

DEFINING THE GENETIC ARCHITECTURE OF HYPERTROPHIC CARDIOMYOPATHY IN EGYPT

This dissertation is submitted to Imperial College London for the degree of Doctor
of Philosophy

by

Mona Allouba

MSc

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

June 2020

Declarations

Declaration of originality

I declare that this thesis is my own work and that it has not been submitted in any form to any university or institution for another degree, diploma or other qualification. Any contribution of others to the work presented here is duly acknowledged or appropriately 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.

This thesis is dedicated to my parents and Mohamed

You have made me everything I am today

Abstract

Hypertrophic Cardiomyopathy (HCM) is an inherited disease of the myocardium characterised by genetic and phenotypic heterogeneity. Despite the substantial progress made in understanding the genetic basis of HCM, several populations remain understudied, which hampers the progression of precision medicine and exacerbates health inequalities.

This research study represents the largest and most comprehensive analysis of HCM in the Middle East and North Africa (MENA) region to date: It includes an ethnically-matched cohort of deeply-phenotyped healthy volunteers (n=400), allowing the genetic architecture of HCM in prospective Egyptian patients enrolled at the Aswan Heart Centre (n=514) to be accurately defined against an appropriate control group. Analysis of genetic variation in well-characterised HCM genes revealed interesting findings. For example, it demonstrated a relatively higher burden of rare homozygous variants in Egyptian patients compared to other well-studied populations, owing to the issue of consanguinity. This opens new avenues for further characterising the effect of homozygosity in HCM-related genes on disease pathogenesis and for identifying novel recessive genes unique to consanguineous populations of MENA ancestry.

In addition, a highly prevalent frameshift variant (c.5769delG) in *MYH7* was identified in the Egyptian HCM cohort. According to the current clinical variant classification guidelines, c.5769delG would be considered a “variant of uncertain significance”. However, it was found to be significantly enriched in Egyptian patients compared to controls (3.31% vs. 0%, $p=0.0004$) and to co-segregate with HCM in a large family (LOD score: 3.01), thereby supporting disease-causality. RNA sequencing analysis confirmed the expression of the variant *MYH7* transcript suggesting a new mechanism of pathogenicity whereby distal *MYH7* truncations cause HCM by escaping nonsense-mediated decay.

Overall, these findings facilitate the implementation of precision medicine by enhancing our understanding of disease mechanisms involved in HCM and improving the utility of clinical genetic testing in the region.

Table of contents

Declarations	2
Declaration of originality	2
Copyright declaration	2
Abstract	4
Table of contents	6
List of Figures	12
List of Tables	15
Contribution of others to work presented in this thesis	17
Acknowledgements	19
List of nonstandard abbreviations and acronyms	21
Awards and publications	25
1 INTRODUCTION	27
1.1 Cardiovascular Diseases and Cardiomyopathy	27
1.2 Overview of Hypertrophic Cardiomyopathy (HCM)	28
1.2.1 Definition	28
1.2.2 Epidemiology	29
1.2.3 Aetiology	30
1.2.4 Diagnosis, prognosis and treatment options	31
1.3 The Genetic Architecture and Disease Mechanism of HCM	33
1.3.1 The genetic basis of HCM	33
1.3.2 Structure and function of the sarcomere	33
1.3.3 Molecular mechanisms of sarcomere dysfunction in HCM	35
1.3.4 Challenges in developing a complete understanding of HCM pathogenesis	36
1.3.5 Implications and challenges of clinical genetic testing for HCM	37
1.4 Next-Generation Sequencing (NGS) and its clinical implications for HCM	39
1.4.1 The evolution of sequencing technology	39
1.4.2 Applications of NGS	40
1.4.3 NGS data analysis and challenges	42
1.5 Limited knowledge of the genetic basis of HCM in Egypt	43

1.5.1	Characteristics of the Egyptian population	44
1.6	Thesis objectives	44
1.7	Plan of investigation	45
1.7.1	Objective 1: To define the genetic architecture of HCM in the Egyptian population	45
1.7.2	Objective 2: To examine the occurrence of genetic features that are unique to Egyptian HCM patients	45
1.7.3	Objective 3: To study the effect of consanguinity on homozygosity and disease severity	45
2	EVALUATING THE BACKGROUND GENETIC VARIATION IN HCM-ASSOCIATED GENES IN AN EGYPTIAN HEALTHY VOLUNTEER (EHVol) COHORT	46
2.1	Aims and hypotheses	46
2.2	Introduction	46
2.2.1	The impact of large-scale reference databases in understanding human genetic variation	46
2.2.2	Underrepresentation of the Middle East and North Africa (MENA) region in genomic reference databases	53
2.2.3	The Egyptian Collaborative Cardiac Genomics (ECCO-GEN) Project	55
2.3	Methods	57
2.3.1	Study population	57
2.3.2	Targeted sequencing	60
2.3.3	Bioinformatics analysis	61
2.3.4	Sequence variant nomenclature	62
2.3.5	Gene selection	62
2.3.6	Calculation of variant frequencies in the EHVol cohort	63
2.3.7	Representation of EHVol genetic variation in gnomAD	64
2.3.8	Rare variant filtering	64
2.3.9	Identification of variants commonly observed in the EHVol cohort with low prevalence in gnomAD	65
2.3.10	Comparison of rare variation between EHVol and CHVol control cohorts	65
2.3.11	Re-filtering of rare variants after combining EHVol genetic data with gnomAD	65
2.3.12	Comparison of unique variants between the EHVol and CHVol cohorts	65
2.4	Results	67
2.4.1	Demographic and clinical characteristics	67
2.4.2	Loss-of-Function (LoF) variants minimally contributed to genetic variation in the EHVol cohort	68
2.4.3	Absence of rare homozygous variants in the EHVol cohort	71
2.4.4	Absence of a subset of EHVol variants from the gnomAD dataset	72
2.4.5	Clinical evidence of Left Ventricular hypertrabeculation in an EHVol control with Myosin Heavy Chain 7 (<i>MYH7</i>) LoF variant	73
2.4.6	LoF Triadin-1 (<i>TRDN</i>) variant associated with Catecholaminergic Polymorphic Ventricular Tachycardia identified at high prevalence in the EHVol cohort	74
2.4.7	Higher frequency of rare variation in HCM genes in EHVol controls compared to CHVols	78

2.4.8	Combining EHVOL data with gnomAD results in a more accurate definition of variant rarity	81
2.4.9	Higher proportion of unique variants in the EHVOL cohort compared to CHVOL controls	82
2.5	Discussion	84
2.5.1	Combining EHVOL data with gnomAD results in a more accurate definition of variant rarity	84
2.5.2	Differences in the frequency of background genetic variation between EHVOL- and gnomAD controls	85
2.5.3	High rate of consanguinity in EHVOL controls	86
2.6	Conclusions	88
3	EVALUATION OF HCM RARE GENETIC VARIATION IN THE EGYPTIAN POPULATION	89
3.1	Aims and hypotheses	89
3.2	Introduction	90
3.2.1	Genetic and non-genetic associations with HCM	90
3.2.2	HCM variant interpretation: current guidelines, implications and challenges	95
3.2.3	Current knowledge of the genetics of HCM in Africa	102
3.3	Methods	105
3.3.1	Study population	105
3.3.2	Targeted sequencing and bioinformatics analysis	107
3.3.3	Gene selection	107
3.3.4	Variant filtering	108
3.3.5	Rare variation in the AhcHCM cohort compared to EHVOL controls and the predominantly Caucasian HCM cohorts	109
3.3.6	Burden testing	110
3.3.7	Sample size calculation for power analysis	111
3.3.8	Assessment of potential bias introduced by gnomAD to define rare variants in Egyptian population	112
3.3.9	Population-matched HCM case versus control comparison of rare variation	112
3.3.10	Distribution of Myosin Regulatory Light Chain 2 (<i>MYL2</i>) and Myosin Essential Light Chain 3 (<i>MYL3</i>) rare variants in the AhcHCM cohort	113
3.3.11	Classification and manual curation of rare variants	113
3.3.12	Comparison of the yield of clinically actionable variants between the AhcHCM- and RbhHCM cohorts	115
3.3.13	Application of the new Etiological Fraction (EF)-based American College of Medical Genetics and Genomics (ACMG)/Association or Molecular Pathology (AMP) rule on curated variants	115
3.3.14	Proportion of AhcHCM and RbhHCM patients with unique variants	116
3.3.15	Genotype-phenotype correlations in AhcHCM patients with distinct genetic aetiologies	116
3.4	Results	118
3.4.1	Cohort description	118

3.4.2	<i>MYH7</i> and Myosin Binding Protein C3 (<i>MYBPC3</i>) account for the majority of rare variation in the AhcHCM cohort.....	119
3.4.3	Key differences in the genetic architecture of HCM between Egyptian and well-characterised predominantly Caucasian cohorts.....	124
3.4.4	“Rare” variants defined using gnomAD introduced bias in comparing HCM genetic architecture between Egyptian and predominantly Caucasian patients.....	130
3.4.5	Significant enrichment of rare variation in Myosin Light Chain genes and <i>MYH7</i> truncating variants in AhcHCM patients compared to RbhHCM.....	132
3.4.6	Significantly higher proportion of Variants of Uncertain Significance (VUSs) in the AhcHCM cohort compared to RbhHCM	138
3.4.7	Application of the new EF-based ACMG/AMP rules increased the yield of clinically actionable variants identified in <i>MYH7</i> in the AhcHCM cohort	142
3.4.8	Higher proportion of AhcHCM patients with unique sarcomeric variants compared to RbhHCM patients	144
3.4.9	Higher burden of homozygous variants in AhcHCM patients compared to RbhHCM	145
3.4.10	A subset of AhcHCM patients presented with double rare variants in sarcomere genes	154
3.4.11	Identification of 4 HCM patients with novel homozygous variants in Cysteine and Glycine-Rich Protein 3 (<i>CSRP3</i>) increases the worldwide number of known <i>CSRP3</i> knockout HCM patients to 7	156
3.4.12	Higher degree of LVH in AhcHCM patients with clinically actionable variants in <i>MYBPC3</i> compared to <i>MYH7</i> -variant carriers.....	157
3.5	Discussion.....	159
3.5.1	Insights into the genetic landscape of HCM in Egypt	159
3.5.2	Enrichment of rare variation in <i>MYL2</i> and <i>MYL3</i> genes and <i>MYH7</i> truncating variants in Egyptian patients compared to predominantly Caucasian patients.....	161
3.5.3	Egyptian HCM patients are more likely to obtain inconclusive genetic results compared to predominantly Caucasian patients	164
3.5.4	A substantial proportion of Egyptian HCM patients did not present with a genetically identifiable aetiology	166
3.5.5	Homozygosity is enriched in AhcHCM patients compared to RbhHCM	168
3.5.6	Identification of 2 novel <i>CSRP3</i> biallelic truncating variants in Egyptian patients provides deeper insights into the mutational spectrum and disease mechanism of <i>CSRP3</i> variants	170
3.5.7	“Non-rare” variants in <i>MYL3</i> could become penetrant when expressed in homozygosity.....	171
3.5.8	Study limitations.....	171
3.6	Conclusions.....	173
4	ANALYSIS OF HCM IN THE EGYPTIAN POPULATION REVEALS A NEW MECHANISM OF PATHOGENICITY	174
4.1	Aims and hypotheses	174
4.2	Introduction	175
4.2.1	Overview of <i>MYH7</i>	175
4.2.2	Conformational changes in myosin facilitate contraction/relaxation	177
4.2.3	HCM-causing variants are enriched in specific regions of <i>MYH7</i>	180

4.2.4	Adapted ACMG/AMP guidelines for the interpretation of <i>MYH7</i> variants in HCM	183
4.2.5	Molecular mechanisms of <i>MYH7</i> variants in HCM	187
4.3	Methods	191
4.3.1	Burden testing of <i>MYH7</i> c.5769delG	191
4.3.2	Calculation of the population-level penetrance of <i>MYH7</i> c.5769delG	191
4.3.3	Genotyping and quality control	192
4.3.4	Relatedness analysis	192
4.3.5	Segregation analysis	193
4.3.6	Age-related penetrance for large AhcHCM family with <i>MYH7</i> fs variant	193
4.3.7	Comparison of disease severity between <i>MYH7</i> frameshift (<i>MYH7</i> fs) variant-carriers and other AhcHCM patients with different genetic aetiologies	194
4.3.8	RNA extraction and sequencing and analysis	194
4.3.9	Allele-specific expression analysis	195
4.3.10	Distribution of rare truncating variants in <i>MYH7</i>	195
4.4	Results	196
4.4.1	Enrichment of a single <i>MYH7</i> frameshift variant (c.5769delG) in the AhcHCM cohort	196
4.4.2	<i>MYH7</i> frameshift variant c.5769delG was classified as a VUS according to the ACMG/AMP guidelines	196
4.4.3	Demographic and clinical characteristics of AhcHCM patients with <i>MYH7</i> c.5769delG variant	196
4.4.4	Enrichment of <i>MYH7</i> c.5769delG variant in AhcHCM patients compared to ethnically-matched controls	199
4.4.5	Segregation analysis of a large AhcHCM family provided supporting evidence for the association of the <i>MYH7</i> fs variant c.5769delG with HCM	200
4.4.6	The <i>MYH7</i> fs variant (c.5769delG) is associated with a late HCM disease-onset compared to other AhcHCM patients with different genetic aetiologies	204
4.4.7	Genetic variability among <i>MYH7</i> fs c.5769delG variant-carriers	205
4.4.8	RNA sequencing analysis confirmed the expression of the variant <i>MYH7</i> transcript	207
4.4.9	Detailed analysis of <i>MYH7</i> fs variant in the Egyptian HCM cohort provided evidence for its re-classification from VUS to P	211
4.4.10	<i>MYH7</i> frameshift variant c.5769delG is more common in AhcHCM patients compared to both LMM/OMGL and RbhHCM patients	215
4.5	Discussion	216
4.5.1	Initial discovery of <i>MYH7</i> fs variant	216
4.5.2	Enrichment of <i>MYH7</i> fs variant in AhcHCM patients provides preliminary evidence for its association with HCM	218
4.5.3	Segregation analysis confirmed the association of <i>MYH7</i> fs with HCM	220
4.5.4	RNA analysis confirmed that <i>MYH7</i> fs is not a LoF variant	222
4.5.5	Detailed analysis of c.5769delG in Egyptian HCM patients provides valuable insights for the application of clinical genetic testing	223
4.5.6	Functional analyses for the elucidation of the molecular mechanism of the <i>MYH7</i> fs variant	224
4.6	Conclusions	227
5	CONCLUDING REMARKS AND FUTURE PERSPECTIVES	228

6	BIBLIOGRAPHY	233
7	APPENDICES	253
7.1	List of rare variants identified in AhcHCM cohort	253
7.2	Permissions to re-use figures/tables from publications	263

List of Figures

Figure 1-1: Age-standardised mortality rate (per 100,000 individuals) due to cardiovascular diseases.....	27
Figure 1-2: The anatomy of the normal heart compared to HCM..	29
Figure 1-3: Scheme of the cardiac sarcomere.....	34
Figure 1-4: Molecular mechanism of cardiac contraction.....	35
Figure 2-1: The 5 major populations analysed in the 1000 Genomes Project.	48
Figure 2-2: The different ancestral populations represented in ExAC.....	50
Figure 2-3: Representation of the size and ethnic composition of populations integrated into the different large-scale genetic reference databases.	52
Figure 2-4: Map of 19 countries spanning the MENA region.	54
Figure 2-5: Workflow of the EHVol study.	58
Figure 2-6: The different variant types included in the analysis of genetic variation in the EHVol cohort.....	64
Figure 2-7: Distribution of analysed EHVol controls (n=400) by city of origin	68
Figure 2-8: Synonymous variants accounted for >50% of genetic variation in the EHVol cohort.	69
Figure 2-9: <i>MYH7</i> and <i>MYBPC3</i> (sarcomere), <i>CACNA1C</i> (syndromic) and <i>ACTN2</i> accounted for the majority of genetic variation in the EHVol cohort.	70
Figure 2-10: Majority of EHVol-only (absent in gnomAD) variants were observed as singletons (AC bin=1) in the EHVol cohort.....	72
Figure 2-11: Long-axis LV CMR (A) 2-chamber (B) 4-chamber and (C) 3-chamber views showing non-compacted myocardium of the LV at the mid-ventricle and apical levels .	74
Figure 2-12: <i>TRDN</i> LoF variant (c.1807G>T) is predicted to elicit NMD.....	78
Figure 2-13: High frequency of rare variation in HCM-associated genes in the EHVol cohort.	80
Figure 3-1: Classes of genetic (top) and experimental (bottom) evidence along with the suggested scoring points used in the ClinGen clinical validity framework.	93
Figure 3-2: Evidence framework from the ACMG/AMP guidelines illustrating the different lines of evidence and strength levels..	97
Figure 3-3: Proposed changes to the ACMG/AMP guidelines for the PM1 rule by Walsh et al., 2019	101
Figure 3-4: Summary of the methodology applied in the analysis of HCM genetic variation	110
Figure 3-5: Simulated power analysis for HCM burden testing.....	111
Figure 3-6: PCA plot for common SNPs between the Egyptian (AhcHCM & EHVol) and HapMap populations.	118
Figure 3-7: Enrichment of rare variants in <i>MYH7</i> , <i>MYBPC3</i> and Myosin light chain genes in the AhcHCM cohort compared to EHVol controls..	121
Figure 3-8: Key differences in the genetic architecture of HCM between Egyptian and Caucasian cohorts.	127
Figure 3-9: Quantile-quantile (Q-Q) plot of Egyptian vs. Caucasian burden testing data based on synonymous variants.....	131

Figure 3-10: Population-matched case versus control comparison of rare variation between AhcHCM ^{EHVoi} excess and RbhHCM ^{CHVoi} excess reduces the bias introduced by gnomAD to define variant rarity and confirms the enrichment of truncating variants in <i>MYH7</i> and non-truncating variants in <i>MYL2</i> , <i>MYL3</i> and <i>JPH2</i> genes in AhcHCM patients compared to RbhHCM.	134
Figure 3-11: Distribution of rare <i>MYL2</i> (top) and <i>MYL3</i> (bottom) variants along the gene.	136
Figure 3-12: Significantly higher yield of VUSs in the (A) AhcHCM cohort compared to (B) RbhHCM.	139
Figure 3-13: Enrichment of rare, missense <i>MYH7</i> variants in the previously reported HCM cluster (red) (residues 167-931) in the AhcHCM cohort.	143
Figure 3-14: Application of the new EF-based ACMG/AMP PM1_strong- and Egypt-specific PS4 rules increased the yield of clinically actionable variants in the AhcHCM cohort.	144
Figure 4-1: Expression level of <i>MYH7</i> in the human heart (dark purple) and skeletal muscle (light purple).	176
Figure 4-2: (A) Schematic diagram of type II myosin protein and (B) the relative locations of the 4 highly conserved 'skip' residues in the myosin tail.	177
Figure 4-3: Crystal structures of myosin in the poststroke (top) and the pre-stroke (bottom) states.	178
Figure 4-4: Myosin IHM conformations in the super relaxed- and disordered relaxed states (SRX and DRX) in (A) healthy and (B) HCM cardiac muscle.	180
Figure 4-5: Graph showing the number and relative positions of <i>MYH7</i> variants identified in HCM	183
Figure 4-6: Summary of adapted ACMG/AMP classification rules by the ClinGen Inherited Cardiomyopathy Expert Panel for <i>MYH7</i> variants identified in patients with inherited cardiomyopathy	186
Figure 4-7: The mechanism of nonsense-mediated decay (NMD)	189
Figure 4-8: NMD-competent and incompetent PTCs.	190
Figure 4-9: Summary of the analysis performed on AhcHCM patients with <i>MYH7</i> fs variant c.5769delG.	193
Figure 4-10: PCA plot for common SNPs between genotyped AhcHCM patients with <i>MYH7</i> frameshift variant (<i>MYH7</i> fs: c.5769delG) and HapMap populations.	197
Figure 4-11: Distribution of AhcHCM <i>MYH7</i> fs-carriers by city of origin	198
Figure 4-12: <i>MYH7</i> fs variant (c.5769delG) co-segregated with HCM in a large AhcHCM family (LOD: 3.01).	202
Figure 4-13: Estimated age-related penetrance for <i>MYH7</i> fs-variant carriers diagnosed with HCM.	204
Figure 4-14: Patients carrying the <i>MYH7</i> frameshift (<i>MYH7</i> fs) variant (c.5769delG) presented with HCM at a significantly older age compared to AhcHCM patients with different genetic aetiologies.	205
Figure 4-15: Carrying the <i>MYH7</i> fs with another pathogenic variant in <i>MYBPC3</i> is associated with early HCM onset.	207

Figure 4-16: The MYH7fs c.5769delG variant is predicted to escape NMD and produce a truncated protein.	209
Figure 4-17: RNA-seq analysis of human myocardial tissue from 2 AhcHCM MYH7fs variant-carriers confirms the expression of the variant <i>MYH7</i> transcript.....	210
Figure 4-19: In contrast to <i>MYH7</i> truncations identified in predominantly Caucasian HCM patients, the AhcHCM variant c.5769delG is predicted to elicit NMD	213
Figure 4-20: Activated classification rules (as per the adapted ACMG/AMP framework by CMP-EP) for the MYH7fs variant (c.5769delG).	214

List of Tables

Table 1-1: Summary of genes causing HCM genocopy conditions	37
Table 2-1: Consortium/Cohort samples included in the ExAC resource	50
Table 2-2: Second round of exclusion criteria.....	59
Table 2-3: List of HCM genes selected for the analysis of background genetic variation in the EHVol and CHVol control cohorts.....	63
Table 2-4: Demographic and clinical characteristics of the EHVol cohort (n=400)	67
Table 2-5: Homozygous variants identified in the EHVol cohort	71
Table 2-6: EHVol-only (absent in gnomAD) singleton LoF variants identified in HCM-associated genes	73
Table 2-7: Common (>1%) variants in the EHVol cohort rarely observed in gnomAD.....	75
Table 2-8: Summary of previously published TRDN variants identified in patients with TKOS .	77
Table 2-9: Rare gnomAD variants re-defined to common after integrating EHVol allele frequencies into the dataset.....	81
Table 2-10: Proportion of unique variants in the EHVol and CHVol cohorts.....	82
Table 2-11: EHVol-unique variants identified in the GME cohort.....	83
Table 3-1: Rules for combining evidence codes to classify variants.....	99
Table 3-2: Summary of the HCM and reference cohorts included in the study.....	105
Table 3-3: Summary of the genes evaluated in this study as well as the number of samples sequenced in each of the study cohorts.	108
Table 3-4: Demographic and clinical characteristics of AhcHCM patients (n=514)	119
Table 3-5: Burden of rare variation in sarcomeric, syndromic and other definitive or likely HCM genes in AhcHCM patients compared to EHVol.	123
Table 3-6: Burden of rare variation in sarcomere genes, other definitive or likely HCM genes and syndromic genes associated with LVH in AhcHCM patients compared to LMM/OMGL and RbhHCM cohorts. (Table continued on next page)	128
Table 3-7: Population-matched cases versus control excess frequencies for sarcomeric- and other definitive or likely HCM genes	135
Table 3-8: Frequency of rare missense <i>MYL2</i> and <i>MYL3</i> variants observed in the AhcHCM cohort compared to LMM/OMGL- and RbhHCM patients.....	137
Table 3-9: Summary of reports of AhcHCM-unique <i>MYL2</i> and <i>MYL3</i> variants from PubMed and ClinVar	138
Table 3-10: Summary of Variants of Uncertain Significance (VUS) prevalent (>3) among AhcHCM patients	141
Table 3-11: OR and EF for missense <i>MYH7</i> variants within the pre-defined HCM cluster (167-931) in the AhcHCM cohort	143
Table 3-12: Significant burden of unique variants in AhcHCM compared to RbhHCM patients	145
Table 3-13: Frequency of rare sarcomeric homozygous variants identified in the AhcHCM cohort	147
Table 3-14: Summary of demographic and clinical characteristics for AhcHCM patients with homozygous and heterozygous variants in sarcomeric HCM genes.....	150
Table 3-15: Frequency of homozygous sarcomere HCM variants identified in the AhcHCM cohort under a recessive disease model (i.e. gnomAD FAF _{popmax} <0.00126).....	153
Table 3-16: Summary of rare double-hits in sarcomere genes identified in the AhcHCM cohort (n=15/514)	155
Table 3-17: Summary of demographic and clinical characteristics for AhcHCM patients with single and double rare variants in sarcomere genes	156

Table 3-18: Summary of homozygous <i>CSRP3</i> variants reported in the literature to date	157
Table 3-19: Summary clinical characteristics for AhcHCM patients with P/LP variants in <i>MYH7</i> compared to <i>MYBPC3</i> -variant carriers	158
Table 4-1: Demographic and clinical characteristics of AhcHCM patients with MYH7fs (n=17)	199
Table 4-2: Significance of the enrichment of <i>MYH7</i> c.5769delG in the AhcHCM cohort over control populations (gnomAD, EHVol, NEA (sub-population of GME) and Iranome)	199
Table 4-3: Estimated population penetrance of <i>MYH7</i> c.5769delG using EHVol and NEA as reference controls (total cohort size: 767) together with OR and EF	200
Table 4-4: Clinical characteristics of the proband (i.e. Patient A) as well as affected family members carrying the MYH7fs variant	203
Table 4-5: Clinical characteristics of MYH7fs c.5769delG variant carriers harbouring a second hit in sarcomere genes	205
Table 4-6: Allele-specific expression (ASE) analysis of the mutant <i>MYH7</i> allele versus WT based on RNA-seq data of myocardial tissue and iPSC-derived CMs from 2 patients	211

Contribution of others to work presented in this thesis

Echocardiographic- and Cardiac Magnetic Resonance imaging measurements for Egyptian HCM patients and healthy volunteers (EHVols) were performed by the Cardiology and Radiology departments at Aswan Heart Centre (AHC). Shehab Anwer created the HCM and EHVols databases in RedCap. Both research nurses and research team members at AHC administered the patient questionnaire and transferred the data to RedCap. Egyptian HCM and EHVols clinical and demographic data were reviewed by the Cardiology and Radiology departments at AHC.

A small subset of DNA samples from Egyptian HCM and healthy volunteer individuals (described in chapter 2 and 3) was prepared and sequenced by Alaa Afify and Aya Galal from AHC and by the Cardiovascular (CV) Genetics and Genomics team at the Imperial College London. Variant calling, filtering, joint-genotyping and variant annotation of Egyptian genetic data were performed by Pantazis Theotokis from the CV Genetics and Genomics team using the customised in-house bioinformatics pipeline developed by the team.

The R script for the principle component analysis was provided by the CV Genetics and Genomics team with only minor modifications made from my side.

Genetic data for the Caucasian healthy volunteers (described in chapter 2) were provided by the UK Digital Heart Project at Imperial College London. Genetic data for the predominantly Caucasian HCM cohort (described in chapter 3), that was prospectively recruited to the Royal Brompton Hospital, were provided by the CV Genetics and Genomics team. Genetic data for up to 6,179 HCM patients referred to the Oxford Molecular Genetics Laboratory (OMGL) and the Partners Laboratory for Molecular Medicine (LMM) was downloaded from Walsh et al., 2017 (Supplementary Table S1A and S1B online). GnomAD whole exome sequencing data (version v2.1.1) (described in chapters 2 and 3) for the HCM-associated genes selected for the analysis were downloaded by Alaa Afify.

Genotyping of a subset of Egyptian HCM patients (described in chapter 4) was performed at the Human Genomics Facility, Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands and I performed the data quality control and analysis.

Induced pluripotent stem cell-derived cardiomyocytes from 2 Egyptian HCM patients were generated by Dr. Ayman Maher, Sara Elshorbagy and Fiesal Mohamed at AHC. RNA extractions from these cells as well as from myocardial tissues were performed by Dr. Ayman Maher and Sara Elshorbagy and RNA sequencing was performed by Aya Galal. I performed the data quality control and all subsequent downstream analysis.

Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Yasmine Aguib, Dr. Paul Barton, Dr. James Ware and Professor Sir Magdi Yacoub for their tremendous support and guidance throughout my PhD studies. Special thanks to Prof and Yasmine for believing in me and offering me the opportunity to pursue my PhD studies in collaboration with the lovely and extremely talented CV Genetics and Genomics team at Imperial College. Yasmine, thank you for being not just my supervisor but my mentor, for always being there when I needed you and for helping me keep optimistic when times were rough. Paul, thank you for your incredible support, for steering me throughout my research study and for your enthusiasm for the project. James, thank you for providing me with invaluable scientific knowledge. Prof, thank you for constantly inspiring me with your strong belief in serving patients and loving humanity. You have awakened in me the passion for expanding my knowledge in the field of cardiovascular genetics.

I would like to thank the Al Alfi Foundation for funding my PhD and for offering me a very generous mobility fund. I also would like to extend my sincere gratitude to the Science and Technology Development Fund for supporting the HCM project.

I am extremely grateful to the CV Genetic and Genomics team all of whom have been so generous with their knowledge, including Roddy Walsh, Nicky Whiffin and Francesco Mazzarotto. Roddy, a million thanks for your never-ending support. I have learned a lot from you and this thesis would not have been possible without your guidance.

Special thanks to Rachel Buchan for teaching me how to prepare DNA libraries for sequencing and for Pantazis Theotokis for performing the quality assessment of the sequencing data.

I express my sincere appreciation to the clinical team at AHC for your help with my project. I would also like to thank Ayman Maher and Sara Shorbagy for providing me with

consistently high-quality RNA to support my analysis. I am most grateful for the love and support of my second family – the AHC Genetics and Genomics team. I will never forget the encouragement and support I received from you and the great memories of our work together. Apart from the constant emotional support you provided me throughout my PhD journey, all of you: Alaa, Marina, Sarah and Aya have offered me great help with my project. Alaa, I owe you a huge debt of gratitude for always being my pillar of strength throughout this project, providing me with any kind of help whenever needed. I will be forever thankful for the strong friendship we have developed throughout the years.

Most importantly of all, I am indebted to the endless support of my family and husband. Mum, dad and Amr, thank you for your boundless faith in me and for your unconditional love and support throughout my life. You have been and always will be my source of inspiration and reason for success. Mohamed, I cannot begin to express my unfailing gratitude and love to you. Without your constant encouragement, support and understanding during the pursuit of my PhD degree, the completion of the thesis would not have been possible. I am extremely grateful for sharing this entire journey with you.

List of nonstandard abbreviations and acronyms

1KG	1000 Genomes Project
AA	Amino Acid
AB	Allelic Balance
AC	Allele Count
ACMG	American College of Medical Genetics
AF	Allele Frequency
AHC	Aswan Heart Centre
AhcHCM	Aswan Heart Centre HCM (Egyptian)
ALT	Alternate Allele
AMP	Association of Molecular Pathology
AR	Allelic Ratio
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
ASA	Alcohol Septal Ablation
ASE	Allele-specific Expression
ATP	Adenosine Triphosphate
B	Benign
BAM	Binary Alignment Map
BH	Blocked Head
BMI	Body Mass Index
BWA	Burrows-Wheeler Aligner
CAD	Coronary Artery Disease
CHVol	Caucasian Healthy Volunteer
CI	Confidence Interval
ClinGen	Clinical Genome Resource
CM	Cardiomyocyte
CMP-EP	Inherited Cardiomyopathy Expert Panel
CMR	Cardiac Magnetic Resonance
CNV	Copy Number Variation
CPVT	Catecholaminergic Polymorphic Ventricular Tachycardia
CRISPR/Cas	Clustered Regularly Interspaced Short Palindromic Repeat
CT	Computed Tomography
CVDs	Cardiovascular Diseases
DCM	Dilated Cardiomyopathy
ddNTP	Dideoxynucleotide Triphosphates
dHPLC	Denaturing High-Pressure Liquid Chromatography
DRX	Disordered Relaxed State

ECCO-GEN	Egyptian Collaborative Cardiac Genomics
ECG	Electrocardiogram
ECHO	Echocardiography
EF	Etiological Fraction
EHVol	Egyptian Healthy Volunteer
EJC	Exon-Junction Complex
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
FAF	Filtering Allele Frequency
FH	Family History
FrH	Free Head
GATK	Genome Analysis Toolkit
GIF	Genomic Inflation Factor
GME	Great Middle East
gnomAD	Genome Aggregation Database
GSA	Global Screening Array
GWAS	Genome-Wide Association Study
H3Africa	Human Heredity and Health in Africa
HCM	Hypertrophic Cardiomyopathy
HF	Heart Failure
HGMD	Human Gene Variant Database
HGVS	Human Genome Variation Society
HOCM	Hypertrophic Obstructive Cardiomyopathy
HWE	Hardy-Weinberg equilibrium
IBD	Identity-by-descent
ICC	Inherited Cardiac Condition
ICD	Implantable Cardiac Defibrillator
IHM	Interacting-Head Motif
INDEL	Insertion-Deletion
iPSC	induced Pluripotent Stem Cell
LB	Likely Benign
LD	Linkage Disequilibrium
LDL	Low-Density Lipoprotein
LGE	Late Gadolinium Enhancement
LMM	Laboratory of Molecular Medicine
LMrM	Light Meromyosin
LP	Likely Pathogenic
LVEDVi	LV end-diastolic volume index
LVEF	Left Ventricular Ejection Fraction
LVESVi	LV end-systolic volume index

LVH	Left Ventricular Hypertrophy
LVMi	Left Ventricular Mass (indexed)
LVMWT	Left Ventricular Maximal Wall Thickness
LVNC	Left Ventricular Non-Compaction
LVOTO	Left Ventricular Outflow Tract Obstruction
LVWT	Left Ventricular Wall Thickness
MAF	Minor Allele Frequency
MENA	Middle East and North Africa
miRNA	micro RNA
ncRNA	non-coding RNA
NEA	North East Africa
NGS	Next-generation sequencing
NHLBI	National, Heart, Lung and Blood Institute
NMD	Nonsense-Mediated Decay
NYHA	New York Heart Association
OMGL	Oxford Molecular Genetic Laboratory
OR	Odds Ratio
P	Pathogenic
PCA	Principal Component Analysis
PRS	Polygenic Risk Score
PTC	Premature Termination Codon
Q-Q	Quantile-Quantile
QC	Quality Control
QD	Quality of Depth
QTRG	Qatar Genome
RCM	Restrictive Cardiomyopathy
REDCap	Research Electronic Data Capture
RFLP	Restriction Fragment Length Polymorphism
RNA-seq	RNA sequencing
ROH	Runs of Homozygosity
S1/2	Subfragment 1/2
SAET	SOLiD Accuracy Enhancement Tool
SAMtools	Sequence Alignment/MAP tools
SCD	Sudden Cardiac Death
SD	Standard Deviation
SHaRE	Sarcomeric Human Cardiomyopathy Registry
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SR	Sarcoplasmic Reticulum

SRX	Super Relaxed State
STAR	Spliced Transcripts Alignment to a Reference
TKOS	Triadin Knockout Syndrome
UTR	Untranslated Region
VEP	Variant Effect Predictor
VUS	Variant of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT	Wild Type

Awards and publications

Awards related to work in this thesis

July 2018: **Highly commended** oral presentation at the National Heart Lung Institute Post Graduate Research Day (Imperial College London).

Publications related to work in this thesis

1. Walsh, R., Mazzarotto, F., Whiffin, N., Buchan, R., Midwinter, W., Wilk, A., Li, N., Felkin, L., Ingold, N., Govind, R., Ahmad, M., Mazaika, E., **Allouba, M.**, Zhang, X., Marvao, A. De, Day, S. M., Ashley, E., Colan, S. D., Michels, M., ... Thomson, K. L. (2019). Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases: the case of hypertrophic cardiomyopathy. *Genome Medicine*, 11(5), 1–18.
2. Aguib, Y.* , **Allouba, M.***, Afify, A.* , Halawa, S., ElKhateb, M., Sous, M., Galal, A., Abdelrahman, E., Shehata, N., El Sawy, A., Maghawry, M., Anwer, S., Kamel, O., El-Mozy, W., Khedr, H., Essam, A., Thabet, N., Theotokis, P., Buchan, R., ... Yacoub, M. (2019). Genomics of Egyptian Healthy Volunteers: The EHVOL Study. *BioRxiv*. <https://doi.org/https://doi.org/10.1101/680520>; npj Genom. Med. (2020) (In press)
3. **Allouba, M.***, Aguib, Y.* , Walsh, R.* , Afify, A., Theotokis, P., Galal, A., Halawa, S., Shorbagy, S., Ibrahim, A. M., Kassem, H. S., Ellithy, A., Buchan, R., Hosny, M., Whiffin, N., Elguindy, A., Anwer, S., Cook, S. A., Ware, J. S., Barton, P. J., & Yacoub, M. (2020). Analysis of HCM in an understudied population reveals a new mechanism of pathogenicity. *MedRxiv*. <https://doi.org/10.1101/2020.03.24.20037358>; Circulation. (2020) (In revision)
4. Zhang, X., Walsh, R., Whiffin, N., Buchan, R., Midwinter, W., Wilk, A., Govind, R., Li, N., Ahmad, M., Mazzarotto, F., Roberts, A., Theotokis, P., Mazaika, E., **Allouba, M.**, Marvao, A. de, Pua, C. J., Day, S. M., Ashley, E., Colan, S. D., ... Ware, J. S. (2020). Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. *BioRxiv*. <https://doi.org/10.1101/2020.03.27.010736>

5. Mazzarotto, F., Hawley, M. H., Beltrami, M., Beekman, L., Boschi, B., Girolami, F., Roberts, A. M., Lodder, E. M., **Allouba, M.**, Romeih, S., Aguib, Y., Baski, A. J., Pantazis, A., Prasad, S.K., Cerbai, E., Yacoub, M.H., O'Regan, D.P., Cook, S. A., Ware, J. S., Funke, B., Olivotto, I., Bezzina, C. R., Barton, P. J. R., & Walsh, R. (2020). The genetic architecture of left ventricular non-compaction reveals both substantial overlap with other cardiomyopathies and a distinct aetiology in a subset of cases. *MedRxiv*. <https://doi.org/10.1101/2020.01.03.19015602>

Oral and poster presentations

1. Oral presentation at Challenges in future pharmaceutical trends Conference (Ahram Canadian Univeristy), Cairo, November 2018: *Defining the Genetic Architecture of Hypertrophic Cardiomyopathy in the Egyptian Population*
2. Oral presentation at the 2nd Arab Association of Genetic Research Conference, Cairo, November 2019: *The Egyptian Collaborative Cardiac Genomics (ECCO-GEN): Defining a Healthy Volunteer Cohort*
3. Oral presentation at the 7th International Cardiovascular Genomic Medicine Conference, York, December 2019: *Discovery of a novel putative pathogenic truncating variant in MYH7 in Egyptian HCM patients*
4. Poster presentation at Heart Failure: Genetics, Genomics and Epigenetics (Keystone Symposia), Utah, April 2016: *Defining the Genetic Architecture of Cardiomyopathy within the Egyptian Population*
5. Poster presentation at the 11th Conference of African Society of Human Genetics (AfSHG), Kigali, September 2018: *Towards Identifying Clinically Actionable Genetic Variants in the Egyptian Population*
6. Poster presentation at the 68th Annual Meeting of the American Society of Human Genetics, San Diego, October 2019: *Defining the Genetic Architecture of Cardiomyopathy within the Egyptian Population*

1 INTRODUCTION

1.1 Cardiovascular Diseases and Cardiomyopathy

Cardiovascular diseases (CVDs) comprise a group of disorders involving the heart and blood vessels ¹. These include: coronary heart disease, rheumatic heart disease, congenital heart disease and cardiomyopathy. Hypertension, hyperlipidaemia, diabetes, obesity and alcohol abuse represent the major determinants of CVDs ¹. According to the World Health Organization (WHO), 31% of global deaths were attributed to CVDs, making it the leading annual cause of mortality worldwide ¹. However, the highest rates of CVD mortality in low- and middle income countries greatly exceeds the rate reported in developed countries (Figure 1-1) ². For example, the rate of age-standardised mortality for CVD in Egypt is around 500 per 100,000 individuals as opposed to 100 and 200 per 100,000 individuals for the United States, France and Japan ².

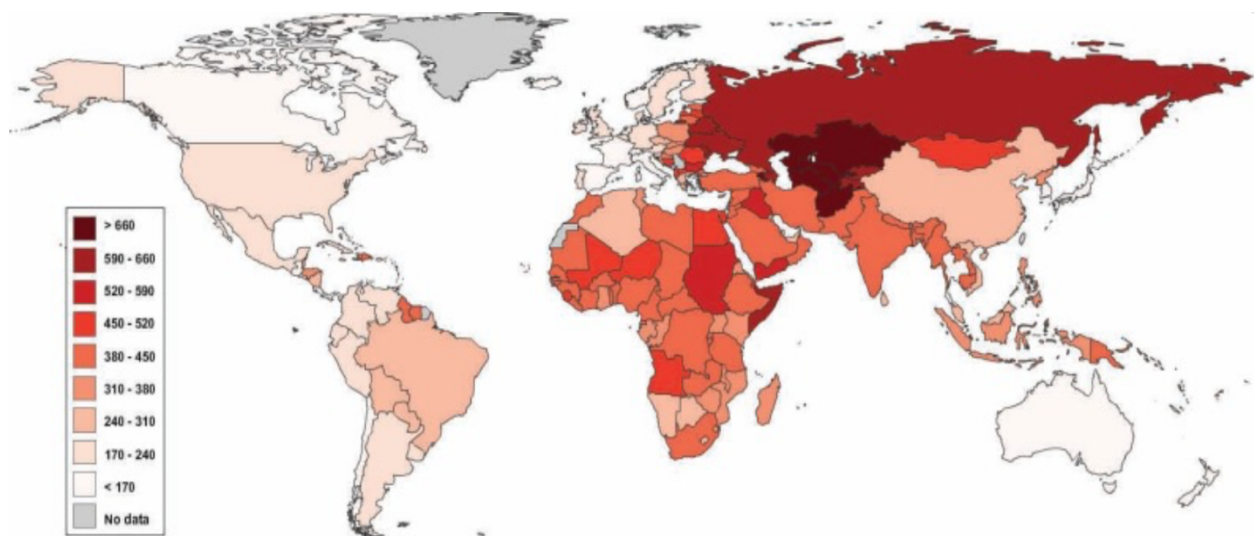


Figure 1-1: Age-standardised mortality rate (per 100,000 individuals) due to cardiovascular diseases. From: Fuster *et al.*, 2010 ².

Cardiomyopathy represents a progressive heart disease that leads to the weakening of the myocardium rendering it incapable of efficiently pumping blood to the rest of the body. It encompasses a variety of myocardial disorders that exhibit both structural and functional abnormalities in the absence of congenital or valvular heart disease, hypertension or coronary artery disease ³. Until recently, identification of the different types of cardiomyopathy was solely based on the morpho-functional phenotype, which may not be consistent with the genetic characteristics. To address this shortcoming, a comprehensive classification system; MOGE(S), was recently proposed ⁴. MOGE(S) classification integrates the morpho-functional phenotype (M), organ(s) involvement (O), genetic inheritance pattern (G), etiological annotation (E) and functional status (S) ⁴.

On the basis of the morpho-functional phenotype, cardiomyopathies can be classified into 5 different subtypes: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), restrictive cardiomyopathy (RCM), arrhythmogenic right ventricular cardiomyopathy (ARVC) and left ventricular non-compaction (LVNC) ⁴. Differential diagnosis of the cardiomyopathy subtypes is achieved via echocardiography and more accurately with cardiac magnetic resonance imaging (MRI) ⁵.

The focus of this research study is on HCM, the most common inherited cardiac disorder.

1.2 Overview of Hypertrophic Cardiomyopathy (HCM)

1.2.1 Definition

HCM is an inherited cardiac disease defined by left ventricular hypertrophy (LVH) in the absence of other identifiable causes, such as hypertension, valvular or coronary artery disease (CAD) ⁶⁻⁸. It constitutes a major cause of sudden cardiac death (SCD) in the young and is the leading cause of SCD among competitive athletes ⁹. The hypertrophy is most commonly asymmetrical targeting predominantly the interventricular septum, which separates the two ventricles ⁸ (Figure 1-2). Hypertrophy of the apex (apical hypertrophy)

or the left ventricular posterior wall are less frequent ⁸. The structural abnormalities of the LV impair diastolic function (i.e. cardiac relaxation) and may predispose patients to progressive heart failure (HF) ^{9–11}. In addition, HCM exhibits morphological heterogeneity with variable degrees of hypertrophy, cardiomyocyte disarray and interstitial fibrosis ⁷. This clinical heterogeneity could be partly explained by the variability in the genetic variants that contribute to the disease ⁷.

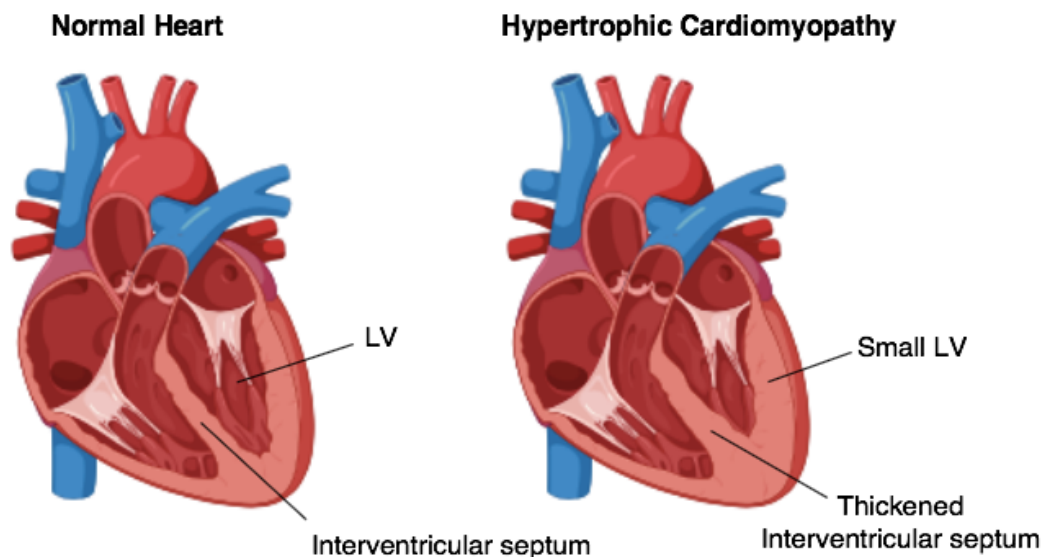


Figure 1-2: The anatomy of the normal heart compared to HCM. The hypertrophied heart (right) shows a thickened interventricular septum causing the inside of the LV to be smaller thereby reducing the blood volume in the LV. LV: left ventricle. Created with BioRender.

1.2.2 Epidemiology

Echocardiographic studies of populations where HCM has been systematically investigated reported that it affects 1 in every 500 individuals in the general population ^{12–14}. Providing a more accurate estimate of the prevalence of HCM is challenging given its age-dependent onset. LVH could develop during adolescence, however, the disease may manifest itself at any later stage in life ¹⁵. In fact, HCM was detected in 0.29% of individuals by the sixth decade of life undergoing cardiac clinical investigation in the United States as opposed to a reported annual incidence of 1.13 patients per 100,000 children (~0.001%) ^{16,17}. A few previous studies have demonstrated that the prevalence

of HCM does not vary across distinct ethnic populations ^{18–20}. It is however worth mentioning that the majority of HCM studies are based on Caucasian patients while data on other ethnic populations is limited. Yet, few studies have demonstrated differences in the clinical phenotypes of Caucasian and non-Caucasian HCM patients. For example, Sheikh *et al.*, 2016 showed that HCM patients of African/Afro-Caribbean origin presented with a higher degree of LVH and echocardiographic (ECG) abnormalities compared to Caucasians ²¹. Another study by the same research group revealed that athletes of African/Afro-Caribbean ancestry showed more ECG-abnormalities compared to Caucasian athletes and thus are more susceptible to false-positive results and invalid disqualification ²².

Moreover, data from published cohorts reported a higher proportion of males with HCM compared to females ^{23,24}. This is intriguing given the autosomal dominant inheritance pattern of the disease and thus warrants further research.

1.2.3 Aetiology

Co-morbid medical conditions, such as diabetes, valvular disease and hypertension largely contribute to HCM development ^{13,25}. Hypertension (HTN), despite being considered as an exclusion criterion for the diagnosis of HCM, is prevalent among a subset of adult HCM patients ²⁶. Hypertensive cardiac disease may be associated with a mild increase in LV wall thickness, yet a small proportion of patients present with substantial hypertrophy ($\geq 16\text{mm}$) ²¹. To discriminate between both cardiac conditions, additional clinical features, such as mitral valve abnormalities and/or family history of HCM/SCD, could be investigated ²⁷. Moreover, genetic testing for variants in sarcomere genes; the contractile apparatus of the myocardium can be further examined to confirm diagnosis.

Patients with a familial form of HCM typically inherit the disease in an autosomal dominant manner albeit with variable penetrance and disease expressivity hindering early diagnosis in some instances ²⁸.

1.2.4 Diagnosis, prognosis and treatment options

According to The European Society of Cardiology guidelines, LV wall thickness (LVWT) of $\geq 15\text{mm}$, that is not attributable solely by loading conditions, such as HTN or valvular disease, confirms the diagnosis of HCM in adults ^{25,29}. Different imaging modalities such as echocardiography, cardiac magnetic resonance imaging (CMR) or computed tomography (CT) could be utilized for the diagnosis. For patients with a lower degree of LVWT; ranging from 13-14mm, additional identifiers including family history and multi-modality cardiac imaging are required to confirm diagnosis of HCM ²⁹. A positive diagnosis of HCM in children requires an LVWT more than two standard deviations from the predicted mean, while for first degree relatives of HCM patients unexplained LVWT should be $\geq 13\text{mm}$ to confirm clinical diagnosis of HCM ²⁹.

Not all patients diagnosed with HCM will have a poor clinical outcome, in fact, epidemiological evidence shows that the annual HCM-related mortality is approximately 1% suggesting a favourable prognostic outcome ³⁰. HCM patients may be entirely asymptomatic with normal life expectancy or suffer from severe symptoms with an increased risk of HF or SCD ^{30,31}.

Asymptomatic patients will mostly present with a non-obstructive form of HCM; where the thickened LV wall does not block the blood flow from the LV to the aorta. However, a small subset of these patients (ca. 5-10%) may progress to a 'burnt out' HCM phase leading to HF due to systolic dysfunction, reduction in left ventricular ejection fraction (LVEF) $< 50\%$, LV dilatation and wall thinning ³². For these patients, cardiac transplantation may be required. Symptomatic patients, on the other hand, frequently present with left ventricular outflow tract obstruction (LVOTO). LVOTO, which is present in 25-30% of HCM patients, is defined as an instantaneous peak Doppler LVOT pressure gradient of $\geq 30\text{mm Hg}$ at rest or during provocation ^{29,30}. It constitutes the major cause of symptoms in HCM patients including angina, syncope and dyspnoea and hence is considered as a significant prognostic factor of the disease ^{30,33}.

The current clinical management of HCM focuses on the alleviation of symptoms as well as risk assessment and prevention of SCD ³⁴. To relieve symptoms in patients with LVOTO, drug therapy including β -adrenergic receptor blockers are administered to stabilise the heart rate and improve diastolic function ^{29,34}. Ca^{2+} channel blockers may be prescribed to patients who are intolerant to beta blockers ³⁵. However, patients refractory to pharmacological intervention or those with LVOTO ≥ 50 mm Hg may be candidates for invasive treatment such as alcohol septal ablation (ASA) or surgical myectomy ^{29,30}. ASA is less invasive compared to surgical myectomy and reduces LVOTO by inducing an infarct in the septum resulting in scar formation and LV modelling ³⁰. However, it is associated with a higher risk of complete heart block compared to surgery as well as a long-term risk for ventricular arrhythmia in the long run ³³. Surgical myectomy involves the removal of a portion of the interventricular septum to remove the obstruction. The clinical benefit of the surgery is immediate and thus has become the gold-standard approach for the treatment of hypertrophic obstructive cardiomyopathy (HOCM) ³³.

Finally, the annual risk of SCD ranges from 0.5%-2% in young HCM adults due to ventricular fibrillation ^{36,37}. Identifying patients at risk of developing SCD; on the basis of having a family history of HCM-related SCD, significant LVOTO or non-sustained ventricular tachycardia, is crucial for their selection for implantable cardiac defibrillator (ICD) therapy ³⁸. Recently, a high level of myocardial fibrosis detected by CMR with late gadolinium enhancement (LGE) has also been associated with an elevated risk of SCD ³⁹. Clinical models predicting the risk of SCD at 5 years based on different clinical parameters have recently been developed ⁴⁰. Yet, determining the level of SCD risk justifying ICD therapy is still an ongoing challenge ⁴⁰.

1.3 The Genetic Architecture and Disease Mechanism of HCM

1.3.1 The genetic basis of HCM

The ground-breaking discoveries by the Seidmans in the 1990s paved the way for the identification of the molecular genetic basis of HCM. To investigate the causative genes involved in familial HCM, the research group performed linkage analysis using Restriction Fragment Length Polymorphism (RFLP) to determine the chromosomal loci that segregated with the disease in families. This approach led to the identification of a missense variant at an evolutionary conserved residue in *MYH7* (Arg403Glu), which encodes the sarcomeric protein *MYH7* (β -myosin heavy chain) ⁴¹. In subsequent years, a plethora of genetic studies reported multiple variants in 7 additional sarcomere genes namely: Myosin-binding protein C (*MYBPC3*), Troponin I (*TNNI3*), troponin T (*TNNT2*), actin (*ACTC1*), regulatory myosin light chain (*MYL2*), essential myosin light chain (*MYL3*) and alpha-tropomyosin (*TPM1*) ^{42–46}. Thus, these discoveries established HCM as a disease of the sarcomere with *MYBPC3* and *MYH7* accounting for up to 80% of pathogenic variants ^{24,47}.

1.3.2 Structure and function of the sarcomere

The sarcomere constitutes the contractile unit within the cardiomyocytes (Figure 1-3). It is composed of thick and thin filaments, which slide past each other to generate force ⁴⁸. The thick filaments are composed of myosin and the thin filaments are composed of three different proteins: actin, troponin and tropomyosin ⁴⁹. The sarcomere's Z disk (two Z lines) defines its ends and supports its mechanical stability ⁵⁰.

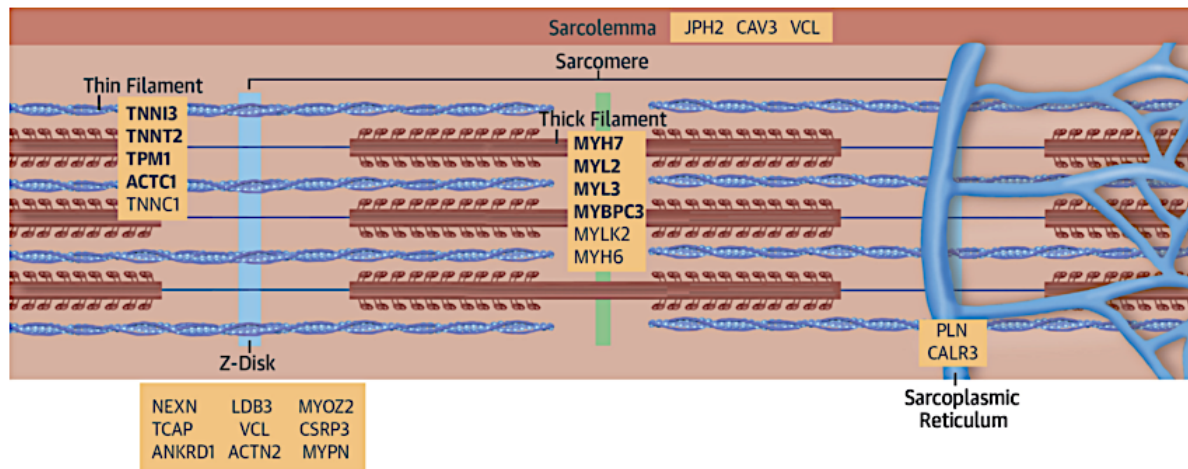


Figure 1-3: Scheme of the cardiac sarcomere. The main thick- and thin-filament genes involved in familial HCM are marked in bold. Reproduced from Burke *et al.*, 2016 under a Creative Commons license ⁶

Upon cardiomyocyte depolarization, calcium (Ca^{2+}) enters the cell via the voltage-gated L-type Ca^{2+} channels in the membrane and stimulates the release of Ca^{2+} stored in the sarcoplasmic reticulum (SR) through Ryanodine Receptor 2 (*RYR2*) causing a rise in intracellular Ca^{2+} levels. The released Ca^{2+} ions bind to the troponin complex (consisting of Troponins C, T and I) resulting in a conformational change that stimulates the Troponin I subunit to displace Tropomyosin which is blocking the interaction between the myosin head and actin ⁴⁹ (Figure 1-4). The accessible myosin head then interacts with actin to hydrolyse Adenosine Triphosphate (ATP) leading to contractile force generation ⁹.

Following contraction, the sarco/endoplasmic reticulum Ca^{2+} -ATPase (*SERCA2a*) induces the ATP-dependent uptake of Ca^{2+} ions in the SR ⁵¹. The troponin complex and tropomyosin rebind actin, thereby inhibiting actin-myosin interactions which leads to cardiac relaxation ⁹.

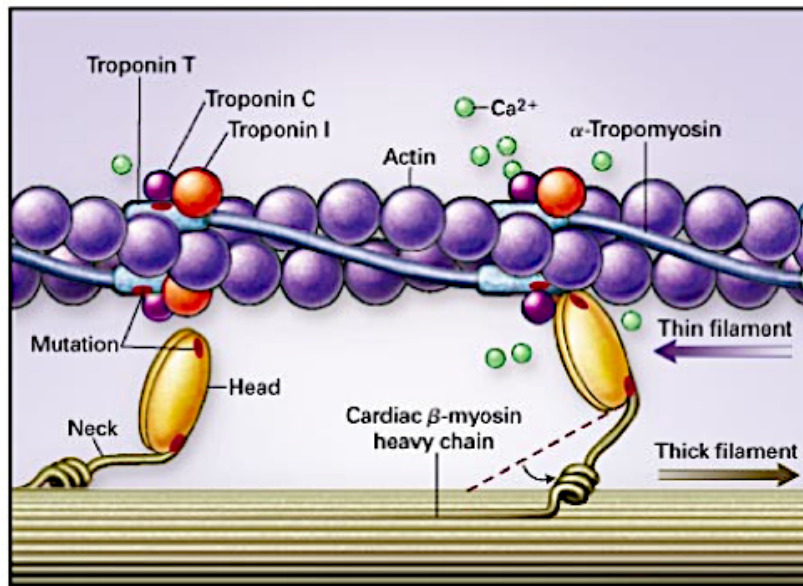


Figure 1-4: Molecular mechanism of cardiac contraction. Contraction is initiated when intracellular Ca^{2+} binds to the troponin complex (Troponins C, T and I) to activate the myosin-actin interaction. This interaction stimulates the thin and thick filaments to slide past each other to generate contractile force. Reproduced with permission from (Kamisago *et al.*, 2000), Copyright Massachusetts Medical Society ⁵².

1.3.3 Molecular mechanisms of sarcomere dysfunction in HCM

With the exception of *MYBPC3*, most variants in sarcomere genes are missense; which substitute a single amino acid for another, and are believed to cause HCM through dominant negative actions ^{14,53}. According to this disease model, the mutant sarcomeric protein incorporates into the sarcomere as a ‘poison peptide’ that interferes with normal sarcomere function ⁵⁴. *In vivo* studies have shown that sarcomeric proteins harbouring HCM variants can affect Ca^{2+} sensitivity, by delaying Ca^{2+} reuptake into the SR ^{55,56}. As a result mutant cardiomyocytes retain abnormal levels of diastolic Ca^{2+} leading to delayed relaxation (i.e. diastolic dysfunction) ^{57,58}. In addition, the observed increased ATPase activity of mutant cardiomyocytes leads to an elevated energy demand, which if unmet can cause premature cardiomyocyte death and fibrosis ⁴⁷. Myocardial fibrosis is a pathological hallmark of HCM that is caused by an elevated number of cardiac fibroblasts and extracellular matrix proteins ⁹. The fibrosis is correlated with the degree of LVH and may contribute to the “burnt out” HCM phase preceding HF as well as to arrhythmias and SCD ^{59,60}. In short, the dual effects of Ca^{2+} sensitivity and increased ATP hydrolysis

promote myocardial hypercontractility and instigate a remodelling process, resulting in cardiomyocyte hypertrophy.

In contrast to all other sarcomere HCM genes, the majority of disease-causing *MYBPC3* (ca. 70%) variants are truncating (i.e. frameshift, essential splice site, nonsense), which would radically change the protein sequence. Thus, mutant transcripts harbouring these truncating variants are believed to be eliminated from cardiomyocytes by cell surveillance mechanisms prior to translation resulting in haploinsufficiency (i.e. an insufficient amount of functional *MYBPC3* is expressed for normal sarcomere performance) ^{47,54}.

1.3.4 Challenges in developing a complete understanding of HCM pathogenesis

Despite the strong evidence of the involvement of the sarcomeric variants in HCM, it still remains challenging to determine the contribution of these variants to the prognosis of HCM patients. For example, Watkins *et al.*, 1992 reported a novel Val606Met variant in *MYH7* as benign as it was prevalent among 18 HCM patients who had nearly normal survival ⁶¹. Years later, other studies described this variant as ‘malignant’ due to its association with SCD ^{62,63}.

Phenotypic heterogeneity is also common among affected individuals within a family who carry the same sarcomeric variant ^{64,65}. This suggests that HCM is a multifactorial disease involving an interaction between causal variants, genetic and/or epigenetic modifiers and the environment. For example, diet, exercise, loading conditions and exposure to toxins may contribute to the HCM phenotype ³⁴.

Adding to this complexity is the fact that up to 70% of patients with diagnosis of HCM do not carry pathogenic variants sarcomere genes ⁶⁶. The elusiveness of the genetic aetiology for these patients provides additional evidence against the notion that HCM is a monogenic disease caused solely by sarcomeric variants. To resolve this missing heritability of HCM, numerous studies have capitalised on candidate gene approaches to

identify novel disease-causing genes. The majority of these studies, however, failed to adequately account for background genetic variation or to provide definitive evidence of variant co-segregation with HCM in families ⁶⁷. Further research is required to understand the variability in disease penetrance and expressivity, which could have a transformative impact on HCM treatment and prevention.

1.3.5 Implications and challenges of clinical genetic testing for HCM

While still searching for the “Holy Grail” underlying the relationship between genotype and phenotype in HCM patients to improve risk stratification, clinical genetic testing has emerged as a cost-effective screening approach for probands and at-risk family members sharing the same disease-causing variant. For the proband, identification of a definitively pathogenic variant can enable clinicians to differentiate HCM from other conditions that may have a similar phenotype. To illustrate, variants in genes associated with LVH may cause different phenotypes mimicking HCM ⁶⁸. These conditions are referred to as HCM genocopies and account for almost 3% of patients with a positive HCM genetic testing result ²⁸. Genocopies include: glycogen storage disorders (e.g. Danon disease), lysosomal storage disorders (e.g. Fabry disease) and adenosine monophosphate-activated protein kinase (*PRKAG2*) cardiomyopathy (Table 1-1) ⁶⁸.

Table 1-1: Summary of genes causing HCM genocopy conditions

Gene	Inheritance	Condition	References
<i>CACNA1C</i>	Autosomal dominant	Timothy syndrome	Boczek <i>et al.</i> ⁶⁹
<i>DES</i>	Autosomal dominant	Desminopathy	Van Spaendonck-Zwarts <i>et al.</i> ⁷⁰
<i>FHL1</i>	X-linked	Emery-Dreifuss muscular dystrophy	Knoblauch <i>et al.</i> ⁷¹
<i>GLA</i>	X-linked	Fabry disease	Elliott <i>et al.</i> ⁷²
<i>LAMP2</i>	X-linked	Danon disease	D'souza <i>et al.</i> ⁷³
<i>PRKAG2</i>	Autosomal dominant	<i>PRKAG2</i> cardiomyopathy	Banankhah <i>et al.</i> ⁷⁴
<i>PTPN11</i>	Autosomal dominant	Noonan syndrome	Tartaglia <i>et al.</i> ⁷⁵
<i>RAF1</i>	Autosomal dominant	Noonan syndrome	Razzaque <i>et al.</i> ⁷⁶
<i>TTR</i>	Autosomal dominant	Transthyretin amyloidosis	González-López <i>et al.</i> ⁷⁷

Distinguishing between sarcomeric HCM and metabolic cardiomyopathy is crucial, given the discrepancies in disease progression and management. For example, patients with Fabry disease could clinically benefit from enzyme replacement therapy with recombinant α -galactosidase, which reduces LVH and improves cardiac function ^{78,79}. Another advantage of genetic testing for patients is that it facilitates risk stratification; for example, HCM patients with pathogenic variants in sarcomere genes are more likely to have worse clinical outcomes compared to non-sarcomere variant carriers ⁸⁰.

Screening of family members of probands with pathogenic variant(s) is another major advantage of genetic testing. Firstly, it enables the identification of at-risk members, who carry pathogenic variants, and thus offers them longitudinal clinical follow-up and reproductive counselling ²⁸. Secondly, it alleviates healthcare expenditures for relatives without pathogenic variants resulting in substantial economic benefits ²⁸.

However, there are still several limitations that reduce the impact of clinical genetic testing for HCM. Firstly, more than 50% of patients receive a negative genetic testing result meaning that no pathogenic variant was detected ⁸¹. Secondly, numerous variants of uncertain significance (VUS) are identified in patients. The uncertainty in the pathogenicity of such variants reduces the clinical utility of genetic testing and could potentially result in an inaccurate communication of the genetic testing result. Therefore, clinical decision-making post genetic testing for HCM must be guided by accurate variant interpretation and an effective communication process ⁸². In addition, a genetic testing result revealing multiple variants (e.g. double hits) in a proband could be difficult to interpret as current understanding of the interaction between different variants is lacking. Identifying genotype-positive, phenotype-negative individuals poses also many uncertainties including how frequently these individuals should be clinically screened or whether they should change their lifestyle choices; e.g. refrain from intensive athletics ⁶. Studying the clinical phenotypes of these individuals could, however, enhance our understanding of the progression of HCM and create a unique opportunity for developing

disease-preventing therapies. Finally, the increasing number of genes reported to be associated with HCM has obscured our understanding of the genetic landscape of HCM.

Yet, despite these limitations, next-generation sequencing technologies have greatly enhanced the utility of clinical genetic testing in HCM and thus continue to provide an unprecedented potential to accelerate patient-centred precision medicine.

1.4 Next-Generation Sequencing (NGS) and its clinical implications for HCM

1.4.1 The evolution of sequencing technology

Until 2003, the application of genetic testing to identify putative pathogenic variants in sarcomere genes in HCM patients depended on expensive and unscalable approaches, such as Sanger sequencing ⁸³. This sequencing method, which was developed by Frederick Sanger in 1977, was based on a “chain termination” approach whereby radioactively labelled DNA dideoxynucleotide triphosphates (ddNTP), representing the 4 DNA nucleotides, are used to terminate the elongation of the DNA ⁸⁴. The ddNTPs lack the 3' hydroxyl group required for bond formation between two nucleotides, hence DNA elongation is halted once a ddNTP is incorporated ⁸⁵. This results in the generation of DNA fragments of variable lengths terminated with the dideoxynucleotides at the 3' end. Following synthesis, the DNA fragments are then separated using gel electrophoresis ⁸⁴. This method was further improved using fluorescent-labelled nucleotides, capillary electrophoresis and more advanced laser fluorescent signal detectors ⁸⁶. In 1987, Applied Biosystems (USA) created the first automated Sanger sequencing machine, which provided a faster sequencing performance ⁸⁷. Automated Sanger sequencing, dominated genomic research for three decades and facilitated the mapping of the first human genome in 2003 ^{87,88}. However, several limitations of Sanger sequencing, including its high costs, low throughput, lengthy process and limit of DNA fragment sizes, encouraged research scientists to develop high-throughput sequencing strategies, which ultimately

led to the development of the “*next-generation*” or “*massive-parallel*” sequencing approach ⁸⁸.

NGS is a broad term that encompasses all technologies capable of sequencing large numbers of fragmented DNA sequences in parallel ⁸⁶. Instead of using ddNTPs to terminate the chain elongation, NGS uses modified dideoxynucleotide triphosphates containing a terminator which blocks further elongation ⁸⁹. The fluorescently labelled terminator emits a signal as each dNTP binds to its complementary nucleotide and is then cleaved to allow for the binding of the subsequent nucleotide ⁹⁰. Thus, terminators are reversible. As a result, billions of sequencing reactions are processed simultaneously thereby substantially enhancing the efficiency and throughput compared to Sanger sequencing ⁸⁴.

