patient's genetic susceptibility to HCM. This may lead to more consistent, specific and predictable correlations between genotype and disease expression and outcome, which could then inform treatment options and clinical management. However, identifying and validating these potential modifier variants will require very large and carefully designed and controlled studies and are therefore unlikely to be clinically relevant for the foreseeable future.

4.5 Conclusion

Early hopes that the knowledge of a disease-causing variant in HCM patients would inform clinical management have largely been confounded. While both the research presented here and other recent studies have highlighted differences in demographics, clinical phenotype and outcomes between patients with different classes of variants, genetic testing remains unlikely to influence treatment or clinical management for the majority of patients. Nevertheless, these studies have underlined intriguing differences between genotype-positive and genotype-negative HCM cases and support emerging ideas that these groups may have distinct underlying disease mechanisms. They also suggest that the implications of a negative HCM genetic test need to be revised from simply being uninformative to being one that indicates both a low likelihood of disease transmission in the patient's family and a more benign prognosis for the patient in question. If supported by future longitudinal studies, these findings may bring substantial benefits for patients, their families and the cost of healthcare provision.

5 Concluding Remarks

Advances in genetic sequencing technologies in recent years have offered unprecedented opportunities to increase our understanding of genetic conditions like HCM and to widely deploy genetic testing as part of the clinical management of patients. However, the relative ease of generating sequencing data has also had the unintended effect of obscuring our understanding of HCM genetics. Candidate gene studies, which can be performed with much less cost and effort compared to earlier linkage studies, have attempted to uncover the apparent missing heritability in HCM but, as most failed to adequately control for rare benign variation in the population, resulted in large numbers of likely false positive associations. The findings here have shown how an overemphasis on the pursuit of novelty in scientific research can lead to a poor allocation of resources and hinder our understanding of the genetic basis of disease, particularly as there were already strong indications that a large proportion of HCM cases lacking sarcomeric variants were likely to be non-Mendelian HCM.

Population databases like ExAC have also underlined the complexity associated with variant interpretation for valid disease genes and difficulties in differentiating pathogenic from rare benign variants. This study has revealed how current approaches to variant interpretation in clinical laboratories are overly conservative and demonstrated that disease and gene-specific approaches to variant analysis, rather than generic techniques, can significantly improve the accuracy and sensitivity of interpreting the results of genetic testing in HCM. It was notable that the increased yield of actionable variants produced by these efforts is likely greater than all previous efforts to identify causative non-sarcomeric variants for this disease.

The genotype-phenotype studies described here highlight both the benefits and limitations of integrated genomics and imaging analysis for HCM. The extensive phenotypic heterogeneity of HCM means that knowledge of a patient's genetic status will not influence clinical management in most cases, and is unlikely to until we get much greater insight into the range of genetic factors that may influence disease expression. However, the findings from this and other studies are changing our understanding of how to interpret the results of genetic testing, in particular for negative results in the absence of prior family history which now likely indicates a non-familial form of the disease, a finding (if confirmed by further studies) that could have major implications for family members and healthcare providers.

6 Bibliography

1. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *Lancet (London, England)* 2013;**381**:242–255.
2. Wilde AAM, Behr ER. Genetic testing for inherited cardiac disease. *Nat Rev Cardiol* 2013;**10**:571–583.
3. Authors/Task Force members, Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P, Hagege AA, Lafont A, Limongelli G, Mahrholdt H, McKenna WJ, Mogensen J, Nihoyannopoulos P, Nistri S, Pieper PG, Pieske B, Rapezzi C, Rutten FH, Tillmanns C, Watkins H. 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* 2014;**35**:2733–2779.
4. Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. 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* 1995;**92**:785–789.
5. Maron BJ, Mathenge R, Casey SA, Poliac LC, Longe TF. Clinical profile of hypertrophic cardiomyopathy identified de novo in rural communities. *J Am Coll Cardiol* 1999;**33**:1590–1595.
6. Lipshultz SE, Sleeper LA, Towbin JA, Lowe AM, Orav EJ, Cox GF, Lurie PR, McCoy KL, McDonald MA, Messere JE, Colan SD. The incidence of pediatric cardiomyopathy in two regions of the United States. *N Engl J Med* 2003;**348**:1647–1655.
7. Nugent AW, Daubeney PEF, Chondros P, Carlin JB, Colan SD, Cheung M, Davis AM, Chow CW, Weintraub RG, National Australian Childhood Cardiomyopathy Study. Clinical features and outcomes of childhood hypertrophic cardiomyopathy: results from a national population-based study. *Circulation* 2005;**112**:1332–1338.
8. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2015;**65**:1249–1254.

9. Baudhuin LM, Kotzer KE, Kluge ML, Maleszewski JJ. What Is the True Prevalence of Hypertrophic Cardiomyopathy? *J Am Coll Cardiol* 2015;**66**:1845–1846.
10. Sen-Chowdhry S, Jacoby D, Moon JC, McKenna WJ. Update on hypertrophic cardiomyopathy and a guide to the guidelines. *Nat Rev Cardiol* 2016;**13**:651–675.
11. Papadakis M, Carre F, Kervio G, Rawlins J, Panoulas VF, Chandra N, Basavarajaiah S, Carby L, Fonseca T, Sharma S. The prevalence, distribution, and clinical outcomes of electrocardiographic repolarization patterns in male athletes of African/Afro-Caribbean origin. *Eur Heart J* 2011;**32**:2304–2313.
12. Sheikh N, Papadakis M, Carre F, Kervio G, Panoulas VF, Ghani S, Zaidi A, Gati S, Rawlins J, Wilson MG, Sharma S. Cardiac adaptation to exercise in adolescent athletes of African ethnicity: an emergent elite athletic population. *Br J Sports Med* 2013;**47**:585–592.
13. Ranasinghe I, Ayoub C, Cheruvu C, Freedman SB, Yiannikas J. Isolated hypertrophy of the basal ventricular septum: characteristics of patients with and without outflow tract obstruction. *Int J Cardiol* 2014;**173**:487–493.
14. McKenna WJ, Stewart JT, Nihoyannopoulos P, McGinty F, Davies MJ. Hypertrophic cardiomyopathy without hypertrophy: two families with myocardial disarray in the absence of increased myocardial mass. *Br Heart J* 1990;**63**:287–290.
15. Varnava A, Baboonian C, Davison F, Cruz L de, Elliott PM, Davies MJ, McKenna WJ. A new mutation of the cardiac troponin T gene causing familial hypertrophic cardiomyopathy without left ventricular hypertrophy. *Heart* 1999;**82**:621–624.
16. Maron BJ, Maron MS, Rowin EJ. Perspectives on the Overall Risks of Living With Hypertrophic Cardiomyopathy. *Circulation* 2017;**135**:2317–2319.
17. Colan SD, Lipshultz SE, Lowe AM, Sleeper LA, Messere J, Cox GF, Lurie PR, Orav EJ, Towbin JA. Epidemiology and cause-specific outcome of hypertrophic cardiomyopathy in children: findings from the Pediatric Cardiomyopathy Registry. *Circulation* 2007;**115**:773–781.
18. McKenna W, Deanfield J, Faruqui A, England D, Oakley C, Goodwin J. Prognosis in hypertrophic cardiomyopathy: role of age and clinical, electrocardiographic and

- hemodynamic features. *Am J Cardiol* 1981;**47**:532–538.
19. Schinkel AFL, Vriesendorp PA, Sijbrands EJG, Jordaens LJLM, Cate FJ ten, Michels M. Outcome and complications after implantable cardioverter defibrillator therapy in hypertrophic cardiomyopathy: systematic review and meta-analysis. *Circ Heart Fail* 2012;**5**:552–559.
 20. O'Mahony C, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi C, Biagini E, Gimeno JR, Limongelli G, McKenna WJ, Omar RZ, Elliott PM, Hypertrophic Cardiomyopathy Outcomes Investigators. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J* 2014;**35**:2010–2020.
 21. Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, Seidman CE, Towbin JA, Udelson JE, Yancy CW, American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* 2011;**124**:e783-831.
 22. Nistri S, Olivotto I, Maron MS, Ferrantini C, Coppini R, Grifoni C, Baldini K, Sgalambro A, Cecchi F, Maron BJ. β Blockers for prevention of exercise-induced left ventricular outflow tract obstruction in patients with hypertrophic cardiomyopathy. *Am J Cardiol* 2012;**110**:715–719.
 23. Nishimura RA, Seggewiss H, Schaff H V. Hypertrophic Obstructive Cardiomyopathy: Surgical Myectomy and Septal Ablation. *Circ Res* 2017;**121**:771–783.
 24. Harris KM, Spirito P, Maron MS, Zenovich AG, Formisano F, Lesser JR, Mackey-Bojack S, Manning WJ, Udelson JE, Maron BJ. Prevalence, clinical profile, and significance of left ventricular remodeling in the end-stage phase of hypertrophic cardiomyopathy. *Circulation* 2006;**114**:216–225.

25. Mearini G, Stimpel D, Geertz B, Weinberger F, Krämer E, Schlossarek S, Mouroto-Filiatre J, Stoehr A, Dutsch A, Wijnker PJM, Braren I, Katus HA, Müller OJ, Voit T, Eschenhagen T, Carrier L. Mybpc3 gene therapy for neonatal cardiomyopathy enables long-term disease prevention in mice. *Nat Commun* 2014;**5**:5515.
26. Mearini G, Stimpel D, Krämer E, Geertz B, Braren I, Gedicke-Hornung C, Précigout G, Müller OJ, Katus HA, Eschenhagen T, Voit T, Garcia L, Lorain S, Carrier L. Repair of Mybpc3 mRNA by 5'-trans-splicing in a Mouse Model of Hypertrophic Cardiomyopathy. *Mol Ther Nucleic Acids* 2013;**2**:e102.
27. Gedicke-Hornung C, Behrens-Gawlik V, Reischmann S, Geertz B, Stimpel D, Weinberger F, Schlossarek S, Précigout G, Braren I, Eschenhagen T, Mearini G, Lorain S, Voit T, Dreyfus PA, Garcia L, Carrier L. Rescue of cardiomyopathy through U7snRNA-mediated exon skipping in Mybpc3-targeted knock-in mice. *EMBO Mol Med* 2013;**5**:1128–1145.
28. Jiang J, Wakimoto H, Seidman JG, Seidman CE. Allele-specific silencing of mutant Myh6 transcripts in mice suppresses hypertrophic cardiomyopathy. *Science* 2013;**342**:111–114.
29. Ma H, Marti-Gutierrez N, Park S-W, Wu J, Lee Y, Suzuki K, Koski A, Ji D, Hayama T, Ahmed R, Darby H, Dyken C Van, Li Y, Kang E, Park A-R, Kim D, Kim S-T, Gong J, Gu Y, Xu X, Battaglia D, Krieg SA, Lee DM, Wu DH, Wolf DP, Heitner SB, Belmonte JCI, Amato P, Kim J-S, Kaul S, et al. Correction of a pathogenic gene mutation in human embryos. *Nature* 2017;**548**:413–419.
30. Ho CY, McMurray JJ V, Cirino AL, Colan SD, Day SM, Desai AS, Lipshultz SE, MacRae CA, Shi L, Solomon SD, Orav EJ, Braunwald E, VANISH trial investigators and executive committee. The Design of the Valsartan for Attenuating Disease Evolution in Early Sarcomeric Hypertrophic Cardiomyopathy (VANISH) Trial. *Am Heart J* 2017;**187**:145–155.
31. Green EM, Wakimoto H, Anderson RL, Evanchik MJ, Gorham JM, Harrison BC, Henze M, Kawa R, Oslob JD, Rodriguez HM, Song Y, Wan W, Leinwand LA, Spudich JA, McDowell RS, Seidman JG, Seidman CE. A small-molecule inhibitor of sarcomere contractility suppresses hypertrophic cardiomyopathy in mice. *Science* 2016;**351**:617–

- 621.
32. Seidman CE, Seidman JG. Identifying sarcomere gene mutations in hypertrophic cardiomyopathy: a personal history. *Circ Res* 2011;**108**:743–750.
 33. Alfares AA, Kelly MA, McDermott G, Funke BH, Lebo MS, Baxter SB, Shen J, McLaughlin HM, Clark EH, Babb LJ, Cox SW, DePalma SR, Ho CY, Seidman JG, Seidman CE, Rehm HL. Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet Med* 2015;**17**:880–888.
 34. Maron BJ, Maron MS, Semsarian C. Genetics of hypertrophic cardiomyopathy after 20 years: clinical perspectives. *J Am Coll Cardiol* 2012;**60**:705–715.
 35. Marston S, Copeland O, Jacques A, Livesey K, Tsang V, McKenna WJ, Jalilzadeh S, Carballo S, Redwood C, Watkins H. Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency. *Circ Res* 2009;**105**:219–222.
 36. Dijk SJ van, Dooijes D, Remedios C dos, Michels M, Lamers JMJ, Winegrad S, Schlossarek S, Carrier L, Cate FJ ten, Stienen GJM, Velden J van der. Cardiac myosin-binding protein C mutations and hypertrophic cardiomyopathy: haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction. *Circulation* 2009;**119**:1473–1483.
 37. Barefield D, Kumar M, Gorham J, Seidman JG, Seidman CE, Tombe PP de, Sadayappan S. Haploinsufficiency of MYBPC3 exacerbates the development of hypertrophic cardiomyopathy in heterozygous mice. *J Mol Cell Cardiol* 2015;**79**:234–243.
 38. Ho CY, López B, Coelho-Filho OR, Lakdawala NK, Cirino AL, Jarolim P, Kwong R, González A, Colan SD, Seidman JG, Díez J, Seidman CE. Myocardial fibrosis as an early manifestation of hypertrophic cardiomyopathy. *N Engl J Med* 2010;**363**:552–563.
 39. Lakdawala NK, Thune JJ, Colan SD, Cirino AL, Farrohi F, Rivero J, McDonough B, Sparks E, Orav EJ, Seidman JG, Seidman CE, Ho CY. Subtle abnormalities in contractile function are an early manifestation of sarcomere mutations in dilated cardiomyopathy. *Circ Cardiovasc Genet* 2012;**5**:503–510.