1.4.2 Applications of NGS

The advancement of NGS technology has led to the identification of numerous putative pathogenic variants associated with HCM (discussed in detail in chapter 3). Currently, 3 NGS strategies are being utilised in HCM genetic testing, which are: gene panel sequencing, whole-exome sequencing (WES) and whole-genome sequencing (WGS), and each has its advantages and limitations.

Sequencing of predefined gene panels allows for analysing variants in specific genes or gene regions suspected to be associated with the disease under study. In this approach, sequencing reads are distributed to specific genomic locations enabling high coverage and ensuring accurate detection of variants residing in these regions ⁸⁶. As a result, many research scientists have attempted to develop gene panels to identify variants in disease-specific genes. For example, Illumina’s TruSight Cardio gene panel allows for the accurate detection of variants in 174 genes involved in inherited cardiac conditions (ICC), including HCM, at excellent coverage ⁹¹. In addition to its high sensitivity in detecting variants, which exceeds that of both WES and WGS, gene panel sequencing allows more samples to be sequenced per run, hence reducing costs ⁹¹. However, targeted gene

panels are known to poorly detect copy number variants (CNVs) (i.e. large insertions, deletions, or duplications of exons, genes or large chromosomal segments), as CNV breakpoints could reside outside the genes targeted for sequencing⁹². Another limitation of targeted gene panels is that they do not allow for the detection of variants in novel genes and require regular review to ensure they include all disease-associated genes.

WES allows for the detection of variants in coding regions (i.e. exons) of all genes (>20,000) in the genome, which comprise roughly 30 million base-pairs that are translated into proteins. Thus, it serves as a comprehensive approach for the detection of protein-altering variants for genes known to be associated with disease as well as for the discovery of novel gene-disease associations. However, WES has 2 main limitations which should be considered before pursuing this sequencing approach. Firstly, the coverage across all target sites is not uniform, which may affect the sensitivity of detecting variants in potentially disease-causing genes⁹³. Secondly, similar to targeted gene panel sequencing, the dependence of WES on exon capture only causes it to poorly detect CNVs, which may extend beyond the protein-coding regions⁹².

On a much broader perspective, WGS allows for variant detection of the entire genome. In contrast to targeted sequencing approaches, WGS provides a uniform coverage of all regions thereby allowing an improved detection of CNVs⁹⁴. Additionally, WGS has become more affordable as its costs have substantially decreased from ~\$100 million per genome, following the completion of the Human Genome Project in 2001, to \$1,906 using the HiSeq platform in 2018⁹⁵. Thus, WGS became only ~3.5 fold more expensive than running a single WES test, which costed \$555 per exome in 2018⁹⁵. However, processing the voluminous data generated from WGS is very labour-intensive, requires considerable expertise and high storage capacity. In addition, it is challenging to distinguish putative disease-causing variants from the thousands of rare benign variants detected by this approach.

1.4.3 NGS data analysis and challenges

Data analysis post NGS comprises 5 steps. These are: a) quality assessment of raw sequences (i.e. reads), b) alignment of reads to the reference genome, c) variant calling, d) variant annotation and e) variant interpretation.

Reads are aligned to the reference genomes to compare the sequence data between the patient samples and the reference sequence. This step outputs the percentage of reads that align to the reference genome as well as the number of reads that align at specific loci; which is referred to as the read depth ⁹⁶. Then, variant calling is performed to identify the differences in the sequences of reads between the patient samples and the reference sequence (i.e. true variants) and differentiates them from sequencing errors or artefacts. Well-established variant calling tools, such as: the Genome Analysis Toolkit (GATK) and the Sequence Alignment/Map tools (SAMtools) have been optimised to accurately identify single nucleotide variants (SNVs), multiple variant substitutions, and small INDELs ^{97,98}. The detected variants must then be annotated before interpreting them. The annotation of detected variants is typically performed by the Variant Effect Predictor (VEP), which determines the effect of the identified variants on the gene-, transcript- and protein-level ⁹⁹.

While the initial steps of NGS data analysis, including variant annotation, are relatively straightforward and have been standardised in various bioinformatics pipelines, accurately interpreting the identified variants associated with HCM remains very challenging.

The inability to interpret variants will undoubtedly affect their clinical actionability and consequently hinder the progression of precision medicine. In addition, communicating false-positive or false-negative genetic testing results could be very devastating or misleading to patients. This explains why clinical laboratories tend to be conservative in classifying a variant as pathogenic and report back VUSs instead ¹⁰⁰. By contrast,

research studies commonly report large numbers of ‘pathogenic’ variants without providing sufficient evidence to support their classification ¹⁰⁰.

As discussed in more detail in chapter 2, the recent development of comprehensive reference population databases has greatly contributed to more accurate variant interpretation. Genotype-phenotype databases, such as ClinVar, also provide a valuable resource for variant interpretation by providing evidence of the pathogenicity of specific variants ¹⁰¹. Yet, inconsistency of the reported pathogenicity of certain variants in ClinVar limits its usefulness. In addition, the recent release of various standards and guidelines for variant interpretation by the American College of Medical Genetics (ACMG) and the Association of Molecular Pathology (AMP) (discussed in detail in chapter 3) has provided a robust framework for interpreting variants implicated in Mendelian diseases ¹⁰².

Finally, as clinical genetic testing becomes deeply integrated into routine medical practice, it will be of immense importance to consider the ACMG/AMP guidelines and to extensively review the literature for supporting evidence of variant interpretation.

1.5 Limited knowledge of the genetic basis of HCM in Egypt

Despite the substantial progress made in identifying the genetic determinants of HCM over the past 20 years, HCM genetic studies on populations from the Middle East and North Africa (MENA) region have been scarce. Given the influence of ethnicity on genetic variation (discussed in more detail in chapter 2), more genetic research in MENA populations is required to improve the accuracy of variant interpretation and subsequently facilitate the application of precision medicine in the region.

This research study focuses on determining the genetic architecture of HCM in Egypt, a historically underrepresented population characterised by ethnic diversity, high degree of consanguinity and high risk of death due to CVDs.

1.5.1 Characteristics of the Egyptian population

The Egyptian population comprises over 100 million individuals mainly of ancient Egyptian origin. Ethnic minorities, such as Nubian, Berber and Bedouin Arabs constitute less than 5% of the population ¹⁰³. According to work published by Shawky *et al.*, 2011, the rate of consanguinity in Egypt is around 35%, which poses a possible risk factor for developing severe forms of inherited diseases owing to homozygosity ^{103,104}. In addition, CVD accounts for a significant proportion of deaths in the Egyptian population (exceeding 500 per 100,000 individuals) ².

Regarding HCM, knowledge on its prevalence in Egypt is lacking but considering the global incidence of the disease (1:500) and the population size (100M in 2020) almost 200,000 Egyptians are expected to be affected by HCM. Despite this estimated high disease burden, the underlying genetic determinants of HCM in the Egyptian population have not been adequately studied.

To address these issues, Professor Sir Magdi Yacoub founded the Aswan Heart Centre (AHC) in 2008 to offer state-of-the-art medical services for the underprivileged free of charge and to promote basic science and applied biomedical research pertaining to CVDs in the region. All Egyptian HCM patients analysed in this study have been prospectively recruited to AHC for clinical phenotyping and genetic testing.

1.6 Thesis objectives

1. To define the genetic architecture of HCM in the Egyptian population
2. To examine the occurrence of genetic features that are unique to Egyptian HCM patients
3. To study the effect of consanguinity on homozygosity and disease severity

1.7 Plan of investigation

1.7.1 Objective 1: To define the genetic architecture of HCM in the Egyptian population

To address this objective, I will compare the burden of rare variation in HCM-associated genes between Egyptian HCM patients and ethnically-matched, clinically phenotyped Egyptian Healthy Volunteers (EHVols). In addition, I will assess similarities and differences in the genetic architecture of HCM between Egyptian and well-characterised, predominantly Caucasian HCM populations.

1.7.2 Objective 2: To examine the occurrence of genetic features that are unique to Egyptian HCM patients

To address this objective, I will investigate whether there are rare variants that are only observed in Egyptian HCM patients and absent from well-characterised, predominantly Caucasian populations. I will then study these variants in more detail by investigating their disease mechanism and their co-segregation with HCM in affected families in order to gain deeper insights into their association with the disease.

1.7.3 Objective 3: To study the effect of consanguinity on homozygosity and disease severity

Given the high prevalence of consanguinity in the Egyptian population (ca. 35%), I will examine whether homozygous variants are enriched in Egyptian HCM patients compared to predominantly Caucasian patients^{103,104}. I will also compare the clinical characteristics of homozygous and heterozygous Egyptian HCM patients to examine whether homozygosity is associated with a more severe disease phenotype.

2 EVALUATING THE BACKGROUND GENETIC VARIATION IN HCM-ASSOCIATED GENES IN AN EGYPTIAN HEALTHY VOLUNTEER (EHVol) COHORT

2.1 Aims and hypotheses

The aim of this chapter is to define the genetic landscape of clinically phenotyped Egyptian Healthy Volunteers (EHVols) by assessing the frequency of common and rare variation in validated HCM genes. The second aim is to evaluate the representation of the EHVol cohort in gnomAD; the largest population reference dataset to date, as well as the Great Middle Eastern (GME) dataset. The third aim is to compare the frequency of rare variation between Egyptian and Caucasian controls in order to investigate if there are differences in background genetic variation between both cohorts. The hypotheses for this chapter are as follows:

- Egyptian background genetic variation is poorly captured in gnomAD
- The underrepresentation of the Egyptian population in gnomAD results in an inaccurate filtering of rare variants in EHVol controls
- Genetic testing of asymptomatic EHVol controls aids in the identification of subclinical phenotypes

2.2 Introduction

2.2.1 The impact of large-scale reference databases in understanding human genetic variation

Large-scale databases of genetic variation discussed below have provided a valuable resource for studying human population evolutionary history, patterns of admixture and prioritising Mendelian variants for pathogenicity. The first reference genetic database was

the International HapMap Project, which was conceived in 2003 with the aim to generate a haplotype map of the human genome to enable the discovery of genetic variants associated with different diseases ¹⁰⁵. A few years later, in 2008, the 1000 Genomes Project (1KG) was launched and it included genetic data compiled by the preceding International HapMap Project. Raw data for the 1KG project were generated using different sequencing approaches such as low coverage WGS (mean depth 7.4X), deep exome sequencing (mean depth 65.7X) and dense microarray genotyping ¹⁰⁶. A subset of the data also included family trios and parent-child duos who were deep sequenced using the Complete Genomics platform at high coverage (mean depth 47X). The data became publicly available over 3 phases with increasing numbers of individuals ^{107–109}. In 2015, the final analysis was published incorporating 26 populations (2,504 individuals) from America, Europe, Africa and Asia and contained biallelic and multiallelic SNPs, indels, and structural variants (i.e. large insertions, deletions, inversions and copy number variants (CNV)) (Figure 2-1) ¹⁰⁹. Research scientists leveraged the availability of the 1KG Project data for imputing genotypes for Genome-wide Association Studies (GWAS), mapping expression Quantitative Trait Loci, analysing population structure, estimating population-specific allele frequencies and identifying putative pathogenic variants ¹¹⁰. For studies focused on the genetics of cardiomyopathy, the 1KG project provided a valuable resource to re-evaluate the classification of variants associated with disease. For example, Golbus *et al.*, 2012, demonstrated that a subset of *MYH7* and *MYBPC3* variants reported in the Human Gene Mutation Database (HGMD) to be pathogenic in HCM were found at high frequencies in the 1KG dataset ^{111,112}. Some of these variants were found to be even more frequent than the reported population prevalence of HCM itself casting doubt on their pathogenicity ^{111,112}. Thus, to reduce the overestimation of variants as being 'actionable', research studies started adopting more stringent criteria for variant classification ¹¹³.

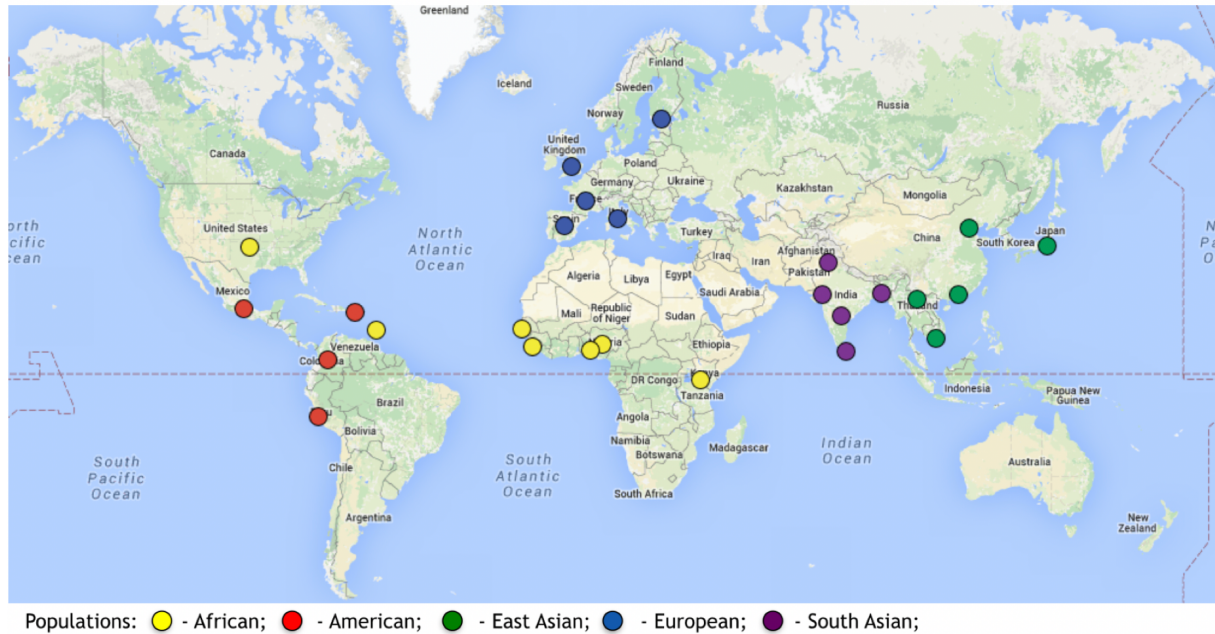


Figure 2-1: The 5 major populations analysed in the 1000 Genomes Project. From: <https://www.internationalgenome.org>

A subsequent effort by the Exome Sequencing Project (ESP) sequenced exomes of 6,515 phenotyped individuals including 4,298 European-Americans and 2,217 African-Americans ¹¹⁴. The ESP was supported by the National, Heart, Lung and Blood Institute (NHLBI) and aimed to study the association of rare protein-coding variants with heart-, lung- and blood disorders ¹¹⁴. It provided more accurate estimates of the frequency of putative pathogenic and highly penetrant variants associated with cardiomyopathy compared to 1KG ^{115–117}. For example, Andreasen *et al.*, 2013 identified in the ESP dataset 94/687 variants (14%) that were previously reported to be associated with HCM ¹¹⁸. These variants were highly prevalent among ESP controls and thus were reclassified as non-disease causing.

Concerning human population history, the study of ESP-, 1KG- data and ancient DNA showed that modern humans migrated from Africa ca. 50 thousand years ago (kya) and spread into South East Asia and Oceania 35-50 kya ¹¹⁹. Consistent with this finding, Fu *et al.*, 2013 reported that single nucleotide variants (SNVs) that arose more than 50 kya were more prevalent among African Americans compared to European Americans ¹¹⁴. In

addition, a higher enrichment of deleterious variants in Mendelian disease genes were observed in European Americans compared to African Americans. This finding highlights the stronger genetic drift (weaker purifying selection) in European Americans as a result of the Out-of-Africa migration ^{120,121}. Similarly, populations which experienced recent bottleneck events, such as the Finnish population (ca. 4 kya), have been shown to exhibit a higher frequency of rare deleterious variants, including loss-of-function (LoF) variants, compared to other European populations ¹²².

Following on from the ESP, the Exome Aggregation Consortium (ExAC) published a massive dataset of protein-coding human genetic variation from 60,706 individuals of different ancestral populations in 2016 ¹²³. Populations represented in ExAC were of European-, African-, East Asian-, South Asian- and Admixed American- (i.e. Latino) ancestry and data was collected and harmonised from different common disease- and population datasets including the 1KG and ESP (Figure 2-2 and Table 2-1) ¹²³. This valuable resource enhanced our understanding of human genetic variation. For example, Lek *et al.*, 2016 reported that the average exome contains >7.4 million non-synonymous variants, the majority of which are novel ¹²³. In addition, Ruderfer *et al.*, 2016, showed that almost 70% of ExAC individuals carried at least 1 rare CNV and estimated an index for the relative intolerance of each gene to CNVs, which may affect gene expression and contribute to disease ¹²⁴.

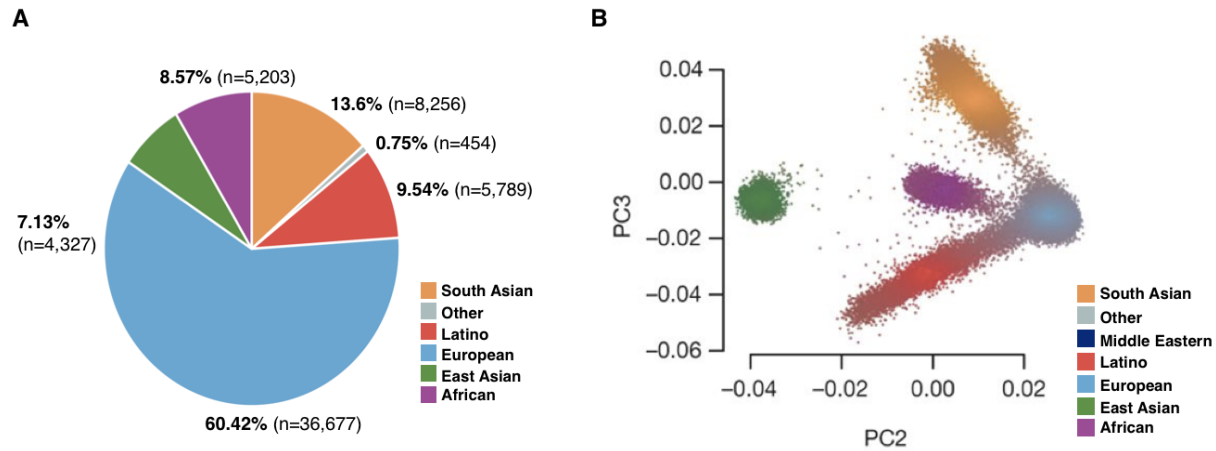


Figure 2-2: The different ancestral populations represented in ExAC. (A) The proportion of each population represented in ExAC. **(B)** Principal component analysis (PCA) showing the distribution of ExAC individuals across the main ancestral populations. From: Lek *et al.*, 2016 (13).

Table 2-1: Consortium/Cohort samples included in the ExAC resource

Consortium/Cohort	Sample size
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

From: Lek *et al.*, 2016 (supplementary table 2 ¹²³).

Concerning rare Mendelian diseases, Lek *et al.*, 2016 showed that only 9 out of 192 variants previously reported to be pathogenic had sufficient evidence for disease causality while the remaining variants were observed at high frequencies in ExAC ¹²³. Similarly, Walsh *et al.*, 2017 showed that a number of variants previously reported to cause cardiomyopathy were present in the ExAC dataset at a frequency that does not conform with penetrant Mendelian diseases (discussed in more detail in the next chapter) ¹⁰⁰.

In addition, the exceptionally higher resolution of ExAC data compared to both 1KG and ESP facilitated the discovery of genes intolerant to LoF for the first time ¹²³. Subsequent studies showed that homozygous LoF variants were enriched in South Asian samples particularly in the Pakistani cohort, where almost 38% of individuals reported to have consanguineous parents ^{125–127}. This finding highlighted the value of studying consanguineous cohorts for gene knockout discovery.

In 2017, an expansion of the ExAC resource, the Genome Aggregation Database (gnomAD), was released (version 1.0). The flagship study was based on the gnomAD version 2.1.1 dataset which includes a total of 125,748 exome data collected from different disease-specific and population genetic studies (Figure 2-3) ¹²⁸. Doubling the sample size of ExAC has enabled the detection of a larger number of predicted LoF variants ¹²⁸. In fact, gnomAD data showed that almost 72% of genes harbour an estimated number of 10 LoF variants on the canonical transcript ¹²⁸. This finding facilitated the development of a comprehensive catalogue of predicted LoF variants as well as the classification of genes with varying degrees of tolerance to LoF variation ¹²⁸. This catalogue provides a valuable resource for identifying putative pathogenic variants and potential drug targets. For example, previous studies on *PCSK9* had demonstrated that individuals carrying LoF variants in this gene were reported to have lower levels of circulating low-density lipoprotein (LDL) compared to non-LoF variant carriers ¹²⁹. As a result of this finding, monoclonal antibodies targeting *PCSK9* were developed and significantly decreased the levels circulating LDLs by approximately 60% leading to a reduction in the risk of adverse cardiovascular events in clinical trials ^{130,131}.

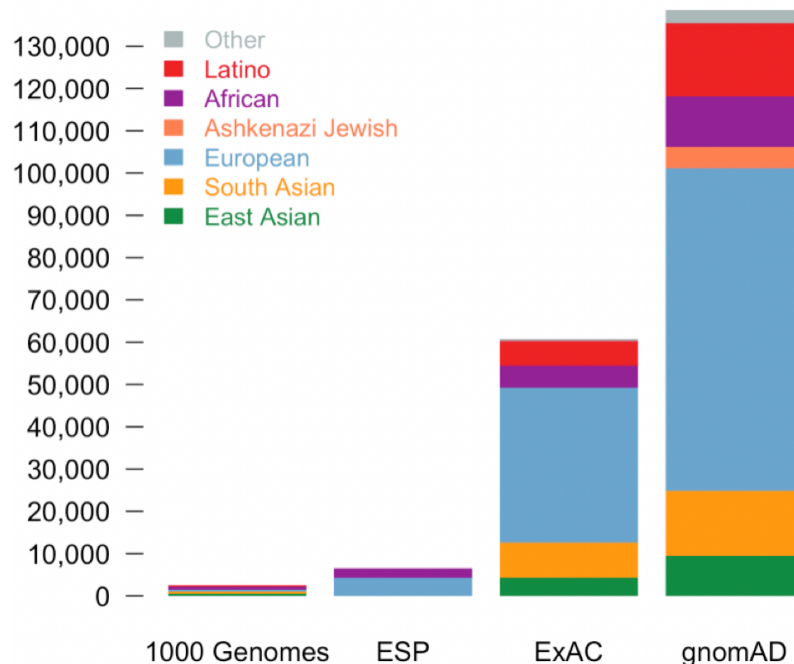


Figure 2-3: Representation of the size and ethnic composition of populations integrated into the different large-scale genetic reference databases. From: <https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/>.

Moreover, gnomAD v2.1.1 includes genome data from 15,708 individuals and version 3 increased this number 5-fold to encompass WGS data from >70K individuals ¹²⁸. Since protein-coding regions represent less than 2% of the genome, it is crucial to study the effect of genetic variation in non-coding regions, such as splice sites, promoters and enhancers on gene function ¹³². In addition to the marked increase in analysed individuals across the different ethnic populations, gnomAD has also added data from the Ashkenazi Jewish population, a genetically isolated population with a high prevalence of autosomal-recessive diseases as a result of consanguinity ¹³³. Thus, the comprehensive nature of the gnomAD resource is being widely used as a reference dataset for studying the genetic architecture of dominant Mendelian diseases, including cardiomyopathy.

2.2.2 Underrepresentation of the Middle East and North Africa (MENA) region in genomic reference databases

The filtering of Mendelian disease variants by frequency in the general population is a crucial step for the discovery of putative pathogenic variants. The accuracy of this filtering process does not only depend on the size of the reference genetic dataset; to be able to uncover rare variants associated with disease, but also on the ethnic diversity of its individuals. As demonstrated above, the frequency of rare variants can be population specific and thus studies focused on populations from the Middle East and North Africa (MENA) region, for example, cannot effectively use reference datasets that lack individuals from this region. Previous and current reference databases almost completely lack data from MENA populations. For example, the 1KG project does not include samples from the MENA region (Figure 2-1). Similarly, looking at the PCA performed on ExAC individuals, a clear separation between East Asian and other samples is observed, which reflects the near-absence of the MENA populations in the dataset (Figure 2-2B).

The Middle Eastern region as well as the vast majority of the African continent remained poorly represented in gnomAD despite the massive increase in sample size ¹²⁸. For example, “African/African-American” individuals constitute the smallest population (n=8,128) in the gnomAD exome dataset (v2.1.1) and their geographic origin is not specified. Inclusion of large-scale populations from the MENA region in population databases is therefore crucial for developing a deep understanding of the impact of genetic variation on disease, especially that individuals from these populations exhibit unique characteristics that are absent in other ethnicities ¹³⁴.



Figure 2-4: Map of 19 countries spanning the MENA region. Map adapted from: <https://mapchart.net/>.

The MENA region encompasses 19 countries spanning horizontally between Iran (east) and Morocco (west) and exhibits linguistic, ethnic and cultural diversity (Figure 2-4). MENA populations have experienced extensive admixture with European, Asian and African populations, which has resulted in their significant genetic diversity ¹³⁵. Consanguinity is a main characteristic of the Arab world with an estimated rate of over 25% of all marriages being consanguineous ¹³⁶. Thus, populations in the MENA region are expected to be enriched for homozygosity. In fact, a study by Yang *et al.*, 2014 showed that Middle Eastern populations had a significantly higher number of long runs of homozygosity (ROHs); long stretches of homozygous genomic regions where identical haplotypes are inherited from each parent, compared to non-Middle Eastern populations ^{135,137}. This finding suggests that the prevalence of recessive diseases is more common among individuals of the MENA region compared to other populations. However, research genetic studies focusing on the influence of consanguinity on the risk of Mendelian recessive disorders are limited.

The evident underrepresentation of MENA individuals in current large-scale reference databases has stimulated research groups to develop a database of background genetic variation of populations from the region. One of the first initiatives to study genetic variation in the Arab world was carried out in Qatar in 2013 where the exomes of 100 individuals from the Qatari population were sequenced ¹³⁸. Following the release of the

exome data, Sivadas *et al.*, 2016 analysed the frequency of candidate variants routinely tested for two common antithrombotic drugs Clopidogrel and Warfarin among Qatari individuals ¹³⁹. Three years later, the Qatar Genome (QTRG) database was published online and included whole genome data from 1,161 Qataris, representing 0.4% of the population ¹⁴⁰. Other large-scale genomic projects such as the “Greater Middle East” (GME) study have provided new insights into the population structure and genetic diversity of individuals from the region, which incorporates countries spanning Morocco in the west to Pakistan in the east ¹³⁴. Analysis of >1,000 individuals from the GME region showed evidence of local admixture between the different sub-populations, which contributed to a high degree of regional genetic diversity, as a result of migration ¹³⁴. In addition, the authors investigated whether GME populations experienced bottlenecks by calculating the rate of decay of linkage disequilibrium (LD) ¹³⁴. LD refers to the non-random segregation of alleles at different loci and it decays due to recombination in response to population growth. The analysis showed that both North East- and North West African populations exhibited a higher rate of LD decay across the different GME sub-populations (high genetic diversity) while Europeans and East Asian populations of the 1KG demonstrated a lower rate of LD owing to their exposure to greater bottlenecks ¹³⁴. Also, despite the elevated rate of consanguinity of the studied GME population, no evidence for purging of recessive alleles was detected. Instead, large rare ROHs were observed, which could be used to study the effect of homozygosity on disease phenotypes. Finally, the most recently released Iranome database includes WES data from 800 individuals from 8 predominant Iranian ethnic groups ¹⁴¹. The authors identified 1,575,702 variants of which almost 20% were novel ¹⁴¹. Only 0.6% of these novel variants were identified in the GME study, which highlights the value of the Iranome database in providing additional insights into genetic variation of the Arab region.

2.2.3 The Egyptian Collaborative Cardiac Genomics (ECCO-GEN) Project

As demonstrated above, several recent efforts have emerged to create a comprehensive catalogue of human genetic variation of populations in the MENA region. To further boost genomic studies in this underrepresented region, the Egyptian Collaborative Cardiac

Genomics (ECCO-GEN) Project was established by Aswan Heart Centre – Magdi Yacoub Foundation in 2015. One of the main objectives of this project is to study genetic and phenotypic variation in 1000 Egyptian healthy Volunteers (EHVols) free of CVD. In this chapter, I will discuss the background genetic variation of 400 EHVol controls in validated HCM-associated genes.

2.3 Methods

2.3.1 Study population

2.3.1.1 The Egyptian healthy volunteer (EHVol) control cohort

EHVols (n=400) were recruited to AHC from December 2015-June 2018 via national advertisement (brochures, flyers, public events). All participants provided informed consent, and the study was reviewed and approved by the local research ethics committee (20151125MYFAHC_Hvol_20161027). A paper-based questionnaire was conducted to gather information regarding demographics, health status, smoking and drinking habits, past medical and surgical history, medication, for the identification of consanguineous marriage (defined as self-reported first-cousin marriage) and family history. Family history of CVD was defined as having ≥ 1 relatives who has or has had heart disease. In the first-round of exclusion, individuals were excluded if they met any of the following criteria: <18 years of age, non-Egyptian nationality, pregnancy, presentation with known cardiovascular or collagen vascular disease, communication difficulties or contraindication to CMR (Figure 2-5). Data was entered into a Research Electronic Data Capture (REDCap) database; and was stored, along with all appropriate documentation, on an access- controlled server ¹⁴².

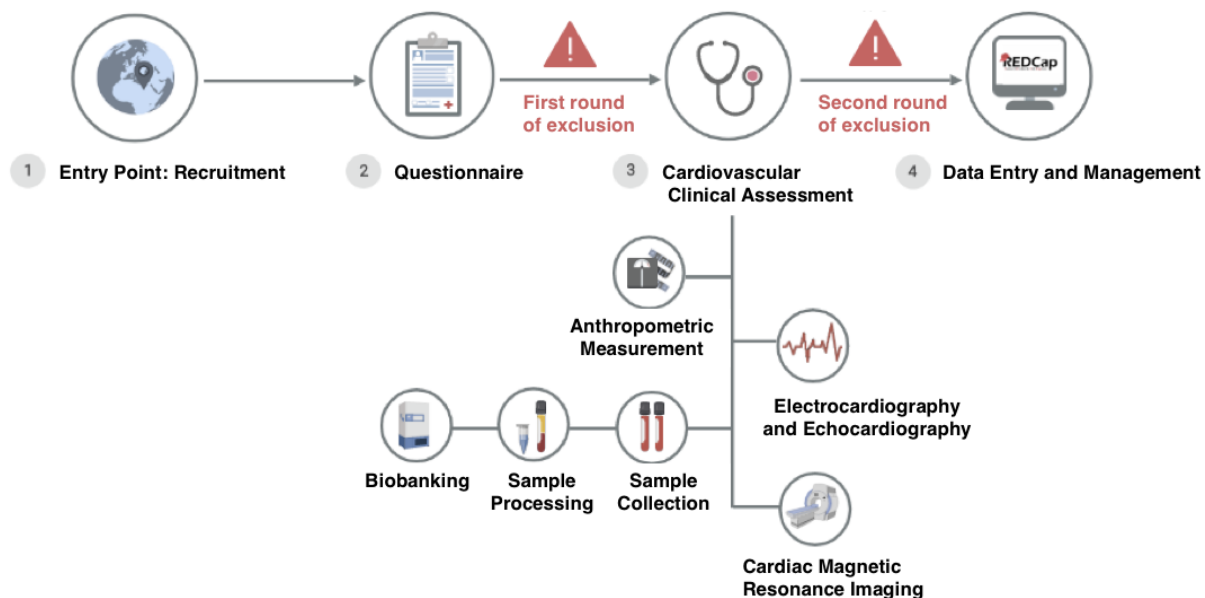


Figure 2-5: Workflow of the EHVol study. Study participants (1) were recruited from the general population via announcements (brochures, flyers, public events); (2) completed a questionnaire providing demographic data, family and clinical history; and (3) underwent detailed cardiovascular clinical assessment and blood sampling. (4) All data were recorded and managed on a local REDCap database. *Rounds of exclusion:* First round of exclusion is based on the basis of demographic and general health questionnaire as described in section 2.3.1.1. A second round of exclusion was based on detailed cardiovascular phenotyping as described under in section 2.3.1.1.1. Created with BioRender.

2.3.1.1.1 Cardiovascular phenotyping

Eligible individuals underwent detailed cardiovascular phenotyping including clinical examination, 12-lead electrocardiogram and CMR. Following phenotyping, a second round of exclusion, on the basis of specific cardiovascular diagnostic criteria, was applied (Table 2-2). Isolated apical noncompaction with normal ECG and normal CMR was not excluded.

Table 2-2: Second round of exclusion criteria

Second Round of Exclusion Criteria
Electrocardiogram (ECG) findings: 1. Abnormal QT interval 2. High-degree atrioventricular block 3. Type I Brugada ECG 4. Atrial fibrillation 5. Non-sustained VT
Cardiac magnetic resonance (CMR) findings*: • Left ventricular mass index (LVM_i) ≥ 99.5 g/m² in males and 85.5 g/m² in females • Left ventricular end-systolic volume (LVESV_i) ≥ 44.5 ml/m² in males and 40.1 ml/m² in females • Ejection fraction $< 52.5\%$ in males and 53.2% in females • LV wall thickness ≥ 15mm (measured at any segment) • Non-compacted-to-compacted myocardium thickness of $\geq 2.3:1$ affecting mid and/or basal segment (i.e. not limited to the apex/apical segments).

Participants were excluded from the study if any of these findings were identified.

*Cut-off thresholds were determined as 3 standard deviations from the mean to be more inclusive of extremes, especially in the absence of normal CMR reference values specific to the Egyptian population¹⁴³. Gadolinium-enhanced CMR imaging was not performed.

2.3.1.1.2 Sample collection

Blood samples were stored at 4°C (max. 5 days) K3EDTA tubes prior to DNA extraction to avoid clotting. DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega, Catalog No. A1620) according to manufacturer's instructions. Concentration of 1µl DNA sample was determined using the NanoDrop 2000 (Thermo Scientific) spectrophotometer. 260/280 and 260/230nm ratios were used to assess the DNA quality. All samples were stored centrally in the AHC Biobank.

2.3.1.2 The Caucasian healthy volunteer (CHVol) control cohort

The CHVols (n=1,054) were recruited prospectively via advertisement for the UK Digital Heart Project at Imperial College London. All volunteers underwent CMR to confirm the absence of cardiac disease.

2.3.1.3 The Genome Aggregation Database (gnomAD)

The gnomAD dataset is an expansion of the ExAC resource and currently contains a total of 125,748 exome- and 15,708 genome data collected from different disease-specific and population genetic studies (more details in 2.2.1). It was developed at the Broad Institute where scientists reprocessed raw sequencing data from each consortium/sequencing cohort through the same pipeline. Variant calling was performed using GATK's HaplotypeCaller. In this study, WES data was downloaded from the gnomAD database (<https://gnomad.broadinstitute.org>; version v2.1.1) and variants with a PASS filter were included in the analysis.

2.3.2 Targeted sequencing

EHVol and CHVol control samples underwent next generation targeted sequencing using the Trusight Cardio Sequencing Kit. TruSight Cardio, which was developed by the Cardiovascular Genetics and Genomics research team at Imperial College, includes 174 genes associated with inherited cardiac conditions (ICCs) ⁹¹. I prepared the targeted DNA libraries for the EHVol samples following the manufacturer's protocol (Illumina Trusight cardio sequencing protocol). To prepare a library, each sample DNA was normalised to the required concentration as recommended by the protocol. Normalised DNA samples were then subjected to 'tagmentation' whereby a transposase enzyme fragments the DNA and simultaneously tagments adapter sequences onto both ends of the template DNA. The tagmented DNA fragments were then amplified by polymerase chain reaction (PCR) to allow for the addition of specific primers required for subsequent cluster generation and sequencing. After purifying the amplified DNA fragments according to the manufacturer's protocol, the fragments were hybridised with capture probes targeting the ICC genes of interest. Then, streptavidin magnetic beads were used to capture the hybridised target regions. The process of hybridisation and capture were performed twice to obtain an enriched DNA library. The quality and concentration of the amplified library were then evaluated using the TapeStation 4200 (Agilent Technologies) and a fluorometric method (Qubit, Invitrogen), respectively. The final step of the DNA library preparation involves the

denaturation of the library and its tagmentation to the adaptors on the flow cell on which the DNA library undergoes bridge amplification resulting in the formation of clonal clusters of amplified target DNA (~1000 Molecules per ~1µm cluster). NGS is then initiated using the sequencing by synthesis method (explained in detail in Chapter 1) to generate millions of sequencing ‘reads’ representing all the fragments. Each read is approximately 100-150bp in length. The DNA libraries of the CHVol cohort were prepared by laboratory scientists of the Cardiovascular Genetics and Genomics team.

2.3.3 Bioinformatics analysis

Variant calling and filtering were performed according to a customised in-house bioinformatics pipeline developed by the Cardiovascular Genetics and Genomics team. A summary of the bioinformatics steps follows. Illumina platforms output sequencing data in Fastq format. The sequencing reads were subjected to quality control using FastQC v0.10.125 and low-quality reads were trimmed via prinseq-lite v0.20.4¹⁴⁴. Trimmed reads were then mapped to hg19 using the Burrows-Wheeler Aligner (BWA) v0.7.10-r789^{145,146}. After alignment, removal of duplicate reads was performed using picard v1.117 (“Picard Tools,” n.d.). The Genome Analysis ToolKit v3.2-2-gec30cee (GATK) was then used for re-alignment of insertions or deletions (Indels) as well as base quality score recalibration¹⁴⁷. Variant calling was performed using GATK’s HaplotypeCaller. Joint genotyping was performed with GATK v4.0.8.1 and hard filters were applied based on GATK best practices workflow for germline short variant discovery¹⁴⁷. Only variants marked as “PASS”, with a quality of depth score (QD) ≥ 4 and an allelic balance (AB) ≥ 0.20 were retained. High coverage of the target regions (>99%) at $\geq 20\times$ read depth was attained. The Ensembl Variant Effect Predictor (VEP) was used for variant annotation with respect to the Ensembl transcripts defined in Table 2-3⁹⁹.

For the prospective UK cohort, samples sequenced on the Illumina platforms followed the same bioinformatics pipeline mentioned above. For samples sequenced on the SOLiD 5500xl platform, reads were aligned in colour space using the LifeScope™ v2.5.1 “Targeted re-sequencing” pipeline. The SOLiD Accuracy Enhancement Tool (SAET) was

used to improve colour call accuracy prior to mapping. Variant calling was performed using the diBayes and SmallIndel packages in LifeScope software as well as GATK. Bases with a read coverage of min. 10 reads, mapping quality ≥ 10 and base quality ≥ 20 were marked as “PASS”. VEP was used for variant annotation with respect to the Ensembl transcripts defined in Table 2-3⁹⁹. An additional filtering step was performed on variants that were solely called by the diBayes/SmallIndel packages (and not confirmed by GATK). Variants with a coverage ≤ 100 and an allelic balance < 0.20 or a coverage < 50 and an allelic balance < 0.30 were filtered and excluded from the analysis due to the high probability of representing false positive sequencing errors.

2.3.4 Sequence variant nomenclature

Variants are described using standard Human Genome Variation Society (HGVS) nomenclature. Genomic coordinates refer to GRCh37, Transcription and protein descriptions use the references in Table 2-3.

2.3.5 Gene selection

I selected 22 HCM genes (summarised in Table 2-3) for the analysis of background genetic variation in both EHVol and CHVol controls as they have been validated for their association with the disease¹⁴⁸.

Table 2-3: List of HCM genes selected for the analysis of background genetic variation in the EHVol and CHVol control cohorts

Acronym	Gene	Ensembl transcript	Ensembl protein
<i>Sarcomere genes</i>			
<i>ACTC1</i>	Actin Alpha Cardiac Muscle 1	ENST00000290378	ENSP00000290378
<i>MYBPC3</i>	Myosin Binding Protein C3	ENST00000545968	ENSP00000442795
<i>MYH7</i>	Myosin Heavy Chain 7	ENST00000355349	ENSP00000347507
<i>MYL2</i>	Myosin Regulatory Light Chain 2	ENST00000228841	ENSP00000228841
<i>MYL3</i>	Myosin Essential Light Chain 3	ENST00000395869	ENSP00000379210
<i>TNNI3</i>	Cardiac Troponin I type 3	ENST00000344887	ENSP00000341838
<i>TNNT2</i>	Cardiac Troponin T	ENST00000367318	ENSP00000356287
<i>TPM1</i>	α -Tropomyosin	ENST00000403994	ENSP00000385107
<i>Other definitive or likely HCM genes</i>			
<i>PLN</i>	Phospholamban	ENST00000357525	ENSP00000350132
<i>ACTN2</i>	α -Actinin 2	ENST00000366578	ENSP00000355537
<i>CSRP3</i>	Cysteine and Glycine-Rich Protein 3	ENST00000533783	ENSP00000431813
<i>JPH2</i>	Junctophilin 2	ENST00000372980	ENSP00000362071
<i>TNNC1</i>	Troponin C1	ENST00000232975	ENSP00000232975
<i>Syndromic genes associated with LVH</i>			
<i>CACNA1C</i>	Calcium Voltage-Gated Channel Subunit Alpha1 C	ENST00000399655	ENSP00000382563
<i>DES</i>	Desmin	ENST00000373960	ENSP00000363071
<i>FHL1</i>	Four and A Half LIM Domains 1	ENST00000370690	ENSP00000359724
<i>GLA</i>	α -Galactosidase	ENST00000218516	ENSP00000218516
<i>LAMP2</i>	Lysosome-associated membrane protein 2	ENST00000200639	ENSP00000200639
<i>PRKAG2</i>	Protein Kinase AMP-Activated Non-Catalytic Subunit Gamma 2	ENST00000287878	ENSP00000287878
<i>PTPN11</i>	Protein Tyrosine Phosphatase Non-Receptor Type 11	ENST00000351677	ENSP00000340944
<i>RAF1</i>	Raf-1 Proto-Oncogene, Serine/Threonine Kinase	ENST00000251849	ENSP00000251849
<i>TTR</i>	Transthyretin	ENST00000237014	ENSP00000237014

2.3.6 Calculation of variant frequencies in the EHVol cohort

For each gene, I calculated the frequency of variants in the EHVol cohort by dividing the sum of the total number of variants identified per HCM gene by the total number of EHVol variants. I divided the variants into 3 types: LoF (i.e. frameshift, splice acceptor, splice donor, stop gained), Missense/Indels (missense, inframe deletion and inframe insertion) and Synonymous (Figure 2-6).

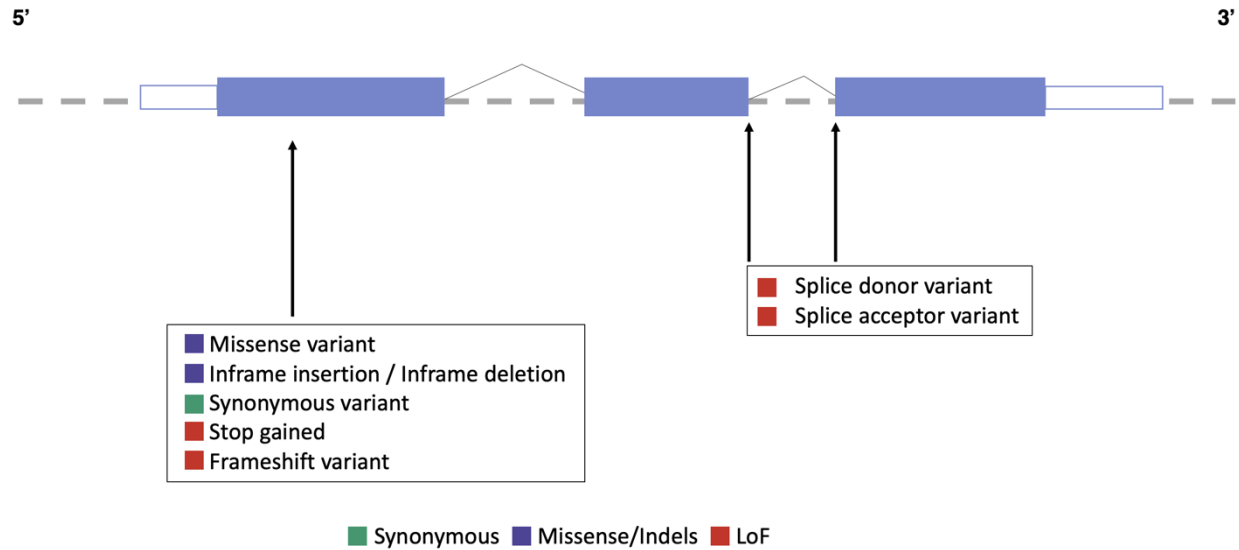


Figure 2-6: The different variant types included in the analysis of genetic variation in the EHVol cohort. Variants were collated into 3 main types: Synonymous, Missense/Indels and LoF (colour-coded in figure above). Reproduced from: <https://www.ensembl.org/>.

2.3.7 Representation of EHVol genetic variation in gnomAD

To assess the representation of EHVol genetic variation in the gnomAD dataset, I calculated the proportion of EHVol protein-coding variants (i.e. LoF, Missense/Indel and Synonymous) present in gnomAD within different observed allele count (AC) bins by dividing the sum of ACs within each bin by the total number of EHVol variants. Then, I calculated the proportion of the different variant types per EHVol AC bin.

2.3.8 Rare variant filtering

I applied a cut-off filtering allele frequency (FAF)_{popmax} of $\leq 4 \times 10^{-5}$ to filter for rare, putatively pathogenic HCM variants in the EHVol cohort. “Popmax” represents the highest FAF across the 5 main gnomAD sub-populations. This cut-off frequency was defined using the statistical framework developed by Whiffin *et al.*, 2017¹⁴⁹. A detailed explanation of this framework is described in section 3.3.4.

2.3.9 Identification of variants commonly observed in the EHVol cohort with low prevalence in gnomAD

To comprehensively examine differences in variant frequencies between EHVol and gnomAD datasets, I identified common EHVol variants (>1%) in all ICC genes (n=174) that were rarely observed in gnomAD ($\text{gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$).

2.3.10 Comparison of rare variation between EHVol and CHVol control cohorts

I calculated the frequency of rare variants ($\text{gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$) in HCM-associated genes in the EHVol- and CHVol cohorts by dividing the sum of rare variants per gene by the total number of controls sequenced. I then performed Fisher's exact test (two-sided) to evaluate the significance of the difference in the overall frequency of rare variation in all HCM-associated genes combined (n=22) between the EHVol- and CHVol cohorts. Only protein-altering variants (i.e. LoF and Missense/Indels) were included in this analysis.

2.3.11 Re-filtering of rare variants after combining EHVol genetic data with gnomAD

To assess whether combining EHVol data with gnomAD (to serve as a sixth gnomAD sub-population) could provide a more accurate definition of rare variation, I re-filtered rare variants using the $\text{FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$ cut-off frequency but this time the "popmax" represented the highest FAF across the 6 sub-populations, including the Egyptian. Thus, I referred to this frequency threshold as Egyptian_adjusted_gnomAD $\text{FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$.

2.3.12 Comparison of unique variants between the EHVol and CHVol cohorts

I compared the proportion of EHVol controls with unique protein-altering variants in HCM-associated genes to CHVol controls. Unique EHVol variants were defined as being absent from both the gnomAD and CHVol cohorts. Similarly, CHVol-unique variants were defined as being absent from gnomAD and the EHVol cohorts.

After identifying EHVol- and CHVol-unique variants, I interrogated the gnomAD database to check whether these variants had been marked by gnomAD filtering processes as low quality. These variants were then excluded from the analysis. Then, I compared the proportion of individuals with unique variants between the EHVol- and CHVol cohorts. I performed Fisher's exact test (two-sided) to evaluate the significance of the difference in the overall frequency of unique variation between the EHVol- and CHVol cohorts. Finally, I interrogated the GME database to check whether EHVol-unique variants were observed in the GME sub-populations.

2.4 Results

2.4.1 Demographic and clinical characteristics

The EHVol cohort included in the analysis consisted of 400 controls who underwent CMR. There were 233 male (58.2%) and 167 female (41.8%) controls. The mean age of EHVol controls at recruitment was 30.1 years (± 10.5 SD). A total of 169 controls (42.25%) were from Aswan and the remaining controls were from different cities across Egypt (Figure 2-7). A summary of demographic- and clinical characteristics is outlined in Table 2-4.

Table 2-4: Demographic and clinical characteristics of the EHVol cohort (n=400)

General characteristics	Total EHVol cohort, n=400
Demographics	
Age at recruitment (years) (mean (sd))	30.1 $\pm$ 10.5
Females n(%)	167 (41.8%)
Males n(%)	233 (58.2%)
Offspring of Consanguineous Marriage* (self-reported) (%)	66 (16.5%)
FH of CVD (%)	64 (16.0%)
BMI (kg/m ²) (%)	< 18.5: 15 (3.8%)
	18.5-25: 151 (37.8)
	25-30: 150 (37.5)
	>30: 83 (20.8%)
Smoking status	
Former smoker (%)	11 (2.8%)
Current smoker (%)	83 (20.8%)
Non-smoker (%)	306 (76.5%)
Alcohol consumption	
No consumption	393 (98.2%)
1 glass/week (%)	3 (0.75%)
2-3 glasses/week (%)	1 (0.25%)
Unspecified (%)	3 (0.75%)
Cardiac characteristics	
CMR Variables	
LVEF (%) (mean (sd))	63.2 $\pm$ 6.2
LVMi (g/m ²) (mean (sd))	46.5 $\pm$ 13.9
LVESVi (ml/m ²) (mean (sd))	28.3 $\pm$ 7.6
LVEDVi (ml/m ²) (mean (sd))	74.9 $\pm$ 12

*Consanguinity was defined as first-cousin marriage

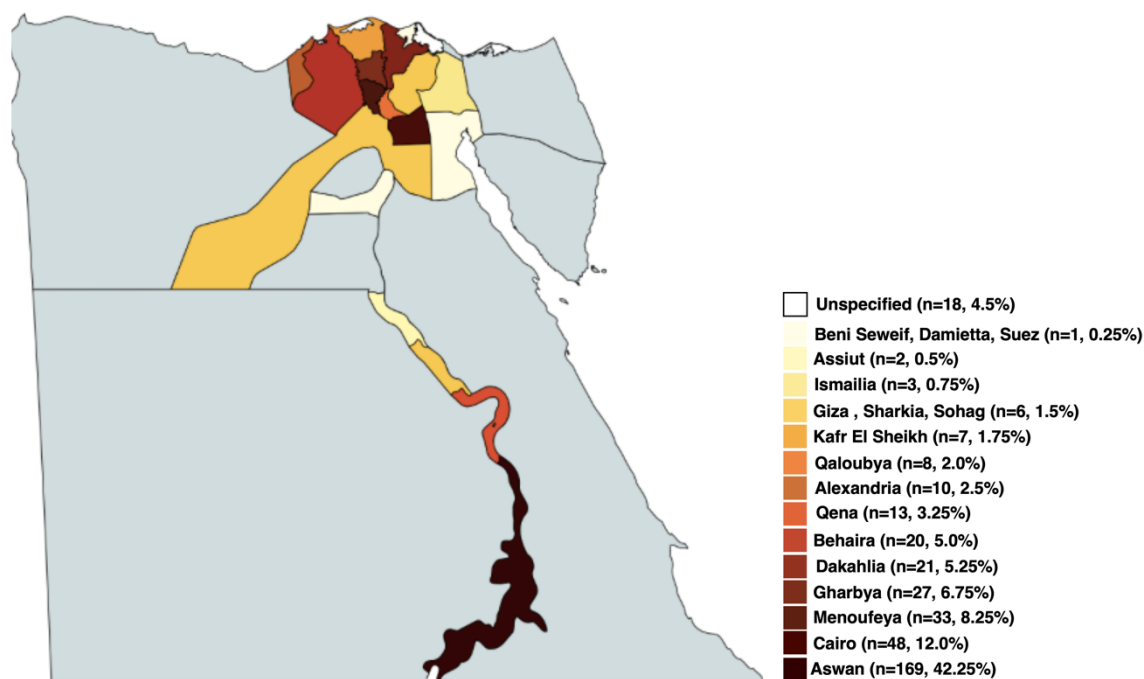


Figure 2-7: Distribution of analysed EHVol controls (n=400) by city of origin. The highest proportion of EHVol controls were from Aswan (42.3%). Map adapted from: <https://mapchart.net/>

2.4.2 Loss-of-Function (LoF) variants minimally contributed to genetic variation in the EHVol cohort

To assess the distribution of genetic variation by variant type in the EHVol cohort, I calculated the proportion of LoF, Missense/Indels and Synonymous variants (Figure 2-8). In this analysis both rare (i.e. $\text{gnomAD } \text{FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$) and common variants were included. A total of 278 distinct variants in HCM-associated genes (listed in Table 2-3) were identified in the EHVol cohort. Of the identified variants, 59.4% were synonymous whereas distinct LoF variants accounted for only 0.7% of genetic variation in this cohort.

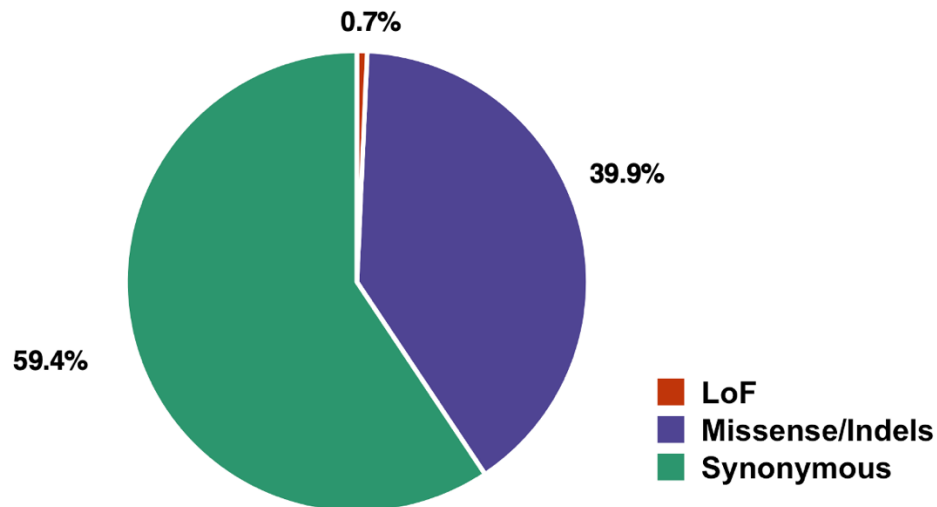


Figure 2-8: Synonymous variants accounted for >50% of genetic variation in the EHVOL cohort. The proportion of EHVOL genetic variation by variant type is shown. Synonymous (i.e. non-protein-altering) variants accounted for 59.4% of variation in the EHVOL cohort.

I also evaluated the distribution of EHVOL genetic variation per HCM-associated gene (Figure 2-9). This analysis was performed for sarcomere, syndromic and other definitive or likely HCM genes, separately. Regarding sarcomere HCM genes, both *MYBPC3* and *MYH7* accounted for the majority of genetic variation in the EHVOL cohort, 16.5% and 14.4%, respectively (Figure 2-9A). Protein-altering variants (i.e. LoF and Missense/Indels) were only observed in *MYBPC3*, *MYH7*, *TNNT2* and *MYL3*. LoF variation was identified in *MYH7* only.

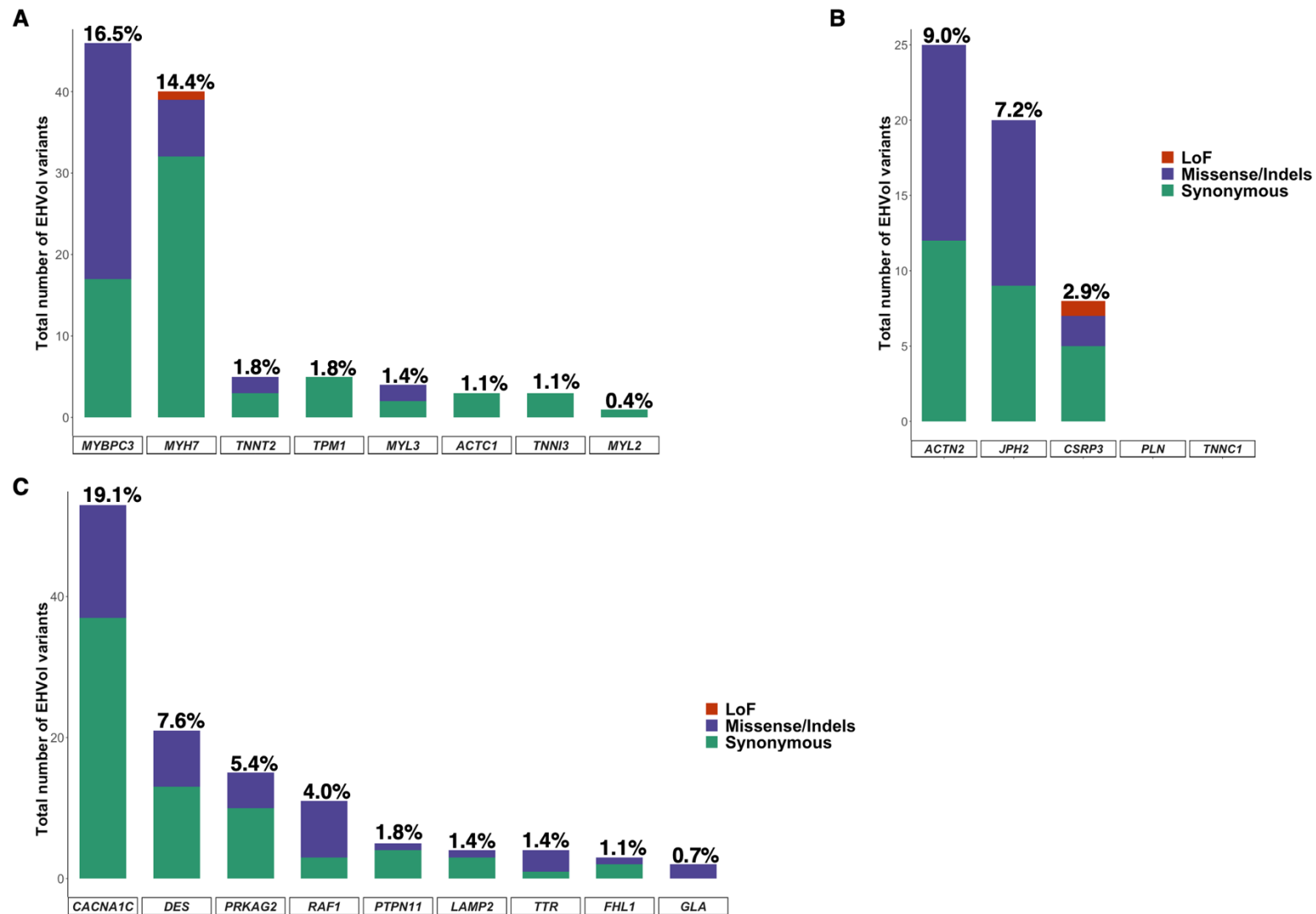


Figure 2-9: *MYH7* and *MYBPC3* (sarcomere), *CACNA1C* (syndromic) and *ACTN2* accounted for the majority of genetic variation in the EHVol cohort. Distribution of variation by variant type in (A) sarcomere genes (B) other definitive or likely HCM genes and (C) syndromic genes is shown. LoF variation was identified in *MYH7* and *CSRP3* genes only.

Protein-altering variants were not identified in *PLN* and *TNNC1* (Figure 2-9B). Of the other likely associated HCM genes, *ACTN2* (9.0%) accounted for the majority of variation followed by *JPH2* (7.2%) and *CSRP3* (2.9%). LoF variation was identified in *CSRP3* only. *CACNA1C* accounted for the majority of variation in syndromic HCM genes (Figure 2-9C). Variation in *DES* (7.6%), *PRKAG2* (5.4%) and *RAF1* (4.0%) was modest. The remaining genes *PTPN11*, *GLA*, *LAMP2*, *TTR* and *FHL1* nominally contributed to genetic variation in the EHVol cohort.

2.4.3 Absence of rare homozygous variants in the EHVol cohort

The prevalence of EHVol controls with consanguineous parents (16.5%) (Table 2-4), prompted me to assess the frequency of homozygous variants in HCM-associated genes in the cohort. I restricted this analysis to protein-altering variants (i.e. LoF and Missense/Indels).

The analysis showed that 3.53% (4/113) of protein-altering variants were identified in the homozygous state (Table 2-5). A total of 24 EHVol controls carried these homozygous variants and only 2 of the homozygotes (8.33%) were offspring of consanguineous parents. To assess whether these variants were rare under a recessive disease model, I applied the appropriate allele frequency cut-off of gnomAD $FAF_{popmax} \leq 0.00126$, which was calculated by the online tool <https://www.cardiodb.org/allelefrequencyapp/>. As expected, the identified homozygous variants were “too common” to be considered disease-causing under a recessive disease model.

Table 2-5: Homozygous variants identified in the EHVol cohort

Gene	Variant type	cDNA variant	Protein variant	gnomAD FAF_{popmax}	EHVol freq (variants/total)
<i>JPH2</i>	Missense	c.1186G>A	p.Ala596Thr	0.3554	3.25% (13/400)
<i>MYBPC3</i>	Missense	c.565G>A	p.Val189Ile	0.003394	0.25% (1/400)
	Missense	c.706A>G	p.Ser236Gly	0.128629	1.5% (6/400)
<i>TNNT2</i>	Missense	c.758A>G	p.Lys253Arg	0.148523	1.00% (4/400)

2.4.4 Absence of a subset of EHVol variants from the gnomAD dataset

To assess the representation of EHVol genetic variation in the gnomAD dataset, I calculated the proportion of EHVol variants (i.e. LoF, Missense/Indel and Synonymous) present in gnomAD within different observed AC bins. 278 variants were identified in the EHVol cohort in HCM-associated genes. Of these 278 variants, 236 (84.9%) were present in gnomAD (Figure 2-10A). The remaining (15.1%) non-gnomAD variants (i.e. EHVol only; n=42) were observed in less than 1% of the EHVol cohort. The majority (88.1%) of EHVol-only variants were singletons (i.e. observed only once in an individual (AC bin=1)).

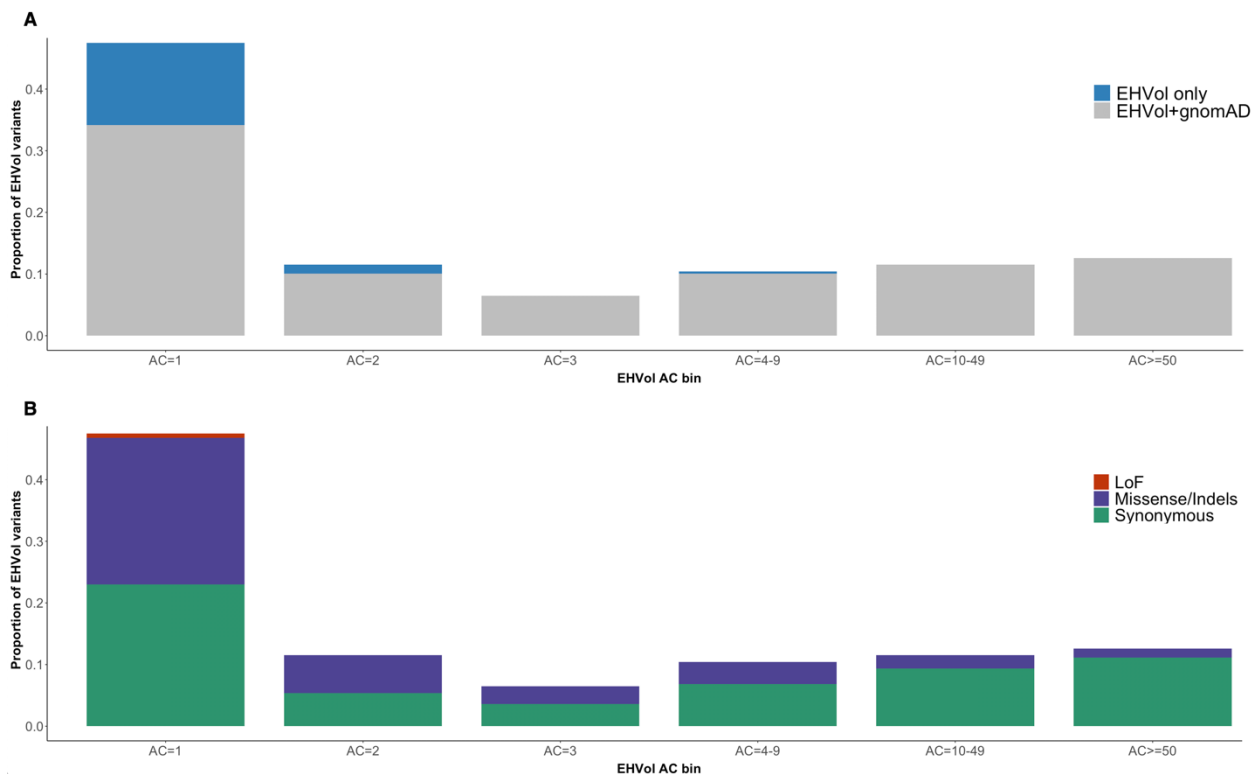


Figure 2-10: Majority of EHVol-only (absent in gnomAD) variants were observed as singletons (AC bin=1) in the EHVol cohort. (A) Bar graph showing the distribution of observed ACs (binned) of variants identified in the EHVol cohort. The proportion of variants that were present in gnomAD is shown in grey and the proportion of variants that were only present in the EHVol cohort is represented in blue. **(B)** Bar graph showing the proportion of all EHVol variants per AC bin by variant type. AC: allele count.

Given the expected lower coverage of WES gnomAD data compared to the targeted EHVol data (i.e. ICC genes on the Trusight Cardio panel) I sought to assess whether any of the identified EHVol-only variants were actually observed in gnomAD in low quality⁹¹. The analysis showed that 5/42 (11.9%) EHVol-only variants were observed in gnomAD

in low quality. This highlights the need to confirm the complete absence of novel variants in gnomAD before reporting them as ethnic-specific variants.

Finally, as demonstrated in Figure 2-10B, EHVol LoF variants (n=2) were only observed as singletons and constituted 5.4% of variants in the AC 1 bin. Both variants were identified in the heterozygous state. Details of these LoF variants are summarised in (Table 2-6).

Clinical data for both LoF-variant carriers were re-evaluated. A healthy phenotype was confirmed for the EHVol control (age 27yrs) carrying the *CSRP3* splice donor variant c.414+1G>T. However, the EHVol control (age 20yrs) with the *MYH7* frameshift variant (c.2337_2341dupGAGGC) showed signs of LVNC.

Table 2-6: EHVol-only (absent in gnomAD) singleton LoF variants identified in HCM-associated genes

Gene	Variant type	cDNA variant	Protein variant	EHVol freq (variants/total)
<i>CSRP3</i>	Splice donor variant	c.414+1G>T	-	0.25% (1/400)
<i>MYH7</i>	Frameshift	c.2337_2341dupGAGGC	p.Leu781ArgfsTer3	0.25% (1/400)

2.4.5 Clinical evidence of Left Ventricular hypertrabeculation in an EHVol control with Myosin Heavy Chain 7 (*MYH7*) LoF variant

LoF variants in *MYH7* have generally been considered non-causative in cardiomyopathies but in a very recent study (available in pre-print), we have provided evidence supporting the pathogenicity of a *MYH7* LoF variant in HCM (discussed in more detail in chapter 4) ¹⁵⁰. Also, another recent pre-print by Mazzarotto *et al.*, 2020 reported an enrichment of LoF *MYH7* variation in LVNC cases compared to controls, suggesting a causal role of this variant class in disease ¹⁵¹. Thus, I sought to evaluate whether the EHVol control carrying the frameshift *MYH7* variant (p.Leu781ArgfsTer3) presented with an LVNC cardiomyopathy phenotype. As depicted in Figure 2-11, the CMR scans of the p.Leu781ArgfsTer3-carrier showed a non-compacted myocardium of the LV at the mid-

ventricle and apical levels. The ratio of non-compacted to compacted (NC/C) endocardial layer was 3.2, which exceeds the widely used ratio threshold for LVNC cardiomyopathy diagnosis. However, in this case, the p.Leu781ArgfsTer3 variant did not cause a LVNC cardiomyopathy phenotype as the individual did not present with LV dysfunction or ECG abnormalities and did not have a family history of LVNC cardiomyopathy. It is worth mentioning that several studies have reported that LV non-compaction/hypertrabeculations can be observed in healthy individuals with normal LV function in response to vigorous physical exercise ^{152,153}. Therefore, diagnosis of LVNC cardiomyopathy should not be based solely on the presence of LV hypertrabeculations.