40. Valente AM, Lakdawala NK, Powell AJ, Evans SP, Cirino AL, Orav EJ, MacRae CA, Colan SD, Ho CY. Comparison of echocardiographic and cardiac magnetic resonance imaging in hypertrophic cardiomyopathy sarcomere mutation carriers without left ventricular hypertrophy. *Circ Cardiovasc Genet* 2013;**6**:230–237.
41. Gray B, Ingles J, Semsarian C. Natural history of genotype positive-phenotype negative patients with hypertrophic cardiomyopathy. *Int J Cardiol* 2011;**152**:258–259.
42. Jensen MK, Havndrup O, Christiansen M, Andersen PS, Diness B, Axelsson A, Skovby F, Køber L, Bundgaard H. Penetrance of hypertrophic cardiomyopathy in children and adolescents: a 12-year follow-up study of clinical screening and predictive genetic testing. *Circulation* 2013;**127**:48–54.
43. Kuliev A, Pomerantseva E, Polling D, Verlinsky O, Rechitsky S. PGD for inherited cardiac diseases. *Reprod Biomed Online* 2012;**24**:443–453.
44. Wordsworth S, Leal J, Blair E, Legood R, Thomson K, Seller A, Taylor J, Watkins H. DNA testing for hypertrophic cardiomyopathy: a cost-effectiveness model. *Eur Heart J* 2010;**31**:926–935.
45. Ingles J, McGaughan J, Scuffham PA, Atherton J, Semsarian C. A cost-effectiveness model of genetic testing for the evaluation of families with hypertrophic cardiomyopathy. *Heart* 2012;**98**:625–630.
46. Elliott P, Baker R, Pasquale F, Quarta G, Ebrahim H, Mehta AB, Hughes DA, ACES study group. Prevalence of Anderson-Fabry disease in patients with hypertrophic cardiomyopathy: the European Anderson-Fabry Disease survey. *Heart* 2011;**97**:1957–1960.
47. D'souza RS, Levandowski C, Slavov D, Graw SL, Allen LA, Adler E, Mestroni L, Taylor MRG. Danon disease: clinical features, evaluation, and management. *Circ Heart Fail* 2014;**7**:843–849.
48. Kelly M, Semsarian C. Multiple mutations in genetic cardiovascular disease: a marker of disease severity? *Circ Cardiovasc Genet* 2009;**2**:182–190.
49. Fourey D, Care M, Siminovitch KA, Weissler-Snir A, Hindieh W, Chan RH, Gollob MH, Rakowski H, Adler A. Prevalence and Clinical Implication of Double Mutations in

- Hypertrophic Cardiomyopathy: Revisiting the Gene-Dose Effect. *Circ Cardiovasc Genet* 2017;**10**:doi:10.1161/CIRCGENETICS.116.001685.
50. Su M, Wang J, Kang L, Wang Y, Zou Y, Feng X, Wang D, Ahmad F, Zhou X, Hui R, Song L. Rare variants in genes encoding MuRF1 and MuRF2 are modifiers of hypertrophic cardiomyopathy. *Int J Mol Sci* 2014;**15**:9302–9313.
 51. Wooten EC, Hebl VB, Wolf MJ, Greytak SR, Orr NM, Draper I, Calvino JE, Kapur NK, Maron MS, Kullo IJ, Ommen SR, Bos JM, Ackerman MJ, Huggins GS. Formin homology 2 domain containing 3 variants associated with hypertrophic cardiomyopathy. *Circ Cardiovasc Genet* 2013;**6**:10–18.
 52. Lechin M, Quiñones MA, Omran A, Hill R, Yu QT, Rakowski H, Wigle D, Liew CC, Sole M, Roberts R. Angiotensin-I converting enzyme genotypes and left ventricular hypertrophy in patients with hypertrophic cardiomyopathy. *Circulation* 1995;**92**:1808–1812.
 53. Ortlepp JR, Vosberg HP, Reith S, Ohme F, Mahon NG, Schröder D, Klues HG, Hanrath P, McKenna WJ. Genetic polymorphisms in the renin-angiotensin-aldosterone system associated with expression of left ventricular hypertrophy in hypertrophic cardiomyopathy: a study of five polymorphic genes in a family with a disease causing mutation in the myosin bindin. *Heart* 2002;**87**:270–275.
 54. Perkins MJ, Driest SL Van, Ellsworth EG, Will ML, Gersh BJ, Ommen SR, Ackerman MJ. Gene-specific modifying effects of pro-LVH polymorphisms involving the renin-angiotensin-aldosterone system among 389 unrelated patients with hypertrophic cardiomyopathy. *Eur Heart J* 2005;**26**:2457–2462.
 55. Amin AS, Giudicessi JR, Tijssen AJ, Spanjaart AM, Reckman YJ, Klemens CA, Tanck MW, Kapplinger JD, Hofman N, Sinner MF, Müller M, Wijnen WJ, Tan HL, Bezzina CR, Creemers EE, Wilde AAM, Ackerman MJ, Pinto YM. Variants in the 3' untranslated region of the KCNQ1-encoded Kv7.1 potassium channel modify disease severity in patients with type 1 long QT syndrome in an allele-specific manner. *Eur Heart J* 2012;**33**:714–723.
 56. Crotti L, Lahtinen AM, Spazzolini C, Mastantuono E, Monti MC, Morassutto C, Parati G, Heradien M, Goosen A, Lichtner P, Meitinger T, Brink PA, Kontula K, Swan H,

- Schwartz PJ. Genetic Modifiers for the Long-QT Syndrome: How Important Is the Role of Variants in the 3' Untranslated Region of KCNQ1? *Circ Cardiovasc Genet* 2016;**9**:330–339.
57. Amin AS, Pinto YM, Ackerman MJ, Wilde AAM. Letter by Amin et al Regarding Article, 'Genetic Modifiers for the Long-QT Syndrome: How Important Is the Role of Variants in the 3' Untranslated Region of KCNQ1?' *Circ Cardiovasc Genet* 2016;**9**:580.
 58. Crotti L, Lahtinen AM, Spazzolini C, Mastantuono E, Cristina Monti M, Morassutto C, Parati G, Heradien M, Goosen A, Lichtner P, Meitinger T, Brink PA, Kontula K, Swan H, Schwartz PJ. Response by Crotti et al to Letter Regarding Article, 'Genetic Modifiers for the Long-QT Syndrome: How Important Is the Role of Variants in the 3' Untranslated Region of KCNQ1?' *Circ Cardiovasc Genet* 2016;**9**:581–582.
 59. Rooij E van, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, Richardson JA, Olson EN. A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. *Proc Natl Acad Sci U S A* 2006;**103**:18255–18260.
 60. Kumarswamy R, Thum T. Non-coding RNAs in cardiac remodeling and heart failure. *Circ Res* 2013;**113**:676–689.
 61. Cheng Y, Ji R, Yue J, Yang J, Liu X, Chen H, Dean DB, Zhang C. MicroRNAs are aberrantly expressed in hypertrophic heart: do they play a role in cardiac hypertrophy? *Am J Pathol* 2007;**170**:1831–1840.
 62. Olivetto I, Maron BJ, Tomberli B, Appelbaum E, Salton C, Haas TS, Gibson CM, Nistri S, Servettini E, Chan RH, Udelson JE, Lesser JR, Cecchi F, Manning WJ, Maron MS. Obesity and its association to phenotype and clinical course in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2013;**62**:449–457.
 63. Pua CJ, Bhalshankar J, Miao K, Walsh R, John S, Lim SQ, Chow K, Buchan R, Soh BY, Lio PM, Lim J, Schafer S, Lim JQ, Tan P, Whiffin N, Barton PJ, Ware JS, Cook SA. Development of a Comprehensive Sequencing Assay for Inherited Cardiac Condition Genes. *J Cardiovasc Transl Res* 2016;**9**:3–11.
 64. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier J-B, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P,

- Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet* 2015;**47**:717–726.
65. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K-I, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;**42**:790–793.
 66. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–1303.
 67. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol* 2016;**17**:122.
 68. Lek M, Karczewski KJ, Minikel E V, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285–291.
 69. Stenson PD, Mort M, Ball E V, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;**133**:1–9.
 70. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;**44**:D862-8.
 71. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E,

- Spector E, Voelkerding K, Rehm HL. 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. *Genet Med* 2015;**17**:405–423.
72. Jarcho JA, McKenna W, Pare JA, Solomon SD, Holcombe RF, Dickie S, Levi T, Donis-Keller H, Seidman JG, Seidman CE. Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1. *N Engl J Med* 1989;**321**:1372–1378.
 73. Driest SL Van, Vasile VC, Ommen SR, Will ML, Tajik AJ, Gersh BJ, Ackerman MJ. Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004;**44**:1903–1910.
 74. Driest SL Van, Ellsworth EG, Ommen SR, Tajik AJ, Gersh BJ, Ackerman MJ. Prevalence and spectrum of thin filament mutations in an outpatient referral population with hypertrophic cardiomyopathy. *Circulation* 2003;**108**:445–451.
 75. Driest SL Van, Jaeger MA, Ommen SR, Will ML, Gersh BJ, Tajik AJ, Ackerman MJ. Comprehensive analysis of the beta-myosin heavy chain gene in 389 unrelated patients with hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004;**44**:602–610.
 76. Solomon SD, Jarcho JA, McKenna W, Geisterfer-Lowrance A, Germain R, Salerni R, Seidman JG, Seidman CE. Familial hypertrophic cardiomyopathy is a genetically heterogeneous disease. *J Clin Invest* 1990;**86**:993–999.
 77. Thierfelder L, Watkins H, MacRae C, Lamas R, McKenna W, Vosberg HP, Seidman JG, Seidman CE. Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. *Cell* 1994;**77**:701–712.
 78. Kimura A, Harada H, Park JE, Nishi H, Satoh M, Takahashi M, Hiroi S, Sasaoka T, Ohbuchi N, Nakamura T, Koyanagi T, Hwang TH, Choo JA, Chung KS, Hasegawa A, Nagai R, Okazaki O, Nakamura H, Matsuzaki M, Sakamoto T, Toshima H, Koga Y, Imaizumi T, Sasazuki T. Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. *Nat Genet* 1997;**16**:379–382.
 79. Poetter K, Jiang H, Hassanzadeh S, Master SR, Chang A, Dalakas MC, Rayment I, Sellers JR, Fananapazir L, Epstein ND. Mutations in either the essential or regulatory

- light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. *Nat Genet* 1996;**13**:63–69.
80. Mogensen J, Klausen IC, Pedersen AK, Egeblad H, Bross P, Kruse TA, Gregersen N, Hansen PS, Baandrup U, Borglum AD. Alpha-cardiac actin is a novel disease gene in familial hypertrophic cardiomyopathy. *J Clin Invest* 1999;**103**:R39-43.
 81. Herman DS, Lam L, Taylor MRG, Wang L, Teekakirikul P, Christodoulou D, Conner L, DePalma SR, McDonough B, Sparks E, Teodorescu DL, Cirino AL, Banner NR, Pennell DJ, Graw S, Merlo M, Lenarda A Di, Sinagra G, Bos JM, Ackerman MJ, Mitchell RN, Murry CE, Lakdawala NK, Ho CY, Barton PJR, Cook SA, Mestroni L, Seidman JG, Seidman CE. Truncations of Titin Causing Dilated Cardiomyopathy. *N. Engl. J. Med.* 2012. p. 619–628.
 82. Pugh TJ, Kelly M a, Gowrisankar S, Hynes E, Seidman M a, Baxter SM, Bowser M, Harrison B, Aaron D, Mahanta LM, Lakdawala NK, McDermott G, White ET, Rehm HL, Lebo M, Funke BH. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet Med* 2014;**16**:601–608.
 83. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015;**526**:68–74.
 84. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. *Circ Cardiovasc Genet* 2012;**5**:391–399.
 85. Andreassen C, Nielsen JB, Refsgaard L, Holst AG, Christensen AH, Andreassen L, Sajadieh A, Haunsø S, Svendsen JH, Olesen MS. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur J Hum Genet* 2013;**21**:918–928.
 86. Lopes LR, Syrris P, Guttman OP, O’Mahony C, Tang HC, Dalageorgou C, Jenkins S, Hubank M, Monserrat L, McKenna WJ, Plagnol V, Elliott PM. Novel genotype-phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy. *Heart* 2014;**101**:294–301.

87. Haas J, Frese KS, Peil B, Kloos W, Keller A, Nietsch R, Feng Z, Müller S, Kayvanpour E, Vogel B, Sedaghat-Hamedani F, Lim W-K, Zhao X, Fradkin D, Köhler D, Fischer S, Franke J, Marquart S, Barb I, Li DT, Amr A, Ehlermann P, Mereles D, Weis T, Hassel S, Kremer A, King V, Wirsz E, Isnard R, Komajda M, et al. Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J* 2015;**36**:1123–35a.
88. Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, Mazzarotto F, Blair E, Seller A, Taylor JC, Minikel E V, Exome Aggregation Consortium, MacArthur DG, Farrall M, Cook SA, Watkins H. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet Med* 2017;**19**:192–203.
89. Walsh R, Buchan R, Wilk A, John S, Felkin LE, Thomson KL, Chiaw TH, Loong CCW, Pua CJ, Raphael C, Prasad S, Barton PJ, Funke B, Watkins H, Ware JS, Cook SA. Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes. *Eur Heart J* 2017;**38**:3461–3468.
90. Hershberger RE, Norton N, Morales A, Li D, Siegfried JD, Gonzalez-Quintana J. Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet* 2010;**3**:155–161.
91. Hershberger RE, Parks SB, Kushner JD, Li D, Ludwigsen S, Jakobs P, Nauman D, Burgess D, Partain J, Litt M. Coding sequence mutations identified in MYH7, TNNT2, SCN5A, CSRP3, LBD3, and TCAP from 313 patients with familial or idiopathic dilated cardiomyopathy. *Clin Transl Sci* 2008;**1**:21–26.
92. Parks SB, Kushner JD, Nauman D, Burgess D, Ludwigsen S, Peterson A, Li D, Jakobs P, Litt M, Porter CB, Rahko PS, Hershberger RE. Lamin A/C mutation analysis in a cohort of 324 unrelated patients with idiopathic or familial dilated cardiomyopathy. *Am Heart J* 2008;**156**:161–169.
93. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, Cooper GM, Dorschner MO, Dulik MC, Ghazani AA, Ghosh R, Green RC, Hart R, Horton C, Johnston JJ, Lebo MS, Milosavljevic A, Ou J, Pak CM, Patel RY, Punj S, Richards CS, Salama J, Strande NT, Yang Y, Plon SE, et al.

- Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet* 2016;**98**:1067–1076.
94. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;**26**:589–595.
 95. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010;**26**:2069–2070.
 96. Adalsteinsdottir B, Teekakirikul P, Maron BJ, Burke MA, Gudbjartsson DF, Holm H, Stefansson K, DePalma SR, Mazaika E, McDonough B, Danielsen R, Seidman JG, Seidman CE, Gunnarsson GT. Nationwide study on hypertrophic cardiomyopathy in Iceland: evidence of a MYBPC3 founder mutation. *Circulation* 2014;**130**:1158–1167.
 97. Strande NT, Riggs ER, Buchanan AH, Ceyhan-Birsoy O, DiStefano M, Dwight SS, Goldstein J, Ghosh R, Seifert BA, Sneddon TP, Wright MW, Milko L V, Cherry JM, Giovanni MA, Murray MF, O'Daniel JM, Ramos EM, Santani AB, Scott AF, Plon SE, Rehm HL, Martin CL, Berg JS. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet* 2017;**100**:895–906.
 98. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, Buchan RJ, Walsh R, John S, Wilkinson S, Mazzarotto F, Felkin LE, Gong S, MacArthur JAL, Cunningham F, Flannick J, Gabriel SB, Altshuler DM, Macdonald PS, Heinig M, Keogh AM, Hayward CS, Banner NR, Pennell DJ, O'Regan DP, San TR, Marvao A de, Dawes TJW, Gulati A, Birks EJ, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med* 2015;**7**:270ra6.
 99. Schafer S, Marvao A de, Adami E, Fiedler LR, Ng B, Khin E, Rackham OJL, Heesch S van, Pua CJ, Kui M, Walsh R, Tayal U, Prasad SK, Dawes TJW, Ko NSJ, Sim D, Chan LLH, Chin CWL, Mazzarotto F, Barton PJ, Kreuchwig F, Kleijn DP V de, Totman T, Biffi C, Tee N, Rueckert D, Schneider V, Faber A, Regitz-Zagrosek V, Seidman JG, et al. Titin-truncating variants affect heart function in disease cohorts and the general

- population. *Nat Genet* 2017;**49**:46–53.
100. Boczek NJ, Ye D, Jin F, Tester DJ, Huseby A, Bos JM, Johnson AJ, Kanter R, Ackerman MJ. Identification and Functional Characterization of a Novel CACNA1C-Mediated Cardiac Disorder Characterized by Prolonged QT Intervals With Hypertrophic Cardiomyopathy, Congenital Heart Defects, and Sudden Cardiac Death. *Circ Arrhythm Electrophysiol* 2015;**8**:1122–1132.
 101. Wang L, Zuo L, Hu J, Shao H, Lei C, Qi W, Liu Y, Miao Y, Ma X, Huang CL-H, Wang B, Zhou X, Zhang Y, Liu L. Dual LQT1 and HCM phenotypes associated with tetrad heterozygous mutations in KCNQ1, MYH7, MYLK2, and TMEM70 genes in a three-generation Chinese family. *Europace* 2016;**18**:602–609.
 102. D’Argenio V, Frisso G, Precone V, Boccia A, Fienga A, Pacileo G, Limongelli G, Paoletta G, Calabrò R, Salvatore F. DNA sequence capture and next-generation sequencing for the molecular diagnosis of genetic cardiomyopathies. *J Mol Diagn* 2014;**16**:32–44.
 103. D’Amico A, Graziano C, Pacileo G, Petrini S, Nowak KJ, Boldrini R, Jacques A, Feng J-J, Porfirio B, Sewry CA, Santorelli FM, Limongelli G, Bertini E, Laing N, Marston SB. Fatal hypertrophic cardiomyopathy and nemaline myopathy associated with ACTA1 K336E mutation. *Neuromuscul Disord* 2006;**16**:548–552.
 104. Kim S-Y, Park Y-E, Kim H-S, Lee C-H, Yang DH, Kim D-S. Nemaline myopathy and non-fatal hypertrophic cardiomyopathy caused by a novel ACTA1 E239K mutation. *J Neurol Sci* 2011;**307**:171–173.
 105. Antonicka H, Mattman A, Carlson CG, Glerum DM, Hoffbuhr KC, Leary SC, Kennaway NG, Shoubbridge EA. Mutations in COX15 produce a defect in the mitochondrial heme biosynthetic pathway, causing early-onset fatal hypertrophic cardiomyopathy. *Am J Hum Genet* 2003;**72**:101–114.
 106. Galmiche L, Serre V, Beinat M, Assouline Z, Lebre A-S, Chretien D, Nietschke P, Benes V, Boddaert N, Sidi D, Brunelle F, Rio M, Munnich A, Rötig A. Exome sequencing identifies MRPL3 mutation in mitochondrial cardiomyopathy. *Hum Mutat* 2011;**32**:1225–1231.
 107. Mohiddin SA, Ahmed ZM, Griffith AJ, Tripodi D, Friedman TB, Fananapazir L, Morell

- RJ. Novel association of hypertrophic cardiomyopathy, sensorineural deafness, and a mutation in unconventional myosin VI (MYO6). *J Med Genet* 2004;**41**:309–314.
108. Palmieri L, Alberio S, Pisano I, Lodi T, Meznaric-Petrusa M, Zidar J, Santoro A, Scarcia P, Fontanesi F, Lamantea E, Ferrero I, Zeviani M. Complete loss-of-function of the heart/muscle-specific adenine nucleotide translocator is associated with mitochondrial myopathy and cardiomyopathy. *Hum Mol Genet* 2005;**14**:3079–3088.
 109. Echaniz-Laguna A, Chassagne M, Ceresuela J, Rouvet I, Padet S, Acquaviva C, Nataf S, Vinzio S, Bozon D, Mousson de Camaret B. Complete loss of expression of the ANT1 gene causing cardiomyopathy and myopathy. *J Med Genet* 2012;**49**:146–150.
 110. Strauss KA, DuBiner L, Simon M, Zaragoza M, Sengupta PP, Li P, Narula N, Dreike S, Platt J, Procaccio V, Ortiz-González XR, Puffenberger EG, Kelley RI, Morton DH, Narula J, Wallace DC. Severity of cardiomyopathy associated with adenine nucleotide translocator-1 deficiency correlates with mtDNA haplogroup. *Proc Natl Acad Sci U S A* 2013;**110**:3453–3458.
 111. Körver-Keularts IMLW, Visser M de, Bakker HD, Wanders RJA, Vansenne F, Scholte HR, Dorland L, Nicolaes GAF, Spaapen LMJ, Smeets HJM, Hendrickx ATM, Bosch BJC van den. Two Novel Mutations in the SLC25A4 Gene in a Patient with Mitochondrial Myopathy. *JIMD Rep* 2015;**22**:39–45.
 112. Geier C, Gehmlich K, Ehler E, Hassfeld S, Perrot A, Hayess K, Cardim N, Wenzel K, Erdmann B, Krackhardt F, Posch MG, Osterziel KJ, Bublak A, Nägele H, Scheffold T, Dietz R, Chien KR, Spuler S, Fürst DO, Nürnberg P, Ozcelik C. Beyond the sarcomere: CSRP3 mutations cause hypertrophic cardiomyopathy. *Hum Mol Genet* 2008;**17**:2753–2765.
 113. Friedrich FW, Wilding BR, Reischmann S, Crocini C, Lang P, Charron P, Müller OJ, McGrath MJ, Vollert I, Hansen A, Linke WA, Hengstenberg C, Bonne G, Morner S, Wichter T, Madeira H, Arbustini E, Eschenhagen T, Mitchell CA, Isnard R, Carrier L. Evidence for FHL1 as a novel disease gene for isolated hypertrophic cardiomyopathy. *Hum Mol Genet* 2012;**21**:3237–3254.
 114. Chiu C, Bagnall RD, Ingles J, Yeates L, Kennerson M, Donald JA, Jormakka M, Lind JM, Semsarian C. Mutations in Alpha-Actinin-2 Cause Hypertrophic Cardiomyopathy. *J Am*

- Coll Cardiol* 2010;**55**:1127–1135.
115. Girolami F, Iascone M, Tomberli B, Bardi S, Benelli M, Marseglia G, Pescucci C, Pezzoli L, Sana ME, Basso C, Marziliano N, Merlini PA, Fornaro A, Cecchi F, Torricelli F, Olivetto I. Novel α -actinin 2 variant associated with familial hypertrophic cardiomyopathy and juvenile atrial arrhythmias: a massively parallel sequencing study. *Circ Cardiovasc Genet* 2014;**7**:741–750.
 116. Osio A, Tan L, Chen SN, Lombardi R, Nagueh SF, Shete S, Roberts R, Willerson JT, Marian AJ. Myozenin 2 is a novel gene for human hypertrophic cardiomyopathy. *Circ Res* 2007;**100**:766–768.
 117. Ruggiero A, Chen SN, Lombardi R, Rodriguez G, Marian AJ. Pathogenesis of hypertrophic cardiomyopathy caused by myozenin 2 mutations is independent of calcineurin activity. *Cardiovasc Res* 2013;**97**:44–54.
 118. Valdés-Mas R, Gutiérrez-Fernández A, Gómez J, Coto E, Astudillo A, Puente DA, Reguero JR, Álvarez V, Morís C, León D, Martín M, Puente XS, López-Otín C. Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. *Nat Commun* 2014;**5**:5326.
 119. Gómez J, Lorca R, Reguero JR, Morís C, Martín M, Tranche S, Alonso B, Iglesias S, Alvarez V, Díaz-Molina B, Avanzas P, Coto E. Screening of the Filamin C Gene in a Large Cohort of Hypertrophic Cardiomyopathy Patients. *Circ Cardiovasc Genet* 2017;**10**:doi:10.1161/CIRCGENETICS.116.001584.
 120. Chung WK, Kitner C, Maron BJ. Novel frameshift mutation in Troponin C (TNNC1) associated with hypertrophic cardiomyopathy and sudden death. *Cardiol Young* 2011;**21**:345–348.
 121. Purevjav E, Arimura T, Augustin S, Huby A-C, Takagi K, Nunoda S, Kearney DL, Taylor MD, Terasaki F, Bos JM, Ommen SR, Shibata H, Takahashi M, Itoh-Satoh M, McKenna WJ, Murphy RT, Labeit S, Yamanaka Y, Machida N, Park J-E, Alexander PMA, Weintraub RG, Kitaura Y, Ackerman MJ, Kimura A, Towbin JA. Molecular basis for clinical heterogeneity in inherited cardiomyopathies due to myopalladin mutations. *Hum Mol Genet* 2012;**21**:2039–2053.

122. Vasile VC, Ommen SR, Edwards WD, Ackerman MJ. A missense mutation in a ubiquitously expressed protein, vinculin, confers susceptibility to hypertrophic cardiomyopathy. *Biochem Biophys Res Commun* 2006;**345**:998–1003.
123. Vasile VC, Will ML, Ommen SR, Edwards WD, Olson TM, Ackerman MJ. Identification of a metavinculin missense mutation, R975W, associated with both hypertrophic and dilated cardiomyopathy. *Mol Genet Metab* 2006;**87**:169–174.
124. Vasile VC, Edwards WD, Ommen SR, Ackerman MJ. Obstructive hypertrophic cardiomyopathy is associated with reduced expression of vinculin in the intercalated disc. *Biochem Biophys Res Commun* 2006;**349**:709–715.
125. Cirino AL, Lakdawala NK, McDonough B, Conner L, Adler D, Weinfeld M, O’Gara P, Rehm HL, Machini K, Lebo M, Blout C, Green RC, MacRae CA, Seidman CE, Ho CY, MedSeq Project*. A Comparison of Whole Genome Sequencing to Multigene Panel Testing in Hypertrophic Cardiomyopathy Patients. *Circ Cardiovasc Genet* 2017;**10**:doi:10.1161/CIRCGENETICS.117.001768.
126. Almomani R, Verhagen JMA, Herkert JC, Brosens E, Spaendonck-Zwarts KY van, Asimaki A, Zwaag PA van der, Frohn-Mulder IME, Bertoli-Avella AM, Boven LG, Slegtenhorst MA van, Smagt JJ van der, IJcken WFJ van, Timmer B, Stuijvenberg M van, Verdijk RM, Saffitz JE, Plessis FA du, Michels M, Hofstra RMW, Sinke RJ, Tintelen JP van, Wessels MW, Jongbloed JDH, Laar IMBH van de. Biallelic Truncating Mutations in ALPK3 Cause Severe Pediatric Cardiomyopathy. *J Am Coll Cardiol* 2016;**67**:515–525.
127. Seidelmann SB, Smith E, Subrahmanyam L, Dykas D, Abou Ziki MD, Azari B, Hannah-Shmouni F, Jiang Y, Akar JG, Marieb M, Jacoby D, Bale AE, Lifton RP, Mani A. Application of Whole Exome Sequencing in the Clinical Diagnosis and Management of Inherited Cardiovascular Diseases in Adults. *Circ Cardiovasc Genet* 2017;**10**:doi:10.1161/CIRCGENETICS.116.001573.
128. Vos J, Otten W, Asperen C van, Jansen A, Menko F, Tibben A. The counsellees’ view of an unclassified variant in BRCA1/2: recall, interpretation, and impact on life. *Psychooncology* 2008;**17**:822–830.
129. Fox JC. SHaRe: advancing the goal of effective treatment for cardiomyopathy. *Eur Heart J* 2015;**36**:1142–1143.