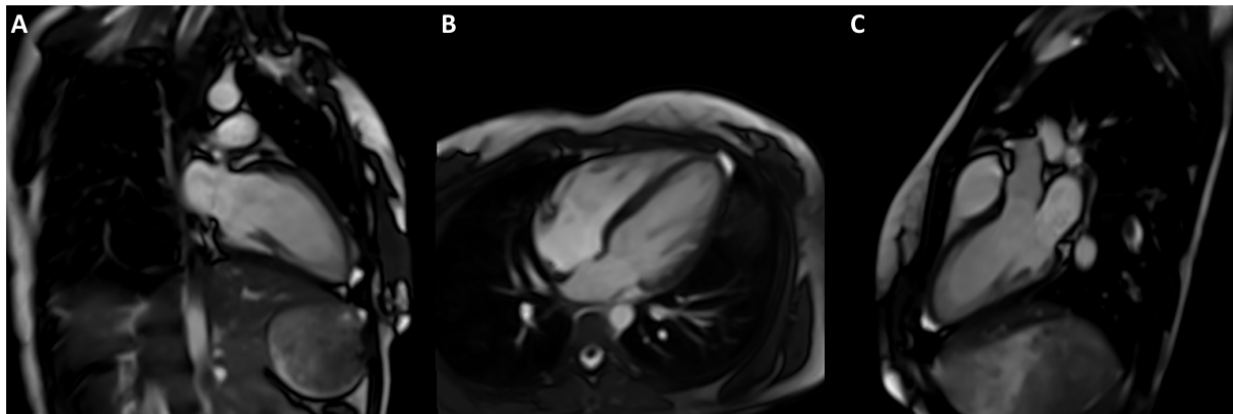


Figure 2-11: Long-axis LV CMR (A) 2-chamber (B) 4-chamber and (C) 3-chamber views showing non-compacted myocardium of the LV at the mid-ventricle and apical levels. The ratio between non-compacted: compacted endocardial layer is 3.2 (Courtesy of Dr. Soha Romeih).

2.4.6 LoF Triadin-1 (*TRDN*) variant associated with Catecholaminergic Polymorphic Ventricular Tachycardia identified at high prevalence in the EHVol cohort

To identify differences in genetic variation between Egyptian and gnomAD sub-populations, I examined whether common EHVol variants (>1%) were rarely observed in gnomAD ($\text{gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$). I performed this analysis on protein-altering variants (LoF and Missense/Indels) in all ICC genes to be able to examine differences in frequencies between both control datasets in a more comprehensive manner. As depicted

in Table 2-7, 19 variants in 14 ICC genes were commonly observed in EHVols. Among these variants 78.9% (n=15) were absent from gnomAD.

Table 2-7: Common (>1%) variants in the EHVol cohort rarely observed in gnomAD

Gene	cDNA variant	Protein variant	Variant type	EHVol freq (variants/total)	gnomAD FAF popmax
<i>ANK2</i>	c.10802T>A	p.Val3601Asp	Missense	3.25% (13/400)	0
	c.7007T>C	p.Val2336Ala	Missense	36.50% (146/400)	0
<i>CACNB2</i>	c.1803T>G	p.Asp601Glu	Missense	38.50% (154/400)	0
<i>GAA</i>	c.868A>G	p.Asn290Asp	Missense	1.50% (6/400)	0
<i>LAMA2</i>	c.5285G>A	p.Arg1762Gln	Missense	1.75% (7/400)	0.00000703
	c.6752C>A	p.Pro2251His	Missense	2.00% (8/400)	0
<i>LAMA4</i>	c.3328G>A	p.Gly1110Ser	Missense	89.75% (359/400)	0
	c.3335C>G	p.Pro1112Arg	Missense	41.25% (165/400)	0
	c.827C>A	p.Ala276Glu	Missense	99.75% (399/400)	0
<i>LDB3</i>	c.1903G>A	p.Val635Ile	Missense	10.25% (41/400)	0
<i>MYH11</i>	c.3866T>C	p.Val1289Ala	Missense	6.75% (27/400)	0
	c.5676G>C	p.Glu1892Asp	Missense	3.75% (15/400)	0
<i>NKX2-5</i>	c.395G>C	p.Arg132Pro	Missense	1.25% (5/400)	2.66E-05
<i>PRKAG2</i>	c.1475T>A	p.Ile492Asn	Missense	1.25% (5/400)	0.00002293
<i>RBM20</i>	c.419C>T	p.Pro140Leu	Missense	1.25% (5/400)	0
<i>SDHA</i>	c.1894G>A	p.Val632Ile	Missense	1.50% (6/400)	0
<i>SEPN1</i>	c.1535C>T	p.Ala512Val	Missense	2.00% (8/400)	0.00000704
<i>TRDN</i>	c.1807G>T	p.Gly603Ter	Stop gained	4.25% (17/400)	0
<i>TTN</i>	c.105180G>C	p.Glu35060Asp	Missense	2.00% (8/400)	0

With the exception of *TRDN*, all identified EHVol ICC variants were benign missense variants or lacked evidence of pathogenicity according to ClinVar

(<https://www.ncbi.nlm.nih.gov/clinvar/>) and/or dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>) databases ¹⁰¹.

The identification of a LoF *TRDN* variant (c.1807G>T) commonly observed among EHVol controls (4.25%) prompted me to study it in more detail. Triadin-1, encoded by *TRDN*, is expressed in both cardiac and skeletal muscle and is alternatively spliced to produce different isoforms ¹⁵⁴. It is a component of the Ca²⁺ release unit in the SR of cardiomyocytes as it interacts with both ryanodine receptors (*RYR1* and *RYR2*) and calsequestrin (*CASQ2*) ¹⁵⁵. Thus, preserved *TRDN* function is crucial for triggering cardiac muscle contraction. Homozygous or compound heterozygous LoF variants in *TRDN* have been shown to cause Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT), which is characterised by syncope and risk of ventricular fibrillation and SCD during physical exercise or emotional stress in the absence of structural cardiac abnormalities ¹⁵⁶. *TRDN*-null mice present with impaired excitation-contraction behaviour and arrhythmias as a result of the loss of the Ca²⁺ release unit ¹⁵⁶. The term “Triadin Knockout Syndrome” (TKOS) is now commonly used to describe patients with biallelic *TRDN* variants ¹⁵⁷. Table 2-8 summarises previously published homozygous or compound heterozygous *TRDN* variants identified in patients with TKOS.

Table 2-8: Summary of previously published *TRDN* variants identified in patients with TKOS

cDNA variant	Protein variant	Variant type	Clinical outcome
c.613C > T & c.22+29A>G	p.Q205* & p.N9fs*5	Nonsense & Splice site	Two siblings (both < 6 years of age) experienced significant arrhythmias after catecholaminergic stimulation ¹⁵⁴
c.del53_56ACAG	p.D18Afs*13	Frameshift	2-year old male died 3 weeks after initial cardiac arrest ¹⁵⁸
c.176C > G & c.613C > T	p.T59R & p.Q205*	Missense & Nonsense	26-year old male experienced exercise-induced syncope throughout his life ¹⁵⁹
c.del 572_576TAAGA	p.K147fs*0	Frameshift	6-year old patient died suddenly ¹⁵⁴
c.del 572_576TAAGA & c.22+29A>G	p.K147fs*0 & p.N9fs*5	Frameshift & Splice site	20-month old male patient experienced exercise-induced syncope and cardiac arrest ¹⁵⁸
c.del53_56ACAG & c.502G>T	p.D18Afs*13 & p.E168*	Frameshift & Nonsense	Two siblings presented with cardiac arrest at 2 years of age. Both had recurrent episodes of VF and underwent ICD implantation ¹⁶⁰
Exon 2 deletion	-	Large deletion	Male patient experienced cardiac arrest at 16 months of age and had recurrent VF ¹⁶¹

VF: ventricular fibrillation, ICD: Internal cardiac defibrillator

TRDN (ENST00000398178) consists of 41 exons, which are variably expressed in the myocardium (Figure 2-12A). The identified *TRDN* variant (c.1807G>T) is located in exon 34. It is a nonsense variant that inserts a premature termination codon (PTC) in this exon and thus is predicted to elicit nonsense-mediated decay (NMD) (Figure 2-12B) ¹⁶². The variant was identified in 17 EHVol controls in the heterozygous state only, which explains why they did not present with a CPVT phenotype.

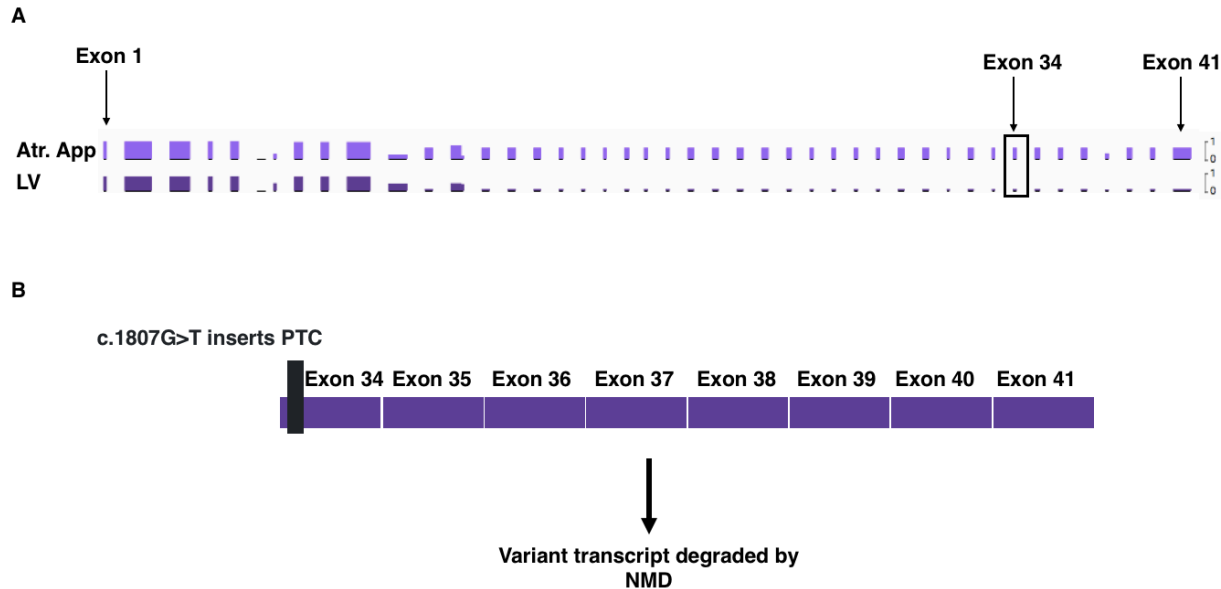


Figure 2-12: *TRDN* LoF variant (c.1807G>T) is predicted to elicit nonsense-mediated decay. (A) Differential expression of *TRDN* exons in atrial muscle (i.e. atrial appendage (Atr. App)) and ventricular muscle (shown here for left ventricle). Adapted from: <https://gnomad.broadinstitute.org/>. **(B)** Location of c.1807G>T in *TRDN* transcript and its predicted effect on expression. PTC: premature termination codon; NMD: nonsense-mediated decay.

2.4.7 Higher frequency of rare variation in HCM genes in EHVol controls compared to CHVols

To compare the background genetic variation between Egyptians and Caucasians, the proportion of EHVol and CHVol controls, respectively, with rare protein-coding variants in HCM-associated genes was calculated (Figure 2-13). Concerning sarcomere genes, both *MYBPC3*, *MYH7* accounted for the majority of rare variation in both cohorts, yet, rare variation in both genes was relatively higher in the EHVol (*MYBPC3*: 6.25%, *MYH7*: 5.75%) cohort compared to CHVol (*MYBPC3*: 2.37%, *MYH7*: 1.99%) controls (Figure 2-13A). LoF variation in *MYH7* was observed in both cohorts. In contrast to the CHVol cohort, EHVols did not carry rare variants in *MYL2*, *MYL3*, *TNNI3*, *ACTC1* and *TNNT2*, possibly due to the smaller Egyptian cohort size (n=400) compared to Caucasian controls (n=1,054).

In addition, the proportion of EHVol controls with rare variants in *ACTN2* was higher compared to Caucasians, 3.75% and 1.04%, respectively (Figure 2-13B). LoF variation was observed in *CSRP3* in the EHVol cohort only.

The highest proportion of rare variants in syndromic HCM genes was attributed to *CACNA1C* in both cohorts (Figure 2-13C). The proportion of EHVol controls with rare variants in *CACNA1C* was higher compared to CHVol controls, 3.01% and 2.28%, respectively. Rare variation in *RAF1*, *PRKAG2* and was modest in both cohorts; <2% and <1% in EHVol and CHVol controls, respectively. In the EHVol cohort, only rare synonymous variants were identified in *PTPN11*, *FHL1*, *LAMP2* and *TTR* genes.

Overall, the frequency of rare variation in the EHVol cohort (30.5%) was significantly higher compared to CHVol (14.52%) controls (Bonferroni corrected p-value = 2.33×10^{-11}).

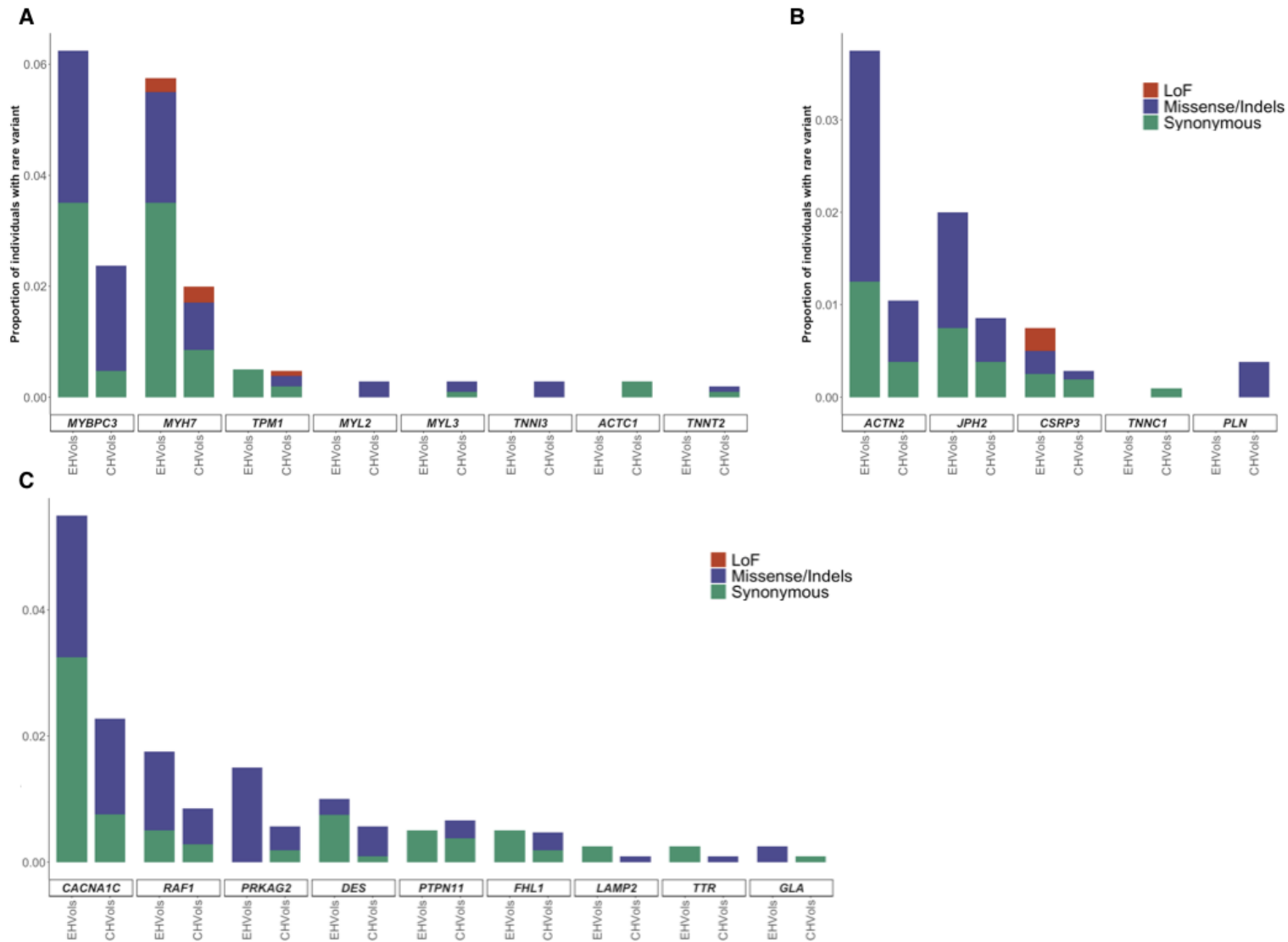


Figure 2-13: High frequency of rare variation in HCM-associated genes in the EHVol cohort. Bar charts represent the proportion of individuals with rare variants in **(A)** sarcomere genes **(B)** other definitive or likely HCM genes **(C)** syndromic genes. The first and second bars represent the EHVol and CHVol cohorts, respectively. Variants are collated by variant type (LoF, Missense/Indels and Synonymous).

2.4.8 Combining EHVol data with gnomAD results in a more accurate definition of variant rarity

The observed higher frequency of rare variation in the EHVol cohort compared to CHVol controls prompted me to assess whether combining EHVol data with the gnomAD dataset will more accurately define variant rarity. Thus, the EHVol cohort was combined with the gnomAD dataset to represent a sixth sub-population. After combining EHVol allele frequencies with the gnomAD dataset, rare variants were re-filtered using the frequency threshold: $\text{Egyptian_adjusted_gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$, where “Popmax” represented the highest FAF across the 6 sub-populations (Egyptian + 5 gnomAD sub-populations). The analysis showed that 14/8569 (0.16%) variants initially considered “rare” in reference to gnomAD (i.e $\text{gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$), were re-defined to “common” after adding Egyptian data to gnomAD (i.e. FAF of these variants was no longer $\leq 4 \times 10^{-5}$) (Table 2-9).

Table 2-9: Rare gnomAD variants re-defined to common after integrating EHVol allele frequencies into the dataset

Gene	cDNA variant	Protein variant	Variant type	gnomAD FAF popmax	Egy_adjusted gnomAD FAF popmax
<i>ACTN2</i>	c.2602G>A	p.Ala868Thr	Missense	3.50E-05	0.001022
<i>CACNA1C</i>	c.5198C>T	p.Ala1733Val	Missense	1.00E-05	0.000444
	c.5568A>G	p.Thr1856%3D	Synonymous	0*	0.001707
	c.5917C>T	p.Arg1973Trp	Missense	0*	0.000444
<i>MYBPC3</i>	c.3780C>T	p.Gly1260%3D	Synonymous	3.00E-06	0.000444
	c.3657G>T	p.Leu1219%3D	Synonymous	1.20E-05	0.002462
	c.3373G>A	p.Val1125Met	Missense	1.20E-05	0.000444
	c.305C>T	p.Pro102Leu	Missense	3.70E-05	0.000444
<i>MYH7</i>	c.4494C>T	p.Phe1498%3D	Synonymous	0*	0.001022
	c.4493T>A	p.Phe1498Tyr	Missense	0*	0.001022
	c.2637C>T	p.Ser879%3D	Synonymous	0*	0.001022
	c.1362G>A	p.Gln454%3D	Synonymous	3.00E-06	0.000444
	c.210C>T	p.Thr70%3D	Synonymous	2.80E-05	0.000444
<i>PRKAG2</i>	c.1475T>A	p.Ile492Asn	Missense	2.20E-05	0.002462

Rare variants (i.e. gnomAD $FAF_{popmax} \leq 4 \times 10^{-5}$) before integration of the EHVOL data into gnomAD are depicted. The FAF was re-calculated after integrating EHVOL allele frequencies into the gnomAD dataset (Egy_adjusted gnomAD FAF_{popmax}). *Variants with gnomAD $FAF_{popmax} = 0$ (and popmax = none) were observed only once (AC=1) in a gnomAD population and thus were not filtered for frequency ¹⁴⁹. Egy: Egypt.

2.4.9 Higher proportion of unique variants in the EHVOL cohort compared to CHVOL controls

To compare the representation of Egyptian and Caucasian background genetic variation in gnomAD, I calculated the proportion of EHVOL- and CHVOL controls, respectively, with unique protein-coding variants. EHVOL-unique variants were defined as being not only absent from gnomAD (i.e. EHVOL-only) but also from the CHVOL cohort. Similarly, CHVOL-unique variants were defined as being absent from both gnomAD and EHVOL cohorts. Then, I assessed the representation of the EHVOL-unique variants in the GME dataset (n=1,111) as it integrates WES data from the ethnically-matched North African sub-population ¹³⁴.

The proportion of EHVOL controls (19.5%) with unique variants was significantly (p-value= 2×10^{-7}) higher compared to CHVOL controls (9.11%) (Table 2-10). This highlights the underrepresentation of Egyptian individuals in the gnomAD dataset compared to Caucasians.

Table 2-10: Proportion of unique variants in the EHVOL and CHVOL cohorts

EHVOL Freq (variants/total)	CHVOL Freq (variants/total)	p-value
19.50% (78/400)	9.11% (96/1054)	2×10^{-7}

A total of 56 (n=278, 20.14%) unique variants were identified in the EHVOL cohort and only 5 of these were identified in the GME cohort (8.93%) (Table 2-11). Of the 5 variants identified in the GME, 2 were exclusively found in the North East African (NEA) sub-population *CSRP3* (p.Arg146Cys) and *MYH7* (p.Lys884del).

Table 2-11: EHVol-unique variants identified in the GME cohort

Gene	cDNA variant	Protein Variant	Variant type	EHVol freq (variants/total)	GME freq (variants/total)
<i>ACTN2</i>	c.1983C>T	p.Ala661=	Synonymous	0.50% (2/400)	0.20% (2/993)
	c.1235C>T	p.Thr412Met	Missense	3.00% (12/400)	1/31% (13/993)
<i>CSRP3</i>	c.436C>T	p.Arg146Cys	Missense	0.25% (1/400)	0.10% (1/993)
<i>JPH2</i>	c.483G>A	p.Thr161=	Synonymous	0.50% (2/400)	0.42% (4/952)*
<i>MYH7</i>	c.2649_2651	p.Lys884del	Indel	0.25% (1/400)	0.10% (1/993)

*Total number of sequenced GME individuals is lower compared to other variant sites due to lower WES coverage

2.5 Discussion

Analysis of background genetic variation in HCM-associated genes in EHVol controls showed that the Egyptian population is underrepresented in large-scale reference genetic databases. This is illustrated in the significantly higher proportion of unique variants in EHVol controls compared to Caucasians (19.5% vs. 9.11%, $p\text{-value} = 2 \times 10^{-7}$). Given the representation of the Egyptian population in the GME study, I expected EHVol-unique variants to be present in this dataset, however, only 8.93% (56/278) of these variants were observed among GME individuals. This highlights the underrepresentation of Egypt in both global and regional reference genetic databases, which impedes our ability to identify genetic risk factors for HCM. One of the main criteria for identifying putative pathogenic variants is to confirm their rarity across different populations. Therefore, it is of paramount importance to assess allele frequencies of Egyptian cases in their matched ancestry control groups.

Variant rarity, however, is not the only criterion to indicate that a variant is putatively pathogenic. Highly prevalent population-specific founder variants, which arise from a common ancestor and are passed on to subsequent generations, will be missed if only the MAF threshold is applied ¹⁶³. Thus, the EHVol resource will also aid in the identification and characterisation of common founder variants identified in Egyptian HCM patients.

Although the cohort size analysed here is still relatively small, the study provides preliminary insights into genetic variation of the underrepresented Egyptian population. Unlike large-scale reference genetic databases, EHVol controls were deeply phenotyped and thus this resource is a “true” representation of genetic variation in CVD-free individuals. In this section, I will discuss main findings obtained from the analysis of 400 EHVol controls.

2.5.1 Combining EHVol data with gnomAD results in a more accurate definition of variant rarity

The analysis showed that the overall frequency of rare variants in HCM-associated genes was significantly higher in the EHVol cohort compared to CHVol controls (30.5%

vs. 14.5%, $p\text{-value} = 2.33 \times 10^{-11}$). This raises the question whether these variants are truly “rare” in the Egyptian population. It is worth mentioning that variant rarity was defined using the frequency reported in gnomAD, which is a limitation of this study. As discussed in more detail in the next chapter, filtering for rare variants based on gnomAD is biased towards the 5 main sub-populations captured in the dataset which lack individuals from the MENA region. Hence, the observed enrichment of rare variation in the healthy EHVol cohort likely reports variants that are common in the Egyptian population rather than a higher burden of HCM-related variation, which is supported by the phenotyping data.

To address this bias, I combined EHVol genetic data with gnomAD and re-selected rare variants based on Egyptian_adjusted_gnomAD $\text{FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$. The analysis showed that combining EHVol genetic data from only 400 controls with gnomAD was sufficient to re-define 14 initially “rare” gnomAD variants to common. This finding highlights further the need to integrate genetic data from MENA populations into gnomAD in order to accurately discriminate benign and rare, putative pathogenic variants.

2.5.2 Differences in the frequency of background genetic variation between EHVol- and gnomAD controls

To identify differences in genetic variation between Egyptian and gnomAD sub-populations, I examined whether common EHVol variants ($>1\%$) were rarely observed in gnomAD. I expanded this analysis to all 174 ICC genes. The analysis showed that 19 protein-altering variants in 14 ICC genes were prevalent among EHVols but were either absent or rarely observed in gnomAD. This finding reflects once again the differences in the frequency of genetic variation between ethnically distinct populations. A study by Manrai *et al.*, 2016 showed that multiple patients of African ancestry received false positive reports stating that *TNNI3* (Lys247Arg) and *MYBPC3* (Gly278Glu) variants were pathogenic in HCM ¹⁶⁴. These variants were subsequently reclassified as “benign” due to their overrepresentation in the general population. Also, a recent study by Jones *et al.*, 2019 reported a higher burden of LoF variants in *JPH2* variation among GME and Iranian individuals compared to gnomAD controls ¹⁶⁵. The

authors also showed that there was a higher incidence of *JPH2* LoF variation in Iranian patients referred to genetic testing compared to individuals referred to Baylor Genetics Laboratories (Texas, United States), which highlights the higher risk of *JPH2*-related cardiomyopathy in individuals of the MENA region ^{165,166}. Thus, ethnic-specific variation should be accounted for in large-scale population genetic studies, including GWAS, in order to accurately identify risk factors for HCM. Currently, 78% of individuals included in GWAS are European, 10% are Asian, 2% are African and $\leq 1\%$ represent all other ethnic populations ¹⁶⁷. In an attempt to capture a much broader range of genetic diversity, the Human Heredity and Health in Africa (H3Africa) consortium, developed an array for GWAS encompassing >2 million variations from different African populations, which could be utilised to identify novel genetic associations with HCM ¹⁶⁸.

It is also worth mentioning that the identified common EHVol variants could potentially act as genetic modifiers of ICCs. Modifier variants are not the critical determinants of disease but may affect disease expressivity and/or severity. Thus, they are expected to be tolerated at high frequencies in the population. The EHVol resource could therefore provide a valuable reference for elucidating the effects of such modifier variants.

2.5.3 High rate of consanguinity in EHVol controls

Another important finding of this study was the observed relatively high rate (16.5%) of self-reported consanguinity among EHVol controls. This finding was expected given the reported high levels of consanguinity in MENA populations, including Egypt where 35% of marriages were found to be consanguineous ^{104,169}. Populations with high rates of consanguinity are enriched for homozygosity and thus may show recessive forms of HCM. For example, homozygous HCM variants in *MYL3* (p.Ala57Asp) and *TNNT2* (p.Ser179Phe) have been reported in a Tunisian and Kuwaiti family, respectively ^{170,171}. Recessive forms of HCM have generally been reported to be less common possibly due to the fact that the majority of HCM studies are being conducted in Western countries exhibiting markedly low rates of consanguinity ($<1\%$) ¹⁶⁹. Therefore, studying the effect of homozygosity in HCM-associated genes in healthy Egyptian

individuals could advance our understanding about the disease mechanism of HCM. In this analysis, 6% of EHVolS carried homozygous protein-altering variants (i.e. Missense/Indels and LoFs). These variants were “too common” (i.e. gnomAD $FAF_{popmax} > 0.00126$) for a recessive disease, which suggests that they are not disease-causing, and this is supported by the EHVol phenotyping data.

2.6 Conclusions

To advance genomic medicine, an unbiased and comprehensive understanding of human genetic diversity is required. Yet, the majority of reference genetic databases still lack information from MENA populations, which limits their utility in interpreting the clinical implications of variants identified in HCM patients. This persistent bias hampers the progression of precision medicine and exacerbates health inequalities in these populations. Despite the limited cohort size analysed in this chapter, the results described here emphasise the effect of ethnicity on genetic variation and the need to assess variant frequencies in ancestry-matched controls. Thus, expanding the ECCO-GEN project to further integrate ancestry-specific genetic and phenotypic data will support HCM variant interpretation and discovery, both of which will translate into better disease prediction and treatment of underrepresented populations.

3 EVALUATION OF HCM RARE GENETIC VARIATION IN THE EGYPTIAN POPULATION

3.1 Aims and hypotheses

The aim of this chapter is to assess the burden of rare variation in validated HCM genes in Egyptian HCM (AhcHCM) patients compared to ethnically-matched controls (i.e. EHVols) in order to study the genetic landscape of HCM. The second aim is to compare the frequency of rare variation between Egyptian and predominantly Caucasian HCM patients in order to identify classes of variants or disease mechanisms that are unique to Egyptians. The hypotheses for this chapter are as follows:

- In HCM, the patterns of rare variation in the understudied Egyptian population are different from other ethnicities. Comparing the burden of rare variation in validated HCM genes in AhcHCM patients to local controls will aid in identifying genes that are relevant to Egyptian HCM as well as their contribution to disease.
- Comparing the genetic landscape of HCM between Egyptian and predominantly Caucasian cohorts will provide more informative application of diagnostics and improved interpretation of variants detected in Egyptian patients.
- The underrepresentation of the Egyptian population in HCM genetic studies results in a higher yield of inconclusive genetic testing results compared to well-characterised predominantly Caucasian HCM patients.
- Egyptian HCM patients are enriched for homozygosity compared to Caucasians due to the high prevalence of consanguinity in Egypt. Studying the

effect of homozygous variants on the HCM phenotype could provide deeper insights into the disease mechanism.

3.2 Introduction

3.2.1 Genetic and non-genetic associations with HCM

The identification of variants in the 8 core sarcomere genes (*MYH7*, *MYBPC3*, *MYL2*, *MYL3*, *TNNT2*, *TNNI3*, *TPM1* and *ACTC1*) through linkage analysis as the underlying genetic cause of HCM in the late 1990s was a breakthrough in genomic medicine^{42,43,45,46,172,173}. Ever since these discoveries, intensive efforts have been made to identify genetic variants in the remaining ca. 50% of HCM patients without sarcomeric variants, mostly through candidate gene studies⁸³. Unlike linkage studies, candidate gene approaches require prior knowledge of the potential role of specific genes in HCM pathogenesis. Candidate genes are selected based on molecular studies proposing the involvement of a specific gene in disease¹⁷⁴. Variants are then verified for disease association by observing their presence in HCM patients and available controls. Subsequent functional validation of the disease-causing effect of the identified variant or analysis of its co-segregation with disease in families was not always performed.

Over the last 2 decades, ca. 250 variants in over 40 additional predominantly non-sarcomere genes have been proposed by candidate gene approaches to be associated with HCM⁶⁷. Many of these genes have been included in gene panels for routine clinical genetic testing of HCM. However, recent studies highlighted that expanding these panels do not necessarily increase the diagnostic yield in patients with HCM²⁸. On the contrary, the use of growing gene panels and broader whole exome- and whole genome sequencing has led to a substantial increase in VUSs which are clinically uninformative.

To reassess the association of variants in proposed genes with HCM, Walsh *et al.*, 2017 leveraged the largest consortium of human genetic data available (>60,000 exomes in ExAC) as well as up to 6,179 sequence data from HCM patients^{28,100}. The

authors set a frequency threshold in ExAC beyond which variants in previously reported candidate HCM genes ($n=20$) were considered non-pathogenic. This population minor allele frequency (MAF) cut-off ($MAF < 1 \times 10^{-4}$) was based on the frequency of the most common HCM-causing variant (*MYBPC3*, Arg502Trp) identified in the analysed clinical HCM cohort. In contrast to previous candidate gene studies, which assessed whether variants were ‘individually’ rare, without recognising that such variants were collectively common in controls, Walsh *et al.* 2017 utilised a massive control population size ($n>6,000$) to compare the aggregate frequency of all rare protein-altering variants in each HCM gene between patients and ExAC individuals. This gene-level ‘burden testing’ approach provides adequate power to identify statistical differences between patients and controls as opposed to analysing single variants.

The analysis showed that out of the 20 analysed HCM genes, only 11 showed a significant burden of rare variation in HCM patients compared to ExAC controls. These included the metabolic cardiomyopathy genes *GLA*, *LAMP2* and *PRKAG2*. As expected, the 8 sarcomere genes showed a demonstrable excess of rare variants in the HCM cohort thereby providing a proof-of-principle for the analytical approach adopted by the authors. However, a few limitations pertaining to technical differences between the targeted and whole-exome sequencing approaches as well as the absence of ancestry information of the HCM cohort could slightly affect the rare variant frequencies reported in the study. In addition, the burden testing approach could have eliminated genes with weak association with HCM even though they harbour isolated pathogenic variants. Despite these limitations, the authors confirmed the causality of a specific set of HCM genes and highlighted the need to develop a deep understanding of gene pathogenicity, in light of extensive reference population data, before inclusion of novel genes in genetic testing panels.

In a subsequent study, Walsh *et al.*, 2017 re-evaluated the association of 31 predominantly non-sarcomere genes with HCM. The authors assessed rare variation in these genes in a prospectively sequenced HCM cohort ($n=804$), previously published HCM cohorts and the published large-scale clinical HCM cohort ($n>6,000$)

in comparison to ExAC ⁶⁷. Analysis of the combined data showed that only 3 genes (*CSRP3*, *FHL1* and *PLN*) out of the 31 were supported by strong evidence of association with HCM based on segregation data (*CSRP3*), the presence of *de novo* variants (*FHL1*) or a significant enrichment of a specific variant in patients versus ExAC (all 3 genes). Thus, this study demonstrated that expanding the gene panels to include more non-sarcomere genes may not improve the yield of clinical genetic testing for HCM. In fact, 2 years later Thomson *et al.*, 2019 analysed WGS data of 51 proposed HCM genes in sarcomere-negative cases and reported that they had negligible diagnostic yield ¹⁷⁵.

In addition, other disease-specific gene curation efforts, such as The Clinical Genome Resource (ClinGen) initiative, have substantially contributed to our understanding of genetic associations with HCM ¹⁷⁶. By establishing an evidence-based framework to assess all reported gene-disease associations in a systematic manner in 2017, the ClinGen initiative has enhanced the integration of genomic information into the clinical setting ¹⁷⁶. This framework provides a semi-quantitative assessment for the strength of genetic and experimental evidence underlying a gene-disease relationship by assigning 0-18 scoring points for each association. Gene-disease associations are then classified as “Definitive” (12-18 points and replicated over time), “Strong” (12-18 points), “Moderate” (7-11 points), “Limited” (1-6 points), “No Reported Evidence” or “Conflicting Evidence” (Figure 3-1). Classifications derived with this framework are then evaluated by disease experts for confirmation ¹⁷⁶. Thus, this detailed framework provides an unbiased evidence assessment of gene-disease associations, which is crucial for precision medicine.

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 ^{Ii}			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 3-1: Classes of genetic (top) and experimental (bottom) evidence along with the suggested scoring points used in the ClinGen clinical validity framework. From Strande *et al.*, 2017, with permission ¹⁷⁷. Genetic evidence (top) is divided into case-level and case-control data and each category is assigned a range of points. For case-level data, points are assigned to each case based on the variant's mode of inheritance, evidence of disease causation, molecular consequence and segregation data. Case-control studies are divided into single- and aggregate variant analysis and points are assigned according to study quality based on different criteria, including: variant detection method, bias and statistical power. Experimental evidence (bottom) is divided into 3 categories and each category is assigned a range of points and *in vivo* data (Models & Rescue) is assigned the highest maximum score that can be achieved.

Such disease-specific curation efforts have substantially enhanced our understanding of the genetic basis of HCM as well as the sensitivity of clinical genetic testing for the disease. However, more than 50% of HCM patients do not present with a genetically identifiable aetiology (i.e. negative genetic testing result) and it remains challenging to identify non-genetic factors or disease modifiers in these patients. The impact of ‘modifier’ factors on the HCM phenotype was first proposed by Fananapazir and Epstein, 1994 after observing distinct clinical manifestations in Caucasian and Korean families with the same missense *MYH7* variant (Arg403Gln) ¹⁷⁸. Similarly, the *TNNI3* Lys183Del variant was initially observed in a patient with apical HCM whereas his son presented with classical hypertrophy ¹⁷⁹. In another study, the same variant was identified in 25 individuals from different HCM families but only 1 individual developed apical hypertrophy ¹⁸⁰. It has also been reported that HCM patients with multiple rare variants in sarcomere genes tend to present with a more severe phenotype even if the “second hit” was classified as a VUS ¹⁸¹. Thus, these studies suggest the contribution of genetic or environmental modifiers to the HCM phenotype.

Concerning common variation, genetic studies have demonstrated that polymorphisms within genes encoding components of the Renin Angiotensin Aldosterone System (RAAS), Endothelin 1 (*ET-1*) and Tumour Necrosis Factor (*TNF*) could partially explain the differences in the degree of LVH. However, findings from these studies explained the variability in LVH in a small proportion of HCM patients ^{182–186}. Currently, only 1 HCM GWAS study conducted on 174 patients and 823 controls reported a significant association between a common intronic variant in Formin Homology 2 Domain Containing 3 (*FHOD3*) and the HCM phenotype ¹⁸⁶.

Epigenetic mechanisms, which include methylation/demethylation of CpG islands, histone modifications, chromatin remodelling and non-coding RNAs (ncRNA), have also been demonstrated to modify HCM expression ^{187,188}. For example, non-coding micro RNAs (miRNA), have been shown to influence cardiac hypertrophy and fibrosis ^{189–191}. However, HCM-causing variants in the genes encoding for miRNAs have not yet been identified ²⁵.

Compelling evidence is accumulating to support the interplay between genetics and other factors including demographic features and co-morbidities. Previous studies have demonstrated differences in HCM expression among male and female patients^{183,192}. Ethnicity appears to also influence the HCM phenotype. A study by Sheikh *et al.* showed that patients of African/Afro-Caribbean ancestry presented with a significantly higher degree of LVH compared to Caucasians²². However, more studies from non-Caucasian ancestries are required to accurately address the influence of ethnicity on the HCM phenotype. In addition, ventricular loading conditions resulting from strenuous physical exercise, obesity or other co-morbid conditions including HTN and CAD, have been demonstrated to aggravate the HCM phenotype^{21,26,193}.

Taken together, these findings reflect the multifactorial nature of HCM, which explains the great inter- and intrafamilial variability in disease presentation and progression. Extensive research studies of large sample sizes, such as GWAS are warranted to understand the interaction between causal variants and modifier factors in HCM. Nonetheless, in our current era of genomic medicine where genetic testing is being routinely performed in the clinical setting, it is imperative to adopt a standardised framework for variant interpretation to allow for accurate diagnosis, risk assessment and potentially the development of targeted therapy.

3.2.2 HCM variant interpretation: current guidelines, implications and challenges

Less than a decade ago, there was a considerable variation in the terminology used for variant classification. Clinical laboratories adopted the use of different terms such as “possibly”, “likely” and “probably” to classify pathogenic variants¹⁹⁴. In addition, the paradigm used for evaluating the evidence for pathogenicity of identified variants was very variable among the different laboratories and numerous variants were reported as pathogenic without adequately controlling for background variation¹⁰⁰. This resulted in the accumulation of large numbers of weak-evidence ‘deleterious variants’ in databases that catalogue variant-disease associations, such as the Human Gene Mutation Database (HGMD)^{112,194}. The lack of standardisation of variant interpretation paradigms became even more evident upon sharing data in the ClinVar database as

many variants were inconsistently classified by the different laboratories. These misclassifications could have serious implications in the clinical setting as illustrated by Gaba *et al.*, 2016, who reported that the decision to implant implantable cardioverter-defibrillators was based on a misclassified variant as pathogenic for long QT ¹⁹⁵.

To ameliorate this issue, the American College of Medical Genetics and Genomics (ACMG) and the Association of Molecular Pathology (AMP) published a set of guidelines to standardise variant interpretation in Mendelian disease genes ¹⁰². These guidelines outline specific criteria to weight variant evidence as well as a set of rules for combining criteria to reach a variant classification of: pathogenic (P), likely pathogenic (LP), uncertain significance (VUS), likely benign (LB) or benign (B) (Figure 3-2 & Table 3-1) ¹⁰².

	BENIGN		PATHOGENIC			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population Data	BS1 - MAF is too high for disorder OR observation in controls is inconsistent with penetrance (BS2)			PM2 – Absent in population databases	PS4 - Prevalence in affecteds statistically increased over controls	
Computational And Predictive Data		BP4 - Multiple lines of computational evidence suggest no impact on gene/gene product BP1 – Missense in gene here only truncating cause disease BP3 – In-frame Indels in repeat w/out known function	PP3 - Multiple lines of computational evidence support a deleterious effect on gene/gene product	PM5 - Novel missense at an amino acid residue where a different pathogenic missense seen before PM4 - Protein length changing variant	PS1 - Same amino acid change as established pathogenic variant	PVS1 - Predicted null variant in a gene where LOF is a known mechanism of disease
Functional Data	BS3 - Well established functional studies show no deleterious effect		PP2 - Missense in gene with low rate of benign missense and path missenses common	PM1 - Mutational hotspot or well-studied functional domain without benign variation	PS3 - Well established functional studies show a deleterious effect	
Segregation Data	BS4 - Non-segregation with disease		PP1 - Co-segregation with disease in multiple affected family members	Increased segregation data →		
De novo Data				PM6 - De novo (without paternity and maternity confirmed)	PS2 - De novo (paternity and maternity confirmed)	

Figure 3-2: Evidence framework from the ACMG/AMP guidelines illustrating the different lines of evidence and strength levels. Reproduced from Richards *et al.*, 2015, with permission ¹⁰². Figure continued on the next page.

	BENIGN		PATHOGENIC			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Allelic Data		BP2 - Observed in <i>cis/trans</i> with a dominant variant		PM3 - For recessive disorders, detected in trans with a pathogenic variant		
Other Database		BP6 - Reputable source w/out shared data = benign	PP5 - Reputable source w/out shared data = pathogenic			
Other data		BP5 - Found in case with an alternate cause	PP4 – Patient's phenotype or FH highly specific for gene			

Table 3-1: Rules for combining evidence codes to classify variants

Pathogenic	(i) 1 Very strong (PVS1) <i>AND</i> (a) ≥ 1 Strong (PS1–PS4) <i>OR</i> (b) ≥ 2 Moderate (PM1–PM6) <i>OR</i> (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) <i>OR</i> (d) ≥ 2 Supporting (PP1–PP5) (ii) ≥ 2 Strong (PS1–PS4) <i>OR</i> (iii) 1 Strong (PS1–PS4) <i>AND</i> (a) ≥ 3 Moderate (PM1–PM6) <i>OR</i> (b) 2 Moderate (PM1–PM6) <i>AND</i> ≥ 2 Supporting (PP1–PP5) <i>OR</i> (c) 1 Moderate (PM1–PM6) <i>AND</i> ≥ 4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) <i>AND</i> 1 moderate (PM1–PM6) <i>OR</i> (ii) 1 Strong (PS1–PS4) <i>AND</i> 1–2 moderate (PM1–PM6) <i>OR</i> (iii) 1 Strong (PS1–PS4) <i>AND</i> ≥ 2 supporting (PP1–PP5) <i>OR</i> (iv) ≥ 3 Moderate (PM1–PM6) <i>OR</i> (v) 2 Moderate (PM1–PM6) <i>AND</i> ≥ 2 supporting (PP1–PP5) <i>OR</i> (vi) 1 Moderate (PM1–PM6) <i>AND</i> ≥ 4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) <i>OR</i> (ii) ≥ 2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) <i>OR</i> (ii) ≥ 2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met <i>OR</i> (ii) the criteria for benign and pathogenic are contradictory

Reproduced from Richards *et al.*, 2015, with permission ¹⁰².

However, several studies have demonstrated that even when adopting the ACMG/AMP guidelines, variant interpretation can be inconsistent across different laboratories ^{196,197}. The main reason behind this discordance is that these guidelines were designed to be broadly applicable to different Mendelian diseases and therefore require disease/gene specific customisation of the evidence criteria. For example, the guidelines did not quantitatively define segregation or set specific thresholds corresponding to increasing degrees of segregation ¹⁹⁶. Thus, Jarvik and Browning 2016 proposed a set of implementable quantitative guidelines for assessing co-segregation of a variant with disease to ensure accurate classification ¹⁹⁶. In addition, the expected allele frequency of putative pathogenic variants (PM2 rule) are not specified by the ACMG/AMP framework. Therefore, several studies have evaluated minor allele frequency (MAF) thresholds to aid in differentiating between putative pathogenic- and benign variants ^{198–201}. In a recent study by Whiffin *et al.*, 2017 the

authors proposed a method for calculating the maximum credible population allele frequency for any causative HCM variant, which incorporates the disease prevalence, inheritance pattern, genetic and allelic heterogeneity and disease penetrance ¹⁴⁹ (explained in detail in section 3.3.4). In addition, the ACMG/AMP guidelines did not provide detailed criteria to guide in the evaluation of functional evidence. As a result, Kelly *et al.*, 2017 have specified that for *MYH7*-associated cardiomyopathies, only mammalian variant-specific knock-in models could be used to provide strong functional evidence ²⁰². Finally, the ACMG/AMP guidelines pose challenges in classifying novel non-truncating variants in the absence of segregation and/or functional data, which, in effect, reduces their clinical actionability. To address this, Walsh *et al.*, 2019 developed a bespoke clustering algorithm to identify sarcomeric gene regions with significant clustering of rare non-truncating variants in a published large-scale HCM cohort (n>6,000) ²⁰³. For clusters with a significant excess of rare variants in HCM cases over ExAC controls the authors calculated 2 related metrics: the odds ratio (OR) with 95% confidence intervals (CI) and the etiological fraction (EF) (detailed information on OR and EF calculation is outlined in section 3.3.13). The EF was interpreted as the proportion of cases with rare variants in the HCM cluster in which the variant caused disease ²⁰³. Based on the EF findings, the authors proposed an adaptation to the ACMG/AMP guidelines by replacing the non-quantitative PM1 and PP2 rules (Figure 3-2) by a single PM1 rule with 3 evidence levels (i.e. supporting, moderate and strong) depending on pre-defined EF bins for rare non-truncating variants located in the HCM clusters (Figure 3-3). For example, rare variants located in an HCM cluster with an $EF \geq 0.95$ could activate the adapted ACMG/AMP PM1_strong rule, which would enable a novel variant identified in this cluster to be reclassified to likely pathogenic ²⁰³.

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

Figure 3-3: Proposed changes to the ACMG/AMP guidelines for the PM1 rule by Walsh *et al.*, 2019²⁰³. Adapted rules were based on the etiological fractions (EFs) derived from the relative frequencies of rare variants in HCM cases compared to ExAC controls.

Ongoing initiatives are attempting to further reduce the discordance in variant interpretation between laboratories. For example, automated tools for variant interpretation according to the ACMG/AMP guidelines, such as InterVar, CardioVAI and CardioClassifier have been developed to reduce subjective interpretation^{204–206}. CardioClassifier is a web tool that aids in variant interpretation for ICCs by automatically annotating variants across 17 computational criteria for 3 distinct lines of evidence (i.e. population-, computational and predictive- and functional data) as defined by the ACMG/AMP framework²⁰⁶. Information for these categories are derived from multiple sources including: expert disease- and gene-specific knowledge and an internal knowledge base of curated data²⁰⁶ (More details on CardioClassifier are outlined in section 3.3.11). Another important feature of this tool is that it is interactive, as it allows the user to input case-specific information (e.g. data on segregation and/or *de novo* inheritance) to further refine variant classification²⁰⁶.

Furthermore, the ClinGen initiative has developed various curation expert panels to adapt the current ACMG/AMP guidelines to specific genes and diseases. For example, the Inherited Cardiomyopathy Expert Panel (CMP-EP) has adapted the ACMG/AMP framework for the interpretation of variants identified in the cardiomyopathy gene *MYH7* (discussed in more detail in section 4.2.4)²⁰². Such efforts further enhance the accuracy of variant interpretation and minimise conflicting classifications across different laboratories.

Another crucial factor to further improve variant interpretation is to share knowledge between different diagnostic laboratories, which typically screen for disease-causing

variants in probands and their families. By collaboratively sharing gene/variant-disease associations through public databases, such as ClinVar and the Sarcomeric Human Cardiomyopathy Registry (ShaRe) (<https://theshareregistry.org/>), the yield of uninterpretable variants identified in clinical genetic testing will be greatly reduced.

Despite the unprecedented advances in improving variant interpretation in Mendelian diseases, including HCM, understudied populations are still far from benefiting from clinical genetic testing. The main reason behind this issue is that, unlike Caucasian populations whose genetic variation has been comprehensively studied, variants in understudied populations have been poorly characterised. As a result of this discrepancy, the detection rate of clinically actionable variants identified through clinical genetic testing is significantly lower compared to Caucasian patients ^{28,207}. Therefore, the aim of this research study is to define the genetic architecture of HCM in the so far neglected Egyptian population, which could advance our global knowledge of the genetic basis of the disease.

3.2.3 Current knowledge of the genetics of HCM in Africa

Africa is considered as the homeland of modern humans, who migrated from the continent within the past 50-100 thousand years and inhabited a broad spectrum of environments around the world (i.e. “Out of Africa” migration) ^{208,209}. Thus, non-African populations share a subset of the genetic variation observed in Africans, who exhibit a high degree of genetic and ethnic diversity ²⁰⁸. Despite this, populations of African descent have been greatly understudied ²¹⁰. Studying the genetic variation among diverse African populations will aid in the understanding of the genetic basis of phenotypic variation of Mendelian genetic diseases, such as HCM, in both African and non-African populations ²¹⁰. In addition, such studies may potentially facilitate the discovery of disease-causing variants whose prevalence varies across the different ethnic populations.

In an attempt to study the genetic basis of HCM in Africa, Shaboodien *et al.*, 2019 reviewed the available literature reporting the genetic determinants of disease in the region. The authors identified a total of 18 HCM genetic studies of which 15 were

conducted in South Africa whereas only 1 study each was conducted in Egypt, Morocco and Tunisia, respectively ²¹¹. The clear bias in HCM genetic studies towards South African populations; with well-established research and medical institutions, has led to the misrepresentation of African genetic and phenotypic diversity. Given the observed genetic variation among even neighbouring African populations, it would be misleading to rely on a single population as a representative of the entire African diaspora ²¹⁰. Nonetheless, these studies provided new insights into the genetic variation of HCM in Africa. For example, the first candidate gene studies to be performed in the region reported 5 *MYH7* variants (Arg249Gln, p.Arg403Trp, p.Glu499Lys, p.Arg719Gln, p.Ala797Thr), which collectively accounted for 37.5% of South African HCM cases of northern-European descent and mixed ancestry ^{212–216}. In addition, long-term follow-up of South African individuals carrying the recurrent variant in *TNNT2* (p.Arg92Trp) were reported to develop minimal hypertrophy ²¹⁷. Despite the non-severe HCM phenotype and low disease penetrance associated with the *TNNT2* variant, it was shown to account for a high incidence of SCD among carriers ²¹⁷. In a study by Kassem *et al.*, 2013, 192 Egyptian HCM cases were screened for 3 sarcomere genes: *MYBPC3*, *MYH7*, and *TNNT2* and a variant detection yield was reported in 40% of cases ²¹⁸. In addition, Jaafar *et al.*, 2016 performed targeted NGS to screen sarcomere HCM genes as well as the *FLNC* gene in 45 Tunisian patients and reported that the prevalence of rare variants in these genes was 27% ¹⁷⁰. A homozygous variant was identified in *MYL3* (p.A57D) as well as 3 novel *FLNC* variants suggesting that this recently reported HCM gene could be disease-causing among Tunisians ¹⁷⁰. Moreover, WES performed on members of a consanguineous Moroccan family with a history of paediatric HCM and SCD revealed a homozygous truncating variant in the *SLC22A5* gene, which is associated with primary carnitine deficiency, as a likely genetic cause of disease ²¹⁹. The small sample sizes analysed in these studies, however, pose challenges in accurately defining the genetic architecture of HCM in Africa.

Similarly, there is a dearth in epidemiological studies describing the prevalence of HCM across the entire African continent with only a few countries reporting the disease prevalence. For example, a study by Abegaz *et al.*, 1990 reported that HCM accounted

for 34% of all diagnosed cardiomyopathy cases in Ethiopia ²²⁰. Amoah *et al.*, 2000 demonstrated that HCM represented the third most common type of cardiomyopathy in Ghana ²²¹. Research scientists from the region have attempted to estimate the prevalence of HCM in Africa by searching for relevant literature studies. Of the 24 published hospital-based studies reporting the incidence of HCM, large-scale studies were only available from Tanzania (n=6,680), Ethiopia (n=6,275) and Malawi (n=3,908), which showed variable incidence rates of 2%, 0.3%, 0.08%, respectively ²²². Also, there is very limited knowledge on the prevalence of HCM in North Africa and so far, research studies reporting the influence of HCM variants on the clinical presentation and natural history of the disease are lacking.

Therefore, this research study serves as an initial step towards elucidating the genetic architecture of HCM in Egypt, the most populous country in the Middle East and North Africa (MENA) region.

3.3 Methods

3.3.1 Study population

3.3.1.1 Egyptian HCM cohort (AhcHCM)

Prospective unrelated Egyptian patients (n=514) with a clinical diagnosis of HCM by echocardiography (ECHO) and/or CMR imaging were recruited to the Aswan Heart Centre (AHC) and are hereafter referred to as AhcHCM patients. All participants provided informed consent, and the study was reviewed and approved by the local research ethics committee (20130405MYFAHC_CMV_20130330). DNA samples were sequenced on the Illumina Miseq or Nextseq platforms using the TruSight Cardio Sequencing kit. A subset of patients (n=174) was sequenced on the Illumina Hiseq platform with an earlier iteration of this panel comprising a 201 gene panel assay, named ICC versions 4-6, from which the TruSight Cardio was later developed. The details of the number of patients sequenced for each gene is summarised in Table 3-3. All patients were prospectively enrolled for research purposes. Ethical approval and informed consents were obtained from all patients.

Table 3-2: Summary of the HCM and reference cohorts included in the study

Cohort	Cohort Type	Cohort Size
Hypertrophic Cardiomyopathy (HCM)		
Clinical		
Laboratory of Molecular Medicine (LMM)	Clinical	632 - 2912
Oxford Molecular Genetics Laboratory (OMGL)	Clinical	807- 3267
Combined Clinical	Clinical	632 - 6179
Prospective		
RbhHCM	Research (prospective)	684
AhcHCM	Research (prospective)	514
Population Reference		
gnomAD	Global reference database	125,748
EHVol	Local reference database	400

3.3.1.2 Royal Brompton Hospital HCM (RbhHCM) cohort

A prospective cohort of predominantly Caucasian HCM patients (RbhHCM, n=684) were recruited by the Cardiovascular Biomedical Research Unit at the Royal Brompton Hospital (RBH), London. HCM diagnosis was confirmed with CMR. All patients were enrolled for research purposes. Of the 684 patients analysed in this study, 396 samples were sequenced on the Illumina Miseq and Nextseq platforms using the TruSight Cardio Sequencing Kit. The remaining 288 samples were sequenced on the Life Technologies SOLiD 5500xl platform using a custom Agilent SureSelect panel of genes associated with ICCs (Table 3-3).

3.3.1.3 Clinical HCM cohorts

Genetic data for up to 6,179 HCM patients referred to the Oxford Molecular Genetics Laboratory (OMGL) and the Partners Laboratory for Molecular Medicine (LMM) was downloaded from Walsh *et al.*, 2017 (Supplementary Table S1A and S1B online) (Table 3-3) ¹⁰⁰. The OMGL cohort comprises 3,367 unrelated patients clinically diagnosed with HCM for whom diagnostic genetic testing was performed ¹⁰⁰. Due to the expansion of the sequencing panel used for the genetic analysis of this cohort over time, not all OMGL patients were analysed for each gene. Data regarding the ethnicity of the OMGL cohort was not provided but patients were recruited from across the UK and thus are expected to be predominantly Caucasian.

Genetic data from the LMM cohort was initially published by Alfares *et al.*, 2015 ²⁸. The cohort comprises 2,912 unrelated HCM patients referred to clinical genetic testing during the time period of 2004-2014. Patients with left ventricular hypertrophy (LVH) secondary to Fabry or Danon disease as well as phenotype-negative individuals with a family history of HCM were excluded ²⁸. The sequencing panel used for the analysis of the LMM patients has also expanded over time, thus the number of patients analysed for each gene is variable. Patients of Caucasian origin constituted 62% of the LMM cohort (refer to Alfares *et al.* ²⁸ supplementary data for more details on ethnicity) but individual-level ethnicity data was not provided.

The clinical HCM cohorts are hereafter referred to as the LMM/OMGL cohort.

3.3.1.4 The EHVol control cohort

EHVol controls (n=400) were recruited from across Egypt as part of the Egyptian Collaborative Cardiac Genomics (ECCO-GEN) Project aiming to study local genetic variation in 1000 healthy individuals (more details on the EHVol cohort are outlined in section 2.3.1).

3.3.2 Targeted sequencing and bioinformatics analysis

All AhcHCM patients and EHVol controls underwent next generation targeted sequencing using the TruSight Cardio Sequencing Kit. The DNA libraries of the RbhHCM cohort were prepared by laboratory scientists of the Cardiovascular Genetics and Genomics team. The steps for targeted sequencing and bioinformatics analysis are described in detail in the previous chapter (sections 2.3.2 and 2.3.3).

3.3.3 Gene selection

To define the genetic architecture of HCM in the Egyptian population, I analysed 22 validated HCM genes, including sarcomere-, other definitive or likely HCM genes and syndromic genes (refer to Table 2-3) ¹⁴⁸. The number of HCM patients from the different cohorts as well as the number of EHVol analysed for each gene are depicted in Table 3-3.

Table 3-3: Summary of the genes evaluated in this study as well as the number of samples sequenced in each of the study cohorts.

#	Gene	HCM					EHVol
		Published		Prospective RbhHCM		Prospective AhcHCM	Illumina
		OMGL	LMM	Illumina	SOLiD	Illumina	
1	ACTC1	1535	2650	396	288	514	400
2	ACTN2	807	632	396	288	514	400
3	CACNA1C	-	-	396	288	514	400
4	CSRP3	1535	632	396	288	514	400
5	DES	-	-	396	288	514	400
6	FHL1	1535	-	329	-	400	400
7	GLA	1535	2165	396	288	514	400
8	JPH2	-	-	396	-	514	400
9	LAMP2	839	2451	396	288	514	400
10	MYBPC3	3267	2912	396	288	514	400
11	MYH7	3200	2912	396	288	514	400
12	MYL2	1535	2650	396	288	514	400
13	MYL3	1535	2650	396	288	514	400
14	PLN	1535	632	396	288	514	400
15	PRKAG2	1535	2438	396	288	514	400
16	PTPN11	-	-	396	288	514	400
17	RAF1	-	-	396	288	514	400
18	TNNC1	-	632	396	288	514	400
19	TNNI3	3135	2912	396	288	514	400
20	TNNT2	3191	2912	396	288	514	400
21	TPM1	1535	2912	396	288	514	400
22	TTR	-	632	396	288	436	400

3.3.4 Variant filtering

I identified rare protein-altering variants in the validated HCM genes using a statistical framework developed by Whiffin *et al.*, 2017 to define the disease-specific ‘maximum credible population’ allele frequency (AF) ¹⁴⁹. The maximum credible population AF was calculated as follows:

$$\text{Maximum credible population AF} = \frac{(\text{disease prevalence} \times \text{maximum allelic contribution})}{\text{Penetrance}}$$

This statistical framework is based on 2 important assumptions. Firstly, the AF of an identified ‘rare’ variant in a dominant Mendelian disease must not exceed the prevalence of the disease itself (1:500 for HCM). Secondly, it must be less frequent in the general population than in the disease cohort for it to be penetrant. In the analysis

of genetic variation of 6,179 LMM/OMGL HCM patients, Walsh *et al.*, 2017 reported that the largest proportion of patients was attributable to the pathogenic variant *MYBPC3* c.1504C4T (p.Arg502Trp) (104/6,179 patients, 1.7%; 95%CI 1.4–2.0%)¹⁰⁰. The upper bound of this proportion (2%) (max. allelic contribution) was integrated in the framework as shown below:

$$\text{Maximum credible population AF } (4 \times 10^{-5}) = \frac{0.002 \text{ (disease prevalence)} \times 0.5 \text{ (diploid)} \times 0.02 \text{ (max.allelic contribution)}}{0.5 \text{ (assumed penetrance for dominant Mendelian diseases)}}$$

Thus, HCM variants with a maximum credible population AF > 4×10^{-5} in the general population were considered “too common” to be disease-causing, assuming a disease penetrance of 50%.

The maximum credible population AF was then compared with the “Filtering AF” (FAF) available in gnomAD, which is the pre-computed AF threshold calculated for all variants in gnomAD. The FAF represents the 95% confidence interval (CI) of the Poisson distribution of a variant in a population (to account for sampling variance in gnomAD allele counts). It was calculated for each of the 5 main gnomAD sub-populations (European (non-Finnish), African, East Asian, South Asian and Admixed American (i.e. Latino)) separately and the highest AF of all gnomAD sub-populations; the “Popmax FAF”, was compared with the maximum credible population AF for HCM (4×10^{-5}). If the Popmax FAF was higher than the maximum credible population AF it was not considered potentially disease-causing.

3.3.5 Rare variation in the AhcHCM cohort compared to EHVol controls and the predominantly Caucasian HCM cohorts

I calculated the frequency of rare variants in the AhcHCM, EHVol, RbhHCM and LMM/OMGL cohorts by dividing the sum of rare variants per HCM gene by the total number of patients sequenced. I then compared the frequency of rare variation between AhcHCM patients and the following cohorts:

- a) EHVOL controls
- b) RbhHCM patients
- c) LMM/OMGL patients

Only protein-altering variants were assessed, and frequencies were calculated for truncating variants (frameshift, essential splice site, nonsense) and non-truncating variants (inframe indels, missense), separately.

3.3.6 Burden testing

I performed burden testing by calculating the excess of rare variation per analysed gene and variant type (truncating and non-truncating) in the AhcHCM cohort (case excess) over the EHVOL controls. I then evaluated the significance of the case excess using Fisher's exact test (one sided for AhcHCM patients vs. EHVols). In addition, I performed burden testing to compare the burden of rare variation per gene between the AhcHCM cohort and RbhHCM as well as LMM/OMGL cohorts. Fisher's exact test (two-sided test) was performed to assess the significance of the difference in gene burden between Egyptian and predominantly Caucasian cohorts. To account for multiple testing, Bonferroni correction was applied depending on the number of genes analysed in each comparison.

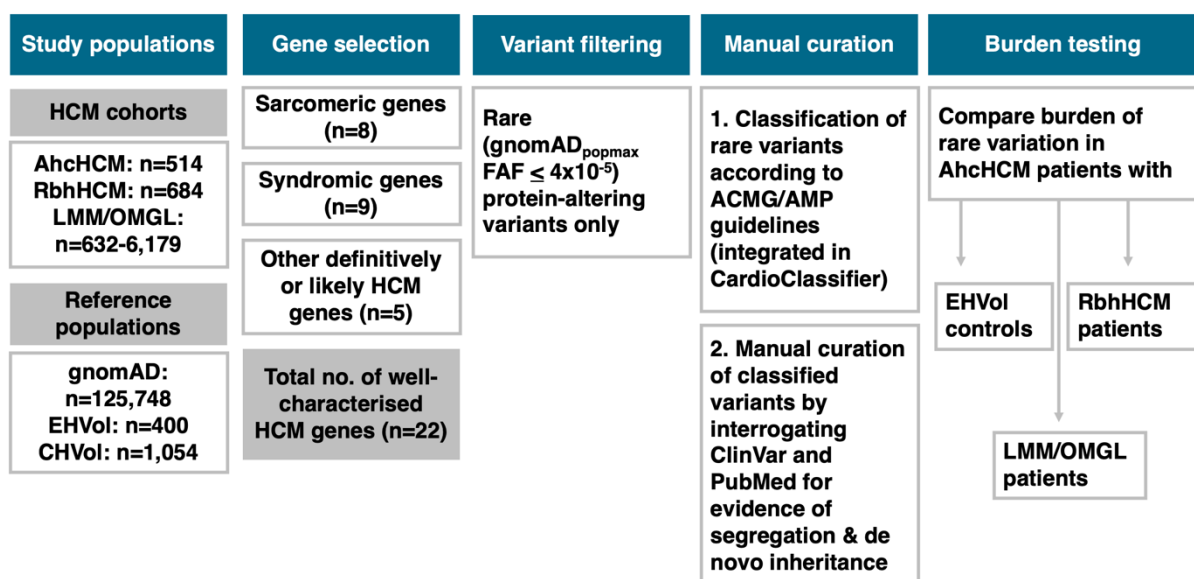


Figure 3-4: Summary of the methodology applied in the analysis of HCM genetic variation

3.3.7 Sample size calculation for power analysis

To evaluate whether the HCM cohort under study was adequately powered for testing the burden of rare variants in HCM genes, I performed a simulated power analysis to demonstrate the changes in power for a given HCM gene burden across different cohort sizes. The power analysis was performed using the `power.prop.test()` function from the “pwr” package (v.1.3-0) (<http://cran.r-project.org/web/packages/pwr/>) in the R statistical environment (v 3.6.3). A background variation frequency of 1% was assumed in this analysis with an alpha level of 0.0023 (Bonferroni corrected p-value for 22 analysed HCM genes - section 3.3.3) and a power of 80%, with one-sided testing (Figure 3-5). For a cohort size of 400 (dark green curve), there is 80% power to detect an HCM gene burden of 5.7% (dashed line) compared to 1% burden in controls.

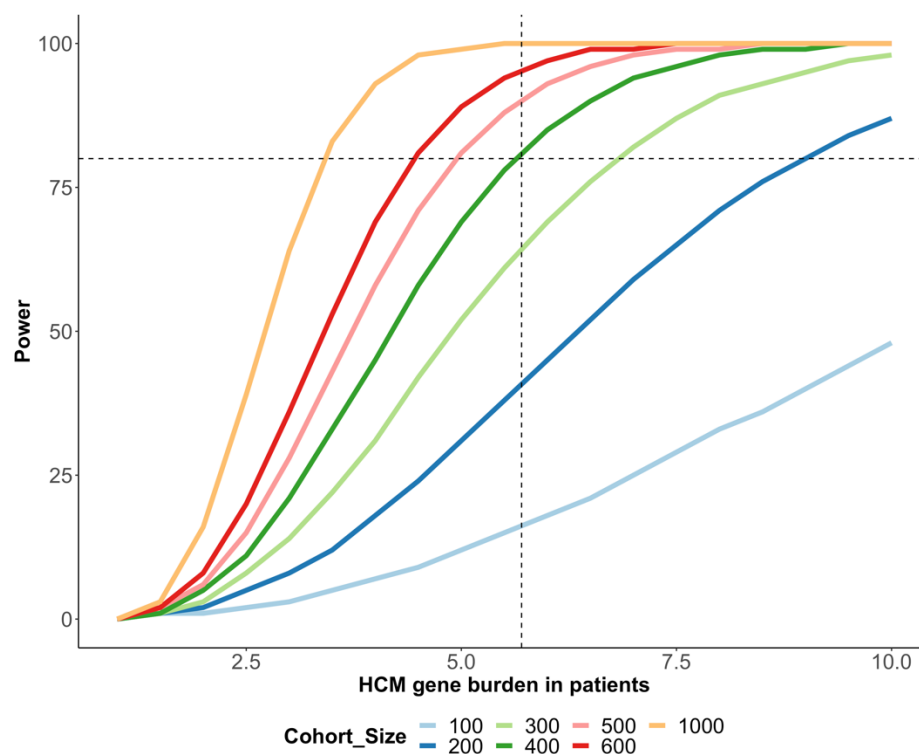


Figure 3-5: Simulated power analysis for HCM burden testing. The HCM gene burden in patients (x-axis) is plotted against power (y-axis) (for a significance level of 0.0023 to corrected for multiple testing for 22 genes) across different cohort sizes (colour-coded in legend). A background variation frequency of 1% was assumed. For a cohort size of 400 (dark green curve), there is 80% power to detect an HCM gene burden of 5.7% (dashed line) compared to 1% burden in controls.

3.3.8 Assessment of potential bias introduced by gnomAD to define rare variants in Egyptian population

To assess if the use of gnomAD to define variant rarity in the Egyptian population introduced bias in the analysis, I calculated the genomic inflation factor (GIF). First, I compared the burden of rare synonymous variants in ICC panel genes (n=174) in the Egyptian (i.e. AhcHCM patients + EHVol controls) cohort to the predominantly Caucasian (RbhHCM patients + CHVol controls) cohort, for whom synonymous data was available. Synonymous variants were selected for this analysis as they are not expected to be disease-causing and thus provide an unbiased approach to analyse inflation. Then, I evaluated the burden testing data for inflation with a quantile-quantile (Q-Q) plot and subsequently calculated the GIF based on the framework proposed by Guo *et al.*, 2018²²³. To calculate GIF, I calculated the observed (P_{obs95}) and expected (P_{obs95}) $-\log_{10}$ p-values at the 95th percentile of all genes. Then, I calculated the highest expected $-\log_{10}$ p-value among genes with p-value =1.0. I estimated the GIF as per the equation below:

$$GIF = \frac{(-\log_{10}(P_{obs95})) - (-\log_{10}(P_{obs0}))}{(-\log_{10}(P_{exp95})) - (-\log_{10}(P_{exp0}))}$$

I repeated this analysis after combining genetic data from the Great Middle Eastern (GME) population with gnomAD to serve as a sixth sub-population. After combining GME allele frequencies with gnomAD dataset, I re-filtered rare variants using the frequency threshold: $GME_adjusted_gnomAD\ FAF_{popmax} \leq 4 \times 10^{-5}$.

3.3.9 Population-matched HCM case versus control comparison of rare variation

To compare the HCM genetic architecture between Egyptian and predominantly Caucasian cohorts in a more accurate manner, I performed a population-matched case versus control comparison. Only sarcomere HCM genes and other definitely (*PLN*) or likely associated HCM genes (*ACTN2*, *CSRP3*, *JPH2* and *TNNC1*) were included in the analysis. For each gene, I compared the AhcHCM_{EHVol} excess with

RbhHCM_{CHVol} excess and evaluated the significance using Fisher's exact test (two-sided testing). For this analysis, I multiplied the case excess proportion by the respective cohort size to obtain an estimate of the number of cases predicted to carry rare, disease-causing variants in a given gene.

3.3.10 Distribution of Myosin Regulatory Light Chain 2 (*MYL2*) and Myosin Essential Light Chain 3 (*MYL3*) rare variants in the AhcHCM cohort

I visualised the distribution of rare variants in *MYL2* and *MYL3* genes in the AhcHCM cohort using the "TrackViewer" package (v 1.22.1) (<https://bioconductor.org/packages/release/bioc/html/trackViewer.html>) in R.

3.3.11 Classification and manual curation of rare variants

I classified rare variants ($\text{gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$) identified in the AhcHCM and RbhHCM cohorts into Pathogenic (P), Likely Pathogenic (LP) and Uncertain Significance (VUS) using CardioClassifier, which integrates the ACMG/AMP guidelines to aid in the interpretation of variants identified in ICC genes^{102,206}. The following rules are automatically activated by CardioClassifier:

- **PM2:** Low frequency in population databases (i.e. FAF in ExAC $\leq 4 \times 10^{-5}$). In my analysis, I used gnomAD as the reference population, so gnomAD data (<https://gnomad.broadinstitute.org>; version v2.1.1) was downloaded for HCM-associated genes and then I re-calculated the gnomAD $\text{FAF}_{\text{popmax}}$ using the R script "frequencyFilter" available in github (https://github.com/ImperialCardioGenetics/frequencyFilter/blob/master/src/pr_ecompute_exac_af_filter.R).
- **PVS1:** truncating variant in a gene that is statistically enriched in LMM/OMGL patients over ExAC controls, such as *CSRP3*, *FHL1*, *MYBPC3*, *TNNI3*, *TNNT2* and *PLN*.

- **PS4:** variant is statistically enriched in LMM/OMGL patients over ExAC controls with the rule activated if the patient count was >2 and the Fisher's exact test p-value was $< 1.79 \times 10^{-6}$ (Bonferroni correction).
- **PM4:** protein length changing variants as a result of inframe insertion/deletions or stop lost.
- **PP3:** missense variant with multiple lines of computational evidence supporting a deleterious effect. The computational prediction algorithms include: SIFT, PolyPhen2 var., LRT, Variant Taster, Variant Assessor, FATHMM, CADD and Grantham scores. This rule is activated if at least 5/8 tools predict a deleterious effect, with only 1 tool predicting benign and <3 with "unknown" classifications or if ≥ 3 tools have unknown variant classifications while all other tools predict a deleterious effect.
- **PM5/PS1:** novel missense variant where a different missense variant at the same amino acid residue is classified as pathogenic or novel missense variants with the same amino acid change as an established pathogenic variant (PS1). Novel variants are defined as "pathogenic" if multiple submitters in ClinVar confer pathogenicity with no conflicting evidence.

I then manually curated rare variants, which activate the PM2 rule, using ClinVar and PubMed resources for evidence regarding the following ACMG/AMP rules:

- **PP1:** co-segregation with disease (supporting ≥ 3 meioses, moderate ≥ 5 , strong ≥ 7)
- **PM2/PM6:** *de novo* inheritance (with/without confirmed maternity and paternity)
- **PS3:** rule relating to evidence of deleterious effects of variants was not applied as functional approaches are not well-validated for HCM variants ²⁰³

3.3.12 Comparison of the yield of clinically actionable variants between the AhcHCM- and RbhHCM cohorts

To compare the yield of clinically actionable variants (i.e. P and LP) between the AhcHCM- and RbhHCM cohorts, I divided the number of patients with P/LP variants by the total number of patients with rare variants in HCM-associated genes. For patients with multiple hits, I prioritised the variant with the highest pathogenicity (i.e. P then LP then VUS).

3.3.13 Application of the new Etiological Fraction (EF)-based American College of Medical Genetics and Genomics (ACMG)/Association or Molecular Pathology (AMP) rule on curated variants

I assessed the effect of the recently proposed EF-based ACMG/AMP rules by Walsh *et al.*, 2019 (more details in section 3.2.2) on the yield of actionable missense variants (i.e. P and LP) identified in *MYH7* in the AhcHCM cohort ²⁰³. I performed this analysis on *MYH7* only as rare missense variants in this gene showed a comparable frequency and variant distribution between the AhcHCM and LMM/OMGL cohorts. First, I calculated the frequency of rare missense *MYH7* variants in the patient-enriched HCM cluster (residues: 167-931) defined by Walsh *et al.*, 2019. To visualise the distribution of rare missense *MYH7* variants in the AhcHCM, LMM/OMGL and EHVol cohorts across the protein, I created a lollipop using the TrackViewer package (v 1.22.1) in R. Then, I calculated the odds ratio (OR) and etiological fraction (EF) for this cluster using the EHVol cohort as the reference control population. The ORs were calculated in R according to the formula provided by Altman, 1991 ²²⁴:

$$OR = \left(\frac{a}{b}\right) / \left(\frac{c}{d}\right)$$

where:

a= number of LMM/OMGL HCM patients with a rare variant in the pre-defined cluster

b= number of gnomAD controls with a rare variant in the pre-defined cluster

c= number of LMM/OMGL HCM patients without a rare variant in the pre-defined cluster

d = number of gnomAD controls without a rare variant in the pre-defined cluster

The 95% CIs for OR values were calculated as:

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

where the standard error (SE) of the log OR was given by:

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

The EF was derived from the OR and is calculated as follows:

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

95% CIs for EF values were calculated in R as described by Hildebrandt *et al.*, 2006²²⁵.

I classified *MYH7* missense variants as per the adopted new EF-based PM1_strong rule, which is activated if these variants are located in a protein region with $EF \geq 0.95$. Finally, I calculated the proportion of AhcHCM patients with VUSs that were upgraded to LP upon activation of the EF-based PM1_strong rule.

3.3.14 Proportion of AhcHCM and RbhHCM patients with unique variants

I identified the proportion of AhcHCM patients with unique sarcomeric variants; i.e. variants absent in EHVol, gnomAD, LMM/OMGL and RbhHCM cohorts. After identifying AhcHCM-unique variants, I interrogated the gnomAD database to check whether these variants had been marked in gnomAD as low quality in order to exclude them from the analysis. I applied the same approach to the RbhHCM cohort. Then, I compared the proportion of patients with unique variants between the AhcHCM and RbhHCM cohorts.

3.3.15 Genotype-phenotype correlations in AhcHCM patients with distinct genetic aetiologies

I compared the clinical phenotype between AhcHCM patients with rare homozygous variants and heterozygous variants as well as between patients with rare variants in *MYH7* and *MYBPC3*. The following parameters were included in the comparison: gender, age (at diagnosis), offspring of consanguinity, FH of HCM and SCD, maximum LVMWT (continuous variable), extreme LVMWT (≥ 30 mm) (proportion of AhcHCM patients) and the proportion of patients with LVOTO. All cardiac dimensions and functions were based on echocardiographic data. Continuous variables (age and

LVMWT) were expressed as mean $\pm$ standard deviation. I evaluated the statistical significance by student t-test. The remaining categorical variables were expressed as the number of patients and the percentage of the total number of patients for which data was available. Statistical significance was assessed by Fisher's Exact test. I performed all statistical analyses in R.

3.4 Results

3.4.1 Cohort description

The homogeneity of the Egyptian cohort (AhcHCM and EHVol) in terms of ethnicity was assessed by principal component analysis (PCA) for common SNPs (mostly exonic SNPs in ICC genes included in the TruSight Cardio Sequencing Panel) between Egyptians and the different HapMap populations to confirm their ancestry. PCA was performed using PLINK software (version 1.9) and the data was visualised using R ²²⁶. As expected, samples from AhcHCM and EHVol cohorts overlapped in 1 distinct cluster that was distinguishable from the HapMap populations (Figure 3-6). A caveat of this PCA analysis, however, is that it was performed using targeted panel data. Genome-wide data could reveal more accurate and distinct clustering.

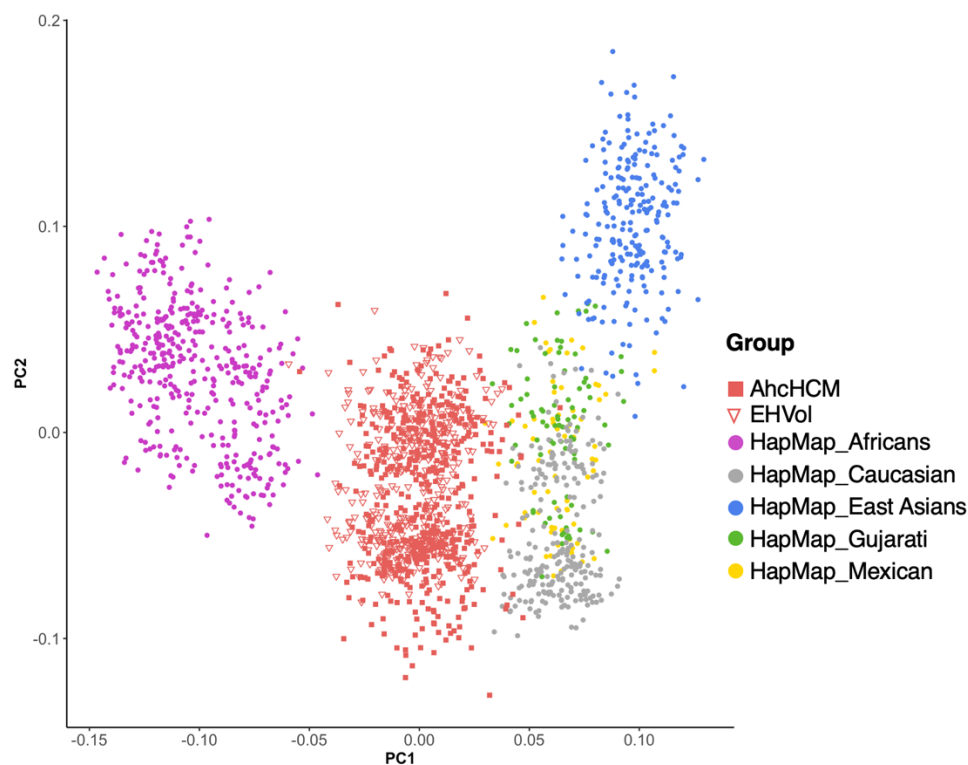


Figure 3-6: PCA plot for common SNPs between the Egyptian (AhcHCM & EHVol) and HapMap populations. x-axis denotes the value PC1, while y-axis denotes the value of PC2, with each dot in the figure representing one individual. Egyptian individuals (red) include both AhcHCM patients (squares) and EHVol controls (triangles). HapMap populations are colour-coded as shown in the legend. Both AhcHCM and EHVol clusters overlap in one distinct cluster that is distinguishable from the HapMap populations which confirms that they are well matched to each other.

Concerning the AhcHCM cohort characteristics, there were 301 male (66.7%) and 150 (33.3%) female patients. The mean age at HCM diagnosis was 34.9 years $\pm$ 17.3. The mean LVMWT (mm) (measured by echocardiography) was 23.8 $\pm$ 7.1. A detailed summary of the demographic- and clinical characteristics is outlined in Table 3-4.

Table 3-4: Demographic and clinical characteristics of AhcHCM patients (n=514)

General characteristics	AhcHCM n (%)	Available data (n)
Demographics		
Age at diagnosis (years) (mean (sd))	34.9 $\pm$ 17.3	430
Male gender n(%)	301 (66.7%)	451
Offspring of Consanguineous Marriage* (self-reported)	85 (30.6%)	278
FH of HCM	110 (25.1%)	438
FH of SCD	78 (17.8%)	438
Cardiac characteristics		
LVMWT (mm) (mean (sd))	23.8 $\pm$ 7.1	285
Extreme LVMWT (>30mm)	52 (18.2%)	285
LVOTO	217 (73.3%)	296
LVEF (%) (Echo)	73.3 $\pm$ 10.9	253
Myectomy	165 (51.2%)	322
ICD implantation	7 (2.8%)	251

*Consanguinity was defined as first-cousin marriage

3.4.2 MYH7 and Myosin Binding Protein C3 (MYBPC3) account for the majority of rare variation in the AhcHCM cohort

To evaluate the overall genetic architecture of HCM in Egypt, I compared the burden of rare variation in HCM genes between AhcHCM patients and ethnically-matched EHVol controls, irrespective of variant classification.

Burden testing was performed by calculating the frequency of rare protein-altering variants in sarcomere, syndromic and other definitive or likely HCM genes in each cohort. Then, I evaluated the significance of the excess burden of rare truncating (i.e. frameshift, essential splice site and nonsense) and non-truncating (i.e. inframe indels and missense) variants in each gene in the AhcHCM cohort over EHVol controls using Fisher's exact test. I applied Bonferroni correction to account for multiple testing: p-value $\leq$ 0.0023 for 22 sarcomere genes.

Of the 8 core sarcomere HCM genes, *MYH7* and *MYBPC3* accounted for the majority of rare variation in AhcHCM patients (Figure 3-7A). Rare variants in both genes showed a significant case excess in the AhcHCM cohort over EHVols (16.98% for *MYH7* and 13.79% for *MYBPC3*). This finding is similar to what has been reported in well-studied predominantly Caucasian HCM populations ¹⁰⁰. Rare variation was less frequent in the other sarcomere genes with only *MYL2* and *MYL3* showing a significant AhcHCM case excess of 2.52% and 2.92%, respectively.

I also assessed the burden of rare variation in other genes with definitive (*PLN*) or moderate evidence of HCM causality (*CSRP3*, *JPH2*, *ACTN2* and *TNNC1*) in the AhcHCM cohort (Figure 3-7C). Non-significant AhcHCM case excess was observed in *CSRP3* and no case excess of rare variation was observed in *ACTN2*. Non-significant case excess was also observed for *JPH2* and *TNNC1*. It should be noted however that for these 5 rare HCM genes, a larger cohort size is required to determine whether or not there is a significant excess in cases.

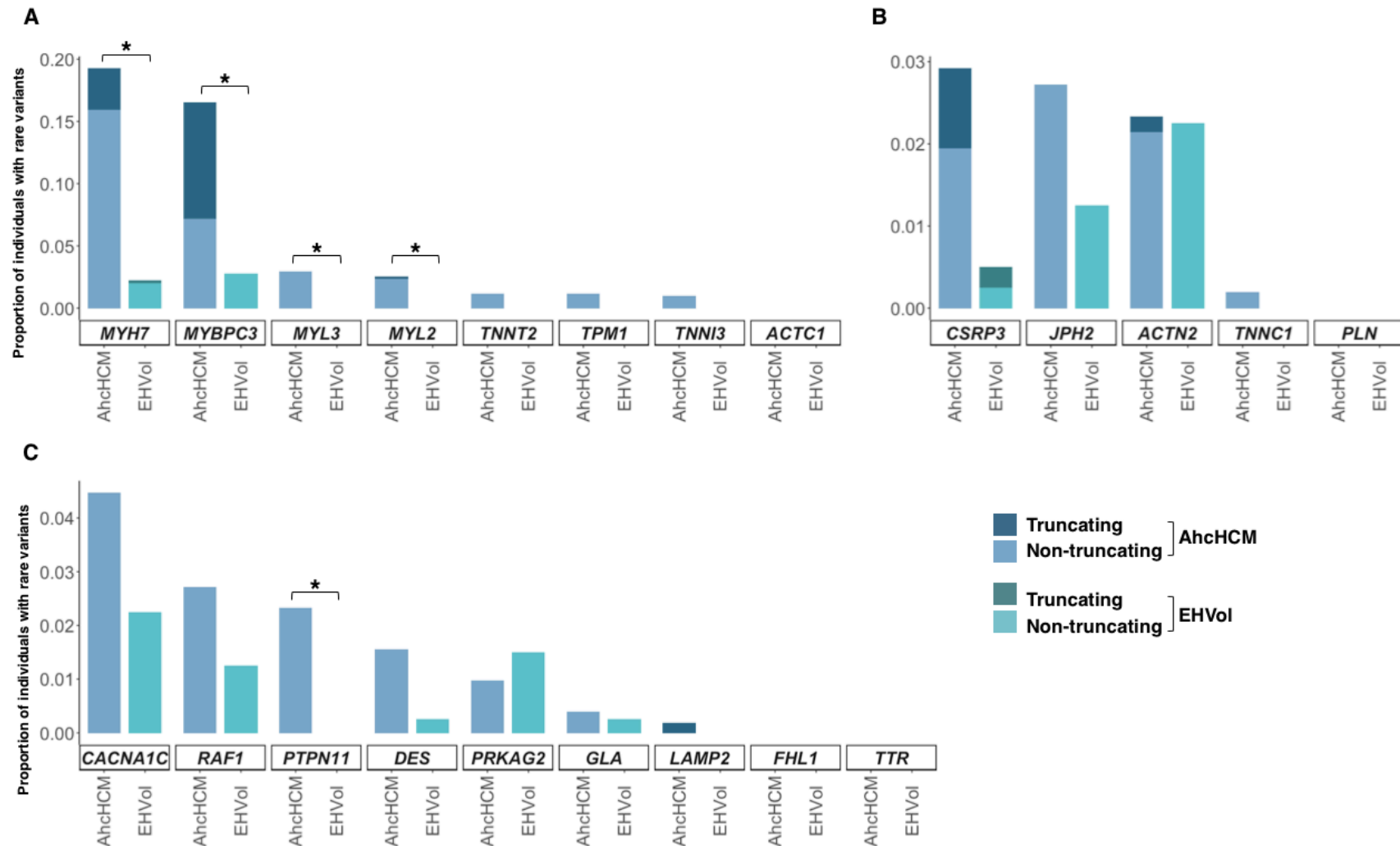


Figure 3-7: Enrichment of rare variants in *MYH7*, *MYBPC3* and Myosin light chain genes in the AhcHCM cohort compared to EHVol controls. Proportion of rare variation (truncating and non-truncating) in **(A)** sarcomeric- **(B)** other definitive or likely HCM genes and **(C)** syndromic genes associated with LVH was compared between AhcHCM patients and EHVol controls. Similar to well-characterised Caucasian HCM populations, *MYH7* and *MYBPC3* accounted for the majority of rare variation in the AhcHCM cohort. *= Significantly enriched genes in AhcHCM cohort compared to EHVol controls with Fisher's exact p-value: p-value ≤ 0.0023 after Bonferroni correction for 22 genes.

In addition, I examined the burden of rare variation in syndromic HCM genes associated with LVH in AhcHCM patients (Figure 3-7B). Rare variants in these genes were mostly attributed to *CACNA1C* (Timothy syndrome) (4.47%), albeit with non-significant AhcHCM case excess. The prevalence of rare variants in *RAF1* and *PTPN11* (Noonan syndrome) was modest, 2.72% and 2.33%, respectively. However, only *PTPN11* showed a significant excess of rare variation (2.33%) in AhcHCM patients over EHVol controls. For the syndromic genes *FHL1* and *TTR* no rare variants were observed in patients, possibly due to the limited AhcHCM cohort size.

To assess whether the excess of rare variation in AhcHCM patients was restricted to a specific variant type, I performed burden testing on truncating and non-truncating variants separately. As illustrated in Table 3-5, only *MYBPC3* and *MYH7* showed a significant excess of both rare truncating- and non-truncating variants in the AhcHCM cohort over EHVol controls. However, truncating variants showed a higher case excess in *MYBPC3* (9.34%) compared to non-truncating variants (4.45%). As expected, a higher case excess in non-truncating variants in *MYH7* (13.92%) was observed compared to truncating variants (3.06%). In addition, a significant excess of rare non-truncating variants in *MYL2* (2.33%), *MYL3* (2.92%), *PTPN11* (2.33%) was observed in AhcHCM patients over EHVol controls.

Table 3-5: Burden of rare variation in sarcomeric, syndromic and other definitive or likely HCM genes in AhcHCM patients compared to EHVol.

Gene	Variant Type	AhcHCM freq. (variants/total)	EHVol freq. (variants/total)	Case Excess	p-value
Sarcomere genes					
<i>ACTC1</i>	Non-truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>MYBPC3</i>	Non-truncating	7.20% (37/514)	2.75% (11/400)	4.45%	0.0017
	Truncating	9.34% (48/514)	0.00% (0/400)	9.34%	<0.0001
<i>MYH7</i>	Non-truncating	15.92% (82/514)	2.00% (8/400)	13.92%	<0.0001
	Truncating	3.31% (17/514)	0.25% (1/400)	3.06%	0.0004
<i>MYL2</i>	Non-truncating	2.33% (12/514)	0.00% (0/400)	2.33%	0.0009
	Truncating	0.19% (1/514)	0.00% (0/400)	0.19%	0.5624
<i>MYL3</i>	Non-truncating	2.92% (15/514)	0.00% (0/400)	2.92%	0.0002
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>TNNI3</i>	Non-truncating	0.97% (5/514)	0.00% (0/400)	0.97%	0.0558
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>TNNT2</i>	Non-truncating	1.17% (6/514)	0.00% (0/400)	1.17%	0.0312
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>TPM1</i>	Non-truncating	1.17% (6/514)	0.00% (0/400)	1.17%	0.0312
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
Other definitive or likely HCM genes					
<i>PLN</i>	Non-truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>ACTN2</i>	Non-truncating	2.14% (11/514)	2.25% (9/400)	-0.11%	0.6359
	Truncating	0.19% (1/514)	0.00% (0/400)	0.19%	0.5624
<i>CSRP3</i>	Non-truncating	1.95% (10/514)	0.25% (1/400)	1.70%	0.0165
	Truncating	0.97% (5/514)	0.25% (1/400)	0.72%	0.1785
<i>JPH2</i>	Non-truncating	2.72% (14/514)	1.25% (5/400)	1.47%	0.0923
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>TNNC1</i>	Non-truncating	0.19% (1/514)	0.00% (0/400)	0.19%	0.5624
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
Syndromic gene associated with LVH					
<i>CACNA1C</i>	Non-truncating	4.47% (23/514)	2.25% (9/400)	2.22%	0.0491
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>DES</i>	Non-truncating	1.56% (8/514)	0.25% (1/400)	1.31%	0.0443
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>FHL1</i>	Non-truncating	0.00% (0/400)	0.00% (0/400)	0.00%	1.0000
	Truncating	0.00% (0/400)	0.00% (0/400)	0.00%	1.0000
<i>GLA</i>	Non-truncating	0.39% (2/514)	0.25% (1/400)	0.14%	0.5932
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>LAMP2</i>	Non-truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
	Truncating	0.19% (1/514)	0.00% (0/400)	0.19%	0.5624
<i>PRKAG2</i>	Non-truncating	0.97% (5/514)	1.50% (6/400)	-0.53%	0.8485
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>PTPN11</i>	Non-truncating	2.33% (12/514)	0.00% (0/400)	2.33%	0.0009
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>RAF1</i>	Non-truncating	2.72% (14/514)	1.25% (5/400)	1.47%	0.0923
	Truncating	0.00% (0/514)	0.00% (0/400)	0.00%	1.0000
<i>TTR</i>	Non-truncating	0.00% (0/436)	0.00% (0/400)	0.00%	1.0000
	Truncating	0.00% (0/436)	0.00% (0/400)	0.00%	1.0000

Table 3-5 (previous page): The frequency of rare variation (non-truncating and truncating) in the AhcHCM and EHVol cohorts are shown. The excess of rare variation in the AhcHCM cohort over EHVol (case excess) is shown for each gene alongside the Fisher's exact test p-value (after Bonferroni correction): $p \leq 0.0023$. Genes with a significant case excess are highlighted in bold.

3.4.3 Key differences in the genetic architecture of HCM between Egyptian and well-characterised predominantly Caucasian cohorts

The second objective of this study was to compare the overall genetic landscape of HCM between the Egyptian- and well-characterised predominantly Caucasian HCM cohorts. To address this, I compared the burden of rare variation in AhcHCM patients with the LMM/OMGL and RbhHCM cohorts.

The majority of HCM patients in the LMM/OMGL ($n =$ up to 6,179) cohort were of Caucasian origin. These patients were recruited to the Partners Laboratory for Molecular Medicine (LMM) (US) and the Oxford Molecular Genetics Laboratory (OMGL) (UK) for clinical genetic testing. Genetic data from this cohort was downloaded from Walsh *et al.*, 2017 (Supplementary Table S1A and S1B online) ¹⁰⁰. Similar to the AhcHCM cohort, RbhHCM patients ($n = 684$) were prospectively recruited by the Cardiovascular Biomedical Research Unit at the Royal Brompton Hospital, London.

As demonstrated in Figure 3-8A, in the 3 clinical cohorts both *MYH7* and *MYBPC3* accounted for the majority of rare variation, which was less frequent in the other sarcomere genes. However, key differences in the genetic landscape of HCM were observed between Egyptian and predominantly Caucasian cohorts. For example, the frequency of non-truncating rare variation in myosin light chain genes (*MYL2* and *MYL3*) was significantly ($p\text{-value} < 0.0001$) higher in AhcHCM (5.25%) patients compared to both LMM/OMGL- (1.84%) and RbhHCM (0.73%) cohorts (Table 3-6). The burden of rare, non-truncating variants in thin filament genes (*ACTC1*, *TNNI3*, *TNNT2* and *TPM1*), however, was higher in the LMM/OMGL cohort (5.44%) compared to AhcHCM patients (3.31%), albeit not significantly.

In addition, the proportion of AhcHCM patients with truncating variants in *MYH7* (3.31%) was significantly higher ($p\text{-value} < 0.0001$) compared to both LMM/OMGL

(0.08%) and RbhHCM (0.29%) patients. This is intriguing as truncating variants in *MYH7* are generally considered non-pathogenic in HCM (discussed in detail in chapter 4). Rare variants in *MYH7* were studied in more detail and a single frameshift variant (c.5769delG) was found to account for the high burden of *MYH7* truncating variants in the AhcHCM cohort (17/514 (3.31%)) (Table 3-6). This truncating variant explains the significantly higher burden of *MYH7* truncations in the AhcHCM cohort compared to both RbhHCM and LMM/OMGL. More details about this variant are discussed in chapter 4.

The frequency of rare non-truncating variants in *ACTN2* was significantly (p-value=0.0019) higher in AhcHCM patients (2.14%) compared to LMM/OMGL patients (0.49%) (Table 3-6). In addition, a significantly (p-value<0.0001) higher burden of rare variation (both truncating and non-truncating) in *CSRP3* was observed in the AhcHCM cohort (2.92%) compared to LMM/OMGL (0.28%) (Table 3-6).

Of the syndromic LVH genes *CACNA1C*, *RAF1*, *PTPN11* and *DES* were analysed in the AhcHCM and RbhHCM cohorts only (Figure 3-8B). Non-truncating variants in *CACNA1C* showed a significantly (p-value=0.0002) higher burden of rare variation in the AhcHCM (4.47%) cohort compared to RbhHCM (1.02%). The burden of rare variants in the remaining genes was relatively higher among AhcHCM patients, albeit not significantly. *PRKAG2*, *GLA*, *LAMP2* and *FHL1* were analysed in all 3 clinical cohorts. Only *PRKAG2* showed a limited non-significant burden of rare variants in the AhcHCM cohort (0.97%) compared to LMM/OMGL (0.93%) patients (Figure 3-8B). It should be noted, however, that the LMM cohort was depleted of patients with LVH secondary to identified syndromes, which could explain the observed higher burden of rare *PRKAG2* variants in AhcHCM patients.

Taken together, this analysis demonstrated that there was a significantly higher enrichment of rare variation in the HCM genes (*MYL2*, *MYL3*, *CSRP3* and *ACTN2*) in AhcHCM patients compared to predominantly Caucasian patients. In addition, truncating variants in *MYH7* were significantly enriched in Egyptian patients. In contrast, rare variation in thin filament genes was higher in Caucasian patients

compared to Egyptians. These findings prompted me to assess whether these discrepancies truly reflect ethnic-specific variation between Egyptian and predominantly Caucasian populations or were confounded by a bias introduced by using gnomAD to define rare variants in the Egyptian cohort.

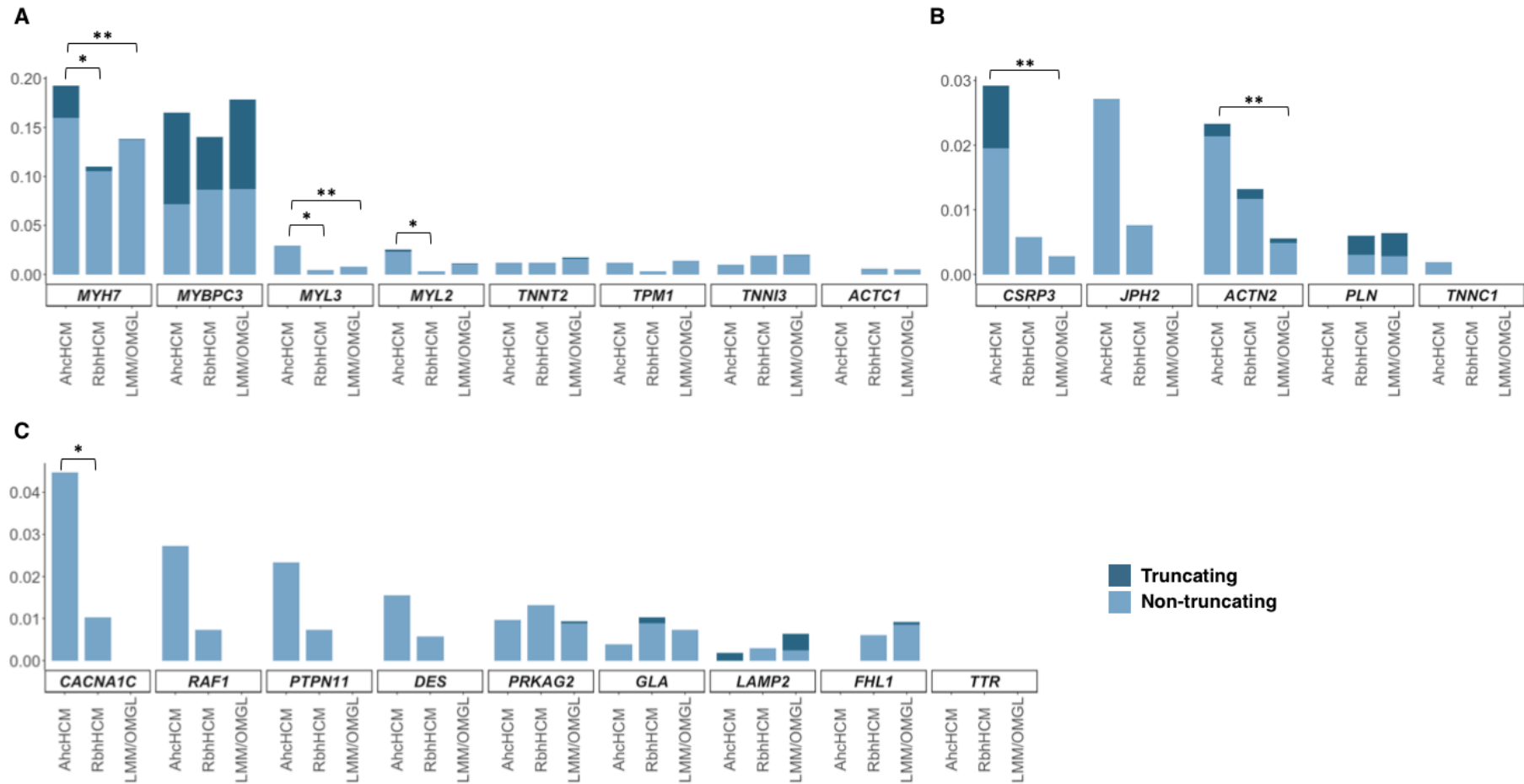


Figure 3-8: Key differences in the genetic architecture of HCM between Egyptian and Caucasian cohorts. Proportion of rare variation (truncating and non-truncating) in **(A)** sarcomeric- **(B)** other definitive or likely HCM genes and **(C)** syndromic genes associated with LVH was compared between AhcHCM patients, RbhHCM and LMM/OMGL patients. A significantly higher burden of truncating variants in *MYH7* and non-truncating variants in *MYL2* and *MYL3* genes was observed in AhcHCM patients compared to Caucasian cohorts. * = Significantly enriched genes in AhcHCM cohort compared to RbhHCM with Fisher's exact p-value: p-value ≤ 0.0023 after Bonferroni correction for 22 genes. ** = Significantly enriched genes in AhcHCM cohort compared to LMM/OMGL with Fisher's exact p-value ≤ 0.0029 after Bonferroni correction for 17 genes.

Table 3-6: Burden of rare variation in sarcomere genes, other definitive or likely HCM genes and syndromic genes associated with LVH in AhcHCM patients compared to LMM/OMGL and RbhHCM cohorts. (Table continued on next page)

Gene	Variant Type	AhcHCM freq. (variants/total)	LMM/OMGL freq. (variants/total)	Freq diff. (AhcHCM- LMM/OMGL)	p-value	RbhHCM freq. (variants/total)	Freq diff. (AhcHCM - RbhHCM)	p-value
Sarcomere genes								
<i>ACTC1</i>	Non-truncating	0.00% (0/514)	0.53% (22/4185)	-0.53%	0.1620	0.58% (4/684)	-0.58%	0.1395
	Truncating	0.00% (0/514)	0.00% (0/4185)	0.00%	1.0000	0.00% (0/684)	0.00%	1.0000
<i>MYBPC3</i>	Non-truncating	7.18% (37/514)	8.69% (537/6179)	-1.51%	0.2861	8.63% (59/684)	-1.45%	0.3912
	Truncating	9.32 % (48/514)	9.16% (566/6179)	0.16%	0.8737	5.41% (37/684)	3.91%	0.0120
<i>MYH7</i>	Non-truncating	15.95% (82/514)	13.74% (840/6112)	2.21%	0.1638	10.53% (72/684)	5.42%	0.0067
	Truncating	3.31% (17/514)	0.08% (5/6112)	3.23%	<0.0001	0.29% (2/684)	3.02%	<0.0001
<i>MYL2</i>	Non-truncating	2.33% (12/514)	1.03% (43/4185)	1.30%	0.0155	0.29% (2/684)	2.04%	0.0016
	Truncating	0.19% (1/514)	0.07% (3/4185)	0.12%	0.3709	0.00% (0/684)	0.19%	0.4290
<i>MYL3</i>	Non-truncating	2.92% (15/514)	0.81% (34/4185)	2.11%	0.0001	0.44% (3/684)	2.48%	0.0005
	Truncating	0.00% (0/514)	0.00% (0/4185)	0.00%	1.0000	0.00% (0/684)	0.00%	1.0000
<i>TNNI3</i>	Non-truncating	0.97% (5/514)	1.93% (117/6047)	-0.96%	0.1700	1.90% (13/684)	-0.93%	0.2346
	Truncating	0.00% (0/514)	0.08% (5/6047)	-0.08%	1.0000	0.00% (0/684)	0.00%	1.0000
<i>TNNT2</i>	Non-truncating	1.17% (6/514)	1.59% (97/6103)	-0.42%	0.5787	1.17% (8/684)	0.00%	1.0000
	Truncating	0.00% (0/514)	0.11% (7/6103)	-0.11%	1.0000	0.00% (0/684)	0.00%	1.0000
<i>TPM1</i>	Non-truncating	1.17% (6/514)	1.39% (62/4447)	-0.22%	0.8416	0.29% (2/684)	0.88%	0.0807
	Truncating	0.00% (0/514)	0.00% (0/4447)	0.00%	1.0000	0.00% (0/684)	0.00%	1.0000

The frequency of rare variation (non-truncating and truncating) in the AhcHCM and LMM/OMGL and RbhHCM cohorts are shown. The differences in the frequency of rare variation per gene between the AhcHCM cohort and the Caucasian cohorts (LMM/OMGL and RbhHCM) are shown alongside the Fisher's exact test p-value (after Bonferroni correction depending on number of genes analysed in each comparison). Genes with a significant burden of rare variation in AhcHCM patients compared to LMM/OMGL or RbhHCM patients (i.e. Freq diff.) are highlighted in bold. Freq: frequency; diff: difference

Gene	Variant Type	AhcHCM freq. (variants/total)	LMM/OMGL freq. (variants/total)	Freq diff. (AhcHCM- LMM/OMGL)	p-value	RbhHCM freq. (variants/total)	Freq diff. (AhcHCM - RbhHCM)	p-value
Other definitive or likely HCM genes								
<i>PLN</i>	Non-truncating	0.00% (0/514)	0.14% (3/2167)	-0.14%	1.0000	0.15% (1/684)	-0.15%	1.0000
	Truncating	0.00% (0/514)	0.18% (4/2167)	-0.18%	1.0000	0.15% (1/684)	-0.15%	1.0000
<i>ACTN2</i>	Non-truncating	2.14% (11/514)	0.49% (7/1439)	1.65%	0.0019	1.17% (8/684)	0.97%	0.2426
	Truncating	0.19% (1/514)	0.07% (1/1439)	0.12%	0.4572	0.15% (1/684)	0.04%	1.0000
<i>CSRP3</i>	Non-truncating	1.95% (10/514)	0.28% (6/2167)	1.67%	0.0002	0.58% (4/684)	1.37%	0.0532
	Truncating	0.97% (5/514)	0.00% (0/2167)	0.97%	0.0003	0.00% (0/684)	0.97%	0.0144
<i>JPH2</i>	Non-truncating	2.72% (14/514)	N/A	N/A	N/A	0.76% (3/396)	1.95%	0.0451
	Truncating	0.00% (0/514)	N/A	N/A	N/A	0.00% (0/396)	0.00%	1.0000
<i>TNNC1</i>	Non-truncating	0.19% (1/514)	0.00% (0/632)	0.19%	0.4485	0.00% (0/684)	0.19%	0.4290
	Truncating	0.00% (0/514)	0.00% (0/632)	0.00%	1.0000	0.00% (0/684)	0.00%	1.0000
Syndromic genes associated with LVH								
<i>CACNA1C</i>	Non-truncating	4.47% (23/514)	N/A	N/A	N/A	1.02% (7/684)	3.45%	0.0002
	Truncating	0.00% (0/514)	N/A	N/A	N/A	0.00% (0/684)	0.00%	1.0000
<i>DES</i>	Non-truncating	1.56% (8/514)	N/A	N/A	N/A	0.58% (4/684)	0.98%	0.1406
	Truncating	0.00% (0/514)	N/A	N/A	N/A	0.00% (0/684)	0.00%	1.0000
<i>FHL1</i>	Non-truncating	0.00% (0/400)	0.85% (13/1535)	-0.85%	0.0832	0.61% (2/329)	-0.61%	0.2033
	Truncating	0.00% (0/514)	0.07% (1/1535)	-0.07%	1.0000	0.00% (0/329)	0.00%	1.0000
<i>GLA</i>	Non-truncating	0.39% (2/514)	0.73% (27/3700)	-0.34%	0.5699	0.88% (6/684)	-0.49%	0.4781
	Truncating	0.00% (0/514)	0.00% (0/3700)	0.00%	1.0000	0.15% (1/684)	-0.15%	1.0000
<i>LAMP2</i>	Non-truncating	0.00% (0/514)	0.24% (8/3290)	-0.24%	0.6083	0.29% (2/684)	-0.29%	0.5097
	Truncating	0.19% (1/514)	0.40% (13/3290)	-0.21%	0.7084	0.00% (0/684)	0.19%	0.4290
<i>PRKAG2</i>	Non-truncating	0.97% (5/514)	0.88% (35/3973)	0.09%	0.8021	1.32% (9/684)	-0.35%	0.7872
	Truncating	0.00% (0/514)	0.05% (2/3973)	-0.05%	1.0000	0.00% (0/684)	0.00%	1.0000
<i>PTPN11</i>	Non-truncating	2.33% (12/514)	N/A	N/A	N/A	0.73% (5/684)	1.60%	0.0255
	Truncating	0.00% (0/514)	N/A	N/A	N/A	0.00% (0/684)	0.00%	1.0000
<i>RAF1</i>	Non-truncating	2.71% (14/514)	N/A	N/A	N/A	0.73% (5/684)	1.98%	0.0087
	Truncating	0.00% (0/514)	N/A	N/A	N/A	0.00% (0/684)	0.00%	1.0000
<i>TTR</i>	Non-truncating	0.00% (0/436)	0.00% (0/632)	0.00%	1.0000	0.00% (0/684)	0.00%	1.0000
	Truncating	0.00% (0/514)	0.00% (0/632)	0.00%	1.0000	0.00% (0/684)	0.00%	1.0000

3.4.4 “Rare” variants defined using gnomAD introduced bias in comparing HCM genetic architecture between Egyptian and predominantly Caucasian patients

As mentioned in section 3.3.4, the definition of variant rarity in the AhcHCM cohort was based on 5 gnomAD populations: African/African American, Admixed American, East Asian, Non-Finnish European and South Asian while data from the MENA region was lacking in the dataset. Thus, variants considered “rare” in AhcHCM patients may actually be common/non-rare in the general Egyptian population. To evaluate the bias in defining variant rarity based on gnomAD, I compared the burden of rare synonymous variants in all ICC genes between the Egyptian (i.e. AhcHCM patients + EHVol controls) and predominantly Caucasian (RbhHCM patients + CHVol controls) cohorts, for whom this data was available. I selected synonymous variants for this analysis as they are not expected to be disease-causing and thus provide an unbiased approach to analyse inflation. I evaluated burden testing data for inflation with a quantile-quantile (Q-Q) plot and then estimated genomic inflation factor (GIF) as per the formula proposed by Guo *et al.* 2018 (refer to section 3.3.8 for details on GIF calculation) ²²³.

As depicted in Figure 3-9A, the resulting distribution of the ICC single-gene p-values was characterised by a GIF of 2.770. This finding reflects an inflation of the expected- (under uniform distribution) and observed p-values (from burden testing data) above the expected GIF = 1, suggesting the presence of bias.

Given the ethnic similarity of the populations from the GME study to the Egyptian population, I then attempted to investigate whether combining GME data with gnomAD to define rare variants would reduce the observed bias. Thus, synonymous GME variants in ICC genes were combined with variants from gnomAD and the GME_adjusted_gnomAD FAF_{popmax} was calculated based on the 5 gnomAD subpopulations and the GME population. Rare synonymous variants in the Egyptian population were defined as: GME_adjusted_gnomAD $FAF_{popmax} \leq 4 \times 10^{-5}$. As shown in Figure 3-9B, this analysis revealed a marginally reduced inflation of the expected and

observed p-values with a corresponding GIF of 2.458. This minimal reduction in bias could be explained by the relatively small sample size of the GME population (n=1,111) compared to gnomAD (n=125,748).

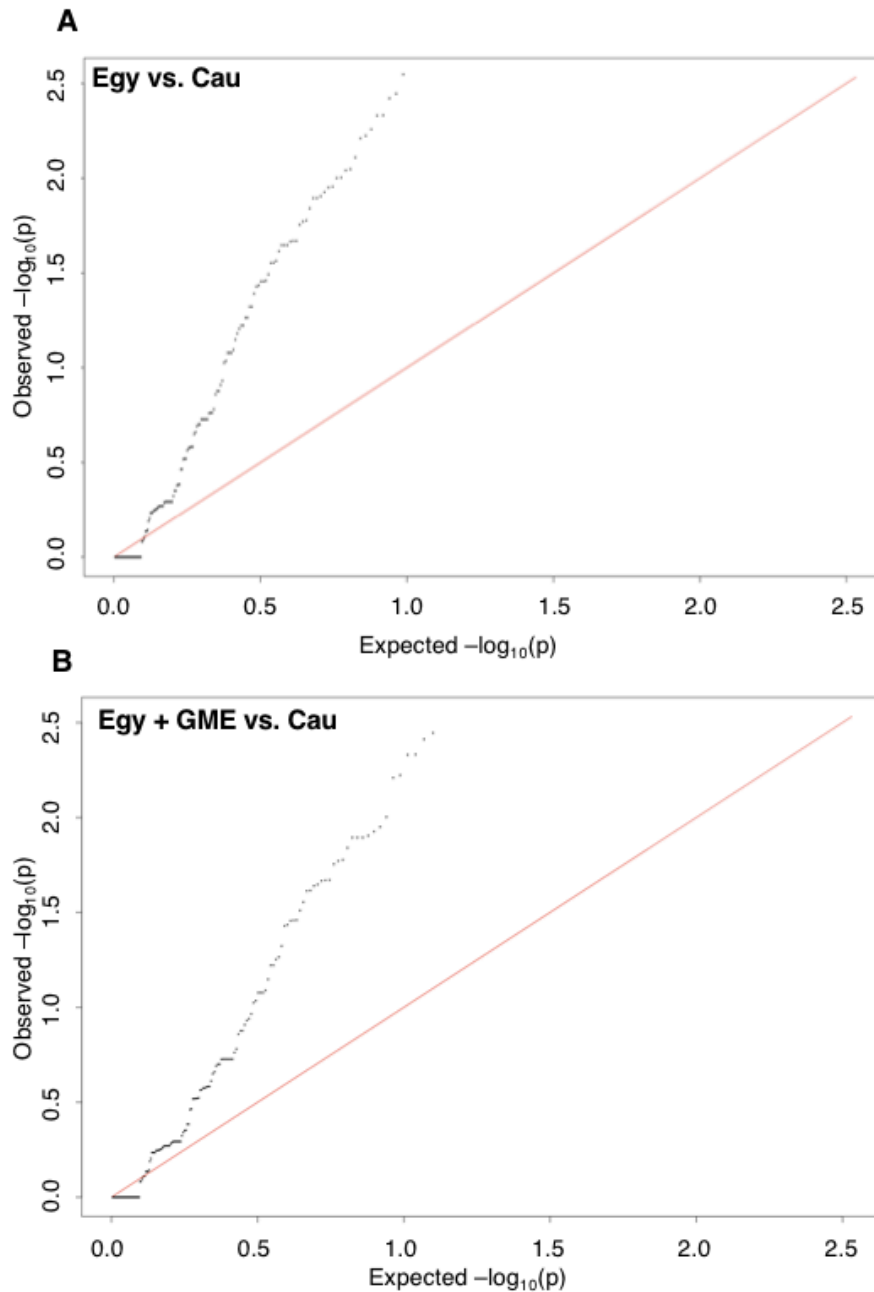


Figure 3-9: Quantile-quantile (Q-Q) plot of Egyptian vs. Caucasian burden testing data based on synonymous variants. Expected $-\log_{10}$ p-values for rare (gnomAD $FAF_{popmax} \leq 4 \times 10^{-5}$) synonymous variants, under the uniform distribution (red line), were plotted against the observed $-\log_{10}$ p-values from the burden testing data for each ICC gene. **(A)** The plot shows a discordance between expected and observed values with a genomic inflation factor (GIF) of 2.770 and **(B)** a reduction of GIF (2.419) was observed after combining GME data with gnomAD. Egy: Egyptian, Cau: Caucasian and GME: Great Middle Eastern.

The observed high GIF confirmed that the biased definition of variant rarity based on gnomAD confounded the comparison of rare variation between Egyptian and predominantly Caucasian HCM cohorts. This bias might explain the observed higher frequency of rare variants in a subset of HCM-associated genes in AhcHCM patients (section 3.4.3).

3.4.5 Significant enrichment of rare variation in Myosin Light Chain genes and *MYH7* truncating variants in AhcHCM patients compared to RbhHCM

To reduce this bias, I used population-matched controls (i.e. AhcHCM_{EHV01} excess vs. RbhHCM_{CHV01} excess) to compare the excess burden of rare variation in sarcomere genes and other definitely (*PLN*) or likely associated HCM genes (*ACTN2*, *CSRP3*, *JPH2* and *TNNC1*) (Figure 3-10). Given the limited number of variants identified in *MYL2* and *MYL3*, I combined the variants as a single group. Similarly, rare variants in *ACTC1*, *TNNI3*, *TNNT2* and *TPM1* were combined as a single group (i.e. thin filament genes).

Consistent with the previous findings described in section 3.4.3, this analysis showed a significantly higher excess burden of *MYH7* truncating variants (*MYH7*tsvs) in AhcHCM patients compared to RbhHCM (3.06% vs. 0.01%, p-value<0.0001) (Table 3-7). In addition, a significantly higher enrichment of non-truncating variants in *MYL2* and *MYL3* genes combined was observed in AhcHCM patients compared to RbhHCM (5.25% vs. 0.35%, p-value<0.0001). These findings suggest that the previously observed significantly higher burden of rare variation of *MYH7*tsvs and *MYL2+MYL3* non-tvs in AhcHCM patients compared to RbhHCM reflect true differences in the genetic architecture of HCM between both populations.

Concerning the other likely associated HCM genes, a significantly higher excess burden of non-truncating *JPH2* variants was observed in AhcHCM patients compared to RbhHCM (1.47% vs. -0.03%, p-value=0.001) (Table 3-7). In addition, the analysis confirmed the enrichment of both truncating and non-truncating *CSRP3* variants in the Egyptian cohort (Table 3-7). Intriguingly, this analysis showed a lower *ACTN2*

AhcHCM_{EHVoi} excess (-0.11%) compared to RbhHCM_{CHVoi} excess (0.51%) even though the previous analysis (Table 3-6) demonstrated a higher burden of *ACTN2* non-truncating variants in AhcHCM patients compared to both LMM/OMGL- (0.49%, p-value=0.0019) and RbhHCM (1.17%, p-value=0.2426) cohorts. This finding suggests that “rare” *ACTN2* non-truncating variants identified in AhcHCM patients may actually be common in Egyptian controls. Thus, integrating population-specific controls into this analysis provides a more accurate comparison of the genetic architecture of HCM between Egyptian and predominantly Caucasian populations.

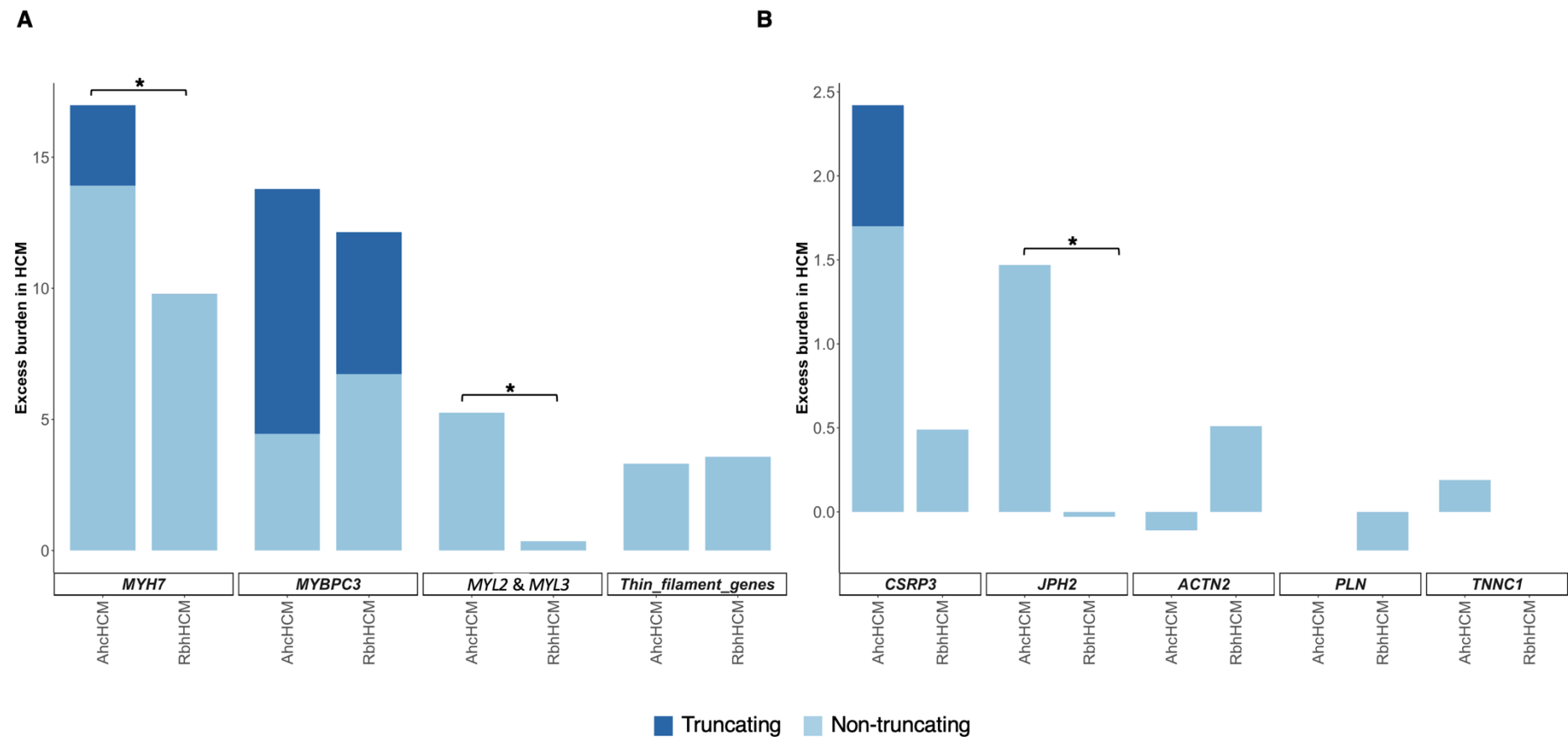


Figure 3-10: Population-matched case versus control comparison of rare variation between AhcHCM_{EHV01} excess and RbhHCM_{CHV01} excess reduces the bias introduced by gnomAD to define variant rarity and confirms the enrichment of truncating variants in *MYH7* and non-truncating variants in *MYL2*, *MYL3* and *JPH2* genes in AhcHCM patients compared to RbhHCM. Excess burden in HCM cases shown for truncating and non-truncating variants in (A) sarcomeric- and (B) other definitive or likely HCM genes.

Table 3-7: Population-matched cases versus control excess frequencies for sarcomeric- and other definitive or likely HCM genes

Gene	Variant type	AhcHCM _{EHVol} excess	RbhHCM _{CHVol} excess	Excess diff.	p-value
Sarcomere genes					
MYH7	Non-truncating	13.92%	9.77%	4.15%	0.03
	Truncating	3.06%	0.01%	3.05%	<0.0001
<i>MYBPC3</i>	Non-truncating	4.45%	6.73%	-2.28%	0.10
	Truncating	9.34%	5.41%	3.93%	0.01
MYL2&MYL3	Non-truncating	5.25%	0.35%	4.90%	<0.0001
<i>Thin filament</i>	Non-truncating	3.31%	3.57%	-0.26	0.87
Other definitive or likely HCM genes					
<i>PLN</i>	Non-truncating	0.00%	-0.23	0.23%	0.43
<i>CSRP3</i>	Non-truncating	1.70%	0.49%	1.21%	0.04
	Truncating	0.72%	0.00%	0.72%	0.03
<i>ACTN2</i>	Non-truncating	-0.11%*	0.51%	-0.62%	0.26
JPH2	Non-truncating	1.47%	-0.03%	1.44%	0.001

The AhcHCM and RbhHCM excess frequencies of rare variants (non-truncating and truncating) per gene over EHVol- and CHVol controls, respectively are shown. The significance of the excess difference between AhcHCM and RbhHCM cohorts was assessed using Fisher's exact test p-value=0.0038 after correcting for multiple testing for 13 genes. *) Negative AhcHCM_{EHVol} excess (-0.11%) for *ACTN2* so the number of individuals estimated to carry rare non-truncating variants in this gene was set to 0 to be able to perform Fisher's exact test.

Given the observed significantly (p-value<0.0001) higher enrichment of rare *MYL2* and *MYL3* variants in AhcHCM patients (AhcHCM_{EHVol} excess: 5.25%) compared to RbhHCM (RbhHCM_{CHVol} excess: 4.90%), I analysed these variants in more detail. In *MYL2* and *MYL3*, 5 and 6 rare missense variants were identified, respectively. All variants were absent from the EHVol cohort, however no significant excess burden of these variants in AhcHCM patients over controls was observed, possibly due to the small cohort size. All variants were also absent in the GME and Iranome databases.

I also analysed the distribution of the identified variants along the gene (Figure 3-11). Consistent with literature findings, the 4/6 *MYL3* variants clustered in exons 4 and 5²⁰³. *MYL2* variants clustered in exon 5.

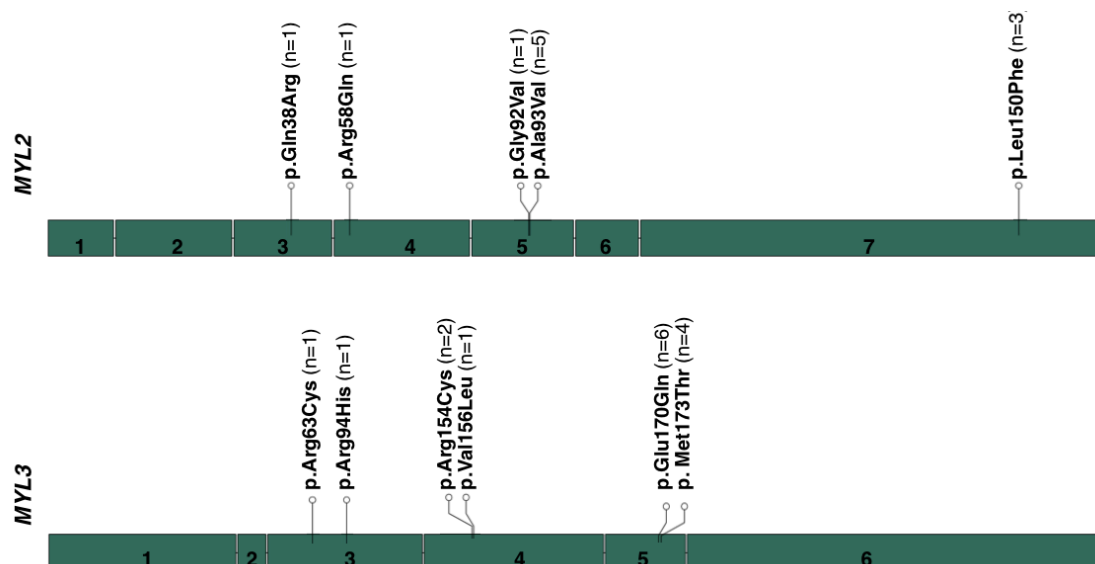


Figure 3-11: Distribution of rare *MYL2* (top) and *MYL3* (bottom) variants along the gene. The protein variant as well as the number of AhcHCM patients with each variant is shown. *MYL2* variants were not confined to a specific region of the gene whereas 4/6 *MYL3* variants clustered in exons 4-5.

I compared the frequency of each variant between the AhcHCM cohort and both LMM/OMGL and RbhHCM cohorts (Table 3-8). All *MYL2* variants were absent from the RbhHCM cohort and 3 variants (p.Gly92Val, p.Arg58Gln, p.Gln38Arg) were observed in the LMM/OMGL cohort. However, only p.Leu150Phe and p.Ala93Val were significantly (p-values: 0.0013 and <0.0001) enriched in AhcHCM patients compared to LMM/OMGL. Of the 6 *MYL3* variants, only 1 variant (p.Arg154Cys) was identified in both the RbhHCM- and LMM/OMGL cohorts. In addition, 2 variants (p.Met173Thr and p.Glu170Gln) were significantly (p-value ≤ 0.0001) enriched in AhcHCM patients compared to LMM/OMGL patients.

Table 3-8: Frequency of rare missense *MYL2* and *MYL3* variants observed in the AhcHCM cohort compared to LMM/OMGL- and RbhHCM patients

Gene	cDNA variant	Protein variant	gnomAD FAF popmax	ACMG classification	AhcHCM freq	LMM/OMGL freq	Freq diff.	p-value	RbhHCM freq	Freq diff.	p-value
<i>MYL2</i>	c.450G>T	p.Leu150Phe	0	VUS	0.58%	0.00%	0.58%	0.0013	0.00%	0.58%	0.0787
	c.278C>T	p.Ala93Val	0	VUS	0.97%	0.00%	0.97%	<0.0001	0.00%	0.97%	0.0144
	c.275G>T	p.Gly92Val	0	VUS	0.19%	0.02%	0.17%	0.2068	0.00%	0.19%	0.4290
	c.173G>A	p.Arg58Gln	0	LP	0.19%	0.17%	0.22%	0.2577	0.00%	0.39%	0.1839
	c.113A>G	p.Gln38Arg	0	VUS	0.19%	0.00%	0.19%	0.1094	0.00%	0.19%	0.4290
<i>MYL3</i>	c.518T>C	p.Met173Thr	0	VUS	0.78%	0.00%	0.78%	0.0001	0.00%	0.78%	0.0337
	c.508G>C	p.Glu170Gln	0	VUS	1.17%	0.00%	1.17%	<0.0001	0.00%	1.17%	0.0061
	c.466G>C	p.Val156Leu	0	VUS	0.19%	0.02%	0.17%	0.2068	0.00%	0.19%	0.4290
	c.460C>T	p.Arg154Cys	2.10E-05	VUS	0.39%	0.07%	0.32%	0.0119	0.15%	0.24%	0.5800
	c.281G>A	p.Arg94His	0	VUS	0.19%	0.10%	0.09%	0.4398	0.00%	0.19%	0.4290
	c.187C>T	p.Arg63Cys	3.00E-06	VUS	0.19%	0.02%	0.17%	0.2068	0.00%	0.19%	0.4290

Variants significantly enriched in AhcHCM patients compared to LMM/OMGL- or RbhHCM patients are shown in bold. Bonferroni correction was applied to account for multiple testing: $p\text{-value} \leq 0.0045$ (11 variants)

It is worth mentioning that 3 *MYL2* and 2 *MYL3* variants, respectively, were exclusively found in the AhcHCM cohort. These variants were classified as VUS as the current ACMG/AMP guidelines interpret uncharacterised missense variants as “uncertain significance”. Thus, I interrogated ClinVar and PubMed to assess if any of these variants were previously reported to contribute to HCM pathogenesis. Table 3-9 summarises the findings from previously published reports (if available).

Table 3-9: Summary of reports of AhcHCM-unique *MYL2* and *MYL3* variants from PubMed and ClinVar

Gene (protein variant)	Authors/ Submitters	Summary of PubMed/ClinVar findings
<i>MYL2</i> (p.Leu150Phe)	Not reported	
<i>MYL2</i> (p.Ala93Val)	Berge and Leren, 2013 ²²⁷	Variant was identified in a Norwegian patient with HCM. Age at disease presentation was not provided
<i>MYL2</i> (p.Gln38Arg)	GeneDx SCV000208853.9	Variant was submitted to ClinVar in 2019 as LP but evidence supporting variant pathogenicity is not provided
<i>MYL3</i> (p.Met173Thr)	Not reported	
<i>MYL3</i> (p.Glu170Gln)	Not reported	

3.4.6 Significantly higher proportion of Variants of Uncertain Significance (VUSs) in the AhcHCM cohort compared to RbhHCM

In the previous analyses, I assessed the burden of rare variation in both Egyptian and predominantly Caucasian cohorts irrespective of variant classification (P, LP, VUS). I then sought to compare the distribution of rare variants in HCM-associated genes by variant classification between the AhcHCM cohort and RbhHCM, for which individual-level data was available. Given that the Egyptian population is poorly represented in HCM genetic studies, I hypothesised that the yield of clinically actionable variants (i.e. P and LP) in AhcHCM patients would be lower compared to the RbhHCM cohort.

I classified rare HCM variants in both cohorts according to the ACMG/AMP rules, which are implemented by CardioClassifier (205). I then manually curated rare variants for published evidence regarding co-segregation and *de novo* inheritance by interrogating both ClinVar and PubMed resources.

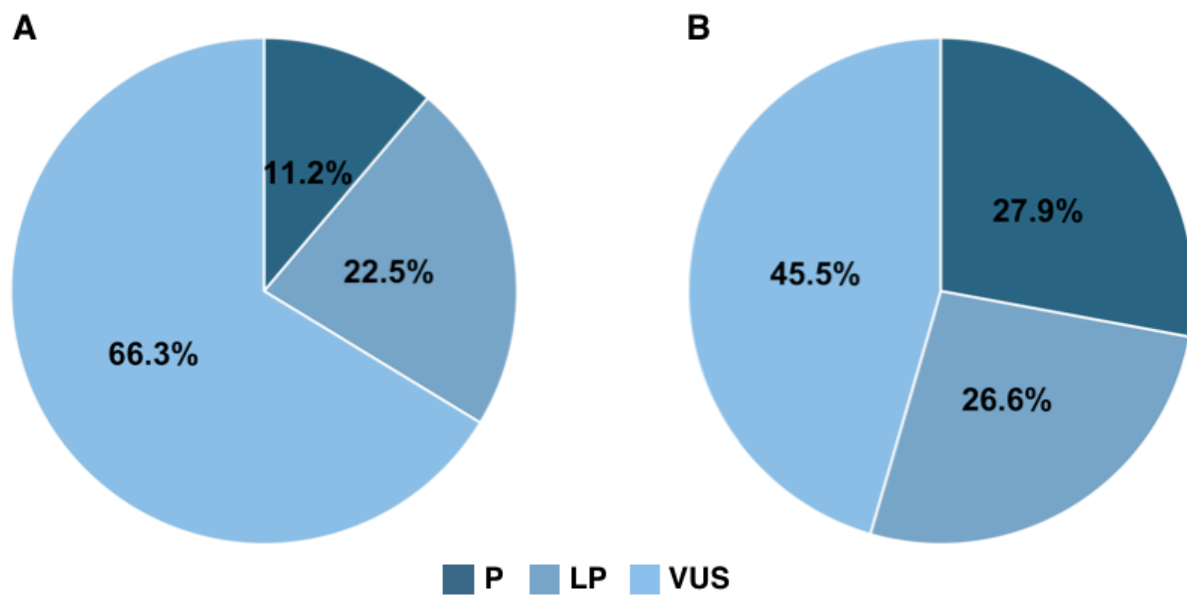


Figure 3-12: Significantly higher yield of VUSs in the (A) AhcHCM cohort compared to (B) RbhHCM. Distribution of rare variants in HCM-associated genes by variant classification (P, LP, VUS) in each cohort is shown.

As demonstrated in Figure 3-12, a significantly ($p\text{-value} < 0.0001$) higher yield of VUSs was observed in the AhcHCM cohort (66.3%) compared to RbhHCM (45.5%). This finding highlights the challenge in interpreting VUSs in understudied populations, which reduces their clinical actionability. As demonstrated in the subsequent section, applying quantitative statistical approaches to interpret VUSs could aid in the identification of specific variants with high likelihood of pathogenicity.

It is important to note that, given the observed bias in defining rare AhcHCM variants based on gnomAD (section 3.4.4), a subset of the identified VUSs could actually be common among EHVol controls. Thus, I examined the frequencies of AhcHCM VUSs in the EHVol cohort. Due to the limited EHVol cohort size ($n=400$), variants identified in at least 2 EHVol controls ($\text{MAF} \geq 0.0025$) were considered common. Only 2 AhcHCM VUSs (*MYBPC3* p.Pro102Leu) and (*PRKAG2* p.Ile492Asn) were observed in 2 and 5 EHVol controls, respectively. Thus, these variants were re-classified to benign variants resulting in a reduction of the yield of VUSs from 66.3% to 65.6%. Expanding the number of sequenced EHVol controls could further reduce the yield of misclassified VUS variants identified in AhcHCM patients.