130. Lopes LR, Murphy C, Syrris P, Dalageorgou C, McKenna WJ, Elliott PM, Plagnol V. Use of high-throughput targeted exome-sequencing to screen for copy number variation in hypertrophic cardiomyopathy. *Eur J Med Genet* 2015;**58**:611–616.
131. Mademont-Soler I, Mates J, Yotti R, Espinosa MA, Pérez-Serra A, Fernandez-Avila AI, Coll M, Méndez I, Iglesias A, Olmo B Del, Riuró H, Cuenca S, Allegue C, Campuzano O, Picó F, Ferrer-Costa C, Álvarez P, Castillo S, Garcia-Pavia P, Gonzalez-Lopez E, Padron-Barthe L, Díaz de Bustamante A, Darnaude MT, González-Hevia JI, Brugada J, Fernandez-Aviles F, Brugada R. Additional value of screening for minor genes and copy number variants in hypertrophic cardiomyopathy. *PLoS One* 2017;**12**:e0181465.
132. Ceyhan-Birsoy O, Pugh TJ, Bowser MJ, Hynes E, Frisella AL, Mahanta LM, Lebo MS, Amr SS, Funke BH. 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* 2016;**4**:143–151.
133. Pilichou K, Lazzarini E, Rigato I, Celeghin R, Bortoli M De, Perazzolo Marra M, Cason M, Jongbloed J, Calore M, Rizzo S, Regazzo D, Poloni G, Illiceto S, Daliento L, Delise P, Corrado D, Tintelen JP Van, Thiene G, Rampazzo A, Basso C, Bauce B, Lorenzon A, Occhi G. Large Genomic Rearrangements of Desmosomal Genes in Italian Arrhythmogenic Cardiomyopathy Patients. *Circ Arrhythm Electrophysiol* 2017;**10**:doi:10.1161/CIRCEP.117.005324.
134. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O’Grady GL, Estrella E, Reddy HM, Zhao F, Weisburd B, Karczewski KJ, O’Donnell-Luria AH, Birnbaum D, Sarkozy A, Hu Y, Gonorazky H, Claeys K, Joshi H, Bournazos A, Oates EC, Ghaoui R, Davis MR, Laing NG, Topf A, Genotype-Tissue Expression Consortium, Kang PB, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 2017;**9**:doi:10.1126/scitranslmed.aal5209.
135. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, Plon SE, Ramos EM, Sherry ST, Watson MS, ClinGen. ClinGen--the Clinical Genome Resource. *N Engl J Med* 2015;**372**:2235–2242.
136. Spaendonck-Zwarts KY van, Rijsingen IAW van, Berg MP van den, Lekanne Deprez RH,

- Post JG, Mil AM van, Asselbergs FW, Christiaans I, Langen IM van, Wilde AAM, Boer RA de, Jongbloed JDH, Pinto YM, Tintelen JP van. Genetic analysis in 418 index patients with idiopathic dilated cardiomyopathy: overview of 10 years' experience. *Eur J Heart Fail* 2013;**15**:628–636.
137. Miyagawa M, Naito T, Nishio S, Kamatani N, Usami S. Targeted exon sequencing successfully discovers rare causative genes and clarifies the molecular epidemiology of Japanese deafness patients. *PLoS One* 2013;**8**:e71381.
 138. Huang X-F, Huang F, Wu K-C, Wu J, Chen J, Pang C-P, Lu F, Qu J, Jin Z-B. Genotype-phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing. *Genet Med* 2015;**17**:271–278.
 139. Bonachea EM, Zender G, White P, Corsmeier D, Newsom D, Fitzgerald-Butt S, Garg V, McBride KL. Use of a targeted, combinatorial next-generation sequencing approach for the study of bicuspid aortic valve. *BMC Med Genomics* 2014;**7**:56.
 140. Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, Benaiche A, Isnard R, Dubourg O, Burban M, Gueffet J-P, Millaire A, Desnos M, Schwartz K, Hainque B, Komajda M, EUROGENE Heart Failure Project. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* 2003;**107**:2227–2232.
 141. Erdmann J, Daehmlow S, Wischke S, Senyuva M, Werner U, Raible J, Tanis N, Dyachenko S, Hummel M, Hetzer R, Regitz-Zagrosek V. Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. *Clin Genet* 2003;**64**:339–349.
 142. Song L, Zou Y, Wang J, Wang Z, Zhen Y, Lou K, Zhang Q, Wang X, Wang H, Li J, Hui R. Mutations profile in Chinese patients with hypertrophic cardiomyopathy. *Clin Chim Acta* 2005;**351**:209–216.
 143. Olivotto I, Girolami F, Ackerman MJ, Nistri S, Bos JM, Zachara E, Ommen SR, Theis JL, Vaubel RA, Re F, Armentano C, Poggesi C, Torricelli F, Cecchi F. Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. *Mayo Clin Proc* 2008;**83**:630–638.

144. García-Castro M, Coto E, Reguero JR, Berrazueta JR, Alvarez V, Alonso B, Sainz R, Martín M, Morís C. [Mutations in sarcomeric genes MYH7, MYBPC3, TNNT2, TNNI3, and TPM1 in patients with hypertrophic cardiomyopathy]. *Rev Esp Cardiol* 2009;**62**:48–56.
145. Millat G, Bouvagnet P, Chevalier P, Dauphin C, Jouk PS, Costa A Da, Prieur F, Bresson J-L, Faivre L, Eicher J-C, Chassaing N, Crehalet H, Porcher R, Rodriguez-Lafrasse C, Rousson R. Prevalence and spectrum of mutations in a cohort of 192 unrelated patients with hypertrophic cardiomyopathy. *Eur J Med Genet* **53**:261–267.
146. Harrison SM, Dolinsky JS, Knight Johnson AE, Pesaran T, Azzariti DR, Bale S, Chao EC, Das S, Vincent L, Rehm HL. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med* 2017;**19**:1096–1104.
147. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, Ware JS. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* 2017;**19**:1151–1158.
148. Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry* 2010;**19**:227–229.
149. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med* 1971;**25**:242–244.
150. Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. *Stat Med* 1989;**8**:845–859.
151. Altman D. Practical Statistics for medical research. London: Chapman and Hall; 1991.
152. Hildebrandt M, Bender R, Gehrman U, Blettner M. Calculating confidence intervals for impact numbers. *BMC Med Res Methodol* 2006;**6**:32.
153. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* 2016;**37**:235–241.
154. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* 2015;**24**:2125–2137.

155. Ye J, Pavlicek A, Lunney EA, Rejto PA, Teng C-H. Statistical method on nonrandom clustering with application to somatic mutations in cancer. *BMC Bioinformatics* 2010;**11**:11.
156. Whiffin N, Walsh R, Govind R, Edwards M, Ahmad M, Zhang X, Tayal U, Buchan R, Midwinter W, Wilk A, Najgebauer H, Francis C, Wilkinson S, Monk T, Brett L, O'Regan D, Prasad SK, Morris-Rosendahl DJ, Barton PJR, Edwards E, Ware JS, Cook SA. CardioClassifier: demonstrating the power of disease- and gene-specific computational decision support for clinical genome interpretation. *bioRxiv* 2017;doi:10.1101/180109.
157. Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF, McLean CY, Tung JY, Yu LPC, Gambetti P, Blevins J, Zhang S, Cohen Y, Chen W, Yamada M, Hamaguchi T, Sanjo N, Mizusawa H, Nakamura Y, Kitamoto T, Collins SJ, Boyd A, Will RG, Knight R, Ponto C, Zerr I, Kraus TFJ, Eigenbrod S, Giese A, Calero M, et al. Quantifying prion disease penetrance using large population control cohorts. *Sci Transl Med* 2016;**8**:322ra9.
158. Kirov G, Rees E, Walters JTR, Escott-Price V, Georgieva L, Richards AL, Chambert KD, Davies G, Legge SE, Moran JL, McCarroll SA, O'Donovan MC, Owen MJ. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry* 2014;**75**:378–385.
159. Alamo L, Ware JS, Pinto A, Gillilan RE, Seidman JG, Seidman CE, Padrón R. Effects of myosin variants on interacting-heads motif explain distinct hypertrophic and dilated cardiomyopathy phenotypes. *Elife* 2017;**6**:doi:10.7554/eLife.24634.
160. Homburger JR, Green EM, Caleshu C, Sunitha MS, Taylor RE, Ruppel KM, Metpally RPR, Colan SD, Michels M, Day SM, Olivotto I, Bustamante CD, Dewey FE, Ho CY, Spudich JA, Ashley EA. Multidimensional structure-function relationships in human β -cardiac myosin from population-scale genetic variation. *Proc Natl Acad Sci U S A* 2016;**113**:6701–6706.
161. Samocha K, Kosmicki J, Karczewski K, O'Donnell-Luria A, Pierce-Hoffman E, MacArthur D, Neale B, Daly M. Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* 2017;doi:10.1101/148353.

162. Brito D, Miltenberger-Miltenyi G, Vale Pereira S, Silva D, Diogo AN, Madeira H. Sarcomeric hypertrophic cardiomyopathy: genetic profile in a Portuguese population. *Rev Port Cardiol* 2012;**31**:577–587.
163. Kaski JP, Syrris P, Esteban MTT, Jenkins S, Pantazis A, Deanfield JE, McKenna WJ, Elliott PM. Prevalence of sarcomere protein gene mutations in preadolescent children with hypertrophic cardiomyopathy. *Circ Cardiovasc Genet* 2009;**2**:436–441.
164. Gimeno JR, Monserrat L, Pérez-Sánchez I, Marín F, Caballero L, Hermida-Prieto M, Castro A, Valdés M. Hypertrophic cardiomyopathy. A study of the troponin-T gene in 127 Spanish families. *Rev Esp Cardiol* 2009;**62**:1473–1477.
165. Miliou A, Anastasakis A, D’Cruz LG, Theopistou A, Rigopoulos A, Rizos I, Stamatelopoulos S, Toutouzas P, Stefanadis C. Low prevalence of cardiac troponin T mutations in a Greek hypertrophic cardiomyopathy cohort. *Heart* 2005;**91**:966–967.
166. Ingles J, Doolan A, Chiu C, Seidman J, Seidman C, Semsarian C. Compound and double mutations in patients with hypertrophic cardiomyopathy: implications for genetic testing and counselling. *J Med Genet* 2005;**42**:e59.
167. Zou Y, Wang J, Liu X, Wang Y, Chen Y, Sun K, Gao S, Zhang C, Wang Z, Zhang Y, Feng X, Song Y, Wu Y, Zhang H, Jia L, Wang H, Wang D, Yan C, Lu M, Zhou X, Song L, Hui R. Multiple gene mutations, not the type of mutation, are the modifier of left ventricle hypertrophy in patients with hypertrophic cardiomyopathy. *Mol Biol Rep* 2013;**40**:3969–3976.
168. Liu W, Liu W, Hu D, Zhu T, Ma Z, Yang J, Xie W, Li C, Li L, Yang J, Li T, Bian H, Tong Q. Mutation spectrum in a large cohort of unrelated Chinese patients with hypertrophic cardiomyopathy. *Am J Cardiol* 2013;**112**:585–589.
169. Otsuka H, Arimura T, Abe T, Kawai H, Aizawa Y, Kubo T, Kitaoka H, Nakamura H, Nakamura K, Okamoto H, Ichida F, Ayusawa M, Nunoda S, Isobe M, Matsuzaki M, Doi YL, Fukuda K, Sasaoka T, Izumi T, Ashizawa N, Kimura A. Prevalence and distribution of sarcomeric gene mutations in Japanese patients with familial hypertrophic cardiomyopathy. *Circ J* 2012;**76**:453–461.
170. Ripoll-Vera T, Gámez JM, Govea N, Gómez Y, Núñez J, Socías L, Escandell Á, Rosell J.

- Clinical and Prognostic Profiles of Cardiomyopathies Caused by Mutations in the Troponin T Gene. *Rev Esp Cardiol (Engl Ed)* 2016;**69**:149–158.
171. Theopistou A, Anastasakis A, Miliou A, Rigopoulos A, Toutouzas P, Stefanadis C. Clinical features of hypertrophic cardiomyopathy caused by an Arg278Cys missense mutation in the cardiac troponin T gene. *Am J Cardiol* 2004;**94**:246–249.
 172. García-Castro M, Reguero JR, Batalla A, Díaz-Molina B, González P, Alvarez V, Cortina A, Cubero GI, Coto E. Hypertrophic cardiomyopathy: low frequency of mutations in the beta-myosin heavy chain (MYH7) and cardiac troponin T (TNNT2) genes among Spanish patients. *Clin Chem* 2003;**49**:1279–1285.
 173. Watkins H, McKenna WJ, Thierfelder L, Suk HJ, Anan R, O'Donoghue A, Spirito P, Matsumori A, Moravec CS, Seidman JG. Mutations in the genes for cardiac troponin T and alpha-tropomyosin in hypertrophic cardiomyopathy. *N Engl J Med* 1995;**332**:1058–1064.
 174. Dhandapany PS, Sadayappan S, Xue Y, Powell GT, Rani DS, Nallari P, Rai TS, Khullar M, Soares P, Bahl A, Tharkan JM, Vaideeswar P, Rathinavel A, Narasimhan C, Ayapati DR, Ayub Q, Mehdi SQ, Oppenheimer S, Richards MB, Price AL, Patterson N, Reich D, Singh L, Tyler-Smith C, Thangaraj K. A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia. *Nat Genet* 2009;**41**:187–191.
 175. Ruklisa D, Ware JS, Walsh R, Balding DJ, Cook SA. Bayesian models for syndrome- and gene-specific probabilities of novel variant pathogenicity. *Genome Med* 2015;**7**:5.
 176. Rehm HL. A new era in the interpretation of human genomic variation. *Genet Med* 2017;**19**:1092–1095.
 177. Christiaans I, Nannenberg EA, Dooijes D, Jongbloed RJE, Michels M, Postema PG, Majoor-Krakauer D, Wijngaard A van den, Mannens MMAM, Tintelen JP van, Langen IM van, Wilde AAM. Founder mutations in hypertrophic cardiomyopathy patients in the Netherlands. *Neth Heart J* 2010;**18**:248–254.
 178. Misalski-Jamka K, Jefferies JL, Mazur W, Głowacki J, Hu J, Lazar M, Gibbs RA, Liczko J, Kłyś J, Venner E, Muzny DM, Rycaj J, Białkowski J, Kluczevska E, Kalarus Z, Jhangiani S,