According to the ACMG/AMP guidelines, variants statistically enriched in patients over controls (PS4 rule) provide strong evidence of pathogenicity. This rule is automatically implemented in CardioClassifier, but it applies to rare variants enriched in LMM/OMGL HCM cohorts versus ExAC. Thus, I sought to assess whether the implementation of an Egypt-specific PS4 rule could reduce the observed high yield of VUSs in the AhcHCM cohort. First, I identified VUSs variants that were common (i.e. ≥ 3) among AhcHCM cases. Then, I performed burden testing to examine whether these variants were statistically enriched in AhcHCM cases over EHVol controls. As depicted in Table 3-10, 14 variants in 6 HCM-associated genes were found to be prevalent (≥ 3) among AhcHCM cases. However, only 1 variant (*MYH7* c.5769delG) was significantly enriched in the AhcHCM (n=17) cohort (case excess: 3.31%, p-value<0.0001). This variant was upgraded to LP leading to an increase in the yield of clinically actionable variants by 4.5%.

It is important to note that the remaining VUSs, which did not show a significant excess in AhcHCM cases, could still be upgraded to LP if evidence from segregation and/or functional studies are provided.

Table 3-10: Summary of Variants of Uncertain Significance (VUS) prevalent (≥ 3) among AhcHCM patients

Gene	cDNA variant	Protein variant	Variant type	gnomAD FAF popmax	AhcHCM freq (variants/total)	EHVol freq (variants/total)	Case Excess	p-value
<i>MYBPC3</i>	c.3617G>T	p.Gly1206Val	missense	0	0.58% (3/514)	0% (0/514)	0.58%	0.1774
	c.2908C>T	p.Arg970Trp	missense	0	0.58% (3/514)	0.25% (1/514)	0.33%	0.4111
	c.2309_2314delACG TGC	p.Asp770_Pro772delin sAla	Inframe indel	0	0.58% (3/514)	0% (0/514)	0.58%	0.1774
<i>MYH7</i>	c.5769delG	p.Ser1924AlafsTer9	frameshift	0	3.31% (17/514)	0% (0/514)	3.31%	<0.0001
	c.3622G>C	p.Asp1208His	missense	0	0.78% (4/514)	0% (0/514)	0.78%	0.0995
	c.2665C>T	p.Leu889Phe	missense	0	0.78% (4/514)	0% (0/514)	0.78%	0.0995
	c.925G>A	p.Asp309Asn	missense	1.10E-05	0.97% (5/514)	0% (0/514)	0.97%	0.0558
	c.665A>G	p.Gln222Arg	missense	0	0.78% (4/514)	0% (0/514)	0.78%	0.0995
<i>MYL2</i>	c.450G>T	p.Leu150Phe	missense	0	0.58% (3/514)	0% (0/514)	0.58%	0.1774
	c.278C>T	p.Ala93Val	missense	0	0.97% (5/514)	0% (0/514)	0.97%	0.0558
<i>MYL3</i>	c.518T>C	p.Met173Thr	missense	0	0.78% (4/514)	0% (0/514)	0.78%	0.0995
	c.508G>C	p.Glu170Gln	missense	0	1.17% (6/514)	0% (0/514)	1.17%	0.0312
<i>PTPN11</i>	c.1403C>T	p.Thr468Met	missense	0	0.58% (3/514)	0% (0/514)	0.58%	0.1774
<i>RAF1</i>	c.601A>G	p.Ile201Val	missense	1.20E-05	0.78% (4/514)	0.25% (1/514)	0.53%	0.2745

Only the *MYH7* frameshift variant c.5769delG (bold) showed a significant case excess over EHVol controls. Bonferroni correction was applied to account for multiple testing: p-value<0.0036 (14 variants)

3.4.7 Application of the new EF-based ACMG/AMP rules increased the yield of clinically actionable variants identified in *MYH7* in the AhcHCM cohort

The study by Walsh *et al.*, 2019 showed that the application of the EF-based ACMG/AMP rule increased the relative yield of clinically actionable variants in sarcomere genes by $\geq 20\%$ in HCM ²⁰³. Thus, I applied this quantitative approach to evaluate whether it would decrease the yield of VUSs in AhcHCM patients. As mentioned in the introduction of this chapter (section 3.2.2), the authors developed a bespoke clustering algorithm to identify sarcomeric gene regions with significant clustering of rare non-truncating variants in a published large-scale LMM/OMGL cohort ($n > 6,000$) compared to ExAC controls. I estimated the pathogenicity of variants clustering in these regions through odds ratio (OR) and etiological fraction (EFs) metrics and variants with $EF \geq 0.95$ were upgraded to LP by activating the new EF-based PM1_strong rule (discussed in detail in sections 3.2.23.3.13) ²⁰³.

Given the observed differences in the HCM genetic architecture between Egyptian and predominantly Caucasian cohorts (e.g. higher frequency of rare *MYL2* and *MYL3* variants in AhcHCM), it would be inaccurate to apply the same proposed EF-based ACMG/AMP rules to all sarcomere genes in the AhcHCM cohort. Also, larger Egyptian case-control cohorts are required to accurately define patient-enriched variant clusters in sarcomere genes. Thus, I applied the new EF-based ACMG/AMP rules on rare missense variants in *MYH7* only, as they showed a similar variant distribution (i.e. enrichment in the *MYH7* HCM cluster 167-931 pre-defined by Walsh *et al.*, 2019) and comparable frequency between the AhcHCM (15.6%) and LMM/OMGL (13.6%) cohorts (Figure 3-13) ²⁰³.

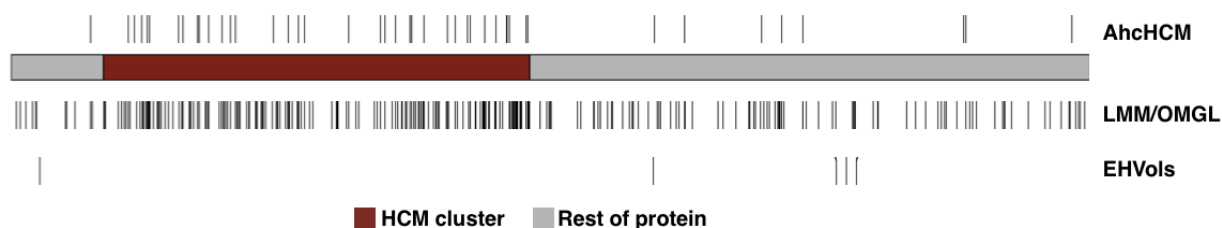


Figure 3-13: Enrichment of rare, missense *MYH7* variants in the previously reported HCM cluster (red) (residues 167-931) in the AhcHCM cohort ²⁰³. The distribution of *MYH7* rare missense variants in LMM/OMGL patients and EHVol controls are also shown.

The analysis showed that the proportion of AhcHCM cases with rare, missense *MYH7* variants in the pre-defined HCM cluster was significantly higher compared to EHVol controls (10.6% vs. 0.15%, $p\text{-value} < 0.0001$) (Table 3-11). The OR and EF for variants within this HCM cluster were calculated. As depicted in Table 3-11, the EF was > 0.95 , thus, *MYH7* missense, VUSs identified within this HCM cluster could be upgraded to LP as they activate the PM1_strong rule proposed by Walsh *et al.*, 2019 ²⁰³.

Of the 80 rare missense *MYH7* variants identified in the AhcHCM cohort (15.6%), 31 variants (6.0%) were clinically actionable (P/LP). Applying the PM1_strong rule increased this proportion to 13.2%.

Table 3-11: OR and EF for missense *MYH7* variants within the pre-defined HCM cluster (167-931) in the AhcHCM cohort

Gene	AhcHCM Freq (variants/total)	EHVol Freq (variants/total)	p-value	OR (95%CI)	EF (95%CI)
<i>MYH7</i>	13.04% (67/514)	0.25% (1/400)	< 0.0001	59.68 (10.25- 2374.28)	0.98 (0.90-1.0)

Significantly higher burden of AhcHCM cases with rare missense *MYH7* variants in the predefined HCM cluster compared to EHVol controls. Rare variants in this cluster are likely to be pathogenic with an $EF \geq 0.95$.

Taken together, applying both Egypt-specific PS4 and new EF-based PM1_strong rule increased the overall yield of clinically actionable variants identified in the AhcHCM cohort increased from 18.7% to 32.3% (Figure 3-14). These findings highlight the importance of using Egyptian-matched controls to accurately interpret variants identified in HCM patients.

Of the remaining patients, 22.8% (n=117) carried VUSs and 44.9% (n=231) did not carry rare variants in HCM-associated genes. This suggests that other undiscovered genetic (e.g. CNVs or non-coding variants), epigenetic (microRNAs, DNA methylation or histone modifications) or environmental factors may have contributed to HCM.

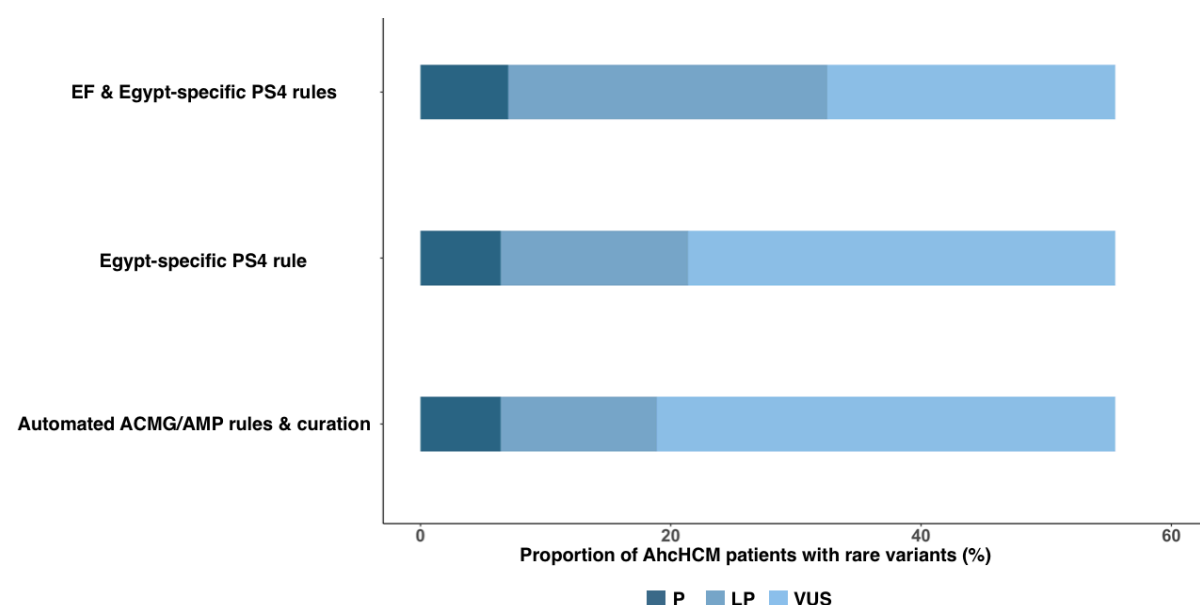


Figure 3-14: Application of the new EF-based ACMG/AMP PM1_strong- and Egypt-specific PS4 rules increased the yield of clinically actionable variants in the AhcHCM cohort. Proportion of AhcHCM patients with actionable variants (LP and P) based on the automated ACMG/AMP guidelines (implemented by CardioClassifier) and manual curation only as well as in addition to the Egypt-specific PS4 rule and finally addition of both EF-based ACMG/AMP PM1_strong rule (applied to *MYH7* missense variants only) and the Egypt-specific PS4 rules.

3.4.8 Higher proportion of AhcHCM patients with unique sarcomeric variants compared to RbhHCM patients

Given the observed underrepresentation of Egyptian patients in genomic research studies (as discussed in section 3.2.3), I attempted to investigate whether the frequency of unique variants was higher in the AhcHCM cohort compared to RbhHCM. Given the limited number of rare variants identified in the HCM minor- (i.e. *ACTN2*, *CSRP3*, *JPH2*, *PLN* and *TNNC1*) and syndromic genes, I restricted this analysis to the core sarcomere HCM (n=8) genes. AhcHCM-unique variants were defined as being absent in controls (EHVol and gnomAD) as well as in predominantly Caucasian HCM patients (RbhHCM and LMM/OMGL).

RbhHCM-unique variants were defined as being absent in Caucasian controls (CHVols) and gnomAD as well as in the AhcHCM cohort. As expected, a significant burden of unique variants in the AhcHCM cohort compared to the RbhHCM cohort was observed (22.2% vs. 3.9%, p-value<0.0001) (Table 3-12). In total, 54 and 27 unique variants in sarcomere genes were identified in the AhcHCM and RbhHCM cohorts, respectively. Of the 54 Egyptian-specific variants, 2 were identified in the GME control cohort (*MYH7* c.5769delG and *TPM1* p.Asp41Glu).

Table 3-12: Significant burden of unique variants in AhcHCM compared to RbhHCM patients

Freq. in AhcHCM (variants/total)	Freq. in RbhHCM (variants/total)	Freq diff. (AhcHCM - RbhHCM)	p-value
22.2% (114/514)	3.9% (27/684)	18.3%	<0.0001

The frequency AhcHCM-unique variants (i.e. absent in gnomAD, EHVols, RbhHCM and LMM/OMGL) compared to RbhHCM-unique variants (i.e. absent in gnomAD, EHVols, AhcHCM and LMM/OMGL) are shown. The difference in the burden of unique variants between AhcHCM- and RbhHCM patients (i.e. Freq. diff.) was evaluated for significance using Fisher's exact test.

Of the AhcHCM patients with unique variants, 65.8% carried VUSs (75 patients/114 variants). The yield of these inconclusive variants could be reduced if the statistical quantitative EF-based ACMG/AMP rules are applied, however, this will require a well-powered Egyptian case-control cohort.

3.4.9 Higher burden of homozygous variants in AhcHCM patients compared to RbhHCM

The reported high rate (~35%) of consanguinity, especially among first-cousins, in the Egyptian population prompted me to investigate whether the prevalence of homozygous variants is increased in the AhcHCM cohort compared to RbhHCM ¹⁰³. Given the limited number of rare variants identified in the HCM minor- (i.e. *ACTN2*, *CSRP3*, *JPH2*, *PLN* and *TNNC1*) and syndromic genes, I restricted this analysis to

the core sarcomere HCM (n=8) genes. Table 3-13 summarises the rare homozygous variants identified in the AhcHCM cohort.

I compared the proportion of AhcHCM patients with rare homozygous variants in all sarcomere genes combined with RbhHCM patients. The analysis showed that there was a significantly higher burden ($p\text{-value} < 0.0001$) of homozygous variants in the AhcHCM cohort (3.11% (16/514)) compared to RbhHCM (0.15% (1/684)).

Table 3-13: Frequency of rare sarcomeric homozygous variants identified in the AhcHCM cohort

Gene	Variant type	cDNA variant	Protein variant	gnomAD FAF popmax	EHVol freq	AhcHCM hom freq (variants/total)	AhcHCM het freq (variants/total)
<i>MYH7</i>	Missense	c.5122G>A	p.Glu1708Lys	0	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Missense	c.4258C>T	p.Arg1420Trp	3.0e-06	0.00% (0/400)	0.19% (1/514)	0.39% (2/514)
	Missense	c.3622G>C	p.Asp1208His	0	0.00% (0/400)	0.19% (1/514)	0.58% (3/514)
	Missense	c.1064C>T	p.Ala355Val	0	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Missense	c.925G>A	p.Asp309Asn	1.1e-05	0.00% (0/400)	0.19% (1/514)	0.78% (4/514)
	Frameshift	c.5769delG	p.Ser1924AlafsTer9	0	0.00% (0/400)	0.19% (1/514)	3.11% (16/514)
<i>MYBPC3</i>	Missense	c.2908C>T	p.Arg970Trp	0	0.25% (1/400)	0.19% (1/514)	0.39% (2/514)
	Missense	c.2458C>G	p.Arg820Gly	0	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Inframe Deletion	c.2441_2443delAGA	p.Lys814del	0	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Nonsense	c.1558G>T	p.Glu520Ter	0	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Splice acceptor variant	c.1227-2A>G	-	0	0.00% (0/400)	0.19% (1/514)	0.19% (1/514)
<i>MYL2</i>	Missense	c.278C>T	p.Ala93Val	0	0.00% (0/400)	0.19% (1/514)	0.78% (4/514)
<i>MYL3</i>	Missense	c.508G>C	p.Glu170Gln	0	0.00% (0/400)	0.58% (3/514)	0.58% (3/514)
<i>TNNI3</i>	Missense	c.485G>T	p.Arg162Leu	0	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)

The relatively higher rate of homozygosity observed in AhcHCM patients compared to RbhHCM patients prompted me to firstly investigate whether rare (i.e. gnomAD $FAF_{popmax} \leq 4 \times 10^{-5}$) heterozygous sarcomeric variants become more pathogenic when expressed in homozygosity. The second aim was to assess whether applying the cut-off allele frequency for recessive inheritance (gnomAD $FAF_{popmax} \leq 0.00126$), which was calculated by the online tool <https://www.cardiodb.org/allelefrequencyapp/>, could identify AhcHCM patients with variants that would be more common than what is expected for autosomal dominant inheritance.

To address the first aim, I sought to compare the clinical characteristics of patients carrying the same variant in the heterozygous and homozygous states. Complete clinical data was only available for the heterozygous and homozygous patients (n=1 each) carrying the *MYBPC3* splice acceptor variant c.1227-2A>G. As expected for truncating *MYBPC3* variants, c.1227-2A>G was classified as likely pathogenic according to the ACMG/AMP guidelines. Evaluation of the clinical data showed that the homozygous c.1227-2A>G-carrier was diagnosed with HCM at a relatively younger age compared to the heterozygous patient (25 years and 40 years, respectively). In addition, the homozygous patient presented with a relatively higher degree of hypertrophy (i.e. LVMWT: 27mm and 20mm, respectively) and had borderline systolic function with EF of 54%. This suggests that the likely pathogenic heterozygous variant c.1227-2A>G could result in a more severe phenotype when expressed in the homozygous state. Missing clinical data for the remaining patients will be collected in the future to assess whether heterozygous patients with putatively disease-causing sarcomeric variants tend to show a less severe HCM phenotype compared to homozygous patients.

In addition, it is worth stating that the *MYL3* missense variant p.Glu170Gln was found to account for the highest proportion of AhcHCM patients with rare homozygous variants (0.58%, 3/514) (Table 3-13). This variant was also observed in the heterozygous state in 3 patients and was absent from EHVol controls as well as from gnomAD and the large-scale LMM/OMGL HCM cohort. Due to the unavailability of evidence supporting the pathogenicity of this novel variant, it was classified as a VUS.

However, its significant enrichment in AhcHCM patients over EHVol controls (1.17% vs. 0%, p-value=0.04) suggests that it could be likely pathogenic (i.e. variant activates PM2 and PS4 ACMG/AMP rules (refer to Figure 3-2)). Since the clinical data for the p.Glu170Gln-carriers was incomplete, I was unable to compare the phenotypes between heterozygous- and homozygous variant-carriers. However, it would be interesting to investigate whether this putatively disease-causing variant could cause a more severe phenotype when expressed in homozygosity.

I then sought to perform a global analysis to investigate whether homozygous variants are generally more pathogenic compared to those expressed in heterozygosity. Thus, I compared the clinical characteristics of homozygous patients with rare sarcomeric variants with all heterozygous variant-carriers identified in the AhcHCM cohort (Table 3-14). As expected, the proportion of homozygous patients with consanguineous parents was higher compared to heterozygous patients (77.8% vs. 28.4%, p-value<0.05). However, there were no differences in the age at disease onset (26.8 ± 17.9 yrs (hom) vs. 35.1 ± 17.3 (het), p-value=0.14) and degree of LVH between both homozygous and heterozygous patients (LVMWT: 26.5 ± 7.5 (hom) vs. 23.5 ± 7.1 (het), p-value=0.50).

It is important to note that comparing the clinical characteristics between homozygous variant-carriers and all heterozygous variant-carriers has several limitations. First of all, a subset of the heterozygous variants may require a “second hit” to express the HCM phenotype, which could explain why these variants did not result in a more severe phenotype when expressed in homozygosity. Secondly, it would be more accurate to compare the clinical characteristics of heterozygous and homozygous patients carrying the same variant or the variant class (e.g. truncating variants in *MYBPC3*) in order to truly assess the effect of the second hit. This analysis, however, will require a larger, sufficiently powered cohort size to accurately assess the effect of homozygosity on disease severity.

Table 3-14: Summary of demographic and clinical characteristics for AhcHCM patients with homozygous and heterozygous variants in sarcomeric HCM genes

Variable*	Total	Homozygous	Heterozygous	p-value
Gender (% Male)	67.1% (314/468)	50.0% (7/14)	67.6% (307/454)	0.25
Age at diagnosis (years) (mean (sd))	34.9±17.4	26.8±17.9	35.1±17.3	0.14
Offspring of consanguineous marriage	28.4% (98/345)	77.8% (7/9)	27.1% (91/336)	<0.05
FH of HCM	25.1% (110/438)	8.3% (1/12)	25.6% (109/426)	0.31
FH of SCD	17.8% (78/438)	16.7% (2/12)	17.8% (76/426)	1.00
LVMWT (mm)	23.6±7.1	26.5±7.5	23.5±7.1	0.23
Extreme LVMWT (>30mm)	17.7% (54/305)	36.4% (4/11)	17.0% (50/294)	0.11
LVOTO	73.3% (217/296)	81.8% (9/11)	73.0% (208/285)	0.73

Proportion of homozygous patients with consanguineous parents was significantly higher (p-value<0.05) compared to heterozygous patients (highlighted in bold). FH: family history; SCD: sudden cardiac death; LVMWT: left ventricular maximal wall thickness; LVOTO: left ventricular outflow tract obstruction. *Cases for which information across the different variables was missing were excluded from the analysis.

To address the second aim, I applied the gnomAD $FAF_{popmax} \leq 0.00126$ for recessive inheritance. This analysis identified 5 homozygous variants, which are depicted in Table 3-15. These variants are not expected to be disease-causing in the heterozygous state as they would be too common to be highly penetrant under an autosomal dominant disease model. However, they could potentially become penetrant when expressed in homozygosity. Clinical data for these patients will be collected in order to be able to assess the effect of these homozygous variants on disease severity. Currently, clinical data was only available for the patients carrying the missense *MYL3* variant p.Ala57Asp. Homozygous p.Ala57Asp-carriers presented with HCM at a relatively younger age (21.5±14.8 vs. 34) and had a greater LVWMT (27±2.8 vs. 16) compared to the heterozygous carrier. This finding suggests that harbouring the p.Ala57Asp in homozygosity may affect the disease severity of HCM, however, segregation analysis of this variant in consanguineous families will provide more definitive evidence on its role in HCM.

Intriguingly, the missense *MYBPC3* variant (p.Glu441Lys), which was identified in 2 homozygous AhcHCM patients, was found to be very commonly observed in the heterozygous state in patients. Collectively, this variant was significantly ($p=0.03$) enriched in AhcHCM patients (het+hom freq= 4.67%) compared to EHVol controls (2.0%). In contrast, the variant was only observed in 1 RbhHCM patient (0.15%, $n=684$) and 6/6,179 patients (0.05%) from the LMM/OMGL HCM cohort. The variant was absent from Caucasian healthy volunteers ($n=1,054$) recruited to the Royal Brompton Hospital (more details in section 2.3.1.2). This suggests that p.Glu441Lys could be more prevalent among Egyptian/Middle Eastern populations. In fact, it was also identified in 1 individual from NEA (0.27%, $n=368$) and 2 Qatari individuals (0.16%, $n=1,213$) included in the al mena database ²²⁸.

I then interrogated PubMed and ClinVar to evaluate whether this variant had previously been reported to be associated with HCM. The literature review showed that the p.Glu441Lys variant had been previously reported in several studies in association with HCM ^{218,229–233}. Olivotto *et al.*, 2008 were the first to observe the variant in 1 HCM patient who carried an additional pathogenic variant (p.Glu258Lys) in *MYBPC3* ²²⁹. Another study by Millat *et al.*, 2010 also reported the expression of p.Glu441Lys in conjunction with another pathogenic sarcomeric variant in HCM patients ²³¹. Similar findings were submitted to ClinVar too (Accessions: SCV000280210.1, SCV000059030.5). In contrast, a study by Marsiglia *et al.*, 2010 reported the expression of p.Glu441Lys in isolation in 1 HCM patient who presented with a milder disease phenotype ²³⁰. Similarly, the LMM genetic testing laboratory reported that the 2 identified patients harbouring the p.Glu441Lys variant only appeared to have a relatively later HCM disease onset and a less severe phenotype compared to another HCM patient who carried an additional pathogenic HCM variant (ClinVar accession: SCV000059030.5). Taken together, these results suggest that p.Glu441Lys may have an intermediate/non-severe effect on the disease phenotype in isolation. However, segregation and/or functional analyses are required to confirm this. It is noteworthy that a subset (31.82%, 7/22) of AhcHCM patients carrying the variant in heterozygosity harboured an additional sarcomeric variant. Once the clinical

data for the AhcHCM cohort is complete, a comparison of the clinical phenotypes between single p.Glu441Lys-variant carriers and compound heterozygotes will be performed to gain deeper insight into the role of the variant in HCM. In addition, the 2 homozygous p.Glu441Lys-variant carriers should be followed-up clinically to assess whether carrying the variant in homozygosity could lead to a more severe disease phenotype.

Table 3-15: Frequency of homozygous sarcomere HCM variants identified in the AhcHCM cohort under a recessive disease model (i.e. gnomAD FAF_{popmax} ≤ 0.00126)

Gene	Variant type	cDNA variant	Protein variant	gnomAD FAF popmax	EHVol freq	AhcHCM hom freq (variants/total)	AhcHCM het freq (variants/total)
MYBPC3	Missense	c.3676C>T	p.Arg1226Cys	0.000064	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Missense	c.2618C>A	p.Pro873His	0.000093	0.00% (0/400)	0.19% (1/514)	0.00% (0/514)
	Missense	c.1321G>A	p.Glu441Lys	0.000216	2.0% (8/400)	0.39% (2/514)	4.28% (22/514)
MYL3	Missense	c.530A>G	p.Glu177Gly	0.000044	0.75% (3/400)	0.78% (4/514)	0.19% (1/514)
	Missense	c.170C>A	p.Ala57Asp	0.000353	0.00% (0/400)	0.39% (2/514)	0.19% (1/514)

This cut-off gnomAD FAF was calculated for recessive inheritance using the online tool: <https://www.cardiodb.org/allelefrequencyapp/>.

3.4.10 A subset of AhcHCM patients presented with double rare variants in sarcomere genes

A study by Burns *et al.*, 2017, reported that patients with multiple rare variants in the 8 sarcomere genes, regardless of variant classification, had an earlier HCM disease onset as well as worse clinical outcomes compared to single variant-carriers¹⁸¹. These findings prompted me to assess the prevalence of multiple rare sarcomere variant-carriers in the AhcHCM cohort and to investigate their effect on the disease phenotype. A total of 15 patients (2.9%) with double-hits (i.e. 2 different rare variants in sarcomere genes) were identified in the AhcHCM cohort (Table 3-16). No patients with three or more variants were identified. A comparison of the clinical characteristics between single-hit (i.e. rare variant in a sarcomere gene) and double-hit variant-carriers showed that patients with an additional variant burden presented with HCM at a significantly younger age (23.8+ 14.9 vs. 33.7+15.0, $p=0.04$) (Table 3-17). This finding further supports the association of carrying multiple rare variants with early HCM disease onset reported by Burns *et al.*, 2017. Analysis of an expanded AhcHCM cohort size could reveal deeper insights into the effect of multiple rare variants on disease presentation and outcome.

Table 3-16: Summary of rare double-hits in sarcomere genes identified in the AhcHCM cohort (n=15/514)

#	Double hit	Variant type	cDNA variant	Protein variant	gnomAD FAF popmax	ACMG classification
1	<i>MYBPC3</i>	Missense	c.2908C>T	p.Arg970Trp	0	VUS
	<i>MYBPC3</i>	Frameshift	c.1444delG	p.Ala482ArgfsTer6	0	LP
2	<i>MYBPC3</i>	Missense	c.2171G>A	p.Arg724Gln	7.00E-06	VUS
	<i>MYBPC3</i>	Nonsense	c.999C>A	p.Tyr333Ter	0	LP
3	<i>MYBPC3</i>	Nonsense	c.1308C>A	p.Cys436Ter	0	LP
	<i>MYBPC3</i>	Missense	c.836G>C	p.Gly279Ala	4.00E-06	VUS
4	<i>MYBPC3</i>	Missense	c.1773G>T	p.Lys591Asn	0	VUS
	<i>MYBPC3</i>	Missense	c.305C>T	p.Pro102Leu	3.70E-05	VUS
5	<i>MYBPC3</i>	Splice donor variant	c.1223+1G>A	-	2.40E-05	LP
	<i>MYH7</i>	Frameshift	c.5769delG	p.Ser1924AlafsTer9	0	VUS
6	<i>MYBPC3</i>	Missense	c.2470G>A	p.Asp824Asn	0	VUS
	<i>MYH7</i>	Frameshift	c.5769delG	p.Ser1924AlafsTer9	0	VUS
7	<i>MYBPC3</i>	Splice donor variant	c.3627+2delT	-	0	LP
	<i>MYH7</i>	Frameshift	c.5769delG	p.Ser1924AlafsTer9	0	VUS
8	<i>MYBPC3</i>	Missense	c.2738G>C	p.Cys913Ser	0	VUS
	<i>MYH7</i>	Missense	c.2770G>A	p.Glu924Lys	0	P
9	<i>MYBPC3</i>	Missense	c.1003C>A	p.Arg335Ser	0	VUS
	<i>MYH7</i>	Missense	c.2348G>A	p.Arg783His	7.00E-06	LP
10	<i>MYBPC3</i>	Missense	c.2534G>A	p.Arg845His	1.90E-05	VUS
	<i>MYH7</i>	Missense	c.2147G>C	p.Gly716Ala	0	LP
11	<i>MYBPC3</i>	Missense	c.130C>T	p.Arg44Cys	0	VUS
	<i>MYH7</i>	Missense	c.2011C>T	p.Arg671Cys	0	VUS
12	<i>MYH7</i>	Frameshift	c.5769delG	p.Ser1924AlafsTer9	0	VUS
	<i>MYH7</i>	Missense	c.925G>A	p.Asp309Asn	1.10E-05	VUS
13	<i>MYH7</i>	Missense	c.3460G>A	p.Gly1154Ser	0	VUS
	<i>TNNI3</i>	Missense	c.485G>T	p.Arg162Leu	0	VUS
14	<i>MYBPC3</i>	Missense	c.2977C>T	p.Leu993Phe	0	VUS
	<i>TNNI3</i>	Missense	c.188G>A	p.Arg63Gln	0	VUS
15	<i>MYH7</i>	Missense	c.3622G>C	p.Asp1208His	0	VUS
	<i>TNNT2</i>	Missense	c.274C>T	p.Arg92Trp	0	P

Table 3-17: Summary of demographic and clinical characteristics for AhcHCM patients with single and double rare variants in sarcomere genes

Variable*	Total	Single variant-carriers (n=199)	Double variant-carriers (n=15)	p-value
Age at diagnosis (years) (mean (sd))	33.1±15.2	33.7±15.0	23.8± 14.9	0.04
Gender (% Male)	65.7% (130/198)	66.3% (122/184)	57.1% (8/14)	0.57
FH of HCM	31.4% (60/191)	32.2% (57/177)	21.4% (3/14)	0.55
FH of SCD	15.7% (30/191)	15.3% (27/177)	21.4% (3/14)	0.46
LVMWT (mm) (mean (sd))	24.5±6.8	24.7±6.9	21.3±4.6	0.07
LVOTO	75.4% (101/134)	76.2% (96/126)	62.5% (5/8)	0.41
Myectomy	60.0% (81/135)	61.1% (77/126)	44.4% (4/9)	0.48
ICD implantation	1.8% (2/112)	0.9% (1/106)	16.7% (1/6)	0.10

Double sarcomere variant-carriers presented with HCM at a significantly (p-value<0.05) younger age compared to single sarcomere variant-carriers (highlighted in bold). *Cases for which information across the different variables was missing were excluded from the analysis.

3.4.11 Identification of 4 HCM patients with novel homozygous variants in Cysteine and Glycine-Rich Protein 3 (*CSRP3*) increases the worldwide number of known *CSRP3* knockout HCM patients to 7

I evaluated whether rare homozygous variants in other likely HCM genes (*ACTN2*, *CSRP3*, *JPH2* and *TNNC1*) were observed in the AhcHCM cohort. A total of 3 variants were identified in *CSRP3* in the homozygous state. The missense variant p.Arg122Gln was observed in 1 (0.19%) and 3 (0.58%) patients in homozygosity and heterozygosity, respectively. In addition, 4 patients carried biallelic truncating variants in *CSRP3*. The splice acceptor variant c.415-1G>C was observed in 1 homozygous AhcHCM patient (0.19%) and the splice donor variant c.414+1G>T was observed in 3 homozygous patients (0.58%) and in 1 EHVol control but in the heterozygous state. Both truncating variants were absent in gnomAD controls as well as in the analysed predominantly Caucasian HCM cohorts. These variants are expected to elicit NMD resulting in the absence or a severe reduction in the amount of generated proteins. To date, only 3 HCM *CSRP3* knockout patients have been reported in the literature (Table 3-18)^{234,235}. With these findings, the total number of *CSRP3* HCM knockout patients has reached 7.

Table 3-18: Summary of homozygous *CSRP3* variants reported in the literature to date

Gene cDNA variant	Protein variant	Summary of findings
<i>CSRP3</i> c.364C>T	p.Arg122X	Female patient was diagnosed with HCM at the age of 54. Echocardiography showed asymmetric IVS hypertrophy of 18mm and maximal LVOTO of 18mmHg under provocation. Severe diastolic dysfunction was observed. Consanguinity of parents was not confirmed but the family history was negative for HCM ²³⁴ .
<i>CSRP3</i> c.369T>A	p.Cys123X	Female patient born to consanguineous parents was diagnosed with HCM (IVS was 13 mm and antero basal septum 15 mm) at the age of 18 years. Echocardiography showed extreme LVOTO (45mmHg at rest). Due to the occurrence of multiple episodes of syncope an ICD was implanted at the age of 31. She suffered from atrial tachycardia at the age of 33 years. Family history of HCM was negative ²³⁵ .
<i>CSRP3</i> c.483dup	p.Lys162GlnfsX52	Female patient born to consanguineous parents was diagnosed with HCM (LVMWT:19mm) at the age of 27 years. Echocardiography showed extreme LVOTO (>50mmHg at rest). Family history of HCM was negative ²³⁵ .

IVS: interventricular septum; ICD: implantable cardiac defibrillator

3.4.12 Higher degree of LVH in AhcHCM patients with clinically actionable variants in *MYBPC3* compared to *MYH7*-variant carriers

Previous studies have reported contradictory findings concerning the clinical outcomes of patients with clinically actionable (P/LP) variants in *MYBPC3* compared to *MYH7* ^{61,236–239}. This prompted me to compare the clinical phenotypes between both patient groups (Table 3-19). The analysis showed that the *MYBPC3*-variant carriers had a significantly higher LVMWT compared to *MYH7*-variant carriers (26.4±5.5 vs. 23.1±6.4, p-value<0.05). *MYBPC3* patients were also more likely to require surgical intervention (myectomy) compared to *MYH7* patients (68.8% vs 39.4%, p-value=0.35). These findings provide preliminary insights into the effect of clinically actionable variants in *MYBPC3* and *MYH7* on the disease phenotype. Expanding this analysis to a larger cohort size will provide supporting evidence of these genetic-phenotypic associations.

Table 3-19: Summary clinical characteristics for AhcHCM patients with P/LP variants in *MYH7* compared to *MYBPC3*-variant carriers

Variable	Total	<i>MYH7</i> -P/LP carriers	<i>MYBPC3</i> -P/LP carriers	p-value
Age at diagnosis (years) (mean (sd))	34.9±17.3	35±16.8	32.2±13.1	0.39
LVMWT (mm)	24.6±6.2	23.1±6.4	26.4±5.5	<0.05
Extreme LVMWT (>30mm)	17.9% (10/56)	12.9% (4/31)	24% (6/25)	0.32
LVOTO	75.9% (44/58)	71.9% (23/32)	80.8% (21/26)	0.35
Surgical myectomy	53.8% (35/65)	39.4% (13/33)	68.8% (22/32)	0.35

Only LVMWT was significantly (i.e. p-value<0.05) higher in *MYBPC3*-variant carriers compared to *MYH7*-variant carriers (highlighted in bold)

3.5 Discussion

The unprecedented advances in NGS technology have fuelled a surge in HCM research studies, which resulted in the rapid increase of reported genetic associations with HCM. These studies, however, did not adequately account for background genetic variation but simply relied on the absence of identified putative pathogenic variants in small control cohorts. The release of the large-scale ExAC reference genetic database in 2016 significantly improved our understanding of HCM variant interpretation as it demonstrated that individually “rare” variants were collectively “common” in the general population. More recent large-scale HCM case-control studies leveraged the ExAC resource to better interpret rare variation in HCM genes, however, the vast majority of these have focused on cases and controls of Caucasian descent. The discordance of non-Caucasian populations, including individuals from the MENA region, in genetic sequencing efforts hinders our ability to interpret variants identified in these populations.

This research study provides an initial milestone towards defining the genetic architecture of HCM in the currently underrepresented Egyptian population, which is characterised by ethnic diversity, elevated levels of consanguinity and high burden of CVD.

In this chapter, I present findings based on burden testing of rare variation in validated HCM-associated genes in Egyptian (AhcHCM) patients compared to ethnically-matched EHVol controls. I also demonstrate differences in the genetic architecture of HCM between Egyptians and predominantly Caucasian populations.

3.5.1 Insights into the genetic landscape of HCM in Egypt

To analyse the genetic landscape of HCM in Egypt, I compared the frequency of rare variation between Egyptian HCM patients and controls. I selected 22 genes that were validated by Ingles *et al.*, 2019 based on a thorough examination of the evidence supporting their association with HCM using the ClinGen framework^{148,176}. Both AhcHCM- and EHVol cohorts were sequenced using the targeted TruSight Cardio Sequencing panel (which includes 174 genes associated with ICCs) with high coverage

⁹¹. Raw data from both cohorts were processed through the same in-house bioinformatics pipeline. Thus, the analysis was not confounded by technical artefacts or ethnic heterogeneity (as confirmed by PCA).

In line with previous HCM genetic studies, the analysis showed that both *MYH7* and *MYBPC3* accounted for the majority of rare variation in the AhcHCM cohort ^{100,238–241}. Non-truncating variants in *MYH7* were more prevalent among patients compared to truncating variants. This finding is in line with the widely accepted belief that missense variants in *MYH7* cause HCM through gain-of-function actions, whereby mutant proteins incorporate into the sarcomere as “poison peptides” ²⁴². As expected, the proportion of patients with truncating variants in *MYBPC3* was higher compared to non-truncating. Previous studies have demonstrated that truncating variants in *MYBPC3* account for the majority of HCM-causing variants in this gene, whereby they are believed to cause disease through haploinsufficiency (i.e. insufficient amount of functional protein for normal sarcomere function) ^{54,243–245}. Intriguingly, missense variants in *MYBPC3* have also been reported to cause haploinsufficiency possibly by affecting sequences involved in splicing, which ultimately result in the generation of premature termination codons (PTC) ²⁴⁶.

Several studies have reported genotype-phenotype correlations in HCM patients with variants in these 2 genes. For example, a study by Coppini *et al.*, 2014 showed that HCM patients with thick filament variants (*MYBPC3* and *MYH7*) exhibited greater maximal LV wall thickness and LVOTO compared to thin-filament (*ACTC1*, *TNNI3*, *TNNT2* and *TPM1*) patients ²⁴⁷. However, there were no significant differences in cardiac- and SCD death rates between both patient groups ²⁴⁷. Studies comparing the prognosis between patients with *MYBPC3* and *MYH7* disease-causing variants have reported inconsistent findings. For example, pathogenic variants in *MYBPC3* were originally believed to be associated with delayed disease onset and improved prognosis compared to *MYH7* variants ^{61,236,237}. However, these findings were based on small study cohorts (13-25 families) casting doubt on their validity. Subsequent studies on larger HCM cohort sizes (n>350 patients each) showed that there was no significant difference in the clinical outcomes of patients with

MYBPC3 and *MYH7* variants ^{238,239}. A recent study by Ho *et al.*, 2018, which leveraged the SHaRe registry, an international database of patients with cardiomyopathy, demonstrated that in an HCM cohort of 4,591 patients, *MYH7*-variant carriers had an increased risk of atrial fibrillation, heart failure and overall composite outcome compared to *MYBPC3*-variant carriers ⁸⁰. In the AhcHCM cohort presented here, the degree of LVH was greater in patients with P/LP variants in *MYBPC3* compared to *MYH7* variant carriers (p-value<0.05) and *MYBPC3* patients appeared more likely to require surgical intervention (myectomy) compared to *MYH7* patients (68.8% vs 39.4%), although this failed to achieve significance (p-value=0.35). A larger cohort size is required to validate these findings.

Consistent with previous studies, rare variants in other sarcomere genes were less frequent, accounting for less than 3% of AhcHCM patients each ^{28,83}. Of these genes, only *MYL2* and *MYL3* genes showed higher frequencies of rare variation in the AhcHCM cohort compared to EHVol controls, that were statistically significant (p-values=0.0009 and 0.0002, respectively, after Bonferroni correction). As discussed below, the burden of rare variation in *MYL2* and *MYL3* genes was higher in AhcHCM patients compared to LMM/OMGL and RbhHCM, hence, these genes were studied in more detail.

3.5.2 Enrichment of rare variation in *MYL2* and *MYL3* genes and *MYH7* truncating variants in Egyptian patients compared to predominantly Caucasian patients

To identify differences in the genetic architecture of HCM between Egyptian and predominantly Caucasian populations, I analysed data from both LMM/OMGL and RbhHCM cohorts. The LMM/OMGL cohort represents the largest published HCM cohort to date with n>6,000 sequenced patients ^{28,100}. However, there are differences between the LMM/OMGL and AhcHCM cohorts in terms of patient recruitment, cohort composition and sequencing gene panels used. To illustrate, LMM/OMGL patients were initially diagnosed with HCM by a consultant cardiologist then referred to clinical genetic testing whereas the AhcHCM cohort comprised prospective patients from outpatient clinics at

AHC. In addition, patients from the LMM cohort, with unclear diagnosis of HCM or with metabolic cardiomyopathies were excluded. Finally, not all LMM/OMGL patients were sequenced using the same gene panels as they expanded over time. Therefore, I also integrated data from the smaller RbhHCM cohort (n=684) into the analysis, as this cohort was comparable to AhcHCM in terms of patient recruitment (i.e. prospective patients from outpatient clinics), cohort composition (e.g. inclusion of patients with metabolic cardiomyopathies) and sequencing approach (i.e. targeted ICC gene panel included in TruSight Cardio Sequencing Kit). Thus, the comparison of rare variation between AhcHCM- and RbhHCM cohorts was not confounded by technical differences.

A main finding of this analysis is that, overall, a higher frequency of rare variation was observed in the Egyptian HCM cohort compared to predominantly Caucasian cohorts. However, it is important to note that rare variants in this analysis were defined using a specific FAF derived from gnomAD, which is not representative of the MENA population. Thus, I sought to assess whether this difference truly reflects a distinct HCM genetic architecture between both cohorts or is more likely to be confounded by a biased definition of variant rarity introduced by gnomAD. Given that rare synonymous variants are not expected to be disease-causing, I compared the frequency of these variants between the Egyptian cohort (including EHVolS) and predominantly Caucasian patients (RbhHCM and CHVolS). The corresponding GIF for the distribution of 174 single-ICC gene p-values was 2.770, demonstrating a presence of bias in the analysis (GIF=1 indicates absence of bias). Given the technical comparability between the analysed cohorts, the observed high level of inflation could only be attributed to population stratification. This finding was not unexpected given that the variant frequency threshold used to define rare variants in the AhcHCM cohort was based on gnomAD, which lacks data from the MENA region. Given the high diversity of MENA populations, “rare” variants identified in gnomAD may be common among MENA individuals ^{134,248}. In fact, I demonstrated that 2 “rare” (i.e. $\text{gnomAD FAF}_{\text{popmax}} \leq 4 \times 10^{-5}$) variants identified in the AhcHCM cohort were commonly observed among EHVol controls casting doubt on their pathogenicity (discussed in detail in next section). This finding highlights the importance of adding MENA-specific

populations to existing reference genetic databases in order to improve variant interpretation and to account for the substantial genetic diversity in these populations. For example, I have shown that the inclusion of only 1,111 individuals from the GME population into the substantially larger gnomAD dataset (n=125,748), was sufficient to reduce the observed inflation to $GIF = 2.419$.

To alleviate this bias, I utilised the EHVol- and CHVol control cohorts to directly compare the excess burden of rare variation between AhcHCM_{EHVol} and RbhHCM_{CHVol} cohorts, respectively. By accounting for ethnic-specific variation, this approach enabled me to re-evaluate the genetic differences between Egyptian and predominantly Caucasian cohorts in a more accurate manner. This analysis confirmed the significant enrichment of truncating variants in *MYH7* as well as non-truncating variants in *MYL2*, *MYL3* and *JPH2* genes.

Given the fact that truncating variants in *MYH7* have generally been considered non-pathogenic in HCM¹⁰⁰, it was intriguing to observe a statistically significant (p-value=0.0004, after Bonferroni correction) burden of these variants in the AhcHCM cohort compared to EHVol controls. This burden was attributed to a single frameshift variant (c.5769delG) that was prevalent among 3.31% of AhcHCM patients and was absent in EHVol controls. Thus, the *MYH7* c.5769delG variant was analysed in detail in the subsequent chapter to investigate its role in HCM.

An enrichment of rare missense variants in *MYL2* and *MYL3* genes was also observed in AhcHCM patients compared to RbhHCM. These essential- (*MYL3*) and regulatory (*MYL2*) light chains are thought to stabilise the α -helical neck region of myosin and to modulate the interaction between actin and myosin during contraction^{249–251}. Few families with variants in these genes have been reported to date, which is why the disease mechanism of *MYL2* and *MYL3* variants is only partially understood^{170,252–257}. In the work presented here, 11 rare missense *MYL2* and *MYL3* variants were identified, including 5 variants that were unique to the Egyptian population. Thus, studying these variants in more detail,

through segregation and/or functional studies, could provide deeper insights into their mechanistic roles in HCM. The *MYL2* variant (p.Arg58Gln) was observed in 0.39% (2/514) and 0.17% (7/4185) of AhcHCM- and LMM/OMGL patients, respectively. This variant was reported to be associated with early onset of HCM, poor prognosis and a risk of SCD ^{253,255}. Clinical data was available for one of the 2 AhcHCM patients carrying the p.Arg58Gln variant. In agreement with literature findings, the AhcHCM patients with the p.Arg58Gln variant presented with HCM at the age of 8 years.

In addition, 4 VUS variants (*MYL2* p.Leu150Phe, p.Ala93Val and *MYL3* p.Met173Thr, p.Glu170Gln) were found to be enriched in AhcHCM patients compared to LMM/OMGL. These variants were also absent in EHVol controls, however, a statistically significant excess burden of these variants in AhcHCM patients over controls was not achieved, possibly due to the limited cohort sizes. Re-evaluating the burden of these variants in a larger-scale Egyptian case-control cohort could provide strong evidence of their pathogenicity (by activating the PS4 rule).

3.5.3 Egyptian HCM patients are more likely to obtain inconclusive genetic results compared to predominantly Caucasian patients

In the second part of the analysis I sought to evaluate whether Egyptian HCM patients were more likely to receive inconclusive genetic results (i.e. VUS) compared to predominantly Caucasian patients, who have been extensively studied. Thus, I compared the distribution of rare variants in HCM-associated genes (n=22) by variant class (P, LP and VUS) between the AhcHCM- and RbhHCM cohort, for which individual-level data was available.

As expected, the yield of VUSs was higher in AhcHCM patients compared to RbhHCM patients (66.2% vs. 45.5%, respectively, p-value<0.0001). This finding highlights the fact that patients from Caucasian-majority populations are more likely to obtain a clinically actionable (P or LP) genetic result compared to underrepresented individuals. That is because HCM genetic research studies and clinical testing have mainly been conducted

in Caucasian populations. Thus, Caucasian patients are more likely to have a disease-causing variant that has previously been well-characterised. Consistent with this notion, a recent study reported that the detection rate of clinically actionable variants for cardiomyopathy was higher in Caucasian patients compared to underrepresented ethnic groups, such as African-Americans and Hispanics ²⁰⁷. Similarly, the proportion of HCM patients with inconclusive genetic results was higher (p-value<0.0001) in non-Caucasian patients (20.6%) in the LMM cohort compared to Caucasians (13.9%) ^{28,203}. These ethnic disparities severely reduce the usefulness of clinical genetic testing for HCM as only patients with previously characterised variants will obtain a positive genetic result.

To address this problem, I utilised the EHVol cohort and the quantitative approach to variant classification developed by Walsh *et al.*, 2019 to re-evaluate the interpretation of VUSs identified in the AhcHCM cohort ^{28,203}. The EHVol genetic dataset enabled me to examine the frequency of AhcHCM VUSs in ethnically-matched controls. Only 2 VUSs, *MYBPC3* p.Pro102Leu and *PRKAG2* p.Ile492Asn, were found to be commonly observed (MAF $\geq$ 0.0025) in EHVol controls suggesting that they are likely benign variants. In addition, the *MYH7* frameshift variant c.5769delG was found to be enriched in AhcHCM patients over EHVol controls (3.31% vs. 0%, p-value<0.0001). Hence, this variant was re-classified to LP by activating the ACMG/AMP PS4 rule.

The new EF-based PM1_strong rule was also applied on AhcHCM VUSs as it provides a conservative (only variants with EF $\geq$ 0.95 activate PM1_strong rule) quantitative approach to variant classification. Walsh *et al.*, 2019 demonstrated that the yield of HCM patients (n=684) with actionable variants increased by 14% after applying the new PM1_strong rule on VUSs variants in 6 HCM genes (*MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *MYL3* and *TPM1*) ^{28,203}. The authors also showed that patients with variants upgraded from VUS to LP/P were more phenotypically similar to patients with well-characterised pathogenic variants, which provides supporting evidence for the validity of the adopted quantitative approach to variant interpretation. In the work presented here, however, the PM1_strong_rule was only applied to missense variants in *MYH7* as they showed a

comparable distribution (i.e. enriched in HCM cluster (167-931)) and frequency between AhcHCM- (13.58%) and LMM/OMGL patients (15.56%). Here, I have demonstrated that applying the Egypt-specific PS4 and EF-based PM1_strong rules increased the yield of clinically actionable variants from 18.7% to 32.3%. Therefore, application of the new EF-based ACMG/AMP rule could substantially improve the yield of clinically actionable variants in currently underrepresented Egyptian patients.

However, a limitation of this approach is that it was based on genetic data from the large-scale LMM/OMGL cohort compared to ExAC. Given the observed differences in HCM genetic architecture between AhcHCM- and predominantly Caucasian patients (e.g. higher frequency of rare *MYL2* and *MYL3* variants in AhcHCM), it would be inaccurate to apply this rule to other genes in the AhcHCM cohort. More genetic research in large-scale Egyptian case-control cohorts is required to be able to validate this quantitative approach.

3.5.4 A substantial proportion of Egyptian HCM patients did not present with a genetically identifiable aetiology

It is also worth stating that 44.9% of the AhcHCM cohort did not present with a genetically identifiable aetiology (i.e. no rare variation in HCM-associated genes). This finding suggests that environmental, epigenetic or other genetic factors (including CNVs and modifiers) may have contributed to HCM pathogenesis. For example, a study by Ceyhan-Birso *et al.*, 2016 reported the prevalence of CNVs, including an intragenic deletion in *MYBPC3*, in only 0.6% of HCM patients ²⁵⁸. In another study by Mademont-Soler *et al.*, 2017, the authors reported that 1.3% of HCM patients (n=303) carried CNVs, which included 2 deletions in *MYBPC3* and *PLN*, respectively ²⁵⁹. Therefore, these studies highlight that CNVs could explain a very small proportion of patients with an unexplained genetic aetiology of HCM, which casts doubt on the added value of incorporating them in routine clinical genetic testing.

In addition, common variants with low effect sizes may explain the disease phenotype in patients with negative genetic testing results. However, evidence on the contribution of

these variants to HCM is limited as it would require large-scale case-control studies to achieve sufficient statistical power to robustly test genetic associations with HCM. Currently, only 1 GWAS study reported a significant association of a common intronic variant in *FHOD3* (rs516514) with HCM ¹⁸⁶. More GWAS studies are required to be able to assess the role of common variants in HCM including disease modifiers, which affect the disease expression of HCM. The genotyping array developed by H3Africa, for example, provides a valuable resource for identifying novel genetic associations with HCM in individuals of African ancestry ¹⁶⁸.

GWAS studies could potentially investigate the combined effects of common variants and other risk factors, such as HTN and obesity, both of which are highly prevalent in MENA populations compared to Europeans, on HCM ^{260,261}. Also, given the high rate of consanguinity in MENA populations, future genetic research studies could investigate the role of common homozygous variants in sarcomere genes in HCM, which are benign when carried by an individual in the heterozygous state ¹³⁶.

Finally, WES or WGS approaches could be carried out to identify novel gene-HCM associations. The screening for additional putative pathogenic genes using these sequencing approaches opens avenues for studying their potential contribution to HCM through large-scale case-control burden testing and/or co-segregation analysis in patients' families ⁹⁴. Also, WGS could allow for the detection of deep intronic variants that could be associated with HCM. In fact, a study by Bagnall *et al.*, 2018 reported that WGS increased the diagnostic yield of genetic testing by 20% and that deep intronic variants accounted for almost 50% of this proportion ²⁶². However, the analysis of the huge genetic data obtained from these approaches is computationally complex to analyse and will require extensive case-control cohorts to be able to identify a significant association of a novel, low disease frequency gene with HCM. Therefore, patients with a non-identifiable genetic aetiology, who have a positive FH of HCM or SCD, could be prioritised for WES or WGS. In the AhcHCM cohort, 27.5% of patients with non-identifiable genetic aetiology (i.e. genotype-negative) had a FH of HCM. Hence, they should be considered for WES

or WGS in the future. Also, genotype-negative patients with healthy parents could provide suitable candidates for these approaches to identify putative pathogenic variants that were inherited *de novo* in an affected child or to test for recessive inheritance.

3.5.5 Homozygosity is enriched in AhcHCM patients compared to RbhHCM

Given the high rate of consanguinity (30.6%) observed in the AhcHCM cohort, I evaluated whether homozygous variants were enriched among Egyptian HCM patients compared to RbhHCM. A significantly ($p\text{-value} < 0.0001$) higher burden of homozygous variants in sarcomere genes was observed in AhcHCM patients (3.11%) compared to RbhHCM patients (0.15%). Previous studies have also identified homozygous variants in sarcomere genes in HCM patients. For example, the first homozygous HCM variant (p.Ser179Phe) was identified in *TNNT2* by Ho *et al.*, 2000^{171,254}. The patient carrying this homozygous variant presented with severe HCM morphological abnormalities and died suddenly at the age of 17 years¹⁷¹. Subsequent studies demonstrated that HCM patients with more than 1 causative variant, including homozygotes, compound heterozygotes (>1 variant in the same gene) or double heterozygotes (>1 variant in different genes), were at a higher risk of developing heart failure and SCD^{263–266}. Patients with these “double hits” constitute a small proportion of HCM cases accounting for only 5% of familial HCM²⁶³. The analysis of the AhcHCM cohort showed that homozygous patients were more likely to have consanguineous parents compared to heterozygous patients (53.3% % vs. 29.3% %, $p\text{-value}=0.08$). However, no differences in the age at disease onset and disease severity were observed between both patient groups. It is worth mentioning, that due to the limited number of homozygotes ($n=21$) with rare sarcomeric variants as well as the incomplete clinical data I could not make any conclusions about the effect of homozygosity on disease severity in the analysed AhcHCM cohort. Also, a more accurate analysis would require a comparison between heterozygous and homozygous patients carrying the same variants or same variant class (e.g. truncating variants in *MYBPC3*). This analysis could be performed once a larger AhcHCM cohort size with complete clinical data is achieved.

In addition, rare sarcomeric homozygous variants were observed in *MYH7*, *MYBPC3*, *MYL2*, *MYL3* and *TNNI3* genes in the AhcHCM cohort. The majority (71.4%) of these variants were missense but 2 probands with homozygous truncating variants in *MYBPC3* were identified. Very few studies of homozygous truncating variants in *MYBPC3* have been reported in the literature, where they have been shown to be associated with severe infantile cardiomyopathy leading to early death ^{267–269}. For example, a study by Xin *et al.*, 2007 reported that 3 neonates with the homozygous splice variant c.3330+2T>G in *MYBPC3* presented with severe HCM and all died within 3-4 months ²⁶⁸. The mutant mRNA transcript was predicted to produce a truncated protein with a complete loss of the C10 domain, which is essential for normal *MYBPC3* function ²⁶⁸. In the work presented here, however, the AhcHCM homozygote carrying the splice acceptor *MYBPC3* variant (c.1227-2A>G) presented with HCM at the age of 25 years (clinical data was not available for the other homozygote *MYBPC3*-truncating variant carrier). Thus, the molecular mechanism of this variant remains to be studied to better understand why the homozygous c.1227-2A>G variant did not result in lethal HCM, as would have been predicted based on the literature findings. It is noteworthy, that the AhcHCM homozygote with the c.1227-2A>G variant was born to consanguineous parents, however, clinical and genetic data were not available for the parents. This highlights the importance of clinical genetic testing coupled with phenotyping of consanguineous parents to be able to provide informative genetic counselling on the disease risk of unborn children. In a country like Egypt, where consanguineous marriages account for almost 35%, premarital genetic counselling may support parents in the decision making around having children ^{103,104}.

Furthermore, the most prevalent (0.58%, n=3) homozygous variant among AhcHCM patients was the missense variant p.Glu170Gln in *MYL3*. This variant is novel as it has not been reported before and was absent from EHVols as well as from the reference population databases gnomAD, GME and Iranome. It was also absent from LMM/OMGL and RbhHCM patients. In addition to the 3 homozygous p.Glu170Gln patients, 3 other patients were found to carry the variant in the heterozygous state. The enrichment of p.Glu170Gln in AhcHCM patients over EHVol controls suggests that it could be likely

pathogenic. Due to incomplete clinical data I was unable to compare the HCM phenotypes between the homozygous and heterozygous groups. However, segregation analysis could provide deeper insights as to whether the variant could be more penetrant when expressed in homozygosity.

3.5.6 Identification of 2 novel *CSRP3* biallelic truncating variants in Egyptian patients provides deeper insights into the mutational spectrum and disease mechanism of *CSRP3* variants

I then evaluated whether rare homozygous variants were identified in genes moderately associated with HCM (*ACTN2*, *CSRP3*, *JPH2* and *TNNC1*) in AhcHCM patients. Intriguingly, 2 homozygous truncating variants were identified in *CSRP3* (c.415-1G>C (n=1) and c.414+1G>T (n=3)). To date, only 3 knockout *CSRP3* HCM patients have been reported in the literature and all had consanguineous parents ^{234,235}. In the AhcHCM cohort, only 1/5 *CSRP3* knockout patients were born to consanguineous parents. The patient carrying the splice acceptor variant c.415-1G>C was diagnosed with HCM (with LVMWT of 18mm) at the age of 47 and had no FH of HCM or SCD. The other 3 *CSRP3* knockout patients with the splice donor variant c.414+1G>T were diagnosed with HCM at an average age of 45±10.6 and presented with an average LVMWT of 16±2.6. All 3 patients had no FH of HCM but 1 patient had a FH of SCD. It is noteworthy that 1 AhcHCM patient carried the c.414+1G>T variant in the heterozygous state. In contrast to the homozygous carriers, the heterozygous patients presented with HCM at a more advanced age (74 years). An EHVol control also carried the c.414+1G>T variant in the heterozygous state and showed no signs of LVH. This suggests that truncating variants in *CSRP3* might be associated with variable penetrance and expressivity when expressed in the heterozygous state but are more pathogenic when expressed in the homozygous state.

3.5.7 “Non-rare” variants in *MYL3* could become penetrant when expressed in homozygosity

Finally, I sought to assess whether applying the cut-off allele frequency for recessive inheritance ($\text{gnomAD FAF}_{\text{popmax}} \leq 0.00126$), could identify AhcHCM patients with variants that would be “too common” to cause HCM in an autosomal dominant fashion.

Only 2 “non-rare” homozygous variants in *MYL3* (p.Glu177Gly and p.Ala57Asp) were identified in 4 and 2 AhcHCM patients, respectively. These variants were identified in only 1 AhcHCM patient each in the heterozygous state. Clinical data was available for patients carrying the p.Ala57Asp variant who presented with HCM at a younger age and showed a greater LVH compared to the heterozygous carrier. A study by Jaafar *et al.*, 2016 reported the p.Ala57Asp variant in a 45-year-old Tunisian female of consanguineous parents presenting with a septal thickness of 28mm and dyspnoea ¹⁷⁰. The authors could not confirm the presence of the p.Ala57Asp variant in the parents but they were believed to be heterozygous carriers. Given that the mother was asymptomatic, Jaafar *et al.*, suggested that the variant could have caused autosomal recessive HCM. However, the presence of p.Ala57Asp in a heterozygous AhcHCM patient, who did not carry any other rare variant in HCM genes, suggests that additional genetic and/or non-genetic factors could explain the variable penetrance observed between the Tunisian and Egyptian heterozygous p.Ala57Asp-carriers.

3.5.8 Study limitations

The rare variant burden analysis performed in this study provided a robust statistical approach to evaluate the genetic architecture of HCM in Egypt and has highlighted key differences compared to predominantly Caucasian populations. However, several limitations persist. Firstly, the limited AhcHCM cohort size is underpowered for testing the burden of rare variation in minor HCM genes (e.g. *ACTN2*, *JPH2*, *TNNC1* and *PLN*). Secondly, the study design did not account for CNV, non-coding variants or epigenetic factors, which may contribute to HCM pathogenesis. Also, common heterozygous variants of small effect sizes were not analysed. In addition, the TruSight Cardio

Sequencing Panel does not include the *FLNC* gene, which is implicated in myofibrillar myopathy but has been recently validated for its definitive association with isolated LVH in families ^{148,270,271}.

The AhcHCM cohort is also predominantly composed of relatively young HCM patients with a mean age at diagnosis of 34.7 ± 17.7 years. This means that the study cohort contained few patients with late disease onset, who could be carrying low penetrant variants associated with mild HCM phenotypes. Finally, not all AhcHCM patients underwent CMR as a subset of patients were screened at AHC's satellite clinics in Cairo, Alexandria and Menofeya, which do not have MRI units. Thus, I was unable to correlate CMR findings pertaining to indexed LVM or degree of myocardial fibrosis (quantified by late gadolinium enhancement; LGE), both of which have been reported to be associated with disease severity, with genetic findings ^{272,273}. Recently, a high level of myocardial fibrosis detected by CMR with LGE has also been associated with an elevated risk of SCD so it could aid in patient risk stratification ³⁹.

3.6 Conclusions

The advent and improvement of novel tools and approaches for the interpretation of HCM genetic variation have substantially improved our understanding of the genetic architecture of the disease. Major contributions to such advances include the development of standardised guidelines and quantitative approaches to variant interpretation, the establishment of comprehensive large-scale reference population databases and knowledge sharing via publicly available resources. However, our knowledge has mainly been derived from studying Caucasian populations with distinct genetic, phenotypic and cultural features compared to MENA populations.

In this study, I demonstrated the influence of ethnicity on genetic variation in HCM. Therefore, it is imperative that future studies address the largely neglected Egyptian population. Such studies could aid in defining the clinical relevance and actionability of variants identified in HCM patients and thus set the foundation for precision medicine in the region.

4 ANALYSIS OF HCM IN THE EGYPTIAN POPULATION REVEALS A NEW MECHANISM OF PATHOGENICITY

4.1 Aims and hypotheses

Analysis of the genetic architecture of HCM in the Egyptian population (chapter 3) revealed that a specific frameshift variant in *MYH7* (c.5769delG) was commonly observed among AhcHCM patients (n=17/514, 3.31%). This finding is intriguing as mutant *MYH7* transcripts harbouring premature termination codons (PTCs), that are introduced by heterozygous variants (i.e. frameshift, nonsense and splice variants), are believed to be eliminated via nonsense-mediated decay (NMD) to prevent disease¹⁰⁰. Also, haploinsufficiency in *MYH7* has not been recognised as a disease mechanism in HCM. Thus, I hypothesised that the c.5769delG variant could have contributed to HCM by escaping NMD and producing a disease-causing, truncated protein. To test this hypothesis, the aims of this chapter are to:

- Evaluate the burden of c.5769delG in AhcHCM patients compared to local and global ethnically-matched controls
- Estimate its population-level penetrance
- Assess the cosegregation of c.5769delG with HCM in a large AhcHCM family
- Compare the disease phenotypes between c.5769delG-carriers and other AhcHCM patients with different genetic aetiologies
- Perform RNA sequencing analysis to study the disease mechanism of c.5769delG
- Compare the burden of c.5769delG in AhcHCM patients with both LMM/OMGL and RbhHCM cohorts

4.2 Introduction

4.2.1 Overview of *MYH7*

The *MYH7* gene is located on chromosome 14q11.2–q13 in humans and encodes for β (or slow) myosin heavy chain. It consists of 40 exons forming a transcript composed of 6,027bp, which produces a 1,935 amino acid protein ²⁷⁴. It is expressed in the heart ventricles as well as in type I muscle fibres found in skeletal muscle tissues (Figure 4-1) ^{275,276}. β -myosin heavy chain forms a component of the larger type II myosin protein ²⁷⁷. Each type II myosin protein is composed of 2 β -myosin heavy chains, 2 essential light chains and 2 regulatory light chains, produced by *MYL3* and *MYL2*, respectively. The heavy chains are composed of 2 globular heads attached to a myosin tail, which consists of an α -helical-coiled coil forming the thick filament's backbone ^{276,278} (Figure 4-2A). Subfragment 1 (S1), at the N-terminus, forms the globular heads, which contain a motor domain and the lever arm ²⁷⁸. The motor domain harbours actin- and ATP binding sites, which are responsible for generating movement ^{275,276,279}. Following the motor domain is the lever arm, which consists of the converter and neck domains. The lever arm transmits the movement from the globular heads thereby promoting the sliding of the thick filament over the thin filament to facilitate contraction ²⁸⁰. Within the lever arm, the neck region contains the regulatory domain, to which the essential- and regulatory light chains bind ²⁷⁸. The lever arm runs into the myosin coiled-coil tail, which consists of subfragment 2 (S2) and light meromyosin (LMrM) ^{281,282}. LMrM constitutes the largest portion of the myosin tail (extending from ca. amino acid residues 1140-1935) and contains binding sites for *MYBPC3* and Titin (*TTN*), a giant protein that spans the sarcomere and mediates sarcomere assembly, stability and maintenance ^{5,275,281,283}. S2 and LMrM are connected via a hinge, which enables the myosin thick filament to interact with the actin thin filament to generate contractile force ²⁸¹. It includes a heptad repeat sequence (*abcdefg*), which represents the hallmark of coiled-coil molecules ^{281,284,285}. The alternating pattern of negatively and positively charged residues in the heptad enable the interaction between neighbouring myosin molecules in the thick filament ²⁸¹. For example, positions *a* and *d*

harbour hydrophobic residues, which stabilise the interaction between the α -helical-coiled coils²⁸⁵. The hydrophilic residues at *e* and *g* positions form salt bridges to stabilise the myosin molecule's hydrophobic core²⁸⁵. These heptad sequences are interrupted by a highly conserved 'skip' residue, at positions Thr1188, Glu1385, Glu1582 and Gly1807, which are believed to provide flexibility into the myosin tail and assist its assembly into the thick filament (Figure 4-2B)^{281,284,286,287}.

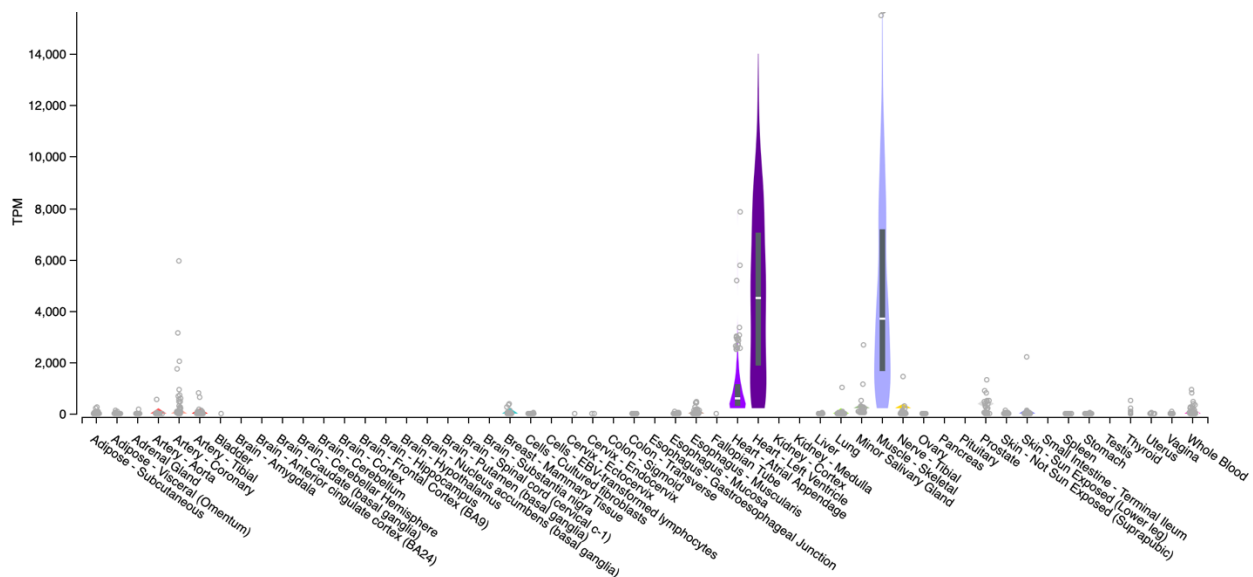


Figure 4-1: Expression level of *MYH7* in the human heart (dark purple) and skeletal muscle (light purple). In the heart, *MYH7* is mainly expressed in the ventricular muscle (shown here for LV). It is also expressed in skeletal muscle tissues. Figure from the GTEx portal (<https://gtexportal.org/>).

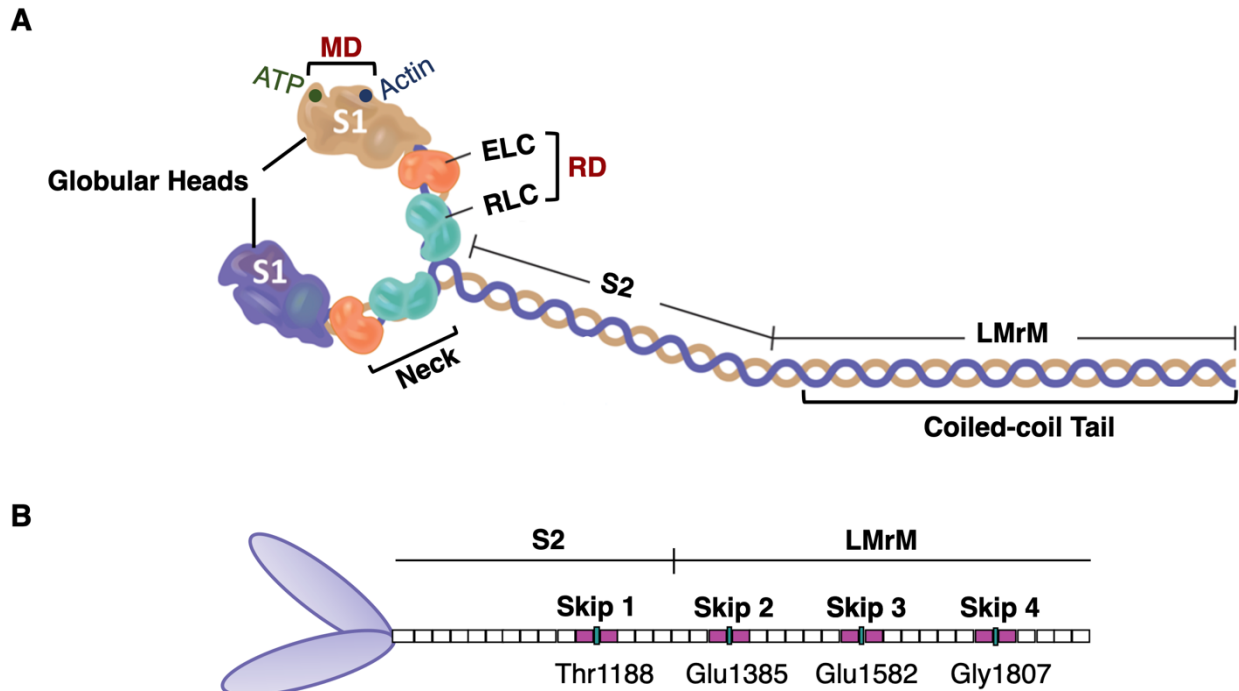


Figure 4-2: (A) Schematic diagram of type II myosin protein and (B) the relative locations of the 4 highly conserved 'skip' residues in the myosin tail. These residues are predicted to provide flexibility into the myosin tail to allow its assembly into the thick filament. S1/S2: subfragment 1/2; MD: motor domain; RD: regulatory domain; ELC: essential light chain; RLC: regulatory light chain, LMrM: light meromyosin. Figure A was adapted from Viswanathan *et al.*, 2017 under a Creative Commons license ²⁸⁸. Figure B was reproduced from Taylor *et al.*, 2015 under a Creative Commons license ²⁸⁴.

4.2.2 Conformational changes in myosin facilitate contraction/relaxation

Cardiac contractile force is generated when the thick filaments slide past the actin-containing thin filaments ²⁸¹. During systolic contraction, the motor domain undergoes a conformational change to hydrolyse ATP resulting in an increased binding affinity between myosin and actin, which is located in the sarcomere's thin filament ²⁷⁸. This interaction results in a powerstroke, which transduces force by rotation of the converter domain and its associated lever arm ca. 70° from the pre-powerstroke position to align with the flat surface of the globular head called the myosin mesa (Figure 4-3) ^{278,279}. Following contraction, the globular heads rebind ATP, disintegrate from actin and the myosin mesa is no longer aligned with the converter domain resulting in a relaxed state ^{278,279}.

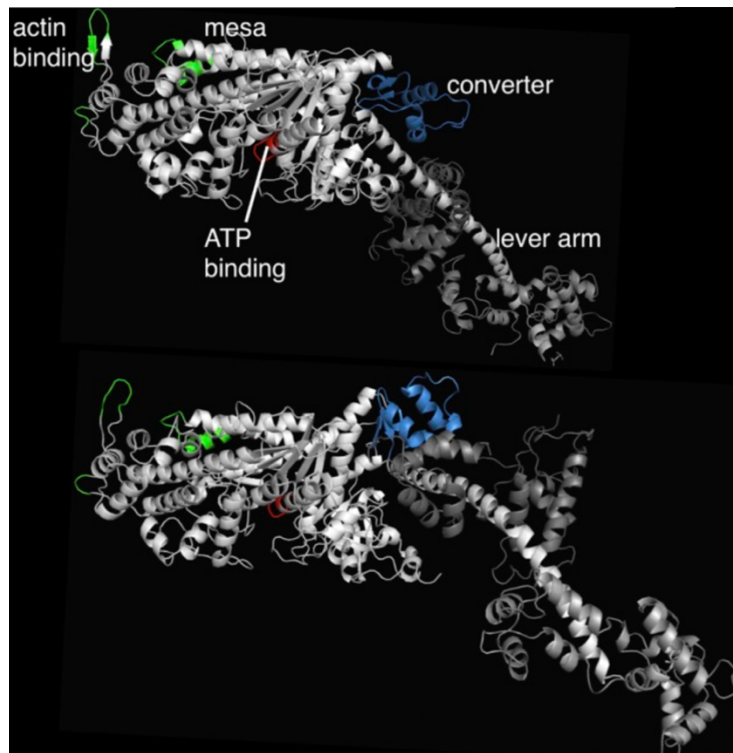


Figure 4-3: Crystal structures of myosin in the poststroke (top) and the pre-stroke (bottom) states. In the poststroke (i.e. relaxed) state, the motor domain falls out of alignment with the converter domain (blue) and is no longer accessible for inter- or intramolecular interactions. In the pre-stroke state, the converter domain and its associated lever arm rotate $\sim 70^\circ$ to align with the flat surface of the myosin mesa leading to force transduction (i.e. powerstroke/contraction). From: Homburger *et al.*, 2016 ²⁷⁹.

A study by Spudich *et al.*, 2014 reported that only 10% of myosin molecules are predicted to be involved in sarcomere force production ²⁸⁹. This suggests that a specific mechanism prevents unwanted myosins to interact with actin causing unnecessary energy expenditure. Indeed, recent studies have demonstrated that during diastolic relaxation, the globular heads of paired myosin molecules interact through the interacting-head motif (IHM) to adopt either an energy-saving, super relaxed state (SRX) conformation or an energy-consuming, disordered relaxed state (DRX) conformation ^{278,290,291}.

The evolutionary conserved IHM forms through the interaction between the actin-binding domain of the 'blocked' myosin head (BH) and the converter domain of the myosin 'free' head (FrH) (Figure 4-4) ²⁷⁸. The IHM induces a SRX state when the FrH attaches itself tightly to the BH thereby inhibiting ATPase activity and preventing the movement of the

converter domain, which is required for myosin-actin binding ^{278,292}. Thus, the adoption of a SRX conformation is believed to facilitate an energy-saving mechanism ²⁷⁸.

The IHM conformational change to the DRX state, in contrast, allows the FrH to sway freely in order to hydrolyse ATP and bind to actin ²⁹³. The swaying myosin FHs consume 5-fold greater ATP compared to those in the SRX state, thus, the DRX conformation has been proposed as an energy-consuming mechanism ²⁷⁸.

The regulated changes in ATP expenditure; from lowest consumption in the SRX state to higher consumption; in the DRX state, is essential to achieve balanced myocardial energetics and regulated contraction-relaxation dynamics ²⁷⁸. Therefore, HCM variants that disrupt the physiological balance of myosins between the 2 IHM conformations could result in elevated ATP consumption leading to hypercontractility, decreased relaxation, elevated metabolic demands and an increased risk of heart failure ²⁷⁷.

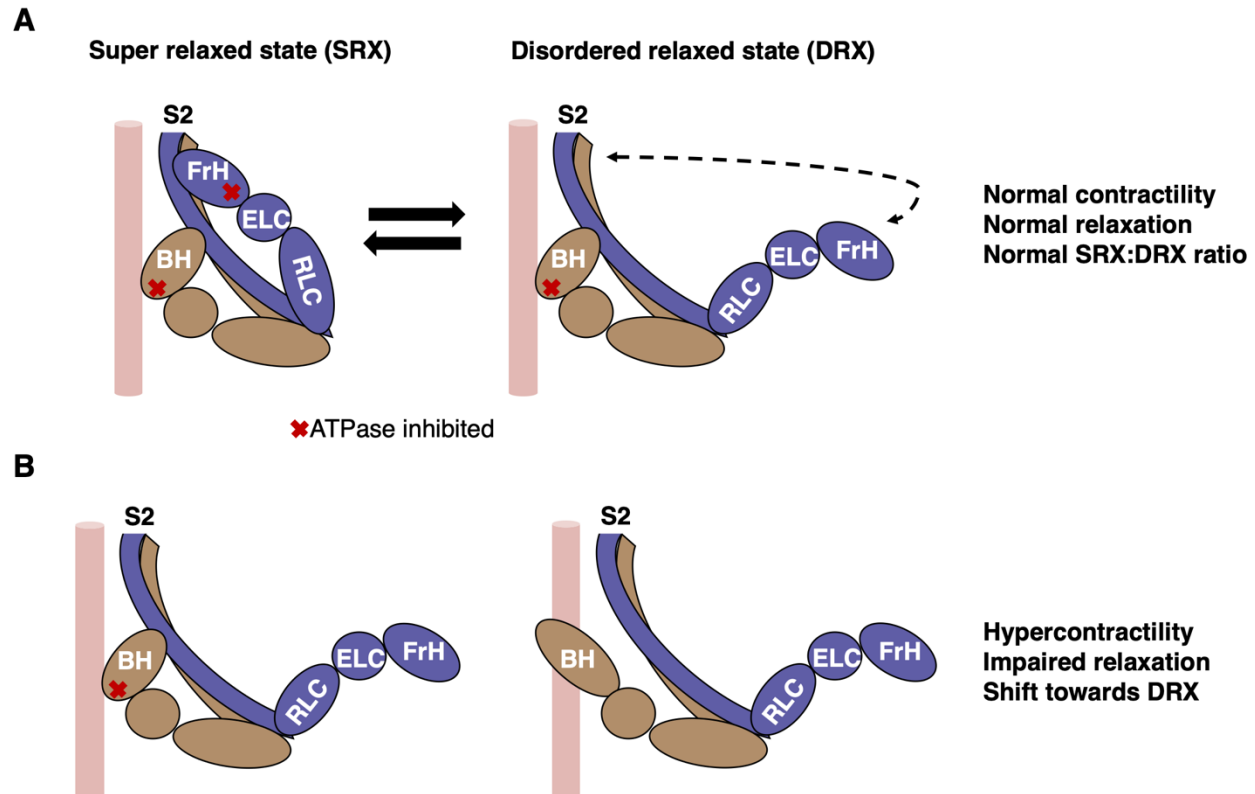


Figure 4-4: Myosin IHM conformations in the super relaxed- and disordered relaxed states (SRX and DRX) in (A) healthy and (B) HCM cardiac muscle. (A) During relaxation, IHM of the healthy cardiac muscle forms through the interaction between the actin-binding domain of the ‘blocked’ myosin head (BH) and the converter domain of the myosin ‘free’ head (FrH). The IHM adopts a SRX state when the FrH attaches itself to the BH to inhibit ATPase activity and myosin-actin interactions by preventing the movement of the converter domain. In the DRX state, FrH sway freely to hydrolyse ATP thereby generating force. **(B)** In HCM, *MYH7* variants destabilise IHM interactions resulting in the reduction of the myosins in the energy-saving SRX state and increase of myosins in the energy-consuming DRX state leading to hypercontractility and impaired relaxation. Reproduced from Alamo et al., 2017 under a Creative Commons license ²⁷⁸.

4.2.3 HCM-causing variants are enriched in specific regions of *MYH7*

Given the distinct properties of the globular head domain (S1), the neck region (S2) and the tail (LMrM) of myosin, it has been proposed that variants may have variable effects on protein function depending on their regional location ²⁷⁴. Recent studies have shown that HCM-causing variants in *MYH7* are typically enriched in the myosin mesa of the globular head and the converter domain, both of which participate in motor domain function and IHM interactions (refer to section 4.2.1) ^{274,278,279,294}. Variants that affect the motor domain function by abnormally increasing sarcomere force, which explains the hypercontractile phenotype observed among HCM patients ²⁹⁵. Also, *MYH7* variants that

alter IHM interactions reduce the SRX to DRX ratio resulting in an increased ATP consumption and altered relaxation, which is another hallmark of HCM ²⁷⁸. In addition, analysis of the distribution of *MYH7* missense variants in 2,913 patients included in the SHaRe database, demonstrated that the proximal S2 domain, which binds to the myosin mesa, is also enriched for HCM variants ²⁷⁹. Variants in this domain typically occur in aspartate and glutamate residues, which are complementary to the positively charged cluster of arginine residue charges on the myosin mesa ²⁹¹. These variants were proposed to disrupt the interaction between myosin and *MYBPC3* or other sarcomeric proteins thereby contributing to HCM pathogenesis ²⁷⁹. Also, a recent study by Walsh *et al.*, 2019, demonstrated that missense *MYH7* variants identified in the LMM/OMGL cohort were significantly enriched within 167-931 amino acid residues, a region that includes the mesa, IHM interactions sites, converter domain and the proximal S2 domain, compared to ExAC controls ²⁰³. Similarly, analysis of 514 AhcHCM patients, showed a similar clustering pattern of *MYH7* missense variants to LMM/OMGL patients (as illustrated in section 3.4.7). Collectively, these observations suggest that specific domains in *MYH7* represent mutational hotspots and that variants in each domain may offer insights into the mechanism by which they lead to HCM pathogenesis.

Moreover, other studies have proposed that charge-changing *MYH7* variants at sites involved in IHM interactions could affect disease severity ^{296–299}. For example, a recent study by Toepfer *et al.*, 2020 showed that HCM patients with *MYH7* VUSs, which changed the charge of AA residues involved in IHM interactions, such as R17C and A226V, were at a higher risk of developing atrial fibrillation and heart failure compared to patients with other *MYH7* VUSs ²⁷⁷. That is because these variants increase the DRX to SRX ratio leading to hypercontractility, impaired relaxation and increased ATP consumption and metabolic demands, which contribute to severe disease phenotypes ²⁷⁷. This explains why application of the ATPase inhibitor MYK-461 decreased the DRX:SRX ratio and hypercontractility in mouse cardiomyocytes ²⁷⁷. Thus, these findings could improve our interpretation of *MYH7* VUSs by evaluating their involvement in IHM interactions.

Finally, the *MYH7* tail, which lacks functional sites, has been associated with almost 5-fold lower density in HCM variants compared to the head domain ²⁷⁴. However, the *MYH7* tail is a highly conserved region of the gene with repeated patterns of the 7 heptad residues ^{274,300}. These heptad repeats are responsible for the coiled-coil structure of the myosin tail ³⁰¹. Thus, disruption of these heptad residues may cause HCM by interfering with normal thick filament structure or assembly in sarcomeres ²⁷⁴. For example, a study by Wolny *et al.*, 2013 showed that the 5 HCM-causing *MYH7* variants residing in the tail region (N1327K, E1356K, R1382W, E1555K, and R1768K) disrupted the helical structure of the thick filaments thereby reducing their ability to incorporate into the sarcomere ³⁰². The authors also suggested that the region close to the variant N1327K could be crucial for the formation of the coiled-coil structure and that this region is in close proximity to a mutational hotspot in the tail ³⁰². However, none of these variants had a significant effect on contraction ³⁰². This suggests that *MYH7* variants residing in the tail domain do not have a direct effect on force generation but rather destabilise thick filament assembly thereby affecting the ability of myosin to interact with other sarcomeric proteins (e.g. *MYBPC3*, *TTN*). Consistent with this notion, Armel and Leinwand 2010, demonstrated that the highly conserved E1356K variant at the *c* position of the heptad sequence may have driven the HCM phenotype by destabilising the coiled-coil structure thereby affecting thick filament assembly ³⁰³. In addition, a study by Colegrave and Peckham, 2014 reported that *MYH7* variants clustered in 2 specific regions in the tail ²⁸². The first region, starting from AA residue 1200-1450, has been shown to contain mostly HCM variants whereas the second region, harbouring the last 200 AA residues, contains DCM, HCM and skeletal muscle variants ²⁸². In contrast, the regions of the *MYH7* tail believed to interact with Myomesin (*MYOM*) (1506-1674) and *MYBPC3* (1554-1581) have been shown to harbour a limited number of variants ^{282,304,305}. These findings suggest that a disruption of the interaction of *MYH7* with these proteins may result in a lethal phenotype. Finally, variants in distal end of the *MYH7* tail, to which *TTN* binds and is responsible for filament formation, have been reported to be rare ²⁸². Taken together, these studies have provided insights into the distribution of HCM variants in the tail region, yet little is known

about the mechanistic effects of these variants ²⁷⁶. Thus, biophysical and biochemical studies are warranted to elucidate their contribution to the HCM phenotype.

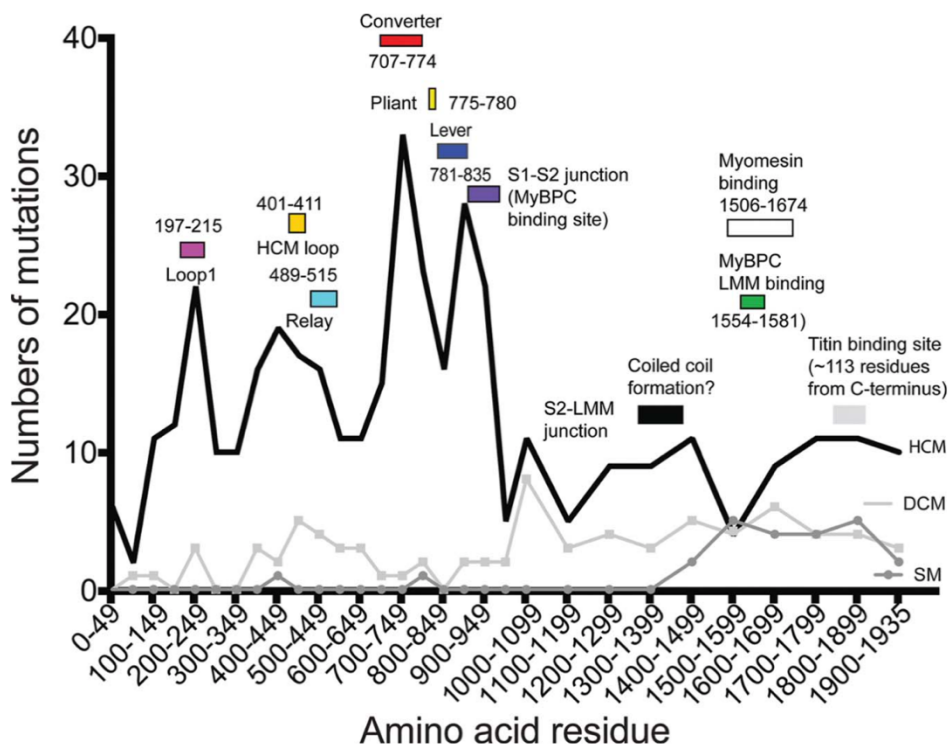


Figure 4-5: Graph showing the number and relative positions of *MYH7* variants identified in HCM (thick black line). The x-axis depicts stretches of 50 amino acids (AA) for sub-fragments 1 and 2 then for 100 AA stretches for the tail region. *MYH7* variants in HCM are enriched in the converter domain (red) and at the S1-S2 junction (purple). From Colegrave and Peckham, 2014, with permission ²⁸².

4.2.4 Adapted ACMG/AMP guidelines for the interpretation of *MYH7* variants in HCM

Missense variants in *MYH7* constitute a major genetic cause of different types of cardiomyopathies, including HCM and DCM ²⁰². However, the mechanism of these variants varies according to the disease context. For example, *MYH7* variants contribute to HCM and DCM pathogenesis by abnormally increasing or reducing cardiac contractility, respectively ³⁰⁶. Since *MYH7* is routinely tested in the clinical setting, ClinGen's Inherited Cardiomyopathy Expert Panel (CMP-EP) sought to re-evaluate the

relevance of the ACMG/AMP guidelines to accurate variant interpretation of *MYH7* variants identified in cardiomyopathies. Of the original 28 ACMG/AMP rules, the panel removed 9 rules, which were regarded as non-applicable to *MYH7*-associated cardiomyopathies, and adjusted 17 rules to enhance their applicability to *MYH7*-associated cardiomyopathies (Figure 4-6) ²⁰². For example, the very strong PVS1 rule “*Predicted null variant in gene where LOF is a known mechanism of disease*” was removed as LoF variants are very rarely observed in patients and there is currently no evidence supporting their contribution to cardiomyopathy ²⁰². Moreover, specific rules, including: PS3, PS4 and PM2, were adjusted to be more applicable to *MYH7*-associated cardiomyopathies. For example, the original ACMG/AMP PS3 rule, which states: “*Well-established functional studies show a deleterious effect*” was adjusted to apply only to variant-specific knock-in models as they provide accurate evidence on variant pathogenicity ²⁰². The PS4 rule related to the significant enrichment of a variant in patients over controls was assigned different strength levels depending on the number of probands identified with the variant. For example, a rare variant identified in ≥ 15 probands with the same disease phenotype would provide strong evidence of pathogenicity (PS4) whereas a variant identified in ≥ 6 probands would activate the PS4_Moderate rule ²⁰². The CMP-EP also recommended activating the PM2 rule if a variant had a FAF of 4×10^{-5} in population databases, which was based on realistic prevalence and penetrance values, rather than following the original, overly conservative ACMG/AMP PM2 rule which is only applied if the variant is absent in population databases ^{202,307}. Finally, the CMP-EP defined certain thresholds for increasing strength of co-segregation (PP1) evidence as the ACMG/AMP guidelines did not provide criteria for an increasing degree of segregation ¹⁰². In the adapted ACMG/AMP framework supporting, moderate and strong evidence of pathogenicity was activated in the presence of 3 meioses/LOD 0.9, 5 meioses/LOD 1.5 and 7 meioses, LOD 2.1, respectively ²⁰².

Collectively, these adaptations to the original ACMG/AMP framework provide a more accurate interpretation of *MYH7* variants associated with inherited cardiomyopathies. Applying these adapted rules in the clinical setting, could therefore greatly reduce the

discordance in variant interpretation among different laboratories and also enhance the actionability of variants identified during clinical screening.

Pathogenic Criteria				Combination of Criteria Needed to Meet Classification							
Rule		Specification Type	Rule Description	Pathogenic				Likely Pathogenic			
VS	PVS1	Removed	~~Null variant in gene with established LOF as disease mechanism~~								
	PS1	No change	Different nucleotide change (same amino acid) as previously established pathogenic variant								
	PS2	Disease/gene	<i>De novo</i> (paternity confirmed) in a patient with disease and no family history								
	PS3	Disease/gene	Functional studies of mammalian knock-in models supportive of a damaging effect on the gene or gene product	2	1	1	1	1	1		
	PS4	Disease/gene	Prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls –OR– Variant identified in ≥15 probands with consistent phenotypes								
	PP1_Strong	Modif. Strength	Variant segregates with ≥7 meiosis								
Moderate	PM1	Disease/gene	Hotspot/est. functional domain (amino acids 181-937) without benign variation								
	PM2	Disease/gene	Absent/extremely rare (<0.004%) from large population studies								
	PM3	Removed	~~Detected in trans with a pathogenic variant (recessive)~~								
	PM4	No change	Protein length changes due to in-frame deletions/insertions of any size in a non-repeat region or stop-loss variants								
	PM5	No change	Missense change at an amino acid residue where a different missense change previously established as pathogenic	3	2	1	1		3	2	
	PM6	Disease/gene	Confirmed <i>de novo</i> without confirmation of paternity								
	PVS1_Moderate	Modif. Strength	Null variant in gene with evidence supporting LOF as disease mechanism								
	PS4_Moderate	Modif. Strength	Variant identified in ≥6 probands with consistent phenotypes								
	PP1_Moderate	Modif. Strength	Variant segregates in ≥5 meiosis								

Figure 4-6: Summary of adapted ACMG/AMP classification rules by the ClinGen Inherited Cardiomyopathy Expert Panel for *MYH7* variants identified in patients with inherited cardiomyopathy. VS: very strong. Removed - not applicable to *MYH7*-associated cardiomyopathy. Modif. strength: Modified rule strength. Numbers under each classification refer to the number of rules with P/LP strength required to classify the variant as P/LP. Reproduced from Kelly et al., 2018 under a Creative Commons license ²⁰².

4.2.5 Molecular mechanisms of *MYH7* variants in HCM

Missense variants in *MYH7* are believed to cause HCM by exerting gain-of-function effects, whereby variants produce an abnormal activated protein that incorporates into the sarcomere as a “poison peptide”²⁴². In contrast, variants that introduce PTCs in transcripts are believed to elicit NMD and thus are considered non-causative in HCM¹⁰⁰. Additionally, *MYH7* has not been recognised to cause HCM through haploinsufficiency, which occurs when the amount of functional protein is insufficient for normal sarcomere function⁴⁷.

The post-transcriptional surveillance mechanism of NMD is essential for RNA quality control and gene regulation³⁰⁸. This mechanism was initially discovered in yeast after recognising that cells often contain extremely low concentrations of mRNAs that are transcribed from alleles harbouring truncating variants³⁰⁹. NMD identifies newly synthesised mRNA transcripts carrying PTCs and subsequently targets them for degradation to prevent the production of truncated proteins which could result in disease.

NMD was initially discovered in β -thalassaemia and has later on been described for other genetic diseases, such as cystic fibrosis, Duchenne muscular dystrophy and cardiomyopathy^{162,310}. For example, in DCM, nonsense variants in Lamin A/C (*LMNA*) were associated with significantly lower levels of mutant transcripts harbouring PTCs, which supports the involvement of NMD³¹⁰. Reduced protein levels in HCM patients with truncating variants in *MYBPC3* also suggested the involvement of NMD leading to haploinsufficiency⁵⁴.

NMD also targets almost 10% of wild type mRNA transcripts to promote crucial cellular processes including differentiation, maturation and apoptosis^{162,311}. This mechanism is conserved among eukaryotes and the signals that trigger its activation or inhibition vary between mutant and non-mutant transcripts³⁰⁸.

4.2.5.1.1 Overview of the NMD pathway

NMD is a complex mechanism that accurately differentiates between normal stop codons and PTCs on a mRNA transcript then recruits specific enzymes that target the PTC-harbouring transcripts for degradation. There are 2 main pathways of NMD: the exon-junction complex (EJC) pathway and the long 3' UTR pathway.

The EJC pathway is initiated during pre-mRNA splicing when the EJC is deposited ~20-24bp upstream of every exon-exon junction ³¹⁰. The function of the EJC is to identify PTCs. During translation, ribosomes scan the mRNA transcript and pause at a stop codon. If the EJC is present >50-55bp downstream of the stop codon, components of the NMD-inducing complex (called the SURF complex), which comprises: suppressor with morphogenetic effect on genitalia (SMG1), up-frameshift protein 1 (UPF1), and the eukaryotic release factors 1 and 3 (eRF1 and eRF3), are recruited to the ribosome (Figure 4-7) ¹⁶². Phosphorylation of UPF1 by SMG1 triggers the dissociation of the eRF1 and eRF3 and the subsequent recruitment of endonuclease SMG6 or the SMG5-SMG7 complex, which promote mRNA degradation ^{162,308,310}. mRNA degradation can occur through 2 main pathways. The recruitment of the endonuclease SMG6 initiates the cleavage of the target transcript in the vicinity of the PTCs leading to the generation of unprotected RNA ends that are then degraded by cellular nucleases ^{162,308}. mRNA degradation can also occur via the recruitment of the SMG5-SMG7 complex which induces RNA decapping and deadenylation by recruiting additional factors that expose the transcript ends to cellular exonucleases ^{162,308}.

In contrast, mRNA transcripts harbouring PTCs in the last exon (and thus downstream of last exon-exon junction) or $\leq 50-55$ bp upstream of the last exon-exon junction are believed to be NMD-incompetent as they fail to elicit NMD and thus produce truncated proteins (Figure 4-8) ¹⁶². These proteins could have deleterious functional effects, which may contribute to severe disease phenotypes.

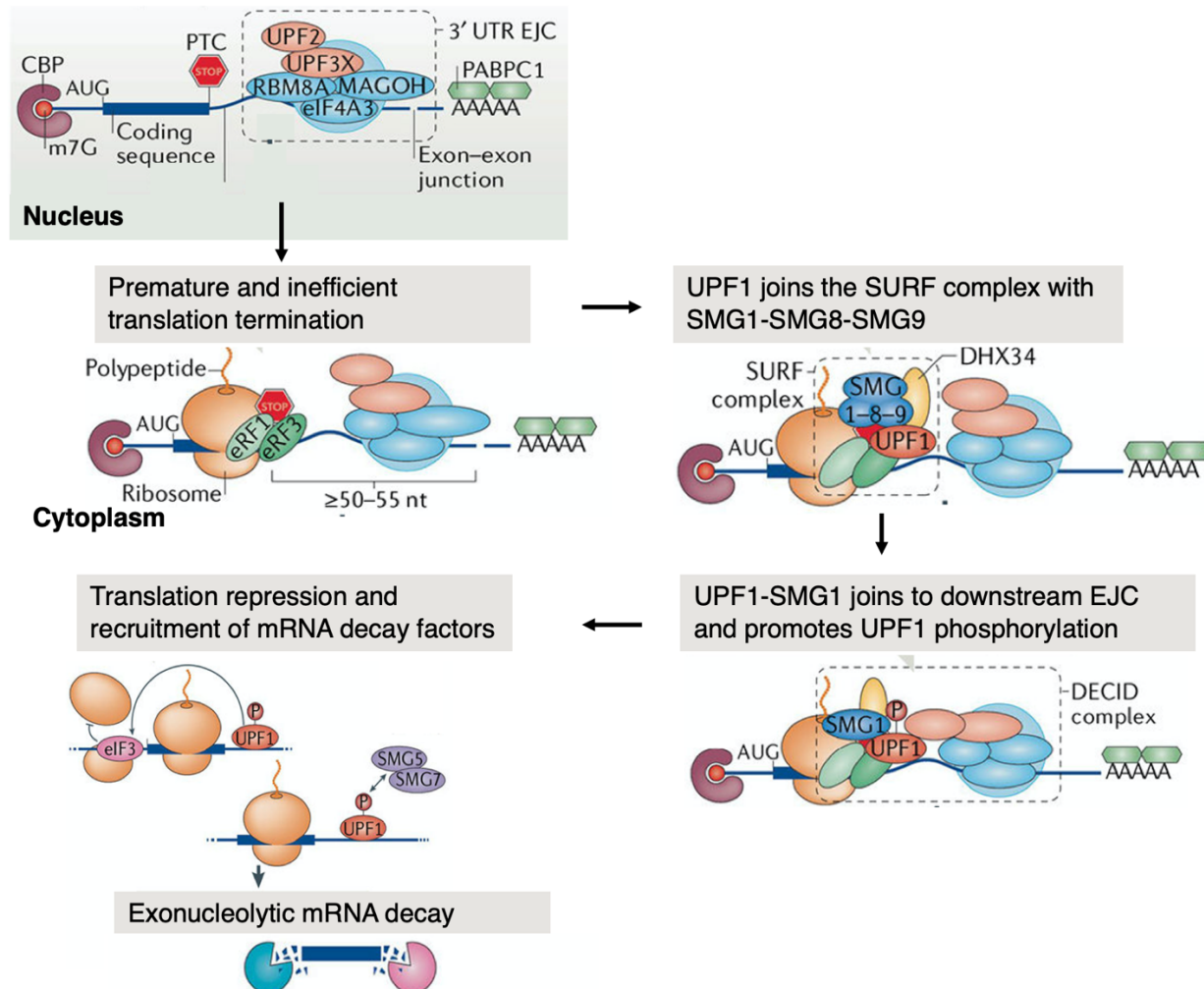


Figure 4-7: The mechanism of nonsense-mediated decay (NMD). During pre-mRNA splicing, the exon junction protein complex (EJC) is deposited ~20-24bp upstream of every exon-exon junction to identify PTCs. After splicing, mRNAs are transported to the cytoplasm for translation. If an EJC detects a PTC (red) 50-55bp upstream of the last exon-exon junction, the translation termination is considered inefficient and the process of nonsense mediated RNA decay (NMD) is initiated. Factors of the NMD-inducing complex (i.e. SURF complex) are recruited to the ribosome and phosphorylation of UPF1 by SMG1 represses translation and induces mRNA decay, which occurs either through exonucleolytic decay (through SMG5-SMG7 complex) or through endonucleolytic decay (by SMG6) (not shown). Adapted from Kurosaki *et al.*, 2019, with permission ¹⁶².

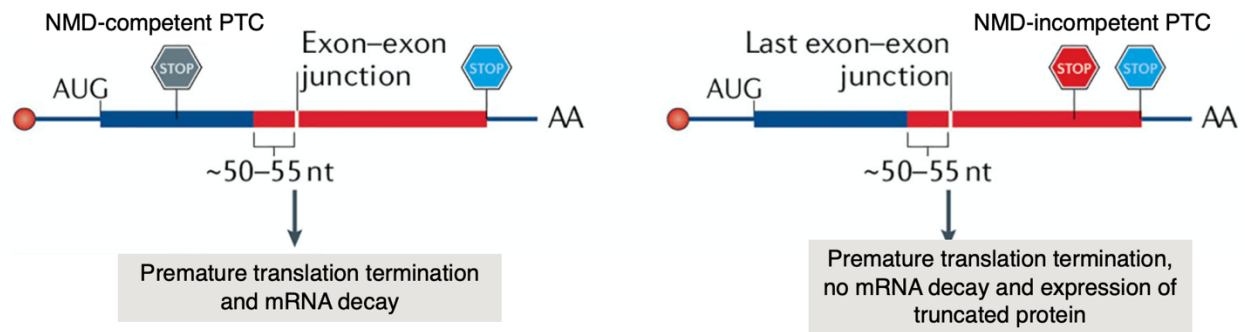


Figure 4-8: NMD-competent and incompetent PTCs. (A) mRNA transcripts harbouring PTCs residing ≥ 50 –55bp upstream of the last exon-exon junction (in the blue region) are considered NMD-competent and thus elicit mRNA decay and the protein is not produced. (B) PTCs located in the last exon (in red region) or ≤ 50 –55bp upstream of the last exon-exon junction (not shown) are considered NMD-incompetent as they fail to elicit NMD and generate truncated proteins with abnormal function. Adapted from Kurosaki *et al.*, 2019, with permission ¹⁶².

Another pathway of NMD is activated when an abnormally long 3' untranslated region (UTR) is downstream of a PTC. The mechanism of this pathway is not well understood, however, it is believed that the accumulation of UPF1 at the 3' UTR site mediates the recruitment of other NMD-inducing factors to trigger mRNA decay ³⁰⁸.

4.3 Methods

4.3.1 Burden testing of *MYH7* c.5769delG

I performed burden testing of the *MYH7* frameshift variant (c.5769delG) by calculating the excess of this variant in the AhcHCM cohort over gnomAD (n=125,064) as well as over EHVol- (n=400), NEA- (n=368) (sub-population of GME) and Iranian- (n=797) controls (Iranome: <http://www.iranome.ir/about>). Due to the relatively lower coverage of WES (gnomAD, NEA and Iranome datasets) data compared to the targeted TruSight Cardio sequencing data (EHVol dataset), not all variant sites in *MYH7* were sequenced at the same coverage level. Thus, for these control populations, I adjusted the total number of individuals sequenced for *MYH7* based on the mean coverage of rare variants in that gene. For populations where the *MYH7* variant was observed, I determined the total number of individuals sequenced for this variant. I then evaluated the significance of the case excess using Fisher's exact test (with one-sided testing).

In addition, I compared the frequency of the c.5769delG variant between AhcHCM patients and the predominantly Caucasian HCM cohorts (LMM/OMGL and RbhHCM) and evaluated statistical significance using Fisher's exact test (with two-sided testing).

4.3.2 Calculation of the population-level penetrance of *MYH7* c.5769delG

I estimated the population-level penetrance of the *MYH7* c.5769delG variant using the online calculator (<http://cardiodb.org/alleleFrequencyApp>) developed by Whiffin *et al.*, 2017 according to the equation described in Minikel *et al.*, 2016^{149,312}. The equation is shown below:

$$Penetrance = baseline\ risk \times \frac{case\ allele\ frequency}{population\ control\ allele\ frequency}$$

where:

baseline risk = prevalence of HCM in the general population

case allele frequency = c.5769delG allele frequency in AhcHCM patients

population control allele frequency= c.5769delG allele frequency in ethnically-matched controls (i.e. EHVol and NEA controls)

4.3.3 Genotyping and quality control

Genotyping of the 16/17 AhcHCM patients as well as their family members, for whom DNA was available, was performed at the Human Genomics Facility (HuGe-F), Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands using the Infinium Global Screening Array (GSA) (Illumina Inc, San Diego, CA., USA). I then assessed the quality of the genotyping data using PLINK software (version 1.9) ²²⁶

Quality Control (QC) criteria included gender checks, missingness (i.e. removal of individuals with >10% missing genotypes) and duplicates. I also performed QC on SNPs to assess for missingness (i.e. removal of SNPs with >5% missing genotyping rate), MAF (i.e. only SNPs with $MAF \geq 5\%$ were included), Hardy-Weinberg equilibrium (HWE) (i.e. removal of SNPs that failed HWE) and SNP pruning for linkage disequilibrium using the PLINK command `indep 50 5 0.2`.

4.3.4 Relatedness analysis

I assessed the relatedness of the 16/17 AhcHCM patients with the *MYH7* variant c.5769delG using PLINK, which estimates the number of alleles at each marker that originated from the same ancestral allele (i.e. identity-by-descent (IBD)) between pairs of samples. I then visualised the output table of the PLINK analysis in R which summarises the probability that 0, 1 or 2 alleles were IBD by the notations $P(Z=0)$, $P(Z=1)$, and $P(Z=2)$, respectively. Unrelated samples would have a Z_0 of 1 and duplicated/twin samples would have a Z_2 of 1. A combined relatedness measure ("PI-HAT"), which is calculated as $P(Z=2) + 0.5 \cdot P(Z=1)$, was also depicted in the output table. PI-HAT quantifies the proportion of a genome that is shared by descent with an ancestor between each pair of samples.

4.3.5 Segregation analysis

To assess the co-segregation of the *MYH7* frameshift variant c.5769delG with HCM, I estimated the Logarithm of the Odds (LOD) score for a large, genotyped AhcHCM family using the equation described by ClinGen for dominant diseases (shown below), where “n” denotes number of segregations/meioses ¹⁷⁶.

$$Z(LOD\ score) = \log_{10}(2^n)$$

I created the pedigree for this family using the Genial Pedigree online tool (<http://www.pedigreedraw.com/#jumbotron>).

4.3.6 Age-related penetrance for large AhcHCM family with MYH7fs variant

To estimate the age-related penetrance in a large AhcHCM family of a proband carrying the *MYH7* frameshift variant c.5769delG, I created Kaplan-Meier plots in R using the Prodlim (v.2019.11.13) (<https://cran.r-project.org/web/packages/prodlim/prodlim.pdf>) and Survival (v.3.1-8) (<https://cran.r-project.org/web/packages/survival/survival.pdf>) packages.

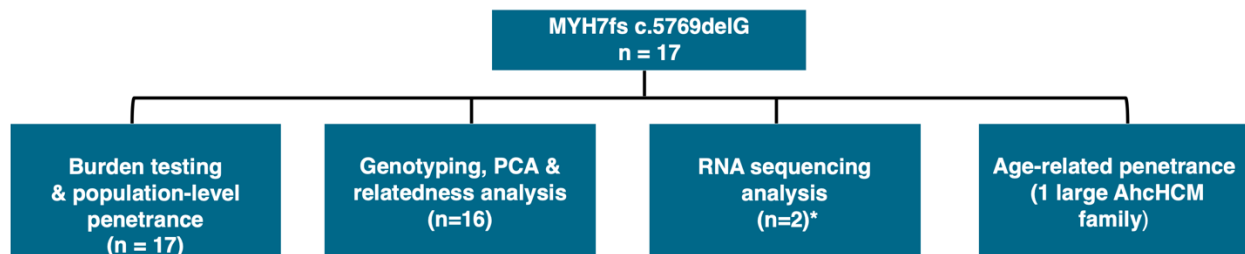


Figure 4-9: Summary of the analysis performed on AhcHCM patients with MYH7fs variant c.5769delG. *From 2 patients RNA-seq was performed on myocardial tissue and iPSC-derived CMs.

4.3.7 Comparison of disease severity between *MYH7* frameshift (*MYH7fs*) variant-carriers and other AhcHCM patients with different genetic aetiologies

I compared the age at disease onset and the left ventricular maximal wall thickness (LVMWT) (mm) between AhcHCM patients carrying the *MYH7fs* c.5769delG variant and other AhcHCM patients with different genetic aetiologies: genotype-positive patients (i.e. defined as harbouring P/LP variants in HCM-associated genes - outlined in section Table 3-3) and AhcHCM patients with missense variants in *MYH7*. *MYH7fs* variant-carriers with double hits were excluded from this analysis to be able to accurately assess the effect of the *MYH7fs* variant on both the age at disease onset and LVMWT. Box plots displaying these comparisons were created using R. I evaluated the statistical significance between AhcHCM *MYH7fs*-carriers and the other 2 patient groups by student t-test.

4.3.8 RNA extraction and sequencing and analysis

RNA was extracted both from myocardial tissue and from induced pluripotent stem cell-derived cardiomyocytes (iPSC-derived CMs) from 2 AhcHCM patients carrying the *MYH7* frameshift variant c.5769delG. RNA was also extracted from a healthy myocardial tissue and CMs derived from a commercial iPS cell line to serve as controls. Wet lab work was performed by my colleagues from the AHC team (refer to page 17 for more details on contributions). RNA was extracted using the TruSeq Stranded Total RNA kit (Illumina).

I assessed the RNA-sequencing (RNA-seq) read quality using FastQC and low-quality reads were trimmed using Trimmomatic^{313,314}. Reads were then mapped to the reference genome hg19 using Spliced Transcripts Alignment to a Reference (STAR)³¹⁵. SAMtools were then used to convert the aligned reads to Binary Alignment Map (BAM) format and duplicated reads were removed using the Picard tools package (<https://broadinstitute.github.io/picard/>)³¹⁶. SNV calling was performed using GATK toolkit and splice junction reads were split followed by base quality score recalibration. Variants were detected using GATK HaplotypeCaller¹⁴⁷. Finally, variants were filtered using the GATK Variant Filtration tool for low quality calls (QD < 2), strong strand bias

(FS >30) SNP clusters (3 NPSs in 35bp window). Mapped reads were visualised using Integrative Genomics Viewer (IGV) ³¹⁷.

4.3.9 Allele-specific expression analysis

To assess whether both WT and mutant *MYH7* alleles were equally expressed in each of the 2 AhcHCM patients, for whom RNA-seq was performed, I used bam-readcount from SAMtools to determine the read counts for each allele ⁹⁸. Then, I calculated the allelic ratio (AR) with regards to the mutant *MYH7* allele (i.e. alternate allele (ALT)) using the equation below:

$$AR = \frac{ALT\ reads\ (i.e.\ with\ MYH7fs)}{Total\ reads}$$

An AR of 0.5 would indicate equal expression of both WT and ALT alleles. A deviation from 0.5 would indicate an allele-specific expression (ASE). To test the significance (p-value < 0.05) of this deviation, I performed a binomial test using R.

4.3.10 Distribution of rare truncating variants in *MYH7*

I used the TrackViewer package (v 1.22.1) in R to examine the distribution of rare truncating variants in *MYH7* in the Egyptian (AhcHCM) and predominantly Caucasian (LMM/OMGL and RbhHCM) cohorts as well as controls (EHVol, CHVol and gnomAD).

4.4 Results

4.4.1 Enrichment of a single *MYH7* frameshift variant (c.5769delG) in the AhcHCM cohort

As previously demonstrated in the chapter 3, a significant ($p = 0.0004$) enrichment of rare truncating variants in *MYH7* (MYH7tv) was observed in AhcHCM patients (3.31%) compared to ethnically-matched EHVol controls (0.25%). This is intriguing as MYH7tvs are believed to be non-causative in HCM¹⁰⁰. Detailed analysis of MYH7tvs observed in the AhcHCM cohort showed that a single frameshift variant (c.5769delG) accounted for all truncations identified in this gene. Thus, this variant was studied in more detail.

4.4.2 *MYH7* frameshift variant c.5769delG was classified as a VUS according to the ACMG/AMP guidelines

According to the automatically activated ACMG/AMP rules implemented in CardioClassifier, the *MYH7* frameshift variant (MYH7fs) c.5769delG only activated the PM2 rule (i.e. low frequency in population databases), as the variant was absent in gnomAD. The variant did not fulfil the criteria for a benign or pathogenic variant so its default classification was VUS¹⁰². The variant has also not been reported in ClinVar or the Human Gene Variant Database (HGMD) and was absent from both the LMM/OMGL ($n > 6,000$) and RbhHCM cohorts ($n = 684$).

It is also worth mentioning that CardioClassifier does not apply the PVS1 rule for rare *MYH7* truncations as they are not considered to be disease-causing.

4.4.3 Demographic and clinical characteristics of AhcHCM patients with *MYH7* c.5769delG variant

PCA for common SNPs between genotyped AhcHCM patients carrying the MYH7fs variant c.5769delG and the different HapMap populations showed that the AhcHCM patient cluster (green) was clearly distinguishable from the HapMap populations (Figure

4-10). The geographical distribution of MYH7fs-carriers was defined on the basis of the city of origin and showed a comparable pattern across 9 different cities across Egypt (Figure 4-11).

I also performed Identity-by-descent (IBD) analysis between each pair of genotyped samples (n=16) to examine the relatedness of MYH7fs-carriers. This analysis confirmed that MYH7fs-carriers were unrelated with the highest PI-HAT value of 0.0466 observed between 1 pair of samples (for reference, 4th degree and 5th degree relatives would have a PI-HAT of 0.0625 and 0.03125, respectively).

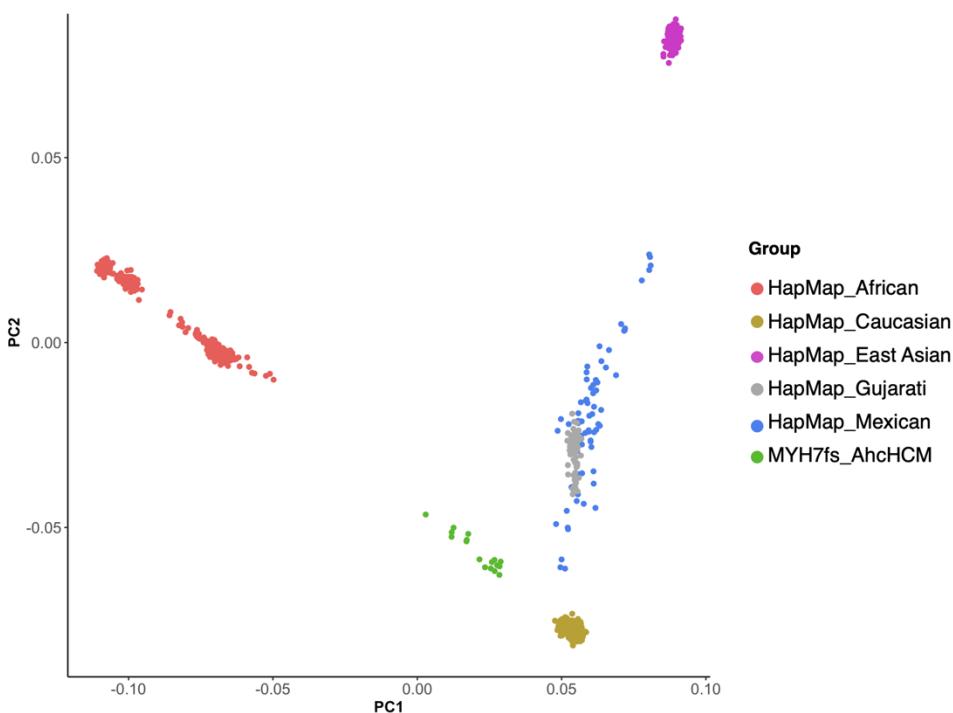


Figure 4-10: PCA plot for common SNPs between genotyped AhcHCM patients with *MYH7* frameshift variant (MYH7fs: c.5769delG) and HapMap populations.

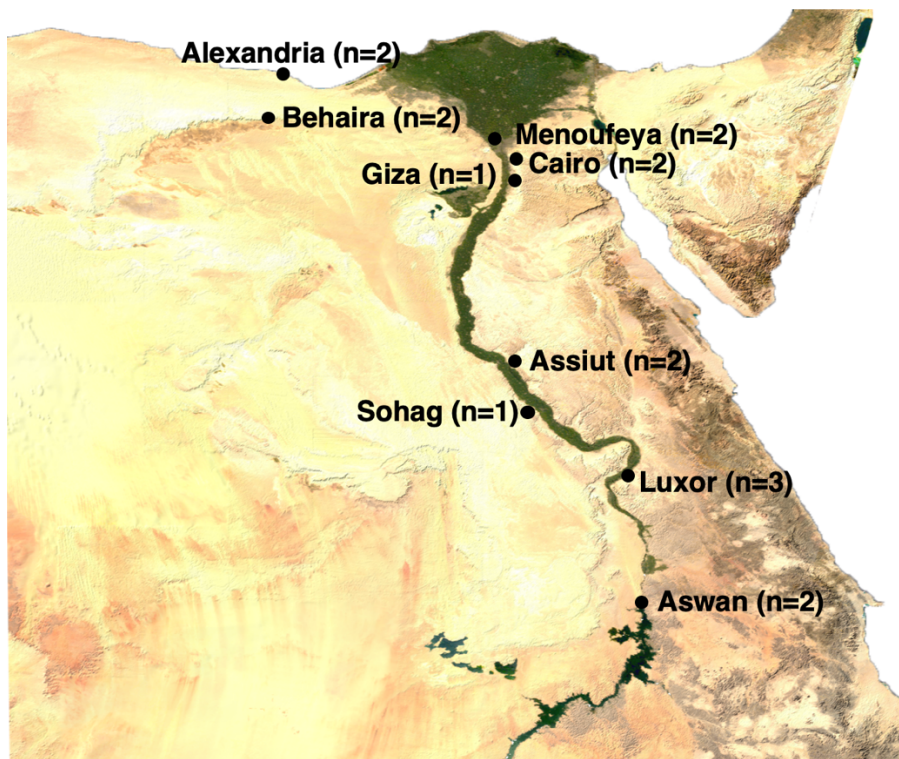


Figure 4-11: Distribution of AhcHCM MYH7fs-carriers by city of origin. Map from: <https://www.cleanpng.com/png-ancient-egypt-eastern-desert-libyan-desert-arabian-1067013/preview.html>

Concerning the characteristics of MYH7fs-carriers, there were 9 male (52.9%) and 8 (47.1%) female patients. The mean age at HCM diagnosis was 42.2 years $\pm$ 16.1 SD. The mean LVMWT (mm) was 25.6 $\pm$ 6 and 23.5% of the MYH7fs-carriers underwent surgical myectomy. A detailed summary of the demographic- and clinical characteristics is outlined in Table 4-1. Cardiac dimensions and function were based on echocardiographic measurements.

Table 4-1: Demographic and clinical characteristics of AhcHCM patients with MYH7fs (n=17)

General characteristics	AhcHCM n (%)	Available data (n)
Demographics		
Age at diagnosis (years) (mean (sd))	42.2± 16.1	17
Male gender n(%)	9 (52.9%)	17
Offspring of Consanguineous Marriage* (self-reported)	5 (35.7%)	14
FH of HCM	3 (17.6%)	17
FH of SCD	1 (5.9%)	17
Cardiac characteristics		
LVMWT (mm) (mean (sd))	25.6±6.0	17
Extreme LVMWT (>30mm)	4 (23.5%)	17
LVOTO	10 (58.8%)	17
Myectomy	4 (23.5%)	17

*Consanguinity was defined as first-cousin marriage.

4.4.4 Enrichment of *MYH7* c.5769delG variant in AhcHCM patients compared to ethnically-matched controls

To assess the association of the *MYH7* frameshift variant (MYH7fs) c.5769delG with HCM, I compared its burden in AhcHCM patients with ethnically-matched controls. As demonstrated in Table 4-2, the frequency of the c.5769delG variant was significantly higher in AhcHCM patients over EHVol controls (p-value=5x10⁻⁵), as well as the Middle Eastern reference datasets NEA (p-value=8x10⁻⁴) and Iranome (p-value= 1x10⁻⁷). The variant was also absent from the ‘al mena’ database, containing >26 million genetic variants from Arab and MENA populations, including Qatar (n=1,213), Iran (n=77) and GME (n=1,002) ²²⁸.

Table 4-2: Significance of the enrichment of *MYH7* c.5769delG in the AhcHCM cohort over control populations (gnomAD, EHVol, NEA (sub-population of GME) and Iranome)

	AhcHCM	gnomAD	EHVol	NEA	Iranome
Freq.	3.31%	0.00%	0.00%	0.27%	0.00%
(variants/total)	(17/514)	(0/125064)	(0/400)	(1/368)	(0/797)
p-value	-	2x10 ⁻⁴¹	5x10 ⁻⁵	8x10 ⁻⁴	1x10 ⁻⁷

In addition, I estimated the population penetrance of the c.5769delG variant and calculated the OR and EF using both EHVol and NEA as the reference control population (Table 4-3). The population penetrance of the c.5769delG variant was estimated to be 5%. This suggests that an Egyptian individual carrying the c.5769delG variant has a 5% chance of developing HCM. It is important to note that the 95% CI calculated for the penetrance was wide (1%- 48%) due to the limited cohort size. Expanding the AhcHCM and EHVol cohort sizes will yield a more accurate estimate of the population-level penetrance.

In addition, the EF, which provides an estimate of the probability that the c.5769delG variant, when identified in an AhcHCM patient, is causative was 0.96. This suggests that, in the absence of functional or segregation data, there is a high probability of the c.5769delG variant being pathogenic when detected in HCM patients.

Table 4-3: Estimated population penetrance of *MYH7* c.5769delG using EHVol and NEA as reference controls (total cohort size: 767) together with OR and EF

AhcHCM freq. (variants/total)	EHVol + NEA freq. (variants/total)	Est. Population Penetrance	OR (95% CI)	EF (95% CI)
3.31% (17/514)	0.13% (1/767)	0.05 (0.01 – 0.48)	26.19 (4.08 – 1092.15)	0.96 (0.76– 1.0)

Est: estimated

4.4.5 Segregation analysis of a large AhcHCM family provided supporting evidence for the association of the MYH7fs variant c.5769delG with HCM

To further evaluate the association of MYH7fs variant c.5769delG with HCM, I performed segregation analysis of a large HCM family (the family of patient A) (Figure 4-12). A total of 16 family members were clinically screened at AHC and I performed targeted sequencing, using the TruSight Cardio Sequencing Kit, on all 16 DNA samples. Of these family members 6 were diagnosed with HCM (4 males and 2 females) and all of them carried the MYH7fs. The affected family members did not carry other rare protein-altering variants in HCM-associated genes ¹⁴⁸. The clinical characteristics of affected family members are summarised in (Table 4-4). In addition, 5 family members with an average

age of 36 ± 21.3 carried the MYH7fs variant but did not show clinical features of HCM at the time of screening, which reflects the variable penetrance of this variant. To assess the co-segregation of the MYH7fs variant with HCM, I calculated the LOD score as per the equation described in section 4.3.5¹⁷⁶. This score is a ratio of the likelihood that the disease and genotyped marker loci are linked on the chromosome, and thus are inherited together, rather than unlinked³¹⁸.

The LOD score for this family was estimated to be 3.01, which corresponds to 10 meioses/segregations. According to the recommendations for *MYH7* variant classification developed by the ClinGen Inherited Cardiomyopathy Expert Panel, ≥ 7 meioses/segregations provide strong evidence of variant co-segregation with disease²⁰². Thus, segregation analysis of this family provided strong evidence of c.5769delG association with HCM.

Table 4-4: Clinical characteristics of the proband (i.e. Patient A) as well as affected family members carrying the MYH7fs variant

Patient ID	Age at disease onset	LVMWT (mm)	LVOTO presence	Clinical history
V. 1	29	35	yes	Dyspnoea, NYHA class III, palpitations, myectomy
IV. 1	58	17	no	Dyspnoea, NYHA class II, palpitations
IV. 2	61	23	no	Dyspnoea, NYHA class II, presyncope
IV. 3	55	32	no	Dyspnoea, NYHA class III, Angina
IV. 4	40	16	yes	Dyspnoea, NYHA class II, PND, myectomy
IV. 5	33	36	yes	Dyspnoea, NYHA class II, Angina, palpitations, myectomy
IV. 6	18	32	yes	Dyspnoea, NYHA class III, Angina, palpitations, myectomy

NYHA class: New York Heart Association Classification; PND: Paroxysmal nocturnal dyspnoea

Then, I estimated the age-related penetrance of the MYH7fs using the ages of AhcHCM-variant carriers at disease onset or last clinical assessment (Figure 4-13). The analysis showed that the age-related penetrance was 8% by age 20 (95% CI:0%-23%), 17% by age 30 (95% CI:0%-35%), 37% by the age of 40-50 (95% CI:0%-61%), 70% by the age of 60 (95% CI:7%-90%) and 84% by the age of 80 (95% CI: 9%-97%).

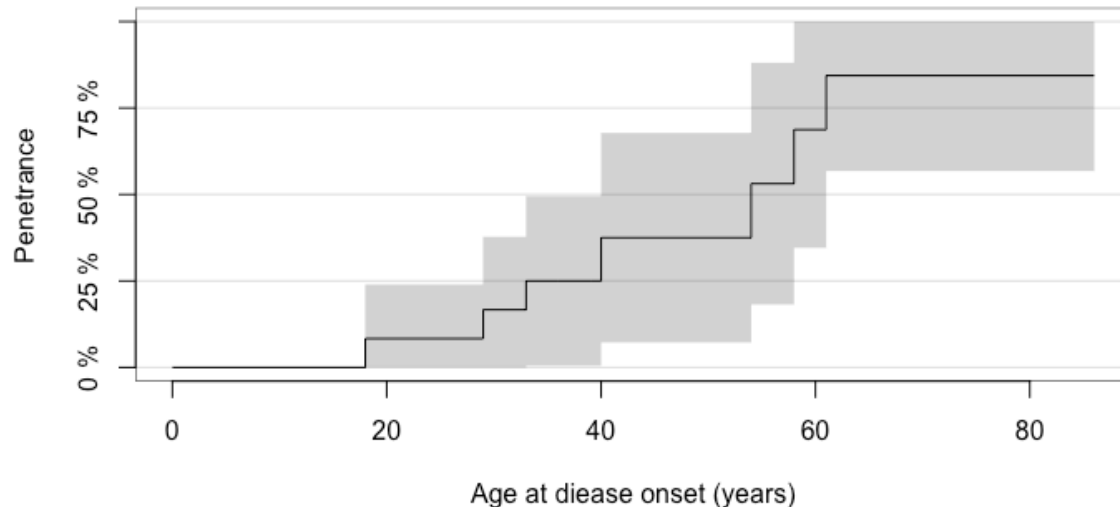


Figure 4-13: Estimated age-related penetrance for MYH7fs-variant carriers diagnosed with HCM. Penetrance analysis was performed on the large AhcHCM family (presented in Figure 4-12). The shaded area (grey) represents the 95% confidence intervals (CI).

4.4.6 The MYH7fs variant (c.5769delG) is associated with a late HCM disease-onset compared to other AhcHCM patients with different genetic aetiologies

I then sought to evaluate whether the MYH7fs variant is associated with a more severe HCM phenotype. Thus, I compared the age at disease onset and LVMWT between AhcHCM patients carrying the MYH7fs c.5769delG variant with genotype-positive patients (i.e. defined as harbouring P/LP variants in HCM-associated genes) as well as with patients harbouring missense variants in *MYH7*. MYH7fs variant-carriers with double hits were excluded from this analysis to be able to accurately assess the effect of the MYH7fs variant on disease severity.

MYH7fs variant (c.5769delG)-carriers were significantly older (43.7 ± 12.1) compared to both genotype-positive (30.8 ± 14.4 , $p\text{-value}=0.004$) and *MYH7*-missense variant carriers (30.8 ± 16.1 , $p\text{-value}=0.005$). No differences in LVMWT were observed between both genotype-positive patients vs. MYH7fs variant-carriers ($23.9\text{mm} \pm 6.9$ vs. $24.5\text{mm} \pm 5.2$, $p\text{-value}=0.750$) and *MYH7*-missense- vs. MYH7fs variant-carriers (23.3 ± 7.3 vs. $24.5\text{mm} \pm 5.2$, $p\text{-value}=0.675$). However, the median LVMWT trended slightly lower in

MYH7fs variant-carriers compared to the other 2 groups (22.5mm vs. 24.5mm (Geno_pos) vs. 23.5mm (MYH7missense)).

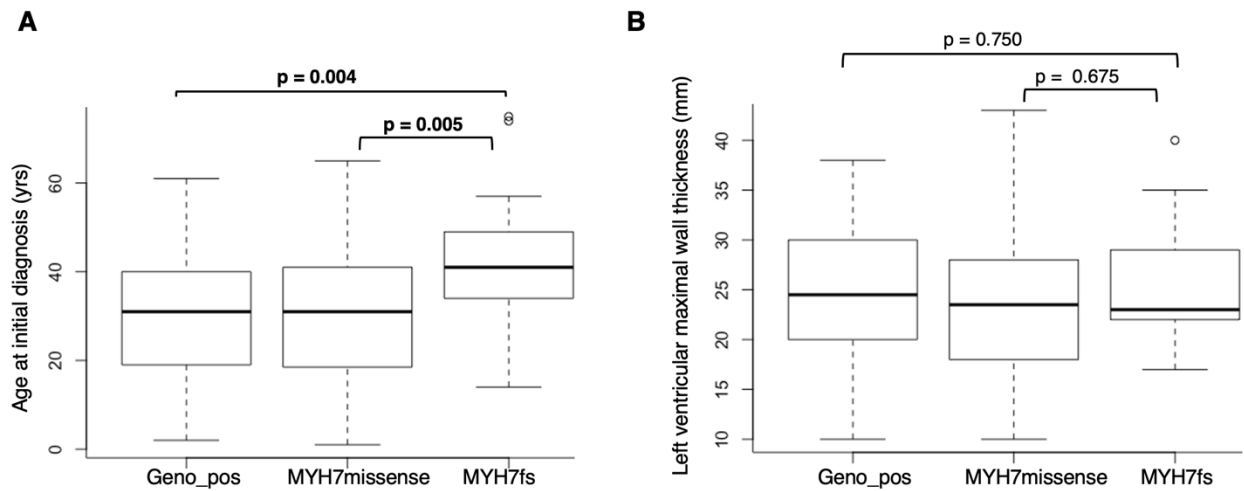


Figure 4-14: Patients carrying the *MYH7* frameshift (MYH7fs) variant (c.5769delG) presented with HCM at a significantly older age compared to AhcHCM patients with different genetic aetiologies. Box plots comparing (A) the age at disease onset and (B) the left ventricular maximal wall thickness (LVMWT) (mm) between AhcHCM patients carrying the *MYH7* frameshift (MYH7fs) variant (c.5769delG), patients carrying missense variants in *MYH7* (MYH7missense) and other genotype-positive AhcHCM patients.

4.4.7 Genetic variability among MYH7fs c.5769delG variant-carriers

It is important to note that a subset (n=5) of MYH7fs variant-carriers harboured an additional rare variant in a sarcomere gene (Table 4-5).

Table 4-5: Clinical characteristics of MYH7fs c.5769delG variant carriers harbouring a second hit in sarcomere genes

Patient #	Gender	Age*	Second hit**	LVMWT (mm)	LVOTO (≥30mmHg)	Myectomy
1	Female	19yrs	<i>MYBPC3</i> c.3627+2delT (LP)	24	yes	yes
2	Female	14yrs	<i>MYBPC3</i> c.1223+1G>A (LP)	28	yes	yes
3	Male	36yrs	<i>MYH7</i> p.Asp309Asn (LP)	20	no	no
4	Female	50yrs	<i>MYBPC3</i> p.Asp824Asn (VUS)	40	no	no
5	Male	75yrs	Homozygous for MYH7fs	29	no	no

*)Age represents age at HCM diagnosis. **)Variants were classified according to the ACMG/AMP guidelines in addition to manual curation for evidence on segregation or de novo inheritance reported in ClinVar or PubMed. Patients, for which family data was available (n=2) are highlighted in bold.

Family data was available for only 2/5 AhcHCM patients (patients 1 and 2) with double hits (Figure 4-15). Both were female, carried a LP variant in *MYBPC3* as the second-hit and were the youngest among the MYH7fs-carriers identified in the AhcHCM cohort.

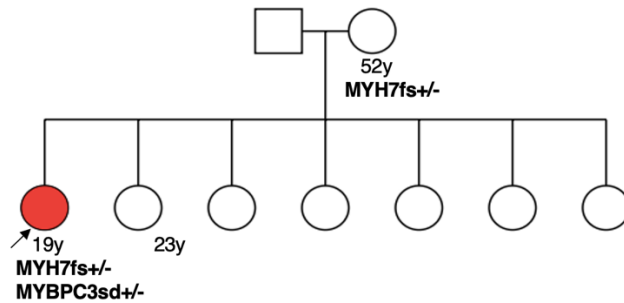
Patient 1, carrying the MYH7fs and *MYBPC3* splice donor variant c.3627+2delT, was diagnosed with HCM at the age of 19 years (Figure 4-15A). This patient presented with a LVMWT of 24mm and underwent surgical myectomy to relieve the LVOTO when she was 20 years old. Her mother, who was clinically screened at AHC at the age of 52 years, carried the MYH7fs variant but presented with a healthy phenotype. These findings suggest that the MYH7fs variant could be associated with variable penetrance (healthy mother) but may exacerbate the disease phenotype when expressed with another LP/P variant. The *MYBPC3* c.3627+2delT variant was previously identified in 0.13% of LMM/OMGL patients (8/6179).