- Al-Khalidi H, Kukulski T, Lupski JR, Craigen WJ, Bainbridge MN. Novel Genetic Triggers and Genotype-Phenotype Correlations in Patients With Left Ventricular Noncompaction. *Circ Cardiovasc Genet* 2017;**10**:doi:10.1161/CIRCGENETICS.117.001763.
179. Sedaghat-Hamedani F, Haas J, Zhu F, Geier C, Kayvanpour E, Liss M, Lai A, Frese K, Pribe-Wolferts R, Amr A, Li DT, Samani OS, Carstensen A, Bordalo DM, Müller M, Fischer C, Shao J, Wang J, Nie M, Yuan L, Haßfeld S, Schwartz C, Zhou M, Zhou Z, Shu Y, Wang M, Huang K, Zeng Q, Cheng L, Fehlmann T, et al. Clinical genetics and outcome of left ventricular non-compaction cardiomyopathy. *Eur Heart J* 2017;**38**:3449–3460.
 180. Niimura H, Bachinski LL, Sangwatanaroj S, Watkins H, Chudley AE, McKenna W, Kristinsson A, Roberts R, Sole M, Maron BJ, Seidman JG, Seidman CE. Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy. *N Engl J Med* 1998;**338**:1248–1257.
 181. Lopes LR, Rahman MS, Elliott PM. A systematic review and meta-analysis of genotype-phenotype associations in patients with hypertrophic cardiomyopathy caused by sarcomeric protein mutations. *Heart* 2013;**99**:1800–1811.
 182. Bos JM, Will ML, Gersh BJ, Kruisselbrink TM, Ommen SR, Ackerman MJ. Characterization of a phenotype-based genetic test prediction score for unrelated patients with hypertrophic cardiomyopathy. *Mayo Clin Proc* 2014;**89**:727–737.
 183. Li Q, Gruner C, Chan RH, Care M, Siminovitch K, Williams L, Woo A, Rakowski H. Genotype-positive status in patients with hypertrophic cardiomyopathy is associated with higher rates of heart failure events. *Circ Cardiovasc Genet* 2014;**7**:416–422.
 184. Velzen HG van, Vriesendorp PA, Oldenburg RA, Slegtenhorst MA van, Velden J van der, Schinkel AFL, Michels M. Value of Genetic Testing for the Prediction of Long-Term Outcome in Patients With Hypertrophic Cardiomyopathy. *Am J Cardiol* 2016;**118**:881–887.
 185. Ingles J, Sarina T, Yeates L, Hunt L, Macciocca I, McCormack L, Winship I, McGaughran J, Atherton J, Semsarian C. Clinical predictors of genetic testing outcomes in hypertrophic cardiomyopathy. *Genet Med* 2013;**15**:972–977.

186. Rowin EJ, Maron MS. The Role of Cardiac MRI in the Diagnosis and Risk Stratification of Hypertrophic Cardiomyopathy. *Arrhythmia Electrophysiol Rev* 2016;**5**:197–202.
187. Rubinshtein R, Glockner JF, Ommen SR, Araoz PA, Ackerman MJ, Sorajja P, Bos JM, Tajik AJ, Valeti US, Nishimura RA, Gersh BJ. Characteristics and clinical significance of late gadolinium enhancement by contrast-enhanced magnetic resonance imaging in patients with hypertrophic cardiomyopathy. *Circ Heart Fail* 2010;**3**:51–58.
188. Weissler-Snir A, Hindieh W, Gruner C, Fourey D, Appelbaum E, Rowin E, Care M, Lesser JR, Haas TS, Udelson JE, Manning WJ, Olivotto I, Tomberli B, Maron BJ, Maron MS, Crean AM, Rakowski H, Chan RH. Lack of Phenotypic Differences by Cardiovascular Magnetic Resonance Imaging in MYH7 (β -Myosin Heavy Chain)- Versus MYBPC3 (Myosin-Binding Protein C)-Related Hypertrophic Cardiomyopathy. *Circ Cardiovasc Imaging* 2017;**10**:doi:10.1161/CIRCIMAGING.116.005311.
189. Coppini R, Ho CY, Ashley E, Day S, Ferrantini C, Girolami F, Tomberli B, Bardi S, Torricelli F, Cecchi F, Mugelli A, Poggesi C, Tardiff J, Olivotto I. Clinical phenotype and outcome of hypertrophic cardiomyopathy associated with thin-filament gene mutations. *J Am Coll Cardiol* 2014;**64**:2589–2600.
190. Page SP, Kounas S, Syrris P, Christiansen M, Frank-Hansen R, Andersen PS, Elliott PM, McKenna WJ. Cardiac myosin binding protein-C mutations in families with hypertrophic cardiomyopathy: disease expression in relation to age, gender, and long term outcome. *Circ Cardiovasc Genet* 2012;**5**:156–166.
191. Maron BJ, Maron MS, Semsarian C. Double or compound sarcomere mutations in hypertrophic cardiomyopathy: a potential link to sudden death in the absence of conventional risk factors. *Heart Rhythm* 2012;**9**:57–63.
192. Chan RH, Maron BJ, Olivotto I, Pencina MJ, Assenza GE, Haas T, Lesser JR, Gruner C, Crean AM, Rakowski H, Udelson JE, Rowin E, Lombardi M, Cecchi F, Tomberli B, Spirito P, Formisano F, Biagini E, Rapezzi C, Cecco CN De, Autore C, Cook EF, Hong SN, Gibson CM, Manning WJ, Appelbaum E, Maron MS. Prognostic value of quantitative contrast-enhanced cardiovascular magnetic resonance for the evaluation of sudden death risk in patients with hypertrophic cardiomyopathy. *Circulation* 2014;**130**:484–495.

193. Helms AS, Day SM. Hypertrophic cardiomyopathy: single gene disease or complex trait? *Eur Heart J* 2016;**37**:1823–1825.
194. Ingles J, Burns C, Bagnall RD, Lam L, Yeates L, Sarina T, Puranik R, Briffa T, Atherton JJ, Driscoll T, Semsarian C. Nonfamilial Hypertrophic Cardiomyopathy: Prevalence, Natural History, and Clinical Implications. *Circ Cardiovasc Genet* 2017;**10**:doi:10.1161/CIRCGENETICS.116.001620.
195. Ko C, Arscott P, Concannon M, Saberi S, Day SM, Yashar BM, Helms AS. Genetic testing impacts the utility of prospective familial screening in hypertrophic cardiomyopathy through identification of a nonfamilial subgroup. *Genet Med* 2017;doi:10.1038/gim.2017.79.

7 Appendices

7.1 Summary of evidence for non-sarcomeric genes implicated in HCM

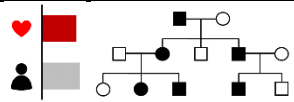
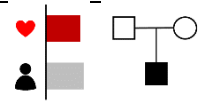
Summary of variant frequency and published evidence for non-sarcomeric genes associated with HCM. The ExAC data is the proportion of individuals with rare variants (defined as MAF < 0.0001). All cohort frequencies also only include rare variants. Clinical refers to the combined data of LMM and OMGL, other studies refer to PubMed IDs. P-values shown are from Fisher's exact test comparing ExAC and HCM cohorts. * = cases were genotype negative for mutations in known sarcomeric HCM genes. # = only exons 2 and 3 of CSRP3 sequenced in this cohort.

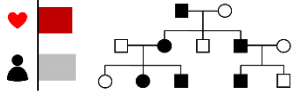
Genes with strong evidence		
Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
CSRP3 (Z-disc): ExAC freq = 0.00324		
12642359	3/200 (0.01500, p=0.029)	c.131T>C/p.L44P - present in 2 affected siblings (and 1 not classifiable). [ExAC AC=0] c.160_164delTCGGAGinsAGGGG/p.S54R,p.E55G - present in 3 affected siblings (father died of SCD). [ExAC AC=0] c.172T>G/p.C58G - present in 3 affected family members (and 1 not classifiable), absent in 3 unaffected (and 1 not classifiable). [ExAC AC=0]
16352453	5/389 (0.01285, p=0.010)	Of these five variants, three are present in patients with unspecified myofilament variants. c.272A>T/p.Q91L - [ExAC AC=1, freq=0.000008] p.K42fs/165 - [ExAC AC=0] c.131T>C/p.L44P - [ExAC AC=0] c.190C>T/p.R64C - [ExAC AC=3, freq=0.000024] c.197A>G/p.Y66C - [ExAC AC=1, freq=0.000008]
18505755	2/352# (0.00568, p=0.319)	c.172T>G/p.C58G - extended pedigree from REF:12642359, present in 8 affected and 3 unaffected, absent in 20 unaffected (approximate LOD = 3.6). Not part of sequenced cohort. [ExAC AC=0] c.136A>C/p.S46R - identified in two probands and segregates in one family, present in 6 affected and 1 unaffected, absent in 6 unaffected. [ExAC AC=4, freq=0.000033] Functional data: MLP protein level reduced by 60% in myocardium of patient with p.C58G variant.
20087448	-	c.50insGCAGATTCTT - present in a family which is double heterozygous with MYH7:p.R633H, cases with both variants are more severely affected than those with one. [ExAC AC=0]
22429680	3/80 (0.03750, p=0.002)	c.128delC/p.A43Vfs165 - [ExAC AC=0] c.170A>C/p.Y57S - [ExAC AC=0] c.131T>C/p.L44P - [ExAC AC=0]
25351510	14/874 (0.01602, p<0.0001)	
Clinical	11/2167 (0.00508, p=0.176)	
This study	6/804 (0.00746, p=0.052)	
Total	44/4866 (0.00904, p<0.0001)	

FHL1 (desmosome): ExAC freq = 0.00170		
Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
22523091	3/121* (0.02479, p=0.001)	c.134delA - detected in single male patient and absent from parents, suggesting <i>de novo</i> mutation. [ExAC AC=0] c.459C>A/p.C153X - present in 2 affected bothers (diagnosed with HCM at 10 and 13) with associated muscle weakness and in sister and mother, both diagnosed with HCM by echocardiography. [ExAC AC=0] c.827G>C/p.C276S - detected in HCM patient diagnosed during adolescence and in unaffected mother. [ExAC AC=0] Functional data: c.134delA and c.827G>C caused reduced <i>FHL1</i> protein levels in C2C12 myoblasts and cardiac myocytes. Altered contractility was observed after transduction into rat engineered heart tissue.
22923418	-	c.672C>G/p.C224W - detected in 12 males and 19 females in 5 different families. [ExAC AC=0] c.736C>T/p.H246Y - detected in 1 affected male and carrier mother. [ExAC AC=0] p.V280M - detected in 2 affected males and carrier mother. [ExAC AC=0] p.A168GfsX195 - detected in 2 affected males and carrier mother. [ExAC AC=0]
23500067	-	c.592C>T/p.Q198X - present in two brothers with unexplained LV hypertrophy and Achilles tendon contractures and in their asymptomatic mother, with history of cardiomyopathy and sudden death in family. Variant is present in exon 6 so doesn't affect isoform 1C. [ExAC AC=0]
23965743	-	c.632delA - detected in male patient with HCM, diffuse muscle pain and muscular hypertrophy. Variant is present in exon 6 so doesn't affect isoform 1C. [ExAC AC=0]
24114807	-	c.599dupT - present in 10 members of a 4 generation family, including 3 affected males and 1 affected female, 5 members with abnormal electrocardiogram and 1 unaffected 2-year old male (approximate LOD = 1.2). Variant is present in exon 6 so doesn't affect isoform 1C. [ExAC AC=0]
25246303	-	c.764G>C/p.C255S - detected in male patient (14) with HCM and X-linked distal myopathy, as well as deceased brother (sudden death at 10 with HCM detected by post-mortem), mother and aunt (maternal uncle also died suddenly with HCM detected by post-mortem but genotype status is unknown). [ExAC AC=0]
Clinical	14/1535 (0.00912, p<0.0001)	
This study	2/405 (0.00494, p=0.092)	
Total	19/2061 (0.00922, p<0.0001)	
PLN (calcium signalling) ExAC freq = 0.00048		
12639993	-	c.116T<G/p.L39X - identified in two unrelated Greek pedigrees - two homozygous individuals developed DCM and heart failure, in 11 heterozygous individuals, 2 had DCM, 4 had left ventricular hypertrophy and 5 had normal echocardiograms. [ExAC AC=1, freq=0.000008] Functional data: In HEK293 cells and adult rat cardiomyocytes, expression of this variant resulted in no PLN inhibition of sarcoplasmic reticulum calcium ATPase and no expression of stable PLN.
12705874	0/87 (0.00000, no excess)	-
16382369	0/53 (0.00000, no excess)	-
16829191	0/101 (0.00000, no excess)	-
17655857	1/252 (0.00397, p=0.117)	c.116T<G/p.L39X - present in proband (F, 61, Greek) and absent in unaffected offspring. [ExAC AC=1, freq=0.000008]
18409188	0/38 (0.00000, no excess)	-

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
21167350	1/1064 (0.00094, p=0.408)	c.116T<G/p.L39X - 3 members of a 3 generation family, proband (M, 51), unaffected daughter (38), affected granddaughter (3) and absent in 3 unaffected. [ExAC AC=1, freq=0.000008]
25351510	7/874 (0.00801, p<0.0001)	
Clinical	7/2167 (0.00323, p=0.0002)	
This study	2/804 (0.00249, p=0.062)	
Total	18/5440 (0.00331, p<0.0001)	

Genes with moderate evidence

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
ACTN2 (Z-disc) ExAC freq = 0.01086		
17097056	2/239* (0.00837, no excess)	p.G111V - [ExAC AC=0] p.R759T - [ExAC AC=0] c.1484C>T/p.T495M - [ExAC AC=26, freq=0.000219]
20022194	3/297 (0.01010, no excess)	c.355G>A/p.A119T - present in 23 members of a 3-generation family, including 7 affected (LOD = 2.82). Gene was identified through linkage study and prioritised for sequencing from 38 genes in locus. Not part of sequenced cohort. [ExAC AC=0] c.1883A>G/p.E628G - present in proband (F, 44) and two sons (16 - borderline LVH, 13 - normal). [ExAC AC=0] c.1748A>C/p.E583A - [ExAC AC=1, freq=0.000008] c.983G>A/p.R328Q did not segregate in family - [ExAC AC=1, freq=0.000008] c.1484C>T/p.T495M - [ExAC AC=26, freq=0.000219] c.1930G>A/p.A644T - [ExAC AC=28, freq=0.000232]
25173926	-	c.683T>C/p.M228T - present in 18 members of a 4-generation family, including 11 affected (5 with LVH) (approximate LOD = 2.1) Identified from large panel sequencing. [ExAC AC=0]
Clinical	10/1439 (0.00695, no excess)	
This study	16/804 (0.01990, p=0.024)	
Total	31/2779 (0.01116, p=0.851)	
CRYAB (cytoskeletal) ExAC freq = 0.00208		
23197161	-	p.Q26X - identified as a <i>de novo</i> variant in a paediatric HCM patient who went on to die of SCD at 17. [ExAC AC=0]
Clinical	-	
This study	2/804 (0.00249, p=0.684)	
Total	2/804 (0.00249, p=0.684)	