Patient 2, carrying the MYH7fs and *MYBPC3* splice donor variant c.1223+1G>A, was diagnosed with HCM at the age of 14 years (Figure 4-15B). The patient presented with a LVMWT of 28mm and also underwent surgical myectomy (at 29 years). Intriguingly, the patient's mother, who was clinically screened at AHC at the age of 48 years, carried the MYBPC3tv but presented with a healthy phenotype. The MYBPC3tv variant was absent in predominantly Caucasian HCM patients but has been reported in ClinVar (Accession no. SCV000208275.9) as a LP/P variant. Thus, carrying the MYBPC3tv *in trans* with the MYH7fs could explain the early disease onset in patient 2. The patient's 4 daughters were also clinically and genetically screened and 2 daughters, aged 2yrs and 9yrs, carried the MYH7fs and MYBPC3tv variants, respectively. All of them were reported to be healthy. However, given the late-onset of HCM (starting from adolescence to late adulthood), it was expected not to identify LVH in the children.

Family data was not available for the remaining 3 AhcHCM patients with double hits. Intriguingly, 1 of these patients carried the MYH7fs variant in the homozygous state. Despite having 2 mutant *MYH7* alleles, the homozygous patient was diagnosed with HCM

at the age of 75. He presented with an elevated LVMWT (29mm) so it could be that he had substantial hypertrophy prior to diagnosis but remained asymptomatic.

A



B

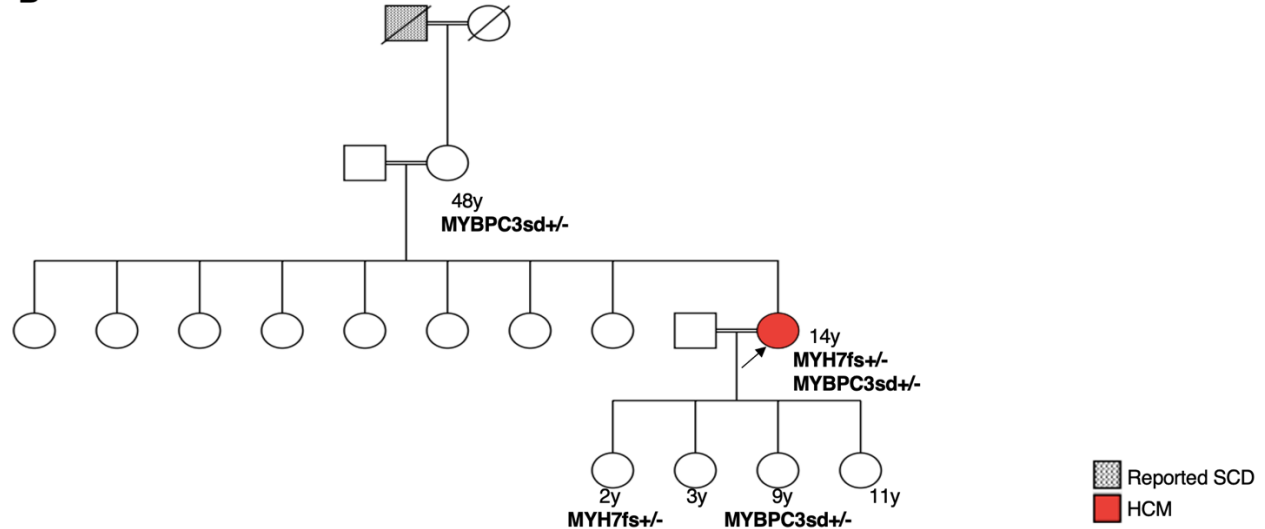


Figure 4-15: Carrying the MYH7fs with another pathogenic variant in MYBPC3 is associated with early HCM onset. Family pedigrees of 2 AhcHCM patients carrying the MYH7fs variant c.5769delG and MYBPC3 LP truncating variants c.3627+2delT (**A**) and c.1223+1G>A (**B**) are shown. fs: frameshift; sd: splice donor.

4.4.8 RNA sequencing analysis confirmed the expression of the variant MYH7 transcript

To gain initial insights into the disease mechanism of the frameshift variant c.5769delG, I studied the predicted effect of the variant on the protein structure in detail. Figure 4-16 depicts the cDNA sequence (for exons 39-40) and the predicted resulting amino acid (AA) sequence for the wildtype (WT) and variant MYH7 alleles. In the variant MYH7 transcript,

nucleotide G in exon 39 was deleted, which resulted into a frameshift leading to the generation of a PTC and thus the predicted protein product is only 4 AA shorter than the WT. Transcripts harbouring PTCs, which could introduce radical changes in the resulting protein sequence, are believed to be degraded by NMD prior to translation. Thus, these PTCs could contribute to HCM pathogenesis through haploinsufficiency (as reported for *MYBPC3*⁴⁷). However, haploinsufficiency of *MYH7* has not been recognised as a disease mechanism for HCM¹⁰⁰.

Given the prevalence of MYH7fs variant among Egyptian HCM patients, I hypothesised that it could have led to the generation of an NMD-incompetent PTC and thus escaped mRNA decay to produce a truncated, disease-causing protein. As explained in the introduction of this chapter, PTCs in the last exon of a gene are believed to escape NMD¹⁶². Indeed, the PTC of the variant *MYH7* transcript is predicted to reside in the last exon of the gene (Figure 4-16B).

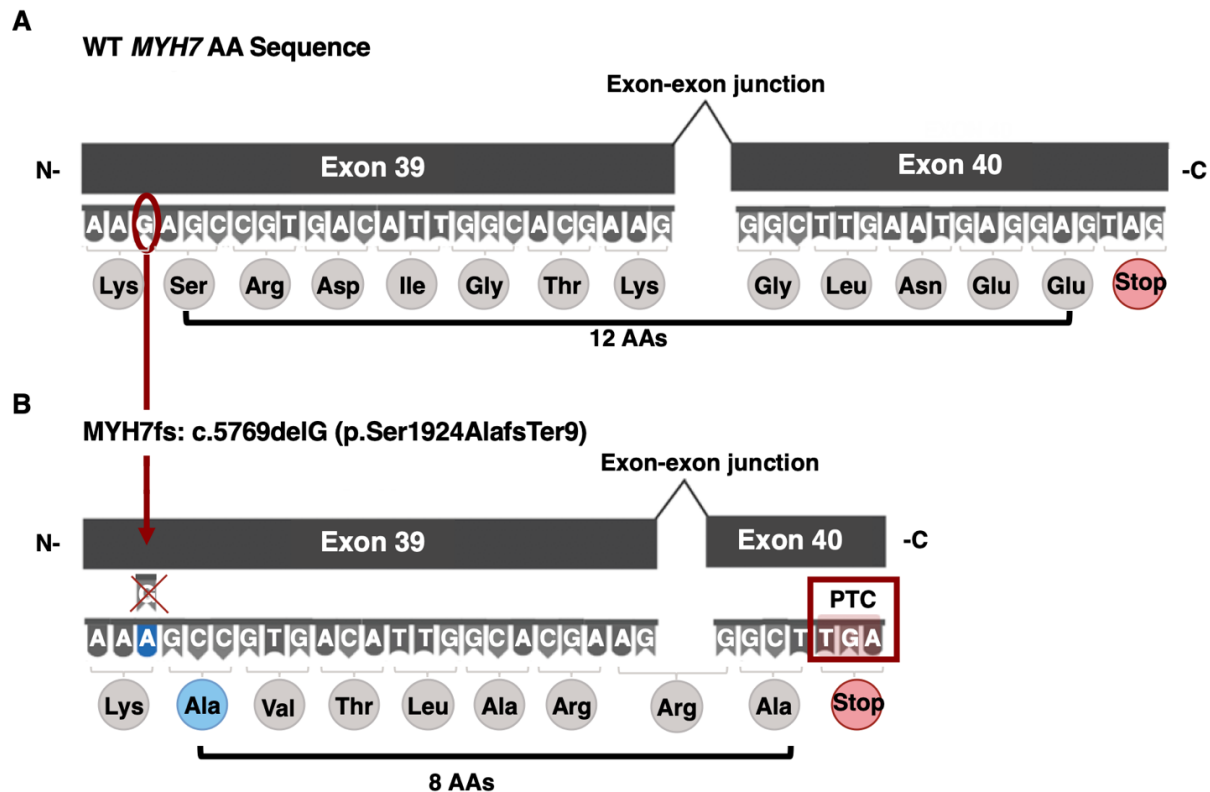


Figure 4-16: The *MYH7*fs c.5769delG variant is predicted to escape NMD and produce a truncated protein. The cDNA sequence and predicted resulting amino acid sequence are shown for wild type (WT) (**A**) and variant *MYH7* (**B**) alleles are shown. The frameshift variant, introduced by the deletion of nucleotide C in the penultimate exon (exon 39), is predicted to result in an NMD-incompetent PTC (located in last exon and thus downstream of last exon-exon junction). Created with BioRender.

To test the hypothesis that the variant *MYH7* transcript harbouring the frameshift variant escaped NMD, RNA sequencing (RNA-seq) of myocardial tissues, which were obtained from 2 AhcHCM patients during surgical myectomy, was performed. RNA-seq was also performed on induced pluripotent stem cell-derived cardiomyocytes (iPSC-derived CMs), which were generated from the same 2 patients. The RNA-seq data was made available for me for downstream analysis.

As illustrated in Figure 4-17, RNA-seq analysis confirmed the expression of the variant *MYH7* transcript caused by the frameshift variant (i.e. deletion of nucleotide C (reverse strand)) in both patients. As expected, the control myocardial tissue did not harbour the

frameshift variant. This confirms the hypothesis that the variant *MYH7* transcript escaped NMD.

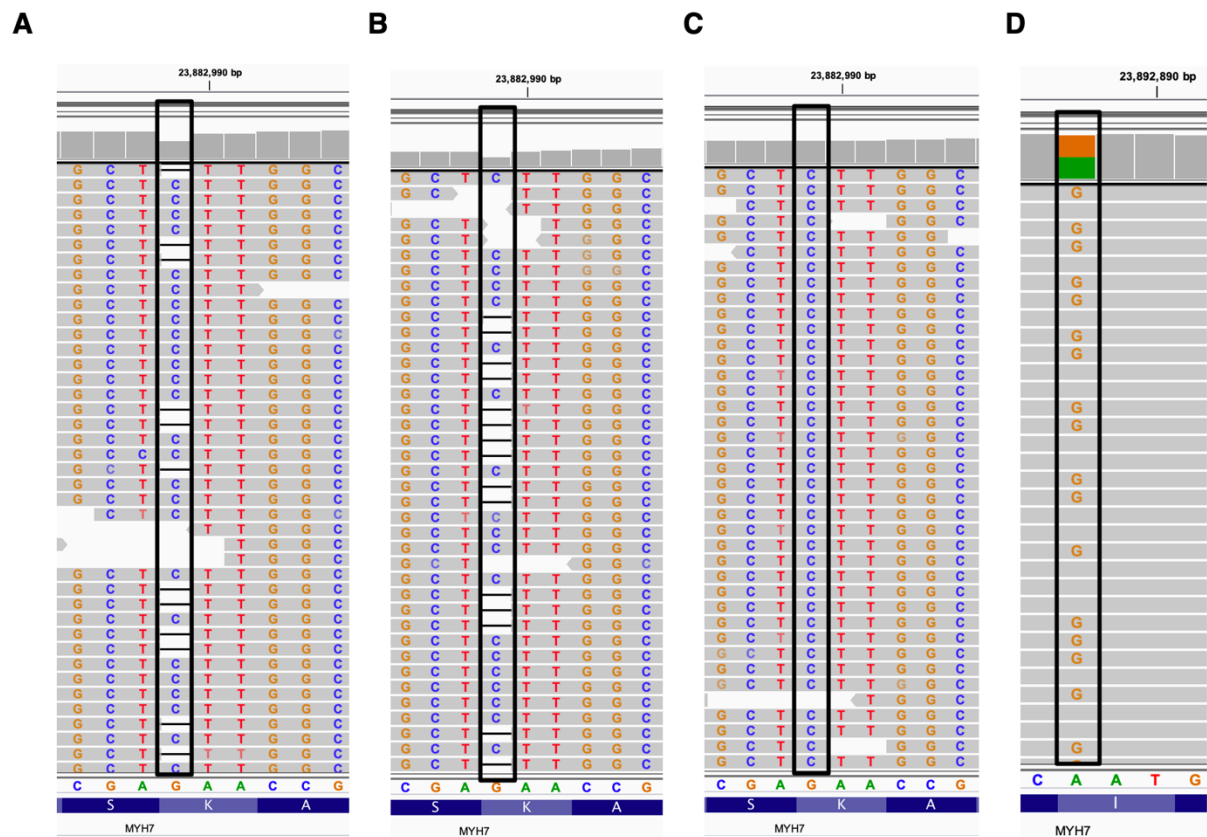


Figure 4-17: RNA-seq analysis of human myocardial tissue from 2 AhcHCM *MYH7*fs variant-carriers confirms the expression of the variant *MYH7* transcript. Mapped RNA-seq reads (reverse strand) of *MYH7* were visualised with Integrative Genomics Viewer (IGV) for 2 different AhcHCM patients with *MYH7*fs c.5769delG (A and B) and 1 healthy control (C). The black box shows the variant site (with black line representing the deleted nucleotide C) in the variant *MYH7* transcript in patients. (D) Equal expression of maternal and paternal transcripts at heterozygous SNP (c.2967T>C) in the healthy control tissue.

To assess the relative expression of the mutant *MYH7* allele to the WT allele in each of the 2 patients, I performed allele-specific expression (ASE) analysis (Table 4-6). The analysis showed that the relative expression of the mutant *MYH7* allele was significantly lower compared to the WT allele (AR ~0.3:0.7 and 0.4:0.6 in myocardial tissue and iPSC-derived-CMs, respectively). In contrast, the AR for the heterozygous SNP (c.2967T>C) in the control tissue was ~0.51 (binomial p-value= 0.9), indicating an equal expression of both maternal and paternal alleles (Figure 4-17D).

As previously described in Rivas *et al.*, 2015, under the strong ASE model, where 1 allele is expressed very little or not expressed at all, the alternate allelic ratio is <0.1 ³¹⁹. This means that the ASE toward the WT *MYH7* allele was not extreme (i.e. AR >0.1) and that the *MYH7* transcript harbouring the frameshift variant was not strongly degraded by NMD.

Table 4-6: Allele-specific expression (ASE) analysis of the mutant *MYH7* allele versus WT based on RNA-seq data of myocardial tissue and iPSC-derived CMs from 2 patients

Myocardial tissue			
	Total no. of reads	Allelic ratio (AR) (ALT allele reads /total reads)	Binomial p-value
Patient A	4989	0.29 (1464/4989)	<0.0001
Patient B	2075	0.28 (583/2075)	<0.0001
iPSC-derived CMs			
Patient A	2615	0.41 (1067/2615)	<0.0001
Patient B	755	0.40 (300/755)	<0.0001

The ARs for patients A and B correspond to the RNA-seq data illustrated in Figures 4-8 A and B, respectively. Significant binomial p-values (< 0.05) are highlighted in bold, which indicate a significant deviation from the expected 0.5 AR (i.e. ASE).

4.4.9 Detailed analysis of MYH7fs variant in the Egyptian HCM cohort provided evidence for its re-classification from VUS to P

Detailed analysis of the MYH7fs variant (c.5769delG) in the Egyptian HCM cohort provided evidence for its reclassification of its original VUS classification according to the ACMG/AMP guidelines to a pathogenic variant (P). In this section, I summarise the lines of evidence from the analysis to support c.5769delG reclassification.

According to the ACMG/AMP guidelines, the moderate PM2 rule could be activated if a variant is absent in large ethnically-matched controls¹⁰². While c.5769delG was absent in EHVol controls, it was observed in an individual from the ethnically-matched North East African (NEA) sub-population of the Great Middle Eastern (GME) study. However, the CMP-EP re-defined this rule for *MYH7* variants identified in cardiomyopathies to apply to

variants with an extremely rare frequency (i.e. $FAF \leq 4 \times 10^{-5}$) in large population databases ²⁰². Even though the combined cohort size of the EHVol and NEA populations is small (n=768), the PM2 rule could still be activated as the variant was also absent in the Iranome dataset (n=77), Qatari individuals (n=1,213) and other sub-populations from the GME study, who share a similar ethnic background to Egyptians as opposed to gnomAD populations (n=634) ²²⁸. Re-analysis of the burden of the c.5769delG variant in a larger-scale EHVol cohort could provide definitive evidence for its rarity in controls.

Also, according to the adapted ACMG/AMP guidelines by the CMP-EP, a *MYH7* variant that is significantly enriched in patients over controls or is identified in ≥ 15 patients with the same phenotypes, could activate the strong PS4 rule ²⁰². Thus, the c.5769delG variant activated the PS4 rule as it was significantly more prevalent in AhcHCM (patients compared to EHVol controls (17/514 (3.31%) vs. (0/400) 0%, p-value= 5×10^{-5}).

Analysis of a large AhcHCM family provided strong evidence (i.e. 7 meioses/LOD 2.1) of variant co-segregation (PP1_strong rule) with HCM ²⁰². In this family a total of 10 meioses were observed, which correspond to a LOD score of 3.01.

In addition, as mentioned in the introduction of this chapter, the CMP-EP has removed the very strong PVS1 rule from the parent ACMG/AMP framework as LoF variation is not considered a disease mechanism for *MYH7* ²⁰². This explains why LoF variants in *MYH7* were not significantly enriched in the large-scale LMM/OMGL HCM cohort compared to ExAC controls ¹⁰⁰. As mentioned in section 4.4.8, LoF variants are believed to give rise to mutant transcripts harbouring PTCs, which are eliminated from the myocardium via NMD ⁴⁷. In addition, haploinsufficiency of *MYH7* is not considered a disease mechanism in HCM ¹⁰⁰. Analysis of the distribution of *MYH7* LoF variants identified in the LMM/OMGL and RbhHCM cohorts showed that they are predicted to produce PTCs in NMD-competent sites (i.e. ≥ 50 -55bp upstream of the last exon-exon junction) and thus elicit mRNA decay (Figure 4-18). In contrast, analysis of the c.5769delG variant showed that it is not a LoF variant as it introduced an NMD-incompetent PTC in the last exon of the

gene, which caused the variant mRNA transcript to escape NMD. RNA sequencing analysis confirmed the expression of the variant transcript in 2 different AhcHCM patients carrying c.5769delG. Thus, it is plausible to assume that the c.5769delG variant caused HCM by producing a mutant transcript that escaped NMD to produce a truncated protein with disease-causing functions. Since the variant is predicted to change the protein length of *MYH7* (i.e. 4 AAs shorter than WT) due to the deletion of nucleotide G in exon 39, the moderate PM4 rule could be activated.

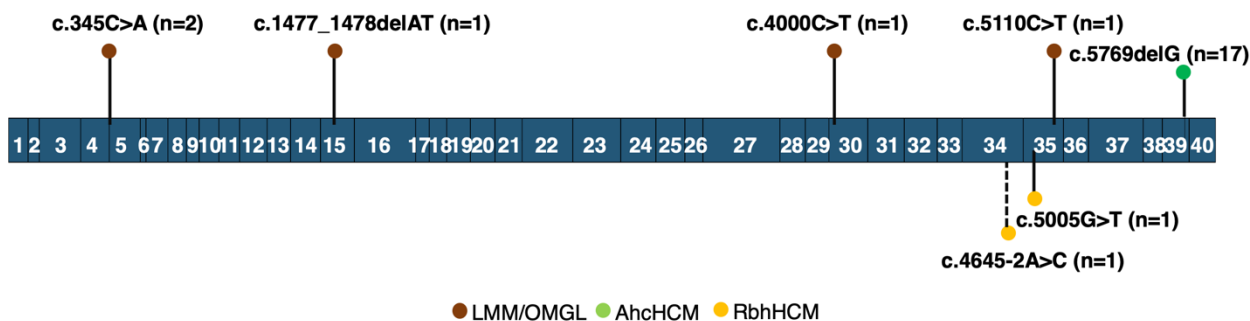


Figure 4-18: In contrast to *MYH7* truncations identified in predominantly Caucasian HCM patients, the AhcHCM variant c.5769delG is predicted to elicit NMD. The distribution of *MYH7*ts in Egyptian and predominantly Caucasian HCM cohorts is shown. Exons are numbered starting from the 5' end of the gene. Cohorts are colour-coded as shown in the figure. Dashed line represents a splice acceptor variant.

Taken together, these observations provide strong evidence for the reclassification of the c.5769delG variant from VUS to P (Figure 4-19).

Pathogenic Criteria				Combination of Criteria Needed to Meet Classification							
Rule		Specification Type	Rule Description	Pathogenic				Likely Pathogenic			
VS	PVS1	Removed	~~Null variant in gene with established LOF as disease mechanism~~								
	PS1	No change	Different nucleotide change (same amino acid) as previously established pathogenic variant								
Strong	PS2	Disease/gene	<i>De novo</i> (paternity confirmed) in a patient with disease and no family history								
	PS3	Disease/gene	Functional studies of mammalian knock-in models supportive of a damaging effect on the gene or gene product	2	1	1	1	1	1		
	PS4	Disease/gene	Prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls –OR– Variant identified in ≥15 probands with consistent phenotypes								
	PP1_Strong	Modif. Strength	Variant segregates with ≥7 meiosis								
Moderate	PM1	Disease/gene	Hotspot/est. functional domain (amino acids 181-937) without benign variation								
	PM2	Disease/gene	Absent/extremely rare (<0.004%) from large population studies								
	PM3	Removed	~~Detected in trans with a pathogenic variant (recessive)~~								
	PM4	No change	Protein length changes due to in-frame deletions/insertions of any size in a non-repeat region or stop-loss variants								
	PM5	No change	Missense change at an amino acid residue where a different missense change previously established as pathogenic		3	2	1	1		3	2
	PM6	Disease/gene	Confirmed <i>de novo</i> without confirmation of paternity								
	PVS1_Moderate	Modif. Strength	Null variant in gene with evidence supporting LOF as disease mechanism								
	PS4_Moderate	Modif. Strength	Variant identified in ≥6 probands with consistent phenotypes								
	PP1_Moderate	Modif. Strength	Variant segregates in ≥5 meiosis								

Figure 4-19: Activated classification rules (as per the adapted ACMG/AMP framework by CMP-EP) for the MYH7fs variant (c.5769delG). Activated rules and the final classification are marked with a check. Adapted from Kelly *et al.*, 2018 under a Creative Commons license ²⁰²

4.4.10 *MYH7* frameshift variant c.5769delG is more common in AhcHCM patients compared to both LMM/OMGL and RbhHCM patients

Finally, I sought to assess the burden of the *MYH7* c.5769delG variant in AhcHCM patients in comparison to both LMM/OMGL and RbhHCM patients. The analysis showed that the c.5769delG was significantly enriched in the AhcHCM cohort compared to both LMM/OMGL and RbhHCM cohorts (n= 0/6112, p-value= 1.04×10^{-19} and n=0/684, p-value= 4.85×10^{-7} , respectively).

4.5 Discussion

Analysis of the genetic architecture of HCM in the Egyptian population (chapter 3) revealed that truncating variants in *MYH7* were significantly enriched in AhcHCM patients compared to EHVol controls. As demonstrated in this chapter, a single frameshift variant (fs) (c.5769delG) accounted for all truncations identified in *MYH7*. The prevalence of the MYH7fs variant in the AhcHCM cohort (n=514) analysed in this study was 3.31% (n=17 patients). Considering the current population size of Egypt (100M individuals) and the global prevalence of HCM (1:500), it could be estimated that ca. 200,000 Egyptians show clinical manifestations of HCM. Of these patients, >6000 were estimated to carry the MYH7fs. It is also worth mentioning that the MYH7fs variant was the most common cause of HCM in the analysed AhcHCM cohort. Taken together, these findings suggest that the MYH7fs variant is associated with HCM.

The aim of this chapter was to assess the evidence for the pathogenicity of the MYH7fs variant.

4.5.1 Initial discovery of MYH7fs variant

The first Egyptian HCM genetic study to report the identification of the MYH7fs variant c.5769delG was conducted by Kassem *et al.*, 2013 ²¹⁸. This study was also the first to screen for the 3 sarcomere genes: *MYH7*, *MYBPC3* and *TNNT2* in Egyptian HCM patients (n=192) ²¹⁸. The cohort was sequenced using denaturing high-pressure liquid chromatography (dHPLC) followed by Sanger sequencing to determine the nature of the sequence change. Sequencing analysis confirmed the presence of the MYH7fs variant in 3 unrelated Egyptian patients. The variant was classified as putative pathogenic due to its absence from 200 ethnically-, age- and sex-matched controls ²¹⁸. In a subsequent study, Kassem *et al.*, 2017 compared the variant detection rate in the same 3 sarcomere genes between dHPLC/Sanger sequencing and NGS targeted deep sequencing using the TruSight Cardio Sequencing Kit ³²⁰. This study was conducted on 150 Egyptian HCM patients. The MYH7fs was detected in 4 additional unrelated Egyptian HCM patients. At

the time of this second study, ExAC data was publicly available so the classification of the variant was not only based on its absence from ethnically-matched controls but also on having a MAF <0.01% in ExAC ³²⁰. According to these criteria, MYH7fs was classified as a putative pathogenic variant.

These 2 studies provided initial insights into the genetics of HCM in the Egyptian population as it had been previously underrepresented in genetic studies. At the time of the first study (in 2013) dHPLC coupled with Sanger sequencing was a common practice to detect putative disease-causing variants. However, it enabled genetic screening for a limited number of genes only and a subset of variants (n=19/50, 38%) were not detected by this technique given its relatively lower sensitivity (95%) compared to NGS ²¹⁸. This explains why, in the second study, the MYH7fs variant was detected in 2 patients by NGS only ³²⁰. In addition, dHPLC relies on the detection of mismatched heteroduplexes, thus, it cannot identify homozygous variants ³²¹. In the second study, the MYH7fs variant was classified as putatively pathogenic based on both its absence in 200 ethnically-matched controls and its rarity in ExAC. However, according to the current widely accepted view, variant rarity in large-scale reference genetic populations is insufficient to declare that it is pathogenic. That is because even healthy controls may carry rare protein-altering variants. For example, the release of the first large-scale reference population dataset, ExAC, revealed that the average exome harbours >7.4 million rare (i.e. MAF <0.1%) non-synonymous variants, the majority of which were novel ¹²³. It is also noteworthy that ExAC does not represent an ethnically-matched control cohort for the Egyptian population and screening for the variant in only 200 Egyptian controls may not yield sufficient power to confirm disease-causality. Also, details on the control group, such as the recruitment method, demographic and clinical characteristics were not provided to determine the degree to which they were matched to the HCM cohort. Therefore, in the present study, EHVol controls were clinically screened for cardiovascular abnormalities and PCA was performed to confirm that they were all of Egyptian origin. In addition, different analytical approaches were adopted to accurately assess the pathogenicity of the MYH7fs.

In the sections below, I discuss the major findings supporting the pathogenicity of the MYH7fs and their implications for clinical genetic testing.

4.5.2 Enrichment of MYH7fs variant in AhcHCM patients provides preliminary evidence for its association with HCM

The first step towards assessing the association of the MYH7fs variant with HCM was to compare its frequency in AhcHCM patients with ethnically-matched controls. In order to provide an accurate frequency of the MYH7fs in the patient cohort, I evaluated the relatedness of MYH7fs-carriers through IBD analysis. Relatedness analysis confirmed that MYH7fs-carriers were unrelated and also their geographical distribution did not show a specific pattern; with carriers originating from the North (e.g. Alexandria) and others from the South (e.g. Aswan).

I performed burden testing against the EHVol control cohort to assess whether the MYH7fs variant was significantly enriched in AhcHCM patients over ethnically-matched controls. The analysis showed that the variant was significantly enriched in AhcHCM patients over EHVol controls (3.31% vs. 0%, $p\text{-value}=5\times 10^{-5}$). Given the limited EHVol cohort size ($n=400$), I sought to investigate whether the variant was also significantly enriched over other ethnically-similar reference populations. The analysis showed that the burden of the MYH7fs in the AhcHCM cohort was significantly higher compared to the North East African (NEA) sub-population ($n=368$) of the Great Middle Eastern (GME) study (3.31% vs. 0.27%, $p\text{-value}=8\times 10^{-4}$). Overall, the frequency of the MYH7fs in both the GME population ($n=985$) and the published Iranian ($n=797$) dataset was significantly lower compared to the AhcHCM cohort (GME: 0.1%, $p=1\times 10^{-7}$; Iran: 0%, $p=1\times 10^{-7}$). The MYH7fs variant was also absent in the 'al mena' database, containing >26 million genetic variants from Arab and MENA populations, including Qatar ($n=1,213$), Iran ($n=77$) and GME ($n=1,002$)²²⁸. Taken together, these observations provided initial evidence supporting the contribution of the MYH7fs variant to HCM.

Intriguingly, the variant was also absent from the extensive gnomAD reference dataset ($n > 125,000$) as well as from the large-scale published predominantly Caucasian LMM/OMGL HCM cohort ($n > 6,000$). Thus, I hypothesised that the MYH7fs variant could be specific to HCM patients of Egyptian or North African origin. To test this hypothesis, I interrogated PubMed to evaluate whether this variant had been previously reported in non-North African or non-Middle Eastern HCM genetic studies. The literature review revealed that the MYH7fs variant was reported in a Norwegian infant (< 1 year) who was referred for clinical genetic testing ²²⁷. However, the criteria used to determine the ethnicity of patients included in the study was not provided.

Moreover, to assess the interpretability or the pathogenicity of the MYH7fs variant, I calculated the odds ratio (OR) for HCM. The OR describes the odds of carrying the MYH7fs variant in the AhcHCM cohort to the odds of carrying the variant in controls. Given that the EHVol and NEA cohorts are ethnically-matched; the NEA cohort consists of 91.3% Egyptians and 8.7% Libyans, I combined both cohorts to increase the size of the control population ¹³⁴. The calculated OR was 26.19 (95% CI: 4.08-1092.15), which confirmed that there was a significant enrichment of the MYH7fs variant in AhcHCM patients over controls. I then derived the etiological fraction (EF) from the OR in order to estimate the probability that the MYH7fs when identified in an HCM patient, is causative. The calculated EF was 0.962 (95% CI: 0.755-0.999), which suggests that, in the absence of additional evidence, such as functional or segregation data, there is a high probability of the MYH7fs variant being pathogenic when detected in HCM patients. It is worth mentioning, however, that the control estimate of the MYH7fs frequency is uncertain given the limited cohort size ($\text{EHVol} + \text{NEA} = 1/767$), which explains why the confidence intervals (CIs) for the calculated OR and EF were very wide. Thus, expanding the EHVol control cohort size will provide a more precise estimate of the observed OR and EF.

Furthermore, I estimated the population-level penetrance of the MYH7fs variant, that is, the probability that a randomly selected Egyptian individual carrying the variant will be affected with HCM. The penetrance was estimated to be 5% (95% CI: 0.6%-48%), which

suggests that an Egyptian individual carrying the c.5769delG variant has only a 5% chance of developing HCM. Population-level penetrance was estimated based on the baseline risk for HCM (1:500) as well as the case and population control allele frequencies³¹². Given the absence of a large-scale North African reference database to date, it remains challenging to accurately estimate the frequency of the MYH7fs in the population. For example, the MYH7fs variant was observed in only 1 individual from the NEA population (n=367), whose cardiac phenotype is not known. The variant was absent from EHVol controls (n=400). However, the limited size of the combined control cohort (i.e. EHVol+NEA) reduces the certainty of the estimated population-level penetrance, which is crucial for appropriate genetic counselling. For example, the identification of a highly penetrant variant in a screened family member may lead to the initiation of more frequent clinical screening, the implementation of preventive treatment and may also affect the carrier's reproductive choices. Therefore, expanding the current EHVol cohort will provide a very valuable resource for estimating the population-level penetrance of the MYH7fs especially that all individuals recruited to the EHVol study will be clinically screened.

However, the enrichment of the MYH7fs variant in Egyptian patients prompted me to study it in more detail.

4.5.3 Segregation analysis confirmed the association of MYH7fs with HCM

I attempted to evaluate whether the MYH7fs variant co-segregates with HCM in a large Egyptian family. Thus, all possible family members of a 29-year old proband with the MYH7fs variant were recruited to AHC for clinical and genetic screening using the TruSight Cardio Sequencing Kit. Family data showed that the variant co-segregated with HCM in this family as all affected members carried the MYH7fs variant. Also, all analysed family members did not carry other rare protein-altering variants in validated HCM genes

¹⁴⁸.

The analysis also showed that of the 11/16 (68.8%) variant-carriers, only 6 presented with clinical features of HCM. These observations were not unexpected as HCM-associated

variants are typically characterised by incomplete disease penetrance ²⁸. The presence of other genetic or environmental factors could explain the variable penetrance among MYH7fs-carriers ³²². It is worth mentioning that multiple consanguineous marriages (i.e. between first cousins) were identified in this family. Thus, 1 possibility for the variable penetrance observed in this family could be attributed to the inheritance of an additional homozygous variant in non-ICC genes, which may modify the expression of the MYH7fs variant. However, this is beyond the scope of this study as it would require an analysis of WES data.

Moreover, of all affected MYH7fs-carriers (n=7), including the proband, 4 underwent surgical myectomy. These observations highlight the variable expressivity of the MYH7fs variant, which is also a common characteristic of HCM-causing variants ²⁸. MYH7fs-carriers who underwent myectomy were significantly younger compared to carriers who did not require surgical intervention (30 ± 9.2 vs. 58 ± 3.0 , p-value=0.005). It is possible that other genetic or non-genetic factors aggravated the clinical phenotype in these patients.

To assess the co-segregation of the MYH7fs variant with HCM, the LOD score was calculated as per the equation described by ClinGen (refer to section 4.3.5) ¹⁷⁶. The calculated LOD score for this family was estimated to be 3.01, which corresponds to 10 meioses/segregations. According to the ClinGen Inherited Cardiomyopathy Expert Panel, ≥ 7 meioses/segregations provide strong evidence of variant co-segregation with disease ²⁰². Therefore, MYH7fs-carriers who remain free of HCM for their entire lives should still be informed that they can pass the disease-causing c.5769delG variant to their offspring.

Finally, data from this family was used to estimate the age-related penetrance of the MYH7fs variant. The analysis showed that at age 40-50, the penetrance of the MYH7fs variant was estimated to be 37%. In contrast, MYH7fs-carriers below the age of 18 were estimated to be negative for the HCM phenotype. Given the age-related penetrance of the MYH7fs, currently young family members might eventually develop HCM.

4.5.4 RNA analysis confirmed that MYH7fs is not a LoF variant

To investigate the disease mechanism of the MYH7fs, I first studied its effect on the cDNA- and predicted amino acid (AA) sequences. The analysis showed that the variant was a deletion of nucleotide G in the penultimate exon (i.e. exon 39), which introduced a frameshift leading to the generation of a distal PTC (i.e. close to the c-terminus). Since PTCs in the last exon of a gene (and thus downstream of the last exon-exon junction) are believed to be NMD-incompetent, I hypothesised that the resulting variant *MYH7* transcript could have escaped NMD ¹⁶². To test this hypothesis, I sought to examine whether the variant *MYH7* transcript was expressed by analysing RNA sequencing data of myocardial tissues and iPSC-derived CMs from 2 different AhcHCM MYH7fs-carriers. As expected, sequencing analysis confirmed the expression of the variant *MYH7* transcript harbouring the deleted nucleotide in exon 39, which was supported further by the ASE data. To illustrate, the alternate allelic ratio (AR) of the variant *MYH7* transcript to the WT transcripts was 0.3:0.7 and 0.4:0.6 in both patients, respectively. These observations demonstrate that the variant *MYH7* transcript harbouring the frameshift variant was not strongly degraded by NMD. In fact, Rivas *et al.*, 2015 explained that if 1 allele is expressed at a very low level or not expressed at all the AR would be <0.1, which was not the case here ³¹⁹.

Taken together, these observations provide new insights into the disease mechanism of the MYH7fs variant c.5769delG. Instead of being degraded by NMD to prevent disease, the variant *MYH7* transcript escaped NMD, which resulted in the expression of a truncated protein with potentially a disease-causing function. Currently, the widely accepted view regarding the disease mechanism of *MYH7* proteins with missense variants is that they function as “poison peptides” that incorporate into the sarcomere to exert a gain-of-function effect on the cardiomyocyte leading to hypercontractility ²⁷⁷.

Since the MYH7fs variant was caused by a nucleotide deletion that was predicted to ultimately give rise to a near full-length protein (i.e. only 4 AAs shorter than the wildtype), it could be postulated that the variant may have behaved as an indel and not a LoF

variant. Thus, the resulting mutant protein may have functioned as a “poison peptide” to contribute to HCM pathogenesis, although further studies are warranted to confirm this.

4.5.5 Detailed analysis of c.5769delG in Egyptian HCM patients provides valuable insights for the application of clinical genetic testing

As discussed above, detailed analysis of the MYH7fs variant (c.5769delG) in the Egyptian HCM cohort provided multiple lines of evidence to support its reclassification from a variant of uncertain significance (VUS) to a pathogenic (P) variant. To summarise, I have shown that the variant was absent (gnomAD, Iranome, al mena, EHVol) or rarely (NEA-GME) observed in population databases (ACMG/AMP PM2 rule). In addition, it was significantly enriched in AhcHCM patients compared to ethnically-matched controls (PS4 rule) and was predicted to act more like an indel, which changed the length of the *MYH7* protein (PM4 rule). Finally, family data demonstrated that the variant strongly co-segregates with HCM (PP1_Strong rule).

These observations provide very valuable insights for the application of clinical genetic testing in Egypt. To illustrate, in the absence of the data provided in this analysis, Egyptian HCM patients carrying the MYH7fs variant would have received an inconclusive result (i.e. VUS) during diagnostic testing, which is not acted upon clinically. This cohort-level study, however, demonstrated that the MYH7fs variant is not only a pathogenic variant but also the most common cause of HCM in the analysed cohort. Hence, the MYH7fs variant should be considered in clinical genetic testing, reproductive counselling and potentially personalised treatment.

Therefore, this study highlights the value of conducting population-specific analysis of HCM genetic variation to facilitate a more equitable application of clinical genetic testing. Since populations from the MENA region are expected to share similar genetic profiles with Egyptians, the interpretation of the MYH7fs variant will no longer be ambiguous if identified in any of these neighbouring populations. Although, there is a possibility that the MYH7fs variant could be a founder variant that is specific to patients of Egyptian

origin. To examine this, haplotype analysis of MYH7fs-carriers will be performed in the future. This analysis could help determine whether the MYH7fs resides on a haplotype that is shared among all carriers, because it was inherited from a common ancestor, or if the variant is simply a recurrent variant.

As discussed above, a clear association of the MYH7fs variant with HCM has been established, yet, the exact molecular mechanism by which the variant causes disease remains to be elucidated. Due to time limitations, I was unable to address this in my work. However, I have outlined below the different technologies that could be utilised in the future to study the molecular mechanism by which the MYH7fs variant causes HCM in more detail.

4.5.6 Functional analyses for the elucidation of the molecular mechanism of the MYH7fs variant

The groundbreaking discovery of adult cell reprogramming to induced pluripotent stem cells (iPSCs) and their differentiation into cardiomyocytes (CMs) enabled the possibility to study the disease mechanism of HCM in an individualised context ³²³. In the past few years, numerous studies have reported the effect of *MYH7* variants on the phenotypes of iPSC-derived CMs generated from patients with HCM. For example, studies have shown that in contrast to healthy controls, iPSC-derived CMs harbouring missense variants in *MYH7* showed an increased level of multinucleation and increased Ca²⁺ sensitivity resulting in hypercontractility ^{324,325}. Also, a recent study by Toeffer *et al.*, 2020 reported that human iPSC-derived CMs with missense variants in *MYH7* significantly increased the DRX/SRX ratio in comparison to healthy controls, which led to elevated ATP hydrolysis, hypercontractility and delayed relaxation ²⁷⁷. Treatment of mutant iPSC-derived CMs with the ATPase inhibitor MYK-461 restored the SRX/DRX ratio to a level that was comparable to healthy controls ²⁷⁷.

Therefore, as a continuation of this project, iPSC-derived CMs from AhcHCM patients harbouring the MYH7fs will be utilized to study the effect of this variant on cardiomyocyte

morphology, electrophysiological properties, contractile and relaxation functions and metabolism. So far, the team at AHC has generated 2 iPSC-derived CM cell lines from patients with the MYH7fs. These cell lines provide a valuable resource for studying the cellular and molecular pathophysiological mechanisms of the MYH7fs variant in the future. In addition, clustered regularly interspaced short palindromic repeat (CRISPR/Cas) technology could be adopted to engineer the MYH7fs variant in a wildtype parent cell line. This technology uses single non-sequence specific protein combined with a small guiding RNA molecule to enable the introduction or correction of variants in genes with high accuracy ³²⁶. Thus, by adopting this approach, the parent cell line could serve as an isogenic control line to which the engineered cell line harbouring the MYH7fs variant will be compared to ³²⁷. Generating a pair of wildtype and mutant isogenic iPSC-derived CM cell lines will allow us to focus on the disease mechanism caused solely by the MYH7fs variant.

To confirm the hypothesis that the variant *MYH7* transcript harbouring the c.5769delG variant escaped NMD, we could potentially introduce a homozygous variant in the isogenic iPS cell line and assess whether the truncated *MYH7* protein accumulates *in vitro*. We could also knockdown the mutant *MYH7* transcript in iPSC-derived CMs harbouring the MYH7fs variant using small interfering RNA (siRNA) to examine whether this could rescue the disease phenotype.

The development of transgenic mouse and zebrafish models harbouring the MYH7fs variant will also provide additional insights into the disease progression of HCM. The first mouse models of HCM were created in 1996 by Vikstrom *et al.*, and Geisterfer-Lowrance *et al.*, and both studies showed that missense variants in *Myh6*, which encodes the mouse cardiac isoform α -myosin heavy chain, gave rise to a cardiac histopathology that was comparable to what has been observed in human patients ^{328,329}. More recently, Toepfer *et al.*, 2020 demonstrated that, similar to mutant human iPSC-derived CMs, the myocardium in transgenic mice harbouring pathogenic *MYH7* missense variants destabilised ILM interactions, which caused the HCM phenotype ²⁷⁷. However, a well-

known limitation of these studies is that adult mice express only the α -myosin heavy chain in the myocardium whereas larger mammals express mainly the β -myosin heavy chain in their ventricle ³³⁰. Thus, other studies developed transgenic rabbits with *MYH7* variants as their myosin composition is similar to that of the human myocardium ^{330,331}. More recently, the zebrafish (*Danio rerio*) has been recognised as a valuable animal model for HCM given the ease of its genetic manipulation (e.g. morpholino or CRISPR/Cas technologies), development of transparent embryos *ex utero*, their small size, low cost and ability to survive without active circulation at early stages of development ³³². Therefore, in the new massive extension of the current AHC a zebrafish facility will be included, which will be utilised to further support the analysis of the disease mechanism of the MYH7fs variant in the future.

4.6 Conclusions

Characterising population-specific variants that underlie HCM will not only deepen our understanding of the genetic basis of the disease but will also enhance the applicability of clinical genetic testing in these populations. In this research study, I have provided multiple lines of evidence supporting the pathogenicity of a highly prevalent *MYH7* truncation in the Egyptian population that was absent from large-scale predominantly Caucasian populations. Therefore, this study highlights the value of studying HCM genetic variation in underrepresented populations to enable a more equitable application of genetic testing. Based on the data presented in this work, cardiologists should be informed of the clinical actionability of the MYH7fs variant (c.5769delG) and aim to facilitate appropriate clinical management strategies. This variant has been submitted to ClinVar (Accession no. SCV001250915) in order to benefit the wider scientific and clinical community.

5 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The overall aim of this research study was to define the genetic architecture of HCM in the Egyptian population using ethnically-matched controls as a reference for background variation. In this section, I highlight the impact of this study on the wider HCM scientific community. I also elaborate on the benefit of the conducted research to the currently understudied Egyptian population.

The clear underrepresentation of MENA populations in HCM genetic studies hampers our ability to fully elucidate the genetic basis of the disease, which consequently increases health inequalities and impedes our ability to practice precision medicine in the region.

Therefore, the different analyses on Egyptian rare genetic variation in well-characterised HCM genes performed in this study provide an initial milestone towards understanding the genetic underpinnings of the disease in this ethnically diverse population. Analysis of only 514 Egyptian patients has already provided new insights into the genetic basis of HCM. For example, the research findings showed that, in contrast to the current widely accepted view regarding the contribution of *MYH7* truncations to HCM, specific truncating variants, can be pathogenic. In addition, it highlighted the value of using an ethnically-matched control population to accurately interpret variants detected in Egyptian HCM patients. The study also showed that 30.6% of Egyptian patients were offspring to consanguineous parents. Since consanguinity is a key characteristic of MENA populations as opposed to European countries, studying genetic variation in these individuals could allow for the elucidation of the effect of homozygosity on the HCM phenotype, which could provide novel insights into the pathogenesis of the disease. More importantly, the issue of consanguinity paves the way for future research into the identification of novel recessive HCM genes pertaining to ethnic-specific populations. Similar to the approach behind the discovery of the novel recessive HCM gene *ALPK3*, WES could be performed on genotype-negative paediatric HCM patients with healthy, consanguineous parents to identify novel genetic associations with HCM ³³³.

Another advantage of studying the genetic architecture of HCM in Egyptian patients is that it could facilitate a more equitable application of clinical genetic testing in the country. Currently, Egyptian patients are more likely to receive inconclusive genetic testing results that are not acted upon clinically and this is borne out in the analysed data. To date, the key benefits of clinical genetic testing are to provide a more accurate diagnosis (e.g. HCM vs. Fabry's disease), to aid in patient risk stratification, and to allow for cascade family screening in order to identify pathogenic variant-carriers at a preclinical stage and thereby offer longitudinal clinical follow-up and reproductive counselling. If Egyptian patients more commonly receive results reporting VUSs they will not benefit from these advantages of clinical genetic testing. Also, patients may feel frustrated for not being able to interpret their ambiguous genetic results. Thus, this study opens new avenues for further characterising VUSs in order to improve their interpretation. Given that Egyptian families are typically large, it is possible to accurately assess the co-segregation of these variants with HCM to gain deeper insights into their pathogenicity as was performed in this study for the *MYH7* frameshift variant c.5769delG.

It is worth mentioning that the true value of HCM genetic studies lies in their ability to translate the obtained findings to the clinic so that patients can benefit from personalised treatment. The marked genetic and phenotypic heterogeneity of HCM is impeding our ability to develop individualised treatment and clinical management strategies following genetic testing. These goals will only be materialised if knowledge on the factors contributing to HCM are shared through worldwide collaborations between different centres of excellence with the aim to elucidate the underlying causes of HCM heterogeneity. Indeed, SHaRe, which brings together cardiologists and geneticists from different internationally renowned HCM specialty centres to develop a massive HCM dataset, has provided new insights into the natural history and burden of adverse outcomes associated with the disease ⁸⁰. For example, SHaRe demonstrated that patients with pathogenic sarcomere mutations presented with an earlier HCM onset and a higher risk of developing complications related to atrial fibrillation and HF ⁸⁰. These young patients (i.e. <40 years), however, tended to develop these complications around

20 years post diagnosis. This means that there may be an extended period of time during which patients could receive effective treatment to prevent or delay the progression to atrial fibrillation or HF provided the genetic factors and the associated disease mechanisms driving these outcomes are accurately identified ⁸⁰. Thus, considering the heterogeneous nature of the clinical course of HCM and its potential treatability, expanding our current knowledge of the disease to other currently underrepresented ethnic populations, such as Egypt, could provide novel genetic and mechanistic insights and uncover therapeutic targets for the disease.

Indeed, this research study revealed a novel pathogenic variant in *MYH7* that was highly prevalent among Egyptian HCM patients but was absent in the large-scale published LMM/OMGL cohort. Given that truncating variants in *MYH7* are generally believed to be non-causative in HCM, the identification of an enriched frameshift variant (c.5769delG) in Egyptian patients provided novel insights into the disease mechanism of HCM. As outlined in chapter 4, detailed analysis of this variant revealed that it strongly contributes to HCM and thus should be reclassified from a VUS to a pathogenic variant. This finding has significant benefits to the Egyptian community. Firstly, patients carrying this variant will no longer receive an inconclusive genetic testing result and instead be informed that they harbour a pathogenic variant that could be passed on to their offspring. Secondly, genetic testing of these patients' families will aid in the identification of at-risk c.5769delG-carriers. It is important to note that AHC represents one of the very few cardiac centres in Egypt that provide medical services free-of-charge. As a result, the waiting list for admission is very long. Thus, this risk stratification process will ensure that only relevant family members will be offered regular clinical surveillance. Also, considering a country like Egypt, where a significant proportion of marriages are consanguineous, genetic testing of related parents followed by reproductive counselling could guide their decision around having children. Of course, the ultimate goal is to prevent the disease from developing at the preclinical phase in at-risk c.5769delG-carriers, which could be possible if the disease mechanism through which the variant drives the HCM phenotype is well studied through *in vitro* and *in vivo* analyses (outlined in section 4.5.6)

New insights into the disease mechanism of this *MYH7* variant could facilitate the discovery of therapeutic drugs. For example, translational research initiatives such as MyoKardia are implementing precision medicine approaches to discover small therapeutic candidates for the treatment of cardiovascular disease including HCM based on the patients' genetic background (<https://www.myokardia.com/about>). The MyoKardia initiative has succeeded in developing the ATPase inhibitor MYK-461 (mavacamten), which has been shown to suppress the development of left ventricular hypertrophy in mice harbouring variants in *MYH7* and is currently in clinical trial phase 3 ²⁷⁷.

Furthermore, analysis of rare variation in the work presented here was focussed on well-characterised HCM genes. While these genes are more likely to yield interpretable results, this approach hinders the possibility of identifying novel, non-ICC genes that could be associated with HCM. In fact, the moderately-sized GWAS study performed by Wooten *et al.*, 2013, has identified a significant association of a common intronic variant in *FHOD3* with HCM ¹⁸⁶. Other large-scale case-control GWAS studies, such as the recently published study by Pirruccello *et al.*, 2020, demonstrated the role of common variants with low effect sizes in influencing cardiac structure and function ³³⁴.

Thus, to expand our knowledge on the genetic determinants of HCM, WES or WGS approaches could be adopted in the future. Using these approaches, well-characterised HCM genes could still be pre-selected for the analysis and, in parallel, novel gene-disease associations could be discovered.

GWAS could also be adopted to study the contribution of common variation to HCM. The advent of large, well-powered GWAS studies has facilitated the construction of polygenic risk scores (PRS), which utilise the small effect sizes of numerous common variants estimated from GWAS data to generate an individual-level risk score to predict disease susceptibility ⁹⁴. Therefore, the PRS approach offers hope for introducing lifestyle changes or pre-emptive therapeutic strategies for individuals at risk of developing disease ³³⁵. One of the main challenges in utilising PRS in the clinical setting, however, is ensuring

their applicability across diverse ancestral populations in order to reduce health inequalities. Since these scores have largely been generated from European-based GWAS cohorts, they may misestimate disease risk in underrepresented populations, such as those of African ancestry ³³⁶. Therefore, as a continuation of this study, the ECCO-GEN project could initiate a collaboration with the H3Africa consortium with the aim to investigate novel genetic associations with HCM through the development of a sufficiently powered African-based GWAS study. It is noteworthy that H3Africa has developed a SNP array encompassing >2 million variations from different African populations, thus, this array will be most suitable for studying common genetic variation in this ancestral group ¹⁶⁸. Insights gained from this study could facilitate the transferability of African-based PRS to health care centres across the region thereby improving health outcomes in these currently severely underserved populations.

To conclude, this research study refines our current understanding of the genetic architecture of HCM in Egypt and opens new opportunities for expanding our knowledge on the pathogenesis of this disease. Subsequent studies arising from this work will help reduce the currently evident health inequalities in this underserved population and potentially improve the health outcomes of Egyptian patients. After all, the main objective of genetic studies is to drive the shift from population- to individual-based treatment approaches in order to offer patients effective, personalised therapies. This hope was recognised a century ago by the father of modern medicine Sir William Osler, who said: *“The good physician treats the disease; the great physician treats the patient who has the disease”*.

6 BIBLIOGRAPHY

1. WHO. Cardiovascular diseases (CVDs). *Fact sheet No. 317* (2017) doi:10.1016/j.molcel.2011.10.016.
2. Fuster, V. & Kelly, B. B. *Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health*. (National Academies Press (US), 2010).
3. Yacoub, M. H. Decade in review—cardiomyopathies: Cardiomyopathy on the move. *Nat. Rev. Cardiol.* **11**, 628–629 (2014).
4. Arbustini, E. *et al.* The MOGE(S) classification of cardiomyopathy for clinicians. *J. Am. Coll. Cardiol.* **64**, 304–318 (2014).
5. McNally, E. M., Golbus, J. R. & Puckelwartz, M. J. Genetic mutations and mechanisms in dilated cardiomyopathy. *J. Clin. Invest.* **123**, 19–26 (2013).
6. Burke, M. A., Cook, S. A., Seidman, J. G. & Seidman, C. E. Clinical and Mechanistic Insights Into the Genetics of Cardiomyopathy. *J. Am. Coll. Cardiol.* **68**, (2016).
7. Kapplinger, J. D., Landstrom, A. P., Callis, T. E. & Ackerman, M. J. Distinguishing Hypertrophic Cardiomyopathy-Associated Mutations from Background Genetic Noise. *J. Cardiovasc. Transl. Res.* **7**, 347–361 (2014).
8. Braunwald, E. & Marian, A. J. Cardiomyopathy Compendium Hypertrophic Cardiomyopathy. *Circ. Res.* **121**, 749–770 (2017).
9. Teekakirikul, P., Padera, R. F., Seidman, J. G. & Seidman, C. E. Hypertrophic cardiomyopathy: Translating cellular cross talk into therapeutics. *J. Cell Biol.* **199**, 417–421 (2012).
10. Morita, H., Seidman, J. & Seidman, C. E. Genetic causes of human heart failure. *J. Clin. Invest.* **115**, 518–526 (2005).
11. Lopes, L. R. *et al.* Novel genotype–phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy. *Heart* **101**, 294–301 (2015).
12. Maron, B. J. *et al.* Prevalence of hypertrophic cardiomyopathy in a general population of young adults: Echocardiographic analysis of 4111 subjects in the CARDIA study. *Circulation* **92**, 785–789 (1995).
13. Helms, A. S. & Day, S. M. Hypertrophic cardiomyopathy: single gene disease or complex trait? *Eur. Heart J.* **37**, 1823–1825 (2016).
14. Kuusisto, J., Sipola, P., Jääskeläinen, P. & Naukkarinen, A. Current perspectives in hypertrophic cardiomyopathy with the focus on patients in the Finnish population: a review. *Ann. Med.* **48**, 496–508 (2016).
15. Maron, B. J. Hypertrophic cardiomyopathy: A systematic review. *JAMA Cardiol.* **287**, 1308–1320 (2002).
16. Maron, B. J., Mathenge, R., Casey, S. A., Poliac, L. C. & Longe, T. F. Clinical profile of hypertrophic cardiomyopathy identified de novo in rural communities. *J. Am. Coll. Cardiol.* **33**, 1590–1595 (1999).
17. Lipshultz, S. E. *et al.* The Incidence of Pediatric Cardiomyopathy in Two Regions of the United States. *N. Engl. J. Med.* **348**, 1647–1655 (2003).
18. Hada, Y. *et al.* Prevalence of hypertrophic cardiomyopathy in a population of adult Japanese workers as detected by echocardiographic screening. *Am. J. Cardiol.* **59**,

- 183–184 (1987).
19. Zou, Y. *et al.* Prevalence of idiopathic hypertrophic cardiomyopathy in China: A population-based echocardiographic analysis of 8080 adults. *Am. J. Med.* **116**, 14–18 (2004).
 20. Maron, B. J. *et al.* Prevalence of Hypertrophic Cardiomyopathy in a General Population of Young Adults. *Circulation* **92**, 785–789 (1995).
 21. Sheikh, N. *et al.* Comparison of hypertrophic cardiomyopathy in Afro-Caribbean versus white patients in the UK. *Heart* **102**, 1797–1804 (2016).
 22. Sheikh, N. *et al.* Epidemiology and Prevention Comparison of Electrocardiographic Criteria for the Detection of Cardiac Abnormalities in Elite Black and White Athletes. *Circulation* **129**, 1637–1649 (2014).
 23. Lopes, L. R., Rahman, M. S. & Elliott, P. M. A systematic review and meta-analysis of genotype-phenotype associations in patients with hypertrophic cardiomyopathy caused by sarcomeric protein mutations. *Heart* **99**, 1800–1811 (2013).
 24. Sedaghat-hamedani, F. *et al.* Clinical outcomes associated with sarcomere mutations in hypertrophic cardiomyopathy: a meta-analysis on 7675 individuals. *Clin. Res. Cardiol.* **107**, (2018).
 25. Ho, C. Y. *et al.* Genetic advances in sarcomeric cardiomyopathies: State of the art. *Cardiovasc. Res.* **105**, 397–408 (2015).
 26. Wolf-maier, K., Cooper, R. S. & Rodri, F. Hypertension Prevalence and Blood Pressure Levels in 6 European Countries , Canada , and the United States. *JAMA* **289**, 2363–2369 (2003).
 27. Rodrigues, J. C. *et al.* Hypertensive heart disease versus hypertrophic cardiomyopathy: multi-parametric CMR predictors beyond end-diastolic wall thickness $\geq 15\text{mm}$. *J. Cardiovasc. Magn. Reson.* **18**, 1–3 (2016).
 28. Alfares, A. A. *et al.* Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet. Med.* **17**, 880–888 (2015).
 29. Zamorano, J. L. *et al.* 2014 ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy: The task force for the diagnosis and management of hypertrophic cardiomyopathy of the European Society of Cardiology (ESC). *Eur. Heart J.* **35**, 2733–2779 (2014).
 30. Veselka, J. Alcohol septal ablation for hypertrophic obstructive cardiomyopathy: is it safe? *Curr. Cardiol. Rev.* **4**, 193–197 (2008).
 31. Maron, B. J. *et al.* Epidemiology of Hypertrophic Cardiomyopathy–Related Death. *Circulation* **102**, 858–865 (2000).
 32. Harris, K. M. *et al.* Prevalence, clinical profile, and significance of left ventricular remodeling in the end-stage phase of hypertrophic cardiomyopathy. *Circulation* **114**, 216–225 (2006).
 33. Williams, L., Gruner, C. & Rakowski, H. Left ventricular outflow tract obstruction in hypertrophic cardiomyopathy: Do we need to stand to be counted. *J. Am. Soc. Echocardiogr.* **24**, 83–85 (2011).
 34. Teekakirikul, P., Zhu, W., Huang, H. C. & Fung, E. Hypertrophic cardiomyopathy: An overview of genetics and management. *Biomolecules* (2019)

doi:10.3390/biom9120878.

35. DM, G. *et al.* A double-blind, placebo-controlled crossover trial of nadolol and verapamil in mild and moderately symptomatic hypertrophic cardiomyopathy. in *Journal of the American College of Cardiology* vol. 21 1672–1679 (1993).
36. Stroumpoulis, K. I. Hypertrophic cardiomyopathy and sudden cardiac death. *World J. Cardiol.* **2**, 289–298 (2010).
37. Elliott, P. M. *et al.* Historical trends in reported survival rates in patients with hypertrophic cardiomyopathy. *Heart* **92**, 785–791 (2006).
38. Uçar, F. M., Yilmaztepe, M. A., Taylan, G. & Aktoz, M. Non-Sustained Ventricular Tachycardia Episodes Predict Future Hospitalization in ICD Recipients with Heart Failure. *Arq. Bras. Cardiol.* **4**, 284–289 (2017).
39. Chan, R. H. *et al.* Prognostic value of quantitative contrast-enhanced cardiovascular magnetic resonance for the evaluation of sudden death risk in patients with hypertrophic cardiomyopathy. *Circulation* **130**, 484–495 (2014).
40. O'Mahony, C. *et al.* A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). *Eur. Heart J.* **35**, 2010–2020 (2014).
41. Geisterfer-Lowrance, A. A. T. *et al.* A molecular basis for familial hypertrophic cardiomyopathy: an alpha/beta cardiac myosin heavy chain hybrid gene. *Cell* **62**, 991–998 (1990).
42. Solomon, S. D. *et al.* Familial hypertrophic cardiomyopathy is a genetically heterogeneous disease. *J. Clin. Invest.* **86**, 993–999 (1990).
43. Thierfelder, L. *et al.* α -tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: A disease of the sarcomere. *Cell* **77**, 701–712 (1994).
44. Kimura A, Harada H, Park JE, Nishi H, Satoh M, Takahashi M, Hiroi S, S. T., Ohbuchi N, Nakamura T, Koyanagi T, Hwang TH, Choo JA, Chung KS, H. A., Nagai R, Okazaki O, Nakamura H, Matsuzaki M, Sakamoto T, Toshima H, K. Y. & Imaizumi T, S. T. Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. *Nat. Genet.* **16**, 379–382 (1997).
45. Poetter K, Jiang H, Hassanzadeh S, Master SR, Chang A, Dalakas MC, R. I. & Sellers JR, Fananapazir L, E. N. 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.* **13:63–69**, 63–69 (1996).
46. Mogensen, J. *et al.* A-Cardiac Actin Is a Novel Disease Gene in Familial Hypertrophic Cardiomyopathy. *J. Clin. Invest.* **103**, 39–43 (1999).
47. Konno, T., Chang, S., Seidman, J. G. & Seidman, C. E. Genetics of Hypertrophic Cardiomyopathy. *Curr. Opin. Cardiol.* **25**, 205–209 (2010).
48. Solaro, R. J. Sarcomere control mechanisms and the dynamics of the cardiac cycle. *J. Biomed. Biotechnol.* **2010**, 1–8 (2010).
49. Galińska-Rakoczy, A. *et al.* Structural Basis for the Regulation of Muscle Contraction by Troponin and Tropomyosin. *J. Mol. Biol.* **379**, 929–935 (2008).
50. Gautel, M. & Djinić-Carugo, K. The sarcomeric cytoskeleton: from molecules to motion. *J. Exp. Biol.* **219**, 135–145 (2016).

51. Cho, G. W., Altamirano, F. & Hill, J. A. Chronic heart failure: Ca²⁺, catabolism, and catastrophic cell death. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1862**, 763–777 (2016).
52. Kamisago, M. *et al.* Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N. Engl. J. Med.* **343**, 1688–1696 (2000).
53. Helms, A. S. *et al.* Hypertrophic Cardiomyopathy. *Circ. Cardiovasc. Genet.* **7**, 434–443 (2014).
54. Marston, S. *et al.* Evidence From Human Myectomy Samples That MYBPC3 Mutations Cause Hypertrophic Cardiomyopathy Through Haploinsufficiency. *Circ. Res.* **105**, 219–222 (2009).
55. Ashrafian, H., Redwood, C., Blair, E. & Watkins, H. Hypertrophic cardiomyopathy: A paradigm for myocardial energy depletion. *Trends Genet.* **19**, 263–268 (2003).
56. Fraysse, B. *et al.* Increased myofilament Ca²⁺sensitivity and diastolic dysfunction as early consequences of Mybpc3 mutation in heterozygous knock-in mice. *J. Mol. Cell. Cardiol.* **52**, 1299–1307 (2012).
57. Guinto, P. J., Haim, T. E., Dowell-Martino, C. C., Sibinga, N. & Tardiff, J. C. Temporal and mutation-specific alterations in Ca²⁺ homeostasis differentially determine the progression of cTnT-related cardiomyopathies in murine models. *AJP Hear. Circ. Physiol.* **297**, H614–H626 (2009).
58. Semsarian, C. *et al.* The L-type calcium channel inhibitor diltiazem prevents cardiomyopathy in a mouse model. *J. Clin. Invest.* **109**, 1013–1020 (2002).
59. Basso, C. *et al.* Hypertrophic cardiomyopathy and sudden death in the young: Pathologic evidence of myocardial ischemia. *Hum. Pathol.* **31**, 988–998 (2000).
60. Choudhury, L. *et al.* Myocardial scarring in asymptomatic or mildly symptomatic patients with hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **40**, 2156–2164 (2002).
61. Watkins, H. *et al.* Characteristics and Prognostic Implications of Myosin Missense Mutations in Familial Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **17**, 1108–1114 (1992).
62. Havndrup, O. *et al.* The Val 606 Met Mutation in the Cardiac Beta-Myosin Heavy Chain Gene in Patients With Familial Hypertrophic Cardiomyopathy Is Associated With a High Risk of Sudden Death at Young Age. *Am. J. Cardiol.* **87**, 1315–1317 (2001).
63. Semsarian, C. *et al.* SUDDEN CARDIAC DEATH IN FAMILIAL HYPERTROPHIC CARDIOMYOPATHY: ARE " BENIGN " MUTATIONS REALLY BENIGN? *Pathology* **29**, 305–308 (1997).
64. Roberts, W. C. *et al.* Dramatically Different Phenotypic Expressions of Hypertrophic Cardiomyopathy in Male Cousins Undergoing Cardiac Transplantation With Identical Disease-Causing Gene Mutation. *Am. J. Cardiol.* **111**, 1818–1822 (2013).
65. Landstrom, A. P. & Ackerman, M. J. Mutation Type is Not Clinically Useful in Predicting Prognosis in Hypertrophic Cardiomyopathy Andrew. *Circulation* **122**, 2441–2450 (2010).
66. Maron, B. J., Maron, M. S., Maron, B. A. & Loscalzo, J. Moving Beyond the Sarcomere to Explain Heterogeneity in Hypertrophic Cardiomyopathy. *J. Am. Coll.*

- Cardiol.* **73**, (2019).
67. Walsh, R. *et al.* Defining the genetic architecture of hypertrophic cardiomyopathy : re-evaluating the role of non-sarcomeric genes. *Eur. Soc. Cardiol.* **38**, 3461–3468 (2017).
 68. Arad, M. *et al.* Glycogen Storage Diseases Presenting as Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **352**, 362–372 (2005).
 69. Boczek, N. J. *et al.* 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. Arrhythmia Electrophysiol.* (2015) doi:10.1161/CIRCEP.115.002745.
 70. Van Spaendonck-Zwarts, K. Y. *et al.* Desmin-related myopathy. *Clin. Genet.* (2011) doi:10.1111/j.1399-0004.2010.01512.x.
 71. Knoblauch, H. *et al.* Contractures and hypertrophic cardiomyopathy in a novel FHL1 mutation. *Ann. Neurol.* (2010) doi:10.1002/ana.21839.
 72. Elliott, P. *et al.* Prevalence of Anderson-Fabry disease in patients with hypertrophic cardiomyopathy: The European Anderson-Fabry Disease Survey. *Heart* (2011) doi:10.1136/heartjnl-2011-300364.
 73. D'souza, R. S. *et al.* Danon disease clinical features, evaluation, and management. *Circ. Hear. Fail.* (2014) doi:10.1161/CIRCHEARTFAILURE.114.001105.
 74. Banankhah, P., Fishbein, G. A., Dota, A. & Ardehali, R. Cardiac manifestations of PRKAG2 mutation. *BMC Med. Genet.* (2018) doi:10.1186/s12881-017-0512-6.
 75. Tartaglia, M. *et al.* PTPN11 mutations in noonan syndrome: Molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity. *Am. J. Hum. Genet.* (2002) doi:10.1086/340847.
 76. Razzaque, M. A. *et al.* Germline gain-of-function mutations in RAF1 cause Noonan syndrome. *Nat. Genet.* (2007) doi:10.1038/ng2078.
 77. González-López, E., López-Sainz, Á. & García-Pavia, P. Diagnosis and Treatment of Transthyretin Cardiac Amyloidosis. Progress and Hope. *Rev. Española Cardiol. (English Ed.)* (2017) doi:10.1016/j.rec.2017.05.036.
 78. Weidemann, F. *et al.* Long-term effects of enzyme replacement therapy on fabry cardiomyopathy. Evidence for a better outcome with early treatment. *Circulation* **119**, 524–529 (2009).
 79. Patel, M. R. *et al.* Cardiovascular events in patients with Fabry disease: Natural history data from the Fabry Registry. *J. Am. Coll. Cardiol.* **57**, 1093–1099 (2011).
 80. Ho, C. Y. *et al.* Genotype and Lifetime Burden of Disease in Hypertrophic Cardiomyopathy. *Circulation* **138**, 1387–1398 (2018).
 81. Mazzarotto, F. *et al.* Contemporary Insights Into the Genetics of Hypertrophic Cardiomyopathy: Toward a New Era in Clinical Testing? *J. Am. Heart Assoc.* **9**, (2020).
 82. Morales, A. & Hershberger, R. E. Variants of Uncertain Significance Should We revisit How They Are Evaluated and Disclosed? *Circ. Genomic Precis. Med.* **11**, 1–4 (2018).
 83. Maron, B. J., Maron, M. S. & Semsarian, C. Genetics of hypertrophic cardiomyopathy after 20 years: Clinical perspectives. *Journal of the American*

- College of Cardiology* vol. 60 705–715 (2012).
84. Alekseyev, Y. O. *et al.* A Next-Generation Sequencing Primer — How Does It Work and What Can It Do? *Acad. Pathol.* **5**, 1–11 (2018).
 85. Heather, J. M. & Chain, B. Genomics The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
 86. Rizzo, J. M. & Buck, M. J. Key Principles and Clinical Applications of " Next-Generation " DNA Sequencing. *Cancer Prev. Res.* **5**, 887–901 (2012).
 87. Barba, M., Czosnek, H. & Hadidi, A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses* **6**, 106–136 (2014).
 88. Black, J. S., Salto-tellez, M., Mills, K. I. & Catherwood, M. A. Pathogenesis The impact of next generation sequencing technologies on haematological research e A review. *Pathogenesis* **2**, 9–16 (2015).
 89. Ambardar, S., Gupta, R., Trakroo, D. & Lal, R. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J. Microbiol.* **56**, 394–404 (2016).
 90. Chen, F., Dong, M., Ge, M., Zhu, L. & Ren, L. The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology. *Genomics. Proteomics Bioinformatics* **11**, 34–40 (2013).
 91. Pua, C. J. *et al.* Development of a Comprehensive Sequencing Assay for Inherited Cardiac Condition Genes. *J. Cardiovasc. Transl. Res.* **9**, 3–11 (2016).
 92. Fromer, M. *et al.* Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet.* **91**, 597–607 (2012).
 93. Meienberg, J., Bruggmann, R., Oexle, K. & Matyas, G. Clinical sequencing: is WGS the better WES? *Hum. Genet.* **135**, 359–362 (2016).
 94. Mazzarotto, F., Olivotto, I. & Walsh, R. Advantages and Perils of Clinical Whole-Exome and Whole-Genome Sequencing in Cardiomyopathy. *Cardiovasc. Drugs Ther.* **34**, 241–253 (2020).
 95. Suwinski, P. *et al.* Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. *Front. Genet.* **10**, 1–16 (2019).
 96. Carter, T. C. & He, M. M. Challenges of Identifying Clinically Actionable Genetic Variants for Precision Medicine. *J. Healthc. Eng.* 1–14 (2016) doi:10.1155/2016/3617572.
 97. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
 98. Li, H. *et al.* The Sequence Alignment / Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 99. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 1–14 (2016).
 100. Walsh, R. *et al.* Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet. Med.* **19**, 192–203 (2017).
 101. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, 862–868 (2016).
 102. Richards, S. *et al.* 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.* **17**, 405–423 (2015).
103. Shawky, R. M., Elsayed, N. S., Ibrahim, D. S. & Seifeldin, N. S. Profile of genetic disorders prevalent in northeast region of Cairo, Egypt. *Egypt. J. Med. Hum. Genet.* **13**, 45–62 (2012).
 104. Shawky, R. M., Y., E.-A. M., M., E. S. & Hamadan, G. E. Consanguineous matings among Egyptian population. *Egypt. J. Med. Hum. Genet.* **12**, 157–163 (2011).
 105. Belmont, J. W. *et al.* The international HapMap project. *Nature* **426**, 789–796 (2003).
 106. Clarke, L. *et al.* The 1000 Genomes Project: Data Management and Community Access. *Nat. Methods* **9**, 459–462 (2012).
 107. Altshuler, D. L. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
 108. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
 109. Auton, A. *et al.* A global reference for human genetic variation. *Nature* vol. 526 68–74 (2015).
 110. Zheng-Bradley, X. & Flicek, P. Applications of the 1000 Genomes Project resources. *Brief. Funct. Genomics* **16**, 163–170 (2017).
 111. Golbus, J. R. *et al.* Population-based variation in cardiomyopathy genes. *Circ. Cardiovasc. Genet.* **5**, 391–399 (2012).
 112. Stenson, P. D. *et al.* The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics* vol. 133 1–9 (2014).
 113. McNally, E. M. & Puckelwartz, M. J. Genetic variation in cardiomyopathy and cardiovascular disorders. *Circ. J.* **79**, 1409–1415 (2015).
 114. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
 115. Amendola, L. M. *et al.* Actionable exomic incidental findings in 6503 participants: Challenges of variant classification. *Genome Res.* **25**, 305–315 (2015).
 116. Tabor, H. K. *et al.* Pathogenic variants for mendelian and complex traits in exomes of 6,517 european and african americans: Implications for the return of incidental results. *Am. J. Hum. Genet.* **95**, 183–193 (2014).
 117. Dorschner, M. O. *et al.* Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am. J. Hum. Genet.* **93**, 631–640 (2013).
 118. Andreassen, C. *et al.* New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur. J. Hum. Genet.* **21**, 918–28 (2013).
 119. Stoneking, M. & Krause, J. Learning about human population history from ancient and modern genomes. *Nature Reviews Genetics* vol. 12 603–614 (2011).
 120. Tennessen, J. A. *et al.* Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science (80-.)*. **337**, 64–69 (2012).
 121. Lohmueller, K. E. *et al.* Proportionally more deleterious genetic variation in

- European than in African populations. *Nature* **451**, 994–997 (2008).
122. E.T., L. *et al.* Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. *PLoS Genet.* **10**, (2014).
 123. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
 124. Ruderfer, D. M. *et al.* Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat. Genet.* **48**, 1107–1111 (2016).
 125. Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nat. Genet.* **47**, 448–452 (2015).
 126. Narasimhan, V. M. *et al.* Health and population effects of rare gene knockouts in adult humans with related parents. *Science* (80-.). **352**, 474–477 (2016).
 127. Saleheen, D. *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
 128. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of- function intolerance across human protein-coding genes. *bioRxiv* (2019) doi:<https://doi.org/10.1101/531210>.
 129. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat. Genet.* **37**, 161–165 (2005).
 130. Stein, E. A. *et al.* Efficacy and safety of evolocumab (AMG 145), a fully human monoclonal antibody to PCSK9, in hyperlipidaemic patients on various background lipid therapies: Pooled analysis of 1359 patients in four phase 2 trials. *Eur. Heart J.* **35**, 2249–2259 (2014).
 131. Desai, N. R. *et al.* AMG145, a Monoclonal Antibody Against Proprotein Convertase Subtilisin Kexin Type 9, Significantly Reduces Lipoprotein(a) in Hypercholesterolemic Patients Receiving Statin Therapy. *Circulation* **128**, 962–969 (2013).
 132. Eilbeck, K., Quinlan, A. & Yandell, M. Settling the score: Variant prioritization and Mendelian disease. *Nature Reviews Genetics* vol. 18 599–612 (2017).
 133. Zlotogora, J., Patrinos, G. P. & Meiner, V. Ashkenazi Jewish genomic variants: integrating data from the Israeli National Genetic Database and gnomAD. *Genet. Med.* **20**, 867–871 (2018).
 134. Scott, E. M. *et al.* Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat. Genet.* **48**, 1071–1076 (2016).
 135. Yang, X. *et al.* The influence of admixture and consanguinity on population genetic diversity in Middle East. *J. Hum. Genet.* **59**, 615–622 (2014).
 136. Al-Gazali, L., Hamamy, H. & Al-Arrayad, S. Genetic disorders in the Arab world. *British Medical Journal* vol. 333 831–834 (2006).
 137. Ceballos, F. C., Hazelhurst, S. & Ramsay, M. Assessing runs of Homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics* **19**, (2018).
 138. Rodriguez-Flores, J. L. *et al.* Exome Sequencing Identifies Potential Risk Variants for Mendelian Disorders at High Prevalence in Qatar. *Hum. Mutat.* **35**, 105–116 (2014).
 139. Sivadas, A., Sharma, P. & Scaria, V. Landscape of warfarin and clopidogrel

- pharmacogenetic variants in Qatari population from whole exome datasets. *Pharmacogenomics* **17**, 1891–1901 (2016).
140. Fakhro, K. A. *et al.* The Qatar genome: A population-specific tool for precision medicine in the Middle East. *Hum. Genome Var.* **3**, (2016).
 141. Fattahi, Z. *et al.* Iranome: A catalog of genomic variations in the Iranian population. *Hum. Mutat.* **40**, 1968–1984 (2019).
 142. Vanderbilt. Software – REDCap. *REDCap - Research Electronic Data Capture* (2017) doi:10.1111/j.1365-2583.2009.00888.x ER.
 143. Kawel-Boehm, N. *et al.* Normal values for cardiovascular magnetic resonance in adults and children. *Journal of Cardiovascular Magnetic Resonance* vol. 17 (2015).
 144. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
 145. Kuhn, R. M. *et al.* The UCSC Genome Browser Database: Update 2009. *Nucleic Acids Res.* **37**, (2009).
 146. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
 147. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
 148. Ingles, J. *et al.* Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circ. Genomic Precis. Med.* **12**, 57–64 (2019).
 149. Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med.* **19**, 1151–1158 (2017).
 150. Allouba, M. *et al.* Analysis of HCM in an understudied population reveals a new mechanism of pathogenicity. *medRxiv* (2020) doi:10.1101/2020.03.24.20037358.
 151. Mazzarotto, F. *et al.* The genetic architecture of left ventricular non-compaction reveals both substantial overlap with other cardiomyopathies and a distinct aetiology in a subset of cases. *medRxiv* (2020) doi:https://doi.org/10.1101/2020.01.03.19015602 .
 152. de la Chica, J. A. *et al.* Association Between Left Ventricular Noncompaction and Vigorous Physical Activity. *J. Am. Coll. Cardiol.* **76**, 1723–1733 (2020).
 153. Gati, S. *et al.* Increased left ventricular trabeculation in highly trained athletes: Do we need more stringent criteria for the diagnosis of left ventricular non-compaction in athletes? *Heart* **99**, 401–408 (2013).
 154. Rooryck, C. *et al.* New Family with Catecholaminergic Polymorphic Ventricular Tachycardia Linked to the Triadin Gene. *J. Cardiovasc. Electrophysiol.* **26**, 1146–1150 (2015).
 155. Hancox, J. C., James, A. F., Walsh, M. A. & Stuart, A. G. Triadin mutations - a cause of ventricular arrhythmias in children and young adults. *J. Congenit. Cardiol.* **1**, (2017).
 156. Chopra, N. & Knollmann, B. C. Triadin regulates cardiac muscle couplon structure and microdomain Ca²⁺ signalling: A path towards ventricular arrhythmias. *Cardiovascular Research* vol. 98 187–191 (2013).
 157. Clemens, D. J. *et al.* International Triadin Knockout Syndrome Registry: The

- Clinical Phenotype and Treatment Outcomes of Patients with Triadin Knockout Syndrome. *Circ. Genomic Precis. Med.* **12**, 65–75 (2019).
158. Altmann, H. M. *et al.* Homozygous/compound heterozygous triadin mutations associated with autosomal-recessive long-QT syndrome and pediatric sudden cardiac arrest: Elucidation of the triadin knockout syndrome. *Circulation* **131**, 2051–2060 (2015).
 159. Roux-buisson, N. *et al.* Absence of triadin, a protein of the calcium release complex, is responsible for cardiac arrhythmia with sudden death in human. *Hum. Mol. Genet.* **21**, 2759–2767 (2012).
 160. M.A., W. *et al.* Compound Heterozygous Triadin Mutation Causing Cardiac Arrest in Two Siblings. *PACE - Pacing Clin. Electrophysiol.* **39**, 497–501 (2016).
 161. O’Callaghan, B. M. *et al.* A unique triadin exon deletion causing a null phenotype. *Hear. Case Reports* **4**, 514–518 (2018).
 162. Kurosaki, T., Popp, M. W. & Maquat, L. E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nature Reviews Molecular Cell Biology* (2019) doi:10.1038/s41580-019-0126-2.
 163. Evans, J. A. Old meets new: Identifying founder mutations in genetic disease. *CMAJ* **187**, 93–94 (2015).
 164. Manrai, A. K. *et al.* Genetic misdiagnoses and the potential for health disparities. *N. Engl. J. Med.* **375**, 655–665 (2016).
 165. Jones, E. G. *et al.* Analysis of enriched rare variants in JPH2-encoded junctophilin-2 among Greater Middle Eastern individuals reveals a novel homozygous variant associated with neonatal dilated cardiomyopathy. *Sci. Rep.* **9**, (2019).
 166. Yang, Y. *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA - J. Am. Med. Assoc.* **312**, 1870–1879 (2014).
 167. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* vol. 177 26–31 (2019).
 168. Mulder, N. *et al.* H3Africa: Current perspectives. *Pharmacogenomics and Personalized Medicine* vol. 11 59–66 (2018).
 169. A.H., B. & M.L., B. Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences of the United States of America* vol. 107 1779–1786 (2010).
 170. Jaafar, N. *et al.* Spectrum of Mutations in Hypertrophic Cardiomyopathy Genes among Tunisian Patients. *Genet. Test. Mol. Biomarkers* **20**, 674–679 (2016).
 171. Ho, C. Y. *et al.* Homozygous mutation in cardiac troponin T: Implications for hypertrophic cardiomyopathy. *Circulation* **102**, 1950–1955 (2000).
 172. Jarcho, J. A. *et al.* Mapping a Gene for Familial Hypertrophic Cardiomyopathy to Chromosome 14q1. *N. Engl. J. Med.* **321**, 1372–1378 (1989).
 173. Kimura, A. *et al.* Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. *Nat. Genet.* **16**, 379–382 (1997).
 174. Marian, A. J. Modifier genes for hypertrophic cardiomyopathy. *Current Opinion in Cardiology* vol. 17 242–252 (2002).
 175. Thomson, K. L. *et al.* Analysis of 51 proposed hypertrophic cardiomyopathy genes from genome sequencing data in sarcomere negative cases has negligible

- diagnostic yield. *Genet. Med.* **21**, 1576–1584 (2019).
176. Rehm, H. L. *et al.* ClinGen - The clinical genome resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).
 177. Strande, N. T. *et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100**, 895–906 (2017).
 178. Fananapazir, L. & Epstein, N. D. Genotype-phenotype correlations in hypertrophic cardiomyopathy: Insights provided by comparisons of kindreds with distinct and identical β -myosin heavy chain gene mutations. *Circulation* **89**, 22–32 (1994).
 179. Watkins, H. *et al.* Mutations in the Genes for Cardiac Troponin T and α -Tropomyosin in Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **332**, 1058–1064 (1995).
 180. Kokado, H. *et al.* Clinical features of hypertrophic cardiomyopathy caused by a Lys183 deletion mutation in the cardiac troponin I gene. *Circulation* **102**, 663–669 (2000).
 181. Burns, C., Bagnall, R., Lam, L., Semsarian, C. & Ingles, J. Multiple Gene Variants in Hypertrophic Cardiomyopathy in the Era of Next Generation Sequencing. *Hear. Lung Circ.* **26**, S307 (2017).
 182. Lechin, M. *et al.* Angiotensin-I converting enzyme genotypes and left ventricular hypertrophy in patients with hypertrophic cardiomyopathy. *Circulation* **92**, 1808–1812 (1995).
 183. Brugada, R. *et al.* Role of candidate modifier genes on the phenotypic expression of hypertrophy in patients with hypertrophic cardiomyopathy. *J. Investig. Med.* **45**, 542–551 (1997).
 184. Tesson, F. *et al.* The influence of the angiotensin I converting enzyme genotype in familial hypertrophic cardiomyopathy varies with the disease gene mutation. *J. Mol. Cell. Cardiol.* **29**, 831–838 (1997).
 185. Patel, R. *et al.* Variants of trophic factors and expression of cardiac hypertrophy in patients with hypertrophic cardiomyopathy. *J. Mol. Cell. Cardiol.* **32**, 2369–2377 (2000).
 186. Wooten, E. C. *et al.* Formin Homology 2 Domain Containing 3 (FHOD3) Variants Associated with Hypertrophic Cardiomyopathy. *Circ. Cardiovasc. Genet.* **6**, 10–18 (2013).
 187. Zhang, W., Song, M., Qu, J. & Liu, G. H. Epigenetic modifications in cardiovascular aging and diseases. *Circ. Res.* **123**, 773–786 (2018).
 188. Gomes, C. P. D. C. *et al.* Regulatory RNAs in Heart Failure. *Circulation* **141**, 313–328 (2020).
 189. Chen, S. *et al.* Cardiac miR-133a overexpression prevents early cardiac fibrosis in diabetes. *J. Cell. Mol. Med.* **18**, 415–421 (2014).
 190. Derda, A. A. *et al.* Blood-based microRNA signatures differentiate various forms of cardiac hypertrophy. *Int. J. Cardiol.* **1**, 115–122 (2015).
 191. Berezin, A. Epigenetics in heart failure phenotypes. *BBA Clin.* **6**, 31–37 (2016).
 192. Olivetto, I. *et al.* Gender-related differences in the clinical presentation and outcome of hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **46**, 480–487 (2005).

193. Olivotto, I. *et al.* Obesity and its association to phenotype and clinical course in hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **62**, 449–457 (2013).
194. Rehm, H. L. A new era in the interpretation of human genomic variation. *Genet. Med.* **19**, 1092–1095 (2017).
195. Gaba, P. *et al.* Implantable cardioverter-defibrillator explantation for overdiagnosed or overtreated congenital long QT syndrome. *Hear. Rhythm* **13**, 879–885 (2016).
196. Jarvik, G. P. & Browning, B. L. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am. J. Hum. Genet.* **98**, 1077–1081 (2016).
197. Harrison, S. M. *et al.* Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet. Med.* **19**, 1096–1104 (2017).
198. Nykamp, K. *et al.* Sherloc: A comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet. Med.* **19**, 1105–1117 (2017).
199. Norton, N. *et al.* Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ. Cardiovasc. Genet.* **5**, 167–174 (2012).
200. Kobayashi, Y. *et al.* Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med.* **9**, 1–14 (2017).
201. Duzkale, H. *et al.* A systematic approach to assessing the clinical significance of genetic variants. *Clin. Genet.* **84**, 453–463 (2013).
202. Kelly, M. A. *et al.* Adaptation and validation of the ACMG / AMP variant classification framework for MYH7 -associated inherited cardiomyopathies : recommendations by ClinGen ' s Inherited Cardiomyopathy Expert Panel. *Genet. Med.* **00**, 1–9 (2018).
203. Walsh, R. *et al.* Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases : the case of hypertrophic cardiomyopathy. *Genome Med.* **11**, 1–18 (2019).
204. Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* **100**, 267–280 (2017).
205. Nicora, G. *et al.* CardioVAL: An automatic implementation of ACMG-AMP variant interpretation guidelines in the diagnosis of cardiovascular diseases. *Hum. Mutat.* **39**, 1835–1846 (2018).
206. Whiffin, N. *et al.* CardioClassifier: disease- and gene-specific computational decision support for clinical genome interpretation. *Genet. Med.* **20**, 1246–1254 (2018).
207. Landry, L. G. & Rehm, H. L. Association of racial/ethnic categories with the ability of genetic tests to detect a cause of cardiomyopathy. *JAMA Cardiol.* **3**, 341–345 (2018).
208. Gomez, F., Hirbo, J. & Tishkoff, S. A. Genetic variation and adaptation in Africa: Implications for human evolution and disease. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).
209. Campbell, M. C. & Tishkoff, S. A. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 403–433 (2008).
210. Tishkoff, S. A. & Williams, S. M. Genetic analysis of African populations: Human evolution and complex disease. *Nature Reviews Genetics* vol. 3 611–621 (2002).

211. Ntusi, N. A. B., Shaboodien, G., Badri, M., Gumedze, F. & Mayosi, B. M. Clinical features, spectrum of causal genetic mutations and outcome of hypertrophic cardiomyopathy in South Africans. *Cardiovasc. J. Afr.* **27**, 152–158 (2016).
212. Moolman, J. C., Brink, P. A. & Corfield, V. A. Identification of a new missense mutation at Arg403, a CpG mutation hotspot, in exon 13 of the β -myosin heavy chain gene in hypertrophic cardiomyopathy. *Hum. Mol. Genet.* **2**, 1731–1732 (1993).
213. Moolman, J. C., Brink, P. A. & Corfield, V. A. Identification of a novel Ala797Thr mutation in exon 21 of the β -myosin heavy chain gene in hypertrophic cardiomyopathy. *Hum. Mutat.* **6**, 197–198 (1995).
214. Posen, B. M., Moolman, J. C., Corfield, V. A. & Brink, P. A. Clinical and prognostic evaluation of familial hypertrophic cardiomyopathy in two South African families with different cardiac I8 myosin heavy chain gene mutations. *Heart* **74**, 40–46 (1995).
215. Moolman-Smook, J. C., De Lange, W. J., Bruwer, E. C. D., Brink, P. A. & Corfield, V. A. The origins of hypertrophic cardiomyopathy-causing mutations in two South African subpopulations: A unique profile of both independent and founder events. *Am. J. Hum. Genet.* **65**, 1308–1320 (1999).
216. Moolman-Smook, J., De Lange, W., Corfield, V. & Brink, P. Expression of HCM causing mutations: Lessons learnt from genotype-phenotype studies of the South African founder MYH7 A797T mutation [4]. *Journal of Medical Genetics* vol. 37 951–956 (2000).
217. Moolman, J. C. *et al.* Sudden death due to troponin T mutations. *J. Am. Coll. Cardiol.* **29**, 549–555 (1997).
218. Kassem, H. S. *et al.* Early results of sarcomeric gene screening from the egyptian national BA-HCM program. *J. Cardiovasc. Transl. Res.* **6**, 65–80 (2013).
219. Lahrouchi, N. *et al.* Exome sequencing identifies primary carnitine deficiency in a family with cardiomyopathy and sudden death. *Eur. J. Hum. Genet.* **25**, 783–787 (2017).
220. Abegaz, B. The impact of echocardiography in the diagnosis of hypertrophic cardiomyopathy. *East Afr. Med. J.* **67**, 556–567 (1990).
221. Amoah, A. G. B. & Kallen, C. Aetiology of heart failure as seen from a national cardiac referral centre in africa. *Cardiology* **93**, 11–18 (2000).
222. Shaboodien, G. *et al.* Genetics of inherited cardiomyopathies in Africa. *Cardiovasc. Diagn. Ther.* (2019) doi:<http://dx.doi.org/10.21037/cdt.2019.10.03>.
223. Guo, M. H., Plummer, L., Chan, Y. M., Hirschhorn, J. N. & Lippincott, M. F. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *Am. J. Hum. Genet.* **103**, 522–534 (2018).
224. Altman, D. G. Practical statistics for medical research. Chapman and Hall. *Stat. Med.* 336–358 (1991) doi:[10.1002/sim.4780101015](https://doi.org/10.1002/sim.4780101015).
225. Hildebrandt, M., Bender, R., Gehrman, U. & Blettner, M. Calculating confidence intervals for impact numbers. *BMC Med. Res. Methodol.* **6**, (2006).
226. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

227. Berge, K. E. & Leren, T. P. Genetics of hypertrophic cardiomyopathy in Norway. *Clin. Genet.* **86**, 355–360 (2014).
228. Koshy, R., Ranawat, A. & Scaria, V. Al mena: A comprehensive resource of human genetic variants integrating genomes and exomes from Arab, Middle Eastern and North African populations. *J. Hum. Genet.* (2017) doi:10.1038/jhg.2017.67.
229. Olivetto, I. *et al.* Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. *Mayo Clin. Proc.* **83**, 630–638 (2008).
230. Marsiglia, J. D. C. *et al.* Screening of MYH7, MYBPC3, and TNNT2 genes in Brazilian patients with hypertrophic cardiomyopathy. *Am. Heart J.* **166**, 775–782 (2013).
231. Millat, G. *et al.* Prevalence and spectrum of mutations in a cohort of 192 unrelated patients with hypertrophic cardiomyopathy. *Eur. J. Med. Genet.* **53**, 261–267 (2010).
232. Coto, E. *et al.* Resequencing the whole MYH7 gene (including the intronic, promoter, and 3' UTR sequences) in hypertrophic cardiomyopathy. *J. Mol. Diagnostics* **14**, 518–524 (2012).
233. Olivetto, I. *et al.* Microvascular function is selectively impaired in patients with hypertrophic cardiomyopathy and sarcomere myofilament gene mutations. *J. Am. Coll. Cardiol.* **58**, 839–848 (2011).
234. Lipari, M. *et al.* Identification of a variant hotspot in MYBPC3 and of a novel CSRP3 autosomal recessive alteration in a cohort of Polish cases affected by hypertrophic cardiomyopathy. *Polish Arch. Intern. Med.* (2020) doi:10.20452/pamw.15130.
235. Janin, A., Bessière, F., Chauveau, S., Chevalier, P. & Millat, G. First identification of homozygous truncating CSRP3 variants in two unrelated cases with hypertrophic cardiomyopathy. *Gene* **676**, 110–116 (2018).
236. Niimura, H. *et al.* Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy. *N. Engl. J. Med.* **338**, 1248–1257 (1998).
237. Maron, B. J. *et al.* Development of left ventricular hypertrophy in adults with hypertrophic cardiomyopathy caused by cardiac myosin-binding protein C gene mutations. *J. Am. Coll. Cardiol.* **38**, 315–321 (2001).
238. Li, Q. *et al.* Genotype-positive status in patients with hypertrophic cardiomyopathy is associated with higher rates of heart failure events. *Circ. Cardiovasc. Genet.* **7**, 416–422 (2014).
239. Weissler-Snir, A. *et al.* 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* **10**, (2017).
240. Bos, J. M. *et al.* Characterization of a phenotype-based genetic test prediction score for unrelated patients with hypertrophic cardiomyopathy. *Mayo Clin. Proc.* **89**, 727–737 (2014).
241. Ingles, J. *et al.* Clinical predictors of genetic testing outcomes in hypertrophic cardiomyopathy. *Genet. Med.* **15**, 972–977 (2013).
242. Montag, J. *et al.* Intrinsic MYH7 expression regulation contributes to tissue level

- allelic imbalance in hypertrophic cardiomyopathy. *J. Muscle Res. Cell Motil.* **38**, 291–302 (2017).
243. Wijnker, P. J. M. *et al.* Comparison of the effects of a truncating and a missense MYBPC3 mutation on contractile parameters of engineered heart tissue. *J. Mol. Cell. Cardiol.* **97**, 82–92 (2016).
 244. Van Dijk, S. J. *et al.* Cardiac myosin-binding protein C mutations and hypertrophic cardiomyopathy haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction. *Circulation* **119**, 1473–1483 (2009).
 245. Barefield, D. *et al.* Haploinsufficiency of MYBPC3 exacerbates the development of hypertrophic cardiomyopathy in heterozygous mice. *J. Mol. Cell. Cardiol.* **79**, 234–243 (2015).
 246. Marston, S., Copeland, O., Gehmlich, K., Schlossarek, S. & Carrier, L. How do MYBPC3 mutations cause hypertrophic cardiomyopathy? *Journal of Muscle Research and Cell Motility* vol. 33 75–80 (2012).
 247. Coppini, R. *et al.* Clinical phenotype and outcome of hypertrophic cardiomyopathy associated with thin-filament gene mutations. *J. Am. Coll. Cardiol.* **64**, 2589–2600 (2014).
 248. Abouelhoda, M. *et al.* Revisiting the morbid genome of Mendelian disorders. *Genome Biol.* **17**, 235 (2016).
 249. Rayment, I. *et al.* Three-dimensional structure of myosin subfragment-1: A molecular motor. *Science* (80-.). **261**, 50–58 (1993).
 250. Trybus, K. M. Role of myosin light chains. *Journal of Muscle Research and Cell Motility* vol. 15 587–594 (1994).
 251. Morano, I. Tuning the human heart molecular motors by myosin light chains. *Journal of Molecular Medicine* vol. 77 544–555 (1999).
 252. Harris, S. P., Lyons, R. G. & Bezold, K. L. In the thick of it: HCM-causing mutations in myosin binding proteins of the thick filament. *Circ. Res.* **108**, 751–764 (2011).
 253. Kabaeva, Z. *et al.* Systematic analysis of the regulatory and essential myosin light chain genes: Genetic variants and mutations in hypertrophic cardiomyopathy. *Eur. J. Hum. Genet.* **10**, 741–748 (2002).
 254. Wang, Y. *et al.* Autosomal Recessive Transmission of MYBPC3 Mutation Results in Malignant Phenotype of Hypertrophic Cardiomyopathy. *PLoS One* **8**, (2013).
 255. Flavigny, J. *et al.* Identification of two novel mutations in the ventricular regulatory myosin light chain gene (MYL2) associated with familial and classical forms of hypertrophic cardiomyopathy. *J. Mol. Med.* **76**, 208–214 (1998).
 256. Lee, W. H. *et al.* Different expressivity of a ventricular essential myosin light chain gene Ala57Gly mutation in familial hypertrophic cardiomyopathy. *Am. Heart J.* **141**, 184–189 (2001).
 257. Andersen, P. S. *et al.* Myosin light chain mutations in familial hypertrophic cardiomyopathy: phenotypic presentation and frequency in Danish and South African populations. *Journal of medical genetics* vol. 38 (2001).
 258. Ceyhan-Birsoy, O. *et al.* Next generation sequencing-based copy number analysis reveals low prevalence of deletions and duplications in 46 genes associated with genetic cardiomyopathies. *Mol. Genet. Genomic Med.* **4**, 143–151 (2016).

259. Mademont-Soler, I. *et al.* Additional value of screening for minor genes and copy number variants in hypertrophic cardiomyopathy. *PLoS One* **12**, (2017).
260. Kostis, R. I. & Panagiotakos, D. B. The epidemic of obesity in children and adolescents in the world. *Central European Journal of Public Health* vol. 14 151–159 (2006).
261. Musaiger, A. O. Overweight and obesity in Eastern Mediterranean Region: Prevalence and possible causes. *Journal of Obesity* vol. 2011 (2011).
262. Bagnall, R. D. *et al.* Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* **72**, 419–429 (2018).
263. Jipin Das, K., Ingles, J., Bagnall, R. D. & Semsarian, C. Determining pathogenicity of genetic variants in hypertrophic cardiomyopathy: Importance of periodic reassessment. *Genet. Med.* **16**, 286–293 (2014).
264. Ortiz, M. F. *et al.* A homozygous MYBPC3 gene mutation associated with a severe phenotype and a high risk of sudden death in a family with hypertrophic cardiomyopathy. *Rev. Esp. Cardiol.* **62**, 572–575 (2009).
265. Zahka, K. *et al.* Homozygous mutation of MYBPC3 associated with severe infantile hypertrophic cardiomyopathy at high frequency among the Amish. *Heart* **94**, 1326–1330 (2008).
266. Gray, B., Yeates, L., Medi, C., Ingles, J. & Semsarian, C. Homozygous mutation in the cardiac troponin i gene: Clinical heterogeneity in hypertrophic cardiomyopathy. *Int. J. Cardiol.* **168**, 1530–1531 (2013).
267. Wessels, M. W. *et al.* Compound heterozygous or homozygous truncating MYBPC3 mutations cause lethal cardiomyopathy with features of noncompaction and septal defects. *Eur. J. Hum. Genet.* **23**, 922–928 (2015).
268. Xin, B., Puffenberger, E., Tumbush, J., Bockoven, J. R. & Wang, H. Homozygosity for a novel splice site mutation in the cardiac myosin-binding protein C gene causes severe neonatal hypertrophic cardiomyopathy. *Am. J. Med. Genet. Part A* **143**, 2662–2667 (2007).
269. Richard, P. *et al.* Hypertrophic cardiomyopathy: Distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* **107**, 2227–2232 (2003).
270. Gómez, J. *et al.* Screening of the Filamin C Gene in a Large Cohort of Hypertrophic Cardiomyopathy Patients. *Circ. Cardiovasc. Genet.* **10**, (2017).
271. Valdés-Mas, R. *et al.* Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. *Nat. Commun.* **5**, (2014).
272. Nouredin, R. A. *et al.* The diagnosis of hypertrophic cardiomyopathy by cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance* vol. 14 (2012).
273. Fernandes, E., Camargo, G. C., Derenne, M. E., Rothstein, T. & Gottlieb, I. Exuberant Pattern of Late Gadolinium Enhancement in Hypertrophic Cardiomyopathy. *Arq. Bras. Cardiol.* (2013) doi:10.5935/abc.20130195.
274. Walsh, R., Rutland, C., Thomas, R. & Loughna, S. Cardiomyopathy: A systematic review of disease-causing mutations in myosin heavy chain 7 and their phenotypic

- manifestations. *Cardiology* (2009) doi:10.1159/000252808.
275. Fiorillo, C. *et al.* MYH7-related myopathies: Clinical, histopathological and imaging findings in a cohort of Italian patients. *Orphanet J. Rare Dis.* **11**, 1–14 (2016).
 276. Tajsharghi, H. *et al.* Myosin storage myopathy associated with a heterozygous missense mutation in MYH7. *Ann. Neurol.* **54**, 494–500 (2003).
 277. Toepfer, C. N., Garfinkel, A. C. & Venturini, G. Myosin Sequestration Regulates Sarcomere Function , Cardiomyocyte Energetics , and Metabolism , Informing the Pathogenesis of Hypertrophic Cardiomyopathy. *Circulation* **141**, 1–29 (2020).
 278. Alamo, L., Ware, J. S., Pinto, A. & Gillilan, R. E. Effects of myosin variants on interacting- heads motif explain distinct hypertrophic and dilated cardiomyopathy phenotypes. *Elife* 1–31 (2017) doi:10.7554/eLife.24634.
 279. Homburger, J. R. *et al.* Multidimensional structure-function relationships in human β -cardiac myosin from population-scale genetic variation. *PNAS* **113**, 6701–6706 (2016).
 280. Navarro-lópez, F. Hypertrophic Cardiomyopathy: Never-Ending Complexity. *Rev Esp Cardiol.* **59**, 994–996 (2006).
 281. Korkmaz, E. N. *et al.* A composite approach towards a complete model of the myosin rod. *Proteins Struct. Funct. Bioinforma.* **84**, 172–189 (2016).
 282. Colegrave, M. & Peckham, M. Structural implications of β -cardiac myosin heavy chain mutations in human disease. *Anat. Rec.* (2014) doi:10.1002/ar.22973.
 283. Roberts, A. M. *et al.* Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* **7**, 270ra6–270ra6 (2015).
 284. Taylor, K. C. *et al.* Skip residues modulate the structural properties of the myosin rod and guide thick filament assembly. *Proc. Natl. Acad. Sci.* **112**, E3806–E3815 (2015).
 285. Moore, J. R., Leinwand, L. & Warshaw, D. M. Invited Review: The Myosins: Exploration of the Development of Our Current Understanding of These Mutations in the Motor. *Circ. Res.* **111**, 375–385 (2014).
 286. McLachlan, A. D. & Karn, J. Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle. *Nature* **299**, 226–231 (1982).
 287. Offer, G. Skip residues correlate with bends in the myosin tail. *J. Mol. Biol.* **216**, 213–218 (1990).
 288. Viswanathan, S. K. *et al.* Hypertrophic cardiomyopathy clinical phenotype is independent of gene mutation and mutation dosage. *PLoS One* **12**, (2017).
 289. Spudich, J. A. Hypertrophic and dilated cardiomyopathy: Four decades of basic research on muscle lead to potential therapeutic approaches to these devastating genetic diseases. *Biophys. J.* **106**, 1236–1249 (2014).
 290. Hooijman, P., Stewart, M. A. & Cooke, R. A new state of cardiac myosin with very slow ATP turnover: A potential cardioprotective mechanism in the heart. *Biophys. J.* (2011) doi:10.1016/j.bpj.2011.02.061.
 291. Nag, S. *et al.* The myosin mesa and the basis of hypercontractility caused by hypertrophic cardiomyopathy mutations. *Nat. Struct. Mol. Biol.* (2017)

- doi:10.1038/nsmb.3408.
292. Stewart, M. A., Franks-Skiba, K., Chen, S. & Cooke, R. Myosin ATP turnover rate is a mechanism involved in thermogenesis in resting skeletal muscle fibers. *Proc. Natl. Acad. Sci.* **107**, 430–435 (2010).
 293. McNamara, J. W. *et al.* Ablation of cardiac myosin binding protein-C disrupts the super-relaxed state of myosin in murine cardiomyocytes. *J. Mol. Cell. Cardiol.* **94**, 65–71 (2016).
 294. Linde, I. H. M. Van Der, Ruivenkamp, C. & Jongbloed, J. D. H. A Dutch MYH7 founder mutation, p.(Asn1918Lys), is associated with early onset cardiomyopathy and congenital heart defects. *Netherlands Hear. J.* **25**, 675–681 (2017).
 295. Tyska, M. J. *et al.* Single-molecule mechanics of R403Q cardiac myosin isolated from the mouse model of familial hypertrophic cardiomyopathy. *Circ. Res.* (2000) doi:10.1161/01.RES.86.7.737.
 296. Woo, A. *et al.* Mutations of the β myosin heavy chain gene in hypertrophic cardiomyopathy: Critical functional sites determine prognosis. *Heart* (2003) doi:10.1136/jmg.40.11.853.
 297. Anan, R. *et al.* Prognostic implications of novel β cardiac myosin heavy chain gene mutations that cause familial hypertrophic cardiomyopathy. *J. Clin. Invest.* (1994) doi:10.1172/JCI116957.
 298. Roopnarine, O. & Leinwand, L. A. Functional analysis of myosin mutations that cause familial hypertrophic cardiomyopathy. *Biophys. J.* (1998) doi:10.1016/S0006-3495(98)77743-4.
 299. Ng, P. C. & Henikoff, S. Predicting the Effects of Amino Acid Substitutions on Protein Function. *Annu. Rev. Genomics Hum. Genet.* (2006) doi:10.1146/annurev.genom.7.080505.115630.
 300. Miroshnichenko, N. S., Balanuk, I. V. & Nozdrenko, D. N. Packing of myosin molecules in muscle thick filaments. *Cell Biol. Int.* (2000) doi:10.1006/cbir.1999.0514.
 301. McNally, E. M. β -Myosin Heavy Chain Gene Mutations in Familial Hypertrophic Cardiomyopathy: The Usual Suspect? *Circ. Res.* **90**, 246–247 (2002).
 302. Wolny, M. *et al.* Cardiomyopathy mutations in the tail of β -cardiac myosin modify the coiled-coil structure and affect integration into thick filaments in muscle sarcomeres in adult cardiomyocytes. *J. Biol. Chem.* **288**, 31952–31962 (2013).
 303. Armel, T. Z. & Leinwand, L. A. A Mutation in the β -Myosin Rod Associated With Hypertrophic Cardiomyopathy Has an Unexpected Molecular Phenotype. *Biochem. Biophys. Res. Commun.* **391**, 1–13 (2011).
 304. Obermann, W. M. J., Gautel, M., Weber, K. & Fürst, D. O. Molecular structure of the sarcomeric M band: Mapping of titin and myosin binding domains in myomesin and the identification of a potential regulatory phosphorylation site in myomesin. *EMBO J.* (1997) doi:10.1093/emboj/16.2.211.
 305. Flashman, E., Watkins, H. & Redwood, C. Localization of the binding site of the C-terminal domain of cardiac myosin-binding protein-C on the myosin rod. *Biochem. J.* (2007) doi:10.1042/BJ20060500.
 306. Robinson, P. *et al.* Alterations in thin filament regulation induced by a human

- cardiac troponin T mutant that causes dilated cardiomyopathy are distinct from those induced by troponin T mutants that cause hypertrophic cardiomyopathy. *J. Biol. Chem.* **277**, 40710–40716 (2002).
307. Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med.* **19**, 1151–1158 (2017).
 308. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell and Bioscience* (2017) doi:10.1186/s13578-017-0153-7.
 309. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural and Molecular Biology* (2009) doi:10.1038/nsmb.1550.
 310. Carrier, L., Schlossarek, S., Willis, M. S. & Eschenhagen, T. The ubiquitin-proteasome system and nonsense-mediated mRNA decay in hypertrophic cardiomyopathy. *Cardiovascular Research* (2010) doi:10.1093/cvr/cvp247.
 311. Kurosaki, T. & Maquat, L. E. Nonsense-mediated mRNA decay in humans at a glance. *J. Cell Sci.* **129**, 461–467 (2016).
 312. Minikel, E. V. *et al.* Quantifying prion disease penetrance using large population control cohorts. *Sci. Transl. Med.* **8**, (2016).
 313. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu170.
 314. Chen, S. *et al.* AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics* (2017) doi:10.1186/s12859-017-1469-3.
 315. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) doi:10.1093/bioinformatics/bts635.
 316. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btp352.
 317. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* (2011) doi:10.1038/nbt.1754.
 318. Duggirala, R., Almasy, L., Williams-Blangero, S., Paul, S. F. D. & Koe, C. *Genome mapping and genomics in human and non-human primates. Genome Mapping and Genomics in Human and Non-Human Primates* (2015). doi:10.1007/978-3-662-46306-2.
 319. Rivas, M. A. *et al.* Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* (80-.). (2015) doi:10.1126/science.1261877.
 320. Kassem, H. S. *et al.* A comparative study of mutation screening of sarcomeric genes (MYBPC3, MYH7, TNNT2) using single gene approach versus targeted gene panel next generation sequencing in a cohort of HCM patients in Egypt. *Egypt. J. Med. Hum. Genet.* **18**, 381–387 (2017).
 321. Mogensen, J. *et al.* Comparison of fluorescent SSCP and denaturing HPLC analysis with direct sequencing for mutation screening in hypertrophic cardiomyopathy. *Journal of medical genetics* 1–5 (2003) doi:10.1136/jmg.40.5.e59.
 322. Marian, A. J. & Braunwald, E. Hypertrophic cardiomyopathy: Genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ. Res.* **121**, 749–770 (2017).
 323. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse

- Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
324. Mosqueira, D., Smith, J. G. W., Bhagwan, J. R. & Denning, C. Modeling Hypertrophic Cardiomyopathy: Mechanistic Insights and Pharmacological Intervention. *Trends in Molecular Medicine* (2019) doi:10.1016/j.molmed.2019.06.005.
 325. Ho, C. Y. *et al.* Assessment of diastolic function with Doppler tissue imaging to predict genotype in preclinical hypertrophic cardiomyopathy. *Circulation* **105**, 2992–2997 (2002).
 326. Ben Jehuda, R., Shemer, Y. & Binah, O. Genome Editing in Induced Pluripotent Stem Cells using CRISPR/Cas9. *Stem Cell Reviews and Reports* 323–336 (2018) doi:10.1007/s12015-018-9811-3.
 327. Bassett, A. R. Editing the genome of hiPSC with CRISPR/Cas9: disease models. *Mammalian Genome* 348–364 (2017) doi:10.1007/s00335-017-9684-9.
 328. Vikstrom, K. L., Factor, S. M. & Leinwand, L. A. Mice expressing mutant myosin heavy chains are a model for familial hypertrophic cardiomyopathy. *Mol. Med.* **2**, 556–567 (1996).
 329. Geisterfer-Lowrance, A. A. T. *et al.* A Mouse Model of Familial Hypertrophic Cardiomyopathy. *Science* (80-.). **272**, 731–734 (1996).
 330. Lowey, S. *et al.* Hypertrophic cardiomyopathy R403Q mutation in rabbit β -myosin reduces contractile function at the molecular and myofibrillar levels. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 11238–11243 (2018).
 331. Marian, A. J. *et al.* A transgenic rabbit model for human hypertrophic cardiomyopathy. *J. Clin. Invest.* **104**, 1683–1692 (1999).
 332. Yang, J., Shih, Y. huan & Xu, X. Understanding cardiac sarcomere assembly with zebrafish genetics. *Anat. Rec.* **297**, 1681–1693 (2014).
 333. Almomani, R. *et al.* Biallelic Truncating Mutations in ALPK3 Cause Severe Pediatric Cardiomyopathy. *J. Am. Coll. Cardiol.* **67**, 515–525 (2016).
 334. Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat. Commun.* **11**, (2020).
 335. De La Vega, F. M. & Bustamante, C. D. Polygenic risk scores: A biased prediction? *Genome Med.* **10**, (2018).
 336. Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J. & Lachance, J. Genetic disease risks can be misestimated across global populations. *Genome Biol.* **19**, (2018).

7 APPENDICES

7.1 List of rare variants identified in AhcHCM cohort

The table below list all rare variants (i.e. gnomAD $FAF_{popmax} \leq 4 \times 10^{-5}$) in HCM-associated genes identified in 514 AhcHCM probands analysed in this study. All variants were classified according to the ACMG/AMP guidelines implemented in the webtool CardioClassifier and were then manually curated using ClinVar and PubMed resources for evidence regarding segregation and *de novo* inheritance (ACMG/AMP rules PP1 and PM2/PM6, respectively). **Activated ACMG/AMP rules:** (p) = supporting, (m) = moderate, (s) = strong.

Gene	cDNA Variant	Protein Variant	Variant Type	gnomAD FAF popmax	AhcHCM count (n=514)	ACMG classification	Activated ACMG/AMP Rules	Curation Evidence
ACTN2	c.265C>T	p.Arg89Trp	missense	3.42E-05	1	VUS	PM2	
ACTN2	c.892C>T	p.Arg298Cys	missense	0	2	VUS	PM2	
ACTN2	c.1266G>C	p.Gln422His	missense	0	1	VUS	PM2	
ACTN2	c.2555G>A	p.Arg852Gln	missense	1.12E-05	1	VUS	PM2	
ACTN2	c.2479A>G	p.Thr827Ala	missense	0	1	VUS	PM2	
ACTN2	c.745G>A	p.Val249Ile	missense	1.12E-05	2	VUS	PM2	
ACTN2	c.829C>G	p.Gln277Glu	missense	0	2	VUS	PM2	
ACTN2	c.721G>A	p.Asp241Asn	missense	0	1	VUS	PM2	
ACTN2	c.2006G>A	p.Gly669Glu	missense	0	1	VUS	PM2	
ACTN2	c.574C>T	p.Arg192Ter	stop gained	0	1	VUS	PM2	
CACNA1C	c.82G>A	p.Ala28Thr	missense	2.00E-05	1	VUS	PM2	
CACNA1C	c.143C>G	p.Ala48Gly	missense	0	1	VUS	PM2	
CACNA1C	c.169G>A	p.Asp57Asn	missense	0	1	VUS	PM2	
CACNA1C	c.1519C>T	p.Arg507Cys	missense	0	2	VUS	PM2	
CACNA1C	c.1552C>T	p.Arg518Cys	missense	0	1	VUS	PM2	
CACNA1C	c.2381C>T	p.Pro794Leu	missense	0	2	VUS	PM2	
CACNA1C	c.2486A>G	p.Asn829Ser	missense	4.00E-05	1	VUS	PM2	
CACNA1C	c.5188A>G	p.Ile1730Val	missense	0	1	VUS	PM2	
CACNA1C	c.5407C>T	p.Arg1803Trp	missense	3.00E-05	1	VUS	PM2	
CACNA1C	c.5917C>T	p.Arg1973Trp	missense	0	2	VUS	PM2	
CACNA1C	c.6001G>A	p.Gly2001Ser	missense	3.50E-05	2	VUS	PM2	
CACNA1C	c.6028A>G	p.Arg2010Gly	missense	0	1	VUS	PM2	
CACNA1C	c.6074C>T	p.Pro2025Leu	missense	0	1	VUS	PM2	
CACNA1C	c.6203A>C	p.Asp2068Ala	missense	0	2	VUS	PM2	
CACNA1C	c.6329G>A	p.Gly2110Asp	missense	0	1	VUS	PM2	
CACNA1C	c.6382C>G	p.Leu2128Val	missense	0	1	VUS	PM2	
CACNA1C	c.6388G>A	p.Asp2130Asn	missense	2.30E-05	2	VUS	PM2	
CSRP3	c.437G>A	p.Arg146His	missense	1.12E-05	1	VUS	PM2	
CSRP3	c.436C>T	p.Arg146Cys	missense	2.92E-06	2	VUS	PM2	

<i>CSRP3</i>	c.415-1G>C	-	splice acceptor variant	0	1	VUS	PM2	
<i>CSRP3</i>	c.414+1G>T	-	splice donor variant	0	4	VUS	PM2	
<i>CSRP3</i>	c.385G>T	p.Ala129Ser	missense	0	1	VUS	PM2	
<i>CSRP3</i>	c.365G>A	p.Arg122Gln	missense	2.92E-06	4	VUS	PM2	
<i>CSRP3</i>	c.272A>T	p.Gln91Leu	missense	0	2	VUS	PM2	
<i>DES</i>	c.44G>A	p.Arg15His	missense	0	1	VUS	PM2	
<i>DES</i>	c.197T>C	p.Leu66Pro	missense	0	2	VUS	PM2	
<i>DES</i>	c.596A>G	p.Gln199Arg	missense	0	1	VUS	PM2	
<i>DES</i>	c.1103C>T	p.Ala368Val	missense	0	2	VUS	PM2	
<i>DES</i>	c.1250A>G	p.Asn417Ser	missense	2.10E-05	2	VUS	PM2	
<i>GLA</i>	c.1112C>T	p.Ser371Phe	missense	0	1	VUS	PM2	
<i>GLA</i>	c.74A>T	p.Asp25Val	missense	0	1	VUS	PM2	
<i>JPH2</i>	c.1987G>C	p.Glu663Gln	missense	0	1	VUS	PM2	
<i>JPH2</i>	c.1886C>T	p.Pro629Leu	missense	0	1	VUS	PM2	
<i>JPH2</i>	c.1732C>A	p.Pro578Thr	missense	3.83E-05	1	VUS	PM2	
<i>JPH2</i>	c.1724C>T	p.Pro575Leu	missense	0	1	VUS	PM2	
<i>JPH2</i>	c.1435C>T	p.Arg479Trp	missense	2.53E-05	1	VUS	PM2	
<i>JPH2</i>	c.1014G>C	p.Glu338Asp	missense	0	1	VUS	PM2	
<i>JPH2</i>	c.713C>T	p.Ser238Phe	missense	0	1	VUS	PM2	
<i>JPH2</i>	c.511_516dup	p.Ser171_Asn172dup	inframe insertion	0	1	VUS	PM2	
<i>JPH2</i>	c.481A>G	p.Thr161Ala	missense	0	3	VUS	PM2	
<i>JPH2</i>	c.296G>A	p.Ser99Asn	missense	0	2	VUS	PM2	
<i>JPH2</i>	c.59G>C	p.Gly20Ala	missense	0	1	VUS	PM2	
<i>LAMP2</i>	c.183+1G>A	-	splice donor variant	0	1	LP	PM2,PVS1	
<i>MYBPC3</i>	c.3814+2T>C	-	splice donor variant	0	1	LP	PM2,PVS1	
<i>MYBPC3</i>	c.3627+2delT	-	splice donor variant	0	5	LP	PM2,PVS1	
<i>MYBPC3</i>	c.3617G>T	p.Gly1206Val	missense	0	3	VUS	PM2,PP3	
<i>MYBPC3</i>	c.3370T>C	p.Cys1124Arg	missense	0	1	VUS	PM2,PP3	
<i>MYBPC3</i>	c.3293G>A	p.Trp1098Ter	nonsense	0	6	LP	PM2,PVS1	

MYBPC3	c.2977C>T	p.Leu993Phe	missense	0	1	VUS	PM2	
MYBPC3	c.2950C>T	p.Gln984Ter	nonsense	0	1	LP	PM2,PVS1	
MYBPC3	c.2908C>T	p.Arg970Trp	missense	0	3	VUS	PM2,PP3	
MYBPC3	c.2827C>T	p.Arg943Ter	nonsense	1.10E-05	1	P	PP1(p),PM2,PS4,PVS1	ClinVar:S CV00005 9178.4
MYBPC3	c.2738G>C	p.Cys913Ser	missense	0	1	VUS	PM2,BP4	
MYBPC3	c.2534G>A	p.Arg845His	missense	1.90E-05	1	VUS	PM2	
MYBPC3	c.2470G>A	p.Asp824Asn	missense	0	1	VUS	PM2	
MYBPC3	c.2458C>G	p.Arg820Gly	missense	0	1	VUS	PM2,PP3	
MYBPC3	c.2441_2443del AGA	p.Lys814del	inframe indel	0	1	VUS	PM2,PM4	
MYBPC3	c.2309_2314del ACGTGC	p.Asp770_Pro772de linsAla	inframe indel	0	3	VUS	PM2,PM4	
MYBPC3	c.2308G>A	p.Asp770Asn	missense	2.30E-05	2	LP	PM2,PS4	
MYBPC3	c.2171G>A	p.Arg724Gln	missense	7.00E-06	1	VUS	PM2	
MYBPC3	c.2005C>A	p.Leu669Met	missense	0	2	VUS	PM2,PP3	
MYBPC3	c.1928-2A>G	-	splice acceptor variant	0	4	P	PM2,PS4,P VS1, PP1(s)	PMID:749 3026
MYBPC3	c.1773G>T	p.Lys591Asn	missense	0	1	VUS	PM2	
MYBPC3	c.1591G>T	p.Gly531Trp	missense	0	1	VUS	PM2,PP3	
MYBPC3	c.1573T>C	p.Tyr525His	missense	0	1	VUS	PM2,PP3	
MYBPC3	c.1558G>T	p.Glu520Ter	nonsense	0	2	LP	PM2,PVS1	
MYBPC3	c.1516delG	p.Asp506ThrfsTer7	frameshift	0	5	LP	PM2 , PVS1	
MYBPC3	c.1483C>G	p.Arg495Gly	missense	0	1	P	PP3,PM2,P M5, PS4	
MYBPC3	c.1444delG	p.Ala482ArgfsTer6	frameshift	0	1	LP	PM2,PVS1	
MYBPC3	c.1418_1423del TTGAGT	p.Phe473_Glu474d el	inframe indel	0	1	VUS	PM2,PM4	
MYBPC3	c.1409G>A	p.Arg470Gln	missense	0	1	VUS	PM2	
MYBPC3	c.1308C>A	p.Cys436Ter	nonsense	0	2	LP	PM2,PVS1	
MYBPC3	c.1227-2A>G	-	splice acceptor variant	0	2	LP	PM2,PVS1	

MYBPC3	c.1223+1G>A	-	splice donor variant	2.40E-05	1	LP	PM2,PVS1	
MYBPC3	c.1210C>T	p.Gln404Ter	nonsense	0	1	LP	PM2,PVS1	
MYBPC3	c.1003C>A	p.Arg335Ser	missense	0	1	VUS	PM2	
MYBPC3	c.999C>A	p.Tyr333Ter	nonsense	0	2	LP	PM2,PVS1	
MYBPC3	c.887_891delGCTCA	p.Ser296ThrfsTer35	frameshift	0	2	LP	PM2,PVS1	
MYBPC3	c.852-2A>T	-	splice acceptor variant	0	1	LP	PM2,PVS1	
MYBPC3	c.836G>C	p.Gly279Ala	missense	4.00E-06	2	VUS	BP4,PM2	
MYBPC3	c.772G>A	p.Glu258Lys	missense	1.20E-05	4	LP	PM2,PS4,P1(p)	PMID:9562578
MYBPC3	c.588G>A	p.Trp196Ter	nonsense	0	1	LP	PM2,PVS1	
MYBPC3	c.587G>A	p.Trp196Ter	nonsense	0	1	LP	PM2,PVS1	
MYBPC3	c.534_541delG GCCGGCG	p.Ala179GlnfsTer59	frameshift	0	7	LP	PM2,PVS1	
MYBPC3	c.416C>G	p.Ser139Ter	nonsense	0	1	LP	PM2,PVS1	
MYBPC3	c.305C>T	p.Pro102Leu	missense	3.70E-05	2	VUS	BP4,PM2	
MYBPC3	c.130C>T	p.Arg44Cys	missense	0	1	VUS	PM2	
MYBPC3	c.89_90delAG	p.Glu30GlyfsTer18	frameshift	0	1	VUS	PM2,PVS1	
MYH7	c.5769delG	p.Ser1924AlafsTer9	frameshift	0	17	VUS	PM2	
MYH7	c.5704G>A	p.Glu1902Lys	missense	0	1	VUS	PM2	
MYH7	c.5135G>A	p.Arg1712Gln	missense	1.70E-05	1	LP	PP3,PP1(p), PM2,PS4	ClinVar: SCV000059603.5 & SCV000208619.11
MYH7	c.5122G>A	p.Glu1708Lys	missense	0	1	VUS	PP3,PM2	
MYH7	c.4258C>T	p.Arg1420Trp	missense	3.00E-06	3	LP	PP3,PM2,PS4	
MYH7	c.4145G>A	p.Arg1382Gln	missense	0	1	VUS	PM2	
MYH7	c.4036C>G	p.Gln1346Glu	missense	0	1	VUS	PP3,PM2	
MYH7	c.3622G>C	p.Asp1208His	missense	0	4	VUS	PP3,PM2	
MYH7	c.3622G>A	p.Asp1208Asn	missense	0	2	VUS	PP3,PM2	
MYH7	c.3460G>A	p.Gly1154Ser	missense	0	1	VUS	PM2	

MYH7	c.2779G>A	p.Glu927Lys	missense	0	2	LP	PM2,PS4,P M1	
MYH7	c.2770G>A	p.Glu924Lys	missense	0	2	P	PP3,PP1(p), PM2,PS4,P S2,PM1	PMID:143 0197
MYH7	c.2680G>A	p.Glu894Lys	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.2669A>C	p.Gln890Pro	missense	0	1	VUS	PM2,PM1	
MYH7	c.2665C>T	p.Leu889Phe	missense	0	4	VUS	PM2,PM1	
MYH7	c.2609G>A	p.Arg870His	missense	0	4	P	PM2,PS4,P P1(s)	PMID:166 50083
MYH7	c.2548G>A	p.Ala850Thr	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.2539_2541del AAG	p.Lys847del	inframe indel	0	2	LP	PM2,PS4,P M4, PP1(p)	ClinVar SCV0000 59453.5
MYH7	c.2470G>A	p.Val824Ile	missense	0	1	VUS	PM2,PP3,P M1	
MYH7	c.2456G>A	p.Arg819Gln	missense	0	2	VUS	PM2,PP3,P M1	
MYH7	c.2389G>A	p.Ala797Thr	missense	2.10E-05	1	P	PM2,PS4,P P1(s)	PMID:111 86938
MYH7	c.2348G>A	p.Arg783His	missense	7.00E-06	1	LP	PM2 PS4,PM1	
MYH7	c.2221G>T	p.Gly741Trp	missense	0	1	P	PP3,PM2,P M5,PP1(m), PS4,PM1	PMID:853 3830
MYH7	c.2156G>A	p.Arg719Gln	missense	0	3	P	PP3,PM2,P M5,PM1,PS 4	
MYH7	c.2155C>T	p.Arg719Trp	missense	0	1	P	PP3,PM2,P M5,PP1(s),P S2,PS4	PMID:982 9907,828 2798,982 2100,129 74739, SHaRe consortiu

								m, PMID: 30297972
MYH7	c.2147G>C	p.Gly716Ala	missense	0	1	LP	PM2,PM5,P M1	
MYH7	c.2146G>A	p.Gly716Arg	missense	0	1	P	PP3,PM2,P M1,PM6,PS 4 ,PP1(s)	PMID:189 53637
MYH7	c.2068A>G	p.Met690Val	missense	0	2	VUS	PM2,PM1	
MYH7	c.2011C>T	p.Arg671Cys	missense	0	1	VUS	PM2,PP3,P M1	
MYH7	c.1988G>A	p.Arg663His	missense	3.00E-06	3	P	PM2,PS4,P M1, PP1(s)	PMID:107 50581
MYH7	c.1816G>A	p.Val606Met	missense	0	3	P	PP3,PM2,P M6, PS4,PP1(s), PM1	ClinVar:S CV00028 0305.1
MYH7	c.1580C>T	p.Pro527Leu	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.1548C>G	p.Asp516Glu	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.1493A>G	p.Gln498Arg	missense	0	2	VUS	PP3,PM2,P M1	
MYH7	c.1412A>G	p.Asn471Ser	missense	0	1	VUS	PM2,PM1	
MYH7	c.1208G>A	p.Arg403Gln	missense	0	1	P	PP3,PM2,P M5,PS4,PP 1(s),PM1	PMID:127 07239,PM ID:275322 57,ClinVa r SCV0006 92503.1 & SCV0000 59359.5, SHaRe consortiu m

MYH7	c.1182C>A	p.Asp394Glu	missense	0	1	VUS	PM2,PM1	
MYH7	c.1135G>A	p.Glu379Lys	missense	0	2	VUS	PP3,PM2,P M1	
MYH7	c.1064C>T	p.Ala355Val	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.1013T>A	p.Val338Glu	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.1012G>A	p.Val338Met	missense	0	2	LP	PP3,PM2,P P1(m),PM1	ClinVar: SCV0001 99244.4
MYH7	c.1003G>A	p.Ala335Thr	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.925G>A	p.Asp309Asn	missense	1.10E-05	5	VUS	PP3,PM2,P M1	
MYH7	c.902T>G	p.Leu301Arg	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.746G>A	p.Arg249Gln	missense	0	1	P	PM2,PP1(s), PS2,PM1	PMID:766 2452,Clin Var: SCV0007 35517.2
MYH7	c.734G>A	p.Gly245Glu	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.704C>G	p.Thr235Ser	missense	0	1	VUS	PP3,PM2,P M1	
MYH7	c.665A>G	p.Gln222Arg	missense	0	4	VUS	PP3,PM2,P M1	
MYH7	c.632C>T	p.Pro211Leu	missense	3.00E-06	1	VUS	PM2,PM1	
MYH7	c.427C>T	p.Arg143Trp	missense	1.90E-05	1	VUS	PP3,PM2,P M5	
MYL2	c.450G>T	p.Leu150Phe	missense	0	3	VUS	PM2	
MYL2	c.402+1G>T	-	splice donor variant	0	1	VUS	PM2	
MYL2	c.278C>T	p.Ala93Val	missense	0	5	VUS	PM2	
MYL2	c.275G>T	p.Gly92Val	missense	0	1	VUS	PP3,PM2,P P1(p)	ClinVar: SCV0001 99348.4

MYL2	c.173G>A	p.Arg58Gln	missense	0	2	LP	PM2,PS4,P P1(s)	
MYL2	c.113A>G	p.Gln38Arg	missense	0	1	VUS	PM2	
MYL3	c.518T>C	p.Met173Thr	missense	0	4	VUS	PM2	
MYL3	c.508G>C	p.Glu170Gln	missense	0	6	VUS	PP3,PM2	
MYL3	c.466G>C	p.Val156Leu	missense	0	1	VUS	PM2	
MYL3	c.460C>T	p.Arg154Cys	missense	2.10E-05	2	VUS	PP3,PM2	
MYL3	c.281G>A	p.Arg94His	missense	0	1	VUS	PM2,PP1(p)	ClinVar: SCV0000 59672.6
MYL3	c.187C>T	p.Arg63Cys	missense	3.00E-06	1	VUS	PP3,PM2	
PRKAG2	c.1475T>A	p.Ile492Asn	missense	2.20E-05	3	VUS	PM2	
PRKAG2	c.1390G>A	p.Asp464Asn	missense	3.40E-05	1	VUS	PM2	
PRKAG2	c.745G>C	p.Glu249Gln	missense	0	1	VUS	PM2,BP4	
PTPN11	c.172A>G	p.Asn58Asp	missense	0	1	VUS	PM2,PP3,P M6	
PTPN11	c.184T>G	p.Tyr62Asp	missense	0	1	VUS	PM2,PP3,P M6	
PTPN11	c.836A>G	p.Tyr279Cys	missense	0	2	VUS	PM2	
PTPN11	c.880G>A	p.Asp294Asn	missense	0	1	VUS	PM2	
PTPN11	c.922A>G	p.Asn308Asp	missense	0	1	P	PM2,PS2,P P1(s)	PMID: 11992261
PTPN11	c.1403C>T	p.Thr468Met	missense	0	3	VUS	PM2	
PTPN11	c.1492C>T	p.Arg498Trp	missense	0	1	LP	PM2,PP3,P S2	
PTPN11	c.1496C>G	p.Ser499Cys	missense	0	1	VUS	PM2	
PTPN11	c.1510A>G	p.Met504Val	missense	0	1	VUS	PM2	
RAF1	c.1913C>T	p.Thr638Met	missense	2.60E-05	1	VUS	PM2	
RAF1	c.1295A>G	p.His432Arg	missense	0	1	VUS	PM2	
RAF1	c.1274A>G	p.Glu425Gly	missense	0	1	VUS	PM2	
RAF1	c.782C>T	p.Pro261Leu	missense	0	1	VUS	PM2,PM6	
RAF1	c.781C>G	p.Pro261Ala	missense	0	2	VUS	PM2	
RAF1	c.770C>T	p.Ser257Leu	missense	0	2	VUS	PM2,PM6	
RAF1	c.766A>G	p.Arg256Gly	missense	0	1	VUS	PM2	
RAF1	c.733C>T	p.Pro245Ser	missense	0	1	VUS	PM2	

<i>RAF1</i>	c.601A>G	p.Ile201Val	missense	1.20E-05	4	VUS	PM2	
<i>TNNC1</i>	c.91G>A	p.Ala31Thr	missense	0	1	VUS	PM2	
<i>TNNI3</i>	c.511G>A	p.Ala171Thr	missense	0	1	VUS	PM2	
<i>TNNI3</i>	c.485G>T	p.Arg162Leu	missense	0	1	VUS	PM2,PM5	
<i>TNNI3</i>	c.470C>T	p.Ala157Val	missense	0	1	P	PM2,PS4,P P1(s)	ClinVar:S CV00005 9951.4
<i>TNNI3</i>	c.356C>A	p.Thr119Asn	missense	2.20E-05	1	VUS	BP4,PM2	
<i>TNNI3</i>	c.188G>A	p.Arg63Gln	missense	0	1	VUS	PM2	
<i>TNNT2</i>	c.833G>A	p.Arg278His	missense	1.20E-05	1	VUS	PM2	
<i>TNNT2</i>	c.807C>A	p.Asn269Lys	missense	0	2	VUS	PM2	
<i>TNNT2</i>	c.784A>G	p.Asn262Asp	missense	0	1	VUS	PM2	
<i>TNNT2</i>	c.274C>T	p.Arg92Trp	missense	0	1	P	PP3,PM2,P M5, PS4,PP1(s)	PMID:906 089,1134 6248
<i>TNNT2</i>	c.268A>G	p.Ile90Val	missense	0	1	VUS	PP3,PM2	
<i>TPM1</i>	c.123T>G	p.Asp41Glu	missense	0	1	VUS	PM2	
<i>TPM1</i>	c.355G>A	p.Ala119Thr	missense	0	1	VUS	PM2	
<i>TPM1</i>	c.574G>A	p.Glu192Lys	missense	0	2	P	PM2,PS4,P P1(s)	ClinVar:S CV00006 0282.4;P MID:2696 0954
<i>TPM1</i>	c.755G>A	p.Ser252Asn	missense	0	1	VUS	PM2	
<i>TPM1</i>	c.842T>C	p.Met281Thr	missense	0	1	VUS	PM2	

7.2 Permissions to re-use figures/tables from publications

License Number	4802380197572
License date	Apr 05, 2020
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
Title	Defining the genetic architecture of hypertrophic cardiomyopathy in Egypt
Institution name	Imperial College London & Aswan Heart Centre
Expected presentation date	Jun 2020
Portions	Figures 2 and 3
Requestor Location	Imperial College London south kensington London, SW7 2AZ United Kingdom Attn: Imperial College London

License Number	4680160045684
License date	Oct 01, 2019
Licensed Content Publisher	Springer Nature
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 et al.
Licensed Content Date	Mar 5, 2015
Licensed Content Volume	17
Licensed Content Issue	5
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Defining the genetic architecture of hypertrophic cardiomyopathy in Egypt
Institution name	Imperial College London & Aswan Heart Centre
Expected presentation date	Jun 2020
Portions	Figure 1 Table 5
Requestor Location	Imperial College London south kensington

License Number	4813580714202
License date	Apr 21, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology
Licensed Content Title	Structural Implications of β -Cardiac Myosin Heavy Chain Mutations in Human Disease
Licensed Content Author	Michelle Peckham, Melanie Colegrave
Licensed Content Date	Aug 14, 2014
Licensed Content Volume	297
Licensed Content Issue	9
Licensed Content Pages	11
Type of Use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	Defining the genetic architecture of hypertrophic cardiomyopathy in Egypt
Institution name	Imperial College London & Aswan Heart Centre
Expected presentation date	Jun 2020
Portions	Figure 2
Requestor Location	Imperial College London south kensington

London, SW7 2AZ
United Kingdom
Attn: Imperial College London

License Number	4816490608171
License date	Apr 26, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Molecular Cell Biology
Licensed Content Title	Quality and quantity control of gene expression by nonsense-mediated mRNA decay
Licensed Content Author	Tatsuaki Kurosaki et al
Licensed Content Date	Apr 16, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Defining the genetic architecture of hypertrophic cardiomyopathy in Egypt
Institution name	Imperial College London & Aswan Heart Centre
Expected presentation date	Jun 2020
Portions	Figures 1 and 6
Requestor Location	Imperial College London south kensington