FLNC (Z-disc) ExAC freq = 0.03211		
Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
28356264	22/448 (0.04911, p=0.059)	p.A1539T - present in 4 affected (and 3 obligate carriers) and absent in 5 affected members of family CAR2 (LOD = 2.33), present in 4 affected and 2 unaffected and absent in 2 unaffected members of family CAR304. [ExAC AC=0] Functional data: Several of the identified variants resulted in the formation of large protein aggregates when expressed in H9C2 rat cardiac myoblasts and rat neonatal cardiac myocytes.
Clinical	-	
This study	-	
Total	22/448 (0.04911, p=0.059)	

MYO2 (Z-disc) ExAC freq = 0.00256		
17347475	0/516 (0.00000, no excess)	c.142T>C/p.S48P - Gene was identified through linkage study and prioritised for sequencing from 30 genes in locus based on cardiac expression and function. Variant detected was present in 6 affected and absent in 4 unaffected members of a 3 generation family (LOD=2.03). [ExAC AC=0] c.738A>G/p.I246M - identified by screening 516 HCM probands and absent in 2 unaffected family members. However, the variant is present at frequency of 0.0019 in ExAC (0.025 in Finnish). [ExAC AC=230, freq=0.001897]
22987565	-	Functional data: Mutant mice (variants p.S48P and p.I246M) showed molecular, cellular, and gross cardiac hypertrophy, preserved systolic function, interstitial fibrosis and mal-aligned Z disks.
18591919	0/438 (0.00000, no excess)	
Clinical	1/632 (0.00158, no excess)	
This study	1/804 (0.00124, no excess)	
Total	2/2390 (0.00084, no excess)	

Genes with weak evidence		
Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
MYH6 (sarcomeric) ExAC freq = 0.02466		
11815426	1/31 (0.03226, p=0.539)	c.2384G>A/p.R795Q - [ExAC AC=4, freq=0.000033]
15998695	-	c.3195G>C/p.Q1065H identified in 1/21 HCM probands and absent in two unaffected offspring. [ExAC AC=28, freq=0.000231]
22429680	-	c.166G>A/p.G56R identified in 1/80 HCM patients. [ExAC AC=8472, freq=0.069796]
25351510	37/874 (0.04233, p=0.002)	
Clinical	-	
This study	27/804 (0.03358, p=0.108)	
Total	65/1709 (0.03803, p=0.001)	

TNNC1 (sarcomeric) ExAC freq = 0.00060		
Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
11385718	-	c.86T>A/p.L29Q - identified in 59 year old patient. [ExAC AC=0]
18572189	3/1025 (0.00293, p=0.027)	c.23C>T/p.A8V - [ExAC AC=0] c.251G>A/p.C84Y - [ExAC AC=0] c.402G>T/p.E134D - [ExAC AC=0] c.435C>A/p.D145E - [ExAC AC=28, freq=0.000232]
		Functional data: Some of the identified variants led to increased calcium sensitivity of force development and force recovery.
21262074	-	c.363dupG - identified in 19 year old male who died of SCD and segregated with HCM in the family, present in 5 affected and 2 unaffected individuals (approximate LOD = 1.2). [ExAC AC=0]
22815480	-	c.91G>T/p.A31S - likely <i>de novo</i> variant in a paediatric HCM patient with multiple episodes of ventricular fibrillation and aborted sudden cardiac death. [ExAC AC=0]
25351510	3/874 (0.00343, p=0.018)	-
26304555	-	Functional data: Knock-in mice with the p.A8V variant (previously detected in a HCM patient) displayed atrial enlargement, papillary muscle hypertrophy, and fibrosis, with intact cardiomyocytes showing reduced diastolic sarcomeric length, increased shortening, and prolonged calcium and contractile transients.
Clinical	0/632 (0.00000, no excess)	
This study	2/804 (0.00249, p=0.088)	
Total	8/3335 (0.00240, p=0.002)	
TRIM55 (Ubiquitin proteasome) ExAC freq = 0.00734		
24865491	14/594 (0.02357, p=0.0002)	14 rare variants detected in a 594 HCM patient cohort (c.1028A>G/p.K343W and c.1174C>A/p.392T are present at >0.0001 in ExAC, c.1568A>G/p.H523W and c.1614delC occur in alternate transcript, NM_033058). 7/14 patients also have sarcomeric mutations.
		Functional data: Patients with rare variants in <i>TRIM55/MURF2</i> and <i>TRIM63/MURF1</i> (as defined in this report) had greater maximum left ventricular wall thickness than patients without (23.8±5.2 vs 21.5±4.7 mm, p=0.006).
Clinical	-	
This study	6/399 (0.01504, p=0.077)	
Total	20/993 (0.02014, p<0.0001)	
TRIM63 (Ubiquitin proteasome) ExAC freq = 0.00514		
24865491	12/594 (0.02020, p<0.0001)	12 rare variants detected in 594 HCM patient cohort (c.378A>T/p.E126D is present at frequency of 0.002 in ExAC East Asians). 8/12 patients also have sarcomeric mutations.
		Functional data: Patients with rare variants in <i>TRIM55/MURF2</i> and <i>TRIM63/MURF1</i> (as defined in this report) had greater maximum left ventricular wall thickness than patients without (23.8±5.2 vs 21.5±4.7 mm, p=0.006).
Clinical	-	
This study	8/804 (0.00995, p=0.075)	
Total	20/1398 (0.01431, p<0.0001)	

Genes with functional data only

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
ANKRD1 (Z-disc): ExAC freq = 0.00354		
19608031	3/384 (0.00781, p=0.159)	c.423C>T/p.Q59X - detected in 2 patients but did not segregate in families. [ExAC AC=0] c.838A>G/p.I280V - [ExAC AC=4, freq=0.000033] c.154C>G/P52A - [ExAC AC=13, freq=0.000107] c.368C>T/p.T123M - [ExAC AC=28, freq=0.000232] Functional data: In COS-7 cells, three missense variants (p.P52A, p.T123M, p.I280V) increased binding to titin and myopalladin. In mature cardiomyocytes, the variants led to abnormal intracellular accumulation of CARP proteins.
23572067	-	Functional data: In engineered heart tissues generated from rat heart cells, p.T123M caused higher force and velocities of contraction and relaxation than wild type. p.P52A and p.I280V were highly unstable, not incorporated into the sarcomere and had no effect on contraction. They were incorporated upon epoxomicin treatment with p.I280V demonstrating prolonged relaxation.
Clinical	4/807 (0.00496, p=0.5409)	
This study	3/804 (0.00373, p=0.764)	
Total	10/1995 (0.00501, p=0.252)	
CAV3 (Calcium signalling): ExAC freq = 0.00138		
14672715	1/146* (0.00685, p=0.184)	c.191C>G/p.T64S - present in 2 affected siblings and absent in unaffected mother (affected father died at 41 and is an obligate carrier) (approximate LOD = 0.6). [ExAC AC=0] Functional data: In transfected COS7 and NIH3T3 cells, the variant reduced the cell surface expression of caveolin-3.
14645200	-	Functional data: Transgenic mice expressing the CAV3 p.P104L variant (a model of autosomal dominant limb-girdle muscular dystrophy) developed HCM, enhanced basal contractility, decreased left ventricular end diastolic diameter, and loss and cytoplasmic mislocalisation of caveolin-3 protein.
25351510	0/874 (0.00000, no excess)	-
Clinical	-	
This study	0/804 (0.00000, no excess)	
Total	1/1824 (0.00055, no excess)	
FHL2 (Cytoskeletal): ExAC freq = 0.00310		
25358972	1/121* (0.00826, p=0.315)	c.512C>T/p.T171M - found in the index patient and her affected son, though her husband (also affected by HCM) was not a carrier. [ExAC AC=5, freq=0.000063] c.559G>T/p.V187L - [ExAC AC=24, freq=0.000231] c.530G>A/p.R177Q - [ExAC AC=484, freq=0.005389] Functional data: Variants do not affect the anti-hypertrophic activity of <i>FHL2</i> but may induce hypercontractility in engineered heart tissues.
Clinical	-	
This study	2/399 (0.00501, p=0.354)	
Total	3/520 (0.00577, p=0.223)	

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
FXN (Cytoskeletal): ExAC freq = 0.00132		
15936968	1/389 (0.00257, p=0.407)	c.118C>T/p.R40C was detected in a patient with a heterozygous MYBPC3:p.R810H variant (previously associated with HCM with homozygosity or compound heterozygosity). The patient had no family history of HCM or SCD, was diagnosed in childhood and had an arrhythmogenic form of HCM with a cardiac phenotype similar to Friedrich ataxia. [ExAC AC=0] Functional data: The variant alters a mitochondrial processing peptide (MPP) cleavage site that prevents the production of the mature form of frataxin. Yeast cells expressing the variant showed increased sensitivity to oxidative stress with an accumulation of the precursor protein.
Clinical	-	
This study	1/804 (0.00124, no excess)	
Total	2/1193 (0.00168, p=0.670)	
JPH2 (Calcium signalling): ExAC freq = 0.00654		
17509612	3/388* (0.00773, p=0.744)	p.S101R - [ExAC AC=0] c.421T>C/p.Y141H - [ExAC AC=1, freq=0.000017] p.S165F - [ExAC AC=0] Functional data: In muscle cell lines, each variant caused protein reorganisation of JPH2, perturbations in intracellular calcium signalling and marked cardiomyocyte hyperplasia.
17476457	0/296 (0.00000, no excess)	c.1513G>A/p.G505S - identified in 4/296 HCM patients and absent in 472 controls, but is present at frequency of 0.015 in 1000 Genomes (poor coverage at this position in ExAC).
22389502	-	Functional data: In primary mouse skeletal myotubes p.Y141H caused hypertrophy and altered calcium signalling through abnormal functional membrane complexes and increased store operated calcium entry.
23973696	2/203 (0.00985, p=0.384)	c.505G>A/p.E169K - identified in a patient with HCM and juvenile-onset paroxysmal AF. [ExAC AC=0] Functional data: Pseudoknock-in mouse models expressing the variant demonstrated a higher incidence of inducible AF than wild type or mice with c.1213G>T/p.A405S and showed an increased incidence of abnormal SR Ca(2+) release events.
Clinical	-	
This study	4/405 (0.00988, p=0.343)	
Total	9/1292 (0.00697, p=0.727)	

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
KLF10 (Transcription regulation): ExAC freq = 0.00486		
22234868	5/923* (0.00542, p=0.808)	c.34G>A/p.A12T - [ExAC AC=0] c.80T>C/p.M27T - [ExAC AC=10, freq=0.000082] c.409G>A/p.E137K - [ExAC AC=2, freq=0.000017] c.610G>A/p.A204T - [ExAC AC=3, freq=0.000025] c.646A>G/p.T216A - [ExAC AC=6, freq=0.000049] c.674G>A/p.S225N - [ExAC AC=74, freq=0.000610] Functional data: p.A12T and p.S225N reversed <i>KLF10</i> repression of <i>SMAD7</i> promoter activity in H9C2 cardiomyocytes (though the latter is a common variant in Finns, ExAC Finnish frequency of 0.01). All variants identified in the cohort reversed <i>KLF10</i> repression of <i>PTTG1</i> promoter activity in AKR2B fibroblasts, though a common variant (p.S249F with a frequency of 0.115 in East Asians in ExAC) also displayed this effect.
Clinical	-	
This study	4/405 (0.00988, p=0.138)	
Total	9/1328 (0.00678, p=0.315)	
LDB3 (Z-disc): ExAC freq = 0.01122		
17097056	3/239* (0.01255, p=0.752)	c.1075G>A/p.D359N - [ExAC AC=7, freq=0.000070] c.1382A>C/p.Y461S - [ExAC AC=2, freq=0.000017] c.1823C>T/p.P608L - [ExAC AC=11, freq=0.000091] c.566C>T/p.S189L - detected in two patients [ExAC AC=68, freq=0.000571] c.1535A>C/p.Q512P - [ExAC AC=28, freq=0.000232]
24730657	-	c.160G>A/p.G54S - identified in a HCM patient from a 50 gene screen. [ExAC AC=5, freq=0.000041] Functional data: Using molecular dynamics simulations, this variant is predicted to affect the binding of <i>LDB3/ZASP</i> to alpha-actinin 2 and calcium voltage channels.
25351510	2/874 (0.00229, no excess)	
Clinical	-	
This study	6/405 (0.01481, p=0.470)	
Total	11/1518 (0.00725, no excess)	
MYLK2 (Enzyme): ExAC freq = 0.00508		
11733062	0/490 (0.00000, no excess)	Two MYLK2 variants - c.260C>T/p.A87V [ExAC AC=16, freq=0.000137] and c.284C>A/p.A95E [ExAC AC=91, freq=0.000785] were identified on the maternal haplotype of a 13 year old white male with early midventricular HCM who had also inherited <i>MYH7</i>:p.E743D from his father. Both parents displayed asymmetrically thickened hearts with echocardiography but were only examined after their son's diagnosis, suggesting that the <i>MYLK2</i> variants exacerbated the phenotype caused by the sarcomeric mutation.
25825456	-	c.970A>G/p.K324E - detected in a complex pedigree with overlapping HCM and LQT phenotypes (along with <i>KCNQ1</i> and <i>MYH7</i> variants). HCM was present in individuals with both or either <i>MYLK2</i> and <i>MYH7</i> variants, with the data suggesting the <i>MYLK2</i> variant might be associated with abnormal cardiac repolarisation in HCM. [ExAC AC=0]
Clinical	-	
This study	10/804 (0.01244, p=0.153)	
Total	10/1294 (0.00773, no excess)	

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
MYOM1 (M-band): ExAC freq = 0.02394		
21256114	1/188 (0.00532, no excess)	c.4468G>A/p.V1490I - present in 3 affected (aged 51-75) and 3 unaffected (aged 23-26) and absent in 5 unaffected (aged 18-52) (approximate LOD = 0.6). [ExAC AC=2, freq=0.000076] Functional data: Variant caused reduced dimerisation of C-terminal myomesin domains.
Clinical	-	 
This study	-	
Total	1/188 (0.00532, no excess)	
MYPN (Z-disc): ExAC freq = 0.01434		
20801532	0/93 (0.00000, no excess)	c.3335C>T/p.P1112L - detected in a HCM patient with a pathogenic MYH7 mutation (p.V606M) and his unaffected mother. [ExAC AC=368, freq=0.003034]
22286171	9/484 (0.01860, p=0.438)	c.458A>G/p.K153R - [ExAC AC=0] c.650C>A/p.A217E - [ExAC AC=0] c.3793G>C/p.A1265P - [ExAC AC=0] c.1229T>C/p.V410A - [ExAC AC=5, freq=0.000041] c.2521C>A/p.P841T - [ExAC AC=0] c.2653C>T/p.R885X - [ExAC AC=1, freq=0.000008] c.1840G>A/p.E614K - present in 3 HCM patients and 3 controls, included here as ExAC frequency is less than 0.0001 [ExAC AC=9, freq=0.000074] c.59A>G/p.Y20C - [ExAC AC=111, freq=0.000916] c.3481C>A/p.L1161I - [ExAC AC=400, freq=0.003296] c.3335C>T/p.P1112L - [ExAC AC=368, freq=0.003034] Functional data: Cardiac-restricted p.Y20C transgenic mice developed HCM and showed other functional effects but this is a common variant (in 111 individuals in ExAC).
Clinical	-	 
This study	12/804 (0.01493, p=0.880)	
Total	21/1381 (0.01521, p=0.731)	
NEXN (Z-disc): ExAC freq = 0.00756		
20970104	1/121* (0.00826, p=0.601)	c.391C>G/p.Q131E - present in 3 affected members and absent in 3 unaffected members of a 3 generation family (approximate LOD = 0.9). [ExAC AC=1, freq=0.000008] c.835C>T/p.R279C - present in 3 affected members and absent in 3 unaffected members of a 3 generation family. [ExAC AC=71, freq=0.000595] Functional data: p.Q131E leads to local accumulations of nexilin in the cytoplasm and decreases the binding of nexilin to alpha-actinin.
Clinical	8/632 (0.01266, p=0.1584)	 
This study	10/804 (0.01244, p=0.146)	
Total	19/1557 (0.01220, p=0.053)	

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
TCAP (Calcium signalling): ExAC freq = 0.00242		
15582318	1/346* (0.00289, p=0.570)	c.410C>T/p.T137I - detected in a female HCM patient and her affected son. [ExAC AC=0] c.458G>A/p.R153H - detected in a male HCM patient, his affected sister and his two unaffected sons (aged 19 and 21). [ExAC AC=20, freq=0.000168] Functional data: With Y2H and GST pull-down assays, the variants detected in HCM patients augment the ability of <i>TCAP</i> to bind with titin and calsarcin-1 (but not MLP) (DCM associated variants impaired these interactions).
16352453	2/389* (0.00514, p=0.246)	Two rare variants detected in this HCM cohort - c.208C>T/p.R70W [ExAC AC=2, freq=0.000021] and c.269C>T/p.P90L (the patient also had the <i>MYBPC3</i> :p.Q998R variant) [ExAC AC=0]. The p.E13del variant detected in two patients is equivalent to c.33_35delGGA/p.E12del, a common variant [ExAC AC=115, freq=0.000952].
25351510	3/874 (0.00343, p=0.476)	-
Clinical	-	
This study	1/804 (0.00124, no excess)	
Total	7/2413 (0.00290, p=0.532)	

Genes with no evidence

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
CALR3 (Calcium signalling): ExAC freq = 0.00456		
17655857	1/252 (0.00397, no excess)	c.218G>A/p.R73Q - patient also had two MYBPC3 missense variants. [ExAC AC=4, freq=0.000032] c.245A>G/p.K82R - [ExAC AC=69, freq=0.000568]
Clinical	-	 
This study	0/804 (0.00000, no excess)	
Total	1/1056 (0.00095, no excess)	
CASQ2 (Calcium signalling): ExAC freq = 0.00508		
17655857	1/252 (0.00397, no excess)	p.D63E - patient also two MYBPC3 variants. [ExAC AC=0]
25351510	6/874 (0.00686, p=0.464)	-
Clinical	-	 
This study	3/804 (0.00373, no excess)	
Total	10/1930 (0.00518, p=0.870)	
LMNA: ExAC freq = 0.00624		
23785128	-	c.1804G>A/p.G602S - [ExAC AC=29, freq=0.000262]
25351510	5/874 (0.00572, no excess)	-
Clinical	-	 
This study	8/804 (0.00995, p=0.174)	
Total	13/1678 (0.00775, p=0.429)	

Study	Cohort frequencies & burden testing	Details of variants, patients/pedigrees and functional validation
OBSCN: ExAC freq = 0.11066		
17716621	-	c.13031G>A/p.R4344Q - [ExAC AC=1005, freq=0.011649] c.13450G>A/p.A4484T - [ExAC AC=125, freq=0.002637]
Clinical	-	
This study	-	
Total	-	
PDLIM3 (Z-disc): ExAC freq = 0.00506		
20801532	1/93* (0.01075, p=0.377)	c.317A>C/p.E106A - identified in a patient genotype-negative for seven HCM genes. [ExAC AC=5, freq=0.000042]
25351510	3/874 (0.00343, no excess)	-
Clinical	-	
This study	7/804 (0.00871, p=0.136)	
Total	11/1771 (0.00621, p=0.495)	
SRI (Calcium signalling): ExAC freq = 0.00214		
16118855	-	c.334T>C/p.F112L - [ExAC AC=72, freq=0.000593]
17655857	0/252 (0.00000, no excess)	-
Clinical	-	
This study	-	
Total	0/252 (0.00000, no excess)	
TRIM54 (Ubiquitin proteasome): ExAC freq = 0.00510		
24865491	12/594 (0.02020, p<0.0001)	-
Clinical	-	
This study	1/399 (0.00251, no excess)	
Total	13/993 (0.01309, p=0.002)	
VCL (Z-disc): ExAC freq = 0.01016		
16236538	0/389* (0.00000, no excess)	c.2923C>T/p.R975W - [ExAC AC=16, freq=0.000132]
		Functional data: Vinculin levels in intercalated discs was markedly reduced using immunohistochemical analysis of myectomy specimen.
16712796	1/228* (0.00439, no excess)	c.829C>A/p.L277M - [ExAC AC=5, freq=0.000041]
		Functional data: Vinculin levels in intercalated discs was markedly reduced using immunohistochemical analysis of myectomy specimen.
16949038	-	Functional data: Vinculin levels in intercalated discs, but not Z-discs, are reduced in 10 patients with obstructive HCM.
23785128	1/63 (0.01587, p=0.475)	c.1543G>A/p.G515S - [ExAC AC=0]
25351510	11/874 (0.01259, p=0.494)	-
Clinical	-	
This study	8/804 (0.00995, no excess)	
Total	21/2358 (0.00891, no excess)	

7.2 List of variants detected in prospective HCM cohort

The table below lists the variants detected in 684 HCM probands of the prospective HCM cohort in this study. Only variants in validated HCM genes and of classes that are potentially pathogenic are displayed. **ExAC FAF** = filtering allele frequency in ExAC. **Classification:** * = variants upgraded by the novel EF-based rules described in Chapter 3. **ACMG rules activated:** (p) = supporting, (m) = moderate, (s) = strong.

Gene	Variant	Protein	Variant Type	ExAC FAF	Classification	ACMG rules activated	Cases	Source for curated evidence
ACTC1	c.76G>A	p.D26N	missense	0	VUS	PM2,PP3,PM1(p)	1	
ACTC1	c.787C>G	p.L263V	missense	0	VUS	PM2,PM1(p)	1	
ACTC1	c.850A>T	p.I284F	missense	0	VUS	PM2,PM1(p)	2	
CRYAB	c.295G>T	p.E99X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
CSRP3	c.131T>C	p.L44P	missense	0	Likely Pathogenic	PM2,PS4,PP3	2	
CSRP3	c.449G>A	p.C150Y	missense	0	Likely Pathogenic	PM2,PS4,PP3	1	
FHL1	c.142G>A	p.G48S	missense	0	VUS	PM2,PM1(p)	1	
FHL1	c.764G>A	p.C255Y	missense	0	VUS	PM2,PP3,PM1(p)	1	
MYBPC3	c.126G>A	p.W42X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.188G>A	p.R63Q	missense	0	VUS	PM2	1	
MYBPC3	c.327_332delCCCTGC	-	inframe	0	VUS	PM2,PM4	1	
MYBPC3	c.442G>A	p.G148R	missense	0	VUS	PM2	2	
MYBPC3	c.451G>A	p.D151N	missense	0	VUS	PM2	1	
MYBPC3	c.484C>T	p.Q162X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.655G>C	p.V219L	missense	0	Likely Pathogenic	PM2,PS4,PP3,PM1(m)	2	
MYBPC3	c.710A>C	p.Y237S	missense	0	Likely Pathogenic	PM2,PS4,PP3,PM1(m)	1	
MYBPC3	c.767T>G	p.V256G	missense	0	VUS	PM2,PP3,PM1(m)	1	
MYBPC3	c.772G>A	p.E258K	missense	1.80E-05	Likely Pathogenic	PM2,PS4,PP1(p),PM1(m)	2	PMID:9562578
MYBPC3	c.806C>A	p.S269X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.821+2T>C	-	splice donor	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1084A>C	p.S362R	missense	0	VUS	PM2	1	
MYBPC3	c.1120C>T	p.Q374X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1201C>T	p.Q401X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	

Gene	Variant	Protein	Variant Type	ExAC FAF	Classification	ACMG rules activated	Cases	Source for curated evidence
MYBPC3	c.1217G>A	p.S406N	missense	0	VUS	PM2,PP3	1	
MYBPC3	c.1218C>A	p.S406R	missense	0	VUS	PM2	1	
MYBPC3	c.1227-2A>G	-	splice acceptor	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1351+1G>A	-	splice donor	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1357_1358delCC	-	frameshift	0	Pathogenic	PM2,PVS1,PP1(s)	1	PMID:12788380
MYBPC3	c.1359delT	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1405C>T	p.Q469X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1408C>T	p.R470W	missense	0	VUS	PM2,PP3,PM1(m)	1	
MYBPC3	c.1457+5G>C	-	splice site	0	VUS	PM2	1	
MYBPC3	c.1471G>A	p.V491M	missense	3.90E-05	VUS	PM2,PP3,PM1(m)	1	
MYBPC3	c.1483C>G	p.R495G	missense	0	Pathogenic	PM2,PS4,PP3,PM5,PM1(m)	4	
MYBPC3	c.1484G>A	p.R495Q	missense	0	Likely Pathogenic	PM2,PS4,PP1(p),PM1(m)	1	PMID:9562578
MYBPC3	c.1503C>G	p.Y501X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.1504C>T	p.R502W	missense	5.00E-06	Pathogenic	PM2,PS4,PP3,PM5,PM1(m)	12	
MYBPC3	c.1505G>A	p.R502Q	missense	0	Pathogenic	PM2,PS4,PP3,PM5,PM1(m)	1	
MYBPC3	c.1518C>A	p.D506E	missense	0	VUS	PM2,PM1(m)	1	
MYBPC3	c.1624G>C	p.E542Q	missense	0	Likely Pathogenic	PM2,PS4,PP1(p),PM1(m)	4	ClinVar:SCV000256661.2.
MYBPC3	c.1624+4A>T	-	splice site	0	Likely Pathogenic	PM2,PS4	3	
MYBPC3	c.1765C>T	p.R589C	missense	0	VUS	PM2,PP3,PM1(m)	2	
MYBPC3	c.1778C>T	p.S593F	missense	0	VUS	PM2,PM1(m)	1	
MYBPC3	c.1828G>A	p.D610N	missense	0	VUS	PM2	1	
MYBPC3	c.1871A>G	p.N624S	missense	3.70E-05	VUS	PM2	1	
MYBPC3	c.1928-2A>G	-	splice acceptor	0	Pathogenic	PM2,PVS1,PS4,PP1(s)	1	PMID:7493026
MYBPC3	c.1955C>G	p.P652R	missense	0	VUS	PM2	1	
MYBPC3	c.2011G>A	p.V671I	missense	0	VUS	PM2	1	
MYBPC3	c.2030C>T	p.P677L	missense	0	VUS	PM2	1	
MYBPC3	c.2198G>A	p.R733H	missense	2.00E-05	VUS	PM2	1	
MYBPC3	c.2267delC	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2308G>A	p.D770N	missense	0	Likely Pathogenic	PM2,PS4,PP3,PM1(m)	2	
MYBPC3	c.2371C>T	p.Q791X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2373dupG	-	frameshift	0	Likely Pathogenic	PM2,PVS1	2	

Gene	Variant	Protein	Variant Type	ExAC FAF	Classification	ACMG rules activated	Cases	Source for curated evidence
MYBPC3	c.2376G>A	p.W792X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2429G>A	p.R810H	missense	2.00E-05	Likely Pathogenic	PM2,PS4,PP3,PM1(m)	3	
MYBPC3	c.2459G>A	p.R820Q	missense	0	Likely Pathogenic *	PM2,PP1(p),PP3,PM1(m)	2	PMID:12628722
MYBPC3	c.2543_2544delCG	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2545delG	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2558delG	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2690_2696del	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2737+2_2737+3delTG	-	splice donor	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2827C>T	p.R943X	nonsense	0	Pathogenic	PM2,PVS1,PS4	1	
MYBPC3	c.2860G>A	p.A954T	missense	0	VUS	PM2	1	
MYBPC3	c.2864_2865delCT	-	frameshift	0	Pathogenic	PM2,PVS1,PS4,PP1(s)	1	PMID:9562578
MYBPC3	c.2905+2T>G	-	splice donor	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.2905+1G>A	-	splice donor	0	Pathogenic	PM2,PVS1,PS4	1	
MYBPC3	c.2905C>T	p.Q969X	nonsense	0	Pathogenic	PM2,PVS1,PP1(s)	1	PMID:9541104
MYBPC3	c.3064C>T	p.R1022C	missense	0	VUS	PM2,PP3	1	
MYBPC3	c.3163A>T	p.K1055X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3190+1G>A	-	splice donor	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3190+5G>A	-	splice site	5.00E-06	Likely Pathogenic	PM2,PS4	1	
MYBPC3	c.3226_3227insT	-	frameshift	0	Pathogenic	PM2,PVS1,PS4	2	
MYBPC3	c.3233G>A	p.W1078X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3257G>A	p.W1086X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3286G>T	p.E1096X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3408C>A	p.Y1136X	nonsense	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3491-3C>G	-	splice site	0	VUS	PM2	1	
MYBPC3	c.3492delC	-	frameshift	0	Likely Pathogenic	PM2,PVS1	1	
MYBPC3	c.3613C>T	p.R1205W	missense	0	Likely Pathogenic	PM2,PS4,PM1(m)	1	
MYBPC3	c.3627+1G>A	-	splice donor	0	Pathogenic	PM2,PVS1,PS4	1	
MYBPC3	c.3697C>T	p.Q1233X	nonsense	0	Pathogenic	PM2,PVS1,PS4,PP1(p)	1	PMID:18957093
MYBPC3	c.3742G>A	p.G1248R	missense	2.00E-05	VUS	PM2,PM1(m)	1	
MYBPC3	c.3763G>C	p.A1255P	missense	0	VUS	PM2,PP3,PM1(m)	1	
MYBPC3	c.3771C>A	p.N1257K	missense	0	Likely Pathogenic	PM2,PS4,PP3,PM1(m)	1	

Gene	Variant	Protein	Variant Type	ExAC FAF	Classification	ACMG rules activated	Cases	Source for curated evidence
MYH7	c.428G>A	p.R143Q	missense	0	Likely Pathogenic	PM2,PS4,PP3	1	
MYH7	c.508G>A	p.E170K	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	2	
MYH7	c.788T>C	p.I263T	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.976G>C	p.A326P	missense	3.90E-05	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.1060G>A	p.G354S	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.1063G>A	p.A355T	missense	0	Pathogenic	PM2,PS4,PP3,PM1(s)	4	
MYH7	c.1075T>G	p.F359V	missense	0	Likely Pathogenic *	PM2,PM1(s)	1	
MYH7	c.1357C>T	p.R453C	missense	0	Pathogenic	PM2,PS4,PP1(s),PP3,PM1(s)	2	PMID:1552912
MYH7	c.1358G>A	p.R453H	missense	0	Likely Pathogenic *	PM2,PP3,PM5,PM1(s)	1	
MYH7	c.1405G>A	p.D469N	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.1447G>A	p.E483K	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.1464C>A	p.F488L	missense	0	Likely Pathogenic *	PM2,PM1(s)	1	
MYH7	c.1477A>T	p.M493L	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.1753A>T	p.I585F	missense	0	Likely Pathogenic *	PM2,PM1(s)	1	
MYH7	c.1988G>A	p.R663H	missense	5.00E-06	Pathogenic	PM2,PS4,PP1(s),PP3,PM1(s)	3	PMID:10750581
MYH7	c.2011C>T	p.R671C	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2146G>A	p.G716R	missense	0	Pathogenic	PM2,PS4,PP1(m),PS2,PP3,PM1(s)	1	PMID:18953637
MYH7	c.2167C>T	p.R723C	missense	5.00E-06	Pathogenic	PM2,PS4,PP1(m),PS2,PP3,PM5,PM1(s)	1	PMID:1430197
MYH7	c.2221G>T	p.G741W	missense	0	Pathogenic	PM2,PS4,PP1(p),PP3,PM1(s)	1	PMID:8533830
MYH7	c.2242T>A	p.S748T	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2243C>T	p.S748F	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2296A>C	p.K766Q	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2302G>A	p.G768R	missense	0	Pathogenic	PM2,PS4,PP3,PM1(s)	2	
MYH7	c.2389G>A	p.A797T	missense	3.40E-05	Pathogenic	PM2,PS4,PP1(s),PM1(s)	3	PMID:11186938
MYH7	c.2539A>G	p.K847E	missense	0	Pathogenic	PM2,PS4,PM1(s)	5	
MYH7	c.2555T>C	p.M852T	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2606G>A	p.R869H	missense	3.40E-05	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2606G>C	p.R869P	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.2609G>A	p.R870H	missense	0	Pathogenic	PM2,PS4,PP1(s),PP3,PM1(s)	1	PMID:16650083
MYH7	c.2717A>G	p.D906G	missense	0	Pathogenic	PM2,PS4,PP1(s),PP3,PM1(s)	5	PMID:12081993
MYH7	c.2722C>G	p.L908V	missense	0	Pathogenic	PM2,PS4,PP1(s),PP3,PM1(s)	1	PMID:1638703

Gene	Variant	Protein	Variant Type	ExAC FAF	Classification	ACMG rules activated	Cases	Source for curated evidence
MYH7	c.2770G>A	p.E924K	missense	0	Pathogenic	PM2,PS4,PS2,PP3,PM1(s)	1	PMID:1430197
MYH7	c.2782G>A	p.D928N	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
MYH7	c.3134G>T	p.R1045L	missense	5.00E-06	VUS	PM2,PP3	2	
MYH7	c.4066G>A	p.E1356K	missense	0	Likely Pathogenic	PM2,PS4,PP3	1	
MYH7	c.4124A>G	p.Y1375C	missense	0	VUS	PM2,PP3	1	
MYH7	c.4135G>A	p.A1379T	missense	0	Pathogenic	PM2,PS4,PP1(s)	1	PMID:11861413
MYH7	c.4258C>T	p.R1420W	missense	0	VUS	PM2,PP3	1	
MYH7	c.4259G>A	p.R1420Q	missense	0	VUS	PM2,PP3	1	
MYH7	c.4317_4319dupTGC	-	inframe	0	VUS	PM2,PM4	1	
MYH7	c.4954G>T	p.D1652Y	missense	1.20E-05	VUS	PM2,PP3	1	
MYH7	c.5075T>A	p.V1692E	missense	0	VUS	PM2	1	
MYH7	c.5134C>T	p.R1712W	missense	0	VUS	PM2,PP3	2	
MYH7	c.5135G>A	p.R1712Q	missense	0	Likely Pathogenic	PM2,PS4,PP3	2	
MYH7	c.5334C>A	p.H1778Q	missense	0	VUS	PM2,PP3	1	
MYH7	c.5341C>T	p.R1781C	missense	0	VUS	PM2,PP3	1	
MYH7	c.5380C>A	p.Q1794K	missense	0	VUS	PM2,PP3	1	
MYH7	c.5470A>G	p.N1824D	missense	0	VUS	PM2	1	
MYH7	c.5471A>G	p.N1824S	missense	1.20E-05	VUS	PM2	1	
MYH7	c.5507C>T	p.S1836L	missense	2.90E-05	VUS	PM2,PP3	1	
MYH7	c.5749G>T	p.V1917F	missense	0	VUS	PM2,PP3	1	
MYH7	c.5769delG	-	frameshift	0	Likely Pathogenic	PM2,PS4	1	
MYH7	c.5779A>T	p.I1927F	missense	3.00E-05	VUS	PM2	1	
MYH7	c.5786C>T	p.T1929M	missense	1.20E-05	VUS	PM2	1	
MYL2	c.221C>T	p.P74L	missense	0	VUS	PM2,PM1(p)	1	
MYL2	c.428C>T	p.P143L	missense	1.20E-05	VUS	PM2,PP3,PM1(p)	1	
MYL3	c.383G>A	p.G128D	missense	0	VUS	PM2,PM1(p)	1	
MYL3	c.445A>G	p.M149V	missense	0	Pathogenic	PM2,PP1(s),PM1(s)	1	PMID:8673105,16267253
MYL3	c.460C>T	p.R154C	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
PLN	c.74G>A	p.R25H	missense	0	VUS	PM2,PP3,PM1(p)	1	
PLN	c.116T>G	p.L39X	nonsense	0	Pathogenic	PM2,PVS1,PS4	1	
TNNI3	c.235C>T	p.R79C	missense	0.00407	VUS	PP3	1	

Gene	Variant	Protein	Variant Type	ExAC FAF	Classification	ACMG rules activated	Cases	Source for curated evidence
TNNI3	c.407G>A	p.R136Q	missense	0	Likely Pathogenic *	PM2,PM1(s)	1	
TNNI3	c.433C>T	p.R145W	missense	0	Pathogenic	PM2,PS4,PP1(p),PP3,PM5,PM1(s)	3	PMID:15607392
TNNI3	c.484C>T	p.R162W	missense	1.20E-05	Pathogenic	PM2,PS4,PM5,PM1(s)	2	
TNNI3	c.485G>A	p.R162Q	missense	5.00E-06	Pathogenic	PM2,PS4,PP1(m),PM1(s)	4	PMID:15607392,22876777
TNNI3	c.557G>A	p.R186Q	missense	0	Likely Pathogenic *	PM2,PM1(s)	1	
TNNI3	c.561G>T	p.E187D	missense	0	Likely Pathogenic *	PM2,PM1(s)	1	
TNNI3	c.587A>G	p.D196G	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
TNNI3	c.602T>C	p.M201T	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
TNNI3	c.608G>T	p.G203V	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	1	
TNNT2	c.236T>A	p.I79N	missense	0	Pathogenic	PM2,PS4,PP3,PM1(s)	2	
TNNT2	c.274C>T	p.R92W	missense	0	Pathogenic	PM2,PS4,PP1(m),PP3,PM5,PM1(s)	2	PMID:11346248,9060892
TNNT2	c.275G>A	p.R92Q	missense	0	Pathogenic	PM2,PM6,PP1(s),PP3,PM5,PM1(s)	1	PMID:8205619,7898523
TNNT2	c.664G>T	p.D222Y	missense	0	VUS	PM2,PP3	1	
TNNT2	c.785A>G	p.N262S	missense	0	VUS	PM2	2	
TNNT2	c.832C>T	p.R278C	missense	0.00044	VUS		5	
TNNT2	c.857G>A	p.R286H	missense	0.00040	VUS		1	
TPM1	c.62G>T	p.R21L	missense	0	Likely Pathogenic *	PM2,PP3,PM1(s)	2	
TPM1	c.842T>C	p.M281T	missense	0	Likely Pathogenic *	PM2,PM1(s)	2	

7.3 Permissions to re-use third-party copyrighted works

Figure 1.1

License Number	4231350909182
License date	Nov 17, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Reviews Cardiology
Licensed Content Title	Genetic testing for inherited cardiac disease
Licensed Content Author	Arthur A. M. Wilde, Elijah R. Behr
Licensed Content Date	Jul 30, 2013
Licensed Content Volume	10
Licensed Content Issue	10
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Figures	Figure 3
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Genetic, genotype and imaging studies of hypertrophic cardiomyopathy

Figure 1.2

License Number	4130810816200
License date	Jun 16, 2017
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Circulation
Licensed Content Title	Perspectives on the Overall Risks of Living With Hypertrophic Cardiomyopathy
Licensed Content Author	Barry J. Maron,Martin S. Maron,Ethan J. Rowin
Licensed Content Date	Jun 13, 2017
Licensed Content Volume	135
Licensed Content Issue	24
Type of Use	Dissertation/Thesis
Requestor type	Individual
Portion	Figures/table/illustration
Number of figures/tables /illustrations	1
Figures/tables /illustrations used	Figure
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	Genetic, genotype and imaging studies of hypertrophic cardiomyopathy

Figure 1.3

License Number	4230290656408
License date	Nov 15, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Reviews Cardiology
Licensed Content Title	Genetic testing for inherited cardiac disease
Licensed Content Author	Arthur A. M. Wilde, Elijah R. Behr
Licensed Content Date	Jul 30, 2013
Licensed Content Volume	10
Licensed Content Issue	10
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Figures	Figure 4
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Genetic, genotype and imaging studies of hypertrophic cardiomyopathy

Figure 1.4, Table 1.1

License Number	4246001484310
License date	Dec 11, 2017
Licensed content publisher	American Association for Clinical Chemistry Inc.
Licensed content title	Clinical chemistry
Licensed content date	Jan 1, 1955
Type of Use	Thesis/Dissertation
Requestor type	Academic institution
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs /tables/figures	2
The requesting person/organization is:	Roddy Walsh
Title or numeric reference of the portion(s)	Figure 1, Table 2.

Figure 2.2

License Number	4159410687678
License date	Jul 31, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Genetics in Medicine
Licensed Content Title	The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing
Licensed Content Author	Trevor J. Pugh, Melissa A. Kelly, Sivakumar Gowrisankar, Elizabeth Hynes, Michael A. Seidman, Samantha M. Baxter
Licensed Content Date	Feb 6, 2014
Licensed Content Volume	16
Licensed Content Issue	8
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Figures	Figure 1
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Genetic, genotype and imaging studies of hypertrophic cardiomyopathy

Figure 3.1

License Number	4170251241003
License date	Aug 15, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Genetics in Medicine
Licensed Content Title	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
Licensed Content Author	Sue Richards, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster
Licensed Content Date	Mar 5, 2015
Licensed Content Volume	17
Licensed Content Issue	5
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Figures	Figure 1.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Genetic, genotype and imaging studies of hypertrophic cardiomyopathy

Figure 2.11

License Number	4151381308767
License date	Jul 17, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	The American Journal of Human Genetics
Licensed Content Title	Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource
Licensed Content Author	Natasha T. Strande, Erin Rooney Riggs, Adam H. Buchanan, Ozge Ceyhan-Birsoy, Marina DiStefano, Selina S. Dwight, Jenny Goldstein, Rajarshi Ghosh, Bryce A. Seifert, Tam P. Sneddon, Matt W. Wright, Laura V. Milko, J. Michael Cherry, Monica A. Giovanni et al.
Licensed Content Date	Jun 1, 2017
Licensed Content Volume	100
Licensed Content Issue	6
Licensed Content Pages	12
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figures 2 and 3
Title of your thesis/dissertation	Genetic, genotype and imaging studies of hypertrophic cardiomyopathy